

UNIVERSIDADE ESTADUAL PAULISTA “JÚLIO DE MESQUITA FILHO”

INSTITUTO DE BIOCÊNCIAS DE BOTUCATU

DEPARTAMENTO: BIOFÍSICA E FARMACOLOGIA

Vinícius Santos de Sousa

**Estudos estruturais teóricos do complexo formado entre o
domínio extracelular da proteína Spike do vírus SARS-CoV-2
ligado ao receptor celular humano ACE2 utilizando
espalhamento de raios-X a baixo ângulo**

Variantes de preocupação B.1.1.7, B.1.351, P.1 e B.1.617.2

Botucatu, São Paulo

Janeiro de 2023

Vinícius Santos de Sousa

Estudos estruturais teóricos do complexo formado entre o domínio extracelular da proteína Spike do vírus SARS-CoV-2 ligado ao receptor celular humano ACE2 utilizando espalhamento de raios-X a baixo ângulo

Variantes de preocupação B.1.1.7, B.1.351, P.1 e B.1.617.2

Trabalho de Conclusão de Curso apresentado ao Curso de Física Médica da Universidade Estadual Paulista “Júlio de Mesquita Filho” como parte dos requisitos necessários para a obtenção do título de Bacharel em Física Médica.

Orientador: Professor Doutor Mario de Oliveira Neto

Botucatu, São Paulo

Janeiro de 2023

FICHA CATALOGRÁFICA ELABORADA PELA SEÇÃO TÉC. AQUIS. TRATAMENTO DA INFORM.
DIVISÃO TÉCNICA DE BIBLIOTECA E DOCUMENTAÇÃO - CÂMPUS DE BOTUCATU - UNESP

BIBLIOTECÁRIA RESPONSÁVEL: ROSANGELA APARECIDA LOBO-CRB 8/7500

Sousa, Vinícius Santos de.

Estudos estruturais teóricos do complexo formado entre o domínio extracelular da proteína Spike do vírus SARS-CoV-2 ligado ao receptor celular humano ACE2 utilizando espalhamento de raios-X a baixo ângulo : variantes de preocupação B.1.1.7, B.1.351, P.1 e B.1.617.2 / Vinícius Santos de Sousa. - Botucatu, 2023

Trabalho de conclusão de curso (bacharelado - Física Médica) - Universidade Estadual Paulista "Júlio de Mesquita Filho", Instituto de Biociências de Botucatu

Orientador: Mario de Oliveira Neto

Capes: 20901003

1. Glicoproteína da espícula de coronavírus. 2. Enzima conversora da angiotensina. 3. SARS-CoV-2. 4. COVID-19 (Doença). 5. Small-angle X-ray scattering.

Palavras-chave: Proteína Spike; Receptor Binding Domain; SARS-CoV-2; SAXS; Variantes de preocupação.

Vinícius Santos de Sousa

Estudos estruturais teóricos do complexo formado entre o domínio extracelular da proteína Spike do vírus SARS-CoV-2 ligado ao receptor celular humano ACE2 utilizando espalhamento de raios-X a baixo ângulo

Variantes de preocupação B.1.1.7, B.1.351, P.1 e B.1.617.2

Trabalho de Conclusão de Curso apresentado ao Curso de Física Médica da Universidade Estadual Paulista “Júlio de Mesquita Filho” como parte dos requisitos necessários para a obtenção do título de Bacharel em Física Médica.

Comissão Examinadora:

Professor Doutor Mário de Oliveira Neto
Orientador

Professora Doutora Flavia Karina Delella
Examinadora

Professor Doutor Paulo Fernando de Arruda Mancera
Examinador

Dedico este trabalho aos meus familiares, amigos, a todos aqueles que perderam seus entes queridos durante a pandemia do Coronavírus e especialmente a todos os 700 mil brasileiros vítimas da COVID-19.

AGRADECIMENTOS

Ao meu orientador Prof. Dr. Mário de Oliveira Neto por todo conhecimento passado a mim, agradecer pelo apoio, pela oportunidade e por acreditar e confiar em mim ao longo da minha graduação me orientando nesse trabalho e também em minhas iniciações científicas.

Ao meu Prof. Dr. Joel Mesa Hormaza por ser um excelente coordenador de curso durante a minha graduação, um ótimo professor e uma pessoa incrível que sempre esteve disposto a conversar comigo e me ajudar com meus problemas durante esses cinco anos na universidade.

Aos meus familiares, minhas tias, minha avó e principalmente a minha mãe Lindaci Lucio e meu primo Igor Duarte por estarem sempre me incentivando durante todo esse período ao longo do curso.

As minhas tão especiais amigas de turma Isabella Scarpa e Maria Gabriela, por batalharem e caminharem comigo durante todos os momentos da graduação, indo nas aulas, realizando trabalhos, participando de reuniões com professores, estudando para provas e por estarem sempre me apoiando, principalmente nos momentos mais difíceis da minha vida na UNESP. Obrigado por todos os momentos incríveis que tive com vocês em Botucatu, seja na vida ou na universidade, por todas as ocasiões que curtimos e também pelas que sofremos juntos, sou muito grato pela existência de vocês.

Aos meus outros especiais amigos de turma Sabrina Spigaroli, Tiago Macedo, Gabriele Martins, Julia Bucci, Gabriel Nardelli, Mariana de Carvalho e Beatriz Magiore pelo companheirismo, ajuda e apoio desde o início e durante esses cinco anos no curso, sou muito grato por ter a companhia e amizade de vocês durante todo esse período, obrigado por tornarem meus dias mais tranquilos e divertidos no dia a dia e na universidade.

Aos meus coleguinhas, Helena Galindo, Heloisa Chain, Ana Carolina Alvarez e Pietro Ferreira da Biologia, Julia Caetano da Nutrição, Amanda Muniz da Biomedicina e Tainara Cateli da Veterinária. Vocês foram essenciais para que eu me formasse e enfrentasse todos os desafios que esse curso me trouxe, sou grato por todos os momentos maravilhosos juntos, sempre vou agradecer a UNESP por ter conhecido vocês.

A todos os alunos da Física Médica das turmas XVI, XVII, XVIII e XIX que me proporcionaram ótimos momentos, especialmente minhas especiais amigas Lorraine Requena, Vitoria Duarte, Júlia Thomazini, Mariana Pelegati, Larissa Picelli e Camila Carvalho, obrigado pelos momentos maravilhosos, por sempre estarem ao meu lado e me ajudarem durante a graduação e situações complicadas que passei.

Aos meus amigos Enzo Mantovani, Beatriz Santos, Eduarda Lima, Alessandra Nathalia e Beatriz Pataro que conheci no meu último ano da graduação e em tão pouco tempo se tornaram pessoas muito especiais pra mim. Obrigado por todas as situações incríveis que vocês me proporcionaram e me ajudar a passar por momentos conturbados que tive no ano de 2022.

Ao Diretório Acadêmico do Instituto de Biociências “Cariamidae”, a Recepção Unificada e a Liga Acadêmica de Física Médica por me proporcionarem um grande aprendizado e muitas experiências na universidade, além de conhecer pessoas incríveis e me presentear com grandes amizades.

A CNPq por me auxiliar com os benefícios e me ajudar a desenvolver as minhas iniciações científicas que foram de grande importância para o desenvolvimento deste trabalho.

A todos os funcionários da Diretoria Técnica Acadêmica que sempre me auxiliaram ao longo da graduação tirando dúvidas e me ajudando com as burocracias que surgiram ao longo do caminho.

A todos os meus amigos de longa data da cidade de Osasco, especialmente para Maria Julia e Tamires Amorim, que são um dos maiores responsáveis por eu ter ingressado na UNESP e estar onde estou hoje, obrigado por me apoiarem durante todo esse período mesmo estando longe.

A todas as outras pessoas que contribuíram e colaboraram de alguma maneira para que eu realizasse esse trabalho de conclusão de curso.

“Uma flor nasceu na rua! É feia. Mas é uma flor. Fudou o asfalto, o tédio, o nojo e o ódio.”

(Carlos Drummond de Andrade)

RESUMO

O surgimento de um novo coronavírus SARS-CoV-2 (agente causador da COVID-19) na China no final de 2019 causou, e segue causando até o momento, preocupações significativas em termos de saúde pública mundial. A região RBD (Receptor Binding Domain) da glicoproteína trimérica Spike do coronavírus quando ligado ao receptor celular humano ACE2 (Angiotensin-Converting Enzyme 2) formam um complexo responsável por permitir a entrada do vírus na célula promovendo a COVID-19. Para a realização do estudo desse complexo e das suas proteínas individuais foi utilizado o espalhamento de raios-X a baixo ângulo (SAXS), sendo ele um método muito útil para a caracterização de proteínas em solução e análise dos parâmetros geométricos das estruturas. Diante disso, o presente trabalho teve como objetivo, realizar uma análise teórica de SAXS a partir de simulações de curvas espalhamento, utilizando estruturas proteicas de alta resolução coletadas no banco de dados PDB (Protein Data Bank), da proteína Spike em sua forma nativa e de suas variantes de preocupação. Além do que, o trabalho também buscou promover um estudo das interfaces de interação e oligomerização da proteína Spike nativa e analisar se os resíduos mutados em suas variantes fazem parte da interface de trimerização e/ou da interface de formação do complexo com a ACE2, verificando a influência dessas mutações na interação dessas possíveis interfaces. Assim, diante das curvas simuladas das estruturas, foram obtidos resultados relacionados à forma das proteínas, baseado no perfil de suas $p(r)$, e dos seus parâmetros estruturais, raio de giro (R_g) e diâmetro máximo ($D_{máx}$). Também foram adquiridos resultados referentes aos aminoácidos da Spike nativa responsáveis pela sua interface de trimerização e interação na formação do complexo entre a glicoproteína e o receptor ACE2. Além disso, foram obtidos dados relativos aos aminoácidos mutados das variantes da proteína Spike e se avaliou a influência dessas mutações na interface de oligomerização do trímero e na interface de formação do complexo com a ACE2. Desse modo, a análise dos resultados mostrou que as mutações não fazem parte da interface de trimerização da proteína, porém, estão presentes na região RBD das variantes de preocupação. Observou-se que as mutações não alteram a forma das estruturas, entretanto, podem ser responsáveis pelas alterações dos parâmetros estruturais das proteínas em solução. Além disso, se verificou que as mutações na RBD influenciam no aumento da afinidade da proteína Spike na formação do complexo com a ACE2 e no escape do vírus do sistema imunológico humano. Logo, é necessário ressaltar a importância de se obter resultados experimentais de SAXS das estruturas estudadas em solução, visando realizar um melhor estudo das proteínas e uma melhor comparação das mesmas com suas estruturas obtidas por meio de cristalização, sendo também fundamental realizar uma análise de todas as mutações existentes na proteína Spike das variantes além das presentes na região RBD, assim possibilitando uma verificação mais concreta da predominância dessas variantes em relação à forma nativa do SARS-CoV-2.

Palavras-Chave: Proteína Spike. Receptor Binding Domain. SARS-CoV-2. SAXS. Variantes de preocupação.

ABSTRACT

The emergence of a new coronavirus SARS-CoV-2 (the causative agent of COVID-19) in China at the end of 2019 caused and continues to cause significant concerns in terms of global public health. The RBD (Receptor Binding Domain) region of the Spike trimeric glycoprotein of the coronavirus when linked to the human cell receptor ACE2 (Angiotensin-Converting Enzyme 2) forms a complex responsible for allowing the entry of the virus into the cell promoting COVID-19. To carry out the study of this complex and its individual proteins, small angle X-rays scattering (SAXS) was used, which is a very useful method for the characterization of proteins in solution and analysis of the geometric parameters of the structures. In view of this, the present work aimed to carry out a theoretical analysis of SAXS from simulations of scattering curves, using high-resolution protein structures collected in the PDB (Protein Data Bank) database, of the Spike protein in its native form and of its worry variants. In addition, the work also sought to promote a study of the interaction and oligomerization interfaces of the native Spike protein and to analyze whether the mutated residues in its variants are part of the trimerization interface and/or the complex formation interface with ACE2, verifying the influence of these mutations on the interaction of these possible interfaces. Thus, in view of the simulated curves of the structures, results related to the shape of the proteins were obtained, based on the profile of their $p(r)$, and their structural parameters, radius of gyration (R_g) and maximum diameter (D_{max}). Results regarding the native Spike amino acids responsible for its trimerization interface and interaction in the formation of the complex between the glycoprotein and the ACE2 receptor were also acquired. Furthermore, data related to the mutated amino acids of the Spike protein variants were obtained and the influence of these mutations on the oligomerization interface of the trimer and on the interface of complex formation with ACE2 was evaluated. Thus, the analysis of the results showed that the mutations are not part of the protein trimerization interface, however, they are present in the RBD region of the variants of concern. It was observed that the mutations do not change the shape of the structures, however, they may be responsible for changes in the structural parameters of the proteins in solution. In addition, it was found that mutations in RBD influence the increase in the affinity of the Spike protein in the formation of the complex with ACE2 and in the escape of the virus from the human immune system. Therefore, it is necessary to emphasize the importance of obtaining experimental SAXS results of the structures studied in solution, aiming to carry out a better study of the proteins and a better comparison of the same with their structures obtained through crystallization, and it is also essential to carry out an analysis of all the existing mutations in the Spike protein of the variants in addition to those present in the RBD region, thus enabling a more concrete verification of the predominance of these variants in relation to the native form of SARS-CoV-2.

Keywords: Binding Domain Receiver. SARS-CoV-2. SAXS. Spike Protein. Variants of concern.

LISTA DE FIGURAS

Figura 1: Diagramas de fita das estruturas do ectodomínio SARS-CoV-2 S (A) junto às subunidades S1 SARS-CoV-2 (B) e S2 SARS-CoV-2 (C) (Walls, <i>et. al.</i> 2020).	18
Figura 2: Configuração de um experimento de SAXS (Kikhney, <i>et. al.</i> 2015).	19
Figura 3: Esquema do vetor de espalhamento q (Itri, <i>et. al.</i> 2015).	20
Figura 4: Intensidades de espalhamento e função de distribuição de pares $p(r)$ para diferentes objetos geométricos (Svergun, <i>et. al.</i> 2003).	23
Figura 5: Estrutura proteica da região RBD da Spike isolada obtida no banco de dados SASBDB através do código SASDJG4 (Custódio, <i>et. al.</i> 2020).	29
Figura 6: Ajustes da estrutura proteica realizados nos programas FoXS e CRY SOL da região RBD da Spike isolada a sua curva experimental.	29
Figura 7: (A) Curvas de SAXS simuladas nos programas FoXS e CRY SOL da estrutura proteica obtida no SASBDB através do código SASDJG4. (B) $p(r)$ geradas a partir das curvas simuladas e da curva experimental de código SASDJG4 depositada no SASBDB.	29
Figura 8: Estrutura proteica da região RBD da Spike isolada obtida no banco de dados SASBDB através do código SASDJG4 (Custódio, <i>et. al.</i> 2020). Estrutura foi colorida do N-Terminal (vermelho) ao C-Terminal (azul).	30
Figura 9: Estrutura proteica da região RBD da Spike isolada em cristalografia, obtida através do complexo SARS-CoV 2 RBD e ACE2, coletada no banco de dados PDB por meio do código 6M0J (Lan, <i>et. al.</i> 2020). Estrutura foi colorida do N-Terminal (vermelho) ao C-Terminal (azul).	30
Figura 10: Estrutura proteica da região RBD da Spike isolada em solução (SASJG4: verde) sobreposta a sua estrutura cristalizada (6M0J: azul claro).	31
Figura 11: Estrutura da proteína trimérica (roxo, verde e azul) da região completa do domínio extracelular da Spike isolada com a região RBD fechada de código 6VXX. A estrutura foi rotacionada 90° em relação ao eixo z.	32
Figura 12: (A) Curvas simuladas nos programas CRY SOL e FoXS da estrutura proteica de código 6VXX. (B) $p(r)$ geradas a partir das curvas simuladas.	32
Figura 13: Estrutura da proteína trimérica (roxo, verde e azul) da região completa do domínio extracelular da Spike isolada com uma região RBD aberta de código 6VYB. A estrutura foi rotacionada 90° em relação ao eixo z.	33

Figura 14: (A) Curvas simuladas nos programas CRY SOL e FoXS da estrutura proteica de código 6VYB. (B) $p(r)$ geradas a partir das curvas simuladas.	33
Figura 15: Estrutura proteica do complexo formado entre a enzima ACE2 (verde) e a região RBD da Spike (azul claro) de código 6M0J.	34
Figura 16: (A) Curvas simuladas nos programas CRY SOL e FoXS da estrutura proteica de código 6M0J. (B) $p(r)$ geradas a partir das curvas simuladas.	34
Figura 17: Estrutura da proteína trimérica (roxo, verde e azul claro) da região completa do domínio extracelular da Spike isolada com uma região RBD aberta (verde) e uma ACE2 ligada (amarela) de código 7A94.	35
Figura 18: Estrutura da proteína trimérica (roxo, verde e azul claro) da região completa do domínio extracelular da Spike isolada com duas regiões RBD abertas (verde e azul claro) e duas ACE2 ligadas (amarela e rosa claro) de código 7A97.	35
Figura 19: Estrutura da proteína trimérica (roxo, verde e azul) da região completa do domínio extracelular da Spike isolada com três regiões RBD abertas (verde e azul claro) e três ACE2 ligadas (amarela, rosa claro e cinza) de código 7A98.	35
Figura 20: Curvas simuladas nos programas FoXS (A) e CRY SOL (B) das estruturas proteicas de código 7A94, 7A97 e 7A98.	36
Figura 21: $p(r)$ geradas a partir das curvas simuladas nos programas FoXS e CRY SOL das estruturas proteicas de código 7A94, 7A97 e 7A98.	36
Figura 22: (A) Curvas simuladas no programa FoXS das estruturas proteicas da região completa do domínio extracelular da Spike isolada com uma região RBD aberta de códigos 6VYB, 7R14, 8DLL, 8DLO e 7WG9. (B) $p(r)$ geradas a partir das curvas simuladas.	37
Figura 23: (A) Curvas simuladas no programa CRY SOL das estruturas proteicas da região completa do domínio extracelular da Spike isolada com uma região RBD aberta de códigos 6VYB, 7R14, 8DLL, 8DLO e 7WG9. (B) $p(r)$ geradas a partir das curvas simuladas.	37
Figura 24: (A) Curvas simuladas no programa FoXS das estruturas proteicas do complexo formado entre a enzima ACE2 e a região RBD da Spike de códigos 6M0J, 8DLK, 8DLNN, 8DLQ e 7TEW. (B) $p(r)$ geradas a partir das curvas simuladas.	38
Figura 25: (A) Curvas simuladas no programa CRY SOL das estruturas proteicas do complexo formado entre a enzima ACE2 e a região RBD da Spike de códigos 6M0J, 8DLK, 8DLNN, 8DLQ e 7TEW. (B) $p(r)$ geradas a partir das curvas simuladas.	38

