



UEPB

**UNIVERSIDADE ESTADUAL DA PARAÍBA
CAMPUS DE CAMPINA GRANDE
CENTRO DE CIÊNCIAS BIOLÓGICAS E DA SAÚDE
DEPARTAMENTO DE BIOLOGIA
CURSO DE CIÊNCIAS BIOLÓGICAS**

LUCAS KELLORRAN SILVA SOUZA

**ANÁLISE *IN SILICO* DA PROTEÍNA GALNS COM AS MUTAÇÕES PRESENTES
NOS PACIENTES COM MPS IV-A DO ESTADO DA PARAÍBA**

CAMPINA GRANDE

2022

LUCAS KELLORRAN SILVA SOUZA

**ANÁLISE *IN SILICO* DA PROTEÍNA GALNS COM AS MUTAÇÕES PRESENTES
NOS PACIENTES COM MPS IV-A DO ESTADO DA PARAÍBA**

Trabalho de Conclusão de Curso em Ciências
Biológicas da Universidade Estadual da
Paraíba, como requisito à obtenção do título
de Bacharel em Ciências Biológicas.

Área de concentração: Genética.

Orientadora: Profa. Dra. Simone Silva dos Santos Lopes

CAMPINA GRANDE

2022

É expressamente proibido a comercialização deste documento, tanto na forma impressa como eletrônica. Sua reprodução total ou parcial é permitida exclusivamente para fins acadêmicos e científicos, desde que na reprodução figure a identificação do autor, título, instituição e ano do trabalho.

S729a Souza, Lucas Kellorran Silva.
Análise *in silico* da proteína GALNS com as mutações presentes nos pacientes com MPS-IVA do estado da Paraíba [manuscrito] / Lucas Kellorran Silva Souza. - 2022.
76 p. : il. colorido.

Digitado.
Trabalho de Conclusão de Curso (Graduação em Ciências Biológicas) - Universidade Estadual da Paraíba, Centro de Ciências Biológicas e da Saúde, 2023.
"Orientação : Profa. Dra. Simone Silva dos Santos Lopes, Coordenação de Curso de Biologia - CCBS."

1. Bioinformática. 2. Modelagem por Homologia. 3. Docking Molecular. 4. Mucopolissacaridoses. I. Título

21. ed. CDD 572

LUCAS KELLORRAN SILVA SOUZA

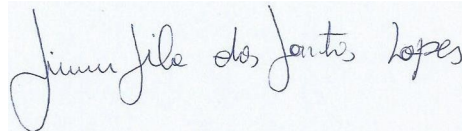
**ANÁLISE *IN SILICO* DA PROTEÍNA GALNS COM AS MUTAÇÕES PRESENTES
NOS PACIENTES COM MPS IV-A DO ESTADO DA PARAÍBA**

Trabalho de Conclusão de Curso em Ciências
Biológicas da Universidade Estadual da
Paraíba, como requisito à obtenção do título
de Bacharel em Ciências Biológicas.

Área de concentração: Genética.

Aprovada em: 28/07/2022.

BANCA EXAMINADORA



Profa. Dra. Simone Silva dos Santos Lopes (Orientador)
Universidade Estadual da Paraíba (UEPB)



Prof. Dr. Carlos Magno dos Santos
Universidade Estadual da Paraíba (UEPB)



Prof. Dr. Sávio Farias Torres
Universidade Federal da Paraíba (UFPB)

