UNIVERSIDADE DE SÃO PAULO
FACULDADE DE FILOSOFIA, CIÊNCIAS E LETRAS DE RIBEIRÃO PRETO
PROGRAMA DE PÓS-GRADUAÇÃO EM FÍSICA APLICADA À MEDICINA E
BIOLOGIA

PATRÍCIA PEREIRA DUZI CARVALHO

**Caracterização da interação dos coronavírus SARS-CoV e SARS-CoV-2 com
o receptor ACE2 por meio de um padrão evolutivo conservado de
aminoácidos**

Ribeirão Preto

2020

PATRÍCIA PEREIRA DUZI CARVALHO

**Caracterização da interação dos coronavírus SARS-CoV e SARS-CoV-2 com
o receptor ACE2 por meio de um padrão evolutivo conservado de
aminoácidos**

Versão corrigida

Dissertação apresentada à Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto, Universidade de São Paulo, para obtenção do título de Mestre em Ciências.

Área de concentração: Física Aplicada à Medicina e Biologia

Orientador: Prof. Dr. Nelson Augusto Alves

Ribeirão Preto

2020

Carvalho , Patrícia Pereira Duzi

Caracterização da interação dos coronavírus SARS-CoV e SARS-CoV-2 com o receptor ACE2 por meio de um padrão evolutivo conservado de aminoácidos / PATRÍCIA PEREIRA DUZI CARVALHO. – Ribeirão Preto, 2020-
79 p. : il. ; 30 cm.

Dissertação de mestrado apresentada ao Programa de Pós-Graduação em Física Aplicada à Medicina e Biologia do Departamento de Física da Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto da Universidade de São Paulo
Orientador: Prof. Dr. Nelson Augusto Alves

1. ACE2. 2. SARS-CoV-2. 3. SARS-CoV. 4. COVID-19. 5. Conservação de aminoácidos. 6. Bolso hidrofóbico 7. Bioinformática.

Dissertação de autoria de Patrícia Pereira Duzi Carvalho, sob o título “**Caracterização da interação dos coronavírus SARS-CoV e SARS-CoV-2 com o receptor ACE2 por meio de um padrão evolutivo conservado de aminoácidos**”, apresentada à Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto, Universidade de São Paulo, para obtenção do título de Mestre em Ciências, na área de concentração Física Aplicada à Medicina e Biologia, aprovada em _____ de _____ de _____ pela comissão julgadora constituída pelos doutores:

Prof. Dr. _____

Instituição: _____

Presidente

Prof. Dr. _____

Instituição: _____

Prof. Dr. _____

Instituição: _____

Prof. Dr. _____

Instituição: _____

Obrigada por me aceitar sem levar em conta a minha aparência. Por aplaudir meu sucesso e me apoiar se fracasso. Por escutar com atenção os meus desabafos e conquistas. Muito obrigada por me apoiar para eu não desistir, me encorajando a acreditar que pode ser possível.

Agradecimentos

A Deus, por estar sempre comigo, protegendo e guiando.

Aos meus pais, Rosimeire e Francisco, pelo apoio e incentivo em todos os momentos da minha vida. Por acreditarem em mim e não medirem esforços para a concretização dos meus sonhos. Sem vocês nada seria possível. Mãe e Pai, amo vocês. A meu esposo Wellington e as minhas filhas Samyra e Maysa por compreenderem as várias horas em que estive ausente desenvolvendo este trabalho. Amo vocês.

A tia Aldenora e minha irmã Vanessa que cuidaram com tanto amor das minhas filhas na minha ausência.

A todos os mestres que contribuíram com a minha formação acadêmica e profissional durante a minha vida.

A dona Fátima, agente de limpeza do bloco de física-RP, que me levou até a sala do professor Nelson justamente no momento que eu ia desistir.

A meu orientador Prof. Dr. Nelson A. Alves, ótimo professor e orientador. Agradeço por sua confiança e incansável dedicação.

Aos integrantes da nossa equipe laboratorial e aos meus irmãos na amizade Alexandre, Emanuel e André que fizeram parte da minha formação.

Aos meus amigos e familiares por todo amor, força, incentivo e apoio incondicional.

A Universidade de São Paulo, por me proporcionar um ambiente criativo e amigável para os estudos. Sou grata a cada membro desta Instituição de ensino.

À CAPES, pois o presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES).

*A busca por um estado permanente é um insulto á própria natureza, enquanto vivos somos
brevidade e inquietude*

[W. Pardo, 2019].

Resumo

CARVALHO, P. P. D. **Caracterização da interação dos coronavírus SARS-CoV e SARS-CoV-2 com o receptor ACE2 por meio de um padrão evolutivo conservado de aminoácidos.** 2020. 79 f. Dissertação (Mestrado em Ciências na área de Física Aplicada à Medicina e Biologia) – FFCLRP - Universidade de São Paulo, São Paulo, 2021.

As proteínas *spike* (S) medeiam a entrada do coronavírus na célula hospedeira. A subunidade S1 das proteínas S contém o domínio de ligação ao receptor (RBD) que é capaz de reconhecer diferentes receptores, destacando sua notável capacidade de adaptação aos seus hospedeiros ao longo da evolução viral. Enquanto o RBD situado na proteína *spike* é determinante para a interação vírus-receptor, os resíduos ativos estão no motivo de ligação ao receptor (RBM), uma região localizada no RBD que desempenha um papel fundamental na ligação da superfície externa de seus receptores. Neste projeto de pesquisa, abordamos a hipótese de que as cepas dos coronavírus SARS-CoV e SARS-CoV-2 capazes de usar a enzima conversora de angiotensina 2 (ACE2) para infectar células hospedeiras adaptaram seu RBM ao longo da evolução viral para explorar topologias conformacionais direcionadas pelos resíduos de aminoácidos YGF que formam um bolso hidrofóbico, importante via de reconhecimento e ligação ao receptor. Mecanismo molecular similar é usado na interação do complexo UBA-ubiquitina, associado à proteólise. Utilizamos neste projeto diferentes gêneros de coronavírus e suas sequências foram analisadas com ClustalW e Jalview com o intuito de investigar se o mecanismo molecular baseado em YGF pode atuar como uma assinatura proteica localizada no RBM, distinguindo os coronavírus capazes de usar o ACE2 como um receptor de entrada celular. Os resíduos analisados podem representar um alvo importante para o desenvolvimento de medicamentos e além disso, a compreensão da organização do coronavírus na ligação ao receptor ACE2.

Palavras-chaves: ACE2. SARS-CoV-2. SARS-CoV. COVID-19. Conservação de aminoácidos. Bolso hidrofóbico. Bioinformática.

Abstract

CARVALHO, P. P. D. **Featuring ACE2 binding SARS-CoV and SARS-CoV-2 through a conserved evolutionary pattern of amino acid residues.** 2020. 79 p. Dissertation (Master of Science in Physics Applied to Medicine and Biology) – Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto, Universidade de São Paulo, São Paulo, 2021.

Spike (S) proteins mediate coronavirus entry into the host cell. The S1 subunit of the S proteins contains the receptor-binding domain (RBD) that is able to recognize different host receptors, highlighting its remarkable capacity to adapt to their hosts along the viral evolution. While RBD in spike proteins is determinant for the virus-receptor interaction, the active residues lie at the receptor-binding motif (RBM), a region located in RBD that plays a fundamental role binding the outer surface of their receptors. In this research project, we approached the hypothesis that strains of the SARS-CoV and SARS-CoV-2 coronaviruses capable of using the angiotensin-converting enzyme 2 (ACE2) to infect host cells adapted their RBM throughout viral evolution to explore topologies conformational directed by the YGF amino acid residues that form a hydrophobic pocket, an important pathway for recognition and binding to the receptor. A similar mechanism is used in the interaction of the UBA-ubiquitin complex, associated with proteolysis. In this project we used different genera of coronaviruses and their sequences were analyzed with ClustalW and Jalview in order to investigate whether the molecular mechanism based on YGF can act as a protein signature located in the RBM, distinguishing coronaviruses capable of using ACE2 as a receptor for cell entry. The analyzed residues may represent an important target for the development of medications and in addition, the understanding of the organization of the coronavirus in the binding to the ACE2 receptor.

Keywords: ACE2. SARS-CoV-2. SARS-CoV. COVID-19. Conservation of amino acids. hydrophobic pocket.

Lista de figuras

- Figura 1.1 – Classificação taxonômica comparada entre coronavírus e humanos. Os coronavírus (CoVs) pertencem à família *Coronaviridae*, ordem *Nidovirales* e subordem *Cornidovirineae*. Sua subfamília *Orthocoronavirinae* contém quatro gêneros baseados em filogenia e denominados α , β , γ e δ -coronavírus. 22
- Figura 1.2 – Genoma de SARS-CoV, MERS-CoV e SARS-CoV-2. As proteínas estruturais são: proteína S, relacionada a ligação ao receptor. Proteína M, define a forma do envelope viral, é a proteína estrutural mais abundante. Proteína E, juntamente com a proteína M formam o envelope viral. É a menor proteína estrutural essencial para patogenicidade e exibe importante função na produção e maturação viral. Proteína N, é a única proteína estrutural que se liga ao genoma viral atuando na formação completa do vírion. Adaptado da referência (1). 23
- Figura 1.3 – Etapas do ciclo viral de SARS-CoV/SARS-CoV-2. Existem dois mecanismos de entrada do vírus na célula hospedeira, via fusão com membrana citoplasmática ou fusão com membrana endossômica da célula hospedeira. Adaptado da referência (2). 24
- Figura 1.4 – Estrutura geral da proteína *spike* trimérica: domínio C-terminal (S1-CTD) em verde, domínio N-terminal (S1-NTD) em magenta, subunidade S2 em ciano, âncora transmembrana (AT) e cauda intracelular (CI), ambos mostrados na região inferior direita. S1-CTD liga-se a diferentes receptores, por exemplo, ACE2, DPP4 e APN. S1-NTD reconhece a molécula de adesão celular relacionada ao CEACAM 1 ou açúcar. Trímero da proteína *spike* adaptado da referência (3). 26

Figura 1.5 – Proteína <i>spike</i> RBD: (a) Diversidade de sequência RBD de diferentes coronavírus. (b) Domínio da proteína <i>spike</i> SARS-CoV. A região S1 que está em cor magenta contém o RBD em ciano responsável pela ligação e afinidade ao receptor do hospedeiro. O RBM que está em vermelho dentro do RBD contém os sítios ativos de ligação ao receptor. A região S2 em cor de laranja é responsável pela fusão da membrana.	27
Figura 2.1 – Diferentes domínios UBA mostrando similaridade no arranjo conformacional. Ambas conformações se organizam em três hélices, induzidas pela presença do trecho superficial hidrofóbico, destacado em amarelo, típico de um bolso hidrofóbico, formado pela sequência de resíduos XGF.	30
Figura 2.2 – Visão detalhada da superfície do domínio UBA (em magenta) da proteína Dsk2p complexada com ubiquitina (em ciano), código de acesso: PDB ID: 1WR1. Os resíduos M342, G343 e F344 são bem conservados nos domínios UBA e formam o núcleo hidrofóbico (em amarelo) do grande trecho de superfície hidrofóbica para interação com ubiquitina.	31
Figura 2.3 – Alinhamento entre diferentes domínios UBA mostrando alta conservação dos resíduos XGF. As setas estão direcionadas para os resíduos que constituem o bolso hidrofóbico.	32
Figura 3.1 – Exemplo de arquivo FASTA de entrada para alinhamento que contém as sequências de aminoácidos no formato de texto. A sequência FASTA de nucleotídeos e proteínas podem ser obtidas através dos bancos de dados UNIPROT e NCBI de acesso livre. O arquivo FASTA é composto pelo cabeçalho e as sequências. Todo cabeçalho FASTA inicia com o sinal de > (maior).	36
Figura 4.1 – (a) Resíduos YGFY na interface do SARS-CoV complexado com ACE2 humano (PDB ID: 2AJF). Esses resíduos (em amarelo) formam um bolso hidrofóbico localizado no RBM (cor magenta). (b) Resíduos YGF e EGF (em amarelo) na interface de SARS-CoV-2 complexados com ACE2 humano (PDB ID: 6LZG). A primeira sequência YGF está localizada num bolso hidrofóbico (em amarelo), enquanto que a segunda sequência EGF (em amarelo) que está na superfície RBM (em magenta). Representação em fita do ACE2 (em azul).	44

- Figura 4.2 – Diagramas de fita de SARS-CoV (a) e SARS-CoV-2 (b) complexados com ACE2 humano (cor azul), onde o RBM é destacado em cor magenta. Os principais resíduos responsáveis pela ligação estrutural são exibidos na representação *stick*. Os resíduos N479/Q493, T487/N501 são essenciais para a ligação ao receptor hospedeiro e os resíduos Y442/L455, L472/F486, D480/S494 são importantes contatos de interação ao receptor. 44
- Figura 4.3 – Redes de contato entre (a) resíduos SARS-CoV e (b) resíduos SARS-CoV-2 localizados nas regiões RBM com ACE2. Os resíduos SARS-CoV e SARS-CoV-2 são na cor magenta, enquanto os resíduos ACE2 humanos são na cor azul. 45
- Figura 4.4 – Alinhamentos de sequência de CoVs humanos restritos a resíduos RBM. A cor púrpura destaca o padrão YGFY seguido pela mutação Y498Q no RBM das cepas SARS-CoV-2. 46
- Figura 4.5 – Ponte salina no receptor hACE2. SARS-CoV (a) resíduos RBM (cor magenta) Y442, N479 e D480 estão localizados perto da ponte salina K31-E35 no receptor hACE2 (em ciano), (b) resíduo RBM T487 está próximo a uma segunda ponte salina D38-K353 no receptor hACE2. SARS-CoV-2 (c) resíduos RBM (cor magenta) L455, Q493 e S494 estão localizados perto da ponte salina K31-E35 no receptor hACE2 (em ciano) e (d) o resíduo N501 está próximo a uma segunda ponte salina D38-K353 no receptor hACE2. Os resíduos Lys31 e Lys353 de hACE2 estão enterrados numa região hidrofóbica e são dois importantes pontos de acesso de ligação ao vírus. Os principais resíduos responsáveis pela ligação estrutural são exibidos na representação *stick*. 47
- Figura 4.6 – Detalhes estruturais após complexação de (a) SARS-CoV e (b) SARS-CoV-2 complexados com ACE2 humano (cor azul), onde o RBM é destacado em cor magenta. Os principais resíduos responsáveis pela ligação estrutural são exibidos na representação *stick*. 48
- Figura 4.7 – Alinhamento de sequência de CoVs de morcego restrito a resíduos RBM de SARS-CoV Tor2. Os resíduos do padrão YGFY estão em cor púrpura. Os últimos três alinhamentos são colocados juntos para comparação direta da sequência de aminoácidos. 50

Figura 4.8 – Alinhamento de sequência de CoVs de civeta de palma e PCoVs de pangolim restrito a resíduos RBM de SARS-CoV Tor2. Os resíduos do padrão YGFY estão em cor púrpura. A última linha inclui a sequência SARS-CoV-2 para comparação.	50
Figura 4.9 – Números de acesso do GenBank para as sequências de coronavírus analisadas neste estudo.	60
Figura A.1 – Organização estrutural do proteassomo. Sua estrutura é organizada em um núcleo funcional 20S semelhante a um cilindro central, que possui atividades proteolíticas e um complexo regulatório 19S. A estrutura 20S adota uma conformação em barril devido a sobreposição de suas quatro fileiras proteicas de formato anelar, sendo duas fileiras externas idênticas α e duas fileiras idênticas β . Cada uma das quatro fileiras da estrutura 20S são constituídas de sete subunidades proteolítica distintas. A estrutura 19S funciona como uma tampa e uma base para 20S, sendo um importante componente no processo de degradação. A tampa 19S reconhece e desenovela a proteína poliubiquitinada, abre o orifício do anel α 20S permitindo o encaixe e a entrada da proteína no barril proteolítico 20S onde ocorre a degradação e liberação de fragmentos peptídicos. A associação dessas duas regiões forma o complexo 26S. A proteína deve ser primeiramente marcada por via ubiquitinação, para a degradação pelo proteassomo 26S. Adaptado da referência (4).	72
Figura A.2 – Sistema ubiquitina-proteassomo. A ubiquitina é ativada pela enzima E1 (1) seguida por sua transferência para uma proteína transportadora de ubiquitina, E2 (2). E2 transfere as porções de ubiquitina ativada para o substrato de proteína que está ligado especificamente a uma ubiquitina ligase E3. A transferência é direta (3) ou através de um intermediário na ligase (4, 4a). A conjugação sucessiva de ubiquitina gera uma cadeia de poliubiquitina que serve como sinal de ligação (5) e degradação para o proteassomo 26S. O substrato é degradado em peptídeos curtos (6), e a ubiquitina livre e reutilizável é liberada por enzimas desubiquitinantes (7). Adaptado da referência (5).	73

Figura C.1–Diagrama de dispersão mostrando o coeficiente de correlação entre a tabela 4.4 (MutaBind2) e tabela 4.6 (BeAtMuSiC). Estes valores possuem alta correlação com $\chi^2 = 0.41$	78
Figura C.2–Diagrama de dispersão mostrando o coeficiente de correlação entre a tabela 4.4 (MutaBind2) e tabela 4.6 (mCSM-PPI2).	79
Figura C.3–Diagrama de dispersão mostrando o coeficiente de correlação entre a tabela 4.6 (BeAtMuSiC) e tabela 4.6 (mCSM-PPI2).	79

Lista de tabelas

Tabela 1.1 – Espécies de coronavírus que podem infectar humanos. A marcação com (X) corresponde ao gênero do coronavírus.	22
Tabela 3.1 – Diferentes algoritmos BLAST para distintas aplicações. Os diferentes domínios de busca dependem da sequência de entrada.	38
Tabela 4.1 – Mutação de ponto único N479 SARS-CoV pelos resíduos E (pangolins), K (morcego e civeta de palma), Q (SARS-CoV-2), R (morcego e civeta de palma) e S (morcego) na complexação.	49
Tabela 4.2 – Mutações múltiplas específicas em G482 e Y484 de SARS-CoV do trecho hidrofóbico, por resíduos que ocorrem em outras cepas, G482D (morcego) variando com Y484F (morcego), Y484H (pangolins), Y484N (morcego), Y484Q (SARS-CoV-2), Y484T (morcego).	51
Tabela 4.3 – Mutações múltiplas no local N479 de SARS-CoV interrompendo o trecho de superfície hidrofóbica com as mutações nos locais 482 e 484. Tomemos, por exemplo, a mutação N479E (pangolins) com G482D (morcego) variando com mutações Y484F (morcego), Y484H (pangolins), Y484N (morcego), Y484Q (SARS-CoV-2), Y484T (morcego).	52
Tabela 4.4 – Alterações na afinidade de ligação do <i>spike</i> RBD de SARS-CoV complexado com ACE2 após mutação, conforme previsto pelo método MutaBind2 (6). Efeitos de previsão são classificados como previsão de baixa confiança: (*), ou previsão de alta confiança: (**). Aqui, adotamos a classificação apresentada no método MutaBind (6) devido à alta similaridade entre as curvas COR (características operacionais do receptor) em ambos os métodos.	53

