

UNIVERSIDADE FEDERAL DE MINAS GERAIS  
PROGRAMA INTERUNIDADES DE PÓS-GRADUAÇÃO EM BIOINFORMÁTICA

ANA PAULA DE ABREU

**POTTER: uma abordagem para engenharia de peptídeos para  
inibição competitiva da proteína *Spike* do SARS-COV-2**

Belo Horizonte

2024

ANA PAULA DE ABREU

**POTTER: uma abordagem para engenharia de peptídeos para  
inibição competitiva da proteína *Spike* do SARS-COV-2**

Tese apresentada ao Programa de Pós-Graduação em  
Bioinformática da Universidade Federal de Minas  
Gerais, como requisito à obtenção do título de  
Doutora em Bioinformática.

Orientador: Adriano de Paula Sabino

Coorientadora: Raquel Cardoso de Melo Minardi

Belo Horizonte

2024

043

Abreu, Ana Paula de.

POTTER: uma abordagem para engenharia de peptídeos para inibição competitiva da proteína Spike do Sars-Cov-2 [manuscrito] / Ana Paula de Abreu. – 2024.

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

Orientador: Adriano de Paula Sabino. Coorientadora: Raquel Cardoso de Melo Minardi.

Tese (doutorado) – Universidade Federal de Minas Gerais, Instituto de Ciências Biológicas. Programa Interunidades de Pós-Graduação em Bioinformática.

1. Bioinformática. 2. COVID-19. 3. SARS-CoV-2. 4. Peptídeos. 5. Simulações de dinâmica molecular. I. Sabino, Adriano de Paula. II. Minardi, Raquel Cardoso de Melo. III. Universidade Federal de Minas Gerais. Instituto de Ciências Biológicas. IV. Título.

CDU: 573:004



UNIVERSIDADE FEDERAL DE MINAS GERAIS

## FOLHA DE APROVAÇÃO

**Ana Paula de Abreu**

**"POTTER: uma abordagem para engenharia de peptídeos para inibição competitiva da proteína Spike do SARS-COV-2"**

Tese aprovada pela banca examinadora constituída pelos Professores:

Prof. Adriano de Paula Sabino - Orientador  
Universidade Federal de Minas Gerais

Prof<sup>a</sup> Raquel Cardoso de Melo Minardi - Coorientadora  
Universidade Federal de Minas Gerais

Prof<sup>a</sup> Mariana Torquato Quezado de Magalhaes  
Universidade Federal de Minas Gerais

Prof. Vinicius Gonçalves Maltarollo  
Universidade Federal de Minas Gerais

Prof<sup>a</sup> Karina dos Santos Machado  
Universidade Federal do Rio Grande

Prof<sup>a</sup> Lucianna Helene Silva dos Santos  
Institut Pasteur de Montevideo

Belo Horizonte, 26 de junho de 2024.



Documento assinado eletronicamente por **Karina dos Santos Machado, Usuário Externo**, em 23/08/2024, às 10:03, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Adriano de Paula Sabino, Professor do Magistério Superior**, em 23/08/2024, às 11:03, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).

---



Documento assinado eletronicamente por **Lucianna Helene Silva dos Santos, Usuário Externo**, em 31/08/2024, às 17:13, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).

---



Documento assinado eletronicamente por **Mariana Torquato Quezado de Magalhaes, Professora do Magistério Superior**, em 06/09/2024, às 10:33, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).

---



Documento assinado eletronicamente por **Raquel Cardoso de Melo Minardi, Coordenador(a) de curso de pós-graduação**, em 11/09/2024, às 10:03, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).

---



Documento assinado eletronicamente por **Vinicius Goncalves Maltarollo, Professor do Magistério Superior**, em 08/10/2025, às 09:53, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).

---



A autenticidade deste documento pode ser conferida no site [https://sei.ufmg.br/sei/controlador\\_externo.php?acao=documento\\_conferir&id\\_orgao\\_acesso\\_externo=0](https://sei.ufmg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0), informando o código verificador **3297090** e o código CRC **B05269A2**.

---

## **Agradecimentos**

A Deus por sempre estar presente e provendo tudo o que foi necessário para a conclusão de mais uma etapa.

A minha família por permitir que meus sonhos fossem possíveis e por prover os meios para que sempre sejam concretizados.

A meu orientador, Professor Doutor Adriano de Paula Sabino por toda a paciência, empenho com que sempre me orientou neste trabalho, pelas boas conversas e pela amizade construída nessa caminhada.

A minha coorientadora, Professor Doutora Raquel Cardoso de Melo-Minardi pela disposição de sempre prover os recursos quando necessários. Muito obrigada por me ter corrigido quando necessário sem nunca me desmotivar.

Aos funcionários da Secretaria do Programa de Pós-Graduação em Bioinformática Tiago e Sheila, pelo empenho em suas profissões e zelo conosco alunos.

Aos meus amigos do Laboratório de Bioinformática e Sistemas Diego Mariano e Frederico Chaves por todos os momentos compartilhados.

Às agências de fomento à pesquisa: Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES e Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq); e à equipe do Laboratório de Bioinformática e Sistemas (LBS) e ao Laboratório de Oncohematologia e Transplantes.

## Resumo

O SARS-CoV-2 é o vírus responsável pela doença respiratória denominada COVID-19, que desencadeou uma grave crise de saúde pública global. Desde 2020 a comunidade científica tem empreendido um esforço significativo no desenvolvimento de agentes profiláticos e terapêuticos seguros e eficazes contra esta doença. Neste contexto, os peptídeos surgiram como alternativa para a inibição do agente causador. No entanto, projetar peptídeos que se liguem eficientemente ainda é um desafio em aberto. Neste trabalho, apresenta-se um algoritmo para engenharia de peptídeos. Nossa estratégia consiste em partir de um peptídeo cuja estrutura é semelhante à região de interação da proteína ACE2 humana com a proteína *Spike*, interação fundamental para a infecção por SARS-CoV-2. Nossa metodologia é baseada em um algoritmo genético que realiza etapas sistemáticas de mutação aleatória, docking proteína-peptídeo (usando a biblioteca PyRosetta) e seleção dos peptídeos mais otimizados com base nos contatos feitos entre a interface peptídeo-proteína. Realizam-se três estudos de caso para avaliar os parâmetros do algoritmo POTTER e os resultados foram comparados com propostas apresentadas na literatura, indicando que os peptídeos gerados pelo POTTER, como o peptídeo HO do Estudo de Caso 1, apresentaram uma energia livre de Gibbs ( $\Delta G$ ) mais favorável (-13.6 kcal/mol) em comparação com o peptídeo inicial (-9.8 kcal/mol).. Além disso, realizaram-se simulações de dinâmica molecular (MD) (três sistemas, 200ns cada) para investigar se nossos peptídeos sugeridos poderiam interagir de forma estável com a proteína *Spike*. Nossos resultados sugerem que nossa metodologia pode ser uma estratégia promissora para a engenharia de peptídeos.

**Palavras-chave:** COVID-19; SARS-CoV-2; peptídeos; *docking*; interações proteína-peptídeo; POTTER; simulações de dinâmica molecular.

## ***Abstract***

*SARS-CoV-2 is the virus responsible for the respiratory disease known as COVID-19, which has triggered a serious global public health crisis. Since 2020, the scientific community has undertaken a significant effort in developing safe and effective prophylactic and therapeutic agents against this disease. In this context, peptides have emerged as an alternative for inhibiting the causative agent. However, designing peptides that bind efficiently remains an open challenge. This work presents an algorithm for peptide engineering. Our strategy consists of starting with a peptide whose structure is similar to the interaction region of the human ACE2 protein with the Spike protein, a fundamental interaction for SARS-CoV-2 infection. Our methodology is based on a genetic algorithm that performs systematic steps of random mutation, protein-peptide docking (using the PyRosetta library), and selection of the most optimized peptides based on the contacts made between the peptide-protein interface. Three case studies were conducted to evaluate the parameters of the POTTER algorithm, and the results were compared with proposals presented in the literature, indicating that the peptides generated by POTTER, such as the HO peptide from Case Study 1, presented a more favorable Gibbs free energy ( $\Delta G$ ) (-13.6 kcal/mol) compared to the initial peptide (-9.8 kcal/mol). Furthermore, molecular dynamics (MD) simulations (three systems, 200ns each) were performed to investigate whether our suggested peptides could interact stably with the Spike protein. Our results suggest that our methodology may be a promising strategy for peptide engineering.*

**Keywords:** *COVID-19; SARS-CoV-2; peptides; docking; protein-peptide interactions; POTTER; molecular dynamics simulations.*

## Lista de Figuras

Figura 1. Níveis estruturais de representação de estruturas de proteínas: estruturas primária, secundária, terciária e quaternária de proteínas. Fonte: Diego Mariano   Wikimedia Commons (CC-BY 4.0) .....	18
Figura 2 <b>A.</b> Estrutura comum a todos os $\alpha$ -aminoácidos, exceto um (à prolina aminoácido cíclico). O grupo R ou cadeia lateral ligado ao carbono $\alpha$ é diferente em cada aminoácido. <b>B.</b> Ligações peptídicas entre dois aminoácidos. Após a ligação peptídica, uma molécula de água é liberada. Cada círculo representa um átomo (H: hidrogênio; C: carbono; O: oxigênio; N; nitrogênio). O quadrado com R indica a cadeia lateral e pode variar em um dos 20 tipos de aminoácidos Formação de uma ligação peptídica (pontilhado vermelho) por condensação. A ligação encontra-se entre o grupo amino (sombreado em amarelo) e o grupo carboxila (sombreado em azul), e libera água (H <sub>2</sub> O) durante seu processo de formação. ....	19
Figura 3. Interação do tipo Ligação de hidrogênio entre uma serina e uma valina. Fonte:[27] (PAIXÃO et al., 2023).....	23
Figura 4.Interações aromáticas também conhecidos como pi-stacking ou empilhamento $\pi$ - $\pi$ . Empilhamento em sanduíche (esquerda), forma de T (centro) e paralelamente deslocado (direita). Fonte: adaptado de Wikimedia Commons / Emily ricq (CC BY-3.0). Disponível em <a href="https://commons.wikimedia.org/wiki/File:Benzene">https://commons.wikimedia.org/wiki/File:Benzene</a> .....	24
Figura 5. Interações aromática entre aminoácidos tirosina com anéis aromáticos, formato sanduiche. Fonte: Diego Mariano   Wikimedia Commons (CC-BY 4.0). Disponível: < <a href="https://creativecommons.org/licenses/by-sa/4.0/">https://creativecommons.org/licenses/by-sa/4.0/</a> >, via Wikimedia Commons .....	25
Figura 6. Taxonomia de Coronaviridae. HCoV, coronavírus humano; MERS-CoV, coronavírus da síndrome respiratória do Oriente Médio; SARS-CoV, coronavírus da síndrome respiratória aguda grave; SARS-CoV-2, síndrome respiratória aguda grave coronavírus-2; SW1, coronavírus da baleia beluga; e BuCoV-HKU11, coronavírus bulbul HKU11.Fonte:[7].....	27
Figura 7. Representação do vírus SARS-CoV-2 com destaque para a proteína Spike (S) que está envolvida no processo de ligação ao receptor na célula hospedeira. A proteína S do vírus SARS-CoV-2 liga-se ao receptor ACE2 presente na membrana celular da célula hospedeira. O corte pela enzima Furina é um passo necessário para ativar a proteína S e permitir a fusão das membranas. O processo de fusão é detalhado, mostrando as subunidades S1 e S2 da proteína S e os objetivos de transição da fusão. TM (domínio transmembranar), HR1 (domínios de	

repetição do heptad 1), HR2 (domínios de repetição do heptad 2), e FP (peptídeo de fusão) são componentes chave nesse processo. A fusão das membranas virais e celulares culmina na entrada do material genético viral na célula hospedeira, permitindo a infecção. A membrana viral é representada em amarelo claro e a membrana celular em amarelo escuro. Fonte: [10] 28

Figura 8. Representação esquemática do genoma e domínio funcional da proteína S do SARS-CoV-2. A. O SARS-CoV-2 é um vírus de RNA conhecido com uma estrutura de 5'-cap e 3'-poli-A e possui 14 supostos quadros de leitura abertos (ORFs) que codificam 27 proteínas. O vírion possui quatro proteínas estruturais, conhecidas como proteínas S, Spike; E, envelope; M, membrana; e N, nucleocapsídeo. B. As proteínas Spike contêm subunidades S1 e S2 e os locais de clivagem para Furina e TMPRSS2. As proteínas Spike medeiam a ligação viral às células hospedeiras após ativação pela enzima TMPRSS2. SP, peptídeo sinal; NTD, domínio N-terminal; RBD, domínio de ligação ao receptor, contém RBM, motivo de ligação central no subdomínio externo; FP, peptídeo de fusão; HR1, domínios de repetição do heptad 1 e heptad 2 HR2; TM, domínio transmembrana; CP, domínio do citoplasma. ....29

Figura 9. Entrada do SARS-CoV-2 na célula: interação entre a proteína S e a membrana ACE2. A proteína S do SARS-CoV-2 possui duas unidades funcionais, S1 e S2. A S1 se liga ao receptor ACE2, resultando na exposição do sítio de clivagem S1/S2. A proteína é então clivada pela furina e pela TMPRSS2, o que provoca mudanças conformacionais que permitem a fusão da membrana, facilitando a inserção do peptídeo de fusão S2. Isso leva à liberação do RNA viral, que após ser internalizado, resulta na replicação e tradução do vírus. Novos vírus são montados e liberados da célula, iniciando um novo ciclo de infecção. Fonte: [34].....30

Figura 10. Estrutura geral do RBD do SARS-CoV-2 ligado ao ACE2 **a**, Topologia geral do monômero Spike do SARS-CoV-2. FP, peptídeo de fusão; HR1, repetição heptad 1; HR2, repetição heptad 2; IC, domínio intracelular; NTD, domínio N-terminal; SD1, subdomínio 1; SD2, subdomínio 2; TM, região transmembrana. **b**, Sequência e estruturas secundárias do RBD do SARS-CoV-2. A sequência RBM é mostrada em vermelho. **c**, Estrutura geral do RBD do SARS-CoV-2 ligado ao ACE2. O ACE2 é mostrado em verde. O núcleo do RBD do SARS-CoV-2 é mostrado em ciano e o RBM em vermelho. As ligações dissulfeto no RBD do SARS-CoV-2 são mostradas como bastões e indicadas por setas. A hélice N-terminal do ACE2 responsável pela ligação é rotulada. Fonte: [2] .....32

Figura 11. Docking molecular, também conhecido como encaixe molecular ou ancoragem. Fonte: Diego Mariano, CC BY-SA 4.0. Disponível: <<https://creativecommons.org/licenses/by-sa/4.0/>>, via Wikimedia Commons .....38

Figura 12. Fluxograma da metodologia POTTER. Onde Gerações (G) denota quantas vezes o processo se repete. Mutantes (M) o número de mutantes gerados aleatoriamente a cada geração, e Repetições (R): número de <i>dockings</i> repetidos para cada mutante. ....	44
Figura 13. Região de interação na interface SARS-CoV-2 RBD-ACE2. Onde os resíduos os K417, G446, Y449, Y453, L455, F456, A475, G476, S477, F486, N487, Y489, F490, Q493, S494, Y495, G496, Q498, T500, N501, G502, V503 ou Y505 da ACE2(cor salmão) utilizados como ponto de partida para propor novos peptídeos que interagem com a proteína S (cor laranja).....	46
Figura 14. Interações de interface entre ACE2 (desenho ciano) e o receptor Spike (desenho verde). Os resíduos do ACE2 mais próximos do Spike são mostrados em bastões laranja (PDB ID: 6M0J). A figura foi gerada usando PyMOL de código aberto (Schrödinger, LLC).....	49
Figura 15. Visão geral do algoritmo POTTER usando os parâmetros: G = 5, M = 6 e R = 3. Em laranja temos a proteína Spike acoplado aos peptídeos mutantes em verde, ciano e vermelho. Em roxo temos o peptídeo otimizado selecionado. ....	52
Figura 16. Frame 0 (0 ns), 5000 (100 ns) e 9999 (200ns) das três simulações MD: Spike complexado com o modelo peptídico (topo); Spike complexada com o peptídeo LDS (meio); e Spike complexado com o peptídeo HO (parte inferior). ....	58
Figura 17. As interações do tipo ligações de hidrogênio são realizadas entre o peptídeo-modelo e o receptor, bem como o peptídeo curado manualmente e o receptor. As interações não foram representadas em escala real. Figura gerada usando ChimeraX.....	60
Figura 18. Perfil de interações estudo de caso 1. Podemos observar no primeiro quadrante A o peptídeo modelo (inicial) onde a ocupação é de 55% e um total de 27 resíduos interagindo entre Proteína Spike e o peptídeo inicial. O peptídeo por MC apresenta uma ocupação de 41%. No modelo LDS apresenta uma ocupação de 36%. No modelo HO apresenta uma ocupação de 71%.....	62
Figura 19. Perfil de interações estudo de caso 2. Podemos observar no primeiro quadrante B o peptídeo modelo (inicial) onde a ocupação é de 47% e um total de 27 resíduos interagindo entre Proteína Spike e o peptídeo inicial. Já o peptídeo MC apresenta uma ocupação de 37%. No modelo LDS apresenta uma ocupação de 40%. No modelo HO apresenta uma ocupação de 68%.....	63
Figura 20. Perfil de interações estudo de caso 3. Podemos observar no primeiro quadrante C o peptídeo modelo (inicial) onde a ocupação é de 42% e um total de 27 resíduos interagindo entre Proteína Spike e o peptídeo inicial. Já o peptídeo MC apresenta uma ocupação de 33%. No	

modelo LDS apresenta uma ocupação de 30%. No modelo HO apresenta uma ocupação de 67%.....	64
Figura 21. Seleção dos melhores peptídeos por gerações para os estudos de caso 1, 2 e 3 de acordo com a maior ocupação (HO). A geração em que cada melhor peptídeo é obtido está destacada por uma seta vermelha. ....	69
Figura 22. Seleção dos melhores peptídeos por gerações para os estudos de caso 1, 2 e 3 de acordo com a menor docking score (LDS). A geração em que cada melhor peptídeo é obtido está destacada por uma seta vermelha. ....	70
Figura 23. Gráfico de linhas 2D-RMSD, RMDS e RMSF para as três simulações MD: pico complexado com o modelo peptídico (A, D e G, respectivamente); Spike complexada com o peptídeo LDS (B, E e H, respectivamente); e Spike complexado com o peptídeo HO (C, F e I, respectivamente).....	72

## Lista de Tabelas

Tabela 1. Aminoácidos mais comumente encontrados nos seres vivos e suas abreviações com 3 e 1 letra. IUPAC[20].	20
Tabela 2. Classificação de cada átomo de cada tipo de aminoácido de acordo com o tipo de contato que podem realizar. Fonte: [24].	22
Tabela 3. Energia de contato. Fonte: adaptado de [22].	26
Tabela 4. Substituições permitidas pelo algoritmo POTTER. Fonte: adaptado de [46].	50
Tabela 5. Dados para os estudos de caso.	53
Tabela 6. Peptídeos selecionados por curadoria manual, menor pontuação de docking e maior ocupação versus peptídeo inicial.	56
Tabela 7. Peptídeos selecionados por curadoria manual, menor pontuação de docking e maior ocupação versus peptídeo inicial, onde demonstramos o perfil de interações de cada um selecionado. Os valores da constante de dissociação ( $K_d$ ) indica que a proteína tem uma alta afinidade pelo ligante, ou seja, o complexo proteína-ligante é estável e se forma facilmente.	65

## Lista de Abreviações

**ACE2:** *Angiotensin Converting Enzyme II* (Enzima Conversora de Angiotensina II)

**CLD:** *Collectrin-Like Domain* (Colectrina)

**CoVs:** *coronavirus* (coronavírus)

**CP:** *Cytoplasmic Domain* (Domínio Citoplasmático)

**CSV:** *Comma-Separated Values*

**FDA:** *Food and Drug Administration*

**FP:** *Fusion Peptide* (Peptídeo de Fusão)

**GISAID:** *Global Initiative on Sharing All Influenza Data* (Iniciativa Global Sobre Compartilhamento de Todos os Dados da Gripe)

**HR1:** *heptal 1*

**HR2:** *heptal 2*

**LDS:** *Lowest Docking Score* (Menor Docking Score)

**MC:** *Manual Curation* (Curadoria Manual)

**MD:** *Molecular Dynamics* (Dinâmica Molecular)

**MERS-CoV:** *Middle East respiratory syndrome coronavirus* (Síndrome Respiratória Médio Oriente dos Coronavírus)

**mRNA:** *Ácido Ribonucleico mensageiro*

**NAMD:** *Nanoscale Molecular Dynamics*

**OH:** *Higher Occupancy* (Maior Ocupação)

**OMS:** *Organização Mundial da Saúde*

**PBC:** *Periodic boundary conditions* (Condições Periódicas de Contorno)

**PDB:** *Protein Data Bank*

**POTTER:** *Peptide optimization tool to enhance receptor-binding*

**RBD:** *Receptor Binding Domain* (Domínio de Ligação ao Receptor)

**RMSD:** *Root Mean Square Deviation*

**RNA:** *Ácido Ribonucleico*

**SARS-CoV:** *Severe acute respiratory syndrome coronavirus* (Síndrome Respiratória Aguda Grave do Coronavírus)

**SARS-CoV-2:** *Severe acute respiratory syndrome coronavirus 2* (Síndrome Respiratória Aguda Grave do Coronavírus 2)

**TM:** *Transmembrane Domain* (Domínio Transmembranar)

**TMPRSS2:** *Transmembrane Serine Protease 2* (Protease Transmembrana Serina Protease 2)

**VMD:** *Visual Molecular Dynamics*

**+ssRNA:** *positive-sense single-stranded RNA viruses*

# Sumário

1. Introdução.....	16
1.1 Proposta.....	17
2. Fundamentação Teórica.....	18
2.1 Proteínas, Peptídeos e aminoácidos.....	18
2.2 Interações intermoleculares.....	21
2.3 O coronavírus SARS-CoV-2 e suas Proteínas Estruturais.....	26
2.4 A Proteína <i>Spike</i> (S) como alvo Principal.....	28
2.5 A Enzima Conversora de Angiotensina 2 (ACE2) como Receptor Viral.....	30
2.6 As Interações Moleculares na interface RBD-ACE2.....	31
2.7 Exemplos de alvos atuais contra SARS-CoV-2.....	33
2.8 O uso de subsequências conservadas para o planejamento de Peptídeos.....	34
2.9 <i>Docking</i> Molecular.....	36
2.10 Dinâmica Molecular.....	38
2.11 Trabalhos Relacionados.....	40
3. Objetivos.....	42
3.1 Objetivo geral.....	42
3.2 Objetivos específicos.....	42
4. Materiais e Métodos.....	43
4.1 Coleta de dados.....	45
4.2 Modelagem molecular e <i>docking</i> .....	46
4.3 Análise de Contatos e Curadoria Manual.....	47
4.4 Implementação do algoritmo.....	48
4.5 Visão Geral do Algoritmo.....	48
4.6 Demonstração do algoritmo.....	51
4.7 Estudo de caso.....	53

4.8 Validação por Dinâmica Molecular.....	53
5. Resultados e discussão .....	55
5.1 Simulações de dinâmica molecular .....	57
5.2 A importância dos complexos proteína-peptídeo .....	59
5.3 Comparação com outras propostas de engenharia de peptídeos descritas na literatura ..	67
5.4 Importância dos parâmetros POTTER .....	68
5.5 Insights obtidos pelas simulações MD .....	72
6. Conclusão .....	74
7. Referências Bibliográficas.....	76
8. Apêndices .....	84
Apêndice 8.1 .....	84
Apêndice 8.2.....	84
Apêndice 8.3.....	84

## 1. Introdução

No início de 2020, um surto do vírus SARS-CoV-2 (Síndrome Respiratória Aguda Grave do Coronavírus 2), responsável por uma doença respiratória chamada COVID-19, devastou a saúde pública global. Até o final de 2023, mais de 6,9 milhões de mortes e 771 milhões de casos foram registrados em todo o mundo (OMS) [1]. A sua rápida propagação internacional exigiu o desenvolvimento de agentes profiláticos e terapêuticos seguros e eficazes contra a infecção do seu agente causador. O primeiro passo foi compreender o mecanismo da infecção viral para desenvolver estratégias de intervenção contra o novo coronavírus [2].

O SARS-CoV-2 é um coronavírus pertencente à família *Coronaviridae*. Seu genoma compreende uma molécula de RNA de fita simples (+ssRNA) contendo aproximadamente 30 kb [3,4]. Os principais genes em seu genoma incluem aqueles que codificam proteínas estruturais, como a proteína *Spike* (S), bem como proteínas de envelope (E), membrana (M) e nucleocapsídeo (N) [5,6]. Em destaque está a proteína S, uma glicoproteína transmembranar que contém a subunidade S1 e a subunidade S2 e que se liga a uma protease transmembrana Serina Protease 2 (TMPRSS2), presente na membrana do hospedeiro [7,8]. Tal como acontece com outras proteínas de fusão de membrana de classe I, a proteína *Spike* é clivada pós-tradução, neste caso pela ação da furina, nos componentes S1 e S2 que permanecem associados após a clivagem. A subunidade S1, que contém o domínio de ligação ao receptor (RBD), capaz de se dobrar funcionalmente de forma independente, pode se ligar à enzima conversora de angiotensina II (ACE2) [9,10]. A subunidade S2, composta pelos domínios de repetição do peptídeo de fusão (FP), *heptal* 1 (HR1) e *heptal* 2 (HR2), domínio transmembranar (TM) e domínio citoplasmático (CP), é responsável por mediar a fusão e a entrada viral [8,10].