Figura 26: (A) Curvas simuladas no programa FoXS das estruturas proteicas da região completa do domínio extracelular da Spike isolada com uma região RBD aberta ligada a uma ACE2 de código 7A94, 8DLJ, 7VXD e 7W98. (B) $p(r)$ geradas a partir das curvas simuladas. 39

Figura 27: (A) Curvas simuladas no programa CRY SOL das estruturas proteicas da região completa do domínio extracelular da Spike isolada com uma região RBD aberta ligada a uma ACE2 de código 7A94, 8DLJ, 7VXD e 7W98. (B) $p(r)$ geradas a partir das curvas simuladas..... 39

Figura 28: Interface de trimerização do domínio extracelular da proteína Spike isolada. 40

Figura 29: Ligações de hidrogênio e pontes salinas nas interfaces da região RBD (azul) ligada a ACE2 (verde). 41

LISTA DE TABELAS

Tabela 1: Parâmetros estruturais calculados da região RDB da Spike isolada obtida no banco de dados SASBDB através do código SASDJG4 junto ao erro percentual das curvas simuladas em relação à curva experimental.....	29
Tabela 2: Parâmetros estruturais calculados da estrutura proteica de código 6VXX a partir de suas curvas simuladas e de suas $p(r)$ geradas.....	32
Tabela 3: Parâmetros estruturais calculados da estrutura proteica de código 6VYB a partir de suas curvas simuladas e de suas $p(r)$ geradas.....	33
Tabela 4: Parâmetros estruturais calculados da estrutura proteica de código 6M0J a partir de suas curvas simuladas e de suas $p(r)$ geradas.....	34
Tabela 5: Parâmetros estruturais calculados das estruturas proteicas de código 7A94, 7A97 e 7A98 a partir de suas curvas simuladas e de suas $p(r)$ geradas.	36
Tabela 6: Parâmetros estruturais calculados das estruturas proteicas de código 6VYB, 7R14, 8DLL, 8DLO e 7WG9 a partir de suas curvas simuladas e de suas $p(r)$ geradas.....	37
Tabela 7: Parâmetros estruturais calculados das estruturas proteicas de código 6M0J, 8DLK, 8DLNN, 8DLQ e 7TEW a partir de suas curvas simuladas e de suas $p(r)$ geradas.	38
Tabela 8: Parâmetros estruturais calculados das estruturas proteicas de código 7A94, 8DLJ, 7VXD e 7W98 a partir de suas curvas simuladas e de suas $p(r)$ geradas.....	39
Tabela 9: Ligações de hidrogênio e pontes salinas nas interfaces SARS-CoV-2 RBD–ACE2.	40
Tabela 10: Mutações na proteína Spike das variantes Alfa (B.1.1.7), Beta (B.1.351), P.1 (Gamma) e Delta (B.1.617.2) do SARS-CoV-2.....	42
Tabela 11: Ligações de hidrogênio e pontes salinas nas interfaces de interação da região RBD (Estrutura 1), da proteína Spike Nativa e das variantes Alfa, Beta, Gamma e Delta do SARS-CoV-2, com o receptor humano ACE2 (Estrutura 2). A coluna Pos. mostra a posição das interações, já a distância entre os átomos dessas interações é dada em Å	43

LISTA DE ABREVIACÕES E SIGLAS

ACE2 - Enzima conversora da Angiotensina 2, do inglês, Angiotensin-Converting Enzyme 2

Dmáx - Diâmetro máximo

DNA - Ácido desoxirribonucleico

HR1 - Repetição Heptad 1, do inglês, Heptad Repeat 1

MERS-CoV - Síndrome respiratória do Oriente Médio coronavírus, do inglês, Middle East Respiratory Syndrome Coronavirus

NTD - Domínio N-terminal, do inglês, N-Terminal Domain

OMS - Organização Mundial da Saúde

PDB - Banco de dados de proteínas, do inglês, Protein Data Bank

p(r) - Função de distribuição de pares

RBD - Domínio de ligação ao receptor, do inglês, Receptor-Binding Domain

Rg - Raio de giro

RMN - Ressonância Magnética Nuclear

RNA - Ácido ribonucleico

SARS-CoV-2 - Síndrome respiratória aguda grave coronavírus-2, do inglês, Severe Acute Respiratory Syndrome Coronavirus 2

SASBDB - Banco de dados biológicos de espalhamento a baixo ângulo, do inglês, Small Angle Scattering Biological Data Bank

SAXS - Espalhamento de raios-X a baixo ângulo, do inglês, Small Angle X-rays Scattering

S1 - Subunidade 1

S2 - Subunidade 2

SUMÁRIO

1. INTRODUÇÃO	17
1.1. Proteína Spike e receptor celular humano ACE2	17
1.2. Variantes de preocupação B.1.1.7, B.1.351, P.1 e B.1.617.2	18
1.3. Espalhamento de raios-X a baixo ângulo (SAXS)	19
1.3.1. Vetor de Espalhamento q	19
1.3.2. Intensidade de Espalhamento $I(q)$	21
1.3.3. Lei de Guinier	21
1.3.4. Função de distribuição de pares $p(r)$	22
1.4. Pacote de programas ATSAS	23
1.4.1. Programa de avaliação e manipulação de modelos CRY SOL	24
1.4.2. Programas de processamento de dados experimentais GNOM e PRIMUS	24
1.5. FoXS Server	25
1.6. Programa PISA (Protein Interfaces, Surfaces and Assemblies)	25
2. OBJETIVO	26
2.1. Objetivo geral	26
2.2. Objetivos específicos	26
3. METODOLOGIA	26
3.1. Obtenção e análise das curvas de SAXS e estruturas de alta resolução	26
3.2. Simulação das curvas de SAXS e cálculo dos parâmetros estruturais	28
4. RESULTADOS	28
4.1. Simulação das curvas de SAXS e cálculos dos parâmetros estruturais da proteína Spike nativa	28
4.1.1. Região RBD da Spike isolada	28
4.1.2. Região completa do domínio extracelular da Spike isolada	31
4.1.3. Complexo formado entre o receptor ACE2 e a região RBD da Spike	33
4.1.4. Complexo formado entre o receptor ACE2 e a região completa do domínio extracelular da Spike isolada	34
4.2. Simulação das curvas de SAXS e cálculos dos parâmetros estruturais das variantes de preocupação da proteína Spike	36
4.2.1. Proteína Spike com uma região RBD aberta	37
4.2.2. Complexo formado entre o receptor ACE2 e a região RBD da Spike	37
4.2.3. Spike com uma região RBD aberta ligada a uma ACE2	38
4.3. Interface das estruturas proteicas	39
4.3.1. Interface de trimerização do domínio extracelular da Spike isolada	40

4.3.2.	Complexo formado entre a enzima ACE2 e a região RBD da Spike Nativa.	40
4.3.3.	Mutações nas variantes de preocupação e suas interfaces de formação do complexo SARS-CoV-2 RBD e ACE2	41
5.	DISCUSSÃO.....	44
6.	CONCLUSÃO	49
7.	REFERÊNCIAS.....	49

1. INTRODUÇÃO

Em dezembro de 2019 um novo e altamente patogênico coronavírus SARS-CoV-2 (Síndrome respiratória aguda grave coronavírus-2, do inglês, *Severe Acute Respiratory Syndrome Coronavirus 2*) iniciou um surto na cidade de Wuhan, localizada província de Hubei na China, se espalhando rapidamente por todo o país e para outros países ao redor do mundo, em que, a sua rápida disseminação internacional representou uma emergência mundial em saúde pública. Como resultado, o epicentro Wuhan e as cidades vizinhas foram isolados para minimizar a disseminação contínua e a OMS (Organização Mundial da Saúde) anunciou uma Emergência de Saúde Pública de importância internacional devido à disseminação rápida e global do SARS-CoV-2 (Lan, *et. al.* 2020).

Análises filogenéticas dos genomas de coronavírus revelaram que o SARS-CoV-2 é um membro do gênero *Betacoronavirus* (pertencente à família *Coronaviridae*), que inclui SARS-CoV, MERS-CoV (Síndrome respiratória do Oriente Médio coronavírus, do inglês, *Middle East Respiratory Syndrome Coronavirus*), coronavírus relacionados à SARS de morcego, bem como outros identificados em humanos e diversas espécies animais (Lan, *et. al.* 2020).

Novos vírus emergentes têm surgido e são principalmente de origem zoonótica, na maioria das vezes sendo o resultado de uma transmissão entre espécies. Isso pode acontecer por meio de vários mecanismos de variação genética, gerando novas características que permitem que o vírus se ligue e entre em uma nova célula hospedeira com maior eficiência, evite o sistema imunológico e modifique sua virulência. Os vírus adquirem mutações ao longo do tempo, o que pode levar ao surgimento de novas variantes, sendo as de preocupação aquelas mais transmissíveis e provocam infecções mais graves de Covid-19 (Shahhosseini N, *et. al.* 2021).

1.1. Proteína Spike e receptor celular humano ACE2

A infecção do vírus SARS-CoV-2 se inicia a partir da interação da glicoproteína Spike (proteína de adsorção viral) em seu envelope com o receptor celular humano ACE2 (Enzima conversora da Angiotensina 2, do inglês, *Angiotensin-Converting Enzyme 2*). Essa ligação forma um complexo heteromérico que desencadeia uma cascata de eventos responsável pela fusão entre a membrana celular viral e humana, assim permitindo a entrada do vírus na célula (Lan, *et. al.* 2020).

Os coronavírus usam a glicoproteína Spike homotrimérica (compreendendo uma subunidade 1 (S1) e uma subunidade 2 (S2) em cada monômero da Spike) no envelope (Figura 1) para se ligar aos seus receptores celulares. A região RBD (Domínio de ligação ao receptor, do inglês, *Receptor-Binding Domain*) da proteína Spike do SARS-CoV-2 e sua interação com o receptor celular ACE2 mostraram que a ligação com o receptor induz a dissociação do S1 com o ACE2, levando o S2 a transitar de uma pré-fusão metaestável para uma pós-fusão estável que é essencial para a fusão da membrana. Portanto, o reconhecimento de proteínas de adsorção viral e de receptores celulares é um determinante importante para a transmissão e tropismo de um vírus em células e tecidos (Wang, *et. al.* 2020).

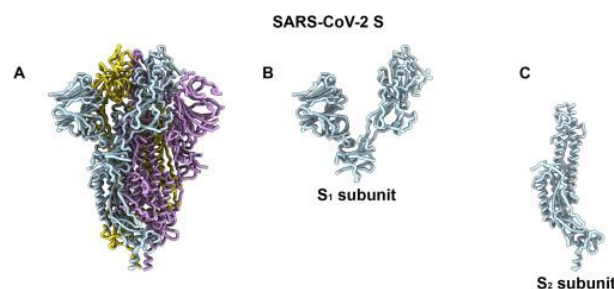


Figura 1: Diagramas de fita das estruturas do ectodomínio SARS-CoV-2 S (A) junto às subunidades S1 SARS-CoV-2 (B) e S2 SARS-CoV-2 (C) (Walls, *et. al.* 2020).

Nos estudos feitos por Benton, dos trimeros da Spike analisados, dois terços foram ligados a ACE2 (Benton, *et. al.* 2020). Das espécies não ligadas, foram observadas partículas de boa qualidade na conformação fechada não ligada e um número considerável de partículas não ligadas com um RBD aberto, bem como alguns em uma conformação intermediária, com uma forma fechada menos compacta, ligada a um único RBD desordenado. Então, a partir dos dados relacionados à representação das superfícies das estruturas obtidas, as duas conformações mais predominantes na amostra são da proteína Spike com uma região RBD aberta (16%) e da Spike com uma única RBD aberta liga a uma única ACE2 (28%), sugerindo que o complexo da Spike ligado a ACE2 totalmente aberto é provavelmente metaestável (Benton, *et. al.* 2020).

1.2. Variantes de preocupação B.1.1.7, B.1.351, P.1 e B.1.617.2

Uma variante de preocupação é aquela para a qual há evidência de aumento na transmissibilidade, maior gravidade na doença (por exemplo, aumento de hospitalizações ou mortes), redução significativa na neutralização por anticorpos gerados durante infecção ou vacinação anterior, eficácia reduzida de tratamentos ou vacinas ou falhas na detecção diagnóstica (Gobeil, *et. al.* 2021).

A variante Alfa ou B.1.1.7 surgiu no Reino Unido em setembro de 2020 e se espalhou pelo mundo, com relatos de aumento da transmissibilidade, virulência e mortalidade. Além da mutação D614G, uma substituição na região RBD da Spike N501Y resultou em uma melhor afinidade com o receptor ACE2. A mutação N501Y, sozinha ou em combinação com a deleção Δ H69/V70 no NTD (Domínio N-terminal, do inglês, *N-Terminal Domain*), ou a mutação na subunidade S2 P681H, não afeta substancialmente a neutralização do soro induzida pelas vacinas atuais (Gobeil, *et. al.* 2021).

A variante Beta ou B.1.351 foi descoberta por volta de outubro de 2020, na África do Sul, momento em que ocorreu a segunda onda da epidemia de SARS-CoV-2 no país. Em particular, a proteína Spike dessa variante contém a substituição D614G e nove mutações adicionais, incluindo um conjunto de mutações no NTD; três substituições, K417N, E484K e N501Y na região RBD; e uma mutação A701V perto do local de clivagem da furina (Wang Y, *et. al.* 2021).

A variante Gamma ou P.1, encontrada no Brasil em dezembro de 2020, emergente da linhagem B.1.1.28, tornou-se uma variante dominante no país. Essa variante, além da mutação D614G, contém 12 mutações na sua Spike, incluindo K417T, E484K e N501Y

no RBD; L18F, T20N, P26S, D138Y e R190S no domínio N-terminal NTD; e H655Y perto do local de clivagem da furina (Wang P, *et. al.* 2021).

Por fim, a variante Delta ou B.1.617.2 surgiu em outubro de 2020 na Índia, se espalhando rapidamente para até 135 países em de agosto de 2021. A variante Delta abriga múltiplas mutações na proteína Spike, incluindo duas substituições L452R e T478K na região RBD; a substituição D614G; um conjunto de mutações no domínio N-terminal NTD; e mais duas substituições perto do sítio de clivagem da furina e no HR1 (Repetição Heptad 1, do inglês, *Heptad Repeat 1*) (Wang Y, *et. al.* 2022).

1.3. Espalhamento de raios-X a baixo ângulo (SAXS)

O espalhamento de raios-X a baixo ângulo SAXS (do inglês, *Small Angle X-rays Scattering*) é um método muito útil para a caracterização estrutural de proteínas ordenadas e desordenadas em solução. Possui a função de fornecer informações sobre os tamanhos e formas de proteínas e complexos em tamanhos moleculares que variam de alguns KDa a GDa sob várias condições experimentais, que variam de extrema (alta pressão, alta temperatura, congelamento criogênico, entre outros) até o seu estado nativo. SAXS é um dos poucos métodos que permitem caracterizar quantitativamente a polidispersidade conformacional, em particular, de macromoléculas completamente ou parcialmente desordenadas, isso inclui proteínas de múltiplos domínios, na presença ou ausência de ligantes (Kikhney, *et. al.* 2015).

A configuração de um experimento de SAXS (Figura 2) se dá a partir de uma solução de partículas que é iluminada por um feixe de raios-X monocromático colimado, no instante em que esses raios-X entram em contato com a solução eles são espalhados e esse espalhamento é registrado por um detector de raios-X (Kikhney, *et. al.* 2015).

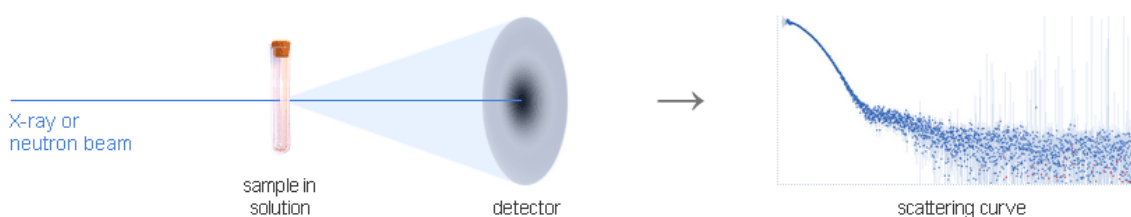


Figura 2: Configuração de um experimento de SAXS (Kikhney, *et. al.* 2015).

Os experimentos mais avançados de SAXS, geralmente, são conduzidos em fontes de sincrotron pelo fato de fornecerem raios-X de alto brilho e intensidade (Svergun, *et. al.* 2013).

1.3.1. Vetor de Espalhamento q

O vetor de espalhamento q é definido pela diferença vetorial entre o feixe de raios-X que emerge da amostra sem interagir com a mesma e aquele que é espalhado e atinge o detector numa dada posição (Itri, *et. al.* 2015) como visto na Figura 3 e definido na equação (1).

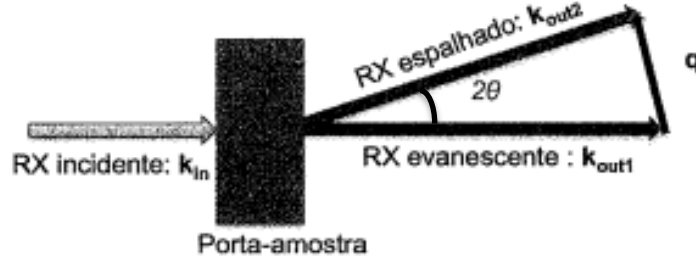


Figura 3: Esquema do vetor de espalhamento q (Itri, *et. al.* 2015).

$$q = k_{out2} - k_{out1}. \quad (1)$$

Os vetores k_{out2} e k_{out1} são chamados vetores de onda, em que sua direção aponta para a direção de propagação da onda eletromagnética. Ambos são definidos como K_i (semelhante ao número de onda), em que, $K_i = \frac{2\pi}{\lambda_i}$ em que λ_i é o comprimento de onda da radiação.