Tabela 4.5 – Alterações na afinidade de ligação do <i>spike</i> RBD de SARS-CoV-2 complexado com ACE2 após mutação, conforme previsto pelo método MutaBind2 (6). Efeitos de previsão são classificados como previsão de baixa confiança: (*), ou previsão de alta confiança: (**). Aqui, adotamos a classificação apresentada no método MutaBind (6) devido à alta similaridade entre as curvas COR (características operacionais do receptor) em ambos os métodos.	54
Tabela 4.6 – Alterações na afinidade de ligação do <i>spike</i> RBD de SARS-CoV e SARS-CoV-2 complexado com ACE2 após mutação. $\Delta\Delta G > 0$ mutação diminui a afinidade de ligação. $\Delta\Delta G < 0$ mutação aumenta a afinidade de ligação, conforme previsto pelo programa BeAtMuSiC (7) e mCSM-PPI2 (8). .	55
Tabela 4.7 – Porcentagem de identidade compartilhada entre diferentes cepas de coronavírus utilizando as sequências do genoma/ <i>spike</i> /RBM respectivamente. As cepas utilizadas foram: SARS-CoV-2 GenBank: MN908947.3 (genoma completo), GenBank: QHD43416.1 (<i>spike</i>); SARS-CoV GenBank: NC_004718 (genoma completo), GenBank: YP_009825051.1 (<i>spike</i>); RaTG13 GenBank: MN996532.2 (genoma completo), GenBank: QHR63300.2 (<i>spike</i>); PCoV GX-P2V GenBank: MT072864.1 (genoma completo), GenBank: QIQ54048.1 (<i>spike</i>); bat-SL-CoV ZC45 GenBank: MG772933 (genoma completo), GenBank: AVP78031.1 (<i>spike</i>); bat-SL-CoV ZXC21 GenBank: MG772934.1 (genoma completo), GenBank: AVP78042.1 (<i>spike</i>). A porcentagem de identidade do RBM de todas as cepas foram calculadas utilizando resultados de alinhamento genéticos obtidos nesse estudo.	58

Lista de abreviaturas e siglas

SARS-CoV	Síndrome Respiratória Aguda Grave
SARS-CoV-2	Síndrome Respiratória Aguda Grave 2
MERS-CoV	Síndrome Respiratória do Oriente Médio
ACE2	Enzima conversora de angiotensina 2
hACE2	Enzima conversora de angiotensina 2 humana
RBD	Domínio de ligação ao receptor
RBM	Motivo de ligação ao receptor
UB	Ubiquitina
UBA	<i>Ubiquitin-associated</i>
Ubl	<i>Ubiquitin-like</i>
RMSD	<i>Root mean square deviation</i>
+ssRNA	RNA de cadeia simples de senso positivo
Proteína S	<i>Spike</i>
Proteína M	Membrana
Proteína E	Envelope
Proteína N	Nucleocapsídeo
AT	Âncora transmembrana
CI	Cauda intracelular
S1	Subunidade de ligação ao receptor

S2	Subunidade de fusão de membrana
S1-NTD	Domínio N-terminal S1
S1-CTD	Domínio C-terminal S1
DPP4	Dipeptidil peptidase 4
APN	Aminopeptidase N
CEACAM 1	Antígeno carcinoembrionário 1
(R_g)	Raio de giração
RMN)	Ressonância magnética nuclear
RNA	Ácido ribonucleico
$\Delta\Delta G$	Variação de energia livre de Gibbs
BLAST	<i>Basic local alignment search tool</i>
UNIPROT	<i>Universal protein</i>
<i>score</i>	Pontuação
PDB	Banco de dados de proteínas
wwPDB	Banco de dados mundial de proteína
NCBI	<i>National center for biotechnology information</i>

Sumário

1	Introdução	21
1.1	<i>Proteína Spike</i>	25
1.2	<i>Objetivo</i>	27
2	O mecanismo molecular na complexação UBA-ubiquitina e seu congênera na complexação RBD-ACE2	29
2.1	<i>Mutações desestabilizadoras do bolso hidrofóbico no domínio UBA . .</i>	33
2.2	<i>Mecanismo molecular na complexação RBD-ACE2</i>	34
3	Métodos	35
3.1	<i>Alinhamento de sequências</i>	36
3.2	<i>GenBank</i>	37
3.3	<i>NCBI</i>	37
3.4	<i>BLAST</i>	37
3.5	<i>CLUSTAL W</i>	38
3.6	<i>Jalview</i>	38
3.7	<i>PyMol</i>	39
3.8	<i>UCSF Chimera</i>	39
3.9	<i>PDB</i>	40
3.10	<i>CSU software</i>	40
3.11	<i>MutaBind2</i>	40
3.12	<i>BeAtMuSiC</i>	42
3.13	<i>mCSM-PPI2</i>	42
4	Resultado e discussão	43
4.1	<i>Motivo de ligação ao receptor da proteína spike em morcegos</i>	56

4.2	<i>Motivo de ligação ao receptor da proteína spike em civetas de palma e pangolins</i>	57
4.3	<i>SARS-CoV e hCoV-NL63: relacionados apenas funcionalmente</i>	59
5	Conclusão	61
	Referências¹	62
A	Autofagia e o sistema ubiquitina-proteassomo	71
B	Algoritmo de alinhamento	74
C	Coefficiente de correlação de Pearson (ρ)	78

¹ De acordo com a Associação Brasileira de Normas Técnicas. NBR 6023.

Capítulo 1

Introdução

Em dezembro de 2019 na cidade de Wuhan, capital da província de Hubei na China, foram relatados vários casos de pneumonia de origem desconhecida (9). O patógeno causava sintomas semelhantes ao coronavírus SARS-CoV (síndrome respiratória aguda grave) causador da pandemia de 2002-2003. O novo coronavírus SARS-CoV-2 (síndrome respiratória aguda grave 2) ficou conhecido como COVID-19, responsável pelo atual surto pandêmico. SARS-CoV-2 provoca uma variedade de doenças relacionadas principalmente ao trato respiratório. Os coronavírus apresentam classificação taxonômica assim como os demais seres vivos (10), pertencendo à grande e diversa família *Coronaviridae*, ordem *Nidovirales* e subordem *Cornidovirineae*, mostrado na figura 1.1 (11). Os gêneros α e β -coronavírus são capazes de infectar mamíferos, incluindo humanos, enquanto γ e δ -coronavírus infectam mamíferos e aves.

O coronavírus SARS-CoV-2 é uma das sete espécies de coronavírus que pode infectar humanos (12). Três espécies podem produzir doenças graves, o SARS-CoV-2, o SARS-CoV e o MERS-CoV (síndrome respiratória do Oriente Médio), sendo o SARS-CoV-2 muito mais letal. Os coronavírus OC43, HKU1, NL63 e 229E causam doenças mais leves (13). SARS-CoV-2, SARS-CoV, MERS-CoV, HCoV-OC43 e HCoV-HKU1 pertencem ao gênero β -coronavírus. HCoV-NL63 e o HCoV-229E pertencem ao gênero α -coronavírus (12) como mostrado na tabela 1.1.

Os representantes da família *Coronaviridae* são vírus que apresentam os maiores genomas de ácido ribonucleico (RNA) conhecidos e possuem material genético constituído por RNA de cadeia simples de senso positivo (+ssRNA). O coronavírus SARS-CoV-2 possui um genoma com média de vinte e nove mil e novecentas bases nitrogenadas traduzindo em média vinte e nove proteínas. O genoma do coronavírus apresenta quatro proteínas

Figura 1.1 – Classificação taxonômica comparada entre coronavírus e humanos. Os coronavírus (CoVs) pertencem à família *Coronaviridae*, ordem *Nidovirales* e subordem *Cornidovirineae*. Sua subfamília *Orthocoronavirinae* contém quatro gêneros baseados em filogenia e denominados α , β , γ e δ -coronavírus.

Coronavírus	gênero β -coronavírus	gênero α -coronavírus
SARS-CoV-2	X	
SARS-CoV	X	
MERS-CoV	X	
HCoV-OC43	X	
HCoV-HKU1	X	
HCoV-NL63		X
HCoV-229E		X

Tabela 1.1 – Espécies de coronavírus que podem infectar humanos. A marcação com (X) corresponde ao gênero do coronavírus.

estruturais principais codificadas na extremidade 3' do genoma (1), mostradas na figura 1.2 e listadas a seguir,

- *Spike* (S), relacionada a ligação ao receptor;
- Membrana (M), define a forma do envelope viral, é a proteína estrutural mais abundante;

Figura 1.3 – Etapas do ciclo viral de SARS-CoV/SARS-CoV-2. Existem dois mecanismos de entrada do vírus na célula hospedeira, via fusão com membrana citoplasmática ou fusão com membrana endossômica da célula hospedeira. Adaptado da referência (2).

- Replicação do ácido nucleico viral;
- Expressão gênica viral realizada pelo processo de transcrição e tradução gênica, formam proteínas necessárias para o metabolismo viral e para formação de sua estrutura;
- Transporte de componentes virais sintetizados para locais específicos no interior celular para a morfogênese viral;
- Montagem de capsídeo a partir de componentes virais sintetizados;
- Encapsidação de ácido nucleico no interior do capsídeo montado;
- Maturação do vírus para se tornar infeccioso geralmente ocorre em resposta ao encapsidação do ácido nucleico;
- Transporte do vírus;
- Liberação do vírus maduro infeccioso para o meio extracelular.

Determinados CoVs utilizam a enzima conversora de angiotensina 2 (ACE2) como receptor de ligação nas células hospedeiras, exemplificados pelas espécies do gênero β -coronavírus, SARS-CoV, SARS-CoV-2 e gênero α -coronavírus, hCoV-NL63 (12, 15, 16, 17). Em particular, o SARS-CoV, assim como o SARS-CoV-2, usam dois mecanismos para a infecção da célula hospedeira. O primeiro processo é por fusão do vírus para o interior celular via endossomo (18, 19, 20, 21, 22), após ligação à ACE2. Outro caminho é por fusão do vírus para o interior celular via protease serinoprotease transmembrana (TMPRSS2), que cliva a proteína spike permitindo sua ligação ao receptor ACE2 (23). A replicação viral promovida pela fusão via TMPRSS2 é 100 vezes maior que a replicação por fusão com membrana endossômica (23). Desse modo, a inibição da ligação do vírus ao receptor hospedeiro não depende somente da diminuição da expressão de ACE2, mas também da inibição da protease TMPRSS2 realizado por inibidores, ocasionando assim, sua investigação como alvo terapêutico (15).

Em contraste, o gênero β -coronavírus MERS-CoV e seu geneticamente relacionado CoV-HKU4 de morcego, utilizam dipeptidil peptidase 4 (DPP4) como receptor viral (24). Outro receptor viral é a aminopeptidase N (APN), reconhecida, por exemplo, pelo gênero α -coronavírus hCoV-229E (25). Há grande variação na infectividade das cepas de SARS-CoV, a qual pode estar conectada à afinidade de ligação ao hACE2 (26). Este comportamento pode apresentar correlação com a gravidade da doença em humanos (27).

1.1 Proteína *Spike*

O grupo de vírus de genoma +ssRNA passa por rápida mutação (28), proporcionando melhor adaptação aos novos hospedeiros, porém são observadas regiões conservadas em alinhamentos múltiplos das cepas de CoVs (29, 30, 31). A entrada do vírus na célula hospedeira ocorre por meio da ligação de suas proteínas *sipike* (S) aos receptores da superfície celular, seguido de fusão da membrana viral e hospedeira. As proteínas *sipike* (S) estão localizadas no envelope viral, são globulares triméricas e consistem em três segmentos, sendo um ectodomínio, uma âncora transmembrana de passagem única e uma cauda intracelular curta (3). O ectodomínio pode ser dividido em uma subunidade S1 de ligação ao receptor e uma subunidade S2 de fusão de membrana (32). A subunidade S1 contém dois domínios independentes, um domínio N-terminal (S1-NTD) e um domínio C-terminal (S1-CTD) como mostrado na figura 1.4. Qualquer um ou ambos os domínios S1

podem funcionar como um domínio de ligação ao receptor responsável pelo reconhecimento e ligação ao receptor da superfície celular. CTD liga-se a diferentes receptores, por exemplo, ACE2, dipeptidil peptidase 4 (DPP4) e aminopeptidase N (APN). S1-NTD reconhece a molécula de adesão celular relacionada ao antígeno carcinoembrionário 1 (CEACAM 1) ou açúcar que funciona em processos imunológicos e comunicação célula-célula (3) exibido na figura 1.4. Na subunidade S2 ocorre fusão da membrana após vários rearranjos estruturais.

O domínio de ligação ao receptor (RBD) é capaz de reconhecer diferentes receptores, destacando sua notável capacidade de se adaptar aos seus hospedeiros ao longo da evolução viral (3), mediando a entrada na célula hospedeira indicado na figura 1.4. Dentro do RBD está localizado o motivo de ligação ao receptor (RBM), região onde se encontra os resíduos ativos de ligação ao receptor (15, 16, 26, 33, 34). A figura 1.5 (a) mostra a diversidade de sequência RBD de diferentes coronavírus. A figura 1.5 (b) mostra o domínio da proteína *spike* SARS-CoV. A região S1 contém o RBD, responsável pela ligação e afinidade ao receptor do hospedeiro. O RBM que está dentro do RBD contém os sítios ativos de ligação ao receptor. A região S2 é responsável pela fusão da membrana. As subunidades S1 de diferentes gêneros de coronavírus e dos mesmos coronavírus que parasitam diferentes espécies hospedeira compartilham pouca ou nenhuma semelhança significativa de sequência. As sequências, estruturas e mecanismos de fusão por membrana das subunidades S2 são conservadas entre os diferentes gêneros de coronavírus (32, 35, 36).

Figura 1.4 – Estrutura geral da proteína *spike* trimérica: domínio C-terminal (S1-CTD) em verde, domínio N-terminal (S1-NTD) em magenta, subunidade S2 em ciano, âncora transmembrana (AT) e cauda intracelular (CI), ambos mostrados na região inferior direita. S1-CTD liga-se a diferentes receptores, por exemplo, ACE2, DPP4 e APN. S1-NTD reconhece a molécula de adesão celular relacionada ao CEACAM 1 ou açúcar. Trímero da proteína *spike* adaptado da referência (3).

Figura 1.5 – Proteína *spike* RBD: (a) Diversidade de sequência RBD de diferentes coronavírus. (b) Domínio da proteína *spike* SARS-CoV. A região S1 que está em cor magenta contém o RBD em ciano responsável pela ligação e afinidade ao receptor do hospedeiro. O RBM que está em vermelho dentro do RBD contém os sítios ativos de ligação ao receptor. A região S2 em cor de laranja é responsável pela fusão da membrana.

1.2 Objetivo

Nesta investigação buscamos caracterizar a interação dos coronavírus SARS-CoV e SARS-CoV-2 com o receptor ACE2 por meio de um padrão evolutivo conservado de aminoácidos. Este estudo revelou uma sequência de resíduos de aminoácidos altamente conservada Tyr-Gly-Phe (YGF) em variantes de coronavírus que empregam esse receptor (37). É provável que cepas de SARS-CoV e SARS-CoV-2 capazes de usar proteínas ACE2 tenham adaptado seu RBM ao longo da evolução para explorar esse mecanismo molecular

de ligação YGF para entrada celular. Sugerimos que essa sequência curta conduz a afinidade de ligação à ACE2, devido à formação de um bolso hidrofóbico, adequado para a ligação ao receptor (33, 38, 39, 40). Este não é um mecanismo molecular novo. Relataremos no próximo capítulo o papel funcional deste mecanismo como via de formação da complexação UBA-ubiquitina, associada à proteólise.

Capítulo 2

O mecanismo molecular na complexação UBA-ubiquitina e seu congênere na complexação RBD-ACE2

O domínio *ubiquitin-associated* (UBA) está presente em muitas proteínas associadas à proteólise mediada pela ubiquitina (41, 42), (ver apêndice A). Sua estrutura secundária se mantém conservada ao longo da evolução eucariótica. O domínio UBA representa a região de ligação à ubiquitina (43). Sua estrutura primária é constituída por cerca de 55 resíduos, mas nem toda essa sequência é conservada (44) sendo 45 a quantidade de resíduos de aminoácidos mais frequentemente encontrados em diversas proteínas envolvidas no complexo ubiquitina-proteassomo. Foi demonstrado experimentalmente que a parte C-terminal do domínio UBA é responsável pela interação com a ubiquitina. Os resíduos do C-terminal estão dispostos em três hélices $\alpha 1$, $\alpha 2$ e $\alpha 3$ (45) e podem ser observados na figura 2.1.

Alinhamentos entre diferentes domínios UBA, exibidos na figura 2.3, mostram alta conservação da sequência XGF. Esta sequência curta induz a formação de um bolso hidrofóbico, constituído por resíduos conservados de caráter hidrofóbicos, onde o resíduo X é frequentemente o resíduo Met. Algumas mutações de ponto único, por exemplo, M→L e F→Y (46), mantém a estabilidade do bolso hidrofóbico. O bolso hidrofóbico exibido na figura 2.1 é uma importante via que induz o reconhecimento e afinidade de ligação à ubiquitina por meio de um trecho de superfície hidrofóbica situado próximo ao bolso hidrofóbico (45, 47, 48, 49, 50, 51, 52, 53). As interações intramoleculares hélice-hélice do domínio UBA são também conduzidas pela presença do bolso hidrofóbico.

Codificada pelo gene RAD23 de *Saccharomyces cerevisiae* as proteínas enzimáticas UV RAD23 (54) e seus homólogos humanos HHR23A (55) e HHR23B (56) são exemplos de proteínas de reparo de excisão de nucleotídeo evolutivamente conservadas. Rad23 e

(a) PDB ID: 1DV0

(b) PDB ID: 2BWB

(c) PDB ID: 2CRN

(d) PDB ID: 2DAI

(e) PDB ID: 2DKL

(f) PDB ID: 2EKK

(g) PDB ID: 2MJ5

Figura 2.1 – Diferentes domínios UBA mostrando similaridade no arranjo conformacional. Ambas conformações se organizam em três hélices, induzidas pela presença do trecho superficial hidrofóbico, destacado em amarelo, típico de um bolso hidrofóbico, formado pela sequência de resíduos XGF.

todos seus homólogos contêm um domínio N-terminal *ubiquitin-like* (Ubl) e dois domínios associados à ubiquitina (UBA) sendo um deles localizado no meio e o outro domínio localizado no C-terminal da proteína. O domínio Ubl interage diretamente com o proteassomo 26 S sugerindo uma ligação entre a degradação da proteína por meio do sistema ubiquitina-proteassomo. Contudo, estudos demonstraram que deleção do domínio N-terminal Ubl de HHR23A não anula a ligação a ubiquitina, sugerindo que a interação entre ubiquitina e HHR23A deve ser confinado à parte C-terminal.

Ilustramos na figura 2.2, o domínio UBA da proteína Dsk2p complexada com ubiquitina (PDB ID: 1WR1), interagindo com a folha β da ubiquitina, promovida pela sequência M342, G343 e F344 que forma o núcleo do trecho de superfície hidrofóbica. A proteína Dsk2p de leveduras apresenta domínio UBA no C-terminal, que se liga a cadeias de poliubiquitina, transferindo proteínas poliubiquitinadas ao proteassoma para degradação (57, 58).

Figura 2.2 – Visão detalhada da superfície do domínio UBA (em magenta) da proteína Dsk2p complexada com ubiquitina (em ciano), código de acesso: PDB ID: 1WR1. Os resíduos M342, G343 e F344 são bem conservados nos domínios UBA e formam o núcleo hidrofóbico (em amarelo) do grande trecho de superfície hidrofóbica para interação com ubiquitina.

A ubiquitina também possui um trecho de superfície hidrofóbica, ao longo de sua folha- β de cinco fitas, que se estende durante a poliubiquitinação, aumentando sua afinidade de ligação (59). Dados sugerem que a ligação entre o domínio UBA e ubiquitina ocorre entre seus dois bolsos superficiais hidrofóbicos (57, 59).

Figura 2.3 – Alinhamento entre diferentes domínios UBA mostrando alta conservação dos resíduos XGF. As setas estão direcionadas para os resíduos que constituem o bolso hidrofóbico.

2.1 Mutações desestabilizadoras do bolso hidrofóbico no domínio UBA

Algumas mutações dos resíduos MGF desestabilizam a estrutura do domínio UBA (60) e interrompem sua ligação à ubiquitina, por causa do desenovelamento parcial. Com a perda da sua estrutura secundária ocorre um aumento do raio de giração (R_g) caracterizado por um maior volume conformacional e menor compactação. Desse modo, a perturbação do domínio UBA causada pela mutação da sequência MGF do seu bolso hidrofóbico, leva a um desfavorecimento energético impossibilitando a ligação com UB (60). Estudos demonstraram que a mutação única em duas posições idênticas, E176A no domínio UBA da proteína HHR23A e L336A no domínio UBA da proteína Ddi1, não anulou as ligações à ubiquitina, pelo fato desta mutação não fazer parte do bolso hidrofóbico. Por outro lado, a mutação única em duas posições idênticas, L198A no domínio UBA da proteína HHR23A e L355A no domínio UBA da proteína Ddi1, resultou em perda de ligação com ubiquitina, pois o resíduo de leucina é altamente conservado e enterrado no interior do bolso hidrofóbico (59).