A glicoproteína *Spike* do SARS-CoV-2 liga-se ao homodímero ACE2, o que facilita a entrada do vírus nas células hospedeiras, formando a interação SARS-CoV-2-RBD/ACE2 [11,12]. A ACE2 é uma enzima metalocarboxipeptidase transmembrana tipo I de passagem única responsável pela maturação do hormônio peptídico angiotensina (Ang), que por sua vez regula a vasoconstrição e a pressão arterial. Está presente nas células epiteliais alveolares tipo II e também é expresso em vários tecidos extrapulmonares, incluindo coração, rim e intestino [2]. Sua estrutura é formada por dois domínios principais - o PD N-terminal e o domínio semelhante à coletrina (CLD) na extremidade C-terminal. Após se ligar ao receptor da enzima ACE2, a partícula viral usa receptores e endossomos da célula hospedeira para entrar nas células. A protease hospedeira, TMPRSS2, facilita a entrada na célula através da proteína S.

Uma vez dentro da célula, são sintetizadas poliproteínas virais que codificam o complexo replicase-transcriptase. O vírus, por sua vez, sintetiza RNA através de sua RNA polimerase dependente de RNA. Proteínas estruturais são sintetizadas, levando à conclusão da montagem e liberação das partículas virais [12,13]. Portanto, a glicoproteína *Spike* desempenha as funções de ligação, fusão e entrada viral e pode ser utilizada como alvo para desenvolvimento de anticorpos, inibidores de entrada e vacinas.

Neste contexto, os peptídeos surgiram como alternativa para inibição desse agente causador. Peptídeos são moléculas compostas por resíduos de aminoácidos em sequência de 2 a 50 monômeros, ligados por ligações peptídicas [14,15]. Eles desempenham papéis cruciais nos seres vivos, incluindo sinalização celular e modulação imunológica [16,17]. Na terapêutica, os peptídeos apresentam diversas características promissoras em comparação com pequenas moléculas e proteínas terapêuticas, incluindo alta compatibilidade estrutural com proteínas alvo e a capacidade de romper interfaces proteína-proteína. Eles também são altamente seletivos e eficazes, ao mesmo tempo que permanecem relativamente seguros e bem tolerados. Por esses motivos, eles têm sido utilizados em estudos de vacinas baseadas em peptídeos com diversas aplicações, como câncer, HIV e números menores para doenças virais infecciosas [10,17]. Portanto, poderiam ser uma alternativa para propor vacinas ou medicamentos para mitigar pandemias globais, como a recente pandemia de COVID-19. No entanto, o planejamento racional de peptídeos que se liguem eficientemente ainda é um desafio em aberto.

## 1.1 Proposta

Neste estudo, é apresentada uma abordagem para a engenharia de peptídeos com potencial para aplicações farmacológicas, partindo-se de um alvo previamente definido. A essa abordagem, foi dado o nome POTTER (*Peptide optimization tool for enhanced receptor-binding*). O objetivo principal deste trabalho é detectar peptídeos que possam interagir com a glicoproteína *Spike* (S), impedindo-a de se ligar à enzima conversora de angiotensina humana 2 (ACE2). Sabe-se que as interações mediadas por peptídeos estão ganhando cada vez mais atenção devido aos seus papéis predominantes nos diversos processos regulatórios que envolvem interações dinâmicas entre proteínas. As estruturas de tais interações fornecem um excelente ponto de partida para sua caracterização e manipulação e podem fornecer pistas para o projeto de inibidores direcionados.

## 2. Fundamentação Teórica

### 2.1 Proteínas, Peptídeos e aminoácidos

As proteínas controlam praticamente todos os processos que ocorrem em uma célula, exibindo uma quase infinita diversidade de funções [18]. As proteínas são macromoléculas em que suas unidades monoméricas são resíduos de aminoácidos [18]. Elas são diferenciadas por sua sequência de aminoácidos e por sua estrutura tridimensional, sendo essa última o determinante das funcionalidades da célula.

As proteínas têm a capacidade de adotar quatro diferentes níveis estruturais distintos (Figura 1). A estrutura primária consiste em uma sequência linear de aminoácidos. Em contraste, a estrutura secundária refere-se à organização local da proteína e é predominantemente composta por hélices  $\alpha$  e folhas  $\beta$ . Na hélice  $\alpha$ , a estrutura se torce, assumindo a forma de um bastão condensado, onde ocorre a interação por meio de ligações de hidrogênio entre o grupo CO de cada resíduo e o grupo NH, localizado quatro resíduos adiante na sequência de aminoácidos. As folhas  $\beta$  são formadas por duas ou mais cadeias polipeptídicas, chamadas de fitas  $\beta$ . Ao contrário das hélices  $\alpha$ , as fitas  $\beta$  apresentam estruturas praticamente estendidas, onde a ligação entre duas ou mais fitas  $\beta$  ocorre por meio de ligações de hidrogênio entre os grupos CO e NH de cada fita [18,19]. A estrutura terciária é definida como o arranjo tridimensional da proteína. Por fim, a estrutura quaternária é formada pela união de duas ou mais estruturas terciárias.

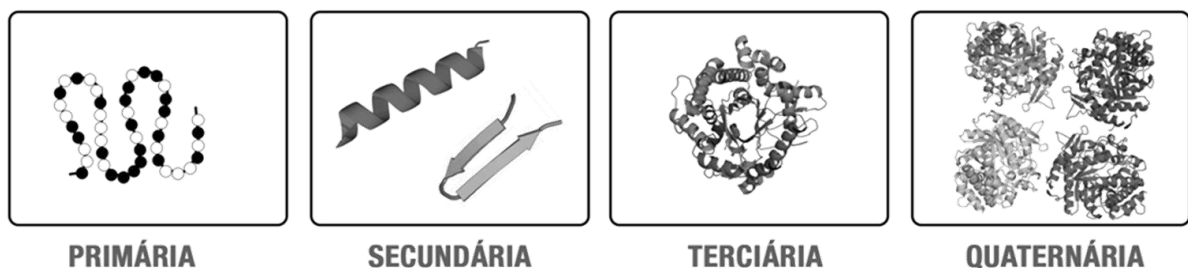


Figura 1. Níveis estruturais de representação de estruturas de proteínas: estruturas primária, secundária, terciária e quaternária de proteínas. Fonte: Diego Mariano | Wikimedia Commons (CC-BY 4.0)

As cadeias polipeptídicas são aminoácidos organizados de forma sequencial e linear (Figura 2). Cada resíduo de aminoácido é unido ao próximo por um tipo específico de ligação covalente. O termo resíduo refere-se à perda de uma molécula de água quando o grupo  $\alpha$ -amino

de um aminoácido atua como nucleófilo para deslocar o grupo hidroxila de outro aminoácido, formando assim a ligação peptídica [18].

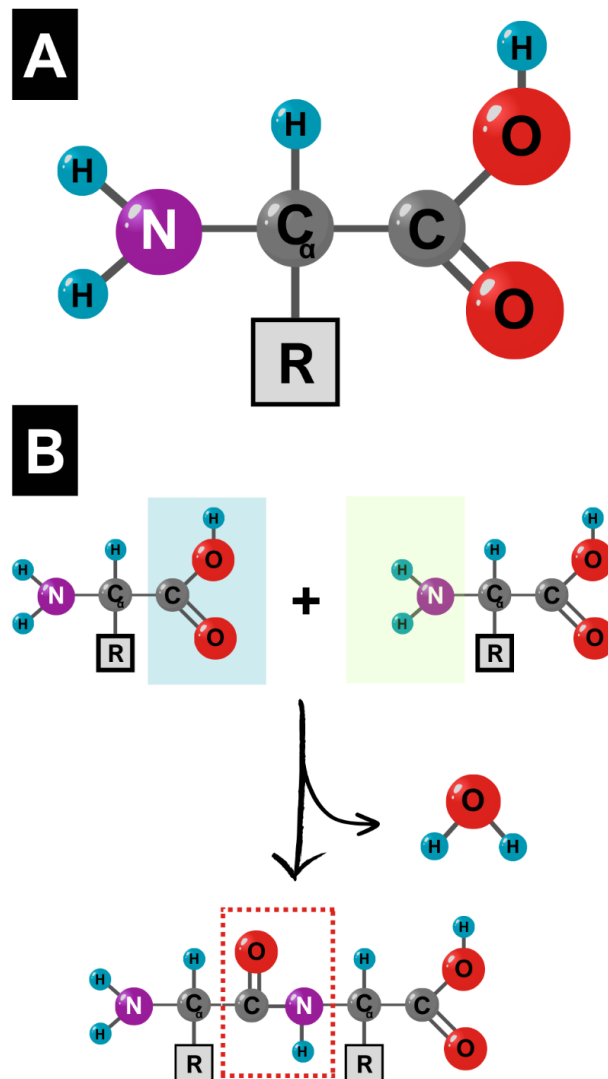


Figura 2 **A**. Estrutura comum a todos os  $\alpha$ -aminoácidos, exceto um (à prolina aminoácido cíclico). O grupo R ou cadeia lateral ligado ao carbono  $\alpha$  é diferente em cada aminoácido. **B**. Ligações peptídicas entre dois aminoácidos. Após a ligação peptídica, uma molécula de água é liberada. Cada círculo representa um átomo (H: hidrogênio; C: carbono; O: oxigênio; N: nitrogênio). O quadrado com R indica a cadeia lateral e pode variar em um dos 20 tipos de aminoácidos. Formação de uma ligação peptídica (pontilhado vermelho) por condensação. A ligação encontra-se entre o grupo amino (sombreado em amarelo) e o grupo carboxila (sombreado em azul), e libera água ( $H_2O$ ) durante seu processo de formação.

Todos os 20 aminoácidos comuns são  $\alpha$ -aminoácidos que possuem uma estrutura básica comum, denominada cadeia principal [19]. Ao mesmo carbono  $\alpha$  está ligado um grupo carboxila, um grupo amino e uma cadeia lateral. O que possibilita as características e propriedades químicas específicas de cada aminoácido e a sua classificação de acordo com suas

propriedades físico-químicas, tais como, cargas, hidrofobicidade, polaridade, etc. [19]. Na Tabela 1 são apresentados os aminoácidos mais comuns aos seres vivos.

Tabela 1. Aminoácidos mais comumente encontrados nos seres vivos e suas abreviações com 3 e 1 letra. IUPAC[20].

<b>Aminoácidos</b>	<b>Símbolo de 3 letras</b>	<b>Abreviação</b>
Alanina	ALA	A
Arginina	ARG	R
Asparagina	ASN	N
Aspartato (Ácido aspártico)	ASP	D
Cisteína	CYS	C
Fenilalanina	PHE	F
Glicina	GLY	G
Glutamato (Ácido glutâmico)	GLU	E
Glutamina	GLN	Q
Histidina	HIS	H
Isoleucina	ILE	I
Lisina	LYS	K
Leucina	LEU	L
Metionina	MET	M
Prolina	PRO	P
Serina	SER	S
Treonina	THR	T
Valina	VAL	V
Triptofano	TRP	W
Tirosina	TYR	Y

Peptídeos e proteínas possuem um grande potencial como agentes terapêuticos devido à sua marcante especificidade, alta eficácia, capacidade direcionada e menor toxicidade em comparação com os medicamentos químicos convencionais. Os peptídeos são estruturas polipeptídicas de 2-50 aminoácidos [19]. Nos últimos anos, o mercado de medicamentos peptídicos e proteicos obteve um crescimento exponencial, com uma taxa anual de expansão de 20%, superando significativamente a taxa de crescimento anual do mercado farmacêutico em

geral, estimada em 9%. Estima-se que o mercado alcance a marca de US\$ 388 bilhões até 2024 [21].

## 2.2 Interações intermoleculares

A complementaridade geométrica e química é responsável pela especificidade no reconhecimento molecular, enquanto as interações entre resíduos distintos são frequentemente mediadas por pares específicos de átomos [22]. Diferentes tipos de interações não covalentes estabilizam as estruturas proteicas e as interações entre moléculas, cujas energias de contato podem ser vistas na Tabela 3. As ligações de hidrogênio exemplificam a ligação entre átomos doadores e receptores de hidrogênio. Por contraste, as interações hidrofóbicas são específicas para átomos classificados como hidrofóbicos, onde moléculas apolares se agrupam para minimizar o contato com a água (REF FASSIO, 2015). Em resíduos de aminoácidos, esse efeito é especialmente observado em grupos de átomos de carbono. Diz-se que um contato hidrofóbico é observado quando um par de átomos de diferentes resíduos, que sejam classificados como hidrofóbicos, estão a uma distância de até 4 Å [23]. O efeito hidrofóbico é especialmente encontrado em um outro tipo de interação existente entre aminoácidos com anéis aromáticos. As interações aromáticas, também conhecidas como *pi-stacking* ou empilhamento  $\pi$ - $\pi$ , se restringem aos átomos caracterizados como aromáticos, com anéis aromáticos que apresentam ligação dupla e ressonância [24]. Existem três arranjos principais para essa interação: em formato de sanduíche, em formato de T e paralelamente deslocado (Figura 4). Por fim, as interações iônicas ocorrem entre átomos com cargas positivas e negativas, enquanto as interações repulsivas se limitam a átomos com cargas semelhantes, em uma distância específica. A **Erro! Fonte de referência não encontrada.** apresenta a classificação de cada átomo c onforme o tipo de contato que podem realizar.

Tabela 2. Classificação de cada átomo de cada tipo de aminoácido de acordo com o tipo de contato que podem realizar. Fonte: [24].

<b>Tipo de átomo</b>	<b>Átomos classificados em um certo tipo</b>
Aceptores	ALA (O), ARG (O), ASN (O, OD1), ASP (O, OD1, OD2), CYS(O), GLN (O, OE1), GLU (O, OE, OE2), GLY (O), HIS (O), ILE(O), LEU (O), LYS (O), MET (O), PHE (O), PRO (O), SER (O), THR (O), TRP (O), TYR (O), VAL (O)
Aromáticos	HIS (CD2, CE1, CG, ND1, NE2), PHE (CD1, CD2, CE1, CE2, CG, CZ), TRP (CD1, CD2, CE2, CE3, CG, CH2, CZ2, CZ3, NE1), TYR (CD1, CD2, CE1, CE2, CG, CZ)
Doadores	ALA (N), ARG (N, NE, NH1, NH2), ASN (N, ND2, OD1), ASP(N), CYS (N), GLN (N), GLU (N), GLY (N), HIS (N, ND1, NE2), ILE (N), LEU (N), LYS (N, NZ), MET (N), PHE (N), PRO (N), SER (N, OG), THR (N, OG1), TRP (N, NE1), TYR (N, OH), VAL(N)
Hidrofóbicos	ALA (CB), ARG (CB, CD, CG, CZ), ASN (CB, CG), ASP (CB,CG), CYS (CB), GLN (CB, CD, CG), GLU (CB, CD, CG), HIS(CB, CD2, CE1, CG), ILE (CB, CD1, CG1, CG2), LEU (CB, CD1,CD2, CG), LYS (CB, CD, CE, CG), MET (CB, CE, CG, SD), PHE(CB, CD1, CD2, CE1, CE2, CG, CZ), PRO (CB, CD, CG), SER(CB), THR (CB, CG2), TRP (CB, CD1, CD2, CE2, CE3, CG,CH2, CZ2, CZ3), TYR (CB, CD1, CD2, CE1, CE2, CG, CZ), VAL(CB, CG1, CG2)
Negativos	ASP (OD1, OD2), GLU (OE1, OE2)
Positivos	ARG (NH1, NH2), HIS (ND1, NE2) e LYS (NZ)

A função de uma proteína é majoritariamente definida por sua conformação [18]. As ligações covalentes são estabelecidas através do compartilhamento de elétrons entre pares de átomos [19]. As interações que ocorrem entre os átomos na cadeia principal de um aminoácido são classificadas como ligações covalentes, as quais possuem uma estabilidade intrínseca. No entanto, as interações que conferem a estrutura tridimensional da proteína são geralmente de natureza mais fraca. Enquanto uma interação covalente pode ter uma energia de ligação em torno de 1000 KJ/mol, uma ligação não covalente tende a ser cerca de 40 KJ/mol [18]. As interações não covalentes são principalmente determinadas pelos átomos presentes nas cadeias

laterais dos aminoácidos. O contato entre os resíduos na estrutura proteica também é conhecido como interação [25].

As interações de van der Waals, por exemplo, derivam da distribuição assimétrica da carga eletrônica em um átomo. Os elétrons se localizam em seus orbitais moleculares do núcleo criando dipolos momentâneos, induzindo cargas opostas em outros átomos e resultando em atração até uma distância mínima determinada pelos raios de van der Waals de cada átomo [25].

As interações do tipo ligações de hidrogênio constituem um tipo crucial de interação, ocorrendo tanto intra quanto intermolecularmente entre um átomo doador de ligação de hidrogênio e um átomo aceptor. Tipicamente, essas ligações envolvem átomos altamente eletronegativos, como oxigênio e nitrogênio, sendo estes frequentemente presentes em resíduos de aminoácidos (Figura 3). Nesse arranjo, um átomo de hidrogênio ligado a um átomo eletronegativo (conhecido como doador) é atraído por outro átomo ainda mais eletronegativo (denominado aceptor). Estudos na literatura têm consistentemente destacado que a distância euclidiana e ângulos entre átomos são métricas eficazes para identificar a presença de ligações de hidrogênio [26]. Em geral, considera-se que uma ligação de hidrogênio ocorre quando os átomos aceptor e doador estão separados por uma distância de até 3,9 Å [24].

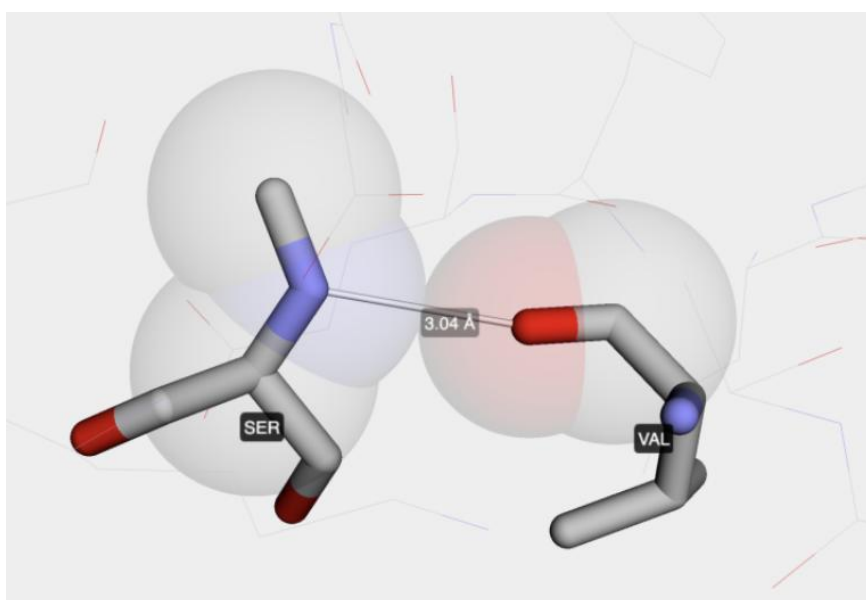


Figura 3. Interação do tipo Ligação de hidrogênio entre uma serina e uma valina. Fonte:[27] (PAIXÃO et al., 2023)

O fenômeno do efeito hidrofóbico ocorre quando moléculas apolares são inseridas em meio aquoso, resultando na formação de um “bolsão” ao redor do soluto, a fim de minimizar o contato com as moléculas de água [24]. Nos resíduos de aminoácidos, esse efeito é

particularmente perceptível nos grupos de átomos de carbono. Define-se um contato hidrofóbico quando um par de átomos de diferentes resíduos, ambos classificados como hidrofóbicos, estão a uma distância de até 4 Å [28].

Além disso, o efeito hidrofóbico é proeminente em outro tipo de interação observada entre aminoácidos com anéis aromáticos, denominada empilhamento aromático ou empilhamento  $\pi$ - $\pi$  [24]. Nesse contexto, as nuvens eletrostáticas dos anéis se alinham, formando uma ligação estável e não covalente, favorecida especialmente pelas partes apolares das moléculas. Existem três principais arranjos em que essa interação pode ocorrer: (i) em formato de sanduíche, quando um anel se sobrepõe ao outro, (ii) em formato de T, quando uma molécula está em um ângulo de 90° em relação à outra, e (iii) paralelamente deslocado, quando as estruturas dos anéis se sobrepõem com uma leve inclinação (Figura 4).

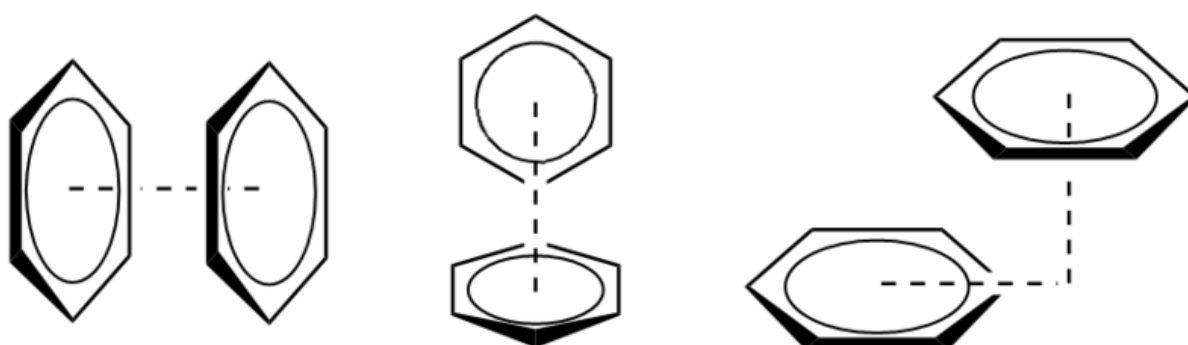


Figura 4. Interações aromáticas também conhecidos como pi-stacking ou empilhamento  $\pi$ - $\pi$ . Empilhamento em sanduíche (esquerda), forma de T (centro) e paralelamente deslocado (direita). Fonte: adaptado de Wikimedia Commons / Emily ricq (CC BY-3.0). Disponível em <https://commons.wikimedia.org/wiki/File:Benzene>

No contexto de estruturas polipeptídicas, essas ligações são estabelecidas pelos aminoácidos fenilalanina, tirosina e triptofano [24]. O critério de distância também pode ser aplicado para detectar esse tipo de interação, com uma distância máxima entre átomos nos anéis de 4 Å ou 4,5 Å (Figura 5).

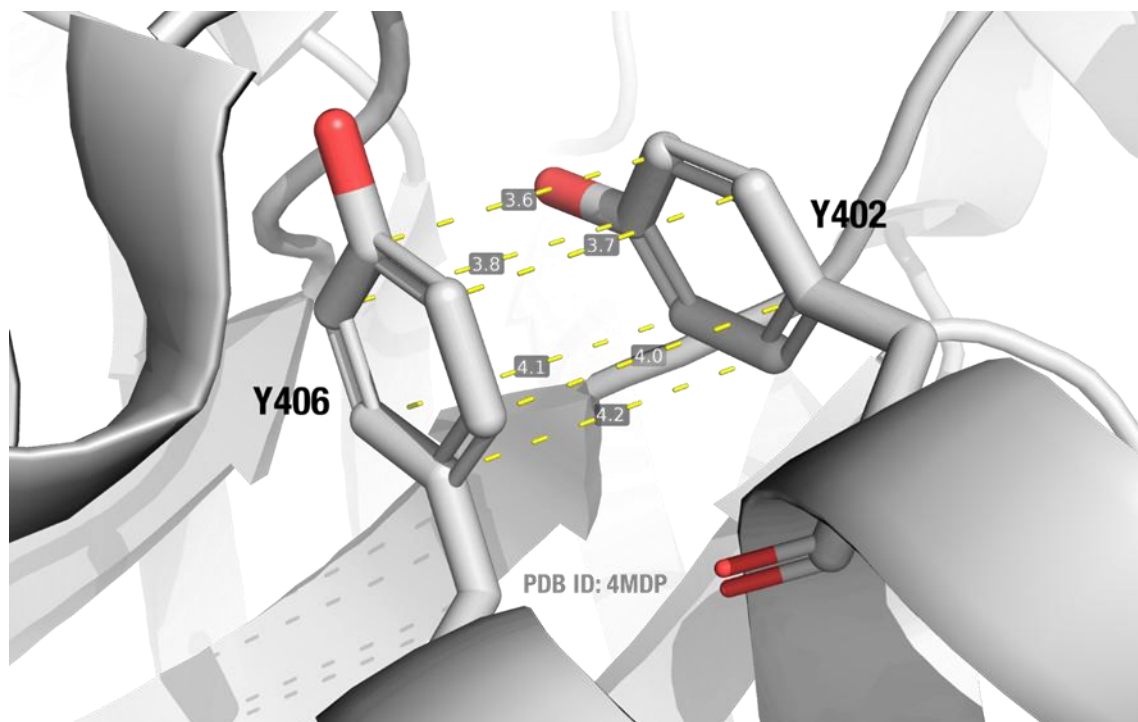


Figura 5. Interações aromática entre aminoácidos tirosina com anéis aromáticos, formato sanduiche.  
 Fonte: Diego Mariano | Wikimedia Commons (CC-BY 4.0). Disponível: <<https://creativecommons.org/licenses/by-sa/4.0/>>, via Wikimedia Commons

Aminoácidos polares carregados têm a capacidade de estabelecer interações atrativas (do tipo iônicas) quando os pares de aminoácidos possuem cargas opostas, ou repulsivas, quando as cargas são idênticas [24]. Três aminoácidos podem adquirir carga positiva (cátion): lisina, arginina e histidina, enquanto aspartato e glutamato podem se tornar negativamente carregados (ânions). O critério de distância, com uma separação de até 6 Å entre os átomos carregados, é utilizado para identificar essas interações. Pares de aminoácidos com cargas opostas podem simultaneamente formar interações iônicas e ligações de hidrogênio, resultando em uma interação mais robusta, conhecida como pontes salinas ou ligações salinas.