Para pequenos ângulos (2θ) podemos considerar um espalhamento elástico onde há uma conservação energia da radiação incidente espalhada, em que, conseqüentemente o comprimento de onda não é alterado durante o processo. A partir da equação do efeito Compton se tem que:

$$\lambda_f - \lambda_i = \Delta\lambda = \lambda_c(1 - \cos(\theta)). \quad (2)$$

Para θ muito pequeno $\cos(\theta) \cong 1$, ou seja, $\Delta\lambda \cong 0$, implicando em $\lambda_f \cong \lambda_i$, e então desprezamos o efeito Compton.

A partir da Figura 3, usando o teorema dos cossenos (Rezende, *et. al.* 2000) podemos deduzir o módulo do vetor espalhamento q :

$$q^2 = k_{out1}^2 + k_{out2}^2 - 2k_{out1}k_{out2}\cos(2\theta),$$

$$q^2 = \left(\frac{2\pi}{\lambda}\right)^2 + \left(\frac{2\pi}{\lambda}\right)^2 - 2\frac{2\pi}{\lambda}\frac{2\pi}{\lambda}\cos(2\theta),$$

$$q^2 = \frac{8\pi^2}{\lambda^2}(1 - \cos(2\theta)).$$

Como $1 - \cos(2\theta) = 2\sin(\theta)^2$ temos que o módulo do vetor de espalhamento q é dado por:

$$q = \frac{4\pi\sin(\theta)}{\lambda}, \quad (3)$$

em que λ é o comprimento de onda do feixe e θ o ângulo de espalhamento.

1.3.2. Intensidade de Espalhamento $I(q)$

No momento em que o detector de raios-X registra o padrão de espalhamento da solução da amostra, o solvente puro também é coletado e subtraído desse espalhamento, deixando apenas o sinal das partículas de interesse (Kikhney, *et. al.* 2015). O padrão de dispersão resultante está relacionado com a forma geral e o tamanho das partículas estudadas.

Devido às orientações aleatórias das partículas em solução, o padrão de espalhamento é isotrópico e, portanto, o padrão de espalhamento registrado, geralmente, por um detector bidimensional pode ser calculado radialmente (Kikhney, *et. al.* 2015).

A intensidade de espalhamento $I(q)$ é representada como uma função da transferência de momento q , representada por:

$$I(q) = \langle A(q)A(q)^* \rangle_\Omega, \quad (4)$$

com $\langle \rangle_\Omega$ representando a média esférica e a amplitude de dispersão $A(q)$ é uma transformada de Fourier da densidade de elétrons em excesso, sendo ela:

$$A(q) = \int \Delta\rho(\mathbf{r})e^{iq\cdot\mathbf{r}}d\mathbf{r}, \quad (5)$$

e $\Delta\rho(\mathbf{r}) = \rho(\mathbf{r}) - \rho_s$, com $\rho(\mathbf{r})$ a densidade eletrônica da partícula, ρ_s a densidade eletrônica do solvente e $\mathbf{r}$ o vetor posição de cada átomo. Esses padrões de dispersão isotrópica são exibidos como curvas unidimensionais de média radial $I(q)$ (Kikhney, *et. al.* 2015).

1.3.3. Lei de Guinier

Uma grandeza que pode ser extraída diretamente das curvas de SAXS é o raio de giro, R_g , da partícula espalhadora (Itri, *et. al.* 2015). Tal grandeza relaciona como a massa de uma dada partícula está distribuída ao redor de seu centro de massa, podendo ser visto analogamente como raio de inércia da mecânica clássica.

O raio de giro está relacionado ao deslocamento do centro de massa da proteína em relação a um eixo, quanto mais volume possuir uma proteína, maior o seu raio de giro, ou seja, podemos dizer, por exemplo, que quando uma proteína está desestruturando, o seu raio de giro está aumentando (Mariano, *et. al.* 2021).

A Equação de Guinier pode ser escrita como:

$$I(q) = I(0)e^{\frac{-q^2 R_g^2}{3}}. \quad (6)$$

A equação de Guinier é válida para ângulos próximos de 0, quando $q \rightarrow 0$ e $q \cdot R_g < 1,3$.

Aplicando logaritmo natural em ambos os lados da equação (6), observamos que:

$$\ln(I(q)) = \ln(I(0)) + \ln e^{\frac{-q^2 R_g^2}{3}},$$

$$\ln(I(q)) = -q^2 \cdot \frac{R_g^2}{3} + \ln(I(0)). \quad (7)$$

Para um sistema monodisperso, a partir da equação (7), podemos determinar experimentalmente o raio de giro da partícula realizado por meio do gráfico de Guinier $\ln(I(q)) \times q^2$. Quando temos $q = 0$ implica que $\ln(I(q)) = \ln(I(0))$, portanto $I(q) = I(0)$, ou seja, $I(0)$ é uma constante que pode ser obtida experimentalmente por Guinier através do coeficiente linear da equação. Também notamos que $R_g^2 / 3$ é proporcional ao coeficiente angular da reta, ou seja, o raio de giro pode ser calculado experimentalmente a partir da inclinação da reta.

1.3.4. Função de distribuição de pares $p(r)$

A $p(r)$ (função de distribuição de pares) contém a mesma informação da intensidade de espalhamento $I(q)$, mas a representação da $p(r)$ no espaço real é mais intuitiva (Svergun, *et. al.* 2003). Sendo r o par de distância entre dois elétrons dentro da partícula (fontes espalhadoras), a função $p(r)$ nos fornece informações sobre a geometria da partícula espalhadora e é proporcional a distribuição de distâncias entre elementos espalhadores dentro de uma dada partícula (Itri, *et. al.* 2015). Para sistemas não interagentes, ela pode ser obtida através de uma transformada inversa de Fourier a partir da função $I(q)$,

$$p(r) = \frac{1}{2\pi^2} \int_0^\infty I(q) q r \sin(qr) dq. \quad (8)$$

Para $p(r) = 0$, $r = D_{\text{máx}}$, assim obtendo o diâmetro máximo $D_{\text{máx}}$ da partícula.

Da $p(r)$ é possível obter o valor do raio de giro da partícula através do valor esperado de r^2 da distribuição de distância da partícula ao redor centro de sua distribuição de densidade de comprimento (Svergun, *et. al.* 2003):

$$R_g^2 = \frac{\int_0^{D_{\text{máx}}} p(r) r^2 dr}{2 \int_0^{D_{\text{máx}}} p(r) dr}. \quad (9)$$

Assim, de uma análise direta de uma curva de SAXS via função $p(r)$, podemos obter informação do $D_{\text{máx}}$, R_g e da geometria da partícula (Itri, *et. al.* 2015).

Além disso, informações sobre a forma da partícula em solução podem ser deduzidas visualmente da $p(r)$. A Figura 4 apresenta padrões de espalhamento e $p(r)$ característicos de objetos geométricos com mesmo $D_{\text{máx}}$ (Svergun, *et. al.* 2003).

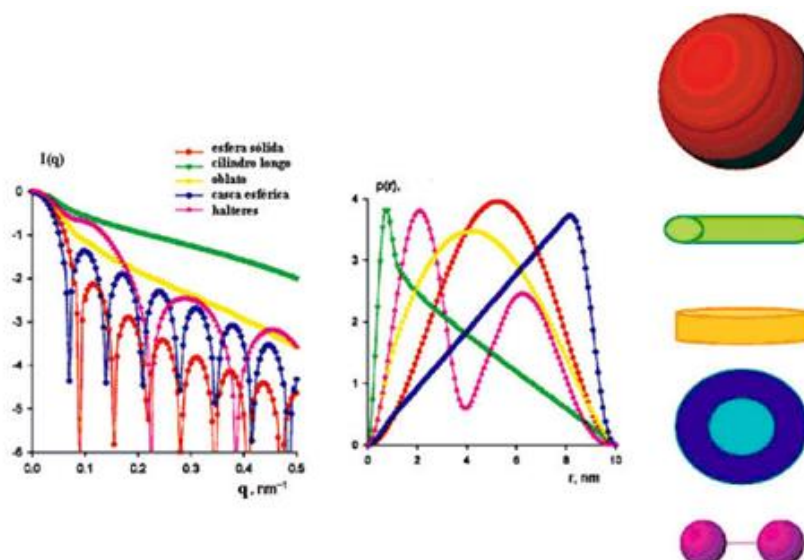


Figura 4: Intensidades de espalhamento e função de distribuição de pares $p(r)$ para diferentes objetos geométricos (Svergun, *et. al.* 2003).

Em geral, as curvas de espalhamento são um pouco menos instrutivas do que as funções $p(r)$ (Svergun, 1992). Partículas globulares (curva vermelha) exibem $p(r)$ em forma de sino com um máximo em torno de $D_{\text{máx}}/2$. Partículas alongadas têm distribuições distorcidas com um pico bem definido em pequenas distâncias, correspondentes ao raio da sua seção transversal (curva verde). Partículas achatadas exibem um pico bastante amplo (curva amarela), deslocado para distâncias menores que $D_{\text{máx}}/2$, porém não tanto quanto o máximo das alongadas. Um máximo deslocado para distâncias maiores que $D_{\text{máx}}/2$ geralmente indica uma partícula oca (curva Azul). Partículas em formato de halteres, que consistem de subunidades bem separadas, podem apresentar mais de um máximo, o primeiro corresponde às distâncias entre as subunidades, já os outros correspondem aos formatos das próprias subunidades (curva rosa). Partículas muito anisométricas (não isométricas) produzem curvas em espalhamento sem características, que decaem muito mais lentamente do que as de partículas isométricas com formatos bem definidos. As distâncias que ocorrem com mais frequência se manifestam como máximos ou ombros nos padrões de dispersão (ombro em $q = 0,1 \text{ nm}^{-1}$ no espalhamento da curva rosa) (Svergun, *et. al.* 2003).

1.4. Pacote de programas ATSAS

O pacote ATSAS consiste em um conjunto abrangente de programas para a análise de dados de espalhamento de raios-X a baixo ângulo de nano partículas ou de soluções de macromoléculas biológicas diluídas. Ele contém aplicações voltadas para avaliação e processamento de dados primários, métodos para a análise de flexibilidade e misturas, modelagem de esferas *ab initio* e validação de modelos. Ademais, conta com abordagens que utilizam informações de cristalografia de raios-X, espectroscopia de RMN (Ressonância Magnética Nuclear) ou modelagem de homologia atômica para construir modelos híbridos com base nos dados do espalhamento (Manalastas-Cantos, *et. al.* 2021). O pacote também contém módulo para modelagem interativa com o sistema gráfico molecular PyMOL.

1.4.1. Programa de avaliação e manipulação de modelos CRY SOL

Utilizando a expansão multipolar para cálculo rápido do padrão de dispersão de média esférica e levando em consideração a camada de hidratação, o programa CRY SOL avalia o espalhamento em solução de macromoléculas com estrutura atômica conhecida. Dadas as coordenadas atômicas usando como entrada um arquivo PDB (Banco de dados de proteína, do inglês, *Protein Data Bank*), ele pode prever a curva de espalhamento da solução ou ajustar a estrutura atômica curva de espalhamento experimental utilizando o volume médio de solvente deslocado por grupo atômico e o contraste da camada de solvatação (Svergun, *et. al.* 1995).

A curva de SAXS de uma proteína em uma solução suficientemente diluída é proporcional ao espalhamento de um único objeto espalhador. A intensidade de espalhamento pode ser descrita como:

$$I(q) = \langle |A_a(q) - \rho_0 A_s(q) + \delta\rho A_b(q)|^2 \rangle_\Omega, \quad (10)$$

em que, $A_a(q)$ é a amplitude espalhada para proteína no vácuo, $A_s(q)$ e $A_b(q)$ são respectivamente, as amplitudes de espalhamento para o volume excluído (solvente) e para a camada de hidratação, ambas com densidade eletrônica respectivamente, ρ_s e ρ_b . A contribuição da amplitude é acrescida por $\delta\rho A_b(q)$ onde $\delta\rho = \rho_b - \rho_s$. O símbolo $\langle \rangle_\Omega$ significa a promediação da média.

Quando a estrutura cristalográfica da proteína é conhecida, a amplitude $I(q)$ pode ser determinada utilizando método de expansão multipolar. O único parâmetro desconhecido que precisa ser ajustado é o volume excluído da proteína, V , e a diferença entre a densidade eletrônica entre o solvente e a camada de hidratação. Esses parâmetros e um fator de escala c , são obtidos num procedimento de ajuste utilizando o programa CRY SOL (Svergun, *et. al.* 1995). Assim a intensidade de espalhamento obtida experimentalmente por SAXS e a $I(q)$ obtida por simulação (10) a partir de um banco de dados de proteínas podem ser comparadas:

$$\chi^2(r_0, \delta\rho) = \frac{1}{N_p} \sum_{i=1}^{N_p} \left[\frac{I_e(q_i) - cI(q_i, r_0, \delta\rho)}{\sigma(q_i)} \right]^2, \quad (11)$$

em que N_p é o número de pontos experimentais, $I(q_i, r_0, \delta\rho)$ é o perfil calculado, $\sigma(q_i)$ são os erros da intensidade experimental $I_e(q_i)$ e c é o fator de escala entre os dados experimentais e o modelo cristalográfico.

1.4.2. Programas de processamento de dados experimentais GNOM e PRIMUS

O GNOM é um programa de transformação indireta para lidar com dados de SAXS. Ele lê curvas de espalhamento em uma dimensão, que podem ser afetadas por distorção instrumental, e utiliza a transformada inversa de Fourier para gerar e avaliar a sua função de distribuição de pares $p(r)$ (para sistemas monodispersos) ou a função de distribuição de tamanhos $D(R)$ (para sistemas polidispersos) (Svergun, 1992).

O programa PRIMUS realiza manipulações utilizando arquivos de curvas experimentais de SAXS e avalia parâmetros oriundos dos gráficos de Guinier e Porod, como raio de giro, volume de Porod, $I(0)$ e peso da molécula (Konarev, *et. al.* 2003).

1.5. FoXS Server

FoXS é um servidor da web capaz de realizar o cálculo rápido e preciso de uma curva teórica de SAXS de uma determinada estrutura atômica e comparar com o seu perfil experimental, sendo isso um componente básico em qualquer modelagem atômica baseada em SAXS. O servidor possui uma capacidade exclusiva de contabilizar vários estados que contribuem para um único perfil de SAXS observado. Os estados múltiplos podem corresponder à heterogeneidade conformacional (conformações múltiplas da mesma proteína ou complexo) e/ou heterogeneidade composicional (conteúdo variável das moléculas de proteína e ligantes no sistema) (Schneidman-Duhovny, *et. al.* 2016).

A entrada para o servidor web é um arquivo de formato PDB e um perfil SAXS experimental caso esteja disponível. O FoXS calcula o perfil teórico e o gera como um arquivo de texto e como um gráfico de $\log(I(q)) \times q$. O perfil é calculado usando a fórmula de Debye (12) para espalhadores esféricos (Glatter, *et. al.* 1982),

$$I(q) = \sum_i^N \sum_j^N f_i(q) f_j(q) \frac{\sin(qr_{ij})}{qr_{ij}}, \quad (12)$$

em que, N é o numero de átomos na molécula, r_{ij} é a distância euclidiana entre esses átomos, e $f_i(q)$ e $f_j(q)$ representam os fatores de forma atômica isotrópica dos átomos i e j . O fator de forma $f_i(q)$ leva em consideração o solvente deslocado, bem como a camada de hidratação (Schneidman-Duhovny, *et. al.* 2010). O ajuste adequado é calculado minimizando a função χ^2 em relação a c , dado por:

$$\chi^2 = \frac{1}{N_p} \sum_{i=1}^{N_p} \left[\frac{I_e(q_i) - cI(q_i)}{\sigma(q_i)} \right]^2, \quad (13)$$

em que N_p é o número de pontos experimentais, $I(q_i)$ é o perfil calculado, $\sigma(q_i)$ são os erros da intensidade experimental $I_e(q_i)$ e c é o fator de escala entre os dados experimentais e o modelo cristalográfico.

1.6. Programa PISA (Protein Interfaces, Surfaces and Assemblies)

O programa PISA, pertencente ao pacote de programas CCP4 (Winn, *et. al.* 2011), é uma ferramenta interativa para a exploração de interfaces macromoleculares (proteínas, DNA/RNA e ligantes), previsão de estruturas quaternárias prováveis (conjuntos), pesquisas em bancos de dados de interfaces e conjuntos estruturalmente semelhantes, bem como pesquisas em várias montagens e parâmetros de entrada PDB (Krissinel, *et. al.* 2007).

2. OBJETIVO

2.1. Objetivo geral

Este trabalho teve como objetivo, a partir de estruturas de proteínas resolvidas que estão depositadas no banco de dados PDB realizar uma análise estrutural teórica de SAXS da forma nativa e das variantes B.1.1.7, B.1.351, P.1 e B.1.617.2 da proteína Spike do vírus Sars-CoV-2, em que, diante dessas estruturas, também será verificado se os resíduos mutados fazem parte das interfaces de oligomerização da Spike e/ou na formação do complexo com a ACE2.

2.2. Objetivos específicos

Neste projeto serão utilizadas informações estruturais da proteína do domínio extracelular (ectodomínio) da proteína Spike do vírus SARS-CoV-2, da enzima ACE2 que atua como receptor na célula humana e do complexo formado entre ambas. Diante disso, este projeto apresenta como objetivos específicos:

- Compreender as interfaces de oligomerização do trímero da proteína Spike isolada, e da interação entre a região RBD da Spike com o receptor celular humano ACE2;
- Verificar em quais resíduos ocorrem às mutações na proteína Spike de cada variante;
- Analisar se os resíduos mutados nas variantes da Spike fazem parte da interface de trimerização e/ou da interface da formação do complexo com a ACE2, verificando a influência dessas mutações na interação dessas possíveis interfaces;
- Realizar estudos envolvendo os conceitos relacionados a estruturas de proteínas, os conceitos teóricos de SAXS e o entendimento de variados programas de análise de dados, como o servidor da web FoXS, os programas pertencentes aos pacotes ATSAS e o programa PISA;
- Construir curvas de SAXS simuladas das proteínas individuais do trímero da Spike nos seus diferentes estados, região RBD aberta e fechada, do complexo formado entre a região RBD da Spike e com o receptor celular humano ACE2, e do complexo formado entre o trímero da região completa do domínio extracelular da Spike com a ACE2. Diante disso, será feita uma análise dessas curvas com o intuito de obter informações teóricas sobre os parâmetros estruturais dessas proteínas.