Experimentos mutagênicos com o domínio UBA da proteína Dsk2p, revelaram que mutações únicas G343A, F344A e L369A, produziram uma grande diminuição de sua afinidade de ligação para a ubiquitina (57). Estes resíduos altamente conservados participam do bolso hidrofóbico e estão enterrados completamente no interior do núcleo no feixe de três hélices.

Assim, o papel do domínio UBA é extremamente importante. Por exemplo, a doença óssea de *Page* (61) está associada a mutação do resíduo P392L no domínio UBA da proteína p62, a qual aumenta a formação de osteoclastos, célula responsável pela formação da matriz óssea. Esta mutação na p62 pode resultar na desestabilização do domínio UBA e perda de função. Como consequência, haverá uma excessiva reabsorção óssea, seguida de abundante formação de osso novo, com eventual substituição da medula óssea normal por tecido vascular e fibroso (61). O domínio UBA é de fato um importante regulador das vias metabólicas, o que motiva sua conservação ao longo da evolução.

2.2 Mecanismo molecular na complexação RBD-ACE2

Enquanto o RBD-*spike* é determinante para a interação vírus-receptor, os resíduos ativos localizados no RBM desempenham um papel fundamental na ligação da superfície externa de seus receptores. A importância do RBM, que faz parte do RBD, é explorada mais detalhadamente aqui em relação à sua topologia estrutural. Assim, em vez de apenas analisarmos resíduos específicos que fazem contato com ACE2 após a ligação, damos um passo adiante e rastreamos a origem molecular que conduz a ligação viral neste receptor celular. Através de recursos computacionais e de bioinformática descobrimos uma região altamente conservada constituída pelos resíduos Tyr-Gly-Phe (YGF) em cepas de coronavírus que usam a enzima conversora de angiotensina 2. Consequentemente, hipotetizamos que a sequência curta YGF é vital para a interação RBD-ACE2 devido à formação de um bolso hidrofóbico adequado à especificidade do receptor. Além disso, reconhecemos no capítulo anterior, que um mecanismo de ligação semelhante é característico da interação entre as proteínas UBA-ubiquitina. Neste sentido, lançamos a hipótese de que as cepas SARS-CoV e SARS-CoV-2 capazes de usar proteínas ACE2 provavelmente adaptaram seu RBM durante a evolução viral com o intuito de explorar tal mecanismo molecular para ataque e infecção da célula hospedeira.

Capítulo 3

Métodos

Usando as ferramentas de bioinformática BLAST e ClustalW realizamos o alinhamento múltiplo para análise de sequências de aminoácidos das cepas de CoVs. O algoritmo Jalview foi usado para examinar e editar vários alinhamentos de sequência destacando a alta conservação dos resíduos YGF entre CoVs que utilizam ACE2. Usando PyMol e Chimera preparamos as figuras que mostram as complexações conformacionais, ligações de hidrogênio, interações hidrofóbicas, pontes salinas e o bolso hidrofóbico. Para todos os cálculos dos diferentes algoritmos de bioinformática aplicados neste estudo, utilizamos o PDB ID: 2AJF (34) para SARS-CoV e PDB ID: 6LZG (62) para SARS-CoV-2. Ambos modelos foram determinados experimentalmente por cristalografia de alta resolução, disponíveis no banco de dados de proteína. Usamos a cadeia E do PDB ID: 2AJF de SARS-CoV referente a proteína *spike*, que inicia no resíduo C323 e termina no resíduo E502. Para SARS-CoV-2 usamos a cadeia B do PDB ID: 6LZG referente a proteína *spike*, que inicia no resíduo T333 e termina no resíduo P527. Estes modelos foram escolhidos por conterem seus respectivos RBM, regiões onde estão localizados importantes sítios ativos envolvidos na complexação com ACE2 e analisados neste estudo.

A análise dos contatos intermoleculares complexadas com hACE2 foi realizada com a ferramenta de bioinformática *Contacts of structural units* (CSU).

Mutações de resíduo único e múltiplo foram introduzidas em posições localizadas no trecho de superfície hidrofóbica para avaliar a importância desses resíduos no estabelecimento de interações específicas com hACE2. As mutações podem afetar a interação com a proteína *spike*, levando a uma afinidade de ligação mais alta, mais baixa ou até mesmo neutra.

Para prever os efeitos das mutações na interface de ligação, utilizamos o método MutaBind2, que calcula a energia livre de ligação após mutação. A mudança de energia livre de ligação após uma única mutação também foi avaliada com o preditor BeAtMuSiC, que é baseado em um conjunto de potenciais estatísticos extraídos de dados mutacionais experimentais e com o algoritmo preditor mCSM-PPI2 analisamos a variação de energia livre de mutação única. Essas ferramentas computacionais previram efeitos muito semelhantes na complexação.

3.1 Alinhamento de sequências

O alinhamento de sequências permite calcular a porcentagem de conservação de sequências proteicas. Através dos resultados de alinhamentos é possível analisar mutações, sequências de proteínas homólogas, funções genômicas e proteicas.

Para realizar o alinhamento de sequências (ver apêndice B) é necessário um arquivo FASTA de entrada que contenha as sequências de nucleotídeos ou aminoácidos no formato de texto, figura 3.1. A sequência FASTA de nucleotídeos e proteínas pode ser obtida através dos bancos de dados UNIPROT (*Universal Protein*) e NCBI (*National Center for Biotechnology Information*) de acesso livre.

```
>SARS-CoV-Tor2/424-494 AY274119.3 GenBank-spike: AAP41037.1 SPIKE_CVHSA P59594.1
NTRNIDATSTGNVNYKYRYLHGKLRPFERDISNVFSPDGKPCPTPALNCYWPLNDYGFYTTTGIGYQPY

>Bat_SARS_Cov_Rs806/2006 GenBank: FJ588692.1 GenBank-spike: ACU31051.1 = Spike_rec_bind 321-555
NTAKQDTGNYYSRKTCLKPFFERDLSSDDGVTLSYDFYFNPVPVAYQAT

>Bat_SARS_Cov_Rs672/2006 GenBank: FJ588686.1 GenBank-spike: ACU31032.1 = Spike_rec_bind 321-554
NTAKQDQGGYYYSRKTCLKPFFERDLSSDENGVRTLSYDFYFNPVPVAYQAT

>Bat_SARS-like_Cov_Rs4255 GenBank: KY417149.1 GenBank-spike: AT098193.1 = Spike_rec_bind 321-554
NTAKQDQGGYYYSRKTCLKPFFERDLSSDENGVRTLSYDFYFNPVPVAYQAT

>Bat_SARS-like_Cov_Rs4081 GenBank: KY417143.1 GenBank-spike: AT098120.1 = Spike_rec_bind 321-554
NTAKQDQGGYYYSRKTCLKPFFERDLSSDENGVRTLSYDFYFNPVPVAYQAT

>Bat_SARS-like_Cov_Rs4247 GenBank: KY417148.1 GenBank-spike: AT098181.1 = Spike_rec_bind 321-556
NTAKQDTGNYYSRKTCLKPFFERDLSSDDGVTLSYDFYFNPVPVAYQAT

>Bat_SARS-like_Cov_Rs4237 GenBank: KY417147.1 GenBank-spike: AT098169.1 = Spike_rec_bind 321-554
NTAKQDQGGYYYSRKTCLKPFFERDLSSDENGVRTLSYDFYFNPVPVAYQAT

>Bat_SARS-like_Cov_A56526 GenBank: KY417142.1 GenBank-spike: AT098108.1 = Spike_rec_bind 321-554
NTAKQDQGGYYYSRKTCLKPFFERDLSSDENGVRTLSYDFYFNPVPVAYQAT

>Bat_SARS_Cov_RP3/2004 GenBank: DQ071615.1 GenBank-spike: AAZ67052.1 = RBD 310-513
NTAKQDQGGYYYSRKTCLKPFFERDLSSDENGVRTLSYDFYFNPVPVAYQAT
```

Figura 3.1 – Exemplo de arquivo FASTA de entrada para alinhamento que contém as sequências de aminoácidos no formato de texto. A sequência FASTA de nucleotídeos e proteínas podem ser obtidas através dos bancos de dados UNIPROT e NCBI de acesso livre. O arquivo FASTA é composto pelo cabeçalho e as sequências. Todo cabeçalho FASTA inicia com o sinal de > (maior).

3.2 GenBank

GenBank é um banco de dados muito utilizado no campo da bioinformática que armazena informações de sequências nucleotídicas e traduções de milhares de espécies. Os dados são obtidos principalmente por meio de apresentações de laboratórios individuais e apresentações em lote de projetos de sequenciamento em grande escala, como projeto genoma e projeto metagenoma. O GenBank é de acesso livre permitindo que toda comunidade científica utilize-o sem restrição. Não é exigido nenhum tipo de registro para acessar os dados depositados no GenBank. O acesso para o uso, cópia, distribuição de informação é ilimitado. Esse banco de dados é mantido pelo *National Center for Biotechnology Information* (NCBI) (63).

3.3 NCBI

O NCBI é um banco de dados de sequências genômicas, de proteínas e de variações evolutivas muito importante para serviços de bioinformática. Os dados revelam regiões de codificação, domínios conservados, mutações, referências, nomes, referências cruzadas de banco de dados. Fontes de informação como, GenBank, PubMed, LocusLink, dados do Projeto Genoma Humano e BLAST fazem parte do NCBI (64, 65).

3.4 BLAST

Desenvolvido em 1990 por David J. Lipman, Stephen Frank Altschul, dentre outros co-autores. BLAST é um algoritmo desenvolvido para uso no campo da bioinformática em análise de sequência de proteínas e nucleotídeos. O sucesso deste algoritmo está relacionado com sua rapidez em gerar saídas de comparação de sequências com estimativa de significância. Atualmente está implementado com NCBI - *National Center for Biotechnology Information* e com WU-BLAST - *University of Washington*. BLAST analisa a conservação de sequências através de alinhamentos que otimizam uma medida de similaridade local. O algoritmo pode ser aplicado em uma variedade de contextos, incluindo pesquisas diretas de banco de dados de sequências de DNA e proteínas, pesquisas de motivos, pesquisas de identificação de genes e na análise de várias regiões de similaridade em longas sequências de DNA (66). O BLAST trabalha com diferentes domínios de busca que depende da

Algoritmo	Aplicação
blastn	Compara duas sequências de nucleotídeos
blastp	Compara duas sequências de aminoácidos
blastx	Compara uma sequência de nucleotídeo com uma sequência de proteína
tblastn	Compara uma sequência de proteína com uma sequência de nucleotídeo
tblastx	Compara os frames tradicionais entre sequência de nucleotídeos

Tabela 3.1 – Diferentes algoritmos BLAST para distintas aplicações. Os diferentes domínios de busca dependem da sequência de entrada.

sequência de entrada como apresentado na tabela 3.1. Realiza o alinhamento ótimo par a par, obtendo o melhor alinhamento possível.

3.5 CLUSTAL W

O algoritmo CLUSTAL W (67), desenvolvido em 1994 por Julie D. Thompson, Toby J. Gibson, Des G. Higgins, é utilizado para realizar o alinhamento múltiplo automático de sequências de nucleotídeos ou aminoácidos para análise de conservação de sequência.

O alinhamento múltiplo não é realizado de forma ótima, mas utilizando vários algoritmos que calculam possíveis aproximações na obtenção de bons resultados, mas não necessariamente o melhor resultado. O CLUSTAL W realiza alinhamento progressivo. A heurística usada no Clustal W se baseia na análise filogenética (68). Primeiramente o algoritmo faz o alinhamento par a par utilizando a matriz de distância que gera uma árvore guia, a partir dos resultados, alinha par a par as sequências evolutivamente mais próximas gerando o alinhamento múltiplo. Enquanto o BLAST foi utilizado para o alinhamento de duas sequências, o CLUSTAL W foi apresentado como um dos principais algoritmos na obtenção de alinhamentos múltiplos de sequências.

3.6 Jalview

Através do alinhamento de múltiplas sequências pode se compreender diversas funções de sequências de ácido nucleico e proteínas relacionadas. Esses programas automáticos não possibilita o usuário editar esse alinhamento múltiplo da melhor maneira viável para uma melhor visualização e compreensão. No entanto, sabe-se que os alinha-

mentos automáticos de múltiplas sequências podem frequentemente ser melhorados pela edição manual. Portanto, são necessárias ferramentas de edição e visualização. Devido ao crescimento nos bancos de dados de sequências, vários alinhamentos podem ser grandes e difíceis de visualizar com eficiência. Em 2004 Michele Clamp, James Cuff, Stephen M Searle, Geoffrey J. Barton criaram o editor de alinhamento Jalview Java que permite a rápida visualização e edição de grandes alinhamentos de múltiplas sequências (69).

3.7 PyMol

PyMOL é um sistema gráfico molecular gratuito. Fornece a maioria dos recursos e desempenho dos gráficos moleculares tradicionais, pacotes escritos em C ou Fortran 4, interpretador Python, que possibilita inúmeros recursos. PyMOL suporta a maioria das representações comuns para estruturas macromoleculares. Rótulos podem ser exibidos para átomos e ligações tracejadas podem ser usadas para indicar interações de ligações de hidrogênio e distâncias. As moléculas podem ser carregadas de arquivos encontrados em banco de dados, por exemplo RCSB PDB que é um banco de dados de proteínas. Através do PyMol é possível obter figuras com qualidade de publicação, caracterizando interações moleculares. É possível criar animações através do carregamento de trajetórias utilizando sequências numeradas de imagens. O PyMOL também funciona por linha de comando, semelhante a um interpretador Python que possui todas as funções de edição e renderização integradas do PyMOL (70).

3.8 UCSF Chimera

Chimera é um programa que fornece serviços de visualização interativa, geração de imagens e animações de alta qualidade de publicação, incluindo mapas de densidade, trajetórias e alinhamentos de sequência. As moléculas podem ser carregadas de arquivos encontrados em banco de dados, por exemplo RCSB PDB que é um banco de dados de proteínas. O programa é gratuito para uso acadêmico, governamental, sem fins lucrativos e pessoal. Está disponível para Microsoft Windows, Linux, Apple Mac OS X, SGI IRIX e HP Tru64 Unix (71).

3.9 PDB

O banco de dados de proteínas (PDB) foi fundado em 1971 para fornecer um repositório de dados estruturais tridimensionais (3D) de macromoléculas biológicas, incluindo proteínas, ácidos nucleicos e grandes complexos macromoleculares. A determinação estrutural é realizada através do uso de técnicas experimentais, como cristalografia de raios-X, RMN e microscopia eletrônica. O acesso é livre sendo realizado através de vários sites, como por exemplo o RCSB PDB. A supervisão do PDB é feita pelo Banco de Dados Mundial de Proteína (wwPDB). O arquivo PDB contém dados de coordenadas 3D, informações sobre o conteúdo químico, como sequência de polímero e química do ligante, informações sobre o experimento usado para derivar a estrutura e algumas descrições qualitativas da estrutura. Os dados contidos no PDB são gerados e enviados por cientistas de todo o mundo a sites nos Estados Unidos, Europa e Ásia. Os avanços na clonagem, expressão, marcação e purificação de proteínas até a determinação da estrutura resultaram em um rápido aumento na taxa de determinação de novas estruturas de proteínas (72).

3.10 CSU software

O software *Contacts of structural units* (CSU) é usado para calcular contatos de várias unidades estruturais principais em proteínas. O arquivo PDB é usado como entrada. Como saída, o programa produz todos os contatos resíduo-resíduo formados por qualquer estrutura secundária, mostrando seus contatos específicos, acessibilidade ao solvente, tipos de contatos atômicos formados pelo resíduo (73).

3.11 MutaBind2

Durante o decorrer da evolução, muitas proteínas passam por mutações que alteram a afinidade de ligação. Essas mutações podem causar perturbações e perda da função. Para avanços na promoção de saúde e biotecnologia o entendimento dos impactos das variantes é extremamente relevante. O programa MutaBind2 usa um método computacional que avalia os efeitos dos resíduos mutantes numa sequência proteica e calcula de forma quantitativa os impactos das mutações na afinidade de ligação. Os impactos das mutações de ponto único e múltiplas nas interações proteína-proteína são estimados. O programa emprega

recursos com descrição das interações das proteínas com o solvente, a conservação evolutiva do local e a estabilidade termodinâmica do complexo e de cada monômero, mostrando os efeitos das mutações que aumentam ou diminuem a afinidade de ligação. MutaBind2 usa o banco de dados SKEMPI2 (74) que possui valores medidos experimentalmente de mudanças na energia livre de ligação em mutações de aminoácidos únicos e múltiplos nos complexos estruturais.

O método MutaBind2 é baseado em campos de força da mecânica molecular, potenciais estatísticos e algoritmos de otimização de cadeia lateral. Alguns procedimentos e parâmetros são essenciais para a qualidade e confiabilidade do modelo no cálculo do valor de $\Delta\Delta G$ de Gibbs, uma medida da estabilidade termodinâmica do complexo proteico. A otimização da geometria através da minimização de energia visa encontrar um conjunto de coordenadas que minimizam a energia potencial do sistema, relaxando ligações, ângulos e contatos das proteínas mutantes e do tipo natural. Desse modo, parâmetros como energia de interação de *Van der Waals*, energias polares de solvatação, áreas de superfície acessíveis ao solvente entre outros, são aplicados entre a proteína do tipo natural e no cálculo da variação da energia livre de ligação após a mutação, através da equação

$$\Delta\Delta G = \Delta G^{\text{mut}} - \Delta G^{\text{nat}}$$

Para uma mutação ser classificada como desestabilizadora os autores identificam a condição $\Delta\Delta G \geq 1,57 \text{ kcal mol}^{-1}$. Para definir a confiança da predição para mutações interfaciais desestabilizadoras é aplicada a curva ROC (Característica de Operação do Receptor), um estimador estatístico de análise do desempenho dos modelos preditos. Nas mutações interfaciais desestabilizadoras de alta confiança os autores identificam a condição $\Delta\Delta G \geq 2,24 \text{ kcal mol}^{-1}$. Assim, mutação desestabilizadora de baixa confiança o $\Delta\Delta G < 2,24 \text{ kcal mol}^{-1}$ e $\Delta\Delta G \geq 1,57 \text{ kcal mol}^{-1}$. A mutação interfacial neutra com alta confiança apresenta $\Delta\Delta G \leq 0,86 \text{ kcal mol}^{-1}$. É considerada mutação neutra de baixa confiança quando $\Delta\Delta G > 0,86 \text{ kcal mol}^{-1}$ e $\Delta\Delta G < 1,57 \text{ kcal mol}^{-1}$ (75).

As predições do MutaBind2 comparadas com as mutações experimentais são medidas pelo coeficiente de correlação de Pearson (ρ). Os valores variam de $\rho=0.61$ a $\rho=0.85$, indicando desde correlação moderada a alta (6, 75).

3.12 BeAtMuSiC

BeAtMuSiC é um algoritmo preditor de mudanças na afinidade de ligação proteína-proteína em mutações pontuais. O cálculo é baseado em um conjunto de potenciais estatísticos adaptados a uma representação do tipo granulação grossa de estruturas de proteínas, o que permite a avaliação rápida de todas as mutações possíveis no complexo de proteínas. O cálculo do impacto das mutações no complexo de interesse é resultado do uso da mecânica molecular combinada com modelo de solvente contínuo devido ao seu menor custo computacional e desempenho considerável. Os resultados do BeAtMuSiC indicam diminuição na afinidade de ligação após mutação se $\Delta\Delta G > 0$ e aumento na afinidade de ligação após mutação se $\Delta\Delta G < 0$.