Outro tipo especial de interação é a ligação dissulfeto, também chamada de ponte dissulfeto, que ocorre entre cadeias polipeptídicas. Essa ligação é formada por uma ligação covalente entre átomos de enxofre presentes nas cadeias laterais de cisteínas. A energia associada às ligações dissulfeto é significativamente maior do que outros tipos de interações, devido à sua natureza covalente (envolvendo o compartilhamento de elétrons), enquanto as outras interações são consideradas não covalentes. **Erro! Fonte de referência não encontrada.**3 apresenta a energia de contato entre os diferentes tipos.

Tabela 3. Energia de contato. Fonte: adaptado de [22].

<b>Tipo de Contato</b>	<b>Energia do contato (Kcal/mol)</b>
van der Waals	0,08
Interação hidrofóbica	0,6
Empilhamento aromático	1,5
Ligação de hidrogênio	2,6
Ponte salina	10
Ligação dissulfeto	85,0

### 2.3 O coronavírus SARS-CoV-2 e suas Proteínas Estruturais

Uma primeira versão do genoma do SARS-CoV-2 foi divulgada em 10 de janeiro de 2020, seguida por genomas adicionais recolhidos de diferentes pacientes. Os dados foram depositados no banco de dados da Iniciativa Global sobre Compartilhamento de Todos os Dados da Gripe (GISAID), que é usado principalmente para compartilhar dados sobre vírus da gripe [29]. Através da disponibilidade dos dados e de análises genéticas de evolução, foi possível realizar sua classificação.

Os coronavírus (CoVs) são um grupo de vírus envelopados com um único genoma de RNA de fita positiva, sendo geneticamente classificados em quatro gêneros: *Alphacoronavirus*, *Betacoronavirus*, *Gammacoronavirus* e *Deltacoronavirus* [9] (Figura 6). Sendo um dos maiores vírus de RNA identificados até agora, os coronavírus (CoVs), pertencentes à família *Coronaviridae*, são divididos em 4 gêneros,  $\alpha$ -,  $\beta$ -,  $\delta$ - e  $\gamma$ -coronavírus, enquanto os  $\beta$ -coronavírus são divididos em linhagens A, B, C e D. Os sete CoVs que podem infectar humanos (HCoVs) incluem HCoV-229E e HCoV-NL63 nos  $\alpha$ -coronavírus, HCoV-OC43 e HCoV-HKU1 na linhagem A de  $\beta$ -coronavírus, SARS-CoV e SARS-CoV-2 no  $\beta$ -coronavírus da linhagem B ( $\beta$ -B coronavírus) e MERS-CoV na linhagem  $\beta$ -coronavírus C [30].

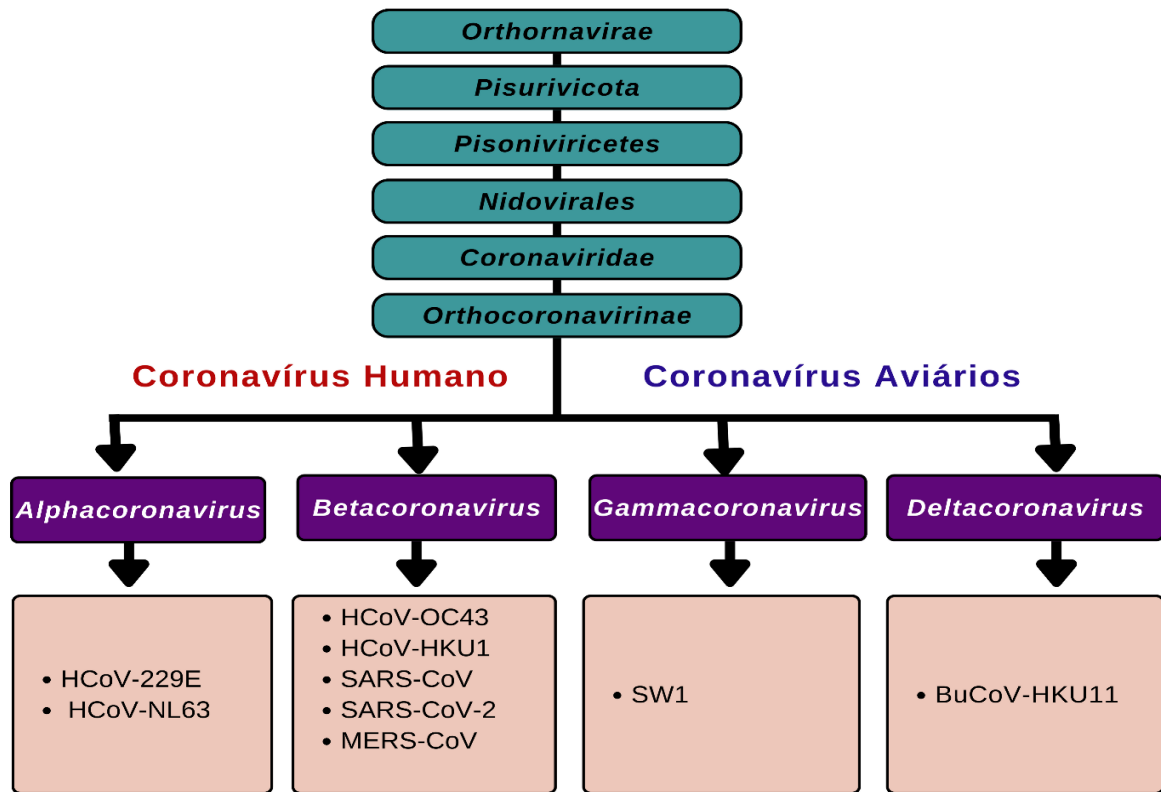


Figura 6. Taxonomia de Coronaviridae. HCoV, coronavírus humano; MERS-CoV, coronavírus da síndrome respiratória do Oriente Médio; SARS-CoV, coronavírus da síndrome respiratória aguda grave; SARS-CoV-2, síndrome respiratória aguda grave coronavírus-2; SW1, coronavírus da baleia beluga; e BuCoV-HKU11, coronavírus bulbul HKU11. Fonte:[7].

O genoma do SARS-CoV-2 pode ser composto por 10 a 12 quadros de leitura aberta (ORFs) putativos que codificam proteínas não estruturais e proteínas estruturais. As proteínas não estruturais (nsps) estão envolvidas no processamento e replicação do vírus, enquanto as proteínas estruturais auxiliam na montagem e liberação de novas cópias virais. As quatro proteínas estruturais significativas são: proteína *Spike* (S), proteína do envelope (E), proteína de membrana (M) e proteína do nucleocapsídeo (N) [6,7]. As suas respectivas funções são: a proteína M atua como organizadora central na montagem e formação do envelope viral, interagindo com outras proteínas estruturais, se ligando às proteínas S e N para completar novas montagens virais. A proteína E é abundantemente expressa no ciclo de replicação nas células infectadas, embora uma pequena porção dela seja incorporada ao envelope viral e contribui principalmente para a montagem e brotamento viral. A proteína N exibe suas funções por interação com a cadeia positiva de RNA do genoma viral, formando assim um complexo helicoidal de ribonucleocapsídeo. Também interage com outras proteínas estruturais da membrana durante a montagem dos vírions. Durante a infecção, a proteína S em sua forma

trimérica é clivada nas subunidades S1 (“cabeça”) e S2 (“caule”). S1 é responsável pela ligação ao receptor, enquanto S2 é responsável pela fusão da membrana [31].

## 2.4 A Proteína *Spike* (S) como alvo Principal

A proteína S, uma glicoproteína transmembrana, contém a subunidade S1 e a subunidade S2 e que se liga a uma protease Transmembrana Serina Protease 2 (TMPRSS2) presentes na membrana do hospedeiro, semelhante ao SARS-CoV. A TMPRSS2 é uma protease semelhante à furina que contém um único domínio transmembranar e um único domínio para os domínios SR TRYPSIN e LDLA que pode clivar a proteína S em S1 e S2 e aumentar a capacidade de fusão viral com a membrana da célula hospedeira [7]. A subunidade S1, que contém o domínio de ligação ao receptor, capaz de dobrar funcionalmente de forma independente, possui a capacidade de se ligar à enzima conversora de angiotensina II [32]. A subunidade S2, composta por peptídeo de fusão (FP), domínios de repetição do *heptad* 1 (HR1) e *heptad* 2 (HR2), domínio transmembranar (TM) e fusão do domínio citoplasmático (CP), é responsável por mediar a fusão e entrada viral (Figura 7 e Figura 8).

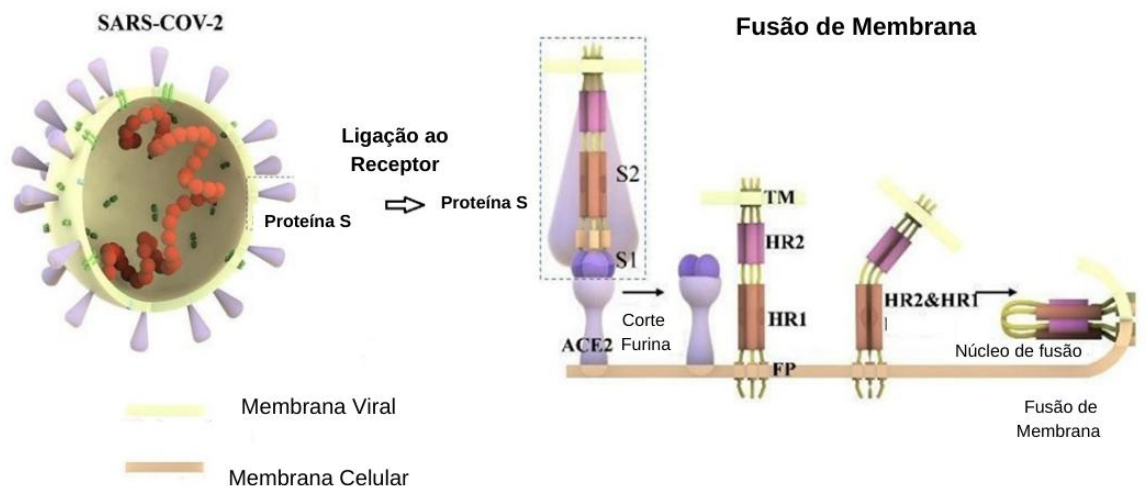


Figura 7. Representação do vírus SARS-CoV-2 com destaque para a proteína Spike (S) que está envolvida no processo de ligação ao receptor na célula hospedeira. A proteína S do vírus SARS-CoV-2 liga-se ao receptor ACE2 presente na membrana celular da célula hospedeira. O corte pela enzima Furina é um passo necessário para ativar a proteína S e permitir a fusão das membranas. O processo de fusão é detalhado, mostrando as subunidades S1 e S2 da proteína S e os objetivos de transição da fusão. TM (domínio transmembranar), HR1 (domínios de repetição do heptad 1), HR2 (domínios de repetição do heptad 2), e FP (peptídeo de fusão) são componentes chave nesse processo. A fusão das membranas virais e celulares culmina na entrada do material genético viral na célula hospedeira, permitindo a infecção. A membrana viral é representada em amarelo claro e a membrana celular em amarelo escuro. Fonte: [10]

Durante a infecção pelo coronavírus, o FP da subunidade S2 é inserido na membrana da célula alvo, expondo o domínio HR1 e HR2 e permitindo que o domínio HR1 e HR2 formem um feixe de 6 hélices que aproxima as bicamadas lipídicas celulares e virais, iniciando o processo de fusão da membrana. Durante esse processo, a proteína S apresenta três estados conformacionais: um estado nativo de pré-fusão, um estado intermediário pré-grampo e um estado estável em grampo pós-fusão. Antes de se ligar ao receptor, a proteína S sai em estado nativo de pré-fusão. Quando a proteína S se liga ao receptor, a subunidade S1 será degradada e será exposta a unidade S2. A proteína S está no estado intermediário pré-grampo. Em seguida, três pares de HR1 e HR2 formam uma estrutura de feixe de seis hélices, e o núcleo de fusão toma forma [32]. Abaixo temos a representação esquemática do genoma e domínio funcional da proteína S do SARS-CoV-2 (Figura 8).

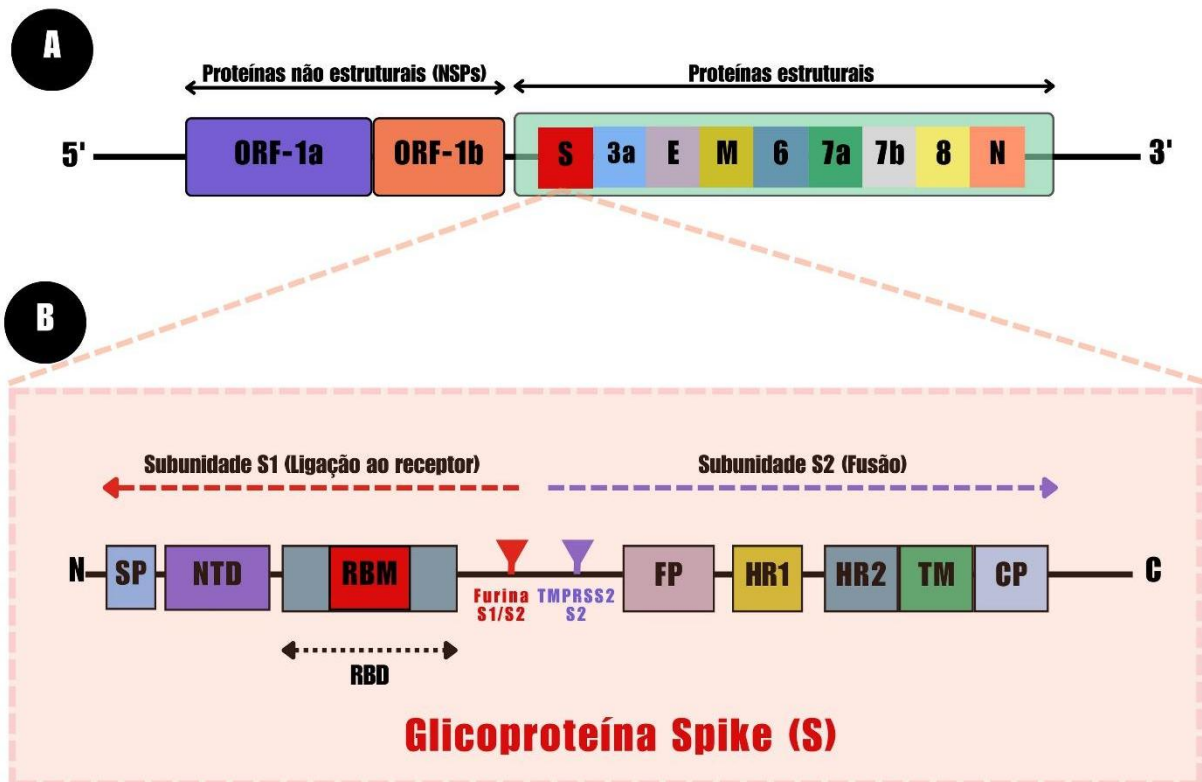


Figura 8. Representação esquemática do genoma e domínio funcional da proteína S do SARS-CoV-2. A. O SARS-CoV-2 é um vírus de RNA conhecido com uma estrutura de 5'-cap e 3'-poli-A e possui 14 supostos quadros de leitura abertos (ORFs) que codificam 27 proteínas. O vírion possui quatro proteínas estruturais, conhecidas como proteínas S, *Spike*; E, envelope; M, membrana; e N, nucleocapsídeo. B. As proteínas *Spike* contêm subunidades S1 e S2 e os locais de clivagem para Furina e TMPRSS2. As proteínas *Spike* medeiam a ligação viral às células hospedeiras após ativação pela enzima TMPRSS2. SP, peptídeo sinal; NTD, domínio N-terminal; RBD, domínio de ligação ao receptor, contém RBM, motivo de ligação central no subdomínio externo; FP, peptídeo de fusão; HR1, domínios de repetição do heptal 1 e heptal 2 HR2; TM, domínio transmembrana; CP, domínio do citoplasma.

## 2.5 A Enzima Conversora de Angiotensina 2 (ACE2) como Receptor Viral

Responsável pela maturação do hormônio peptídico angiotensina (Ang), que por sua vez regula a vasoconstrição e a pressão arterial, a ACE2 é uma enzima metalocarboxipeptidase transmembrana tipo I de passagem única. Ela está presente em células epiteliais alveolares tipo II e também é expressa em diversos tecidos extrapulmonares, incluindo coração, rim e intestino [18]. Sua estrutura é formada por dois domínios principais - o N-terminal PD e o domínio semelhante à colectrina (CLD) na extremidade C-terminal. A proteína S do SARS-CoV-2 se liga ao homodímero de ACE2, o que facilita a entrada do vírus nas células hospedeiras, formando a interação SARS-CoV-2-RBD/ACE2 [33]. Após a ligação ao receptor da enzima ACE2, a partícula do vírus usa receptores e endossomos da célula hospedeira para entrar nas células. A protease do hospedeiro, TMPRSS2, facilita a entrada na célula através da proteína S. Uma vez dentro da célula, são sintetizadas poliproteínas virais que codificam o complexo replicase-transcriptase. O vírus por sua vez sintetiza RNA por meio de sua RNA polimerase dependente de RNA (Figura 9). As proteínas estruturais são sintetizadas levando à conclusão da montagem e liberação de partículas virais [13].

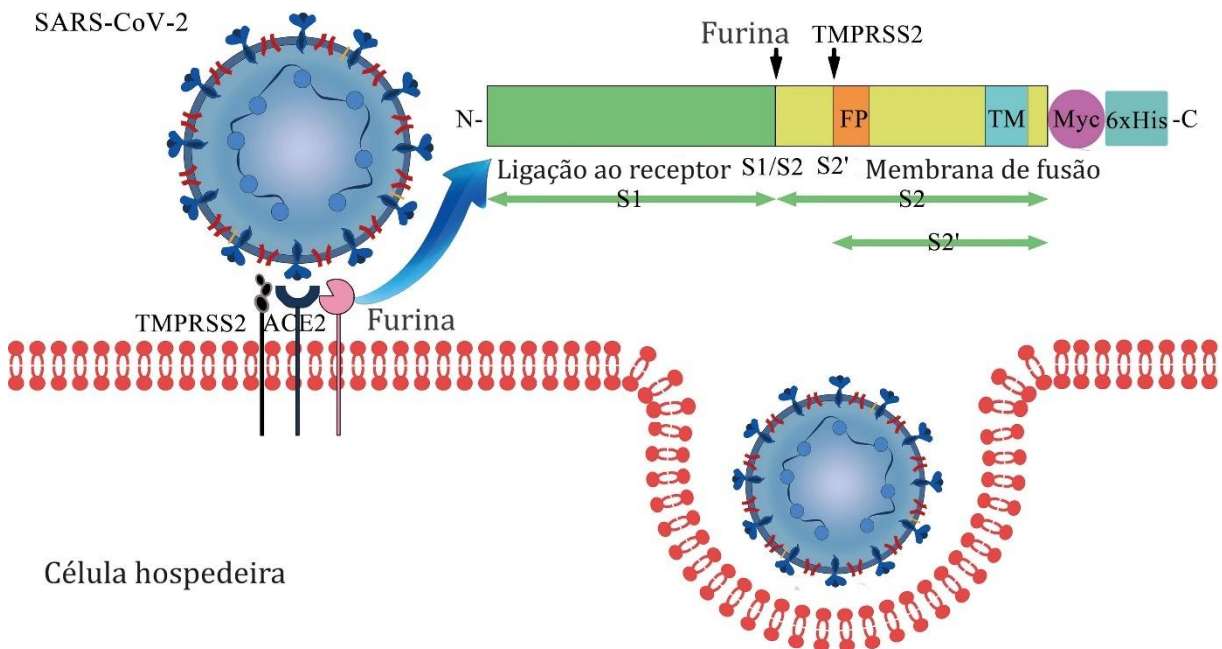


Figura 9. Entrada do SARS-CoV-2 na célula: interação entre a proteína S e a membrana ACE2. A proteína S do SARS-CoV-2 possui duas unidades funcionais, S1 e S2. A S1 se liga ao receptor ACE2, resultando na exposição do sítio de clivagem S1/S2. A proteína é então clivada pela furina e pela TMPRSS2, o que provoca mudanças conformacionais que permitem a fusão da membrana, facilitando a inserção do peptídeo de fusão S2. Isso leva à liberação do RNA viral, que após ser internalizado, resulta na replicação e tradução do vírus. Novos vírus são montados e liberados da célula, iniciando um novo ciclo de infecção. Fonte: [34]

## 2.6 As Interações Moleculares na interface RBD-ACE2

Utilizando o alinhamento de sequências entre os RBDs do SARS-CoV-2 e do SARS-CoV (~75% de identidade de sequência) com o RBD do MERS-CoV (~24% de identidade de sequência) revelaram diferenças nas sequências de aminoácidos primários que levam a diferentes receptores do hospedeiro, ou seja, ACE2 para SARS-CoV-2 e SARS-CoV e dipeptidil peptidase 4 para MERS-CoV [12].

Acredita-se que ligantes à base de peptídeos têm potencial para impedir a entrada do vírus nas células humanas por possivelmente bloquear a interação entre a proteína S e ACE2, oferecendo uma nova oportunidade para intervenção terapêutica. Logo, o foco era o desenvolvimento de peptídeos inibidores baseados no complexo SARS-CoV-2 RBD/hACE2. Para esse objetivo, utilizou-se a estrutura 6M0J para o modelo do projeto.

Através do uso de cristalografia de raio X com uma resolução de 2,45 Å, foram descritas as interações do SARS-CoV-2 RBD e ACE2 [2]. O RBD é caracterizado por uma folha  $\beta$  antiparalela de cinco fitas torcidas com uma longa inserção entre as fitas  $\beta_4$  e  $\beta_7$ . A maioria das interações com o domínio peptidase N-terminal (PD) da ACE2 são mediadas através desta longa inserção. O ACE2 PD N-terminal possui dois lóbulos, formando o sítio de ligação do substrato peptídico entre eles. O RBM estendido no RBD do SARS-CoV-2 entra em contato com o lado inferior do pequeno lóbulo do ACE2, com uma superfície externa côncava no RBM que acomoda a hélice N-terminal do ACE2. O RBD e o ACE2 contribuem com um total de 17 e 20 resíduos, respectivamente, na interface de ligação. Existem 14 resíduos compartilhados pelos RBD do SARS-CoV e do SARS-CoV-2 durante a sua ligação ao ACE2, e deles 8 são idênticos (Figura 10).