3. METODOLOGIA

3.1. Obtenção e análise das curvas de SAXS e estruturas de alta resolução

A curva de SAXS e estrutura proteica da região RBD da Spike isolada foram obtidas através do banco de dados biológicos de espalhamento a baixo ângulo SASBDB (do inglês, *Small Angle Scattering Biological Data Bank*) (Kikhney, *et. al.* 2019) sendo

elas obtidas através do código SASDJG4 (Custódio, *et. al.* 2020). Já as estruturas proteicas utilizadas foram obtidas através do banco de dados PDB (Berman, *et. al.* 2000), sendo elas:

- As proteínas do domínio extracelular da Spike nativa isolada através dos códigos 6VXX (Walls, *et. al.* 2020) e 6VYB (Walls, *et. al.* 2020), em que pode ser verificada a sua possível interface de trimerização;
- O complexo formado entre a enzima ACE2 e o domínio de ligação RBD da Spike nativa através do código PDB 6M0J (Lan, *et. al.* 2020), em que pode ser verificada a sua interface de interação;
- Os complexos formados entre a enzima ACE2 e a proteína do domínio extracelular da Spike nativa isolada através dos códigos PDB 7A94 (Benton, *et. al.* 2020), 7A97 (Benton, *et. al.* 2020) e 7A98 (Benton, *et. al.* 2020);
- As proteínas do domínio extracelular da Spike isolada com uma região RBD aberta 7R14 (Alfa ou B.1.1.7) (Wrobel, *et. al.* 2022), 8DLL (Beta ou B.1.351) (Mannar, *et. al.* 2022), 8DLO (Gamma ou P.1) (Mannar, *et. al.* 2022) e 7WG9 (Delta ou B.1.617.2) (Cui, *et. al.* 2022) em que pode ser verificada a sua possível interface de trimerização e quais resíduos mutados realizam essas interações;
- Os complexos formados entre a enzima ACE2 e o domínio de ligação RBD da Spike através dos códigos 8DLK (Alfa ou B.1.1.7) (Mannar, *et. al.* 2022), 8DLN (Beta ou B.1.351) (Mannar, *et. al.* 2022), 8DLQ (Gamma ou P.1) (Mannar, *et. al.* 2022), e 7TEW (Delta ou B.1.617.2) (Saville, *et. al.* 2022) em que pode ser verificada a sua interface de interação e quais resíduos mutados fazem parte dessas interações;
- Os complexos formados entre uma enzima ACE2 ligada a uma região RBD aberta da proteína do domínio extracelular da Spike isolada através dos códigos 8DLJ (Alfa ou B.1.1.7) (Mannar, *et. al.* 2022), 7VXD (Beta ou B.1.351) (Wang Y, *et. al.* 2021) e 7W98 (Delta ou B.1.617.2) (Wang Y, *et. al.* 2022).

Os códigos PDB 6VSB, 6VYB, 7R14, 8DLL, 8DLO e 7WG9, referentes ao ectodomínio da Spike isolada, possuem cerca 1284 resíduos em sua sequência, tendo somente o domínio de ligação RBD (resíduos 329 a 544) da Spike disponível no complexo formado com a ACE2, de códigos PDB 6M0J, DLK, 8DLN, 8DLQ e 7TEW.

Inicialmente, a partir de arquivos do PDB, serão analisadas as interfaces das proteínas e suas mutações em cada variante, buscando verificar se os resíduos mutados nas variantes da Spike que fazem parte da interface de trimerização e/ou da interface de formação do complexo com a ACE2. Assim, utilizando o programa PISA, será feita uma verificação de suas interfaces para analisar a influência dessas mutações na interação dessas possíveis interfaces.

É importante destacar que as possíveis interfaces de oligomerização, assim como os complexos formados *a priori* podem ser biologicamente ativos ou apenas artefatos da cristalização, devido a isso ressaltamos a importância desse estudo.

3.2. Simulação das curvas de SAXS e cálculo dos parâmetros estruturais

Para gerar as curvas de SAXS a partir das estruturas de alta resolução foram utilizados os programas do pacote ATSAS (Manalastas-Cantos, *et. al.* 2021) e o servidor da web FoXS (Schneidman-Duhovny, *et. al.* 2016). As curvas das proteínas foram simuladas utilizando o programa CRY SOL (Svergun, *et. al.* 1995) e o servidor da web FoXS (Schneidman-Duhovny, *et. al.* 2016), em que se obteve um espalhamento teórico dessas estruturas. Dessa forma, visando adquirir mais informações, foi efetuada uma comparação entre os dados das curvas obtidas pelos diferentes programas.

A partir das curvas teóricas geradas pelos programas CRY SOL e FoXS, foram calculados o R_g teórico a partir de uma análise de Guinier utilizando o programa PRIMUS (Konarev, *et. al.* 2003), e gerada a função de pares de distâncias $p(r)$ teórica utilizando o programa GNOM (Svergun, 1992), em que, a partir dela também se obteve o R_g , assim como, o seu $D_{máx}$ teórico.

4. RESULTADOS

4.1. Simulação das curvas de SAXS e cálculos dos parâmetros estruturais da proteína Spike nativa

O espalhamento de raios-X a baixo ângulo tornou-se uma técnica amplamente utilizada para a caracterização estrutural de moléculas em solução. O perfil SAXS de uma macromolécula, $I(q)$, é calculado subtraindo o perfil SAXS do tampão do perfil SAXS da macromolécula no tampão. O perfil pode ser convertido em uma distribuição aproximada de distâncias atômicas emparelhadas da macromolécula $p(r)$ por meio de uma transformada inversa de Fourier. A modelagem atômica baseada em SAXS pode ser usada em uma ampla gama de aplicações, como comparação de soluções e estruturas cristalinas, modelagem de regiões ausentes na estrutura de alta resolução e determinação de estados biologicamente relevantes do cristal.

4.1.1. Região RBD da Spike isolada

O ectodomínio da proteína Spike fortemente glicosilada forma um trímero onde cada protômero consiste em uma subunidade S2 central e uma subunidade S1 distal. Cada subunidade S1 compreende um domínio de ligação ao receptor RBD que quando isolada e em solução possui certa conformação com parâmetros a serem analisados.

Na Figura 6 são apresentados os resultados relacionados aos ajustes, realizados nos programas FoXS e CRY SOL, da estrutura proteica de alta resolução da região RBD isolada (Figura 5) à curva experimental. Já na Figura 7 são exibidas as curvas simuladas da mesma estrutura proteica nos programas CRY SOL e FoXS e as $p(r)$ geradas a partir dessas curvas e da curva experimental. Na Tabela 1 é são expostos os parâmetros estruturais das estruturas, R_g e $D_{máx}$, obtidos através da curva experimental e das curvas simuladas, em que, também foram exibidos os erros relacionados aos dados calculados das curvas simuladas em relação aos dados da curva experimental.



Figura 5: Estrutura proteica da região RBD da Spike isolada obtida no banco de dados SASBDB através do código SASDJG4 (Custódio, et. al. 2020). A estrutura é mostrada como cartoon.

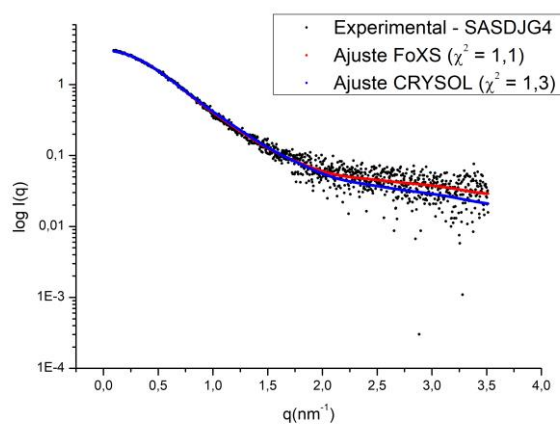


Figura 6: Ajustes da estrutura proteica realizados nos programas FoXS e CRY SOL da região RBD da Spike isolada a sua curva experimental. A curva e o a estrutura foram obtidas no banco de dados SASBDB através do código SASDJG4 (Custódio, et. al. 2020).

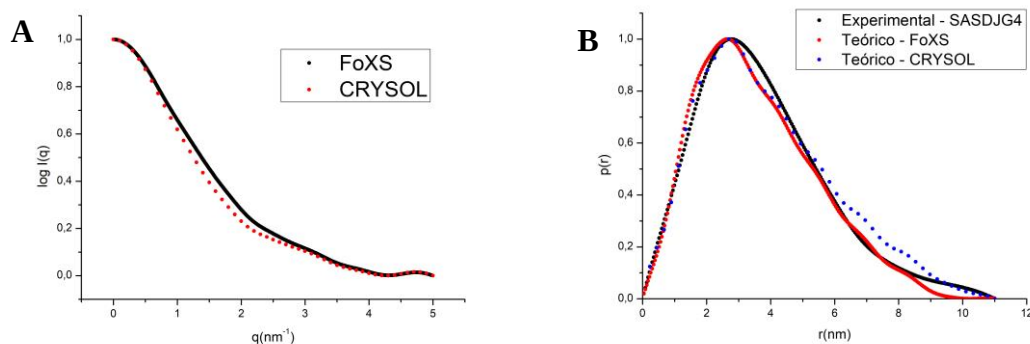


Figura 7: (A) Curvas de SAXS simuladas nos programas FoXS e CRY SOL da estrutura proteica obtida no SASBDB através do código SASDJG4. (B) $p(r)$ geradas a partir das curvas simuladas e da curva experimental de código SASDJG4 depositada no SASBDB.

Tabela 1: Parâmetros estruturais calculados da região RBD da Spike isolada obtida no banco de dados SASBDB através do código SASDJG4 junto ao erro percentual das curvas simuladas em relação à curva experimental.

Parâmetros	Experimental	CRY SOL	FoXS	Erro CRY SOL (%)	Erro FoXS (%)
Rg (Guinier) (nm)	2,9	2,8	2,8	3,4	3,4
Rg ($p(r)$) (nm)	3,0	3,1	2,9	3,3	3,3
Dmáx (nm)	11	11	11	0	0

No programa PyMOL, foi realizada a sobreposição da estrutura da região RBD isolada em solução (Figura 8) com sua estrutura cristalizada (Figura 9) sendo essa sobreposição exposta na Figura 10. Nas Figuras 8 e 9 são mostradas as estruturas secundárias (Hélices Alfa e Folhas Beta) com múltiplas cores da região RBD, sendo possível notar o N-Terminal (azul) e C-Terminal (vermelho) em ambas. Já na Figura 10 são mostrados os domínios sobrepostos de ambas as estruturas terciárias com duas diferentes cores (SASDJG4: verde e 6M0J: azul claro).

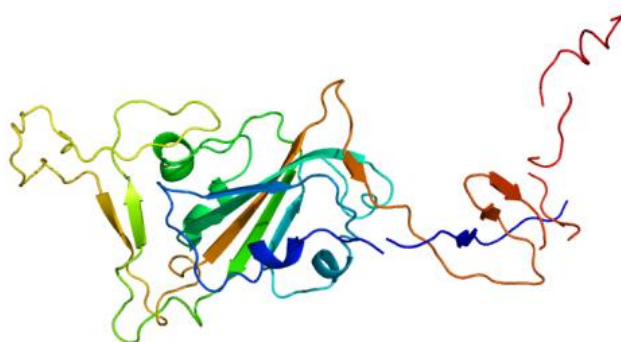


Figura 8: Estrutura proteica da região RBD da Spike isolada obtida no banco de dados SASBDB através do código SASDJG4 (Custódio, *et. al.* 2020). Estrutura foi colorida do N-Terminal (vermelho) ao C-Terminal (azul). A estrutura é mostrada como cartoon.



Figura 9: Estrutura proteica da região RBD da Spike isolada em cristalografia, obtida através do complexo SARS-CoV 2 RBD e ACE2, coletada no banco de dados PDB por meio do código 6M0J (Lan, *et. al.* 2020). Estrutura foi colorida do N-Terminal (vermelho) ao C-Terminal (azul). A estrutura é mostrada como cartoon.

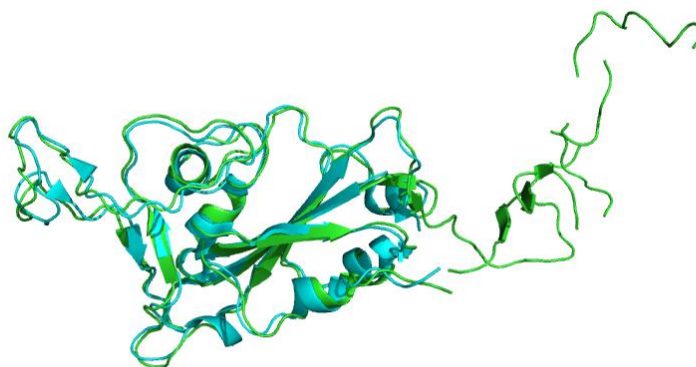


Figura 10: Estrutura proteica da região RBD da Spike isolada em solução (SASJG4: verde) sobreposta a sua estrutura cristalizada (6M0J: azul claro). As estruturas são mostradas como *cartoon*.

4.1.2. Região completa do domínio extracelular da Spike isolada

A proteína Spike é uma proteína de fusão trimérica que existe em uma conformação de pré-fusão metaestável que sofre um rearranjo estrutural substancial para fundir a membrana viral com a membrana da célula hospedeira. Para se engajar em um receptor de célula hospedeira, o RBD em S1 sofre movimentos em sua conformação que ocultam ou expõem transitoriamente os determinantes da ligação ao receptor. Esses dois estados são chamados de conformação fechada, onde corresponde ao estado inacessível ao receptor, como mostrado na Figura 11, e conformação aberta, onde corresponde ao estado acessível ao receptor, como mostrado na Figura 13.

Nas Figuras 11 e 13 são apresentadas as curvas de SAXS simuladas nos programas CRY SOL e FoXS relacionadas às estruturas proteicas apresentadas, respectivamente, nas Figuras 12 e 14, e as $p(r)$ geradas com base nessas curvas. Já nas Tabelas 2 e 3 são exibidos os parâmetros estruturais teóricos, R_g e $D_{\text{máx}}$, obtidos através das curvas e $p(r)$ geradas.

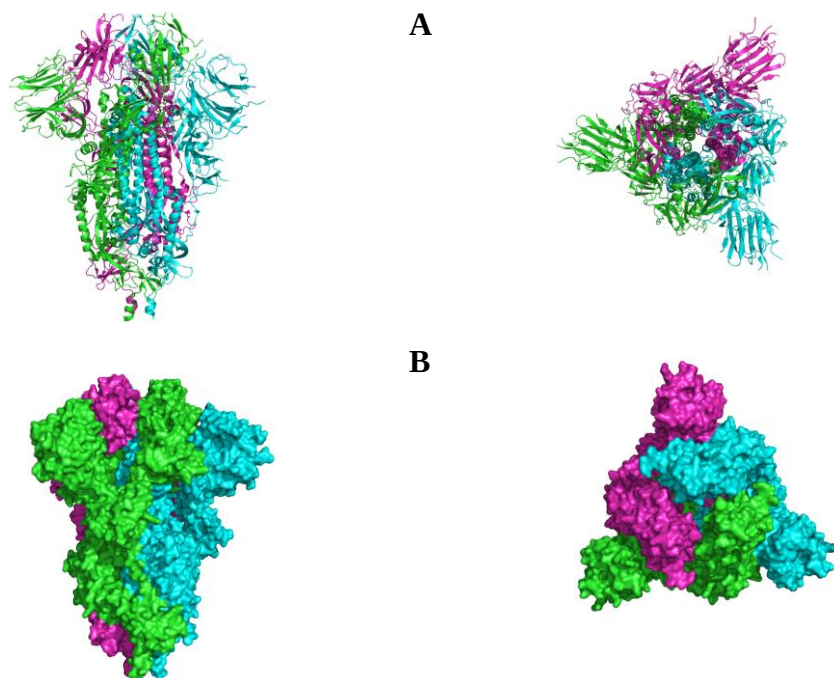


Figura 11: Estrutura da proteína trimérica (roxo, verde e azul) da região completa do domínio extracelular da Spike isolada com a região RBD fechada de código 6VXX. A estrutura foi rotacionada 90° em relação ao eixo z. As estruturas são mostradas como cartoon (A) e superfície (B).

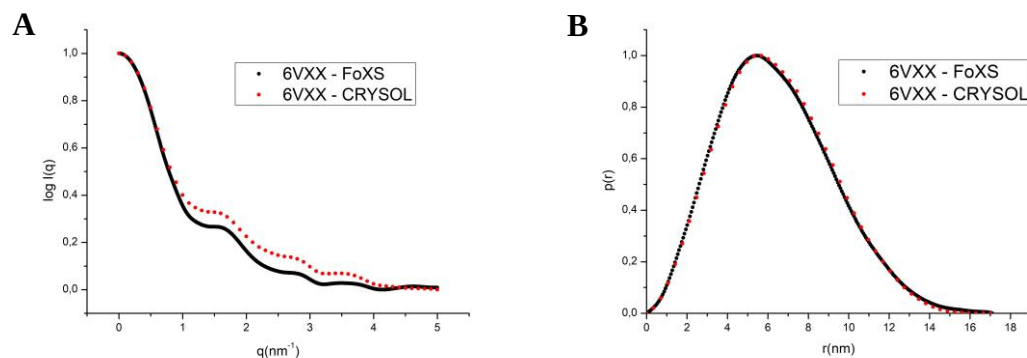


Figura 12: (A) Curvas simuladas nos programas CRY SOL e FoXS da estrutura proteica de código 6VXX. (B) $p(r)$ geradas a partir das curvas simuladas.

Tabela 2: Parâmetros estruturais calculados da estrutura proteica de código 6VXX a partir de suas curvas simuladas e de suas $p(r)$ geradas.