3.13 mCSM-PPI2

mCSM-PPI2 é um algoritmo projetado para prever os efeitos de mutações de ponto único na afinidade de ligação de interação proteína-proteína. mCSM-PPI2 usa os dados derivados da versão 2.0 do banco de dados SKEMPI, que compila dados experimentais sobre mudanças nos parâmetros termodinâmicos e cinéticos na mutação para complexos proteína-proteína, que têm estruturas 3D depositadas no PDB. SKEMPI 2.0 inclui novas mutações identificadas na literatura. A afinidade de ligação dos complexos proteína-proteína foi usada para calcular a energia livre de Gibbs de ligação

$$\Delta G = -RT \ln(K_D)$$

onde R é a constante de gás ideal, T é a temperatura (em Kelvin) e K_D é a constante de dissociação de equilíbrio do complexo proteína-proteína (em molar). A mudança na afinidade de ligação após a mutação foi calculada da seguinte forma

$$\Delta\Delta G_{\text{nat-mut}} = \Delta G_{\text{tipo natural}} - \Delta G_{\text{tipo mutante}}$$

O programa mCSM define $\Delta\Delta G$ em relação ao mutante para o tipo natural. Uma vez que a formulação de energia livre de Gibbs é uma função de estado termodinâmico, uma mudança na afinidade de ligação de uma mutação de uma proteína de tipo natural para seu mutante ($\Delta\Delta G_{\text{nat} \rightarrow \text{mut}}$) deve ser equivalente à energia livre de ligação da mudança negativa de mutação reversa, do mutante para a proteína de tipo natural ($-\Delta\Delta G_{\text{mut} \rightarrow \text{nat}}$).

Capítulo 4

Resultado e discussão

As sequências RBD SARS-CoV/SARS-CoV-2 são altamente similares, observado pelo RMSD de 1,2 Å entre átomos de C α alinhados. Mesmo no RBM, que possui maior variação de sequência, as estruturas conformacionais entre SARS-CoV/SARS-CoV-2 apresentam RMSD de 1,3 Å (76). SARS-CoV e SARS-CoV-2 utilizam seu RBD de forma semelhante para ligação ao receptor ACE2 (39).

Aqui, investigamos a ocorrência e a importância da sequência específica de resíduos de aminoácidos YGF para as cepas SARS-CoV e SARS-CoV-2 capazes de usar proteínas ACE2 como receptores. A figura 4.1 mostra este mecanismo molecular. As sequências RBM YGFY (481-484) da figura 4.1 (a) e RBM YGF (495-497) da figura 4.1 (b) formam um bolso hidrofóbico exibido em amarelo. Essa conformação induz a ligação à ACE2 em consequência das posições atômicas favoráveis.

Em SARS-CoV/SARS-CoV-2, apresentados na figura 4.2, os resíduos N479/Q493, T487/N501 são essenciais para a ligação ao receptor hospedeiro (27, 34, 77) e os resíduos Y442/L455, L472/F486, D480/S494 são importantes contatos de interação ao receptor (33). A figura 4.3 (a) exhibe os resíduos RBM SARS-CoV em contato direto com hACE2, determinado pelas propriedades hidrofóbicas-hidrofílicas dos resíduos de interface, previsto pelo programa CSU (73). Esta rede de contatos destaca a importância dos resíduos que estão localizados perto da sequência YGF e contribui para a estabilização do SARS-CoV complexado com hACE2. Por exemplo, Y475 faz contatos de ligação de hidrogênio com Q24, F28 e Y83; N479 com K31 e H34; Y486 com Y41, N330 e R357; e T487 com Y41.

Os seres vivos em geral, inclusive os vírus, passam por mutações como resposta à seleção natural. Os vírus formados por RNA, como exemplo do coronavírus, acumulam mutações a cada ciclo reprodutivo (28). A figura 4.4 compara as sequências de aminoácidos

Figura 4.1 – (a) Resíduos YGFY na interface do SARS-CoV complexado com ACE2 humano (PDB ID: 2AJF). Esses resíduos (em amarelo) formam um bolso hidrofóbico localizado no RBM (cor magenta). (b) Resíduos YGF e EGF (em amarelo) na interface de SARS-CoV-2 complexados com ACE2 humano (PDB ID: 6LZG). A primeira sequência YGF está localizada num bolso hidrofóbico (em amarelo), enquanto que a segunda sequência EGF (em amarelo) que está na superfície RBM (em magenta). Representação em fita do ACE2 (em azul).

Figura 4.2 – Diagramas de fita de SARS-CoV (a) e SARS-CoV-2 (b) complexados com ACE2 humano (cor azul), onde o RBM é destacado em cor magenta. Os principais resíduos responsáveis pela ligação estrutural são exibidos na representação *stick*. Os resíduos N479/Q493, T487/N501 são essenciais para a ligação ao receptor hospedeiro e os resíduos Y442/L455, L472/F486, D480/S494 são importantes contatos de interação ao receptor.

de cepas de SARS-CoV e SARS-CoV-2 humanas alinhadas com RBM de SARS-CoV Tor2, uma cepa epidêmica isolada de humanos durante a epidemia de SARS. A cepa Tor2 humana possui alta afinidade para hACE2 (26). Destacamos nesta figura em cor púrpura a sequência hidrofóbica YGFY típica de SARS-CoV, ocorrendo nas posições 481-484 na proteína *spike*. A sequência mutada correspondente, ocorre agora nas posições 495-498 na proteína *spike* SARS-CoV-2. Nessa mutação ocorreu substituição de um aminoácido hidrofóbico de SARS-CoV por um hidrofílico em SARS-CoV-2.

Importantes resíduos de interface de ligação de SARS-CoV foram mutados no SARS-CoV-2. Os alinhamentos de sequência mostram o mapeamento, Y442→L455, L472→F486, N479→Q493, D480→S494, T487→N501. Essas mutações não apresentam uma mudança drástica em seu caráter hidrofóbico (78), preservando assim a estrutura topológica geral de ligação do vírus ao receptor. Mutações nestes resíduos de interface afetam o reconhecimento do receptor ACE2 e a variedade de hospedeiros infectados pelo vírus (38). Dados sugerem que SARS-CoV-2 preservou suas interações favoráveis com hACE2, sendo observadas entre os resíduos L455, Q493 de SARS-CoV-2 e E35, K31 de hACE2 (79) figura 4.2.

Ilustramos na figura 4.5, em SARS-CoV/SARS-CoV-2, os resíduos Y442/L455, N479/Q493 e D480/S494 localizados perto da ponte salina K31-E35 no receptor hACE2 e os resíduos T487/N501 próximos à uma segunda ponte salina D38-K353 no receptor hACE2. Essas pontes salinas são importantes regiões de interações devido a presença dos resíduos

Figura 4.4 – Alinhamentos de sequência de CoVs humanos restritos a resíduos RBM. A cor púrpura destaca o padrão YGFY seguido pela mutação Y498Q no RBM das cepas SARS-CoV-2.

Lys31 e Lys353 de hACE2 promovendo dois importantes pontos de acesso de ligação ao vírus. Estas pontes salinas de hACE2 estão enterradas numa região hidrofóbica viabilizando o aumento da energia livre que favorece a complexação receptor/vírus. Observando a

Figura 4.5 – Ponte salina no receptor hACE2. SARS-CoV (a) resíduos RBM (cor magenta) Y442, N479 e D480 estão localizados perto da ponte salina K31-E35 no receptor hACE2 (em ciano), (b) resíduo RBM T487 está próximo à uma segunda ponte salina D38-K353 no receptor hACE2. SARS-CoV-2 (c) resíduos RBM (cor magenta) L455, Q493 e S494 estão localizados perto da ponte salina K31-E35 no receptor hACE2 (em ciano) e (d) o resíduo N501 está próximo à uma segunda ponte salina D38-K353 no receptor hACE2. Os resíduos Lys31 e Lys353 de hACE2 estão enterrados numa região hidrofóbica e são dois importantes pontos de acesso de ligação ao vírus. Os principais resíduos responsáveis pela ligação estrutural são exibidos na representação *stick*.

figura 4.6 (a), na interface RBM SARS-CoV/hACE2 o ponto de acesso 31 de hACE2 precisa de suporte do resíduo Y442 de SARS-CoV. Já na interface RBM SARS-CoV-2/hACE2 (b), o resíduo L455 de SARS-CoV-2 (correspondente ao resíduo Y442 de SARS-CoV) tem uma cadeia lateral menos volumosa, fornecendo menos suporte para K31 de hACE2. Consequentemente a estrutura do ponto de acesso 31 de hACE2 foi reorganizado: a ponte salina entre K31-E35 de hACE2 se quebra e ambos resíduos formam ligação de hidrogênio com Q493 de SARS-CoV-2 (80). Na interface RBM SARS-CoV/hACE2, o ponto de acesso 353 de hACE2 requer suporte da cadeia lateral de T487 de SARS-CoV, através de interação hidrofóbica. Em comparação, na interface RBM SARS-CoV-2/hACE2, o N501 de SARS-CoV-2 tem sua conformação fixada por meio de ligação de hidrogênio entre

Figura 4.6 – Detalhes estruturais após complexação de (a) SARS-CoV e (b) SARS-CoV-2 complexados com ACE2 humano (cor azul), onde o RBM é destacado em cor magenta. Os principais resíduos responsáveis pela ligação estrutural são exibidos na representação *stick*.

sua cadeia lateral e o ponto de acesso K353 de hACE2, porém sua cadeia lateral fornece menos suporte que T487 (resíduo correspondente de SARS-CoV). Consequentemente, na interface RBM SARS-CoV-2/hACE2, o resíduo K353 de hACE2 assume uma conformação ligeiramente diferente mantendo a ponte salina com D38 de hACE2. Assim, ambos os pontos de acesso se ajustaram ao suporte reduzido de resíduos de RBD próximos, se tornando estáveis nesta interface (80).

Mutações virais naturais que circundam pontos de acesso de ligação do vírus promovem impactos significativos na afinidade de ligação, estrutura e resposta imune do hospedeiro. Por exemplo, mutação K479N no RBM facilitou a transmissão de SARS-CoV de civetas de palma para humanos. Na mutação S487T no RBM facilitou a transmissão de SARS-CoV de humano para humano. Essas duas mutações contribuíram significativamente para a epidemia de SARS em 2002 a 2003 (25, 26, 78). Em relação a SARS-CoV-2, mutação nos resíduos L472→F486 desencadeia a formação de outra sequência baseada em GF figura 4.1: (b) no RBM de SARS-CoV-2, produzindo uma pequena superfície hidrofóbica que não atrapalha o mecanismo molecular proposto para a ligação a ACE2. Nenhuma outra sequência baseada em GF aparece em sua RBD.

A figura 4.3 (b) exhibe os resíduos de RBM SARS-CoV-2 em contato direto com hACE2 conforme previsto pelo programa CSU, mostrando novamente a importância dos resíduos próximos ao bolso hidrofóbico. Os detalhes das interfaces de ligação proteína-proteína podem ser bastante diferentes entre as cepas, provavelmente relacionados ao seu grau de infectividade. Provavelmente exista certas particularidades no complexo de

ligação entre as diferentes cepas do tipo SARS com o receptor hospedeiro, levando ao seu grau de infectividade. Foi observado que as mutações no resíduo RBM T487 têm um papel importante na transmissão humano→humano e animal→humano do SARS-CoV (25, 26, 77). Em dezembro de 2020 foi descoberto uma variante de SARS-CoV-2, linhagem B.1.1.7, também conhecida como VOC 202012/01 ou 20I/501Y.V1. A variante aumentou a transmissão do SARS-CoV-2 no ano de 2021. A cepa mutante B.1.1.7 carrega uma mutação na proteína S, N501Y, que afeta a conformação do RBD (81, 82). Estudos apontam que a mutação N501Y é responsável pelo aumento da infectividade. Foi demonstrado que a mutação N501Y provocou aumento da força de interação entre RBM SARS-CoV-2/ACE2 de ratos (83).

Agora, investigamos a relevância do bolso hidrofóbico formado pela sequência YGFY na promoção da estabilidade do *spike* RBD de SARS-CoV complexado com ACE2 realizando uma série de mutações usando o método MutaBind2 (6) para avaliar a mudança na afinidade de ligação através da $\Delta\Delta G$ em SARS-CoV. A presença de bolsos hidrofóbicos nas proteínas não são tão comuns e, quando presentes, costumam ser locais de ligação entre proteínas (59). Para validação de nossas principais conclusões, realizamos mutações únicas, conforme previsto pelas ferramentas de bioinformática BeAtMuSiC (7) e mCSM-PPI2 (8). Inicialmente, investigamos a influência da mutação N479 pelos resíduos E, K, Q, R e S na complexação (tabela 4.1). As mutações N479E, N479K, N479Q, N479R e N479S foram observadas, respectivamente, em cepas de pangolim (ver figura 4.8), cepas de morcego e civeta de palma (ver figura 4.7 e 4.8), SARS-CoV-2 humano (ver figura 4.4), cepas de morcego e civeta de palma (ver figura 4.7 e 4.8), e cepas de morcego (figura 4.7).

Natural	N479 (SARS-CoV)
Variante	
E	N479E
K	N479K
Q	N479Q
R	N479R
S	N479S

Tabela 4.1 – Mutação de ponto único N479 SARS-CoV pelos resíduos E (pangolins), K (morcego e civeta de palma), Q (SARS-CoV-2), R (morcego e civeta de palma) e S (morcego) na complexação.

SARS-CoV_Tor2	424	NTRN	I	D	A	T	S	T	G	N	Y	N	K	R	L	R	H	G	K	L	R	P	F	E	R	D	I	S	N	V	F	S	P	D	G	K	P	C	T	P	-	P	A	L	N	C	Y	W	P	L	N	D	Y	G	F	Y	T	T	G	I	G	Y	O	P	Y	494		
Rhinolophus_affinis_CoV_LYRa3	428	NTRN	I	D	A	T	S	S	G	N	F	N	K	R	L	R	H	G	K	L	R	P	F	E	R	D	I	S	N	V	F	S	P	D	G	K	P	C	T	P	-	P	A	F	N	C	Y	W	P	L	N	D	Y	G	F	Y	T	T	G	I	G	Y	O	P	Y	496		
Rhinolophus_affinis_CoV_LYRa11	428	NTRN	I	D	A	T	S	S	G	N	F	N	K	R	L	R	H	G	K	L	R	P	F	E	R	D	I	S	N	V	F	S	P	D	G	K	P	C	T	P	-	P	A	F	N	C	Y	W	P	L	N	D	Y	G	F	Y	T	T	G	I	G	Y	O	P	Y	496		
Bat_SARS-like_CoV_Rs3367	425	NTRN	I	D	A	T	Q	T	G	N	Y	N	K	R	L	R	H	G	K	L	R	P	F	E	R	D	I	S	N	V	F	S	P	D	G	K	P	C	T	P	-	P	A	F	N	C	Y	W	P	L	N	D	Y	G	F	Y	T	T	G	I	G	Y	O	P	Y	495		
Bat_SARS-like_CoV_Rs4874	424	NTRN	I	D	A	T	Q	T	G	N	Y	N	K	R	L	R	H	G	K	L	R	P	F	E	R	D	I	S	N	V	F	S	P	D	G	K	P	C	T	P	-	P	A	F	N	C	Y	W	P	L	N	D	Y	G	F	Y	T	T	G	I	G	Y	O	P	Y	494		
Bat_SARS-like_CoV_WIV1	425	NTRN	I	D	A	T	Q	T	G	N	Y	N	K	R	L	R	H	G	K	L	R	P	F	E	R	D	I	S	N	V	F	S	P	D	G	K	P	C	T	P	-	P	A	F	N	C	Y	W	P	L	N	D	Y	G	F	Y	T	T	G	I	G	Y	O	P	Y	495		
SARS-like CoV WIV16	424	NTRN	I	D	A	T	Q	T	G	N	Y	N	K	R	L	R	H	G	K	L	R	P	F	E	R	D	I	S	N	V	F	S	P	D	G	K	P	C	T	P	-	P	A	F	N	C	Y	W	P	L	N	D	Y	G	F	Y	T	T	G	I	G	Y	O	P	Y	494		
Bat CoV RaTG13	437	NSKH	I	D	A	K	E	G	G	N	F	N	L	R	L	R	F	R	K	A	N	L	K	P	F	E	R	D	I	S	T	E	I	Y	Q	A	G	S	T	P	C	N	G	Q	T	G	L	N	C	Y	P	L	R	Y	G	F	Y	P	T	D	G	V	G	H	O	P	Y	508
Bat SARS-like CoV Rs7327	425	NTRN	I	D	A	T	S	T	G	N	Y	N	K	R	L	R	H	G	K	L	R	P	F	E	R	D	I	S	N	V	F	S	P	D	G	K	P	C	T	P	-	P	A	F	N	C	Y	W	P	L	N	D	Y	G	F	Y	T	T	G	I	G	Y	O	P	Y	495		
Bat SARS-like CoV Rs9401	425	NTRN	I	D	A	T	S	T	G	N	Y	N	K	R	L	R	H	G	K	L	R	P	F	E	R	D	I	S	N	V	F	S	P	D	G	K	P	C	T	P	-	P	A	F	N	C	Y	W	P	L	N	D	Y	G	F	Y	T	T	G	I	G	Y	O	P	Y	495		
CoV BtRs-BetaCoV/YN2018B	425	NTRN	I	D	A	T	S	T	G	N	Y	N	K	R	L	R	H	G	K	L	R	P	F	E	R	D	I	S	N	V	F	S	P	D	G	K	P	C	T	P	-	P	A	F	N	C	Y	W	P	L	N	D	Y	G	F	Y	T	T	G	I	G	Y	O	P	Y	495		
Bat SARS-like CoV Rs5HC014	425	NTRN	I	D	A	T	S	T	G	N	Y	N	K	R	L	R	H	G	K	L	R	P	F	E	R	D	I	S	N	V	F	S	P	D	G	K	P	C	T	P	-	P	A	F	N	C	Y	W	P	L	N	D	Y	G	F	Y	T	T	G	I	G	Y	O	P	Y	495		
Bat SARS-like CoV Rs4084	425	NTRN	I	D	A	T	S	T	G	N	Y	N	K	R	L	R	H	G	K	L	R	P	F	E	R	D	I	S	N	V	F	S	P	D	G	K	P	C	T	P	-	P	A	F	N	C	Y	W	P	L	N	D	Y	G	F	Y	T	T	G	I	G	Y	O	P	Y	495		
Bat SARS-like CoV Rs4231	424	NTRN	I	D	A	T	S	T	G	N	Y	N	K	R	L	R	H	G	K	L	R	P	F	E	R	D	I	S	N	V	F	S	P	D	G	K	P	C	T	P	-	P	A	F	N	C	Y	W	P	L	N	D	Y	G	F	Y	T	T	G	I	G	Y	O	P	Y	494		
SARSr-CoV_BUKY72	428	NTRN	I	D	A	T	S	T	G	N	Y	N	K	R	L	R	H	G	K	L	R	P	F	E	R	D	I	S	N	V	F	S	P	D	G	K	P	C	T	P	-	P	A	F	N	C	Y	W	P	L	N	D	Y	G	F	Y	T	T	G	I	G	Y	O	P	Y	497		
SARS-like_CoV_BB9904/BGR/2008	432	NTRN	I	D	A	T	S	T	G	N	Y	N	K	R	L	R	H	G	K	L	R	P	F	E	R	D	I	S	N	V	F	S	P	D	G	K	P	C	T	P	-	P	A	F	N	C	Y	W	P	L	N	D	Y	G	F	Y	T	T	G	I	G	Y	O	P	Y	498		
Bat CoV_BM48-31/BGR/2008	429	NTRN	I	D	A	T	S	T	G	N	Y	N	K	R	L	R	H	G	K	L	R	P	F	E	R	D	I	S	N	V	F	S	P	D	G	K	P	C	T	P	-	P	A	F	N	C	Y	W	P	L	N	D	Y	G	F	Y	T	T	G	I	G	Y	O	P	Y	495		
CoV_BtRI-BetaCoV/SC2018	427	NTRN	I	D	A	T	S	T	G	N	Y	N	K	R	L	R	H	G	K	L	R	P	F	E	R	D	I	S	N	V	F	S	P	D	G	K	P	C	T	P	-	P	A	F	N	C	Y	W	P	L	N	D	Y	G	F	Y	T	T	G	I	G	Y	O	P	Y	495		
BtRs-BetaCoV/HuB2013	427	NTRN	I	D	A	T	S	T	G	N	Y	N	K	R	L	R	H	G	K	L	R	P	F	E	R	D	I	S	N	V	F	S	P	D	G	K	P	C	T	P	-	P	A	F	N	C	Y	W	P	L	N	D	Y	G	F	Y	T	T	G	I	G	Y	O	P	Y	495		
Bat SARS-CoV_HKU3-1	428	NTRN	I	D	A	T	S	T	G	N	Y	N	K	R	L	R	H	G	K	L	R	P	F	E	R	D	I	S	N	V	F	S	P	D	G	K	P	C	T	P	-	P	A	F	N	C	Y	W	P	L	N	D	Y	G	F	Y	T	T	G	I	G	Y	O	P	Y	495		
Bat SARS-like_CoV_bat-SL-CoVZC45	433	NTRN	I	D	A	T	S	T	G	N	Y	N	K	R	L	R	H	G	K	L	R	P	F	E	R	D	I	S	N	V	F	S	P	D	G	K	P	C	T	P	-	P	A	F	N	C	Y	W	P	L	N	D	Y	G	F	Y	T	T	G	I	G	Y	O	P	Y	495		
Bat SARS-like_CoV_bat-SL-CoVZXC21	432	NTRN	I	D	A	T	S	T	G	N	Y	N	K	R	L	R	H	G	K	L	R	P	F	E	R	D	I	S	N	V	F	S	P	D	G	K	P	C	T	P	-	P	A	F	N	C	Y	W	P	L	N	D	Y	G	F	Y	T	T	G	I	G	Y	O	P	Y	495		
Bat SARS-CoV_R11/2004	428	NTRN	I	D	A	T	S	T	G	N	Y	N	K	R	L	R	H	G	K	L	R	P	F	E	R	D	I	S	N	V	F	S	P	D	G	K	P	C	T	P	-	P	A	F	N	C	Y	W	P	L	N	D	Y	G	F	Y	T	T	G	I	G	Y	O	P	Y	495		
Bat SARS-like_CoV_R104092	421	NTRN	I	D	A	T	S	T	G	N	Y	N	K	R	L	R	H	G	K	L	R	P	F	E	R	D	I	S	N	V	F	S	P	D	G	K	P	C	T	P	-	P	A	F	N	C	Y	W	P	L	N	D	Y	G	F	Y	T	T	G	I	G	Y	O	P	Y	495		
Bat CoV Anlong-103	420	NTRN	I	D	A	T	S	T	G	N	Y	N	K	R	L	R	H	G	K	L	R	P	F	E	R	D	I	S	N	V	F	S	P	D	G	K	P	C	T	P	-	P	A	F	N	C	Y	W	P	L	N	D	Y	G	F	Y	T	T	G	I	G	Y	O	P	Y	495		
Bat CoV Anlong-112	420	NTRN	I	D	A	T	S	T	G	N	Y	N	K	R	L	R	H	G	K	L	R	P	F	E	R	D	I	S	N	V	F	S	P	D	G	K	P	C	T	P	-	P	A	F	N	C	Y	W	P	L	N	D	Y	G	F	Y	T	T	G	I	G	Y	O	P	Y	495		
Bat SARS-CoV_Rp3/2004	428	NTRN	I	D	A	T	S	T	G	N	Y	N	K	R	L	R	H	G	K	L	R	P	F	E	R	D	I	S	N	V	F	S	P	D	G	K	P	C	T	P	-	P	A	F	N	C	Y	W	P	L	N	D	Y	G	F	Y	T	T	G	I	G	Y	O	P	Y	495		
CoV BtRs-BetaCoV/YN2018C	428	NTRN	I	D	A	T	S	T	G	N	Y	N	K	R	L	R	H	G	K	L	R	P	F	E	R	D	I	S	N	V	F	S	P	D	G	K	P	C	T	P	-	P	A	F	N	C	Y	W	P	L	N	D	Y	G	F	Y	T	T	G	I	G	Y	O	P	Y	495		
Bat CoV_Rp/Shaanxi2011	427	NTRN	I	D	A	T	S	T	G	N	Y	N	K	R	L	R	H	G	K	L	R	P	F	E	R	D	I	S	N	V	F	S	P	D	G	K	P	C	T	P	-	P	A	F	N	C	Y	W	P	L	N	D	Y	G	F	Y	T	T	G	I	G	Y	O	P	Y	495		
Bat SARS-CoV_Rm1/2004	428	NTRN	I	D	A	T	S	T	G	N	Y	N	K	R	L	R	H	G	K	L	R	P	F	E	R	D	I	S	N	V	F	S	P	D	G	K	P	C	T	P	-	P	A	F	N	C	Y	W	P	L	N	D	Y	G	F	Y	T	T	G	I	G	Y	O	P	Y	495		
SARS-CoV_Tor2	424	NTRN	I	D	A	T	S	T	G	N	Y	N	K	R	L	R	H	G	K	L	R	P	F	E	R	D	I	S	N	V	F	S	P	D	G	K	P	C	T	P	-	P	A	L	N	C	Y	W	P	L	N	D	Y	G	F	Y	T	T	G	I	G	Y	O	P	Y	494		
SARS-CoV_PC4-115	424	NTRN	I	D	A	T	S	T	G	N	Y	N	K	R	L	R	H	G	K	L	R	P	F	E	R	D	I	S	N	V	F	S	P	D	G	K	P	C	T	P	-	P	A	P	N	C	Y	W	P	L	N	D	Y	G	F	Y	T	T	G	I	G	Y	O	P	Y	494		
SARS-CoV_PC4-145	424	NTRN	I	D	A	T	S	T	G	N	Y	N	K	R	L	R	H	G	K	L	R	P	F	E	R	D	I	S	N	V	F	S	P	D	G	K	P	C	T	P	-	P	A	P	N	C	Y	W	P	L	N	D	Y	G	F	Y	T	T	G	I	G	Y	O	P	Y	494		
SARS-CoV_PC4-241	424	NTRN	I	D	A	T	S	T	G	N	Y	N	K	R	L	R	H	G	K	L	R	P	F	E	R	D	I	S	N	V	F	S	P																																			