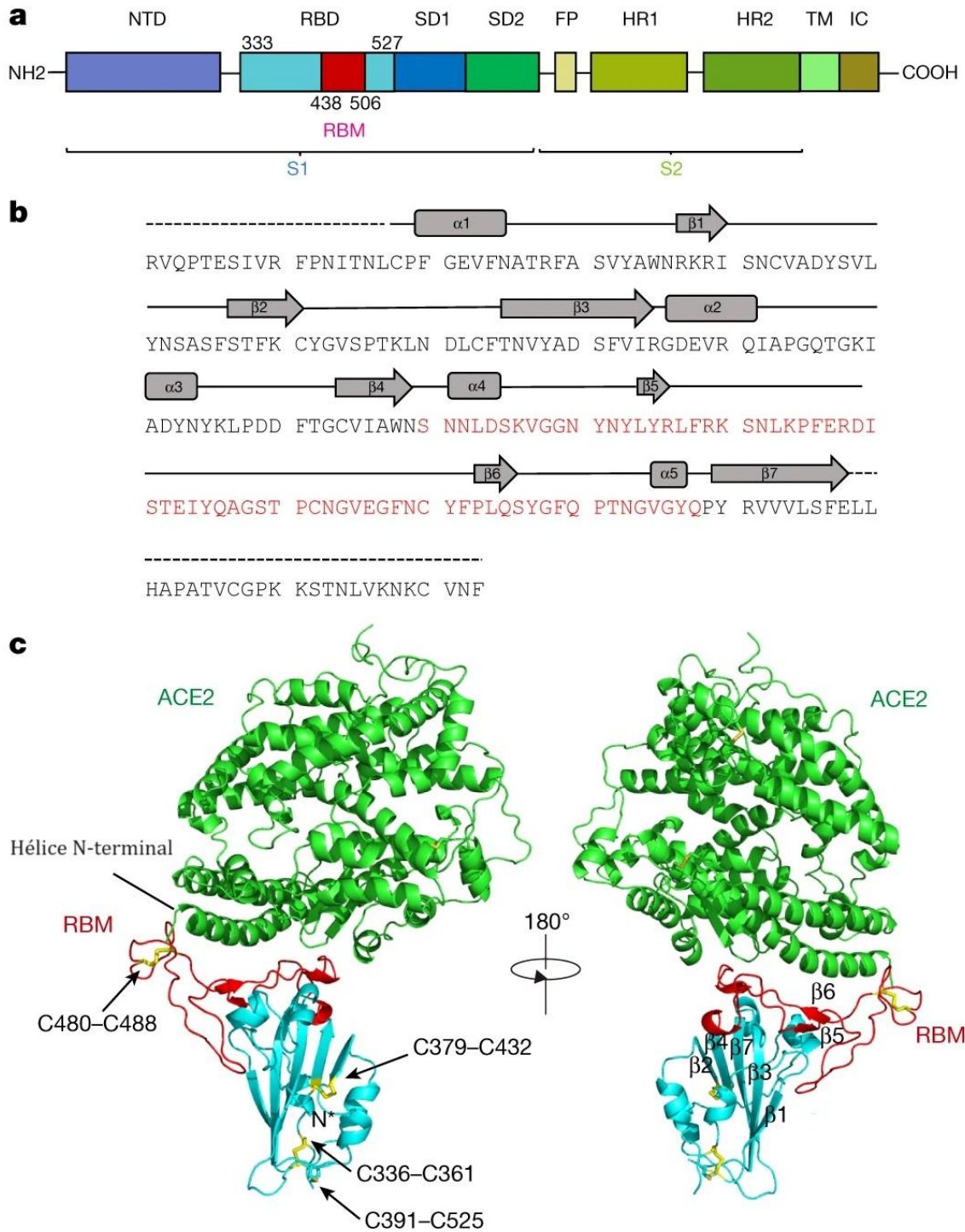


Figura 10. Estrutura geral do RBD do SARS-CoV-2 ligado ao ACE2 **a**, Topologia geral do monômero Spike do SARS-CoV-2. FP, peptídeo de fusão; HR1, repetição heptad 1; HR2, repetição heptad 2; IC, domínio intracelular; NTD, domínio N-terminal; SD1, subdomínio 1; SD2, subdomínio 2; TM, região transmembrana. **b**, Sequência e estruturas secundárias do RBD do SARS-CoV-2. A sequência RBM é mostrada em vermelho. **c**, Estrutura geral do RBD do SARS-CoV-2 ligado ao ACE2. O ACE2 é mostrado em verde. O núcleo do RBD do SARS-CoV-2 é mostrado em ciano e o RBM em vermelho. As ligações dissulfeto no RBD do SARS-CoV-2 são mostradas como bastões e indicadas por setas. A hélice N-terminal do ACE2 responsável pela ligação é rotulada. Fonte: [2]

## 2.7 Exemplos de alvos atuais contra SARS-CoV-2

Atualmente, as principais estratégias antivirais contra o SARS-CoV-2 podem ser amplamente divididas em dois tipos: estratégias diretas que visam o vírus e estratégias indiretas que modulam o hospedeiro. Em ambas, há medicamentos aprovados e candidatos experimentais em ensaios clínicos. A descoberta de novos alvos celulares para o coronavírus levou a novas estratégias clínicas que merecem investigação na COVID-19 [35].

Entre os antivirais, destaca-se como exemplo o Remdesivir, aprovado pela FDA para tratar pacientes com COVID-19 e pneumonia que apresentam escassez de oxigênio [35]. Sua funcionalidade baseia-se na capacidade de inibir a replicação viral nas células epiteliais associadas à respiração [36]. Estudos indicam que a combinação de Remdesivir com Baricitinibe reduziu o tempo de recuperação de pacientes hospitalizados com pneumonia por COVID-19 [36].

Outra opção são alguns medicamentos anti-HIV, como Lopinavir e Ritonavir, direcionados ao vírus RNA, que atuam como inibidores de protease, mas não demonstraram benefícios significativos em pacientes gravemente afetados [37]. Recentemente, os anticorpos monoclonais Tocilizumabe e Sarilumabe, inicialmente usados para tratar artrite reumatoide, foram reaproveitados para tratar a liberação da tempestade de citocinas em casos críticos de COVID-19 [38]. Ensaios clínicos mostraram melhorias na taxa de sobrevivência quando administrados dentro de 24 horas após a entrada nas unidades de terapia intensiva (UTI) [39].

O Casirivimab em combinação com Imdevimab, conhecido como REGEN-COV™, é um coquetel de anticorpos monoclonais eficaz contra o SARS-CoV-2, reduzindo hospitalizações ou a mortes em pacientes não hospitalizados em 70% [40]. Foi aprovado pelo FDA para tratamento de casos leves a moderados em pacientes adultos e pediátricos com alto risco de gravidade [30].

Outra opção seria o Paxlovid, que se trata de uma formulação combinada que contém nirmatrelvir, um inibidor de protease de segunda geração, e ritonavir, um modulador farmacológico. O Paxlovid é indicado no tratamento de infecções causadas pelo SARS-CoV-2. A administração de Paxlovid ocorre por via oral, durante um período de cinco dias, e é recomendada para pacientes que apresentam os primeiros sinais da infecção. Estudos não demonstraram associação entre o uso de Paxlovid e elevações nos níveis de aminotransferases séricas ou ocorrência de lesão hepática de caráter clinicamente aparente [41].

Como modificadores imunológicos existem vacinas contra a COVID-19 com diferentes mecanismos de funcionamento.

- Vacinas de DNA que têm por princípio a clonagem do quadro de leitura aberta (ORF) do pico SARS-CoV-2 em um DNA plasmídico.
- Vacinas de vetores virais, onde a proteína Spike ORF é clonada no genoma do adenovírus, formando um vírus recombinante infeccioso. A AZD1222 (Oxford-AstraZeneca) é uma vacina contra SARS-CoV-2 baseada em adenovírus recombinante construída a partir do vetor de adenovírus de chimpanzé símio com deficiência de replicação (ChAdOx2) que expressa a glicoproteína de pico SARS-COV-2 de comprimento total [42].
- Vacina de mRNA, por meio de síntese química o mRNA do pico do SARS-CoV-2 é sintetizado e depois envolvido com nanopartículas lipídicas e posteriormente injetado no corpo humano. Dentre elas, destacam-se a BNT162b2 (Pfizer-BioNTech) e mRNA-1273 (Moderna biotechnologies Inc.) [43].
- Vacina Proteica, podendo-se utilizar a parte total ou subunidade da proteína Spike com um adjuvante específico antes de ser injetada no sistema humano, como exemplo podemos citar a NVX-CoV2373 (Novavax) [44].
- Vacina de Vírus Inativado, na qual o vírus SARS-CoV-2 é quimicamente inativado, misturado com adjuvante específico e injetado por via intramuscular. Em destaque temos Vacina COVID-19 (Vero Cell) Inativada (CoronaVac (anteriormente PiCoVacc)) desenvolvida pela Sinovac Biotech Ltd [45].

## 2.8 O uso de subsequências conservadas para o planejamento de Peptídeos

O planejamento racional de peptídeos terapêuticos frequentemente se baseia na identificação de regiões funcionais conservadas em proteínas-alvo. Um elevado grau de conservação de uma subsequência é importante em projetos como o desenvolvimento de vacina peptídica sintética ou medicamentos terapêuticos. A razão para isso é que, ao longo do tempo evolutivo, incluindo o surgimento de novas cepas virais, uma subsequência conservada geralmente indica um motivo que tem uma função importante para a replicação e/ou sobrevivência do vírus, logo ela não sofre variação e é mantida. Outro fator a ser considerado é que, sob a pressão seletiva de vacinas e medicamentos, a evolução de vírus de RNA pode ser acelerada, tornando essas estratégias terapêuticas obsoletas, talvez em um curto espaço de

tempo. Vírus de RNA como o SARS-CoV-2 são caracterizados por altas taxas de mutação, ocorrendo tanto no RNA genômico quanto no antígeno. A edição de RNA dentro do hospedeiro atua como uma fonte adicional de mutação além das mutações induzidas pela polimerase e pode ter um impacto duradouro na evolução de patógenos virais [47].

Diante da necessidade de propor uma alternativa terapêutica utilizando peptídeos, este projeto baseou-se em substituições conservadoras, onde as substituições sugeridas fossem selecionadas com base em resíduos de aminoácidos de propriedades semelhantes, considerando critérios evolutivos e propensões da estrutura secundária. Tais substituições fossem consideradas mutações alternativas na engenharia de peptídeos, isto é, resíduos de aminoácidos com características similares. Assim, para esse estudo foi considerado o estudo de Simon French e Barry Robson [46], que forneceram uma base quantitativa para definir substituições conservativas.

Os autores determinaram, de forma objetiva e quantitativa, as propriedades que influenciam as substituições conservativas utilizando uma técnica de escala multidimensional. O método permitiu a análise das relações evolutivas sem suposições ou vieses iniciais, oferecendo uma abordagem baseada em dados para compreender a conservação de propriedades em sequências de proteínas. Foi utilizado dados extensivos de várias famílias de proteínas, visando calcular a média dos efeitos familiares específicos como conservação da hidrofobicidade e do volume molecular dos resíduos de aminoácidos, focando nas propriedades geralmente conservadas nas proteínas, em vez de instâncias específicas.

A matriz de similaridade de Dayhoff *et al.*, [48] representou a proporção de duas probabilidades: a probabilidade de que dois resíduos de aminoácidos na mesma posição em duas proteínas compartilhassem uma ancestralidade comum versus a probabilidade de que sua semelhança tivesse ocorrido por acaso. Derivada da comparação de sequências dentro de várias famílias de proteínas, essa matriz forneceu informações sobre as pressões seletivas atuantes sobre as proteínas em geral. Ela serviu como uma ferramenta para avaliar as relações evolutivas entre os aminoácidos com base em suas frequências de substituição, indicando quais resíduos tinham maior probabilidade de serem conservados.

## 2.9 Docking Molecular

O *docking* molecular, também conhecido como ancoragem molecular, é um experimento computacional que visa simular a interação entre duas moléculas para formar complexos estáveis, como os complexos proteína-proteína ou, neste caso proteína-peptídeo [49]. Essa técnica oferece informações sobre o comportamento das moléculas em nível atômico dentro do sítio ativo de uma proteína, descrevendo a conformação e orientação do ligante, bem como estimando a afinidade de ligação (Figura 11). Identificar o local de ligação antes do processo de encaixe aumenta muito a eficiência, sendo esta informação geralmente conhecida antes que os ligantes sejam ancorados [50]. O objetivo principal da pesquisa de *docking* é gerar diversas conformações de ligação ao ligante no sítio de ligação ao receptor e estimar a energia de ligação no complexo receptor-ligante [51 e 53]. A energia da ligação entre o ligante (peptídeo) e o sítio ativo da proteína é determinada pelas interações não covalentes, pelos termos entálpicos e termos relacionados à dessolvatação dos compostos entre grupos funcionais na molécula bioativa e no sítio ativo da proteína (receptor) [52].

Por meio do *docking* molecular, várias conformações espaciais do ligante são produzidas, permitindo ao analista identificar a interação mais provável no ligante alvo. Cada conformação espacial produz energias livres de ligação (entre o ligante e o alvo), com a menor energia sendo considerada a explicação mais plausível para a conformação de interação [51].

A utilização do *docking* molecular no exame da afinidade de moléculas biologicamente ativas com seus receptores envolve dois elementos-chave: o algoritmo de busca e a função de avaliação (ou *scoring function*). O algoritmo de busca é responsável por explorar o espaço conformacional e de orientação do peptídeo dentro do sítio ativo da proteína alvo. A função de avaliação, por sua vez, calcula uma pontuação (*score*) que estima a energia livre de ligação ( $\Delta G$ ) para cada pose gerada, identificando os modos de ligação mais favoráveis, caracterizados pela menor energia [54].

Existem vários tipos de funções de avaliação disponíveis nos programas de *docking* [53]. Elas podem ser classificadas em categorias principais:

- Funções baseadas em campo de força: Utilizam parâmetros de mecânica molecular para calcular as energias das interações de van der Waals e eletrostáticas.
- Funções empíricas: Utilizam termos de energia ponderados, cujos coeficientes são ajustados a partir de dados experimentais de afinidade de ligação de complexos conhecidos.
- Funções baseadas em conhecimento: Derivam potenciais estatísticos a partir da frequência de interações atômicas observadas em um banco de dados de estruturas cristalográficas.
- Funções baseadas em aprendizado de máquina (*machine learning*): Recentemente, novas funções têm sido desenvolvidas utilizando inteligência artificial, que aprendem a prever a afinidade de ligação a partir de um grande conjunto de dados de complexos proteína-ligante, demonstrando alta precisão.

Além de selecionar a função de avaliação, os pesquisadores também podem determinar o nível de flexibilidade no experimento. A metodologia permite diferentes configurações: o mais simples considera ligante e proteína rígidos; abordagens mais complexas permitem a flexibilidade do ligante com uma proteína rígida; ou a flexibilidade de ambos, o que aumenta

consideravelmente o custo computacional, mas representa de forma mais fiel a dinâmica biológica [57].

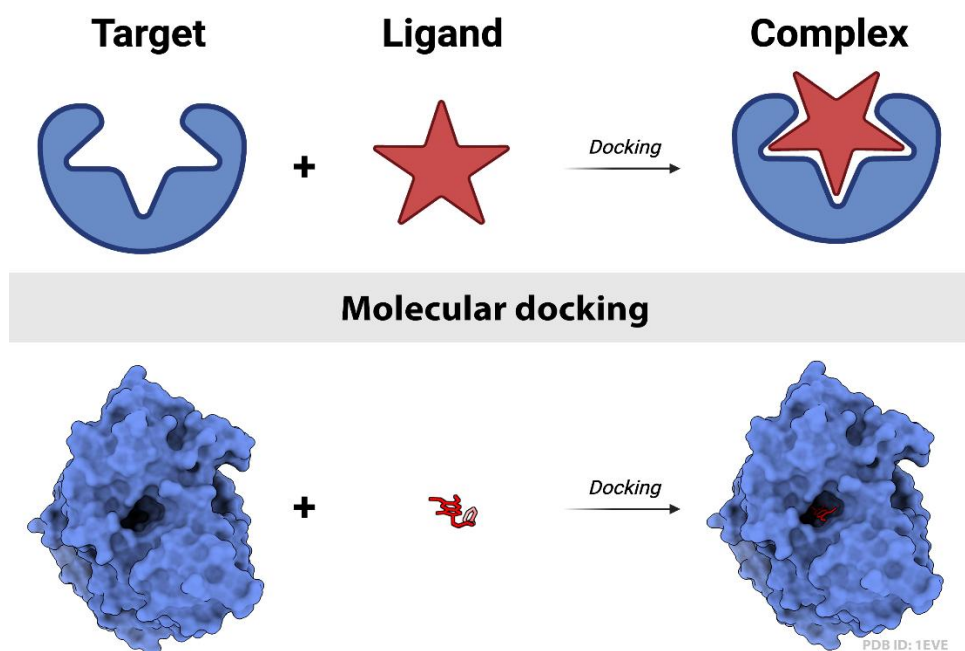


Figura 11. *Docking* molecular, também conhecido como encaixe molecular ou ancoragem. Fonte: Diego Mariano, CC BY-SA 4.0. Disponível: <<https://creativecommons.org/licenses/by-sa/4.0/>>, via Wikimedia Commons

## 2.10 Dinâmica Molecular

A Simulação de Dinâmica Molecular (DM) é uma técnica computacional versátil, empregada para o estudo do comportamento de macromoléculas biológicas. [58–60]. Os princípios fundamentais da Mecânica Clássica sustentam a abordagem da DM, oferecendo informações valiosas sobre os comportamentos dinâmicos microscópicos e dependentes do tempo dos átomos de um sistema [60,61]. Para deduzir as propriedades macroscópicas desejadas, a utilização da mecânica estatística torna-se crucial, pois ela permite calcular atributos macroscópicos observáveis (como pressão, energia interna, volume, temperatura, entropia, energia livre etc.) a partir de dados microscópicos. Essa técnica computacional envolve essencialmente a determinação das posições atômicas de cada átomo do sistema em vários pontos de tempo consecutivos, com intervalos tipicamente da ordem de *femto* segundos (1 fs =  $10^{-15}$  segundos). A coleção de coordenadas do núcleo de cada átomo no sistema, junto com sua progressão temporal, delinea a trajetória do sistema.

Baseada na Mecânica Clássica, a dinâmica molecular utiliza as equações do movimento, inicialmente postuladas por Newton e posteriormente generalizadas por Hamilton e Lagrange [49]. De acordo com a Segunda Lei de Newton, o movimento da  $i$ -ésima partícula dentro de um sistema, sujeito a uma força, é delineado de acordo com a Equação 1:

$$F_i(t) = m_i a_i \quad (1)$$

Onde  $F_i$  denota a força atuando sobre a  $i$ -ésima partícula de massa  $m_i$  no sistema em um instante de tempo específico  $t$ , levando à aceleração  $a_i$ . Em termos práticos, para cada átomo dentro do sistema, calcula-se a força vetorial resultante originada dos demais átomos. Posteriormente, a aceleração instantânea é determinada, influenciando a nova posição e velocidade do átomo no próximo incremento de tempo  $t + \delta t$ . Esse processo é repetido iterativamente durante toda a simulação, gerando posições e velocidades atualizadas para instâncias de tempo  $t + 2\delta t, t + 3\delta t \dots t + n\delta t$ .

A Mecânica Molecular (MM) considera as moléculas como conjuntos de átomos regidos por forças newtonianas, visualizando-as como partículas conectadas por forças harmônicas ou elásticas. O conjunto de potenciais de interação entre as partículas é denominado como um campo de força.[49,60]. A equação 2 ilustra um campo de força padrão, em que o potencial total,  $V_{(r)}$ , surge do somatório resultante dos potenciais relacionados às interações presentes no sistema.

$$V_{(r)} = \sum V_{(l)} + \sum V_{(\theta)} + \sum V_{(\varphi)} + \sum V_{(vdW)} + \sum V_{(elet)} \quad (2)$$

Os vários componentes que contribuem para esse potencial incluem  $V_l$  para alongamento de ligação,  $V_\theta$  para deformação angular,  $V_\varphi$  para potencial de torção,  $V_{vdW}$  para interações de van der Waals e  $V_{elet}$  para interações de Coulomb [49].

Ao empregar a aproximação do potencial harmônico, os aspectos associados ao alongamento da ligação e à deformação angular são delineados por meio das Equações 3 e 4.

$$V_{(l)} = \frac{1}{2} K_r (r - r_0)^2 \quad (3)$$

$$V_{(\theta)} = \frac{1}{2} K_\theta (\theta - \theta_0)^2 \quad (4)$$

Onde  $r$  e  $\theta$  denotam o comprimento e o ângulo da ligação, com  $r_0$  e  $\theta_0$  denotando suas respectivas posições de equilíbrio. Os parâmetros  $K_r$  e  $K_\theta$  significam as constantes de força pertinentes ao sistema [54]. A inclusão de interações entre partículas não ligadas é facilitada pelos potenciais de Lennard-Jones ( $V_{vdW}$ ) e Coulomb ( $V_{elet}$ ), conforme as Equações 5 e 6.

$$V_{(vdW)} = 4\varepsilon_{ij} \left[ \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left( \frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] \quad (5)$$

$$V_{(elet)} = \frac{q_i q_j}{4\pi\varepsilon_0\varepsilon_r r_{ij}} \quad (6)$$

Uma vez que o campo de força é especificado, a fase subsequente envolve a integração numérica das equações de movimento. Inúmeros algoritmos são utilizados para esta finalidade, entre os quais o algoritmo de Verlet e suas variantes (como o Leap-frog) têm importância significativa em simulações de DM [62].

## 2.11 Trabalhos Relacionados

Diversas abordagens computacionais foram propostas para o planejamento e a otimização de peptídeos com potencial terapêutico contra o SARS-CoV-2. A seguir, são descritos alguns trabalhos que contextualizam a presente tese, seja por focarem em alvos semelhantes ou por empregarem estratégias computacionais correlatas.

Rongsong Ling et al. [10], em seu estudo, propuseram o desenvolvimento de peptídeos antivirais direcionados ao núcleo de fusão do SARS-CoV-2, com o intuito de evitar a fusão da membrana viral [10]. Os peptídeos antivirais desenvolvidos tinham como alvo específico o núcleo de fusão do SARS-CoV-2, concentrando-se nas regiões HR1 e HR2. Estes peptídeos foram projetados para competir pela ligação à região HR1, impedindo a formação do núcleo de fusão e, conseqüentemente, inibindo a fusão da membrana, processo crucial para a entrada do vírus nas células hospedeiras [10]. A metodologia envolveu modelagem por homologia com o *software* SWISS-MODEL, alinhamento de sequências para identificação dos domínios, *docking* proteína-proteína com GRAMM-X e simulações de dinâmica molecular para avaliação da energia de ligação [10]. O estudo ressaltou a importância de direcionar o núcleo de fusão viral como uma nova estratégia no desenvolvimento de fármacos antivirais.

Em outro estudo B. Robson [67], investigou o desenvolvimento de uma vacina sintética e um antagonista peptidomimético preventivo contra o vírus SARS-CoV-2, destacando a utilização de ferramentas de bioinformática [67]. Ferramentas como o XTRACTOR foram empregadas para gerar estruturas proteicas 3D preditivas do vírus, constituindo um recurso valioso nos esforços de pesquisa para combater a COVID-19 [68,69].

A análise do genoma utilizou a proteína de superfície do vírus (isolado do mercado de frutos do mar de Wuhan), bem como a identificação de sequências de aminoácidos conservadas entre os coronavírus, foi conduzida com o auxílio de ferramentas bioinformáticas. Além disso, a pesquisa identificou a sequências de aminoácidos bem conservadas, como o motivo KRSFIEDLLFNKV, que correspondem a locais de clivagem (S1/S2) do vírus SARS. O autor reconhece, contudo, limitações em sua abordagem, como a dependência de estruturas preditivas e a não consideração da dinâmica de mutações virais, que poderiam afetar a eficácia das intervenções propostas [67].

No campo de bancos de dados para interações, destaca-se a ferramenta Propedia [14], que consiste em um banco de dados com mais de 19.000 estruturas de alta resolução de complexos proteína-peptídeo. A ferramenta utiliza um algoritmo de agrupamento híbrido para organizar os dados com base na similaridade de sequência e estrutura da interface, oferecendo suporte ao planejamento racional de peptídeos ao permitir que os usuários analisem as interações sob diversas perspectivas [14, 15].

## 3. Objetivos

### 3.1 Objetivo geral

Desenvolver e avaliar um algoritmo de código-fonte aberto, denominado POTTER, para a engenharia e otimização de peptídeos com potencial de inibição sítio-específica, utilizando como estudo de caso a interface de interação entre a proteína *Spike* do SARS-CoV-2 e o receptor humano ACE2.

### 3.2 Objetivos específicos

Para alcançar o objetivo geral, foram definidos os seguintes objetivos específicos:

- Implementar um algoritmo genético para otimização de peptídeos inibitórios, partindo da estrutura da região de interação SARS-CoV-2-RBD/ACE2.
- Desenvolver uma função de avaliação no algoritmo baseada em métricas estruturais, como o número de contatos intermoleculares (ligações de hidrogênio, interações atrativas, repulsivas e empilhamento aromático) e a ocupação do sítio ativo.
- Avaliar o desempenho do algoritmo POTTER através de três estudos de caso com diferentes parâmetros de execução, comparando o perfil dos peptídeos otimizados com o peptídeo modelo original.
- Validar a estabilidade da interação dos peptídeos otimizados com a proteína alvo por meio de simulações de dinâmica molecular (*in silico*).

## 4. Materiais e Métodos

A metodologia apresentada visa o desenvolvimento de peptídeos para a inibição competitiva da proteína *Spike* do SARS-CoV-2, elemento crucial para a infecção viral. Para este fim, foi desenvolvido um algoritmo de código-fonte aberto, denominado POTTER (*Peptide optimization tool for enhanced receptor-binding*), que utiliza princípios de algoritmos genéticos de forma aprimorada para a otimização de sequências de aminoácidos.

Diferentemente de algoritmos genéticos tradicionais, a abordagem aqui descrita inicia o procedimento com três informações: i) a estrutura da proteína alvo acoplada a seu ligante natural; ii) a sequência de aminoácidos da região que se deseja bloquear; e iii) a região de seu ligante natural que faz contato com a região delimitada. Desta forma, o procedimento inicia apenas com um indivíduo na iteração 0, e não com uma população propriamente dita.