Parâmetros	CRY SOL	FoXS
Rg (Guinier) (nm)	4,9	4,8
Rg ($p(r)$) (nm)	4,9	5,0
Dmáx (nm)	17	17,1

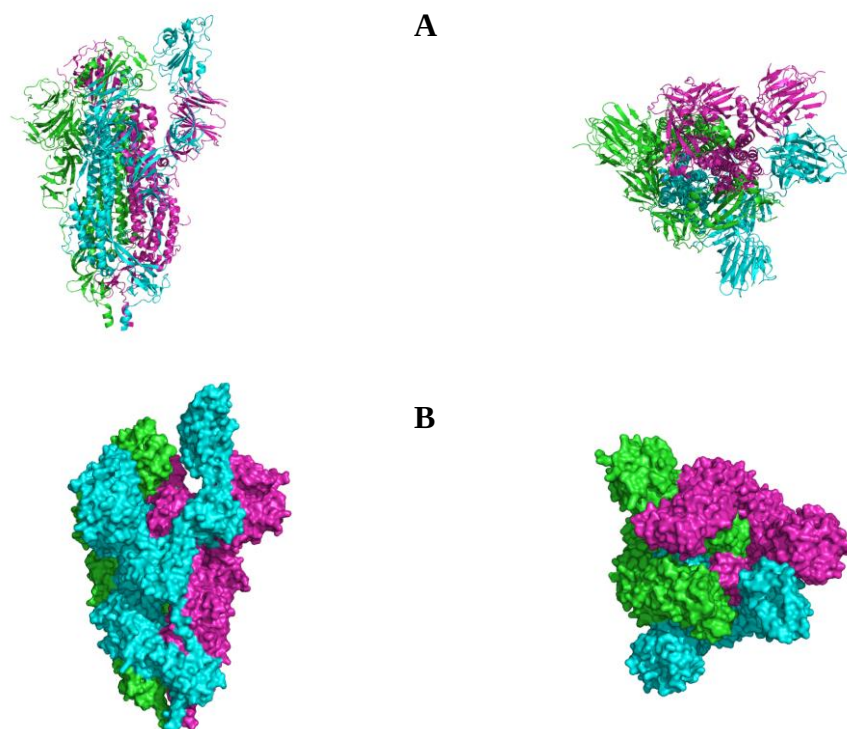


Figura 13: Estrutura da proteína trimérica (roxo, verde e azul) da região completa do domínio extracelular da Spike isolada com uma região RBD aberta de código 6VYB. A estrutura foi rotacionada 90° em relação ao eixo z. As estruturas são mostradas como cartoon (A) e superfície (B).

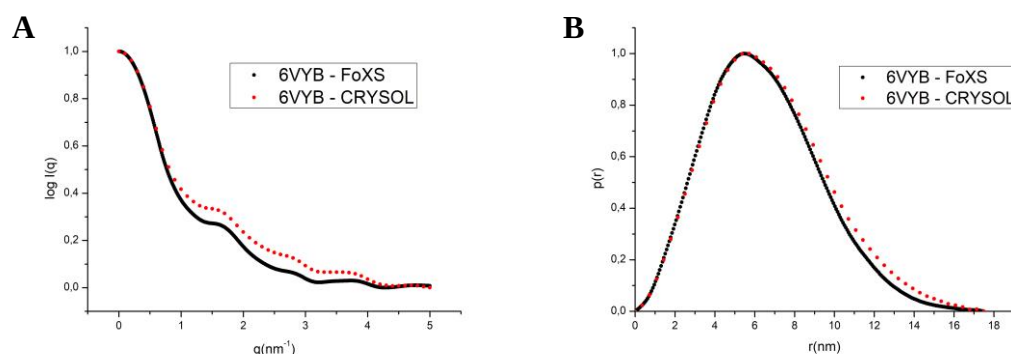


Figura 14: (A) Curvas simuladas nos programas CRY SOL e FoXS da estrutura proteica de código 6VYB. (B) $p(r)$ geradas a partir das curvas simuladas.

Tabela 3: Parâmetros estruturais calculados da estrutura proteica de código 6VYB a partir de suas curvas simuladas e de suas $p(r)$ geradas.

Parâmetros	CRY SOL	FoXS
Rg (Guinier) (nm)	5,0	4,9
Rg ($p(r)$) (nm)	5,2	5,0
Dmáx (nm)	17,5	17,5

4.1.3. Complexo formado entre o receptor ACE2 e a região RBD da Spike

A ligação da região RBD da Spike ao receptor ACE2 forma um complexo que é um passo inicial crítico para que o SARS-CoV-2 entre nas células-alvo, sendo possível obter as estruturas resolvidas desses complexos por meio de cristalografia.

Na Figura 16 são exibidas as curvas de SAXS simuladas nos programas CRY SOL e FoXS, relativa ao complexo proteico exibido na Figura 15 e as $p(r)$ criadas a partir dessas curvas. Já na Tabela 4 são mostrados os parâmetros estruturais teóricos, R_g e $Dmáx$, obtidos através das curvas e $p(r)$ geradas.

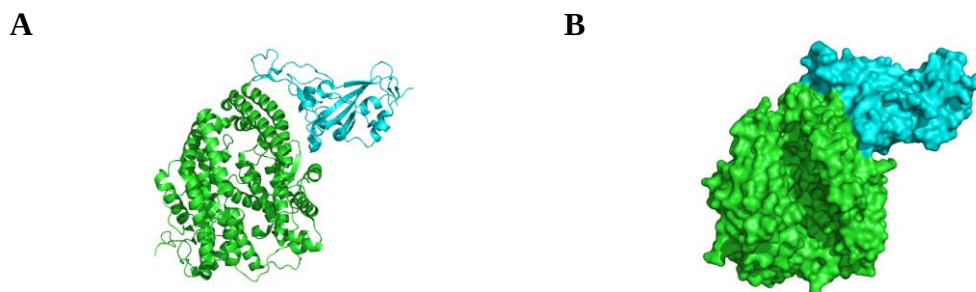


Figura 15: Estrutura proteica do complexo formado entre a enzima ACE2 (verde) e a região RBD da Spike (azul claro) de código 6M0J. As estruturas são mostradas como cartoon (A) e superfície (B).

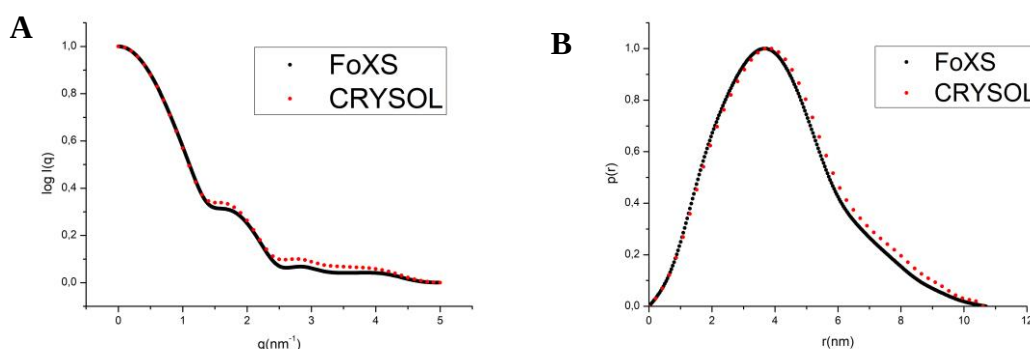


Figura 16: (A) Curvas simuladas nos programas CRY SOL e FoXS da estrutura proteica de código 6M0J. (B) $p(r)$ geradas a partir das curvas simuladas.

Tabela 4: Parâmetros estruturais calculados da estrutura proteica de código 6M0J a partir de suas curvas simuladas e de suas $p(r)$ geradas.

Parâmetros	CRY SOL	FoXS
R_g (Guinier) (nm)	3,2	3,1
R_g ($p(r)$) (nm)	3,3	3,2
$Dmáx$ (nm)	10,6	10,7

4.1.4. Complexo formado entre o receptor ACE2 e a região completa do domínio extracelular da Spike isolada

Nas membranas do vírus, as proteínas Spike existem tanto em uma forma fechada quanto em uma forma aberta, na qual a região S1 se abre para expor o RBD para a ligação de ACE2, em que, essa exposição ocorre de diferentes maneiras, sendo elas, com uma RBD aberta ligada a uma ACE2, duas RBD abertas ligadas duas ACE2 e três RBD abertas ligadas a três ACE2.

Na Figura 20 são apresentadas as curvas de SAXS simuladas nos programas CRY SOL e FoXS relacionadas às estruturas proteicas apresentadas nas Figuras 17, 18 e 19, sendo as $p(r)$ geradas a partir dessas curvas exibidas na Figura 21. Já na Tabela 5

são mostrados os parâmetros estruturais teóricos, R_g e $D_{máx}$, obtidos através das curvas e $p(r)$ geradas.



Figura 17: Estrutura da proteína trimérica (roxo, verde e azul claro) da região completa do domínio extracelular da Spike isolada com uma região RBD aberta (verde) e uma ACE2 ligada (amarela) de código 7A94. As estruturas são mostradas como *cartoon* (A) e *superfície* (B).

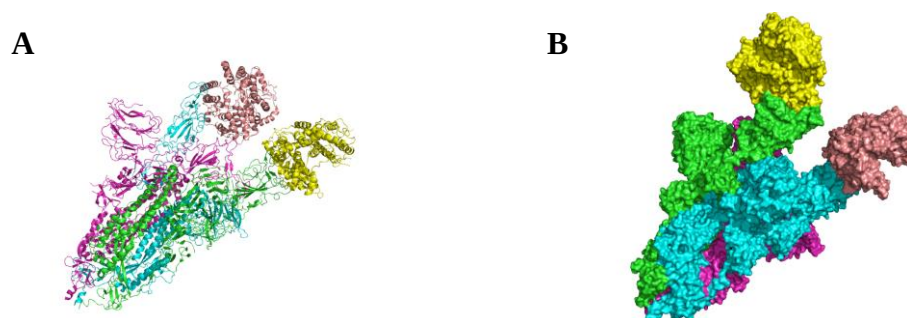


Figura 18: Estrutura da proteína trimérica (roxo, verde e azul claro) da região completa do domínio extracelular da Spike isolada com duas regiões RBD abertas (verde e azul claro) e duas ACE2 ligadas (amarela e rosa claro) de código 7A97. As estruturas são mostradas como *cartoon* (A) e *superfície* (B).



Figura 19: Estrutura da proteína trimérica (roxo, verde e azul) da região completa do domínio extracelular da Spike isolada com três regiões RBD abertas (verde e azul claro) e três ACE2 ligadas (amarela, rosa claro e cinza) de código 7A98. As estruturas são mostradas como *cartoon* (A) e *superfície* (B).

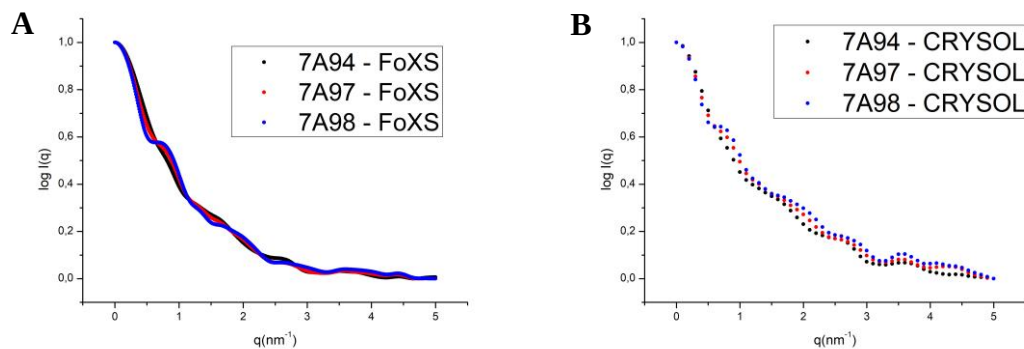


Figura 20: Curvas simuladas nos programas FoXS (A) e CRY SOL (B) das estruturas proteicas de código 7A94, 7A97 e 7A98.

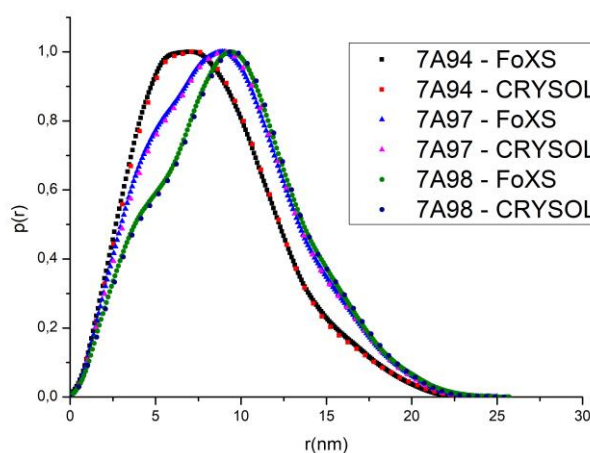


Figura 21: $p(r)$ geradas a partir das curvas simuladas nos programas FoXS e CRY SOL das estruturas proteicas de código 7A94, 7A97 e 7A98.

Tabela 5: Parâmetros estruturais calculados das estruturas proteicas de código 7A94, 7A97 e 7A98 a partir de suas curvas simuladas e de suas $p(r)$ geradas.

Códigos PDB	7A94		7A97		7A98	
Parâmetros	CRY SOL	FoXS	CRY SOL	FoXS	CRY SOL	FoXS
Rg (Guinier) (nm)	6,3	6,3	7,0	7,0	7,3	7,3
Rg ($p(r)$) (nm)	6,5	6,4	7,0	7,0	7,4	7,4
Dmáx (nm)	24,3	24,4	25,1	25,0	25,4	25,3

4.2. Simulação das curvas de SAXS e cálculos dos parâmetros estruturais das variantes de preocupação da proteína Spike

As conformações mais predominantes do ectodomínio da proteína Spike do vírus SARS-CoV-2 quando está ativo são os da Spike com uma região RBD aberta (16%) e da Spike com uma única região RBD aberta ligada a uma única ACE2 (28%) (Benton, *et. al.* 2020). Desse modo, por conta dos predomínios dessas conformações, foi realizada uma análise de SAXS das variantes de preocupação, dessas duas conformações, junto a um estudo do complexo formado entre a região RBD isolada da Spike e o receptor ACE2, por conta de sua relevância quando se trata da infecção do vírus na célula.

4.2.1. Proteína Spike com uma região RBD aberta

Nas Figuras 22 e 23, são apresentadas as curvas simuladas nos programas FoXS e CRY SOL, respectivamente, e as $p(r)$ geradas a partir dessas curvas, sendo elas relacionadas às estruturas de código 6VYB (Nativa), 7R14 (Alfa), 8DLL (Beta), 8DLO (Gamma) e 7WG9 (Delta). Já na Tabela 6 é possível observar os parâmetros estruturais teóricos, R_g e $D_{\text{máx}}$, obtidos através das curvas e $p(r)$ geradas.

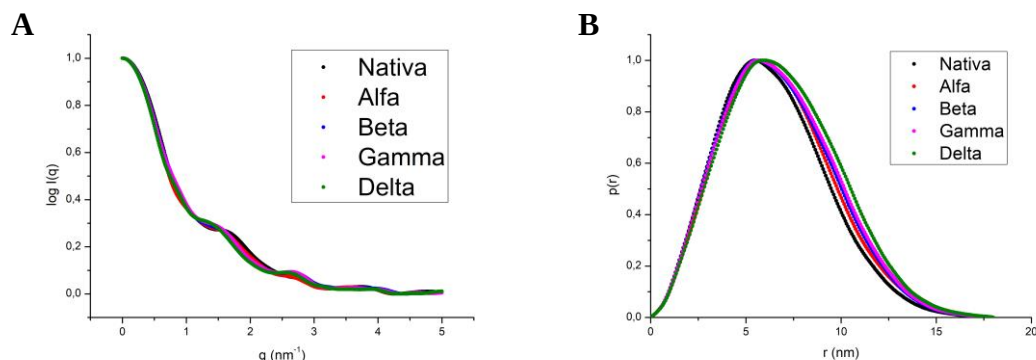


Figura 22: (A) Curvas simuladas no programa FoXS das estruturas proteicas da região completa do domínio extracelular da Spike isolada com uma região RBD aberta de códigos 6VYB, 7R14, 8DLL, 8DLO e 7WG9. (B) $p(r)$ geradas a partir das curvas simuladas.

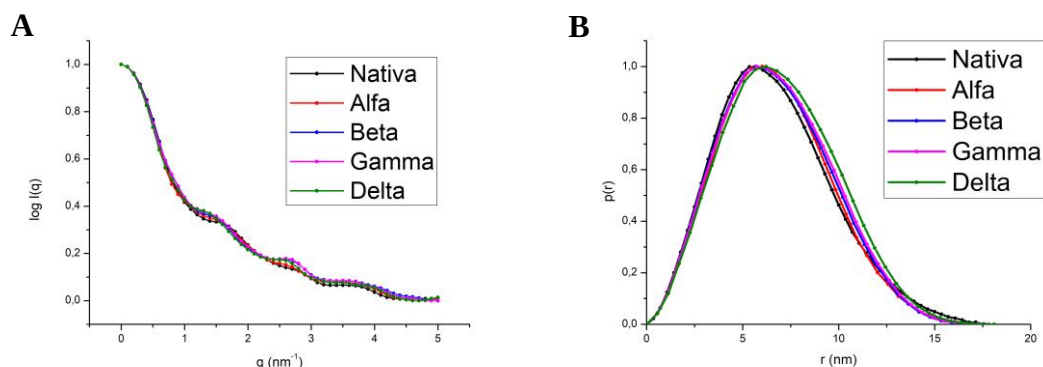


Figura 23: (A) Curvas simuladas no programa CRY SOL das estruturas proteicas da região completa do domínio extracelular da Spike isolada com uma região RBD aberta de códigos 6VYB, 7R14, 8DLL, 8DLO e 7WG9. (B) $p(r)$ geradas a partir das curvas simuladas.

Tabela 6: Parâmetros estruturais calculados das estruturas proteicas de código 6VYB, 7R14, 8DLL, 8DLO e 7WG9 a partir de suas curvas simuladas e de suas $p(r)$ geradas.