resíduos Y481 e F483 são conservados por meio de todas as cepas de SARS-CoV que analisamos neste trabalho, enquanto G482 e Y484 são mutados em algumas cepas (G482 em morcegos e Y484 em morcegos e pangolins). Portanto, mutamos G482 e Y484 por resíduos plausíveis, ou seja, aqueles que ocorrem em outras cepas para avaliar as mudanças na afinidade de ligação. Para realizar essa mutação dupla, fixamos G482D, uma mutação observada em cepas de morcegos, seguida por Y484F, Y484N e Y484T, observada em morcegos; Y484H, observado em pangolins; e Y484Q, observado em SARS-CoV-2 humano (tabela 4.2). As mudanças de energia livre para essas mutações duplas indicam fortemente a desestabilização do SARS-CoV humano complexado com hACE2. É interessante notar que G482D junto com Y484Q diminui a afinidade de ligação no SARS-CoV humano porque eles diminuem a hidrofobicidade do bolso YGFY inicial. Esta conclusão é apoiada pela

Natural	G482 (SARS-CoV)	Y484 (SARS-CoV)
Variante		
D, F	G482D,	Y484F
D, H	G482D,	Y484H
D, N	G482D,	Y484N
D, Q	G482D,	Y484Q
D, T	G482D,	Y484T

Tabela 4.2 – Mutações múltiplas específicas em G482 e Y484 de SARS-CoV do trecho hidrofóbico, por resíduos que ocorrem em outras cepas, G482D (morcego) variando com Y484F (morcego), Y484H (pangolins), Y484N (morcego), Y484Q (SARS-CoV-2), Y484T (morcego).

estrutura conformacional mutante G482D-Y484Q obtida pela ferramenta de bioinformática MutaBind2, que primeiramente minimiza a energia em 100 passos que não apresenta obstáculos estéricos entre esses resíduos e aqueles na interface com ACE2. Também utilizamos a ferramenta de previsão MutaBind2 às mutações únicas G482A e G482V para avaliar os efeitos estéricos. A quantidade fornecida por esta mutação foi de 1,07 e 1,78 kcal/mol, respectivamente para G482A e G482V. A configuração conformacional obtida por MutaBind2 para G482V indica que esta mutação contém impedimento estérico em Y436:CZ da proteína *spike*, o que poderia explicar porque essa mutação não é observada nas cepas que analisamos. Também é apresentado as mudanças na afinidade de ligação para mutações únicas de Y484, um resíduo altamente conectado com hACE2, exibido na figura 4.3 (a). As mutações Y484H, Y484N e Y484T de caráter hidrofílicos, interrompem a hidrofobicidade

do bolso YGFY desestabilizando a complexação. Em seguida, mutamos N479 seguido por mutações nos locais 482 e 484 para analisar as consequências sobre a afinidade de ligação desta mutação tripla, interrompendo o trecho de superfície hidrofóbica. Novamente, fixamos por exemplo, as mutações N479E e G482D (tabela 4.3). Os impactos deste conjunto de mutações podem ser vistos na tabela 4.4. O que foi considerado uma mutação neutra, N479E mostra alto efeito desestabilizador no novo ambiente conformacional.

Natural	N479	G482	Y484
	(SARS-CoV)	(SARS-CoV)	(SARS-CoV)
Variante			
E, D, F	N479E,	G482D,	Y484F
E, D, H	N479E,	G482D,	Y484H
E, D, N	N479E,	G482D,	Y484N
E, D, Q	N479E,	G482D,	Y484Q
E, D, T	N479E,	G482D,	Y484T

Tabela 4.3 – Mutações múltiplas no local N479 de SARS-CoV interrompendo o trecho de superfície hidrofóbica com as mutações nos locais 482 e 484. Tomemos, por exemplo, a mutação N479E (pangolins) com G482D (morcego) variando com mutações Y484F (morcego), Y484H (pangolins), Y484N (morcego), Y484Q (SARS-CoV-2), Y484T (morcego).

Efeitos desestabilizadores semelhantes no *spike* de SARS-CoV humano RBD complexado com hACE2 são obtidos para importantes resíduos de interface de ligação, T487, Y442, L472 e D480, quando substituem os resíduos que formam o bolso hidrofóbico. As mutações triplas L472P, G482D, Y484F e D480T, G482D, Y484F estabilizaram a complexação por promoverem a hidrofobicidade do bolso, exibido na tabela 4.4. Agora, repetimos o procedimento acima para investigar o papel dos resíduos YGFQ para a afinidade de ligação de SARS-CoV-2 RBD/hACE2 humano. As mutações Q498 não alteraram a afinidade de ligação para a complexação (tabela 4.5), por exemplo, a $\Delta\Delta G$ previsto para a mutação Q498Y é de cerca de 0,16 kcal/mol. Também analisamos o impacto de mutações na estabilidade de complexações duplas da sequência YGFQ, consulte a tabela 4.5. Para este fim, fixamos, por exemplo, G496D e substituímos Q498 pelos resíduos F, H, N, Y e T que aparecem em cepas de outras espécies. As mudanças previstas na afinidade de ligação por mutações indicam desestabilização de novas complexações.

	$\Delta\Delta G$ (kcal/mol)	Efeito na complexação
SARS-CoV		
N479E	0.76	neutro (**)
N479K	0.66	neutro (**)
N479Q	0.40	neutro (**)
N479R	0.08	neutro (**)
N479S	0.63	neutro (**)
G482A	1.07	neutro (*)
G482D	1.85	altamente desestabilizador (*)
G482V	1.78	altamente desestabilizador (*)
G482D, Y484F	2.40	altamente desestabilizador (**)
G482D, Y484H	2.68	altamente desestabilizador (**)
G482D, Y484N	3.24	altamente desestabilizador (**)
G482D, Y484Q	2.47	altamente desestabilizador (**)
G482D, Y484T	3.26	altamente desestabilizador (**)
Y484F	0.64	neutro (**)
Y484H	1.57	altamente desestabilizador (*)
Y484N	2.19	altamente desestabilizador (*)
Y484Q	1.44	neutro (*)
Y484T	1.92	altamente desestabilizador (*)
N479E, G482D, Y484F	4.58	altamente desestabilizador (**)
N479E, G482D, Y484H	4.90	altamente desestabilizador (**)
N479E, G482D, Y484N	4.69	altamente desestabilizador (**)
N479E, G482D, Y484Q	4.58	altamente desestabilizador (**)
N479E, G482D, Y484T	4.67	altamente desestabilizador (**)
T487S	1.23	neutro (*)
T487N	2.12	altamente desestabilizador (*)
T487D	2.78	altamente desestabilizador (**)
T487A	1.15	neutro (*)
T487V	1.03	neutro (*)
T487I	2.48	altamente desestabilizador (**)
T487V, G482D, Y484F	3.27	altamente desestabilizador (**)
T487V, G482D, Y484H	3.13	altamente desestabilizador (**)
T487V, G482D, Y484N	3.67	altamente desestabilizador (**)
T487V, G482D, Y484Q	2.29	altamente desestabilizador (**)
T487V, G482D, Y484T	3.74	altamente desestabilizador (**)
Y442L	1.31	neutro (*)
Y442S	2.46	altamente desestabilizador (**)
Y442W	0.94	neutro (*)
Y442R	1.24	neutro (*)
Y442W, G482D, Y484F	2.94	altamente desestabilizador (**)
Y442W, G482D, Y484H	3.96	altamente desestabilizador (**)
Y442W, G482D, Y484N	4.17	altamente desestabilizador (**)
Y442W, G482D, Y484Q	3.71	altamente desestabilizador (**)
Y442W, G482D, Y484T	3.86	altamente desestabilizador (**)
L472P	0.03	neutro (**)
L472F	0.18	neutro (**)
L472P, G482D, Y484F	1.49	neutro (*)
L472P, G482D, Y484H	3.80	altamente desestabilizador (**)
L472P, G482D, Y484N	3.46	altamente desestabilizador (**)
L472P, G482D, Y484Q	3.04	altamente desestabilizador (**)
L472P, G482D, Y484T	3.48	altamente desestabilizador (**)
D480G	0.26	neutro (**)
D480R	-0.32	neutro (**)
D480S	0.14	neutro (**)
D480P	-0.2	neutro (**)
D480T	0.01	neutro (**)
D480T, G482D, Y484F	1.3	neutro (*)
D480T, G482D, Y484H	1.81	altamente desestabilizador (*)
D480T, G482D, Y484N	2.06	altamente desestabilizador (*)
D480T, G482D, Y484Q	1.65	altamente desestabilizador (*)
D480T, G482D, Y484T	2.54	altamente desestabilizador (**)

Tabela 4.4 – Alterações na afinidade de ligação do *spike* RBD de SARS-CoV complexado com ACE2 após mutação, conforme previsto pelo método MutaBind2 (6). Efeitos de previsão são classificados como previsão de baixa confiança: (*), ou previsão de alta confiança: (**). Aqui, adotamos a classificação apresentada no método MutaBind (6) devido à alta similaridade entre as curvas COR (características operacionais do receptor) em ambos os métodos.

	$\Delta\Delta G$ (kcal/mol)	Efeito na complexação
SARS-CoV-2		
Q493E	0.51	neutro (**)
Q493K	0.69	neutro (**)
Q493N	-0.07	neutro (**)
Q493R	0.25	neutro (**)
Q493S	0.89	neutro (*)
G496D	0.96	neutro (*)
G496A	1.10	neutro (*)
G496V	1.69	altamente desestabilizador (*)
G496D, Q498F	2.75	altamente desestabilizador (**)
G496D, Q498H	2.52	altamente desestabilizador (**)
G496D, Q498N	1.72	altamente desestabilizador (*)
G496D, Q498Y	2.12	altamente desestabilizador (*)
G496D, Q498T	1.68	altamente desestabilizador (*)
Q498F	1.34	neutro (*)
Q498H	1.06	neutro (*)
Q498N	0.88	neutro (*)
Q498Y	0.16	neutro (**)
Q498T	0.86	neutro (**)
Q493S, G496D, Q498F	3.63	altamente desestabilizador (**)
Q493S, G496D, Q498H	4.33	altamente desestabilizador (**)
Q493S, G496D, Q498N	2.96	altamente desestabilizador (**)
Q493S, G496D, Q498Y	3.44	altamente desestabilizador (**)
Q493S, G496D, Q498T	2.85	altamente desestabilizador (**)
N501S	0.1	neutro (**)
N501T	-1.24	neutro (**)
N501D	0.89	neutro (*)
N501A	0.16	neutro (**)
N501V	-0.31	neutro (**)
N501I	1.36	neutro (*)
N501D, G496D, Q498F	3.7	altamente desestabilizador (**)
N501D, G496D, Q498H	3.38	altamente desestabilizador (**)
N501D, G496D, Q498N	3.66	altamente desestabilizador (**)
N501D, G496D, Q498Y	4.37	altamente desestabilizador (**)
N501D, G496D, Q498T	4.26	altamente desestabilizador (**)
L455Y	1.51	neutro (*)
L455S	2.71	altamente desestabilizador (**)
L455W	-0.73	neutro (**)
L455R	2.66	altamente desestabilizador (**)
L455Y, G496D, Q498F	1.57	altamente desestabilizador (*)
L455Y, G496D, Q498H	1.62	altamente desestabilizador (*)
L455Y, G496D, Q498N	2.06	altamente desestabilizador (*)
L455Y, G496D, Q498T	1.93	altamente desestabilizador (*)
L455Y, G496D, Q498Y	2.10	altamente desestabilizador (*)
F486P	1.35	neutro (*)
F486L	0.27	neutro (**)
F486L, G496D, Q498F	3.2	altamente desestabilizador (**)
F486L, G496D, Q498H	2.66	altamente desestabilizador (**)
F486L, G496D, Q498N	2.20	altamente desestabilizador (*)
F486L, G496D, Q498Y	2.71	altamente desestabilizador (**)
F486L, G496D, Q498T	2.05	altamente desestabilizador (*)
S494G	0.1	neutro (**)
S494R	0.55	neutro (**)
S494D	0.3	neutro (**)
S494P	0.04	neutro (**)
S494T	0.18	neutro (**)
S494P, G496D, Q498F	3.80	altamente desestabilizador (**)
S494P, G496D, Q498H	3.04	altamente desestabilizador (**)
S494P, G496D, Q498N	2.61	altamente desestabilizador (**)
S494P, G496D, Q498Y	3.25	altamente desestabilizador (**)
S494P, G496D, Q498T	2.11	altamente desestabilizador (*)

Tabela 4.5 – Alterações na afinidade de ligação do *spike* RBD de SARS-CoV-2 complexado com ACE2 após mutação, conforme previsto pelo método MutaBind2 (6). Efeitos de previsão são classificados como previsão de baixa confiança: (*), ou previsão de alta confiança: (**). Aqui, adotamos a classificação apresentada no método MutaBind (6) devido à alta similaridade entre as curvas COR (características operacionais do receptor) em ambos os métodos.

Adicionamos a tabela 4.6 com o valor das variações na afinidade de ligação em mutações únicas, previsto pelos programas BeAtMuSiC e mCSM-PPI2.

Medimos o coeficiente de correlação de Pearson (ρ) entre as predições de mutações realizadas em SARS-CoV (ver apêndice C).

Todas as mutações duplas e triplas envolvendo o bolso hidrofóbico e importantes resíduos de interface de ligação desestabilizaram a complexação.