O processo, sumarizado na Figura 12, inicia-se com a realização de várias simulações de *docking* (ancoragem) molecular para determinar a melhor posição de ligação da parte selecionada do ligante natural com a região que se deseja bloquear. Esta parte do ligante é considerada a sequência de aminoácidos de um peptídeo, o qual este procedimento visa otimizar. A melhor sequência avaliada segundo a pontuação de *docking* passa por uma análise de contatos automatizados, que visa determinar os resíduos que fazem contatos. Estes resíduos são selecionados para dar origem a um novo peptídeo, denominado “semente”.

Partindo da semente, o algoritmo entra em um ciclo iterativo. Primeiramente, criam-se “**M**” novos peptídeos, aplicando-se mutações nas quais um único resíduo da base é substituído por outro, utilizando-se uma matriz de substituição conservativa (Tabela 4). Este processo gera uma diversidade de sequências que são então avaliadas e selecionadas ao longo de “**G**” gerações.

Os novos peptídeos são avaliados através de simulações de *docking*, da quantidade de contatos com a região de interesse da proteína alvo e da contagem de interações por tipo (hidrofóbica, empilhamento aromático, atrativa, repulsiva, ponte salina ou ligação de hidrogênio). O melhor peptídeo de cada geração é utilizado como semente para a criação da nova geração. Este procedimento é repetido por “**R**” vezes para cada mutante. Todos os peptídeos gerados durante o processo são arquivados para um ranqueamento final. Vale ressaltar que os resíduos dos novos peptídeos que não realizam contatos com a proteína alvo são deletados, permitindo assim a variação no tamanho e a diversidade da população final.

A abordagem descrita maximiza a eficiência e a precisão da otimização, utilizando algoritmos genéticos de maneira mais focada e dirigida. Ao restringir as mutações com base em uma tabela de substituições, garante-se que as sequências geradas mantenham a viabilidade biológica, ao passo que a utilização do ligante natural como ponto de partida permite que a otimização seja alcançada mais rapidamente.

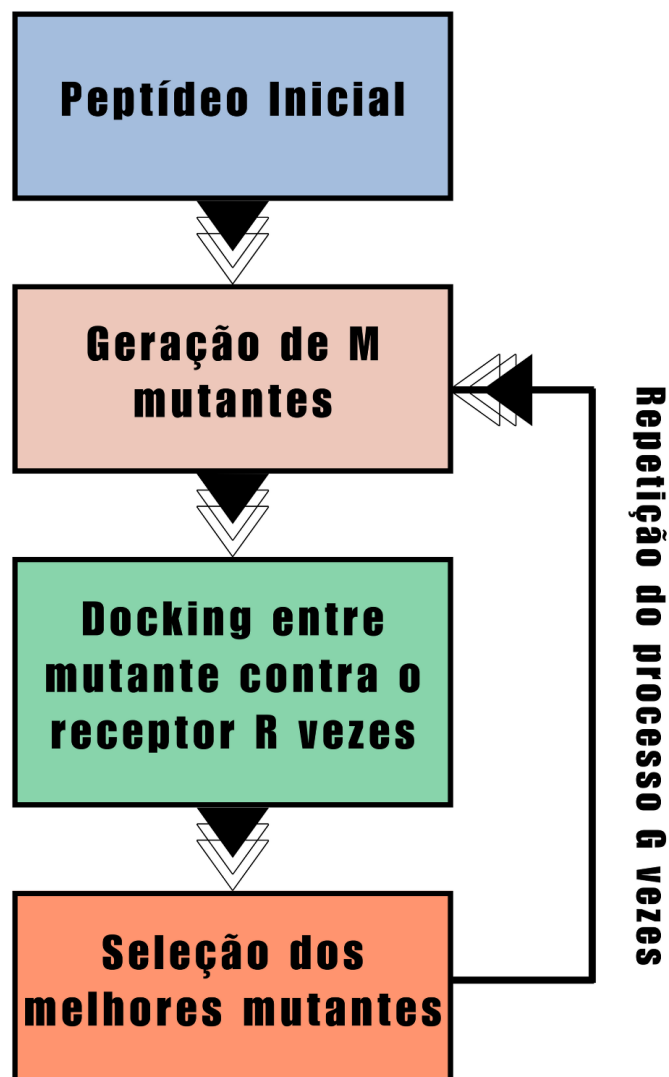


Figura 12. Fluxograma da metodologia POTTER. Onde Gerações (G) denota quantas vezes o processo se repete. Mutantes (M) o número de mutantes gerados aleatoriamente a cada geração, e Repetições (R): número de *dockings* repetidos para cada mutante.

#### 4.1 Coleta de dados

Com a proposta de projetar peptídeos com potencial para uma provável inibição competitiva presentes nas interfaces SARS-CoV-2 RBD-ACE2 da estrutura 6M0J, essa região foi admitida como ponto de partida para propor novos peptídeos. A estrutura da proteína *Spike* do SARS-CoV-2 foi recuperada do *Protein Data Bank* (PDB) [67]. Para este trabalho, escolheu-se a estrutura 6M0J [2], que contém a estrutura da proteína *Spike* do SARS-CoV2 ligada ao seu receptor, a proteína ACE2. Ao analisar a estrutura, foi possível isolar a região da proteína ACE2 que interage com a proteína *Spike* do SARS-CoV2. Em seguida, usou-se o *software* PyMOL [70] para separar essa sequência do ACE2 como um peptídeo que foi usado como modelo e ponto de partida no processo de otimização. Também foi separada a estrutura da proteína *Spike* para ser usada como alvo em nosso algoritmo.

Com as estruturas definidas, foi examinada a literatura relacionada [2] para identificar o domínio de ligação ao receptor (RBD) da proteína *Spike* do SARS-CoV2, o qual foi utilizado para definir o sítio de ligação. A escolha dos resíduos da interface foi baseada em um critério de distância: utilizou-se um corte de 6 Å, localizado aproximadamente entre os resíduos 333 a 527. Essa escolha de resíduos (K417, G446, Y449, Y453, L455, F456, A475, G476, S477, F486, N487, Y489, F490, Q493, S494, Y495, G496, Q498, T500, N501, G502, V503 e Y505) foi fundamental, pois eles representam a interface de contato biologicamente relevante. Estes mesmos resíduos foram utilizados como referência para o cálculo da métrica de "ocupação" em cada geração do algoritmo, permitindo avaliar quantitativamente a capacidade de cada peptídeo gerado de interagir com o sítio-alvo. Dentro desse intervalo, encontra-se o núcleo do domínio de ligação ao receptor (RBD) e, no mesmo, está localizado motivo de ligação ao receptor (RBM). Foram selecionados os resíduos que resultam na sequência STIEEQAKTFLDKFNHEAEDLFYQSSL presentes na interface SARS-CoV-2 RBD-ACE2, que foram adotados como ponto de partida para propor novos peptídeos (Figura 13). Os resíduos desta região também foram utilizados como referência para contagem e classificação de contatos com cada peptídeo gerado. Esta abordagem permitiu que houvesse uma medida adicional de quão bem cada peptídeo proposto se ligava e possivelmente inibindo eficazmente o RBD. As estruturas foram analisadas utilizando os *softwares* PyMOL e ChimeraX [71].



protocolo foi aplicado às estruturas utilizando PyRosetta [73], um pacote Python projetado especificamente para tais tarefas de modelagem molecular. Optou-se por utilizar essas ferramentas devido à sua capacidade de considerar a flexibilidade dos peptídeos e da proteína S, o que está mais próximo da realidade do que o *docking* rígido.

Iniciou-se enviando nosso modelo (peptídeo-RBD-ACE2) ao processo de *redocking* para estabelecer uma pontuação de base de ancoragem com a qual todas as pontuações de peptídeos subsequentes seriam comparadas. Este primeiro experimento foi realizado uma vez, e 50 modelos foram gerados e avaliados. Para calcular a pontuação, usou-se a função de pontuação REF15 no PyRosetta [73].

O *Docking* também foi aplicado a cada novo peptídeo criado. Em cada *docking* aplicado, N modelos foram gerados e pontuados. Os resultados foram armazenados em um arquivo CSV que foi posteriormente enriquecido com informações de contato para classificação final. Todas as estruturas (modelos) geradas foram salvas para análise de contato.

### 4.3 Análise de Contatos e Curadoria Manual

Para cada complexo peptídeo-proteína gerado, foi realizada uma análise quantitativa das interações intermoleculares, seguida por uma etapa qualitativa de curadoria manual. Para a análise de contatos, utilizou-se a biblioteca Signa (ainda não publicada), que considera as definições apontadas pela ferramenta *web nApoli* [24].

Os critérios para a identificação de cada tipo de interação foram os seguintes:

- Ligações de hidrogênio: Consideraram-se pares de átomos acceptor-doador a uma distância de até 3,9 Å.
- Interações hidrofóbicas: Consideraram-se pares de átomos apolares a uma distância de até 4 Å.
- Empilhamento aromático: Consideraram-se os átomos em anéis aromáticos a uma distância variando de 2-4 Å.
- Interações atrativas (iônicas): Considerou-se um par de átomos em resíduos com cargas diferentes (positiva e negativa).

- Interações repulsivas: Consideraram-se átomos de aminoácidos com a mesma carga (positiva-positiva ou negativa-negativa).
- Ponte salina: Consideraram-se pares de resíduos que realizam ligações de hidrogênio e interações atrativas ao mesmo tempo.

Adicionalmente, para complementar a avaliação puramente quantitativa do algoritmo, foi instituído um processo de Curadoria Manual (MC). Nesta etapa, os complexos gerados pelo POTTER foram primeiramente filtrados, mantendo-se apenas aqueles com pontuação de *docking* negativa, indicando uma interação energeticamente favorável. Em seguida, essas estruturas pré-selecionadas foram inspecionadas visualmente por um especialista utilizando o *software* PyMOL [70]. O critério final de seleção foi a análise da pose: o peptídeo cuja conformação e posicionamento mais se assemelhavam à interação biologicamente conhecida na interface RBD-ACE2 foi selecionado como o melhor candidato pela curadoria manual. Para uma análise complementar, utilizou-se a ferramenta *web* Arpeggio para detectar contatos e a ferramenta *web* VTR [74,75] para realizar alinhamentos estruturais.

#### 4.4 Implementação do algoritmo

O algoritmo POTTER foi implementado em linguagem Python, utilizando bibliotecas nativas para o pré-processamento de dados e a biblioteca PyRosetta [76] para modelagem e *docking* molecular.). O código-fonte está disponível em <https://github.com/LBS-UFMG/potter>.

#### 4.5 Visão Geral do Algoritmo

Nesta seção, descreve-se uma visão geral do algoritmo, ao qual se deu o nome de POTTER. O algoritmo visa retornar um peptídeo otimizado que possa se ligar melhor à proteína *Spike* do SARS-COV-2. Assim, o algoritmo recebeu como entrada (i) o peptídeo inicial (utilizamos a hélice-alfa extraída da ACE2) e (ii) o receptor inicial (a proteína *Spike*). Além disso, recebe três parâmetros de execução:

Gerações (G): quantas vezes o ciclo de otimização se repete.

Mutantes (M): número de peptídeos mutantes gerados aleatoriamente a cada geração.

Repetições (R): número de simulações de *dockings*(pose) realizadas para cada mutante.

Portanto, foram executadas as seguintes etapas: a sequência inicial do peptídeo é modelada usando PyRosetta [76]. Em seguida, realiza-se o *docking* do peptídeo contra o receptor R vezes. A pose do peptídeo com menor pontuação PyRosetta foi selecionada.

Em seguida, foi calculada a “ocupação” da pose selecionada. A ocupação é a porcentagem de resíduos peptídicos interagindo com o receptor. Esta métrica avalia se cada um dos aminoácidos do peptídeo tem pelo menos um átomo dentro de 6 Å do resíduo R413 de um dos seguintes resíduos do receptor: K417, G446, Y449, Y453, L455, F456, A475, G476, S477, F486, N487, Y489, F490, Q493, S494, Y495, G496, Q498, T500, N501, G502, V503 ou Y505 (Figura 14). O algoritmo foi executado G vezes, sendo cada execução chamada de “geração”. Além disso, na primeira geração, foram removidos resíduos peptídicos que estavam a mais de 6 Å do receptor. Essa etapa de refinamento inicial, visa focar o processo de otimização apenas na região central de interação, permitindo também que o algoritmo explore peptídeos de tamanhos variados e potencialmente menores nas gerações subsequentes.

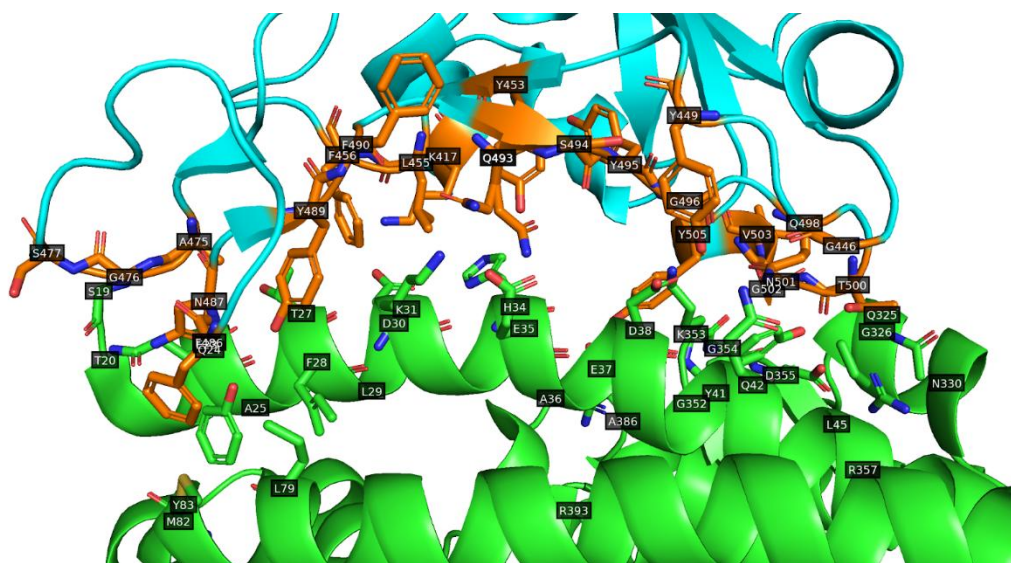


Figura 14. Interações de interface entre ACE2 (desenho ciano) e o receptor *Spike* (desenho verde). Os resíduos do ACE2 mais próximos do *Spike* são mostrados em bastões laranja (PDB ID: 6M0J). A figura foi gerada usando PyMOL de código aberto (Schrödinger, LLC).

Para cada geração restante (de 2 a G), foram feitas x mutações aleatórias na estrutura do peptídeo. O valor x foi definido aleatoriamente e pode variar de 1 resíduo até o tamanho máximo da sequência peptídica na geração 2. Esse processo foi repetido M vezes, gerando M mutantes do mesmo tamanho com mutações aleatórias. Esses mutantes foram incluídos em uma lista chamada “população”.

A Tabela 4 demonstra as substituições permitidas utilizadas ao longo deste estudo. As opções sugeridas basearam-se em substituições conservadoras na evolução, que foram aceitas nas propensões da estrutura secundária [46,67]. Elas têm sido consideradas mutações alternativas no planejamento de peptídeos, ou seja, resíduos de aminoácidos com propriedades semelhantes [46,67].

Tabela 4. Substituições permitidas pelo algoritmo POTTER. Fonte: adaptado de [46].

Aminoácidos	Substituições permitidas	Aminoácidos	Substituições permitidas
A	E, S, T	M	F, W, Y
C	S, T, V	N	G, D, Q
D	E	P	G
E	A, D	Q	N, E
F	M, W, Y	R	H, K
G	N, P	S	A, T
H	K, R	T	A, I, S
I	L, V	V	A, I, L
K	H, R	W	F, M, Y
L	I, V	Y	F, M, W

Para cada elemento da população, o peptídeo mutante foi modelado e acoplado ao receptor R vezes, em ambos os casos utilizando PyRosetta. Observe que todas as poses R foram incluídas na etapa seguinte (nas etapas 2 a G, a pontuação de encaixe não foi utilizada).

A seguir, foram calculadas a taxa de ocupação do novo complexo e os contatos realizados entre o peptídeo e o receptor. Foram avaliados os seguintes contatos: (i) Hidrofóbico, (ii) Empilhamento Aromático, (iii) Interação Atrativa, (iv) Interação do tipo Ligação de Hidrogênio, (v) Repulsivo e (vi) Pontes Salinas. Os contatos não foram utilizados pelo algoritmo para seleção, mas foram armazenados em um relatório que pode ser utilizado na tomada de decisão dos usuários do algoritmo.

O peptídeo com maior ocupação foi selecionado como ponto de partida para a próxima geração. Em seguida, a lista “população” foi excluída e o processo foi repetido G vezes.

#### **4.6 Demonstração do algoritmo**

Por exemplo, se executarmos o POTTER usando os parâmetros: gerações ( $G=5$ ), mutantes por geração ( $M=6$ ) e réplicas por mutante ( $R=3$ ), o algoritmo seria executado de forma semelhante à visão geral exibida na Figura 15. Em primeiro lugar, o peptídeo inicial sofre mutação aleatória, correspondendo ao parâmetro  $M=6$ . Cada estrutura será acoplada ao receptor (a proteína *Spike*) três vezes, conforme o parâmetro  $R=3$ . Em seguida, o algoritmo calcula os contatos realizados entre cada resíduo peptídico (ocupação). O peptídeo com maior percentual de ocupação é selecionado para uso na próxima geração. Este processo será repetido por cinco gerações. Na última geração, levanta-se a hipótese de que se obteria um peptídeo otimizado que poderia se ligar melhor ao receptor e ser um potencial inibidor.

Para entender melhor como o algoritmo funciona e seus possíveis usos no projeto de peptídeos, foram conduzidos três estudos de caso com diferentes parâmetros, discutidos na próxima seção.

**Parâmetros usados neste exemplo:**

Gerações (máx.): 5

Mutantes por geração: 6

Réplicas por mutante: 3

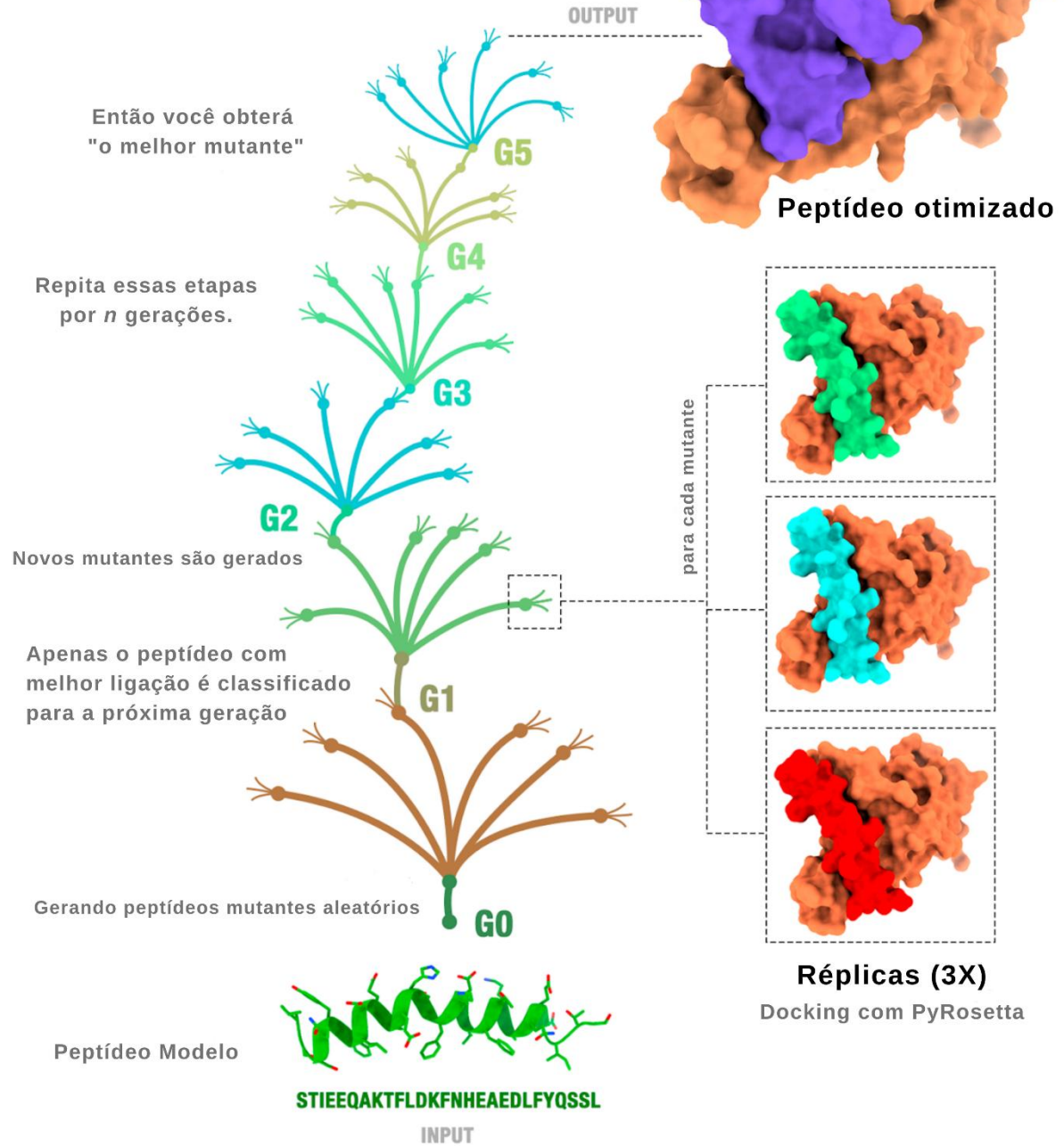


Figura 15. Visão geral do algoritmo POTTER usando os parâmetros:  $G = 5$ ,  $M = 6$  e  $R = 3$ . Em laranja temos a proteína *Spike* acoplado aos peptídeos mutantes em verde, ciano e vermelho. Em roxo temos o peptídeo otimizado selecionado.

#### 4.7 Estudo de caso

Para avaliar o algoritmo POTTER, realizaram-se três estudos de caso (Tabela 5). Os parâmetros para cada estudo foram escolhidos de forma a explorar diferentes estratégias de busca no espaço de sequências: o Estudo de Caso 1 buscou um equilíbrio entre o número de mutantes e o número de réplicas de *docking*; o Estudo de Caso 2 priorizou um maior número de gerações com um número moderado de mutantes; e o Estudo de Caso 3 utilizou uma abordagem de exploração intensiva, com um número muito elevado de mutantes por geração.

Em todos os estudos de caso, começou-se com a sequência peptídica STIEEQAKTFLDKFNHEAEDLFYQSSL (também chamada de modelo peptídico). Essa sequência foi obtida da região da estrutura ACE2 que interage com a proteína *Spike* (PDB ID: 6M0J) [2]. A sequência foi extraída usando a ferramenta de código aberto PyMOL. Os parâmetros utilizados nos estudos de caso como: Gerações (quantas vezes o processo se repete), Mutantes (número de mutantes gerados aleatoriamente a cada geração), Repetições (número de dockings repetidos para cada mutante), estão detalhados na Tabela 5.

Tabela 5. Dados para os estudos de caso.

#	PDB	Sequência peptídica inicial	Repetições	Mutações	Gerações
1	6M0J	STIEEQAKTFLDKFNHEAEDLFYQSSL	50	50	17
2	6M0J	STIEEQAKTFLDKFNHEAEDLFYQSSL	3	100	100
3	6M0J	STIEEQAKTFLDKFNHEAEDLFYQSSL	3	1000	14

#### 4.8 Validação por Dinâmica Molecular

Para a validação dos três estudos de caso, os três sistemas foram preparados usando a ferramenta VMD [75]. Estimaram-se os estados de protonação usando o conjunto de *software* APBS-PDB2PQR [71]. As simulações seguiram todas as etapas clássicas de uma preparação, solvatação, neutralização, minimização em duas etapas, equilíbrio em duas etapas e, finalmente, simulação de produção. Os sistemas foram solvatados utilizando uma caixa cúbica de 10 Å. Além disso, adicionaram-se íons Na<sup>+</sup> e Cl<sup>-</sup> usando a ferramenta de extensão VMD. Após a preparação, realizaram-se 10.000 etapas de simulação MD para cada sistema usando a

ferramenta NAMD na temperatura de 310 K [77]. Cada etapa consiste em um quadro de 20 ps, totalizando 200 ns para cada sistema. As condições de contorno periódicas (PBC) foram estabelecidas em um corte de 10 Å de cada fronteira.

As análises de RMSD foram realizadas utilizando a ferramenta VMD, através do comando *Measure rmsd* (“VMD Measure Documentation”) [75]. Gráficos 2D-RMDS (quadros todos contra todos) foram projetados usando MDAnalysis [78,79], *numpy* [80] e biblioteca Matplotlib [81].