Programas	Parâmetros CRY SOL			Parâmetros FoXS		
Códigos PDB	R_g (Guinier) (nm)	R_g ($p(r)$) (nm)	$D_{\text{máx}}$ (nm)	R_g (Guinier) (nm)	R_g ($p(r)$) (nm)	$D_{\text{máx}}$ (nm)
Nativa (6VYB)	5	5,2	17,5	4,9	5	17,5
Alfa (7R14)	5,2	5,2	17,5	5,1	5,1	17,5
Beta (8DLL)	5,2	5,2	17,6	5,1	5,2	17,8
Gamma (8DLO)	5,3	5,3	17,8	5,1	5,2	17,9
Delta (7WG9)	5,4	5,4	18,1	5,3	5,4	18

4.2.2. Complexo formado entre o receptor ACE2 e a região RBD da Spike

Nas Figuras 24 e 25, são mostradas as curvas simuladas nos programas FoXS e CRY SOL, respectivamente, e as $p(r)$ geradas a partir dessas curvas, sendo elas referentes às estruturas de código 6M0J (Nativa), 8DLK (Alfa), 8DLN (Beta), 8DLQ (Gamma) e

7TEW (Delta). Já na Tabela 7 é possível observar os parâmetros estruturais teóricos, R_g e $D_{máx}$, obtidos através das curvas e $p(r)$ geradas.

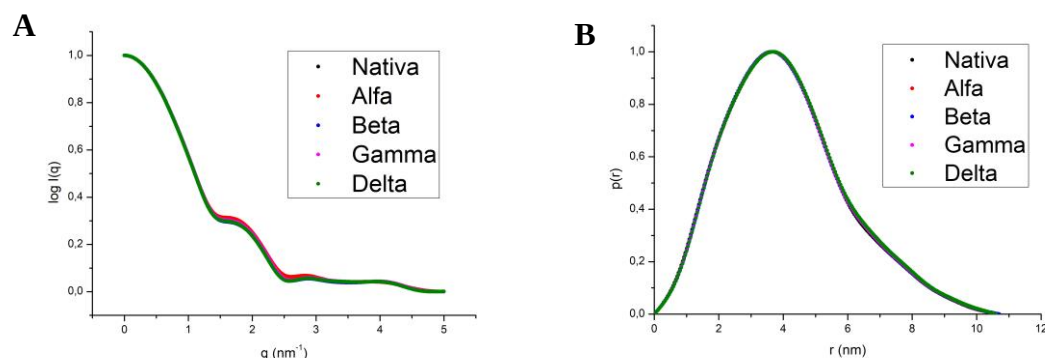


Figura 24: (A) Curvas simuladas no programa FoXS das estruturas proteicas do complexo formado entre a enzima ACE2 e a região RBD da Spike de códigos 6M0J, 8DLK, 8DLNN, 8DLQ e 7TEW. (B) $p(r)$ geradas a partir das curvas simuladas.

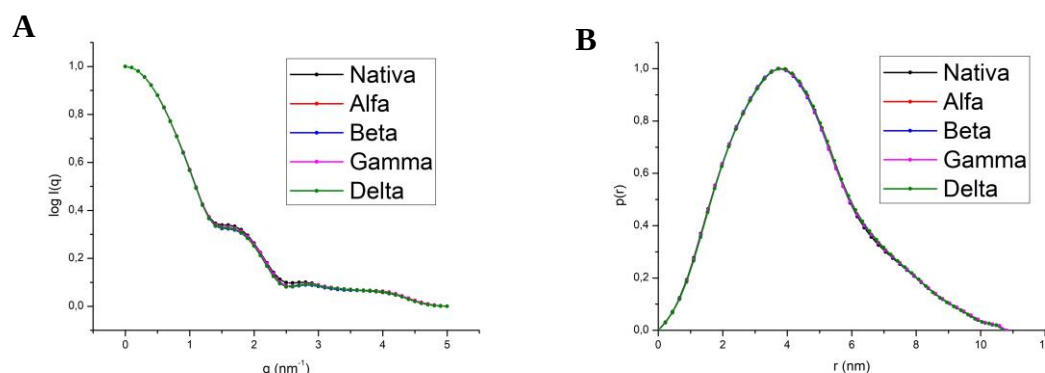


Figura 25: (A) Curvas simuladas no programa CRY SOL das estruturas proteicas do complexo formado entre a enzima ACE2 e a região RBD da Spike de códigos 6M0J, 8DLK, 8DLNN, 8DLQ e 7TEW. (B) $p(r)$ geradas a partir das curvas simuladas.

Tabela 7: Parâmetros estruturais calculados das estruturas proteicas de código 6M0J, 8DLK, 8DLNN, 8DLQ e 7TEW a partir de suas curvas simuladas e de suas $p(r)$ geradas.

Programas	Parâmetros CRY SOL			Parâmetros FoXS		
Códigos PDB	Rg (Guinier) (nm)	Rg (p(r)) (nm)	Dmáx (nm)	Rg (Guinier) (nm)	Rg (p(r)) (nm)	Dmáx (nm)
Nativa (6M0J)	3,2	3,3	10,8	3,1	3,2	10,7
Alfa (8DLK)	3,2	3,3	10,8	3,1	3,2	10,7
Beta (8DLN)	3,2	3,3	10,8	3,1	3,2	10,7
Gamma (8DLQ)	3,2	3,3	10,8	3,1	3,2	10,6
Delta (7TEW)	3,2	3,3	10,7	3,1	3,2	10,6

4.2.3. Spike com uma região RBD aberta ligada a uma ACE2

Nas Figuras 26 e 27, são exibidas as curvas simuladas nos programas FoXS e CRY SOL, respectivamente, e as $p(r)$ geradas a partir dessas curvas, sendo elas relacionadas às estruturas de código 7A94 (Nativa), 8DLJ (Alfa), 7VXD (Beta) e 7W98 (Delta). Já na Tabela 8 é possível observar os parâmetros estruturais teóricos, R_g e $D_{máx}$, obtidos através das curvas e $p(r)$ geradas.

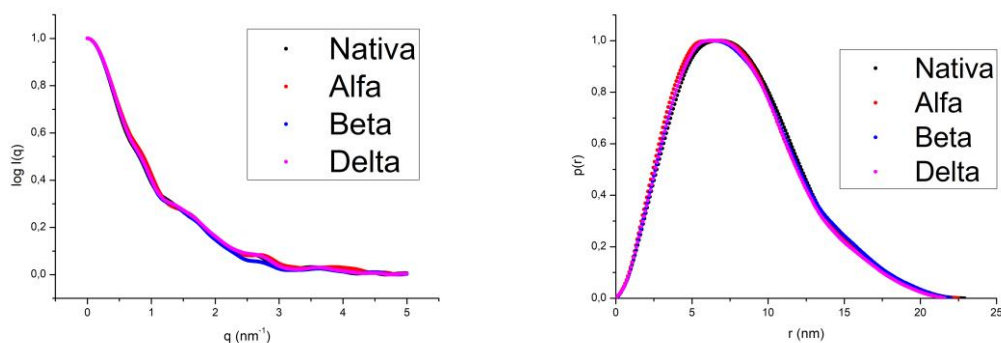


Figura 26: (A) Curvas simuladas no programa FoXS das estruturas proteicas da região completa do domínio extracelular da Spike isolada com uma região RBD aberta ligada a uma ACE2 de código 7A94, 8DLJ, 7VXD e 7W98. (B) $p(r)$ geradas a partir das curvas simuladas.

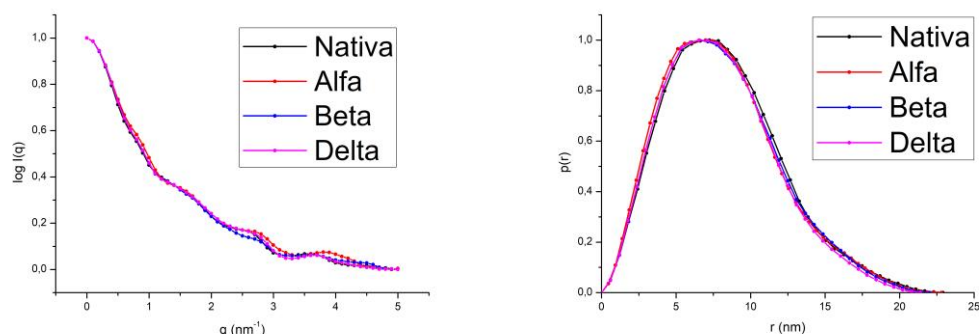


Figura 27: (A) Curvas simuladas no programa CRY SOL das estruturas proteicas da região completa do domínio extracelular da Spike isolada com uma região RBD aberta ligada a uma ACE2 de código 7A94, 8DLJ, 7VXD e 7W98. (B) $p(r)$ geradas a partir das curvas simuladas.

Tabela 8: Parâmetros estruturais calculados das estruturas proteicas de código 7A94, 8DLJ, 7VXD e 7W98 a partir de suas curvas simuladas e de suas $p(r)$ geradas.

Programas		Parâmetros CRY SOL			Parâmetros FoXS		
Códigos PDB	Rg (Guinier) (nm)	Rg (p(r)) (nm)	Dmáx (nm)	Rg (Guinier) (nm)	Rg (p(r)) (nm)	Dmáx (nm)	
Nativa (7A94)	6,3	6,4	22,9	6,3	6,4	22,9	
Alfa (8DLJ)	6,2	6,3	22,8	6,2	6,3	22,5	
Beta (7VXD)	6,3	6,3	22	6,3	6,3	22	
Gamma*	-	-	-	-	-	-	
Delta (7W98)	6,2	6,2	21,4	6,2	6,3	21,6	

*Não existe estrutura da variante Gamma da Spike com uma RBD aberta ligada a uma ACE2 depositada no PDB.

4.3. Interface das estruturas proteicas

As interações entre proteínas desempenham papéis cruciais em muitas funções biológicas. Analisar as características dessas interfaces é essencial para encontrar fatores que contribuam para a afinidade e especificidade das interações entre as proteínas. Uma abordagem geral para estudar a interação entre duas proteínas é obter uma estrutura cristalina no banco de dados PDB e então investigar as propriedades atômicas a partir dos aminoácidos de sua interface.

4.3.1. Interface de trimerização do domínio extracelular da Spike isolada

Usando o programa PyMOL é possível observar a partir das estruturas proteicas de códigos 6VXX e 6VYB que a interface de trimerização da proteína Spike está localizada na Hélice Central (Resíduos 985 a 1035) (Walls, *et. al.* 2020), no subdomínio S2 da estrutura trimérica, em que, na interface de conexão entre seus monômeros existem 3 ligações de hidrogênio, sendo elas as interações D994-R995, Q1005-T1006 e E1017-R1019 mostradas na Figura 29.

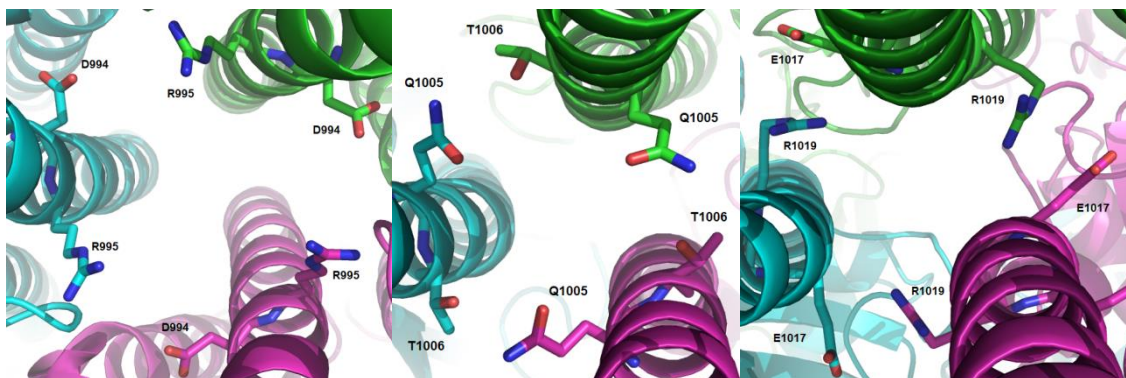


Figura 28: Interface de trimerização do domínio extracelular da proteína Spike isolada. As estruturas são mostradas como sticks e cartoon.

4.3.2. Complexo formado entre a enzima ACE2 e a região RBD da Spike Nativa

Utilizando o programa PyMOL é possível notar a partir da estrutura proteica de código 6M0J que existem 15 ligações de hidrogênio e 2 pontes salinas na interface SARS-CoV-2 RBD e ACE2, sendo elas mostradas na Tabela 9 e na Figura 28.

Tabela 9: Ligações de hidrogênio e pontes salinas nas interfaces SARS-CoV-2 RBD–ACE2.

Interações	SARS-CoV-2 RBD	ACE2
Ligações de Hidrogênio	A475	S19
	N487	Q24
	K417	D30
	Q493	K31
	Y505	E37
	Y449	D38
	T500	Y41
	N501	Y41
	Q498	Q42
	Y449	Q42
	Y489	Y83
	N487	Y83
	G496	K353
	G502	K353
	Y505	R393
Pontes Salinas	K417	D30
	K417	D30

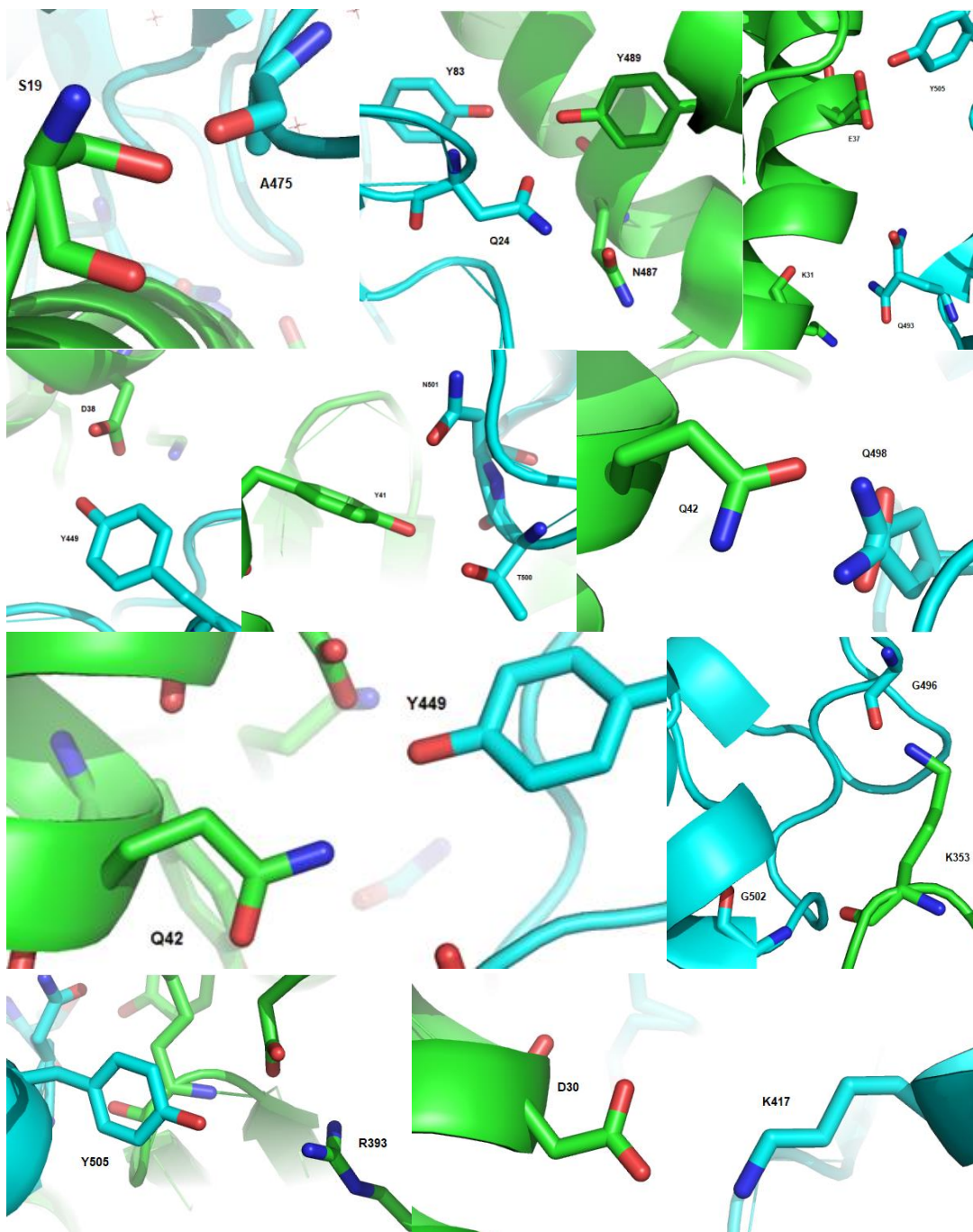


Figura 29: Ligações de hidrogênio e pontes salinas nas interfaces da região RBD (azul) ligada a ACE2 (verde). As estruturas são mostradas como sticks e cartoon.

4.3.3. Mutações nas variantes de preocupação e suas interfaces de formação do complexo SARS-CoV-2 RBD e ACE2

Todos os vírus se modificam com o tempo, ou seja, quanto mais oportunidades um vírus tiver de se espalhar, maior a chance dele se replicar e mudar. A mutação é algo evolutivo e natural, ainda mais se o vírus em questão for um vírus de RNA (Ácido Ribonucleico) em seu material genético, como é o caso do SARS-CoV-2, em que suas mutações resultam na substituição de aminoácidos em posições-chave de sua estrutura proteica culminando em consequências nocivas para o desempenho de suas funções.

Na Tabela 10 são elucidadas as mutações na proteína Spike das variantes Alfa (B.1.1.7), Beta (B.1.351), P.1 (Gamma) e Delta (B.1.617.2) do SARS-CoV-2.

Tabela 10: Mutações na proteína Spike das variantes Alfa (B.1.1.7), Beta (B.1.351), P.1 (Gamma) e Delta (B.1.617.2) do SARS-CoV-2. Mutações sem preenchimento representam as não compartilhadas entre as cepas, em Azul são as na região RBD da proteína Spike, em amarelo a mutação compartilhada entre todas as cepas, em verde compartilhada apenas entre as variantes Beta e Gamma. Mutações com letra vermelha representam as compartilhadas entre Alfa, Beta e Gamma, com letra laranja apenas entre Beta e Gamma, já as mutações com diferentes substituições, porém no mesmo resíduo, são representados em branco para as variantes Beta e Gamma e em verde para as variantes Alfa e Delta.