	BeAtMuSiC	mCSM-PPI2		BeAtMuSiC	mCSM-PPI2
	$\Delta\Delta G$	$\Delta\Delta G$		$\Delta\Delta G$	$\Delta\Delta G$
SARS-CoV	(kcal/mol)	(kcal/mol)	SARS-CoV-2	(kcal/mol)	(kcal/mol)
G482D	1.15	-1.428	G496D	0.65	-0.910
G482A	0.67	0.861	G496A	0.84	0.829
G482V	0.40	0.838	G496V	0.70	1.174
Y484F	0.45	0.872	Q498F	-0.42	0.466
Y484H	1.17	1.670	Q498H	0.13	0.415
Y484N	1.94	-0.333	Q498N	0.69	-1.103
Y484Q	1.65	1.599	Q498Y	-0.14	-1.961
Y484T	1.64	1.581	Q498T	0.39	1.363
N479E	0.86	0.805	Q493E	1.37	0.092
N479K	0.66	1.448	Q493K	1.06	0.808
N479Q	0.57	0.743	Q493N	1.36	-0.497
N479R	0.29	0.797	Q493R	0.76	0.552
N479S	0.91	-0.610	Q493S	1.86	-0.509
T487S	1.11	1.308	N501S	1.00	0.968
T487N	0.65	1.391	N501T	0.49	-1.036
T487D	1.05	1.176	N501D	0.68	0.825
T487A	1.32	0.985	N501A	0.80	1.075
T487V	0.36	-0.902	N501V	0.19	-0.132
T487I	0.07	-0.918	N501I	0.08	-1.197
Y442L	1.29	1.278	L455Y	-0.46	-1.288
Y442S	1.87	0.243	L455S	1.48	-0.218
Y442W	0.57	0.324	L455W	-0.27	-0.108
Y442R	1.26	1.315	L455R	0.70	1.289
L472P	-0.03	0.490	F486P	0.92	0.954
L472F	0.30	-0.212	F486L	0.72	0.935
D480G	0.32	0.550	S494G	0.05	0.323
D480R	0.19	0.194	S494R	0.07	-0.010
D480S	0.16	0.621	S494D	0.42	-0.888
D480P	-0.04	-0.138	S494P	0.47	0.127
D480T	0.08	-0.307	S494T	-0.02	-0.342

Tabela 4.6 – Alterações na afinidade de ligação do *spike* RBD de SARS-CoV e SARS-CoV-2 complexado com ACE2 após mutação. $\Delta\Delta G > 0$ mutação diminui a afinidade de ligação. $\Delta\Delta G < 0$ mutação aumenta a afinidade de ligação, conforme previsto pelo programa BeAtMuSiC (7) e mCSM-PPI2 (8).

4.1 Motivo de ligação ao receptor da proteína *spike* em morcegos

Não são todas as cepas de SARS-CoV isoladas do morcego que usam o ACE2 como receptor de ligação ao hospedeiro. Portanto, o conjunto de sequências de aminoácidos exibido na figura 4.7 pode exemplificar a relação de sucesso entre a evolução do vírus e o mecanismo de ligação. Este conjunto destaca em cor púrpura os resíduos de aminoácidos preservados na sequência YGFY, características do SARS-CoV humano. Para comparação, também exibimos cepas de CoV com mutações nesse padrão SARS-CoV para explorar a relação entre o mecanismo molecular hipotético e o reconhecimento do receptor celular. Foi demonstrado que LYRa11 (16), Rs3367 (84), Rs4874 (85), WIV1 e WIV16 (16, 86), têm a capacidade de usar ACE2 para a entrada celular, bem como RaTG13, em linha com nossa hipótese. Além disso, a mutação de ponto único próximo Y→F nas próximas seis cepas Rs7327, Rs9401, YN2018B, RsSHC014, Rs4084 e Rs4231, não interfere, como esperado, no mecanismo de ligação. Esta observação é apoiada por estudos de entrada na célula hospedeira para Rs7327 (16, 85), Rs9401, RsSHC014, Rs4084 e Rs4231 (85), por estarem em um grupo de provável uso de receptor ACE2. Além disso, essa mutação substitui um resíduo hidrofóbico por outro de maior hidrofobicidade, reforçando a topologia conformacional para ligação ao receptor. Identificamos no próximo grupo constituído por BtKY72, BB9904/BGR/2008 e BM48-31/BGR/2008, respectivamente, as mutações Y487T, Y488T e Y485T, diminuindo a hidrofobicidade inicial do bolso hidrofóbico esperado nestas cepas. Parece improvável que esta mutação e as deleções de resíduos de aminoácidos associadas à sequência Tor2 RBM afetem o mecanismo de ligação baseado em YGF para BtKY72 e BB9904/BGR/2008. Infelizmente, não há dados experimentais disponíveis sobre seus receptores. É importante observar que o resíduo F492 em BM48-31/BGR/2008 produz outro IGF de sequência hidrofóbica nos resíduos 490-492 (figura 4.7). Especulamos que esta dupla ocorrência pode interromper o mecanismo molecular acima mencionado por causa das indicações de que BM48-31/BGR/2008 não interage, pelo menos com ACE2 humano (16). Nenhuma outra sequência GF ocorre no RBD destas cepas. As cepas de CoV seguintes na figura 4.7 não contêm tais sequências específicas de resíduos de YGF no RBM nem em seu RBD. Encontramos a sequência de duas letras GF em Rf1/2004, mas localizada no RBD e com GF cercado por resíduos hidrofílicos. Embora tenhamos considerado apenas

parte das sequências que melhor se alinham com RBM de Tor2, foi demonstrado que os *spike* de HuB2013, HKU3, CoVZC45, CoVZXC21, Rf1, Rf4092 e Shaanxi2011 não usam hACE2, um resultado que não é apenas uma consequência de deleções no RBD (16). Suporte adicional foi apresentado contra HKU3 no uso de hACE2 (87). Parece improvável que Rm1/2004 infecte hACE2 por causa de sua energia livre de ligação desfavorável (88). Outro resultado conclui que Rp3 é incapaz de infectar hACE2 ou mesmo ACE2 de morcego (89). Colocamos juntos os alinhamentos envolvendo Tor2, RaTG13 e SARS-CoV-2 no final da figura 4.7 para comparação posterior. Todo o genoma de RaTG13 compartilha 96% de identidade de sequência de aminoácidos com o SARS-CoV-2 sendo considerado o genoma mais estreitamente relacionado a este CoV (90). Considerando sua proteína *spike* e RBM, RaTG13 compartilha, respectivamente, 97% e 76% de identidade de aminoácidos com o SARS-CoV-2. Para comparação, RaTG13 compartilha 79%, 77% e 53% de identidade, respectivamente, para todo o genoma, proteína *spike* e RBM com SARS-CoV Tor2 mostrado na tabela 4.7. Portanto, o SARS-CoV-2 é mais semelhante ao RaTG13 do que às cepas do SARS-CoV em todas as regiões.

4.2 Motivo de ligação ao receptor da proteína *spike* em civetas de palma e pangolins

Para explorar o papel do mecanismo de ligação baseado em YGF, exibimos sequências de resíduos comparativas para civetas e pangolins, novamente alinhadas com RBM de SARS-CoV Tor2, figura 4.8. Esta figura mostra que o padrão YGFY característico de SARS-CoV humano é mantido para os dados coletados, mas com uma mutação de ponto único Y→H para pangolins hospedeiros PCoV. Vale observar que outra sequência GF de duas letras não é encontrada nem mesmo no RBD dessas cepas, o que poderia promover outro bolso hidrofóbico interrompendo a ligação ao receptor ACE2. Incluímos o SARS-CoV-2 na última linha da figura 4.8 para uma comparação direta. PCoV GX-P2V compartilha 79%, 77% e 50% de identidade de aminoácidos com Tor2, respectivamente para o genoma inteiro, proteína *spike* e RBM alinhado com Tor2. Em relação ao SARS-CoV-2, o PCoV GX-P2V compartilha 85%, 92% e 75% da identidade de aminoácidos, respectivamente para o genoma inteiro, proteína *spike* e RBM mostrado na tabela 4.7. Acredita-se que o SARS-CoV humano passou de civetas de palma para humanos na

PORCENTAGEM DE IDENTIDADE COMPARTILHADA ENTRE DIFERENTES CEPAS DE CORONAVÍRUS		
Genoma completo		
GenBank	SARS-CoV-2 Wuhan-Hu-1 (MN908947.3)	SARS-CoV-Tor2 (NC_004718)
SARS-CoV-Tor2 (NC_004718)	82%	
Bat SARS CoV RaTG13 (MN996532.2)	96%	79%
PCoV GX-P2V (MT072864.1)	85%	79%
Bat SARS-like CoV bat-SL-CoVZC45 (MG772933)	89%	85%
Bat SARS-like CoV bat-SL-CoVZXC21 (MG772934.1)	89%	85%
<i>Spike</i>		
GenBank	SARS-CoV-2 Wuhan-Hu-1 (QHD43416.1)	SARS-CoV-Tor2 (YP_009825051.1)
SARS-CoV-Tor2 (YP_009825051.1)	76%	
Bat SARS CoV RaTG13 (QHR63300.2)	97%	77%
PCoV GX-P2V (QIQ54048.1)	92%	77%
Bat SARS-like CoV bat-SL-CoVZC45 (AVP78031.1)	81%	77%
Bat SARS-like CoV bat-SL-CoVZXC21 (AVP78042.1)	80%	77%
RBM		
GenBank	SARS-CoV-2 Wuhan-Hu-1 (QHD43416.1)	SARS-CoV-Tor2 (YP_009825051.1)
SARS-CoV-Tor2 (YP_009825051.1)	50%	
Bat SARS CoV RaTG13 (QHR63300.2)	76%	53%
PCoV GX-P2V (QIQ54048.1)	75%	50%
Bat SARS-like CoV bat-SL-CoVZC45 (AVP78031.1)	39%	32%
Bat SARS-like CoV bat-SL-CoVZXC21 (AVP78042.1)	32%	30%

Tabela 4.7 – Porcentagem de identidade compartilhada entre diferentes cepas de coronavírus utilizando as sequências do genoma/*spike*/RBM respectivamente. As cepas utilizadas foram: SARS-CoV-2 GenBank: MN908947.3 (genoma completo), GenBank: QHD43416.1 (*spike*); SARS-CoV GenBank: NC_004718 (genoma completo), GenBank: YP_009825051.1 (*spike*); RaTG13 GenBank: MN996532.2 (genoma completo), GenBank: QHR63300.2 (*spike*); PCoV GX-P2V GenBank: MT072864.1 (genoma completo), GenBank: QIQ54048.1 (*spike*); bat-SL-CoV ZC45 GenBank: MG772933 (genoma completo), GenBank: AVP78031.1 (*spike*); bat-SL-CoV ZXC21 GenBank: MG772934.1 (genoma completo), GenBank: AVP78042.1 (*spike*). A porcentagem de identidade do RBM de todas as cepas foram calculadas utilizando resultados de alinhamento genéticos obtidos nesse estudo.

epidemia de 2002-2003 porque suas sequências de genoma são altamente semelhantes (26, 77, 91). Os alinhamentos de aminoácidos mostram um RBM quase idêntico entre o SARS-CoV humano, representado pela cepa Tor2, e dados coletados de cepas de civeta de palma. Esta identificação também inclui o mecanismo molecular baseado em YGF capaz de usar proteínas ACE2. No entanto, esses alinhamentos exibem alta similaridade entre os pangolins e o SARS-CoV-2, o que também apóia as conclusões anteriores sobre os pangolins serem a provável origem do SARS-CoV-2 (90, 92). No entanto, com base em nossos dados relacionados à ligação ao receptor do hospedeiro e seus alinhamentos RBM e proteína S, não podemos descartar a cepa do tipo RaTG13 de morcego como também a possível origem do SARS-CoV-2.

4.3 SARS-CoV e hCoV-NL63: relacionados apenas funcionalmente

Embora não haja muitos dados experimentais disponíveis identificando a proteína de ligação ao receptor viral para CoVs, está bem estabelecido que o SARS-CoV humano e o hCoV-NL63 empregam ACE2 como o receptor celular para infectar as células hospedeiras (93, 94). Curiosamente, os domínios SARS-CoV e hCoV-NL63 não apresentam grande similaridade de sequência. Por exemplo, suas subunidades *spike*-S1 compartilham apenas 10% na similaridade. Outros recursos separam SARS-CoV e hCoV-NL63 (95). Os SARS-CoVs são classificados como β -coronavírus com subgênero sarbecovírus, enquanto hCoV-NL63 está no gênero α -coronavírus e subgênero setracovírus. Embora o hCoV-NL63 também entre na célula por endocitose, seu receptor funcional requer proteoglicanos de sulfato de heparan para a ligação inicial, representando um fator extra importante para o ACE2 atuar como um receptor funcional (21, 95). Além disso, a proteína *spike*-S1 de SARS-CoV liga-se mais eficientemente a ACE2 do que o *spike*-S1 correspondente de NL63 (NL63-S) (96). Isso pode estar ligado ao fato de que o SARS-CoV e o NL63-S contatam hACE2 de maneira diferente, uma conclusão baseada nos resultados experimentais de que o NL63-S não se liga ao hACE2 por meio de um único e grande domínio (94, 97). Na verdade, diferentes regiões RBD foram identificadas no NL63-S. Uma dessas regiões foi posicionada nos resíduos 476-616 e compreendendo três regiões RBM descontínuas, RBM1 (resíduos 497-501), RBM2 (resíduos 530-540) e RBM3 (resíduos 575-594) (98, 99, 100).

Um RBD ligeiramente diferente foi identificado para este CoV (97). Ele estaria localizado nos resíduos 482-602, também com três regiões RBM descontínuas, que circundam uma cavidade rasa na interface de ligação hCoV-NL63-ACE2. Curiosamente, seu alinhamento de proteína *spike* com Tor2 não mostra o padrão de resíduo esperado no RBM correspondente de Tor2 nem nas regiões RBD mencionadas acima de NL63-S. Isso pode ajudar a explicar a via incomum de ligação a ACE2 para este CoV.

A lista de códigos de acesso do GenBank para as proteínas *spike* analisadas neste trabalho está disponível na figura 4.9.

Name	Spike Glycoprotein GenBank	Name	Spike Glycoprotein GenBank	Name	Spike Glycoprotein GenBank
SARS-CoV-Tor2	AAP41037.1	SARS-CoV PC4-115	AAV49719.1	Rhinolophus affinis CoV LYRa3 (SARS-like)	AHX37569.1
SARS-CoV BJ01	AAP30030.1	SARS-CoV PC4-145	AAV49721.1	Rhinolophus affinis CoV LYRa11 (SARS-like)	AHX37558.1
SARS-CoV CUHK-W1	AAP13567.1	SARS-CoV PC4-241	AAV49723.1	Bat SARS-like CoV Rs3367	AGZ48818.1
SARS-CoV Urbani	AAP13441.1	SARS-CoV civetD14	AAU04661.1	Bat SARS-like CoV Rs4874	ATO98205.1
SARS-CoV Frankfurt-1	AAP33697.1	SARS-CoV PC4-137	AAV49720.1	Bat SARS-like CoV WIV1	AGZ48828.1
SARS-CoV CUHK-AG01	AAP94737.1	SARS-CoV civetD20	AAU04664.1	SARS-like CoV WIV16	ALK02457.1
SARS-CoV TW6	AAR87567.1	SARS-CoV PC4-127	AAU93318.1	Bat SARS CoV RaTG13	QHR63300.2
SARS-CoV TW11	AAR87512.1	SARS-CoV B039	AAV97993.1	Bat SARS-like CoV Rs7327	ATO98218.1
SARS-CoV HKU-39849	ADC35483.1	SARS-CoV PC4-205	AAU93319.1	Bat SARS-like CoV Rs9401	ATO98231.1
SARS-CoV-2 Wuhan-Hu-1	QHD43416.1	SARS-CoV PC4-199	AAV49722.1	CoV BtRs-BetaCoV/YN2018B	QDF43825.1
SARS-CoV-2 CA-CDC-0139	QJV58828.1	SARS-CoV civetD10	AAU04649.1	Bat SARS-like CoV RsSHC014	AGZ48806.1
SARS-CoV-2 France	QJT72590.1	SARS-CoV civetD19	AAU04662.1	Bat SARS-like CoV Rs4084	ATO98132.1
SARS-CoV-2 WHU/HnCoV011	QIU82034.1	SARS-CoV A022	AAV91631.1	Bat SARS-like CoV Rs4231	ATO98157.1
SARS-CoV-2 CA-CZB-1248	QKE49268.1	Civet SARS-CoV 007/2004	AAU04646.1	SARS-CoV strain BtKY72	AP040579.1
SARS-CoV-2 CA-CZB-1033	QQJ38756.1	SARS-CoV A001	AAV97984.1	SARS-like CoV BatCoV/BB9904/BGR/2008	ALJ94036.1
SARS-CoV-2 France/10070SK	QJT73010.1	PCoV GX-P2V	QIQ54048.1	Bat CoV BM48-31/BGR/2008 (SARS-like)	ADK66841.1
SARS-CoV-2 CA-CZB0103	QJE38330.1	PCoV GX-P4L	QIA48614.1	CoV BtRI-BetaCoV/SC2018	QDF43815.1
SARS-CoV-2 BetaCoV/Wuhan/IPBCAMS-WH-05	QHU36864.1	PCoV GX-P5L	QIA48632.1	BtRs-BetaCoV/HuB2013	AIA62310.1
SARS-CoV-2 CruiseA-18	QII57278.1			Bat SARS-CoV HKU3-1	AAV88866.1
SARS-CoV-2 CA-CZB-1105	QJS54754.1			Bat SARS-like CoV bat-SL-CoVZC45	AVP78031.1
				Bat SARS-like CoV bat-SL-CoVZXC21	AVP78042.1
				Bat SARS-CoV Rf1/2004	ABD75323.1
				Bat SARS-like CoV Rf4092	ATO98145.1
				Bat CoV Anlong-103	ARI44799.1
				Bat CoV Anlong-112	ARI44804.1
				Bat SARS-CoV Rp3/2004	AAZ67052.1
				CoV BtRs-BetaCoV/YN2018C	QDF43830.1
				Bat CoV Rp/Shaanxi2011	AGC74165.1
				Bat SARS-CoV Rm1	ABD75332.1

Figura 4.9 – Números de acesso do GenBank para as sequências de coronavírus analisadas neste estudo.

Capítulo 5

Conclusão

Analizamos uma série de cepas de CoV para apoiar a hipótese de que as cepas SARS-CoV e SARS-CoV-2 compartilham um mecanismo molecular evolutivo comum para a ligação inicial ao ACE2.

Além disso, investigamos o mecanismo molecular baseado em YGF que forma uma superfície hidrofóbica favorecendo a ligação ao hospedeiro. Esta sequência baseada em YGF pode atuar como uma assinatura proteica para distinguir CoVs capazes de usar ACE2 como receptor celular sempre que esta sequência de resíduos estiver localizada na região RBM do CoV. Por exemplo, bat-SL-CoV ZC45 e ZXC21, com identidade geral do genoma de 89% compartilhado com SARS-CoV-2, podem ser colocados sob suspeita em suas afinidades de ligação à ACE2, devido à falta de tal assinatura proteica YGF. Obviamente, conforme exemplificado pelo hCoV-NL63, não podemos descartar que outro mecanismo molecular possa atuar auxiliando nessa ligação. Deve ser acentuado que a ocorrência de mais de uma sequência do tipo XGF, principalmente com X sendo um resíduo hidrofóbico, no RBM, ou mesmo na região RBD, pode ocasionar a interrupção do mecanismo molecular proposto para a ligação ACE2. Esta ocorrência pode introduzir bolsos hidrofóbicos promovendo um novo reconhecimento ligante-substrato.