## 5. Resultados e discussão

Inicialmente, analisou-se a estrutura da proteína *Spike* do SARS-CoV-2 complexada com a ACE2. O objetivo inicial era propor um peptídeo que se ligasse à proteína *Spike*, competindo com a ligação que ocorre com a ACE2. A hipótese é que poder-se-ia propor um peptídeo que se liga à proteína *Spike* analisando uma estrutura obtida experimentalmente (PDB: 6M0J). Assim, a parte da estrutura da ACE2 que se liga à proteína *Spike* seria um ponto de partida. Utilizou-se a ferramenta PyMOL para detectar os resíduos que conectam as duas proteínas. Percebeu-se que os principais resíduos que interagem com a proteína *Spike* estão em uma alfa hélice. Extraíu-se assim, a sequência STIEEQAKTFLDKFNHEAEDLFYQSSL correspondente a esta região. Após extrair essa sequência, realizou-se a modelagem peptídica utilizando *ColabFold*. A estrutura obtida foi denominada "modelo peptídico".

A estrutura da proteína *Spike* (cadeia E) foi extraída do PDB 6M0J. Em seguida, o modelo peptídico foi acoplado à estrutura da ACE2 usando PyRosetta. Por fim, a estrutura da *Spike* e o modelo peptídico foram utilizados como entrada para o algoritmo POTTER, que foi executado em três experimentos com os seguintes parâmetros: (i) réplicas: 50, mutantes: 50 e gerações: 17; (ii) réplicas: 3; mutantes: 100; gerações: 100; e (iii) réplicas: 3, mutantes: 1000, gerações: 14.

Para cada experimento, avaliaram-se todas as gerações e selecionaram-se os mutantes que apresentaram melhor ligação com a proteína *Spike*. O programa define o melhor peptídeo utilizando a porcentagem de resíduos em contato (maior ocupação - HO). No entanto, muitos outros fatores podem ser utilizados para escolher o melhor peptídeo. O algoritmo retorna uma série de métricas, como contatos e pontuação de encaixe. Portanto, realizou-se uma etapa subsequente de curadoria manual para selecionar o melhor peptídeo, conforme descrito na seção de Metodologia (item 4.3).

Nesta fase, filtraram-se os complexos cuja pontuação de *docking* retornada pela ferramenta PyRosetta foi negativa. Analisou-se, então, manualmente cada uma das estruturas usando a ferramenta PyMOL. O peptídeo em uma pose mais próxima do sítio de ligação foi selecionado como a melhor pose pela curadoria manual (MC). Além disso, considerou-se um caso em que o melhor peptídeo foi selecionado com base no Menor *Docking Score* (LDS). Os resultados estão resumidos na Tabela 6:

Tabela 6. Peptídeos selecionados por curadoria manual, menor pontuação de *docking* e maior ocupação versus peptídeo inicial.

Estudo de caso	CS1	CS2	CS3
<b>PDB</b>	6M0J	6M0J	6M0J
<b>Receptor</b>	Spike (6M0J:E)	Spike (6M0J:E)	Spike (6M0J:E)
<b>Sequência peptídica inicial</b>	STIEEQAKTFLDKFNHEAEDLFYQSSL	STIEEQAKTFLDKFNHEAEDLFYQSSL	STIEEQAKTFLDKFNHEAEDLFYQSSL
<b>Número de resíduos</b>		27	27
<b>Ocupação</b>		55%	47%
<b>Docking score</b>		54	170
<b>Parâmetros</b>			
<b>Replicações (por docking)</b>		50	3
<b>Mutantes (por geração)</b>		50	100
<b>Gerações(máx.)</b>		17	100
<b>Curadoria Manual (MC)</b>			
<b>Melhor peptídeo (MC)</b>	AIIEEQWGRAAEVFWAS SLASYQ	ATVEENSRTYIDHFNRTDD YWATVETFD	TEEQAKTFLDFDLFWQSSLN
<b>Geração (MC)</b>		6	14
<b>Docking score (MC)</b>		-194.71	-177.7
<b>Ocupação (MC)</b>		41%	37%
<b>Identidade</b>		29%	29%
<b>Menor Docking Score (LDS).</b>			
<b>Melhor peptídeo (LDS)</b>	AIIEEQWGRAAEVFWAS SLASYQ	SAIDAATKTYVAKMEHDDE DWFSSLAWE	TEEQAKTFLDFDLFWQSSLN
<b>Geração (LDS)</b>		6	54
<b>Docking score (LDS)</b>		-239	-233.44
<b>Ocupação (LDS)</b>		36%	40%
<b>Identidade</b>		29%	39%
<b>Maior Ocupação (HO)</b>			
<b>Melhor peptídeo (HO)</b>	ESIDAGFPRAAEVFWAS LSSMN	TTIEAQAHSMIERWPRDTAE WWTAIATMD	TEENAKSFVDFDLFYQATLQ
<b>Geração (HO)</b>		16	87
<b>Docking score (HO)</b>		1432	302
<b>Ocupação (HO)</b>		71%	68%
<b>Identidade</b>		23%	17%

Observou-se que, em alguns casos, o peptídeo selecionado pelo critério de Menor *Docking Score* (LDS) possuía uma sequência idêntica ou semelhante ao selecionado pela Curadoria Manual (MC). Nesses casos, a análise do especialista considerou que, embora a sequência fosse promissora, a melhor pose (conformação) indicada pelo *software* de *docking* não era a mais biologicamente plausível, justificando a seleção manual de uma pose alternativa.

As sequências sugeridas pelo algoritmo POTTER apresentaram semelhanças de sequência (identidade) com o peptídeo original variando entre 17% e 57%. Este resultado denota a importância da escolha do peptídeo inicial no processo de otimização. Observou-se também uma diferença no tamanho dos peptídeos selecionados em relação à sequência inicial. Isso ocorre porque o algoritmo POTTER remove, nas gerações iniciais, os resíduos que não interagem com a proteína alvo, permitindo a exploração de peptídeos menores e potencialmente mais eficientes.

Além disso, constatou-se que diferentes parâmetros de entrada (mutações, repetições e gerações), produziram peptídeos com sequências bastante distintas. Este fato sugere que a exploração do espaço de parâmetros é um passo fundamental na aplicação do algoritmo, uma vez que diferentes estratégias de busca (exploração vs. exploração) podem levar a ótimos locais distintos.

## 5.1 Simulações de dinâmica molecular

Para avaliar as ligações entre peptídeos e receptores, realizaram-se simulações de dinâmica molecular (MD) usando o programa NAMD [70]. Devido às restrições impostas pelos altos custos computacionais para a realização de simulações MD, foram escolhidos apenas os complexos sugeridos no estudo de caso 1 para a validação por DM. A escolha deste caso se deu por ele representar um equilíbrio entre o número de mutantes gerados e a quantidade de réplicas de *docking* testadas, fornecendo resultados representativos e diversificados para uma análise de estabilidade mais aprofundada. Construíram-se três sistemas: (i) proteína *Spike* complexada com o modelo peptídico; (ii) proteína *Spike* complexada com o peptídeo de menor pontuação de *docking* (LDS); e (iii) proteína *Spike* complexada com o peptídeo de maior ocupação (HO).

A simulação MD foi iniciada com base na hipótese de que o peptídeo já estava conectado à região de ligação da proteína *Spike*. O objetivo era verificar se o peptídeo permaneceria conectado à proteína após um certo tempo (definiu-se arbitrariamente o tempo para 200ns).

Portanto, realizaram-se 200ns (10.000 *frames*) de simulações MD para cada sistema. A Figura 16 ilustra três *frames* de cada simulação MD (*frame* 0 - o primeiro; *frame* 5.000 - o meio; e *frame* 9.999 - o último). O peptídeo permaneceu conectado à estrutura da proteína durante todo o período de simulação MD nos três casos.

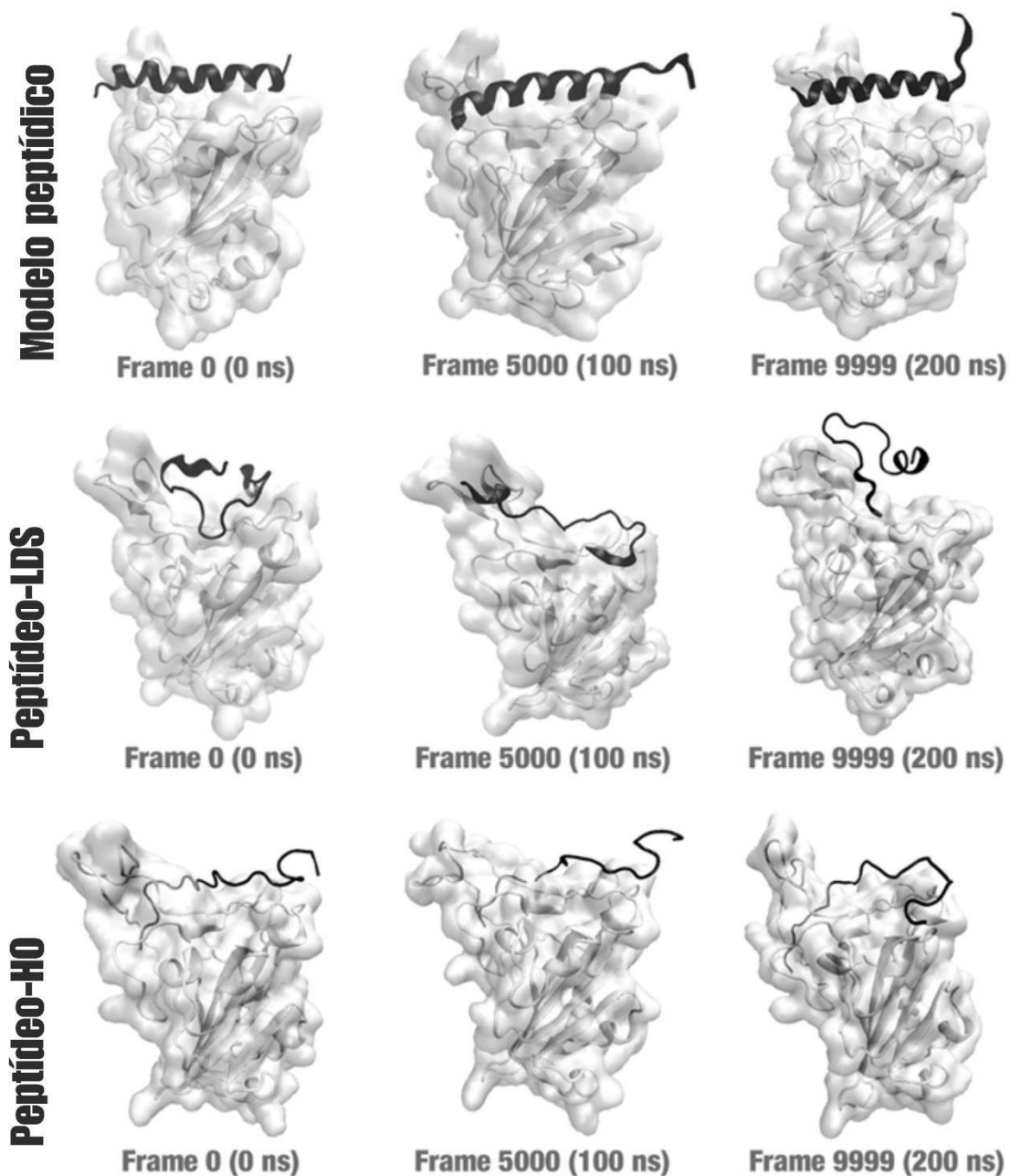


Figura 16. *Frame* 0 (0 ns), 5000 (100 ns) e 9999 (200ns) das três simulações MD: *Spike* complexado com o modelo peptídico (topo); *Spike* complexada com o peptídeo LDS (meio); e *Spike* complexada com o peptídeo HO (parte inferior).

Observa-se nesta figura que as mutações feitas no peptídeo alteraram sua estrutura secundária. Inicialmente, o peptídeo, que era um fragmento da estrutura ACE2, manteve sua estrutura de alfa-hélice. As mutações MC/LDS e HO mantiveram apenas 29% e 23% de identidade de sequência, respectivamente. Logo, a estrutura 3D assumiu novas conformações. Isso permitiu que o peptídeo se ligasse a mais resíduos da proteína *Spike* ou com uma energia de ligação mais alta (se considerarmos o *docking score*). No entanto, isto causou maior mobilidade do peptídeo. Notavelmente, observou-se que no complexo da *Spike* com o peptídeo HO, o peptídeo entranhou a proteína S. Observou-se que, no complexo da *Spike* com o peptídeo HO, o peptídeo se inseriu mais profundamente em uma cavidade na superfície da proteína S. Os detalhes serão discutidos mais detalhadamente nas próximas seções.

## 5.2 A importância dos complexos proteína-peptídeo

Neste estudo, propôs-se um método de engenharia de peptídeos que pudessem interagir com proteínas, neste caso, a proteína *Spike* do SARS-COV-2. Encontrar um peptídeo que interaja com essa proteína pode ser uma opção terapêutica no combate à COVID-19. Vários artigos relataram estratégias para inibir o SARS-COV-2 usando peptídeos. Contudo, esta não é uma tarefa trivial. A presente abordagem foi baseada numa estratégia que parte de um ponto simples: um peptídeo produzido a partir de um fragmento da proteína ACE2. Utilizando uma estratégia baseada em um algoritmo genético, a estratégia (POTTER) tenta encontrar um peptídeo otimizado para esta ligação.

A Figura 17 ilustra nosso ponto de partida (estrutura 6M0J), como extraímos o peptídeo modelo (estrutura cor de salmão) e, finalmente, como obtivemos um peptídeo mutado otimizado. Neste exemplo, as linhas indicam ligações de hidrogênio que conectam as interfaces de peptídeos e proteínas.

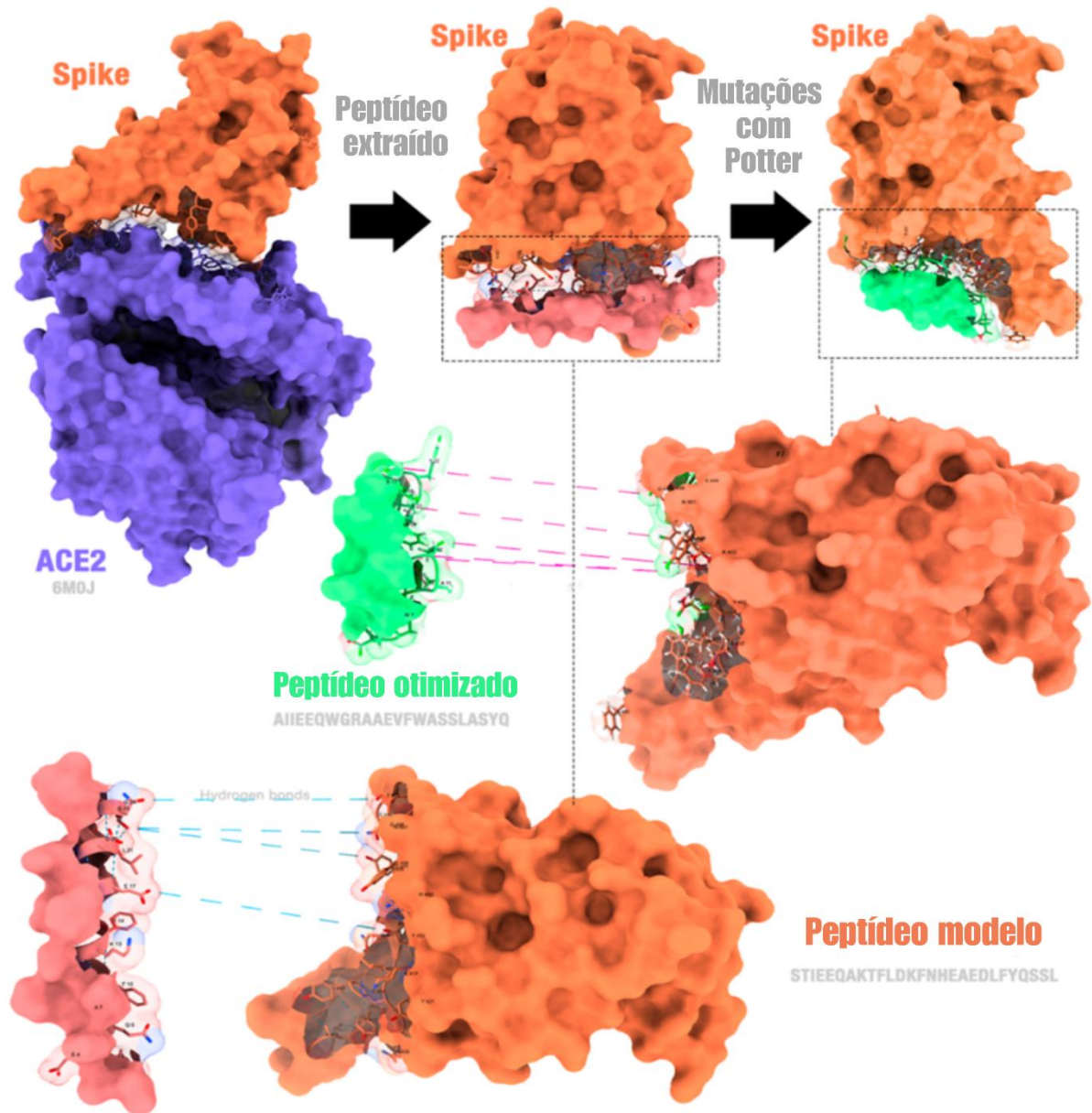


Figura 17. As interações do tipo ligações de hidrogênio são realizadas entre o peptídeo-modelo e o receptor, bem como o peptídeo curado manualmente e o receptor. As interações não foram representadas em escala real. Figura gerada usando ChimeraX.

A seleção do melhor peptídeo (ou seja, aquele mais otimizado para ligação à *Spike*) foi baseada em diversas estratégias. Na realização da Curadoria Manual (MC), o critério adotado foi a localização exata do peptídeo nas interfaces SARS-CoV-2 RBD-ACE2, aliada a um maior número de interações iônicas, interações do tipo ligações de hidrogênio e empilhamento aromático. Considerou-se este parâmetro mais robusto, pois, diferentemente dos outros critérios puramente quantitativos como o Menor *Docking Score* (LDS) e a Maior Ocupação (HO), a

curadoria manual incorpora uma análise qualitativa da pose, garantindo que a orientação do peptídeo seja biologicamente relevante para a inibição. Quanto aos demais parâmetros, os modelos peptídicos também estão na interface, mas nem todos possuem a mesma cobertura, como é o caso do *Lowest Docking Score* (LDS). Alguns parâmetros estão resumidos na Tabela 7.

Na verdade, o *docking score* é um critério que tem sido amplamente criticado na literatura [82]. Muitas vezes, a melhor pose nem sempre é aquela com a menor pontuação de encaixe. Isso porque essa pontuação é definida com base em critérios estabelecidos pela ferramenta de *docking* (no caso, PyRosetta). No entanto, apesar de suas limitações, ele não deve ser descartado. Combinado com outras estratégias, como o número de contatos realizados ou a análise manual do local de ligação, pode-se utilizar este critério na tomada de decisão. Foi o que se fez para selecionar o melhor peptídeo com base na curadoria manual.

Uma premissa geralmente aceita é que as proteínas que interagem têm um alto grau de complementaridade superficial, mas a complementaridade eletrostática também é observada [83]. As interfaces proteína-peptídeo são ambientes complexos; seria inviável detalhar as interações entre os resíduos com base em análises físico-químicas. Uma das alternativas é considerar os estudos já conhecidos sobre estrutura e interações proteína-peptídeo, bem como a preferência de contato resíduo-resíduo, ou seja, derivar potenciais de interação a partir das propensões dos resíduos em interagirem entre si, como observado na coleção de estruturas conhecidas de complexos de proteínas [84].

Observa-se que resíduos com cargas opostas tendem a entrar em contato entre si. A orientação relativa dos grupos carregados de ambos os resíduos sugere atração eletrostática entre eles. Isto se deve aos grupos carregados que estão espacialmente próximos uns dos outros nos pares observados, sugerindo que eles interagem entre si. Há três aminoácidos que podem adquirir carga positiva (cátion): lisina, arginina e histidina, e há dois aminoácidos que podem adquirir carga negativa (ânions): aspartato e glutamato. As interações atrativas observadas nas interfaces proteína-peptídeo são devido a aminoácidos polares carregados com cargas opostas, entre Arg-Glu, Arg-Asp, Lys-Glu. Pares de aminoácidos positiva e negativamente carregados podem realizar interações iônicas e ligações de hidrogênio ao mesmo tempo, proporcionando uma interação mais forte denominada de ponte salina.

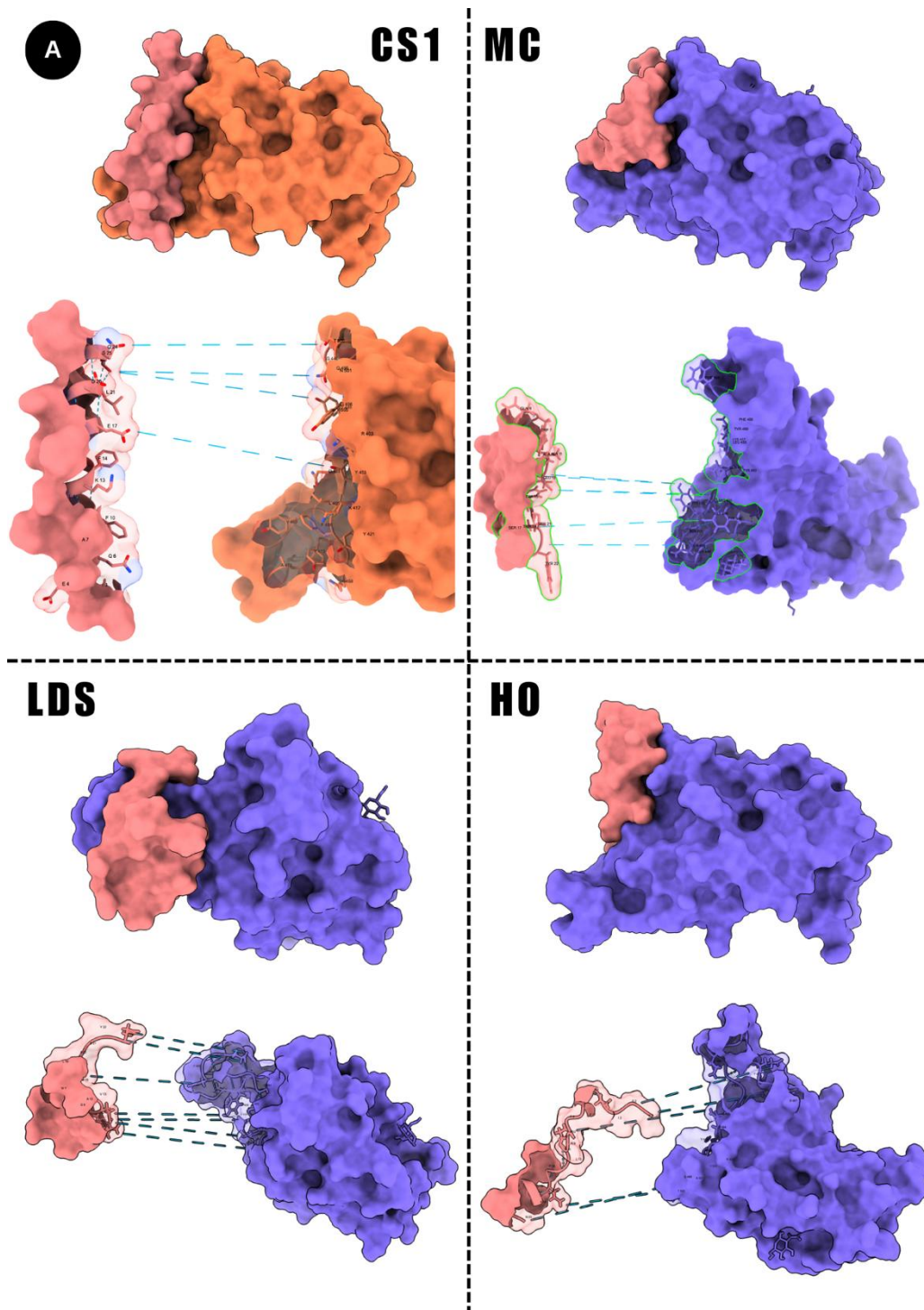


Figura 18. Perfil de interações estudo de caso 1. Podemos observar no primeiro quadrante A o peptídeo modelo (inicial) onde a ocupação é de 55% e um total de 27 resíduos interagindo entre Proteína *Spike* e o peptídeo inicial. O peptídeo por MC apresenta uma ocupação de 41%. No modelo *LDS* apresenta uma ocupação de 36%. No modelo *HO* apresenta uma ocupação de 71%.