B.1.1.7	B.1.351	P.1	B.1.617.2
ΔH 69	L 18 F	L 18 F	T 19 R
ΔV 70	D 80 A	T 20 N	ΔE 156
ΔS 144	D 215 G	P 26 S	ΔF 157
-	ΔL 241	D 80 A	R 158 G
-	ΔL 242	D 138 Y	L 452 R
-	ΔA 243	R 190 S	T 478 K
-	K 417 N	K 417 T	-
-	E 484 K	E 484 K	-
N 501 Y	N 501 Y	N 501 Y	-
A 570 D	-	-	-
D 614 G	D 614 G	D 614 G	D 614 G
P 681 H	A 701 V	H 655 Y	P 681 R
T 716 I	-	T 1027 I	D 950 N
S 982 A	-	V 1176 F	-
D 1118 H	-	-	-

Na Tabela 11 são mostradas as interações entre os aminoácidos da interface da região RBD ligada ACE2 da proteína nativa e das variantes de preocupação geradas no programa PISA a partir da estrutura de código 6M0J (Nativa), 8DLK (Alfa), 8DLN (Beta), 8DLQ (Gamma) e 7TEW (Delta).

Tabela 11: Ligações de hidrogênio e pontes salinas nas interfaces de interação da região RBD (Estrutura 1), da proteína Spike Nativa e das variantes Alfa, Beta, Gamma e Delta do SARS-CoV-2, com o receptor humano ACE2 (Estrutura 2). A coluna Pos. mostra a posição das interações, já a distância entre os átomos dessas interações é dada em Å. *Interações sem cores preenchidas representam que todas as cepas às compartilham, já as coloridas representam apenas as específicas compartilhadas entre elas. Fontes vermelhas representam as mutações na região RBD das variantes que interagem com a ACE2.*

Nativa	Pos.	Estrutura 1	Distância [Å]	Estrutura 2
Ligações de Hidrogênio	1	B:LYS 417[NZ]	2.86	A:ASP 30[OD1]
	2	B:TYR 449[OH]	2.83	A:ASP 38[OD1]
	3	B:TYR 449[OH]	3.09	A:GLN 42[OE1]
	4	B:ALA 475[O]	2.81	A:SER 19[OG]
	5	B:ASN 487[OD1]	2.87	A:GLN 24[NE2]
	6	B:ASN 487[ND2]	2.82	A:TYR 83[OH]
	7	B:TYR 489[OH]	3.54	A:TYR 83[OH]
	8	B:GLN 493[OE1]	3.45	A:LYS 31[NZ]
	9	B:GLY 496[O]	3.22	A:LYS 353[NZ]
	10	B:GLN 498[NE2]	2.57	A:GLN 42[OE1]
	11	B:THR 500[OG1]	2.73	A:TYR 41[OH]
	12	B:ASN 501[N]	3.85	A:TYR 41[OH]
	13	B:GLY 502[N]	2.79	A:LYS 353[O]
	14	B:TYR 505[OH]	3.66	A:GLU 37[OE1]
	15	B:TYR 505[OH]	3.53	A:ARG 393[NH2]
Pontes Salinas	1	B:LYS 417[NZ]	3.76	A:ASP 30[OD1]
	2	B:LYS 417[NZ]	2.86	A:ASP 30[OD2]

Alfa	Pos.	Estrutura 1	Distância [Å]	Estrutura 2
Ligações de Hidrogênio	1	B:LYS 417[NE]	3.20	A:ASP 30[OD2]
	2	B:TYR 449[OH]	3.61	A:ASP 38[OD1]
	3	B:TYR 449[OH]	3.02	A:GLN 42[OE1]
	4	B:TYR 453[OH]	3.24	A:HIS 34[ND1]
	5	B:ALA 475[O]	3.12	A:SER 19[N]
	6	B:ASN 487[OD1]	2.47	A:TYR 83[OH]
	7	B:TYR 489[OH]	3.28	A:TYR 83[OH]
	8	B:GLN 493[OE1]	3.45	A:LYS 31[NZ]
	9	B:GLN 498[NE2]	2.57	A:GLN 42[OE1]
	10	B:THR 500[OG1]	2.60	A:TYR 41[OH]
	11	B:TYR 501[OH]	3.66	A:TYR 41[OH]
	12	B:GLY 502[N]	3.30	A:LYS 353[O]
	13	B:TYR 505[OH]	3.08	A:GLU 37[OE2]
	14	B:TYR 505[OH]	3.77	A:ARG 393[NH2]
Pontes Salinas	1	B:LYS 417[NZ]	3.52	A:ASP 30[OD1]
	2	B:LYS 417[NE]	2.56	A:ASP 30[OD2]

Gamma	Pos.	Estrutura 1	Distância [Å]	Estrutura 2
Ligações de Hidrogênio	1	B:TYR 449[OH]	3.52	A:ASP 38[OD1]
	2	B:TYR 449[OH]	3.11	A:GLN 42[NE2]
	3	B:TYR 453[OH]	3.12	A:HIS 34[ND1]
	4	B:ALA 475[O]	3.43	A:SER 19[N]
	5	B:ASN 487[ND2]	3.00	A:TYR 83[OH]
	6	B:TYR 489[OH]	3.59	A:TYR 83[OH]
	7	B:GLN 493[OE1]	2.40	A:LYS 31[NZ]
	8	B:GLN 498[NE2]	2.97	A:GLN 42[OE1]
	9	B:THR 500[O]	3.79	A:ASN 330[ND2]
	10	B:THR 500[OG1]	2.74	A:TYR 41[OH]
	11	B:TYR 501[N]	3.64	A:TYR 41[OH]
	12	B:GLY 502[N]	2.90	A:LYS 353[O]
	13	B:TYR 505[OH]	3.81	A:GLU 37[OE1]
	14	B:TYR 505[OH]	3.87	A:ARG 393[NH2]

Beta	Pos.	Estrutura 1	Distância [Å]	Estrutura 2
Ligações de Hidrogênio	1	B:TYR 449[OH]	3.55	A:ASP 38[OD1]
	2	B:TYR 449[OH]	3.15	A:GLN 42[NE2]
	3	B:TYR 453[OH]	3.04	A:HIS 34[ND1]
	4	B:ALA 475[O]	3.32	A:SER 19[N]
	5	B:ASN 487[OD1]	2.24	A:TYR 83[OH]
	6	B:TYR 489[OH]	3.74	A:TYR 83[OH]
	7	B:GLN 493[OE1]	2.40	A:LYS 31[NZ]
	8	B:GLN 498[OE1]	2.51	A:GLN 42[NE2]
	9	B:THR 500[O]	2.90	A:ASN 330[ND2]
	10	B:THR 500[OG1]	2.26	A:TYR 41[OH]
	11	B:TYR 501[OH]	3.67	A:TYR 41[OH]
	12	B:GLY 502[N]	3.08	A:LYS 353[O]
	13	B:TYR 505[OH]	3.22	A:GLU 37[OE2]
	14	B:TYR 505[OH]	3.49	A:ARG 393[NH2]
Delta	Pos.	Estrutura 1	Distância [Å]	Estrutura 2
Ligações de Hidrogênio	1	B:LYS 417[NE]	2.76	A:ASP 30[OD2]
	2	B:GLY 446[O]	3.07	A:GLN 42[NE2]
	3	B:TYR 449[OH]	2.66	A:ASP 38[OD1]
	4	B:TYR 449[OH]	3.04	A:GLN 42[NE2]
	5	B:ALA 475[O]	3.42	A:SER 19[N]
	6	B:ALA 475[O]	3.76	A:GLN 24[NE2]
	7	B:ASN 487[OD1]	3.82	A:GLN 24[NE2]
	8	B:ASN 487[OD1]	2.37	A:TYR 83[OH]
	9	B:TYR 489[OH]	3.38	A:TYR 83[OH]
	10	B:GLN 493[OE1]	2.65	A:LYS 31[NZ]
	11	B:GLY 496[O]	2.65	A:LYS 353[NZ]
	12	B:GLN 498[NE2]	3.10	A:GLN 42[OE1]
	13	B:THR 500[OG1]	2.68	A:TYR 41[OH]
	14	B:ASN 501[N]	3.70	A:TYR 41[OH]
	15	B:GLY 502[N]	2.55	A:LYS 353[O]
	16	B:TYR 505[OH]	2.70	A:GLU 37[OE2]
	17	B:TYR 505[OH]	3.79	A:ARG 393[NH2]
Pontes Salinas	1	B:LYS 417[NE]	3.45	A:ASP 30[OD1]
	2	B:LYS 417[NE]	2.76	A:ASP 30[OD2]

5. DISCUSSÃO

Junto a um estudo do programa FoXS e dos programas do pacote ATSAS foi feita uma análise dos parâmetros estruturais da região RBD da Spike isolada em solução.

Observando os dados da Tabela 1 se verifica que os valores de R_g obtidos utilizando o método de Guinier (Itri, *et. al.* 2015) coincidem com os valores obtidos através da $p(r)$. Além do mais, ao analisar os parâmetros estruturais calculados a partir da curva experimental se observa que, quando comparados aos das curvas simuladas, além de valores de R_g e $D_{\text{máx}}$ muito próximos entre si, possuem erros percentuais idênticos e muito baixos.

Ao comparar as $p(r)$ exibidas na Figura 7 é possível notar uma semelhança entre suas formas, porém, se nota que a curva experimental possui um perfil muito mais suave do que as curvas simuladas, isso ocorre por conta maior número de pontos existentes (1236 pontos) na sua curva de SAXS do que as curvas geradas no FoXS (256 pontos) e no CRY SOL (51 pontos). Além do mais, ao contrastar as $p(r)$ geradas com as elucidadas na Figura 4, se nota que, os seus máximos estão bem definidos e bastante deslocados para distâncias menores que $D_{\text{máx}}/2$, se assemelhando ao da curva verde relacionada ao cilindro longo mostrada nessa mesma Figura (Svergun, *et. al.* 2003). Ou seja, em razão das $p(r)$ teóricas e experimentais da estrutura da região RBD da Spike isolada apresentarem um formato semelhante, é admissível dizer que essa proteína possui um formato alongado em solução.

Os resultados gerados referentes à estrutura da região RBD isolada em solução (Figura 5) evidenciam que os programas FoXS e CRY SOL, ao compará-los com os resultados obtidos através da curva experimental da estrutura, realizam simulações de SAXS precisas de proteínas em solução a partir de estruturas de alta resolução existentes, permitindo que seja realizado um estudo teórico de estruturas em solução a partir das suas proteínas resolvidas por cristalografia depositadas no PDB. Além disso, observando os ajustes realizados na Figura 6, é possível notar que o valor de χ^2 do programa FoXS ($\chi^2 = 1,1$) é menor que o do CRY SOL ($\chi^2 = 1,3$), mostrando que o FoXS é mais adequado que o CRY SOL para a realização de ajustes da estrutura de alta resolução a sua curva de SAXS experimental, evidenciando que a função χ^2 utilizada no FoXS, elucidada na equação (13), é mais apropriada para ajustar dados de dispersão teóricos aos experimentais do que a utilizada no CRY SOL, mostrada na equação (11).

Ao realizar a sobreposição da estrutura em solução com a estrutura cristalizada da região RBD da Spike isolada, foi possível realizar uma análise comparativa de sua estrutura quando diluída com sua estrutura em cristalografia. Diante dos resultados obtidos na Figura 10, se percebe que as duas estruturas são muito parecidas em ambos os casos, porém, na estrutura em solução é possível observar que foi feita uma maior expressão de aminoácidos, tanto na região C-Terminal (vermelho) quanto na região N-Terminal (azul), da estrutura em relação à mesma quando cristalizada. De outro modo, quando se compara as estruturas isoladas nas Figuras 8 (solução) e 9 (cristalizada) pode-se notar que a estrutura cristalizada expressa uma maior e mais detalhada quantidade de Hélices Alfa e Folhas Beta do que a estrutura diluída. Desse modo, sabendo que as

técnicas de SAXS e cristalografia são muito úteis para a determinação de estruturas proteicas (Svergun, *et. al.* 2003), se nota que, no caso da região RBD isolada, combinando as duas técnicas, é possível obter uma estrutura muito mais completa dessa região. Diante disso, se conclui que a cristalografia permite um melhor detalhamento dos aminoácidos e dos arranjos de suas estruturas secundárias (Folhas Beta e Hélices Alfa), já o SAXS permite uma melhor visualização superficial da estrutura, gerando informações complementares das regiões da proteína que relutam em cristalizar.

Observando as $p(r)$ geradas a partir das curvas simuladas no FoXS e CRY SOL relacionadas às duas conformações, fechada (Figura 12) e aberta (Figura 14), do trímero da Spike, ao complexo SARS-CoV-2 RBD-ACE2 (Figura 16), e às três estruturas complexo formado entre o trímero da proteína Spike e a ACE2 (Figura 21), é possível notar que, as curvas em ambos os programas, para ambas as estruturas, são muito semelhantes e praticamente se sobrepõem entre si. O resultado dessa semelhança entre as curvas se complementa ao observar os parâmetros estruturais exibidos nas Tabelas 2, 3, 4 e 5 onde os resultados de R_g e $D_{máx}$, para todas as estruturas analisadas, são muito próximos entre si quando se compara os dados obtidos, tanto no CRY SOL quanto no FoXS. Além disso, se percebe ao verificar as tabelas que os valores de R_g obtidos utilizando o método de Guinier (Itri, *et. al.* 2015) concordam com os valores obtidos através da $p(r)$.

Ao contrastar as $p(r)$ exibidas nas Figuras 12, 14 e 16 com as mostradas na Figura 4 se observa que, para as conformações fechada e aberta da Spike, e para o complexo SARS-CoV-2 RBD-ACE2, as funções possuem seus picos bastante amplos e deslocados para distâncias menores que $D_{máx}/2$, se parecendo ao da curva amarela relacionada ao oblató mostrado na mesma figura, ou seja, é possível presumir que essas estruturas possuem um formato achatado em solução (Svergun, *et. al.* 2003). Já ao comparar as $p(r)$ exibidas na Figura 21 com as exibidas na Figura 4, para a estrutura 7A94 se observa que, as funções possuem os seus picos bem amplos, porém bastante deslocados para distâncias menores que $D_{máx}/2$, o que dificulta assemelhar essa $p(r)$ a curva verde ou amarela e deduzir a sua forma, logo, é válido supor que em solução a estrutura da região completa do ectodomínio da Spike isolada ligada a uma ACE2 possua um formato achatado, porém levemente alongado (Figura 4) (Svergun, *et. al.* 2003). Além do mais, ao realizar a mesma comparação para as estruturas 7A97 e 7A98, se nota que, além do lento decaimento nas suas curvas de espalhamento, as $p(r)$ de ambas (Figura 21) possuem seus máximos bem definidos e deslocados para distâncias menores que $D_{máx}/2$, porém, as duas apresentam deformações em $r = 5nm$ e não têm características bem definidas, dificultando assimilar essas curvas com alguma das exibidas nas $p(r)$ da Figura 4, sugerindo que ambas as estruturas sejam anisométricas (Svergun, *et. al.* 2003) e não possuam uma forma bem definida em solução.

Os dados teóricos referentes ao R_g e $D_{máx}$ fornecem informações sobre a geometria das proteínas estudadas. Acerca da região inteira do ectodomínio da Spike isolada, comparando os parâmetros estruturais calculados na Tabela 2 (RBD fechado) com a tabela 3 (RBD aberto) é possível verificar que em solução o R_g e o $D_{máx}$ da estrutura com a região RBD aberta é maior do que os da estrutura com a região RBD fechada,

mostrando que o R_g e o $D_{m\acute{a}x}$ dessa estrutura aumentam quando uma de suas regi\~oes RBD \u00e9 aberta, ou seja, essa estrutura se expande quando modifica sua conforma\~ao de fechada para aberta. J\u00e1 em rela\~ao a essa mesma estrutura aberta, por\u00e9m ligadas a enzima ACE2, comparando os par\u00e2metros estruturais calculados na Tabela 5 \u00e9 poss\u00edvel notar que o R_g e o $D_{m\acute{a}x}$ dos complexos aumentam quanto mais ACE2 ligadas a regi\u00e3o RBD da Spike existem, ou seja, se nota que o R_g e $D_{m\acute{a}x}$ dos complexos formados aumentam conforme as regi\u00e3o RBD abrem para se ligarem a ACE2, constatando que existe uma amplia\~ao dessa estrutura quanto mais ACE2 se conectam a regi\u00e3o RBD da Spike.

Uma observa\~ao importante referente \u00e0s $p(r)$ \u00e9 o fato das curvas geradas pelos PDB 7A97 e 7A98 possuirem um formato relativamente semelhante, se sobrepondo em suas regi\u00e3o decrescentes e apresentarem picos mais pr\u00f3ximos entre si, em rela\~ao \u00e0 curva gerada pelo PDB 7A94. J\u00e1 em rela\~ao aos par\u00e2metros estruturais calculados na Tabela 5 se observa que o aumento no R_g e $D_{m\acute{a}x}$ \u00e9 muito maior, cerca de 0,6nm, quando se compara a estrutura de c\u00f3digo 7A94 com a de c\u00f3digo 7A97, em rela\~ao ao aumento dos par\u00e2metros quando se compara a prote\u00edna de c\u00f3digo 7A97 com a de c\u00f3digo 7A98, sendo ele cerca de 0,3nm, ou seja, de metade do aumento quando se compara entre as estruturas 7A94 e 7A97.

A fun\~ao de distribui\~ao de pares $p(r)$ e os par\u00e2metros estruturais, R_g e $D_{m\acute{a}x}$, relacionados \u00e0s estruturas estudadas das variantes de preocupa\~ao do SARS-CoV-2 permitem uma an\u00e1lise importante acerca das suas diferen\~as estruturais em solu\~ao, em que, ao analisar dados das Tabelas 6, 7 e 8, se nota que os valores de R_g obtidos utilizando o m\u00e9todo de Guinier (Itri, *et. al.* 2015) coincidem com os valores obtidos atrav\u00e9s da $p(r)$, tanto no CRY SOL, quanto no FoXS. Assim sendo, quando se examina o perfil de suas $p(r)$ (Figuras 22, 23, 24, 25, 26 e 27), percebe-se que o formato de todas as curvas para todas as estruturas \u00e9 igual ao da prote\u00edna nativa, ou seja, se presume que as muta\~oes nos amino\u00e1cidos n\u00e3o influenciam na forma da prote\u00edna em solu\~ao, assim sendo, a conforma\~ao aberta da regi\u00e3o completa do ectodom\u00ednio da Spike isolada, o complexo SARS-CoV-2 RBD e ACE2 e estrutura completa do ectodom\u00ednio da Spike isolada ligada a uma ACE2 conservam os seus formatos independentes da sua variante.