Referências¹

- 1 MALIK, Y. A. Properties of coronavirus and sars-cov-2. *The Malaysian Journal of Pathology*, College of Pathologists, Academy of Medicine of Malaysia, v. 42, n. 1, p. 3–11, 2020. PMID: 32342926.
- 2 MENÉNDEZ-ARIAS, L.; GAGO, F. *Structure and Physics of Viruses: An Integrated Textbook*, edited by GM Mateu. [S.l.]: Dordrecht: Springer, 2013. 26–8 p. DOI 10.1007/978-94-007-6552-8.
- 3 LI, F. Structure, function, and evolution of coronavirus spike proteins. *Annual review of virology*, Annual Reviews, v. 3, p. 237–61, 2016. DOI 10.1146/annurev-virology-110615-042301.
- 4 KOSTOVA, Z.; WOLF, D. H. For whom the bell tolls: protein quality control of the endoplasmic reticulum and the ubiquitin–proteasome connection. *The EMBO journal*, John Wiley & Sons, Ltd, v. 22, n. 10, p. 2309–2317, 2003.
- 5 CIECHANOVER, A. The ubiquitin proteolytic system: from a vague idea, through basic mechanisms, and onto human diseases and drug targeting. *Neurology*, AAN Enterprises, v. 66, n. 1 suppl 1, p. S7–19, 2006.
- 6 ZHANG, N. et al. Mutabind2: predicting the impacts of single and multiple mutations on protein–protein interactions. *Iscience*, Elsevier, p. 100939, 2020. DOI 10.1038/s41423-020-0458-z.
- 7 DEHOUCK, Y. et al. Beatmusic: prediction of changes in protein–protein binding affinity on mutations. *Nucleic acids research*, Oxford University Press, v. 41, n. W1, p. W333–W339, 2013.
- 8 RODRIGUES, C. H. et al. mcsm-ppi2: predicting the effects of mutations on protein–protein interactions. *Nucleic acids research*, Oxford University Press, v. 47, n. W1, p. W338–W344, 2019.
- 9 CHAN, J. F.-W. et al. Genomic characterization of the 2019 novel human-pathogenic coronavirus isolated from a patient with atypical pneumonia after visiting wuhan. *Emerging microbes & infections*, Taylor & Francis, v. 9, n. 1, p. 221–36, 2020. DOI 10.1080/22221751.2020.1719902.
- 10 CLAVERIE, J.-M. Viruses take center stage in cellular evolution. *Genome biology*, BioMed Central, v. 7, n. 6, p. 110, 2006. DOI 10.1128/aem.70.10.5842-5846.2004.
- 11 GORBALENYA, A. et al. The species severe acute respiratory syndrome-related coronavirus: classifying 2019-ncov and naming it sars-cov-2. *Nature Microbiology*, Nature Publishing Group, v. 5, n. 4, p. 536–44, 2020. DOI 10.1038/s41564-020-0695-z.

¹ De acordo com a Associação Brasileira de Normas Técnicas. NBR 6023.

- 12 JAIMES, J. A. et al. Phylogenetic analysis and structural modeling of sars-cov-2 spike protein reveals an evolutionary distinct and proteolytically-sensitive activation loop. *Journal of molecular biology*, Elsevier, v. 432, p. 3309–25, 2020. DOI 10.1016/j.jmb.2020.04.009.
- 13 WEISS, S. R.; NAVAS-MARTIN, S. Coronavirus pathogenesis and the emerging pathogen severe acute respiratory syndrome coronavirus. *Microbiology and molecular biology reviews*, Am Soc Microbiol, v. 69, n. 4, p. 635–64, 2005. DOI 10.1128/MMBR.69.4.635-664.2005.
- 14 SIMMONS, G. et al. Proteolytic activation of the sars-coronavirus spike protein: cutting enzymes at the cutting edge of antiviral research. *Antiviral research*, Elsevier, v. 100, n. 3, p. 605–614, 2013.
- 15 HOFFMANN, M. et al. Sars-cov-2 cell entry depends on ace2 and tmprss2 and is blocked by a clinically proven protease inhibitor. *Cell*, Elsevier, v. 181, p. 271–80, 2020. DOI 10.1016/j.cell.2020.02.052.
- 16 LETKO, M.; MARZI, A.; MUNSTER, V. Functional assessment of cell entry and receptor usage for sars-cov-2 and other lineage b betacoronaviruses. *Nature microbiology*, Nature Publishing Group, v. 5, n. 4, p. 562–69, 2020. DOI 10.1038/s41564-020-0688-y.
- 17 WALLS, A. C. et al. Structure, function, and antigenicity of the sars-cov-2 spike glycoprotein. *Cell*, Elsevier, v. 180, p. 281–92, 2020. DOI 10.1016/j.cell.2020.02.058.
- 18 WANG, H. et al. Sars coronavirus entry into host cells through a novel clathrin-and caveolae-independent endocytic pathway. *Cell research*, Nature Publishing Group, v. 18, n. 2, p. 290–301, 2008. DOI 10.1038/cr.2008.15.
- 19 WANG, S. et al. Endocytosis of the receptor-binding domain of sars-cov spike protein together with virus receptor ace2. *Virus research*, Elsevier, v. 136, n. 1-2, p. 8–15, 2008. DOI 10.1016/j.virusres.2008.03.004.
- 20 YUAN, Y. et al. Cryo-em structures of mers-cov and sars-cov spike glycoproteins reveal the dynamic receptor binding domains. *Nature communications*, Nature Publishing Group, v. 8, n. 1, p. 1–9, 2017. DOI 10.1038/ncomms15092.
- 21 MILEWSKA, A. et al. Entry of human coronavirus nl63 into the cell. *Journal of virology*, Am Soc Microbiol, v. 92, n. 3, p. e01933–17, 2018. DOI 10.1128/JVI.01933-17.
- 22 OU, X. et al. Characterization of spike glycoprotein of sars-cov-2 on virus entry and its immune cross-reactivity with sars-cov. *Nature communications*, Nature Publishing Group, v. 11, n. 1, p. 1–12, 2020. DOI 10.1038/s41467-020-15562-9.
- 23 MATSUYAMA, S. et al. Protease-mediated enhancement of severe acute respiratory syndrome coronavirus infection. *Proceedings of the National Academy of Sciences*, National Acad Sciences, v. 102, n. 35, p. 12543–12547, 2005.
- 24 YANG, Y. et al. Receptor usage and cell entry of bat coronavirus hku4 provide insight into bat-to-human transmission of mers coronavirus. *Proceedings of the National Academy of Sciences*, National Acad Sciences, v. 111, n. 34, p. 12516–21, 2014. DOI 10.1073/pnas.1405889111.

- 25 LI, F. Receptor recognition mechanisms of coronaviruses: a decade of structural studies. *Journal of virology*, Am Soc Microbiol, v. 89, n. 4, p. 1954–64, 2015. DOI 10.1128/JVI.02615-14.
- 26 CUI, J.; LI, F.; SHI, Z.-L. Origin and evolution of pathogenic coronaviruses. *Nature Reviews Microbiology*, Nature Publishing Group, v. 17, n. 3, p. 181–92, 2019. DOI 10.1038/s41579-018-0118-9.
- 27 LI, W. et al. Receptor and viral determinants of sars-coronavirus adaptation to human ace2. *The EMBO journal*, John Wiley & Sons, Ltd Chichester, UK, v. 24, n. 8, p. 1634–43, 2005. DOI 10.1038/sj.emboj.7600640.
- 28 SANJUÁN, R.; DOMINGO-CALAP, P. Mechanisms of viral mutation. *Cellular and molecular life sciences*, Springer, v. 73, n. 23, p. 4433–48, 2016. DOI 10.1007/s00018-016-2299-6.
- 29 AHOLA, T.; KARLIN, D. G. Sequence analysis reveals a conserved extension in the capping enzyme of the alphavirus supergroup, and a homologous domain in nodaviruses. *Biology direct*, Springer, v. 10, n. 1, p. 16, 2015. DOI 10.1186/s13062-015-0050-0.
- 30 LAU, S. K. et al. Complete genome sequence of bat coronavirus hku2 from chinese horseshoe bats revealed a much smaller spike gene with a different evolutionary lineage from the rest of the genome. *Virology*, Elsevier, v. 367, n. 2, p. 428–39, 2007. DOI 10.1016/j.virol.2007.06.009.
- 31 WOO, P. C. et al. Discovery of seven novel mammalian and avian coronaviruses in the genus deltacoronavirus supports bat coronaviruses as the gene source of alphacoronavirus and betacoronavirus and avian coronaviruses as the gene source of gammacoronavirus and deltacoronavirus. *Journal of virology*, Am Soc Microbiol, v. 86, n. 7, p. 3995–4008, 2012. DOI 10.1128/JVI.06540-11.
- 32 BENIAC, D. R. et al. Architecture of the sars coronavirus prefusion spike. *Nature structural & molecular biology*, Nature Publishing Group, v. 13, n. 8, p. 751–52, 2006. DOI 10.1038/nsmb1123.
- 33 WAN, Y. et al. Receptor recognition by the novel coronavirus from wuhan: an analysis based on decade-long structural studies of sars coronavirus. *Journal of virology*, Am Soc Microbiol, v. 94, n. 7, 2020. DOI 10.1128/JVI.00127-20.
- 34 LI, F. et al. Structure of sars coronavirus spike receptor-binding domain complexed with receptor. *Science*, American Association for the Advancement of Science, v. 309, n. 5742, p. 1864–1868, 2005. DOI 10.1126/science.1113611.
- 35 LI, F. Evidence for a common evolutionary origin of coronavirus spike protein receptor-binding subunits. *Journal of virology*, Am Soc Microbiol, v. 86, n. 5, p. 2856–2858, 2012. DOI 10.1128/JVI.06882-11.
- 36 SONG, W. et al. Cryo-em structure of the sars coronavirus spike glycoprotein in complex with its host cell receptor ace2. *PLoS pathogens*, Public Library of Science, v. 14, n. 8, p. e1007236, 2018. DOI 10.1371/journal.ppat.1007236.
- 37 CARVALHO, P. P.; ALVES, N. A. Featuring ace2 binding sars-cov and sars-cov-2 through a conserved evolutionary pattern of amino acid residues. *arXiv preprint arXiv:2008.12830*, 2020.

- 38 WU, K. et al. Mechanisms of host receptor adaptation by severe acute respiratory syndrome coronavirus. *Journal of Biological Chemistry*, ASBMB, v. 287, n. 12, p. 8904–11, 2012. DOI 10.1074/jbc.M111.325803.
- 39 LAN, J. et al. Structure of the sars-cov-2 spike receptor-binding domain bound to the ace2 receptor. *Nature*, Nature Publishing Group, v. 581, n. 7807, p. 215–20, 2020. DOI 10.1038/s41586-020-2180-5.
- 40 HE, J. et al. Molecular mechanism of evolution and human infection with sars-cov-2. *Viruses*, Multidisciplinary Digital Publishing Institute, v. 12, n. 4, p. 428, 2020. DOI 10.3390/v12040428.
- 41 PICKART, C. M.; COHEN, R. E. Proteasomes and their kin: proteases in the machine age. *Nature reviews Molecular cell biology*, Nature Publishing Group, v. 5, n. 3, p. 177–87, 2004. DOI 10.1038/nrm1336.
- 42 FINLEY, D. Recognition and processing of ubiquitin-protein conjugates by the proteasome. *Annual review of biochemistry*, Annual Reviews, v. 78, p. 477–513, 2009. DOI 10.1146/annurev.biochem.78.081507.101607.
- 43 SU, V.; LAU, A. F. Ubiquitin-like and ubiquitin-associated domain proteins: significance in proteasomal degradation. *Cellular and molecular life sciences*, Springer, v. 66, n. 17, p. 2819–33, 2009. DOI 10.1007/s00018-009-0048-9.
- 44 HOFMANN, K. The uba domain: a sequence motif present in multiple enzyme classes of the ubiquitination pathway. *Trends Biol Sci*, Elsevier, v. 21, p. 172–173, 1996. DOI 10.1016/s0968-0004(96)30015-7.
- 45 DIECKMANN, T. et al. Structure of a human dna repair protein uba domain that interacts with hiv-1 vpr. *Nature structural biology*, Nature Publishing Group, v. 5, n. 12, p. 1042–47, 1998. DOI 10.1038/4220.
- 46 GEETHA, T.; WOOTEN, M. W. Structure and functional properties of the ubiquitin binding protein p62. *FEBS letters*, Elsevier, v. 512, n. 1-3, p. 19–24, 2002. DOI 10.1016/S0014-5793(02)02286-X.
- 47 WILKINSON, C. R. et al. Proteins containing the uba domain are able to bind to multi-ubiquitin chains. *Nature cell biology*, Nature Publishing Group, v. 3, n. 10, p. 939–43, 2001. DOI 10.1038/ncb1001-939.
- 48 BERTOLAET, B. L. et al. Uba domains mediate protein-protein interactions between two dna damage-inducible proteins. *Journal of molecular biology*, Elsevier, v. 313, n. 5, p. 955–63, 2001. DOI 10.1006/jmbi.2001.5105.
- 49 MADURA, K. The ubiquitin-associated (uba) domain: on the path from prudence to prurience. *Cell Cycle*, Taylor & Francis, v. 1, n. 4, p. 233–42, 2002. DOI 10.4161/cc.1.4.130.
- 50 RAASI, S. et al. Diverse polyubiquitin interaction properties of ubiquitin-associated domains. *Nature structural & molecular biology*, Nature Publishing Group, v. 12, n. 8, p. 708–14, 2005. DOI 10.1038/nsmb962.
- 51 LONG, J. et al. Ubiquitin recognition by the ubiquitin-associated domain of p62 involves a novel conformational switch. *Journal of Biological Chemistry*, ASBMB, v. 283, n. 9, p. 5427–40, 2008. DOI 10.1074/jbc.M704973200.

- 52 TSE, M. K. et al. Structural analysis of the uba domain of x-linked inhibitor of apoptosis protein reveals different surfaces for ubiquitin-binding and self-association. *PloS one*, Public Library of Science, v. 6, n. 12, p. e28511, 2011. DOI 10.1371/journal.pone.0028511.
- 53 CABE, M. et al. Pbl and uba domains of p62 are essential for aggresome-like induced structure formation. *Biochemical and biophysical research communications*, Elsevier, v. 503, n. 4, p. 2306–11, 2018. DOI 10.1016/j.bbrc.2018.06.153.
- 54 WATKINS, J. F. et al. The *saccharomyces cerevisiae* dna repair gene rad23 encodes a nuclear protein containing a ubiquitin-like domain required for biological function. *Molecular and cellular biology*, Am Soc Microbiol, v. 13, n. 12, p. 7757–65, 1993. DOI 10.1128/mcb.13.12.7757.
- 55 WANG, Q. et al. Ubiquitin recognition by the dna repair protein hhr23a. *Biochemistry*, ACS Publications, v. 42, n. 46, p. 13529–35, 2003. DOI 10.1021/bi035391j.
- 56 SUGASAWA, K. et al. Hhr23b, a human rad23 homolog, stimulates xpc protein in nucleotide excision repair in vitro. *Molecular and cellular biology*, Am Soc Microbiol, v. 16, n. 9, p. 4852–61, 1996. DOI 10.1128/mcb.16.9.4852.
- 57 OHNO, A. et al. Structure of the uba domain of dsk2p in complex with ubiquitin: molecular determinants for ubiquitin recognition. *Structure*, Elsevier, v. 13, n. 4, p. 521–532, 2005.
- 58 FUNAKOSHI, M. et al. Budding yeast dsk2p is a polyubiquitin-binding protein that can interact with the proteasome. *Proceedings of the National Academy of Sciences*, National Acad Sciences, v. 99, n. 2, p. 745–750, 2002.
- 59 MUELLER, T. D.; FEIGON, J. Solution structures of uba domains reveal a conserved hydrophobic surface for protein–protein interactions. *Journal of molecular biology*, Elsevier, v. 319, n. 5, p. 1243–1255, 2002.
- 60 BUDHIDARMO, R.; DAY, C. L. The ubiquitin-associated domain of cellular inhibitor of apoptosis proteins facilitates ubiquitylation. *Journal of Biological Chemistry*, ASBMB, v. 289, n. 37, p. 25721–36, 2014. DOI 10.1074/jbc.M113.545475.
- 61 ROODMAN, G. D.; WINDLE, J. J. et al. Paget disease of bone. *The Journal of clinical investigation*, Am Soc Clin Investig, v. 115, n. 2, p. 200–208, 2005.
- 62 WANG, Q. et al. Structural and functional basis of sars-cov-2 entry by using human ace2. *Cell*, Elsevier, v. 181, n. 4, p. 894–904, 2020.
- 63 BENSON, D. A. et al. Genbank. *Nucleic acids research*, Oxford University Press, v. 33, n. suppl_1, p. D34–D38, 2005.
- 64 PRUITT, K. D.; TATUSOVA, T.; MAGLOTT, D. R. Ncbi reference sequence (refseq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic acids research*, Oxford University Press, v. 33, n. suppl_1, p. D501–D504, 2005.
- 65 SHERRY, S. T. et al. dbsnp: the ncbi database of genetic variation. *Nucleic acids research*, Oxford University Press, v. 29, n. 1, p. 308–311, 2001.

- 66 ALTSCHUL, S. F. et al. Basic local alignment search tool. *Journal of molecular biology*, Elsevier, v. 215, n. 3, p. 403–410, 1990.
- 67 THOMPSON, J. D.; HIGGINS, D. G.; GIBSON, T. J. Clustal w: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic acids research*, Oxford University Press, v. 22, n. 22, p. 4673–4680, 1994.
- 68 GIBAS, C.; JAMBECK, P. *Desenvolvendo Bioinformática: ferramentas de software para aplicações em biologia*. [S.l.]: Campus, 2001.
- 69 CLAMP, M. et al. The jalview java alignment editor. *Bioinformatics*, Oxford University Press, v. 20, n. 3, p. 426–427, 2004.
- 70 DELANO, W. L. et al. Pymol: An open-source molecular graphics tool. *CCP4 Newsletter on protein crystallography*, Citeseer, v. 40, n. 1, p. 82–92, 2002.
- 71 PETTERSEN, E. F. et al. Ucsf chimera—a visualization system for exploratory research and analysis. *Journal of computational chemistry*, Wiley Online Library, v. 25, n. 13, p. 1605–1612, 2004.
- 72 BERMAN, H. et al. The worldwide protein data bank (wwpdb): ensuring a single, uniform archive of pdb data. *Nucleic acids research*, Oxford University Press, v. 35, n. suppl.1, p. D301–D303, 2007.
- 73 SOBOLEV, V. et al. Automated analysis of interatomic contacts in proteins. *Bioinformatics (Oxford, England)*, v. 15, n. 4, p. 327–332, 1999.
- 74 JANKAUSKAITE, J. et al. Skempi 2.0: an updated benchmark of changes in protein–protein binding energy, kinetics and thermodynamics upon mutation. *Bioinformatics*, Oxford University Press, v. 35, n. 3, p. 462–469, 2019.
- 75 LI, M. et al. Mutabind estimates and interprets the effects of sequence variants on protein–protein interactions. *Nucleic acids research*, Oxford University Press, v. 44, n. W1, p. W494–W501, 2016.
- 76 PRAJAPAT, M. et al. Update on the target structures of sars-cov-2: a systematic review. *Indian journal of pharmacology*, Wolters Kluwer–Medknow Publications, v. 52, n. 2, p. 142, 2020.
- 77 QU, X.-X. et al. Identification of two critical amino acid residues of the severe acute respiratory syndrome coronavirus spike protein for its variation in zoonotic tropism transition via a double substitution strategy. *Journal of Biological Chemistry*, ASBMB, v. 280, n. 33, p. 29588–95, 2005. DOI 10.1074/jbc.M500662200.
- 78 WIMLEY, W. C.; WHITE, S. H. Experimentally determined hydrophobicity scale for proteins at membrane interfaces. *Nature structural biology*, Nature Publishing Group, v. 3, n. 10, p. 842–8, 1996. DOI 10.1038/nsb1096-842.
- 79 YI, C. et al. Key residues of the receptor binding motif in the spike protein of sars-cov-2 that interact with ace2 and neutralizing antibodies. *Cellular & Molecular Immunology*, Nature Publishing Group, p. 1–10, 2020. DOI 10.1038/s41423-020-0458-z.