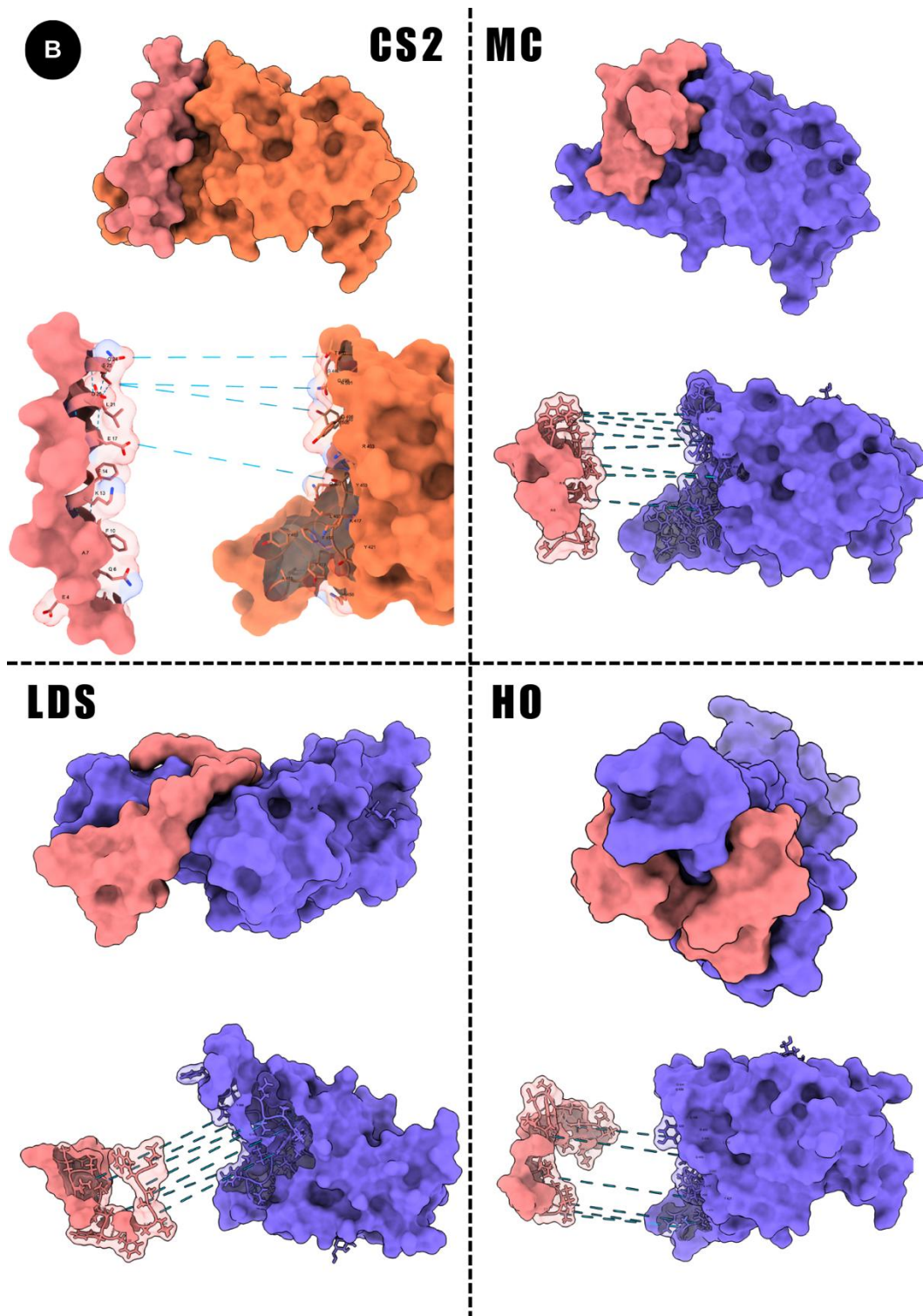


Figura 19. Perfil de interações estudo de caso 2. Podemos observar no primeiro quadrante B o peptídeo modelo (inicial) onde a ocupação é de 47% e um total de 27 resíduos interagindo entre Proteína *Spike* e o peptídeo inicial. Já o peptídeo MC apresenta uma ocupação de 37%. No modelo LDS apresenta uma ocupação de 40%. No modelo HO apresenta uma ocupação de 68%.

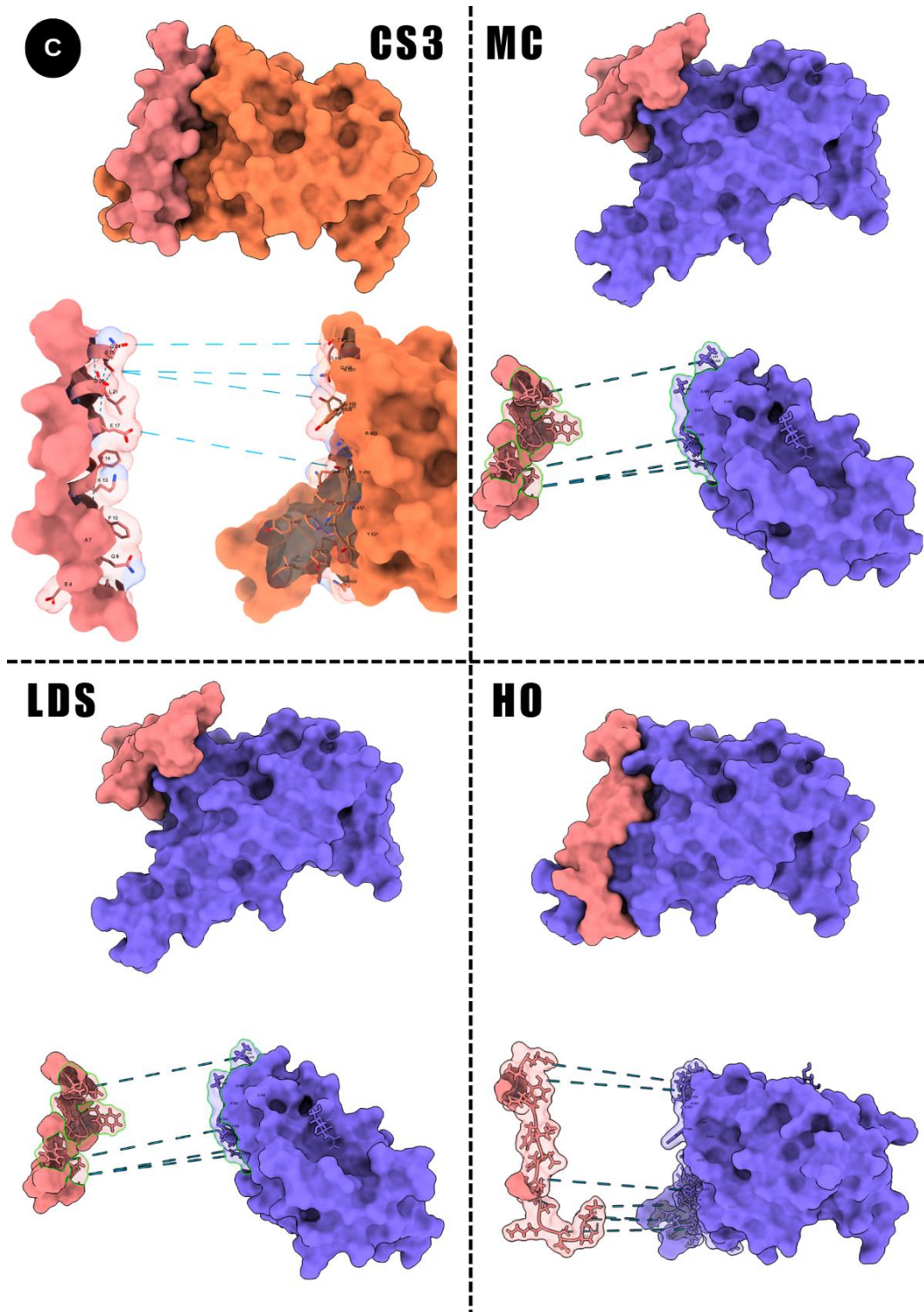


Figura 20. Perfil de interações estudo de caso 3. Podemos observar no primeiro quadrante C o peptídeo modelo (inicial) onde a ocupação é de 42% e um total de 27 resíduos interagindo entre Proteína *Spike* e o peptídeo inicial. Já o peptídeo MC apresenta uma ocupação de 33%. No modelo LDS apresenta uma ocupação de 30%. No modelo HO apresenta uma ocupação de 67%.

Nas interações de empilhamento, as nuvens eletrostáticas dos anéis se alinham, resultando em uma interação estável e não covalente. Essa forma de interação é especialmente

favorecida pelo efeito hidrofóbico originado pelas regiões apolares das moléculas. Nota-se a presença de resíduos como tirosina e triptofano, esses dois aminoácidos possuem cadeias laterais anfipáticas com volume estérico e flexibilidade suficientes, permitindo a formação da interface de ligação para otimizar a complementaridade, uma vez que o efeito hidrofóbico desempenha um papel importante no enovelamento e estabilidade da proteína.

Vários estudos relatam o papel central do efeito hidrofóbico no dobramento e estabilidade de proteínas é também uma tendência de moléculas apolares de se agruparem para minimizar seu contato com a água. Nos dados apresentados observa-se uma quantidade significativa de interações hidrofóbicas na interface proteína-peptídeos, isso sugere que o efeito hidrofóbico também estabiliza os complexos proteicos. Glaser, *et al.* (2001) relata que pares de resíduos polares e carregados sugerem que o efeito hidrofóbico deriva pelo menos parcialmente de suas preferências de emparelhamento. Em nosso estudo as interações do tipo hidrofóbicas apresentam uma predominância entre resíduos aromáticos da proteína com resíduos polares dos peptídeos. Os resíduos de aminoácidos predominantes são a tirosina e o triptofano. Esses dois aminoácidos apresentam cadeias laterais anfipáticas e com volume estérico e flexibilidade suficientes permitindo que a interface de ligação se molde para otimizar a complementaridade.

Kiran, *et al.* (2023) relata em seu estudo que a composição de resíduos das interfaces proteicas dificilmente difere de outras regiões da proteína [85]. Os resíduos hidrofóbicos tendem a ocorrer com mais frequência em grandes superfícies de contato, enquanto os resíduos polares prevalecem em pequenas superfícies em interações proteína-peptídeos.

A relação entre Constante de dissociação ( $K_d$ ) e Energia livre de Gibbs ( $\Delta G$ ) fornece uma perspectiva importante sobre a estabilidade de interações moleculares, como as que ocorrem entre receptores e ligantes. Valores menores de  $K_d$  indica que o complexo se dissocia com dificuldade, resultando em um  $\Delta G$  negativo e indicando a formação do complexo é espontânea. Os valores apresentados na Tabela 7, foram calculados utilizando a *Web Server Prodigy* [86,87].

Tabela 7. Peptídeos selecionados por curadoria manual, menor pontuação de docking e maior ocupação versus peptídeo inicial, onde demonstramos o perfil de interações de cada um selecionado. Os valores da constante de dissociação ( $K_d$ ) indica que a proteína tem uma alta afinidade pelo ligante, ou seja, o complexo proteína-ligante é estável e se forma facilmente.

Estudo de caso	CS1	CS2	CS3
<b>PDB</b>	6M0J	6M0J	6M0J
<b>Receptor</b>	Spike (6M0J:E)	Spike (6M0J:E)	Spike (6M0J:E)
<b>Sequência peptídica inicial</b>	STIEEQAKTFLDKFNHEAEDLFYQSSL	STIEEQAKTFLDKFNHEAEDLFYQSSL	STIEEQAKTFLDKFNHEAEDLFYQSSL
<b>ΔG (kcal/mol)</b>		-9.8	-9.8
<b>K d (M) em °C</b>	1.9x10 <sup>-08</sup>	1.9x10 <sup>-08</sup>	1.9x10 <sup>-08</sup>
<b>Interações</b>			
<b>Atrativas</b>		5	5
<b>Aromáticas</b>		0	0
<b>Pontes de Hidrogênios</b>		21	21
<b>Hidrofóbicas</b>		14	14
<b>Curadoria Manual (MC)</b>			
<b>Melhor peptídeo (MC)</b>	AIIEEQWGRAAEVFWAS SLASYQ	ATVEENSRTYIDHFNRAATDD YWATVETFD	TEEQAKTFLDFDLFWQSSLN
<b>ΔG (kcal/mol)</b>		-10	-9.5
<b>K d (M) em °C</b>	4.5x10 <sup>-07</sup>	1.20x10 <sup>-10</sup>	1.00x10 <sup>-07</sup>
<b>Interações</b>			
<b>Atrativas</b>		0	4
<b>Aromáticas</b>		2	11
<b>Pontes de Hidrogênios</b>		6	5
<b>Hidrofóbicas</b>		29	37
<b>Menor Docking Score (LDS).</b>			
<b>Melhor peptídeo (LDS)</b>	AIIEEQWGRAAEVFWAS SLASYQ	SAIDAATKTYVAKMEHDDE DWFSSLAWE	TEEQAKTFLDFDLFWQSSLN
<b>ΔG (kcal/mol)</b>		-10	-9.5
<b>K d (M) em °C</b>	4.5x10 <sup>-07</sup>	7.20x10 <sup>-08</sup>	1.00x10 <sup>-07</sup>
<b>Interações</b>			
<b>Atrativas</b>		0	4
<b>Aromáticas</b>		2	11
<b>Pontes de Hidrogênios</b>		6	5
<b>Hidrofóbicas</b>		29	37
<b>Maior Ocupação (HO)</b>			
<b>Melhor peptídeo (HO)</b>	ESIDAGFPRAAEFYESA LSSMN	TTIEAQAHSMIERWPRDTAE WWTAIATMD	TEENAKSFVDFDLFYQATLQ
<b>ΔG (kcal/mol)</b>		-13.6	-12.3
<b>K d (M) em °C</b>	1.00x10 <sup>-10</sup>	9.70x10 <sup>-11</sup>	1.00x <sup>-09</sup>
<b>Interações</b>			
<b>Atrativas</b>		1	0
<b>Aromáticas</b>		2	11
<b>Pontes de Hidrogênios</b>		12	10
<b>Hidrofóbicas</b>		45	36

### 5.3 Comparação com outras propostas de engenharia de peptídeos descritas na literatura

Diversas metodologias computacionais e experimentais foram desenvolvidas para projetar novos peptídeos ou melhorar os existentes que podem inibir a proteína *Spike* do SARS-CoV-2, ligando-se ao seu domínio de ligação ao receptor (RBD). Por exemplo, Valiente *et al.* (2021) empregaram métodos *in vitro* para criar inibidores do peptídeo D que imitam a hélice de ligação  $\alpha 1$  da ACE2. Seus melhores projetos exibiram afinidades de ligação ao RBD de 29 e 31 nM, respectivamente, e bloquearam a infecção de células Vero por SARS-CoV-2 com valores IC50 de 5,76 e 6,56  $\mu\text{M}$  [88]. Métodos *in silico* também têm sido amplamente utilizados, com graus variados de sucesso. Sitthiyotha e Chunsrivirod (2021) utilizaram simulações de dinâmica molecular para orientar o projeto de peptídeos com afinidades de ligação aprimoradas ao RBD do SARS-CoV-2 em comparação com o receptor ACE2 humano, identificando mutações que melhoram a ligação do peptídeo ao RBD [89]. Da mesma forma, Sun *et al.* (2023) empregaram simulações de dinâmica molecular de múltiplas réplicas para investigar a ligação competitiva de proteínas ACE2 projetadas, especificamente variantes 3N39 e 3N94, ao RBD do SARS-CoV-2. Ao modificar os tipos de resíduos próximos à interface de ligação, às proteínas ACE2 projetadas alteram a distribuição do potencial eletrostático e reconfiguram a rede de ligações de hidrogênio, resultando na ligação modificada ao RBD [90].

O estudo de B. Robson, descreveu o uso de sistemas Q-UDEL para facilitar o acesso a sequências de aminoácidos que estão bem conservadas em muitos coronavírus, incluindo SARS-CoV-2, e para interagir com ferramentas de bioinformática padrão disponíveis publicamente [91]. A sequência KRSFIEDLLFNKV foi considerada particularmente bem conservada em seu estudo e corresponde à região ao redor de um dos locais de clivagem conhecidos do vírus SARS que, acredita-se, serem necessários para ativar o vírus para entrar nas células [78,91,92]. Sua proposta é um epítopo específico de vacina sintética e um agente peptidomimético. Infelizmente, na região utilizada, regiões da superfície das proteínas ficam expostas e são mais suscetíveis a mutações [84,93]. Isso demonstra que a proposta do autor pode se tornar ineficiente devido ao escape de mutações, ou tais mutações alteram uma vez o mecanismo de entrada viral, pois se sabe que os coronavírus reconhecem diferentes receptores e possuem mecanismos de entrada de acordo com o gênero ao qual pertencem.

Neste estudo, apresenta-se uma ferramenta que propõe a engenharia de peptidomiméticos baseados na região de interação SARS-CoV-2-RBD/ACE2, uma região menos suscetível a mutações, para que no desenvolvimento de futuros protótipos terapêuticos, não se tornem obsoletos ou ineficazes ao longo do tempo. Outro fator é a capacidade de

direcionar a engenharia para um gênero específico. A ferramenta POTTER utiliza o receptor como ponto de partida, mitigando assim possíveis interferências ou interações indesejadas. A ferramenta POTTER foi projetada para atingir interações específicas, usando o receptor ACE2 como ponto de partida. Esta abordagem é semelhante a outros estudos onde varreduras computacionais são usadas para criar peptídeos a partir da análise da região de interação RBD/ACE2, minimizando assim potenciais interferências ou interações indesejadas.

#### 5.4 Importância dos parâmetros POTTER

Conforme discutido anteriormente, realizaram-se três estudos de caso com POTTER utilizando parâmetros diferentes. Determinar os melhores parâmetros é uma tarefa complexa. Portanto, realizaram-se estudos de caso variando o número de gerações, mutações por geração e réplicas para cada experimento de *docking*. Aumentar consideravelmente todos esses parâmetros acarreta um alto custo computacional, o que pode inviabilizar a execução do POTTER.

A Figura 21 e 22 ilustram os resultados das pontuações de *docking* e da ocupação para cada geração nos três estudos de caso. Nota-se que para pontuações de encaixe, valores mais baixos indicam melhor interação. Por outro lado, quanto maior a ocupação (número de resíduos realizando contato), melhor será a interação. A seta vermelha indica em qual geração o melhor peptídeo foi selecionado. Nota-se que os peptídeos com melhores resultados não são encontrados na última geração. A cada geração, o peptídeo escolhido como melhor resultado é selecionado para a próxima geração, mas não é utilizado para comparações nessa geração. Apenas mutantes deste peptídeo são utilizados. Isto se deve a um comportamento estabelecido pelo POTTER para remover o viés do algoritmo genético, no qual o algoritmo atinge um mínimo de energia logo no início. Se o peptídeo descartado em uma dada geração for, no final, o melhor peptídeo em todo o processo, o usuário poderá selecioná-lo após a conclusão da execução do algoritmo, pois todos os resultados são arquivados em formato CSV.



Figura 21. Seleção dos melhores peptídeos por gerações para os estudos de caso 1, 2 e 3 de acordo com a maior ocupação (HO). A geração em que cada melhor peptídeo é obtido está destacada por uma seta vermelha.



Figura 22. Seleção dos melhores peptídeos por gerações para os estudos de caso 1, 2 e 3 de acordo com a menor *docking score* (LDS). A geração em que cada melhor peptídeo é obtido está destacada por uma seta vermelha.

A figura 22 mostra os resultados alcançados em cada estudo de caso e aponta os efeitos que podemos esperar de cada parâmetro do algoritmo. Como podemos ver nos gráficos LDS e HO no estudo de caso 1 ( $G=17$ ,  $M=50$ ,  $R=50$ ), equilibrar a exploração do espaço de sequências (número de peptídeos mutantes) com a profundidade da análise de cada mutante (número de réplicas testadas) resultou em um comportamento pouco consistente do algoritmo, com a pontuação de *docking* mais baixa sendo alcançada relativamente cedo e um ligeiro aumento geral na ocupação ao longo das gerações. Porém, após a décima geração, a pontuação de *docking* aumentou, atingindo um patamar. O maior valor de ocupação só foi encontrado na décima sexta geração. Recomenda-se que recomendamos ter um valor de pontuação de *docking* baixo e um valor de ocupação alto para escolher o melhor peptídeo.

O estudo de caso 2 ( $G=100$ ,  $M=100$ ,  $R=3$ ) mostra como o uso de poucas replicações com poucos modelos pode levar a resultados instáveis (considerando a pontuação de *docking*). Enquanto o primeiro estudo de caso usa 2.500 *dockings* por geração (50 réplicas para 50 mutantes levando a mais gerações, mas a uma avaliação menos exaustiva de cada mutante, o que pode fazer com que bons resultados sejam perdidos em algumas gerações. Por outro lado, ao diminuir a demanda computacional por geração, é possível iterar por mais gerações. Resumidamente, menos réplicas reduzem consideravelmente o custo do processamento computacional, o que permite ao algoritmo atingir mais gerações em menos tempo. Cada estudo de caso foi executado em um computador com 32 núcleos e 96 GB de RAM. Os tempos de execução refletem o tempo total decorrido do início ao fim dos experimentos em um ambiente de computação de alto desempenho, incluindo não apenas o tempo de processamento, mas também tempos de espera em filas de execução. O estudo de caso 1 levou aproximadamente 4 meses para 17 gerações; o estudo de caso 2, 45 dias para 100 gerações; e o estudo de caso 3, 5 meses para 14 gerações.

Finalmente, no estudo de caso 3 ( $G=14$ ,  $M=1000$ ,  $R=3$ ), podemos ver como uma abordagem de exploração intensiva do espaço de sequências (número de mutações elevado) pode ser benéfica para encontrar as melhores soluções antecipadamente. Usando 1000 mutantes em cada geração, identificaram-se bons candidatos apresentando pontuações mínimas de *docking* e ocupação razoavelmente boa dentro de duas gerações. Também é importante notar que, como os resultados são encontrados precocemente, a identidade da sequência original também é a mais elevada.

## 5.5 Insights obtidos pelas simulações MD

As simulações MD demonstraram que os peptídeos sugeridos por POTTER para o estudo de caso 1 foram mantidos na região mais próxima do sítio de ligação da proteína *Spike* durante os 200ns de simulação. Estes resultados corroboram parcialmente os resultados de *docking* e ocupação, indicando que os peptídeos propostos podem ser bons alvos para inibir a proteína *Spike*. No entanto, mais experimentos *in silico*, *in vitro* e *in vivo* devem ser realizados para confirmar isso. Além disso, os resultados das simulações MD devem ser avaliados cuidadosamente. Em particular, ao se comparar a mobilidade do peptídeo quando se liga à proteína. A Figura 23 apresenta os gráficos de linhas 2D-RMSD, RMSD e RMSF dos três sistemas analisados.

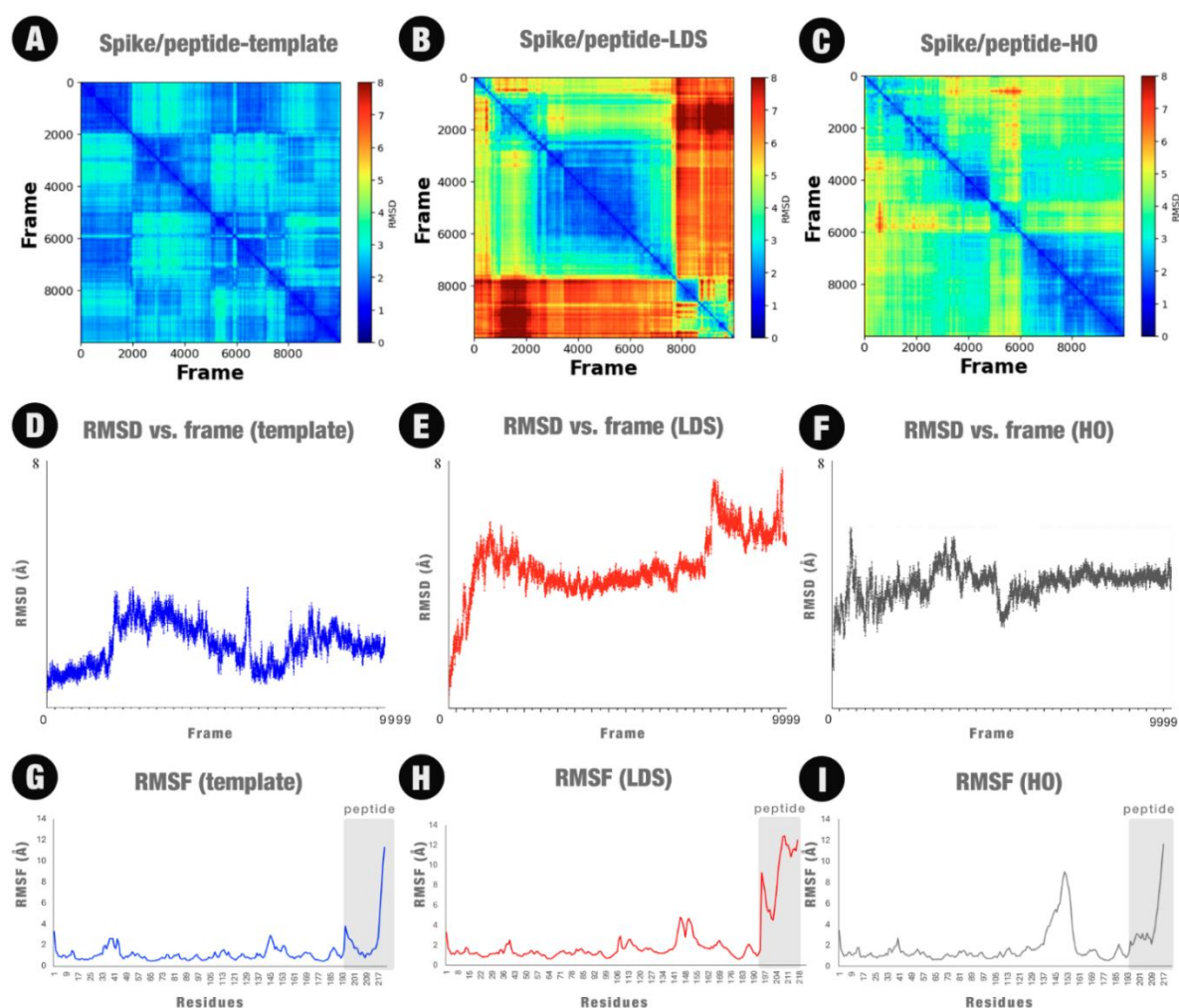


Figura 23. Gráfico de linhas 2D-RMSD, RMSD e RMSF para as três simulações MD: pico complexado com o modelo peptídico (A, D e G, respectivamente); *Spike* complexada com o peptídeo LDS (B, E e H, respectivamente); e *Spike* complexado com o peptídeo HO (C, F e I, respectivamente).