Analisando as $p(r)$ (Figuras 22 e 23) referentes \u00e0 conforma\~ao da prote\u00edna Spike, em sua forma nativa e de suas variantes, com uma \u00fanica regi\u00e3o RBD aberta, pode-se notar que o pico das curvas das variantes est\u00e1 deslocado para a direita em rela\~ao \u00e0 nativa, evidenciando um aumento do R_g das variantes nessa conforma\~ao. Tamb\u00e9m \u00e9 poss\u00edvel notar essa informa\~ao na Tabela 6, onde se verifica, no CRY SOL e no FoXS, um aumento do R_g e do $D_{m\acute{a}x}$ das variantes em rela\~ao a nativa, em que, as estruturas da tabela est\u00e3o dispostas em ordem crescente em rela\~ao aos valores dos seus par\u00e2metros, sendo essa ordem, da menor para a maior, dada por: Nativa, Alfa, Beta, Gamma, Delta. Ao observar as $p(r)$ (Figuras 24 e 25) associadas ao complexo formado entre a regi\u00e3o RBD da Spike, em sua forma nativa e suas variantes, ligado a ACE2, pode-se notar que todas as curvas se sobrep\u00f5em e possuem pontos de pico semelhantes, apontando um R_g constante para todas as prote\u00ednas. Assim como, \u00e9 poss\u00edvel notar essa informa\~ao na tabela 7, onde se verifica, em ambos os programas, o R_g e $D_{m\acute{a}x}$

semelhante para todas as formas da proteína, se observando apenas uma pequena diminuição no $D_{máx}$ da variante Delta em relação as outras. Ao verificar as $p(r)$ (Figuras 26 e 27), referentes à conformação da proteína Spike, em sua forma nativa e de suas variantes, com uma única região RBD aberta ligada a uma ACE2, pode-se notar que os picos das curvas das variantes estão deslocados para a esquerda em relação à nativa, evidenciando uma diminuição do R_g das variantes nessa conformação. Também é possível observar essa informação na Tabela 8, onde se percebe, nos dois programas utilizados, uma diminuição do R_g e do $D_{máx}$ das variantes em relação a nativa, em que, as estruturas da tabela estão dispostas em ordem decrescente em relação aos valores dos seus parâmetros, sendo essa ordem, da maior para a menor, dada por: Nativa, Alfa, Beta, Delta. Diante disso, as variações presente no R_g e no $D_{máx}$ das variantes comparados a proteína em sua forma nativa sugerem que as mutações são responsáveis por alterar os parâmetros geométricos nessas estruturas, levando a crer que essas alterações nas variantes influenciam no aumento da afinidade da região RBD da Spike com a ACE2, facilitando a formação do complexo SARS-CoV-2 RBD e ACE2, consequentemente favorecendo a entrada do vírus da célula humana, permitindo que as variantes de preocupação sejam mais perigosas e transmissíveis que o vírus em sua forma original (Banho CA, *et. al.* 2022) (Davies NG, *et. al.* 2021) (Li, *et. al.* 2022).

As mutações nos aminoácidos do ectodomínio da proteína Spike do SARS-CoV-2 resultam diretamente em alterações nas suas características funcionais, principalmente quando ocorrem em sua região RBD que é o foco principal deste estudo.

Ao analisar a tabela 10, relativa às mutações nos aminoácidos das variantes de preocupação da proteína Spike, se observa que, todas as variantes possuem mutações tanto na região RBD (resíduos 329 a 544), quanto nas outras regiões da Spike, porém não existem mutações que façam parte de sua interface de trimerização (Figura 28). As variantes Beta e Gamma são as que mais compartilham mutações entre si, sendo a mutação N501Y (letra vermelha) presente em todas as variantes, exceto na Delta, a única variante que não compartilha nenhuma mutação com as outras na região RBD. Desse modo, ao verificar as interações (Tabela 11) da região RBD da proteína Spike, em sua forma nativa e de suas variantes, com a enzima ACE2, primeiramente se nota que, a única mutação na região RBD (Tabela 10) responsável pela interação com a ACE2 é a substituição no resíduo N501Y. Essa substituição, presente nas cepas Alfa, Beta e Gamma, indica um aumento na força de interação da região RBD com a ACE2 das variantes em comparação a forma nativa, causada pela diminuição na distância de interação (letra vermelha) entre o resíduo Y501(Tyr) da RBD com o Y41(Tyr) do receptor humano, em relação à interação N501(Asn)---Y41(Tyr) existente na nativa. Do mesmo modo, o aminoácido Tyr, por ser menos flexível que o resíduo Asn (Huang, *et. al.* 2003), gera um aumento da rigidez na região RBD da Spike. Além disso, a mutação N501Y (Gobeil, *et. al.* 2021) causa uma alteração no espaçamento entre os resíduos da região RBD, ou seja, é possível que ela seja responsável pela nova interação Y453---H34 e pelas ausências das interações N487(Asn)---Q24(Gln) e G496(Gly)---K353(Lys) nas três variantes, quando comparado à forma nativa da proteína. Assim sendo, a interação dessas cepas com a ACE2 passa a depender de um aminoácido rígido (Y453)

e não depende mais de seus resíduos flexíveis (N487 e G496) (Huang, *et. al.* 2003). Diante disso, se conclui que a mutação N501Y é responsável por gerar uma maior afinidade das variantes de preocupação com a ACE2 e por diminuir a flexibilidade das regiões RBD delas, em que, esse aumento de rigidez na região dificulta que anticorpos se liguem a proteína e impeçam a entrada do vírus SARS-CoV-2 na célula humana, aumentando a transmissibilidade e o escape do sistema imunológico dessas cepas em relação à proteína nativa.

A substituição E484K (letra laranja), observada na Tabela 10, se encontra apenas nas variantes Beta e Gamma, em que, nelas também existe uma mutação no resíduo K417 (letra Branca) também elucidada na Tabela 10, porém com uma substituição diferente, sendo K417N na Beta e K417T na Gamma. A mutação E484K consiste na substituição de um aminoácido de cadeia lateral polar negativa E(Glu) por um de cadeia lateral polar positiva K(Lys), em que, apesar de não interagir diretamente com a ACE2, esse novo resíduo pode aumentar previsivelmente a ligação com o receptor (Han, *et.al.* 2021), culminando numa maior afinidade entre ele com a região RBD. Já as mutações K417N (Beta) e K417T (Gamma), são mutações que um aminoácido de cadeia lateral positiva K(Lys), em ambos os casos, é substituído respectivamente por um de cadeia lateral polar não carregada, N(Asn) e T(Thr), sendo os dois mais flexíveis que o aminoácido K(Lys) (Huang, *et. al.* 2003). Observando a Tabela 11 se nota que essas mutações destruíram as pontes de hidrogênio e, principalmente, as fortes interações iônicas realizadas pelas pontes salinas que existiam entre o resíduo K(Lys) da região RBD com o aminoácido de cadeia lateral polar negativa D(Asp) na ACE2, reduzindo drasticamente o número de interações entre a proteína Spike dessas variantes com receptor humano ACE2, consequentemente diminuindo a afinidade entre eles, tornando as variantes Beta e Gamma aquelas que possuem menos interações com a ACE2 entre as mencionadas neste trabalho. O resíduo N417(Asn) da Beta sofre uma N-Glicosilação, em que, esse processo é responsável por aumentar o escape do vírus do sistema imune, já o resíduo T417(Thr) da Gamma, por não sofrer nenhuma glicosilação, permite que a RBD dessa cepa possua uma maior afinidade com a ACE2 em relação a variante Gamma. Além disso, as cepas Beta e Gamma compartilham uma nova interação entre si, T500---N330, sendo essa uma interação exclusiva dessas variantes. Sendo assim, apesar da nova interação e da mutação E484K, a afinidade entre a região de formação do complexo SARS-CoV-2 RBD e ACE2 das variantes Beta e Gamma é menor do que da Alfa, mostrando que ambas as variantes tem uma transmissibilidade menor do que a surgida no Reino Unido (Alfa) (Banho CA, *et. al.* 2022) (Davies NG, *et. al.* 2021). Entretanto, por conta do processo de glicosilação sofrido em seu aminoácido mutado, a cepa Beta possui um escape imunológico muito maior do que a Alfa. Já em relação à Gamma, pode-se dizer que ela está no meio termo, é mais transmissível que a Beta, porém não tanto quanto Alfa, e escapa mais do sistema imune que a Alfa, porém não tanto quanto a Beta.

A variante Delta, além de manter todas as interações que a forma original da região RBD realiza com a ACE2, apresenta duas novas, G446---Q42 e A475---Q24 (Tabela 11), sendo assim a cepa que possui mais interações na região RBD com a enzima ACE2 entre as citadas neste trabalho. As substituições L452R e T478K (Tabela 10), presentes

na variante Delta, apesar de não interagirem com a ACE2 são mutações em que ambos os aminoácidos mutados, L(Leu) e T(Thr), são substituídos respectivamente por dois aminoácidos de cadeia lateral com grupo polar positivo R(Arg) e K(Lys), onde, essas cadeias laterais carregadas possibilitam um aumento da interação da região RBD com a ACE2, aumentando a afinidade entre eles. Os aminoácidos R(Arg) e K(Lys) são muito menos flexíveis que os resíduos L(Leu) e T(Thr) (Huang, *et. al.* 2003), além do mais, o aminoácido R(Arg) sofre um processo de N-Glicosilação em sua estrutura, mostrando que os resíduos R(Arg) e K(Lys) também são responsáveis por aumentar potencialmente o escape do vírus do sistema imunológico humano. Além disso, o aminoácido K478(Lys) exibe uma alteração conformacional na região RBD (Wang, Y, *et. al.* 2022), sendo ele o responsável por gerar as novas interações G446(Gly)---Q42(Glu) e A475(Ala)---Q24(Glu), aumentando o número de interações entre o complexo SARS-CoV RBD e ACE2, consequentemente aumentando sua afinidade. Portanto, embora a mutação N501Y não esteja presente na variante Delta, essa cepa possui um número de interações com a ACE2 maior que o da proteína na forma nativa e de todas as outras variantes de preocupação, além disso, possui uma região RBD altamente carregada e rígida, permitindo uma enorme afinidade entre a região RBD e a ACE2. Isso revela o porquê dessa variante, além ser a mais transmissível, é a que tem o maior escape do sistema imune entre todas as variantes citadas, além disso, possui uma carga viral mil vezes maior que a cepa original do vírus (Li, *et. al.* 2022).

Verificando as mutações nas outras regiões da Spike, além da RBD, elucidadas na Tabela 10, nota-se que a substituição D614G (amarelo) está presente em todas as variantes, já a mutação L18F (verde) existe apenas nas variantes Beta e Gamma, por fim, se percebe que, entre as variantes Alfa e Delta existe uma mutação no mesmo resíduo (letra verde), porém com uma substituição diferente, sendo P681H na Alfa e P681R na Delta.

6. CONCLUSÃO

É preciso ressaltar que existe a necessidade de se obter resultados experimentais de SAXS das estruturas estudadas em solução, visando realizar uma melhor análise dessas proteínas e uma comparação mais adequada de suas estruturas quando obtidas por meio de cristalização. Também é necessário evidenciar que é fundamental realizar uma análise de todas as mutações existentes na proteína Spike das variantes além das presentes na região RBD, assim possibilitando um estudo mais concreto da predominância dessas variantes em relação a forma nativa do SARS-CoV-2 e os motivos pelos quais elas são responsáveis pelo aumento de casos na pandemia, aumento de hospitalizações ou mortes, e maior resistência a vacinas e tratamentos.

7. REFERÊNCIAS

Banho CA, *et. al.* Impact of SARS-CoV-2 Gamma lineage introduction and COVID-19 vaccination on the epidemiological landscape of a Brazilian city. **Commun Med.** 2022, v.2, 41.

- Benton DJ, et. al. Receptor binding and priming of the spike protein of SARS-CoV-2 for membrane fusion. **Nature**. 2020, v.588, 327–330.
- Berman HM, et. al. The Protein Data Bank. **Nucleic Acids Research**. 2000, v.28, 235–242.
- Cui Z, et. al. Structural and functional characterizations of infectivity and immune evasion of SARS-CoV-2 Omicron. **Cell**. 2022, v.185(5), 860-871, e13.
- Custódio TF, et. al. Selection, biophysical and structural analysis of synthetic nanobodies that effectively neutralize SARS-CoV-2. **Nat Commun**. 2020, v.11, 5588.
- Davies NG, et. al. Estimated transmissibility and impact of SARS-CoV-2 lineage B.1.1.7 in England. **Science**. 2021, v.372, 6538.
- Glatzer O, et. al. Small angle x-ray scattering. **London: Academic Press A**. 1982, v.39, 500.
- Gobeil SM, et. al. Effect of natural mutations of SARS-CoV-2 on spike structure, conformation, and antigenicity. **Science**. 2021, v.373, 6555.
- Han P, et. al. Molecular insights into receptor binding of recent emerging SARS-CoV-2 variants. **Nat Commun**. 2021, v.12(1), 6103.
- Huang F, et. al. A Conformational Flexibility Scale for Amino Acids in Peptides. **Angewandte Chemie International Edition**. 2003, v.42(20), 2269–2272.
- Itri R, et. al. Espalhamento de raios-X a baixos ângulos (SAXS) aplicado ao estudo de proteínas não interagentes em solução. **Estrategias y avances em el estudio de toxinas de interés para la biomedicina, Benemérita Universidad Autónoma de Puebla**. 2015, v.1, p.171-190, e1.
- Kikhney AG, et. al. A practical guide to small angle X-ray scattering (SAXS) of flexible and intrinsically disordered proteins. **FEBS Letters**. 2015, v.589(19), 2570-2577.
- Kikhney AG, et. al. SASBDB: Towards an automatically curated and validated repository for biological scattering data. **Protein Science**. 2019, v.29(1), 66-75.
- Krissinel E, et. al. Inference of macromolecular assemblies from crystalline state. **J Mol Biol**. 2007, v.372(3), 774-97.
- Lan J, et. al. Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. **Nature**. 2020, v.581(7807), 215-220.
- Li B, et. al. Viral infection and transmission in a large, well-traced outbreak caused by the SARS-CoV-2 Delta variant. **Nat Commun**. 2022, v.13, 460.
- Manalastas-Cantos K, ATSAS 3.0: expanded functionality and new tools for small-angle scattering data analysis. **J Appl Crystallogr**. 2015, v.54, 343-355.

- Mannar D, et. al. SARS-CoV-2 variants of concern: spike protein mutational analysis and epitope for broad neutralization. **Nat Commun.** 2022, v.13, 4696.
- Mariano D. et. al. Dinâmica molecular: como mostrar um filme completo em uma folha de papel?. **BIOINFO – Revista Brasileira de Bioinformática.** 2021, Cap.11, e1.
- Rezende EQF, et. al. Geometria Euclidiana Plana e Construções Geométricas. **Campinas: Editora UNICAMP.** 2000, v.1, 73-80.
- Saville JW, et. al. Structural and biochemical rationale for enhanced spike protein fitness in delta and kappa SARS-CoV-2 variants. **Nat Commun.** 2022, v.13, 742.
- Schneidman-Duhovny D, et. al. FoXS: a web server for rapid computation and fitting of SAXS profiles. **Nucleic Acids Research.** 2010, v.38, W540–W544.
- Schneidman-Duhovny D, et. al. FoXS, FoXSDock and MultiFoXS: Single-state and multi-state structural modeling of proteins and their complexes based on SAXS profiles, **Nucleic Acids Research.** 2016, v.44, W424–W429.
- Shahhosseini N, et. al. Mutation Signatures and In Silico Docking of Novel SARS-CoV-2 Variants of Concern. **Microorganisms.** 2021, v.9(5), 92.
- Svergun DI. Determination of the regularization parameter in indirect-transform methods using perceptual criteria. **J Appl Crystallogr.** 1992, v.25, 495-503.
- Svergun DI, et. al. CRY SOL - a Program to Evaluate X-ray Solution Scattering of Biological Macromolecules from Atomic Coordinates. **J Appl Crystallogr.** 1995, v.28(6), 768-773.
- Svergun DI, et. al. Small-angle scattering studies of biological macromolecules in solution. **Rep Prog Phys.** 2003, v.66(10), 1735–1782.
- Svergun DI, et. al. Small Angle X-Ray and Neutron Scattering from Solutions of Biological macromolecules. **IUCr Texts on Crystallography.** 2013, No.19, Oxford University Press.
- The PyMOL Molecular Graphics System, Version 1.2r3pre, Schrödinger, LLC.
- Walls AC, et. al. Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. **Cell.** 2020, v.181(2), 281-292, e6.
- Wang P, et. al. Increased resistance of SARS-CoV-2 variant P.1 to antibody neutralization. **Cell Host Microbe.** 2021, v.29(5), 747-751, e4.
- Wang Q, et. al. Structural and Functional Basis of SARS-CoV-2 Entry by Using Human ACE2. **Cell.** 2020, v.181(4), 894-904, e9.

Wang Y, et. al. Conformational dynamics of the Beta and Kappa SARS-CoV-2 spike proteins and their complexes with ACE2 receptor revealed by cryo-EM. **Nat Commun.** 2021, v.12, 7345.

Wang, Y, et. al. Structural basis for SARS-CoV-2 Delta variant recognition of ACE2 receptor and broadly neutralizing antibodies. **Nat Commun.** 2022, v.13, 871.

Winn MD, et. al. Overview of the CCP4 suite and current developments. **Acta Crystallogr D Biol Crystallogr.** 2011, v.67(4), 235-42.

Wrobel AG, et. al. Evolution of the SARS-CoV-2 spike protein in the human host. **Nat Commun.** 2022, v.13, 1178.