- 80 SHANG, J. et al. Structural basis of receptor recognition by sars-cov-2. *Nature*, Nature Publishing Group, v. 581, n. 7807, p. 221–224, 2020.
- 81 GALLOWAY, S. E. et al. Emergence of sars-cov-2 b. 1.1. 7 lineage—united states, december 29, 2020–january 12, 2021. *Morbidity and Mortality Weekly Report*, Centers for Disease Control and Prevention, v. 70, n. 3, p. 95, 2021.
- 82 PEREIRA, F. Sars-cov-2 variants combining spike mutations and the absence of orf8 may be more transmissible and require close monitoring. *Biochemical and biophysical research communications*, Elsevier, 2021.
- 83 GU, H. et al. Adaptation of sars-cov-2 in balb/c mice for testing vaccine efficacy. *Science*, American Association for the Advancement of Science, v. 369, n. 6511, p. 1603–1607, 2020.
- 84 GE, X.-Y. et al. Isolation and characterization of a bat sars-like coronavirus that uses the ace2 receptor. *Nature*, Nature Publishing Group, v. 503, n. 7477, p. 535–8, 2013. DOI 10.1038/nature12711.
- 85 HU, B. et al. Discovery of a rich gene pool of bat sars-related coronaviruses provides new insights into the origin of sars coronavirus. *PLoS pathogens*, Public Library of Science San Francisco, CA USA, v. 13, n. 11, p. e1006698, 2017. DOI 10.1128/JVI.02582-15.
- 86 YANG, X.-L. et al. Isolation and characterization of a novel bat coronavirus closely related to the direct progenitor of severe acute respiratory syndrome coronavirus. *Journal of virology*, Am Soc Microbiol, v. 90, n. 6, p. 3253–6, 2016. DOI 10.1128/JVI.02582-15.
- 87 GRALINSKI, L. E.; MENACHERY, V. D. Return of the coronavirus: 2019-ncov. *Viruses*, Multidisciplinary Digital Publishing Institute, v. 12, n. 2, p. 135, 2020. DOI 10.3390/v12020135.
- 88 ARMIJOS-JARAMILLO, V. et al. Sars-cov-2, an evolutionary perspective of interaction with human ace2 reveals undiscovered amino acids necessary for complex stability. *Evolutionary Applications*, Wiley Online Library, 2020. DOI 10.1101/2020.03.21.001933.
- 89 HOFFMANN, M. et al. Differential sensitivity of bat cells to infection by enveloped rna viruses: coronaviruses, paramyxoviruses, filoviruses, and influenza viruses. *PloS one*, Public Library of Science, v. 8, n. 8, p. e72942, 2013. DOI 10.1371/journal.pone.0072942.
- 90 ZHANG, T.; WU, Q.; ZHANG, Z. Probable pangolin origin of sars-cov-2 associated with the covid-19 outbreak. *Current Biology*, Elsevier, v. 30, p. 1346–51, 2020. DOI 10.1016/j.cub.2020.03.022.
- 91 LI, F. Structural analysis of major species barriers between humans and palm civets for severe acute respiratory syndrome coronavirus infections. *Journal of virology*, Am Soc Microbiol, v. 82, n. 14, p. 6984–91, 2008. DOI 10.1128/JVI.00442-08.
- 92 LAM, T. T.-Y. et al. Identifying sars-cov-2-related coronaviruses in malayan pangolins. *Nature*, Nature Publishing Group, v. 583, p. 282–5, 2020. DOI 10.1038/s41586-020-2169-0.

- 93 HOFMANN, H. et al. Human coronavirus nl63 employs the severe acute respiratory syndrome coronavirus receptor for cellular entry. *Proceedings of the National Academy of Sciences*, National Acad Sciences, v. 102, n. 22, p. 7988–93, 2005. DOI 10.1073/pnas.0409465102.
- 94 HOFMANN, H. et al. Highly conserved regions within the spike proteins of human coronaviruses 229e and nl63 determine recognition of their respective cellular receptors. *Journal of virology*, Am Soc Microbiol, v. 80, n. 17, p. 8639–52, 2006. DOI 10.1128/JVI.00560-06.
- 95 MILEWSKA, A. et al. Human coronavirus nl63 utilizes heparan sulfate proteoglycans for attachment to target cells. *Journal of virology*, Am Soc Microbiol, v. 88, n. 22, p. 13221–30, 2014. DOI 10.1128/JVI.02078-14.
- 96 GLOWACKA, I. et al. Differential downregulation of ace2 by the spike proteins of severe acute respiratory syndrome coronavirus and human coronavirus nl63. *Journal of virology*, Am Soc Microbiol, v. 84, n. 2, p. 1198–205, 2010. DOI 10.1128/JVI.01248-09.
- 97 WU, K. et al. Crystal structure of nl63 respiratory coronavirus receptor-binding domain complexed with its human receptor. *Proceedings of the National Academy of Sciences*, National Acad Sciences, v. 106, n. 47, p. 19970–4, 2009. DOI 10.1073/pnas.0908837106.
- 98 LI, W. et al. The s proteins of human coronavirus nl63 and severe acute respiratory syndrome coronavirus bind overlapping regions of ace2. *Virology*, Elsevier, v. 367, n. 2, p. 367–74, 2007. DOI 10.1016/j.virol.2007.04.035.
- 99 LIN, H.-X. et al. Identification of residues in the receptor-binding domain (rbd) of the spike protein of human coronavirus nl63 that are critical for the rbd–ace2 receptor interaction. *Journal of General Virology*, Microbiology Society, v. 89, n. 4, p. 1015–24, 2008. DOI 10.1099/vir.0.83331-0.
- 100 LIN, H.-X. et al. Characterization of the spike protein of human coronavirus nl63 in receptor binding and pseudotype virus entry. *Virus research*, Elsevier, v. 160, n. 1-2, p. 283–93, 2011. DOI 10.1016/j.virusres.2011.06.029.
- 101 CACCAMO, A. et al. *p62 improves AD-like pathology by increasing autophagy (Retraction of Vol 22, Pg 856, 2017)*. [S.l.]: NATURE PUBLISHING GROUP MACMILLAN BUILDING, 4 CRINAN ST, LONDON N1 9XW, ENGLAND, 2020.
- 102 PANKIV, S. et al. p62/sqstm1 binds directly to atg8/lc3 to facilitate degradation of ubiquitinated protein aggregates by autophagy. *Journal of biological chemistry*, Elsevier, v. 282, n. 33, p. 24131–24145, 2007.
- 103 BJØRKØY, G. et al. p62/sqstm1 forms protein aggregates degraded by autophagy and has a protective effect on huntingtin-induced cell death. *The Journal of cell biology*, Rockefeller University Press, v. 171, n. 4, p. 603–614, 2005.
- 104 KOMATSU, M. et al. Homeostatic levels of p62 control cytoplasmic inclusion body formation in autophagy-deficient mice. *Cell*, Elsevier, v. 131, n. 6, p. 1149–1163, 2007.
- 105 BITTO, A. et al. P62/sqstm1 at the interface of aging, autophagy, and disease. *Age*, Springer, v. 36, n. 3, p. 1123–1137, 2014.

- 106 CHANG, Y.-G. et al. Solution structure of the ubiquitin-associated domain of human bmsc-ubp and its complex with ubiquitin. *Protein science*, Wiley Online Library, v. 15, n. 6, p. 1248–1259, 2006.
- 107 ALGORITMO Needleman-Wunsch. (<http://https://pt.slideshare.net/mcastrosouza/algoritmo-needlemanwunsch>).
- 108 LÖYTYNOJA, A.; GOLDMAN, N. An algorithm for progressive multiple alignment of sequences with insertions. *Proceedings of the National Academy of Sciences*, National Acad Sciences, v. 102, n. 30, p. 10557–10562, 2005.
- 109 SOUZA, M. C. de. Algoritmo needleman-wunsch. 2015.
- 110 TURNER, P. Xmgrace, version 5.1. 25. *Center for Coastal and Land-Margin Research, Oregon Graduate Institute of Science and Technology*, 2005.

Apêndice A

Autofagia e o sistema ubiquitina-proteassomo

Em células eucarióticas, a autofagia utiliza-se do sistema ubiquitina-proteassomo para degradação de proteínas mal formadas, agregados proteicos e organelas danificadas. São exemplos de agregados proteicos os emaranhados neurofibrilares e depósito β -amilóide associado à doença de Alzheimer.

A autofagia pode ser mediada pelas chaperonas que direcionam proteínas para degradação no lisossomo (101) e pela proteína p62, conhecida como sequestossomo-1 (SQSTM1) (102, 103, 104), o primeiro receptor seletivo de autofagia identificado. A ação do lisossomo é limitada sendo necessária outra via de degradação, os proteassomos (105). Sua estrutura é organizada em um núcleo funcional 20S semelhante a um cilindro central, o qual possui atividades proteolíticas e um complexo regulatório 19S. A estrutura 20S adota uma conformação em barril devido a sobreposição de suas quatro fileiras proteicas de formato anelar, sendo duas fileiras externas idênticas α e duas fileiras idênticas β . Cada uma das quatro fileiras da estrutura 20S são constituídas de sete subunidades proteolítica distintas. A estrutura 19S funciona como uma tampa e uma base para 20S, sendo um importante componente no processo de degradação. A tampa 19S reconhece e desenovela a proteína poliubiquitinada, abre o orifício do anel α 20S permitindo o encaixe e a entrada da proteína no barril proteolítico 20S onde ocorre a degradação e liberação de fragmentos peptídicos, conforme esquematizado na figura A.1. A associação dessas duas regiões forma o complexo 26S (5). A proteína deve ser primeiramente marcada por via ubiquitinação, para a degradação pelo proteassomo 26S.

A ubiquitina é uma proteína evolutivamente conservada de 76 resíduos de aminoácidos responsável por marcar proteínas e direcioná-las para degradação ou regulação de suas funções (106). Para que a conjugação da ubiquitina com a proteína marcada seja efetiva-

Figura A.1 – Organização estrutural do proteassomo. Sua estrutura é organizada em um núcleo funcional 20S semelhante a um cilindro central, que possui atividades proteolíticas e um complexo regulatório 19S. A estrutura 20S adota uma conformação em barril devido a sobreposição de suas quatro fileiras proteicas de formato anelar, sendo duas fileiras externas idênticas α e duas fileiras idênticas β . Cada uma das quatro fileiras da estrutura 20S são constituídas de sete subunidades proteolítica distintas. A estrutura 19S funciona como uma tampa e uma base para 20S, sendo um importante componente no processo de degradação. A tampa 19S reconhece e desenovela a proteína poliubiquitinada, abre o orifício do anel α 20S permitindo o encaixe e a entrada da proteína no barril proteolítico 20S onde ocorre a degradação e liberação de fragmentos peptídicos. A associação dessas duas regiões forma o complexo 26S. A proteína deve ser primeiramente marcada por via ubiquitinação, para a degradação pelo proteassomo 26S. Adaptado da referência (4).

mente realizada, três etapas devem ser precedidas. Na primeira etapa, a enzima (E1) ativa a ubiquitina numa reação de consumo de ATP. Na segunda etapa, a proteína transportadora de ubiquitina ou enzima de conjugação de ubiquitina (E2) envia a ubiquitina ativada de E1 para o substrato de proteína que está ligado a proteína-ubiquitina ligase (E3). Essa transferência é direta, porém em alguns casos pode acontecer a transferência da ubiquitina de E2 para um resíduo Cys de sítio ativo no E3 antes da sua ligação ao substrato ligado

no E3. A terceira e última etapa no processo de conjugação tem por finalidade, fixar a ubiquitina ao substrato, através de uma ligação isopeptídica covalente entre a ubiquitina com o grupo amino da cadeia lateral da lisina presente no substrato. Após sucessivas conjugações de ubiquitina ativada, forma-se a poliubiquitina, que serve como um marcador de proteína. Somente após a marcação, torna-se possível a degradação da proteína pelo proteassomo 26S. O proteassomo 26S degrada proteínas maiores em pequenos peptídeos e a ubiquitina livre é liberada por enzimas desubiquitinantes, mostrado na figura A.2 (5). A conjugação da ubiquitina (UB) com proteínas enzimáticas eucarióticas auxilia nos processos de degradação pelo proteassomo, autofagia (41, 42) e regulação funcional.

Figura A.2 – Sistema ubiquitina-proteassomo. A ubiquitina é ativada pela enzima E1 (1) seguida por sua transferência para uma proteína transportadora de ubiquitina, E2 (2). E2 transfere as porções de ubiquitina ativada para o substrato de proteína que está ligado especificamente a uma ubiquitina ligase E3. A transferência é direta (3) ou através de um intermediário na ligase (4, 4a). A conjugação sucessiva de ubiquitina gera uma cadeia de poliubiquitina que serve como sinal de ligação (5) e degradação para o proteassomo 26S. O substrato é degradado em peptídeos curtos (6), e a ubiquitina livre e reutilizável é liberada por enzimas desubiquitinantes (7). Adaptado da referência (5).

Apêndice B

Algoritmo de alinhamento

O alinhamento pode ser realizado entre duas sequências, conhecido como alinhamento por par, ou alinhamento múltiplo, utilizando três ou mais sequências. Quando o alinhamento é global, o algoritmo calcula todo comprimento das sequências. Quando o alinhamento é local o algoritmo calcula apenas trechos das sequências em busca de alinhamento global ótimo.

O algoritmo de alinhamento usa o paradigma de soluções dinâmicas, que tem por finalidade calcular soluções de todos possíveis subproblemas de um problema, iniciando a partir dos cálculos simples para os mais complexos. Todos os resultados são armazenados numa matriz, a solução dos problemas mais simples não serão revisados. Somente solucionando os subproblemas é que se pode obter a solução dos problemas mais complexos. Esse procedimento é conhecido como subestrutura ótima, que obtém a melhor pontuação geral de similaridade de sequência, através dos melhores resultados obtidos do cálculo de cada subestrutura anterior. Para uma pontuação precisa (*score*) na matriz, no objetivo de encontrar os alinhamentos mais similares, são adicionados pesos estatísticos para importantes parâmetros: *match* (recompensa por apresentar caracteres iguais), *mismatch* (penalidades por mutações) e *gap penalty* (penalidades por deleções ou inserções).

A fim de elucidar o algoritmo de alinhamento, tomemos como exemplo (107) a sequência 1 nomeada 1s de tamanho t_s1, a sequência 2 nomeada 2s de tamanho t_s2 e uma matriz nomeada M. A matriz terá t_s2 + 1 linhas por t_s1 + 1 colunas. O início é zero. M[0][0] corresponde ao elemento da primeira linha e primeira coluna, M[0][1] corresponde ao elemento da primeira linha e segunda coluna e assim sucessivamente. O M[0][0] tem valor zero. De forma sucessiva os elementos da primeira linha seguirão a equação

$$M[0][i] = M[0][i - 1] + \text{gap penalty}(i > 0)$$

De forma sucessiva os elementos da primeira coluna seguirão a equação

$$M[j][0] = M[j-1][0] + gap\ penalty(j > 0)$$

Iniciando a primeira linha com $gap\ penalty = -1$, sucessivamente aplica-se a equação

$$M[0][1] = M[0][0] + gap\ penalty = 0 + (-1) = -1$$

$$M[0][2] = M[0][1] + gap\ penalty = -1 + (-1) = -2$$

Iniciando a primeira coluna com $gap\ penalty = -1$, sucessivamente aplica-se a equação

$$M[1][0] = M[0][0] + gap\ penalty = 0 + (-1) = -1$$

$$M[2][0] = M[1][0] + gap\ penalty = -1 + (-1) = -2$$

O alinhamento será entre as sequências -KFGL e -KGL e os valores dos parâmetros são ajustados para $match = 1$, $mismatch = -1$ e $gap\ penalty = -1$

$$M = \begin{pmatrix} & - & \mathbf{K} & \mathbf{F} & \mathbf{G} & \mathbf{L} \\ & - & 0 & -1 & -2 & -3 & -4 \\ \mathbf{K} & -1 & ? & & & \\ \mathbf{G} & -2 & & & & \\ \mathbf{L} & -3 & & & & \end{pmatrix}$$

É necessário calcular o valor oriundo da célula diagonal superior esquerda, do topo e da esquerda da matriz. Calculando os valores para $[1][1]$ representado pelo ponto de interrogação na matriz anterior. O valor do *score* será o mesmo valor de *match* para resíduo idêntico. Para resíduos diferentes, o valor do *score* será o mesmo valor de *mismatch*. Calculando o valor diagonal

$$diagonal = M[i-1][j-1] + score$$

Para $M[1][1]$ ($i=1$ e $j=1$)

$$diagonal = M[0][0] + score$$

$$diagonal = 0 + (+1)(match) = 1$$

Calculando o topo

$$topo = M[i - 1][j] + gap\ penalty$$

Para $M[1][1]$ $gap\ penalty=-1$

$$topo = M[0][1] + (-1) = -1 + (-1) = -2$$

Calculando a esquerda

$$esquerda = M[i][j - 1] + gap\ penalty$$

Para $M[1][1]$ $gap\ penalty=-1$

$$esquerda = M[1][0] + (-1) = (-1) + (-1) = -2$$

Através desses cálculos, obtivemos os valores para diagonal=1, topo=-2 e esquerda=-2. O valor da célula $M[1][1]$ é a máxima desses valores

$$M[1][1] = \max(1, -2, -2) = 1$$

$$M = \begin{pmatrix} & - & \mathbf{K} & \mathbf{F} & \mathbf{G} & \mathbf{L} \\ & - & 0 & -1 & -2 & -3 & -4 \\ \mathbf{K} & -1 & \mathbf{1} & & & \\ \mathbf{G} & -2 & & & & \\ \mathbf{L} & -3 & & & & \end{pmatrix}$$

Calculando a célula $M[1][2]$

$$diagonal = -1 + (-1)(mismatch) = -2$$

$$topo = -2 + (-1) = -3$$

$$esquerda = 1 + (-1) = 0$$

$$M[1][2] = \max(-2, -3, 0) = 0$$

$$M = \begin{pmatrix} & - & K & F & G & L \\ & - & 0 & -1 & -2 & -3 & -4 \\ K & -1 & 1 & 0 & & & \\ G & -2 & & & & & \\ L & -3 & & & & & \end{pmatrix}$$

Calculando e preenchendo toda matriz

$$M = \begin{pmatrix} & - & K & F & G & L \\ & - & 0 & -1 & -2 & -3 & -4 \\ K & -1 & 1 & 0 & -1 & -2 \\ G & -2 & 0 & 0 & 1 & 0 \\ L & -3 & -1 & -1 & 0 & 2 \end{pmatrix}$$

Mostrado em vermelho na última matriz temos os valores ótimos que dará origem as sequências para o alinhamento. O alinhamento é iniciado pela última célula preenchida, no exemplo é a $M[3][4]$ que possui o valor 2. O alinhamento $M[3][4]$ se originou da diagonal, logo as duas sequências receberão resíduo L, sequência 1=L e sequência 2=L. $M[2][3]$ se originou da diagonal, logo as duas sequências receberão resíduo G, sequência 1=GL e sequência 2=GL. $M[1][2]$ se originou da esquerda, a sequência 1 receberá o resíduo F e a sequência 2 receberá um *gap*, sequência 1=FGL e sequência 2=-GL. $M[1][1]$ se originou da diagonal, logo as duas sequências receberão o resíduo K, sequência 1=KFGL e sequência 2=K-GL. No caso do alinhamento se originar do topo, coloca-se um *gap* na sequência 1 e o resíduo na sequência 2. No final dos cálculos obtivemos a melhor possibilidade de alinhamento, o alinhamento ótimo (108, 109).

Sequência1 : KFGL

Sequência2 : K-GL

Apêndice C

Coeficiente de correlação de Pearson (ρ)

Através da ferramenta de plotagem de dados XMGRACE (110) foi calculado o coeficiente de correlação de Pearson entre as previsões de mutações realizadas em SARS-CoV. Entre a tabela 4.4 (MutaBind2) e tabela 4.6 (BeAtMuSiC) $\rho = 0.64$ (figura C.1), entre a tabela 4.4 (MutaBind2) e tabela 4.6 (mCSM-PPI2) $\rho = 0.09$ (figura C.2) e entre a tabela 4.6 (BeAtMuSiC) e tabela 4.6 (mCSM-PPI2) $\rho = 0.3$ (figura C.3). Sendo assim, os valores da tabela 4.6 (mCSM-PPI2) apresenta menor correlação entre as demais estimativas, pois a tabela 4.4 (MutaBind2) e tabela 4.6 (BeAtMuSiC) possui alta correlação, com $\chi^2 = 0.41$ (figura C.1).

Figura C.1 – Diagrama de dispersão mostrando o coeficiente de correlação entre a tabela 4.4 (MutaBind2) e tabela 4.6 (BeAtMuSiC). Estes valores possuem alta correlação com $\chi^2 = 0.41$.

Figura C.2 – Diagrama de dispersão mostrando o coeficiente de correlação entre a tabela 4.4 (MutaBind2) e tabela 4.6 (mCSM-PPI2).

Figura C.3 – Diagrama de dispersão mostrando o coeficiente de correlação entre a tabela 4.6 (BeAtMuSiC) e tabela 4.6 (mCSM-PPI2).