Quando se observa o gráfico 2D-RMSD, pode-se ver que a ligação entre o peptídeo inicial e a *Spike* tem uma mobilidade menor (Figura 23A) em comparação com o complexo *Spike*-LDS e o complexo *Spike* -HO. Embora um platô não seja claramente observado no gráfico RMSD vs. *frame* (Figura 23D), o pico na variação do RMSD foi de 4 Å, que é inferior aos picos obtidos nos outros sistemas. Além disso, podem-se observar *clusters* que representam as conformações mais adotadas com base nos quadrados mais azuis formados na região próxima à diagonal central. Na Figura 23A-C, as cores quentes podem indicar maior mobilidade, enquanto as cores frias podem indicar menor mobilidade.

Como o peptídeo selecionado na curadoria manual (MC) para o Estudo de Caso 1 possui a mesma sequência do peptídeo LDS, a análise de sua simulação (Figura 23B, E, H) é a mesma. Observa-se que este peptídeo apresentou maior variação no RMSD na comparação dos 10.000 *frames*. Confirmam-se visualmente as variações nas poses deste peptídeo quando se analisa a Figura 16 (linha do meio). Tais alterações nas posições de ligação à proteína podem indicar que o peptídeo possui alta flexibilidade no sítio de ligação.

Isso sugere que, embora o processo de otimização tenha sido bem-sucedido em aumentar as métricas de afinidade previstas (*in silico*), ele pode ter introduzido uma maior flexibilidade conformacional. Este resultado indica um possível *trade-off* entre a afinidade de ligação e a estabilidade estrutural do complexo, um fator que deverá ser mais explorado em estudos futuros com simulações de maior duração e outras técnicas biofísicas.

## 6. Conclusão

Neste estudo, propôs-se um método para a engenharia e otimização de peptídeos com vistas a conferir capacidade de inibição direcionada a proteínas específicas. Para tal finalidade, foi desenvolvido um algoritmo de código-fonte aberto combinando *docking* molecular, algoritmos genéticos e análise automatizada de contatos. Testes *in silico*, como a dinâmica molecular, corroboram a efetividade do algoritmo, o que sugere que este é muito promissor para tal finalidade.

A metodologia possui alta especificidade com o sítio selecionado para inibição, uma vez que o peptídeo modelo utilizado como ponto de partida para propor novos peptídeos é selecionado de acordo com o alvo de interesse. A seleção do melhor peptídeo (ou seja, aquele mais otimizado) pode ser baseada em diversas estratégias, como perfil de ligação, a porcentagem de resíduos em contato no sítio de ligação especificado e a pontuação de *docking*. Os três estudos de caso apresentados foram submetidos a simulações de dinâmica molecular\*\*, nas quais\*\* o peptídeo permaneceu conectado à estrutura da proteína durante todo o período de simulação em todos os casos.

Do ponto de vista computacional, vislumbram-se pelo menos dois caminhos que podem ser tomados por trabalhos futuros: o primeiro deles diz respeito à otimização computacional do método. Durante o desenvolvimento, observou-se que reduzir o número de réplicas reduz consideravelmente o custo do processamento computacional, o que permite ao algoritmo atingir mais gerações em menos tempo. Esta abordagem, no entanto, reduziria a exploração de poses para um mesmo peptídeo, diminuindo assim a probabilidade de a melhor conformação encontrada ser de fato a melhor conformação possível. Uma alternativa para melhorar seu desempenho é gerar uma versão paralelizada das funções, tornando assim o processo mais ágil e utilizando os recursos computacionais de forma mais eficiente.

O segundo caminho consiste em utilizar algoritmos de *docking* que sejam mais computacionalmente eficientes ou com maior assertividade no processo, garantindo assim melhores resultados. Para isso, é possível testar algoritmos e funções de pontuação mais recentes, como métodos baseados em inteligência artificial.

Como vantagem, a ferramenta desenvolvida utiliza o receptor como ponto de partida, mitigando assim possíveis interferências ou interações indesejadas e foi projetada para atingir

interações específicas. Além disso, como perspectivas para o futuro, pretende-se sintetizar os peptídeos propostos, seguidos de experimentos para verificar as interações entre peptídeos e ligantes.

Adicionalmente, a ferramenta POTTER, que foi originalmente projetada para a otimização de peptídeos contra a proteína *Spike* do SARS-CoV-2, pode ser adaptada para diversas outras aplicações na pesquisa biomédica. Por exemplo, poderia ser utilizada no desenvolvimento de vacinas, permitindo o *design* de peptídeos miméticos de epítomos de antígenos para imunização contra diferentes patógenos. Poderia, ainda, ser aplicada na criação de terapias direcionadas, projetando peptídeos capazes de se ligar seletivamente a receptores celulares específicos, úteis em tratamentos contra o câncer ou em estudos de modulação de vias metabólicas e sinalização celular. Espera-se também que as estratégias computacionais aqui aplicadas possam ser estendidas ao projeto de peptídeos utilizados para aplicações farmacêuticas.

## 7. Referências Bibliográficas

1. WHO Coronavirus (COVID-19) Dashboard.
2. Lan J, Ge J, Yu J *et al.* Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature* 2020;**581**:215–20.
3. Amoretti M, Amsler C, Bonomi G *et al.* Production and detection of cold antihydrogen atoms. *Nature* 2002;**419**:456–9.
4. Carvalho PPD, Alves NA. Featuring ACE2 binding SARS-CoV and SARS-CoV-2 through a conserved evolutionary pattern of amino acid residues. *Journal of Biomolecular Structure and Dynamics* 2022;**40**:11719–28.
5. Jackson CB, Farzan M, Chen B *et al.* Mechanisms of SARS-CoV-2 entry into cells. *Nat Rev Mol Cell Biol* 2022;**23**:3–20.
6. Troyano-Hernández P, Reinoso R, Holguín Á. Evolution of SARS-CoV-2 Envelope, Membrane, Nucleocapsid, and Spike Structural Proteins from the Beginning of the Pandemic to September 2020: A Global and Regional Approach by Epidemiological Week. *Viruses* 2021;**13**:243.
7. Kumar M, Al Khodor S. Pathophysiology and treatment strategies for COVID-19. *Journal of Translational Medicine* 2020;**18**:353.
8. Xia S, Zhu Y, Liu M *et al.* Fusion mechanism of 2019-nCoV and fusion inhibitors targeting HR1 domain in spike protein. *Cell Mol Immunol* 2020;**17**:765–7.
9. Zhu Y, Yu D, Yan H *et al.* Design of Potent Membrane Fusion Inhibitors against SARS-CoV-2, an Emerging Coronavirus with High Fusogenic Activity. *J Virol* 2020;**94**:e00635-20.
10. Ling R, Dai Y, Huang B *et al.* In silico design of antiviral peptides targeting the spike protein of SARS-CoV-2. *Peptides* 2020;**130**:170328.
11. Yan R, Zhang Y, Li Y *et al.* Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2. *Science* 2020;**367**:1444–8.
12. Tai W, He L, Zhang X *et al.* Characterization of the receptor-binding domain (RBD) of 2019 novel coronavirus: implication for development of RBD protein as a viral attachment inhibitor and vaccine. *Cell Mol Immunol* 2020;**17**:613–20.

13. Chen Y, Liu Q, Guo D. Emerging coronaviruses: Genome structure, replication, and pathogenesis. *J Med Virol* 2020;**92**:418–23.
14. Martins PM, Santos LH, Mariano D *et al.* Propedia: a database for protein–peptide identification based on a hybrid clustering algorithm. *BMC Bioinformatics* 2021;**22**:1.
15. Martins P, Mariano D, Carvalho FC *et al.* Propedia v2.3: A novel representation approach for the peptide-protein interaction database using graph-based structural signatures. *Front Bioinform* 2023;**3**, DOI: 10.3389/fbinf.2023.1103103.
16. Lau JL, Dunn MK. Therapeutic peptides: Historical perspectives, current development trends, and future directions. *Bioorganic & Medicinal Chemistry* 2018;**26**:2700–7.
17. Hamley IW. Peptides for Vaccine Development. *ACS Appl Bio Mater* 2022;**5**:905–44.
18. David L. Nelson (Autor), Michael M. Cox (Autor), Carla Dalmaz (Tradutor), Carlos Termignoni. *Princípios de Bioquímica de Lehninger*.
19. Voet D, Voet JG, Pratt CW. *Principles of Biochemistry*. 4. ed., internat. student version. Singapore: Wiley, 2013.
20. Nomenclature of  $\alpha$ -amino acids (Recommendations, 1974). *Biochem J* 1975;**149**:497.b1.
21. Qin L, Cui Z, Wu Y *et al.* Challenges and Strategies to Enhance the Systemic Absorption of Inhaled Peptides and Proteins. *Pharm Res* 2023;**40**:1037–55.
22. Pereira JG de C. Caracterização dos aminoácidos da interface proteína-proteína com maior contribuição na energia de ligação e sua predição a partir dos dados estruturais. 2012.
23. Silva MFM, Martins PM, Mariano DCB *et al.* Proteingo: Motivation, user experience, and learning of molecular interactions in biological complexes. *Entertainment Computing* 2019;**29**:31–42.
24. Fassio AV, Santos LH, Silveira SA *et al.* nAPOLI: A Graph-Based Strategy to Detect and Visualize Conserved Protein-Ligand Interactions in Large-Scale. *IEEE/ACM Transactions on Computational Biology and Bioinformatics* 2020;**17**:1317–28.
25. Martins PM, Mayrink VD, de A. Silveira S *et al.* How to compute protein residue contacts more accurately? *Proceedings of the 33rd Annual ACM Symposium on Applied Computing*. New York, NY, USA: Association for Computing Machinery, 2018, 60–7.

26. da Silveira CH, Pires DEV, Minardi RC *et al.* Protein cutoff scanning: A comparative analysis of cutoff dependent and cutoff free methods for prospecting contacts in proteins. *Proteins* 2009;**74**:727–43.
27. Paixão VM, Santos LM dos, Bastos LL *et al.* *Bioinformática com Jupyter Notebook*. Alfahelix Publicações, 2023.
28. Silva M, Martins P, Mariano D *et al.* Proteingo: Motivation, user experience, and learning of molecular interactions in biological complexes. *Entertainment Computing* 2018;**29**, DOI: 10.1016/j.entcom.2018.11.001.
29. Zhang Y-Z, Holmes EC. A Genomic Perspective on the Origin and Emergence of SARS-CoV-2. *Cell* 2020;**181**:223–7.
30. Zhou P, Yang X-L, Wang X-G *et al.* A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* 2020;**579**:270–3.
31. Shang J, Wan Y, Luo C *et al.* Cell entry mechanisms of SARS-CoV-2. *Proceedings of the National Academy of Sciences* 2020;**117**:11727–34.
32. Xia S, Liu M, Wang C *et al.* Inhibition of SARS-CoV-2 (previously 2019-nCoV) infection by a highly potent pan-coronavirus fusion inhibitor targeting its spike protein that harbors a high capacity to mediate membrane fusion. *Cell Res* 2020;**30**:343–55.
33. Mol CD, Dougan DR, Schneider TR *et al.* Structural Basis for the Autoinhibition and STI-571 Inhibition of c-Kit Tyrosine Kinase \*. *Journal of Biological Chemistry* 2004;**279**:31655–63.
34. Chen F, Zhang Y, Li X *et al.* The Impact of ACE2 Polymorphisms on COVID-19 Disease: Susceptibility, Severity, and Therapy. *Front Cell Infect Microbiol* 2021;**11**, DOI: 10.3389/fcimb.2021.753721.
35. Fernandes Q, Inchakalody VP, Merhi M *et al.* Emerging COVID-19 variants and their impact on SARS-CoV-2 diagnosis, therapeutics and vaccines. *Annals of Medicine* 2022;**54**:524–40.
36. Lu H. Drug treatment options for the 2019-new coronavirus (2019-nCoV). *Biosci Trends* 2020;**14**:69–71.

37. Cao B, Wang Y, Wen D *et al.* A Trial of Lopinavir-Ritonavir in Adults Hospitalized with Severe Covid-19. *N Engl J Med* 2020;**382**:1787–99.
38. Soin AS, Kumar K, Choudhary NS *et al.* Tocilizumab plus standard care versus standard care in patients in India with moderate to severe COVID-19-associated cytokine release syndrome (COVINTOC): an open-label, multicentre, randomised, controlled, phase 3 trial. *Lancet Respir Med* 2021;**9**:511–21.
39. Branch-Elliman W, Ferguson R, Doros G *et al.* Subcutaneous sarilumab for the treatment of hospitalized patients with moderate to severe COVID19 disease: A pragmatic, embedded randomized clinical trial. *PLOS ONE* 2022;**17**:e0263591.
40. Phase 3 Trial Shows REGEN-COV™ (casirivimab with imdevimab) Antibody Cocktail Reduced Hospitalization or Death by 70% in Non-hospitalized COVID-19 Patients. *BioSpace*.
41. Amani B, Amani B. Efficacy and safety of nirmatrelvir/ritonavir (Paxlovid) for COVID-19: A rapid review and meta-analysis. *Journal of Medical Virology* 2023;**95**:e28441.
42. Folegatti PM, Ewer KJ, Aley PK *et al.* Safety and immunogenicity of the ChAdOx1 nCoV-19 vaccine against SARS-CoV-2: a preliminary report of a phase 1/2, single-blind, randomised controlled trial. *Lancet* 2020;**396**:467–78.
43. Pardi N, Hogan MJ, Porter FW *et al.* mRNA vaccines - a new era in vaccinology. *Nat Rev Drug Discov* 2018;**17**:261–79.
44. Tian J-H, Patel N, Haupt R *et al.* SARS-CoV-2 spike glycoprotein vaccine candidate NVX-CoV2373 immunogenicity in baboons and protection in mice. *Nat Commun* 2021;**12**:372.
45. Gao Q, Bao L, Mao H *et al.* Development of an inactivated vaccine candidate for SARS-CoV-2. *Science* 2020;**369**:77–81.
46. French S, Robson B. What is a conservative substitution? *J Mol Evol* 1983;**19**:171–5.
47. Zhao L, Hall M, de Cesare M *et al.* The mutational spectrum of SARS-CoV-2 genomic and antigenomic RNA. *Proceedings of the Royal Society B: Biological Sciences* 2022;**289**:20221747.
48. A Model Evolutionary Change in Proteins: Dayhoff, Schwartz, and | PDF | Life Sciences | Molecular Biophysics. *Scribd*.

49. Verli H. *Bioinformática : da biologia à flexibilidade molecular*. Sociedade Brasileira de Bioquímica e Biologia Molecular, 2014.
50. Meng J, Zhu Q, Zhang L *et al*. Genome and Transcriptome Analyses Provide Insight into the Euryhaline Adaptation Mechanism of *Crassostrea gigas*. *PLOS ONE* 2013;**8**:e58563.
51. Kitchen DB, Decornez H, Furr JR *et al*. Docking and scoring in virtual screening for drug discovery: methods and applications. *Nat Rev Drug Discov* 2004;**3**:935–49.
52. Andrei CC. *Da química medicinal a química combinatoria e modelagem molecular : um curso pratico /*. 2022.
53. Rodrigues RP, Mantoani SP, Almeida JR de *et al*. Estratégias de Triagem Virtual no Planejamento de Fármacos. *Revista Virtual de Química* 2012;**4**:739–76.
54. Morgon NH, Coutinho KR. *Métodos de química teórica e modelagem molecular*. 2007.
55. Goodarzi M, Freitas MP, Ferreira EB. Influence of Changes in 2-D Chemical Structure Drawings and Image Formats on the Prediction of Biological Properties Using MIA-QSAR. *QSAR & Combinatorial Science* 2009;**28**:458–64.
56. Ventura RG. *Introdução à Termodinâmica da Engenharia Química* Smith, Van Ness 7<sup>a</sup> ed.
57. Kovács A, Kenny B, Lipkowitz (ed): Reviews in computational chemistry, vol 26. *Struct Chem* 2009;**20**:951–2.
58. Dabrowski-Tumanski P, Stasiak A, Sulkowska JI. In Search of Functional Advantages of Knots in Proteins. *PLOS ONE* 2016;**11**:e0165986.
59. Sułkowska JI, Noel JK, Onuchic JN. Energy landscape of knotted protein folding. *Proceedings of the National Academy of Sciences* 2012;**109**:17783–8.
60. Xu Y, Li S, Yan Z *et al*. Stabilizing Effect of Inherent Knots on Proteins Revealed by Molecular Dynamics Simulations. *Biophysical Journal* 2018;**115**:1681–9.
61. Faísca PFN. Knotted proteins: A tangled tale of Structural Biology. *Computational and Structural Biotechnology Journal* 2015;**13**:459–68.
62. Namba AM, Silva VB da, Silva CHT de P da. Dinâmica molecular: teoria e aplicações em planejamento de fármacos. *Eclética Química* 2008;**33**:13–23.

63. Hsu S-TD, Lee Y-TC, Mikula KM *et al.* Tying up the Loose Ends: A Mathematically Knotted Protein. *Front Chem* 2021;**9**, DOI: 10.3389/fchem.2021.663241.
64. Waterhouse A, Bertoni M, Bienert S *et al.* SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Research* 2018;**46**:W296–303.
65. The spike glycoprotein of the new coronavirus 2019-nCoV contains a furin-like cleavage site absent in CoV of the same clade. *Antiviral Research* 2020;**176**:104742.
66. Tovchigrechko A, Vakser IA. GRAMM-X public web server for protein–protein docking. *Nucleic Acids Research* 2006;**34**:W310–4.
67. Robson B. Computers and viral diseases. Preliminary bioinformatics studies on the design of a synthetic vaccine and a preventative peptidomimetic antagonist against the SARS-CoV-2 (2019-nCoV, COVID-19) coronavirus. *Computers in Biology and Medicine* 2020;**119**:103670.
68. Robson B, Boray S. Data-mining to build a knowledge representation store for clinical decision support. Studies on curation and validation based on machine performance in multiple choice medical licensing examinations. *Computers in Biology and Medicine* 2016;**73**:71–93.
69. Robson B, Boray S. Interesting things for computer systems to do: Keeping and data mining millions of patient records, guiding patients and physicians, and passing medical licensing exams. *2015 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. 2015, 1397–403.
70. Phillips JC, Hardy DJ, Maia JDC *et al.* Scalable molecular dynamics on CPU and GPU architectures with NAMD. *J Chem Phys* 2020;**153**:044130.
71. Jurrus E, Engel D, Star K *et al.* Improvements to the APBS biomolecular solvation software suite. *Protein Sci* 2018;**27**:112–28.
72. Bisong E. *Building Machine Learning and Deep Learning Models on Google Cloud Platform: A Comprehensive Guide for Beginners.*, 2019.
73. Raveh B, London N, Zimmerman L *et al.* Rosetta FlexPepDock ab-initio: Simultaneous Folding, Docking and Refinement of Peptides onto Their Receptors. *PLOS ONE* 2011;**6**:e18934.

74. Pimentel V, Mariano D, Cantão LXS *et al.* VTR: A Web Tool for Identifying Analogous Contacts on Protein Structures and Their Complexes. *Front Bioinform* 2021;**1**, DOI: 10.3389/fbinf.2021.730350.
75. Humphrey W, Dalke A, Schulten K. VMD: Visual molecular dynamics. *Journal of Molecular Graphics* 1996;**14**:33–8.
76. Chaudhury S, Lyskov S, Gray JJ. PyRosetta: a script-based interface for implementing molecular modeling algorithms using Rosetta. *Bioinformatics* 2010;**26**:689–91.
77. Nelson MT, Humphrey W, Gursoy A *et al.* NAMD: a Parallel, Object-Oriented Molecular Dynamics Program. *The International Journal of Supercomputer Applications and High Performance Computing* 1996;**10**:251–68.
78. Gowers RJ, Linke M, Barnoud J *et al.* *MDAnalysis: A Python Package for the Rapid Analysis of Molecular Dynamics Simulations*. Los Alamos National Laboratory (LANL), Los Alamos, NM (United States), 2019.
79. MDAnalysis: Um kit de ferramentas para a análise de simulações de dinâmica molecular - Michaud-Agrawal - 2011 - Journal of Computational Chemistry - Wiley Online Library.
80. Harris CR, Millman KJ, van der Walt SJ *et al.* Array programming with NumPy. *Nature* 2020;**585**:357–62.
81. Hunter JD. Matplotlib: A 2D Graphics Environment. *Computing in Science & Engineering* 2007;**9**:90–5.
82. Weng G, Gao J, Wang Z *et al.* Comprehensive Evaluation of Fourteen Docking Programs on Protein–Peptide Complexes. *J Chem Theory Comput* 2020;**16**:3959–69.
83. Neduva V, Linding R, Su-Angrand I *et al.* Systematic Discovery of New Recognition Peptides Mediating Protein Interaction Networks. *PLOS Biology* 2005;**3**:e405.
84. Glaser F, Steinberg DM, Vakser IA *et al.* Residue frequencies and pairing preferences at protein–protein interfaces. *Proteins: Structure, Function, and Bioinformatics* 2001;**43**:89–102.
85. Kumar A K, Rathore RS. Categorization of hotspots into three types - weak, moderate and strong to distinguish protein-protein versus protein-peptide interactions. *J Biomol Struct Dyn* 2023:1–13.
86. Servidor Web Prodigy.

87. Xue LC, Rodrigues JP, Kastritis PL *et al.* PRODIGY: a web server for predicting the binding affinity of protein–protein complexes. *Bioinformatics* 2016;**32**:3676–8.
88. Valiente PA, Wen H, Nim S *et al.* Computational Design of Potent D-Peptide Inhibitors of SARS-CoV-2. *J Med Chem* 2021;**64**:14955–67.
89. Sitthiyotha T, Chunsrivirod S. Computational design of SARS-CoV-2 peptide binders with better predicted binding affinities than human ACE2 receptor. *Sci Rep* 2021;**11**:15650.
90. Sun J, Liu X, Zhang S *et al.* Molecular insights and optimization strategies for the competitive binding of engineered ACE2 proteins: a multiple replica molecular dynamics study. *Phys Chem Chem Phys* 2023;**25**:28479–96.
91. Robson B. *Preliminary Bioinformatics Studies on the Design of Synthetic Vaccines and Preventative Peptidomimetic Antagonists against the Wuhan Seafood Market Coronavirus. Possible Importance of the KRSFIEDLLFNKV Motif.*, 2020.
92. Robson B. COVID-19 Coronavirus spike protein analysis for synthetic vaccines, a peptidomimetic antagonist, and therapeutic drugs, and analysis of a proposed achilles' heel conserved region to minimize probability of escape mutations and drug resistance. *Computers in Biology and Medicine* 2020;**121**:103749.
93. Li F. Structure, Function, and Evolution of Coronavirus Spike Proteins. *Annual Review of Virology* 2016;**3**:237–61.

## 8.Apêndices

### Apêndice 8.1

Artigo 1: **An Approach for Engineering Peptides for Competitive Inhibition of the SARS-COV-2 Spike Protein**

Disponível: DOI: [10.3390/molecules29071577](https://doi.org/10.3390/molecules29071577)

Artigo 2: **From In-Person to the Online World: Insights Into Organizing Events in Bioinformatics**

Disponível: DOI: [10.3389/fbinf.2021.711463](https://doi.org/10.3389/fbinf.2021.711463)

### Apêndice 8.2

**Bioinformática com Jupyter Notebook**

2023 | Livro

DOI: [10.51780/978-65-992753-7-1](https://doi.org/10.51780/978-65-992753-7-1)

**BIOINFO #03 - Revista Brasileira de Bioinformática e Biologia Computacional**

2023-09 | Livro

DOI: [10.51780/978-65-992753-8-8](https://doi.org/10.51780/978-65-992753-8-8)

**LBS Tech - Desenvolvimento web**

2023-03-20 | Livro

DOI: [10.51780/978-65-992753-6-4](https://doi.org/10.51780/978-65-992753-6-4)

**Using Natural Language Processing for Context Identification in COVID-19 Literature**

2023 | Capítulo de livro

DOI: [10.1007/978-3-031-42715-2\\_7](https://doi.org/10.1007/978-3-031-42715-2_7)

### Apêndice 8.3

2017-2021: Membro da comissão organizadora do Curso de Verão em Bioinformática (CVBioinfo), composta por alunos de pós-graduação em bioinformática da Universidade Federal de Minas Gerais (Brasil).