RESUMO

ANÁLISE *IN SILICO* DA PROTEÍNA GALNS COM AS MUTAÇÕES PRESENTES NOS PACIENTES COM MPS IV-A DO ESTADO DA PARAÍBA

Lucas Kellorran Silva Souza¹

Simone Silva dos Santos Lopes²

O acúmulo de glicosaminoglicanos (GAGs) não totalmente degradados em lisossomos é uma característica das mucopolissacaridoses (MPSs), um grupo de doenças de armazenamento lisossomal. A MPS tipo IV-A está incluído nesta categoria e é causado por mutações no gene GALNS, que produz um precursor de 522-aminoácidos que é processado em uma enzima oligomérica, a enzima GALNS, que tem sulfato de queratana e sulfato de condroitina como substratos. O primeiro passo para compreender as interações proteína-molécula nos processos celulares é compreender a estrutura tridimensional da proteína, pois a função de um polipeptídeo é determinada por sua estrutura e sequência. A estrutura tridimensional de uma proteína pode ser determinada usando a modelagem por homologia, que usa como referência uma ou mais proteínas homólogas de estrutura terciária conhecida para extrapolar a estrutura tridimensional da proteína-alvo. Assim, o objetivo deste trabalho é se utilizar dessa metodologia para tentar prever a estrutura tridimensional das variantes da enzima GALNS identificadas nos pacientes com MPS IVA do estado da Paraíba, e avaliar se essas abordagens são eficazes em determinar tanto as estruturas, quanto a relação real de cada uma das variantes com os substratos da enzima. Até agora, foram identificadas seis mutações em pacientes do estado da Paraíba. Como o efeito da c.120+1G>A sobre a proteína ainda é desconhecido, ela não foi incluída nas análises. As variantes selecionadas para o estudo foram, portanto, p.D45G, p.G301C, p.R386C, p.S341R, e p.V239F. Os softwares Modeller e I-Tasser foram usados para construir os modelos, que posteriormente foram refinados com o 3D Refine. A qualidade dos modelos 3D gerados foi validada pelo gráfico de Ramachandran e com o programa PROCHECK, e pelo Verify3D. O alinhamento dos modelos mutados com a proteína selvagem foi realizado pelo TM-Align, e a visualização de todas as proteínas foi realizada pelo PyMol. Mesmo prevendo que as proteínas mutantes exibiriam características estruturais distintas, não foi possível detectar tais distinções nos alinhamentos gerados pelo TM-Align. No entanto, usando o docking molecular, observou-se que diferentes variantes interagem com substratos de maneiras diferentes dependendo do modelo, isto é, se havia sido gerado pela I-Tasser ou pela Modeller. Estas descobertas sugerem que, embora não

sejamos capazes de distinguir visualmente as proteínas mutantes e do tipo selvagem com esse processo de modelagem, há mudanças significativas que alteram a forma como o modelo interage com os substratos. Os resultados desse trabalho geram novos questionamentos, como se alguma das técnicas de modelagem foi eficaz para identificar a estrutura tridimensional das variantes do GALNS e se os dockings foram suficientes para desvendar a forma que essas enzimas se relacionam aos substratos. Novas pesquisas são necessárias para responder a estas perguntas.

Palavras-Chave: Bioinformática; Modelagem por homologia; *Docking* molecular; Mucopolissacaridoses.

¹ **Graduando em Ciências Biológicas**

² **Professora efetiva da Universidade Estadual da Paraíba, Doutora em Genética e Biologia Molecular, orientadora do Trabalho de Conclusão de Curso.**

ABSTRACT

***IN SILICO* ANALYSIS OF GALNS PROTEIN WITH MUTATIONS PRESENT IN PATIENTS WITH MPS IV-A IN THE STATE OF PARAÍBA**

Lucas Kellorran Silva Souza¹

Simone Silva dos Santos Lopes²

The accumulation of not fully degraded glycosaminoglycans (GAGs) in lysosomes is a feature of mucopolysaccharidoses (MPSs), a group of lysosomal storage disorders. MPS IVA is included in this category and is caused by mutations in the GALNS gene, which produces a 522-amino acid precursor that is processed into an oligomeric enzyme, the GALNS enzyme, which has keratan sulfate and chondroitin sulfate as substrates. The first step in understanding protein-molecule interactions in cellular processes is to understand the three-dimensional structure of the protein, as the function of a polypeptide is determined by its structure and sequence. The three-dimensional structure of a protein can be determined using homology modeling, which uses one or more homologous proteins of known tertiary structure as a reference to extrapolate the three-dimensional structure of the target protein. Thus, the objective of this work is to use this methodology to try to predict the three-dimensional structure of the GALNS enzyme variants identified in patients with MPS IVA in the state of Paraíba, and to assess whether these approaches are effective in determining both the structures and the real relationship of each of the variants with the enzyme substrates. So far, six mutations have been identified in patients in the state of Paraíba. As the effect of c.120+1G>A on the protein is still unknown, it was not included in the analyses. The variants selected for the study were, therefore, p.D45G, p.G301C, p.R386C, p.S341R, and p.V239F. Modeller and I-Tasser software were used to build the models, which were later refined with 3D Refine. The quality of the 3D models generated was validated by the Ramachandran plot and with the PROCHECK program, and by Verify3D. Alignment of the mutated models with the wild-type protein was performed by TM-Align, and visualization of all proteins was performed by PyMol. Even though predicting that the mutant proteins would exhibit distinct structural characteristics, it was not possible to detect such distinctions in the alignments generated by TM-Align. However, using molecular docking, it was observed that different variants interacted with substrates in different ways depending on the model, that is, whether it had been generated by I-Tasser or Modeller. These findings suggest that while we are not able to visually distinguish mutant and wild-type proteins with this modeling process, there

are significant changes that alter the way the model interacts with substrates. The results of this work raise new questions, such as whether any of the modeling techniques were effective in identifying the three-dimensional structure of the GALNS variants and whether the dockings were sufficient to reveal the way in which these enzymes are related to the substrates. New research is needed to answer these questions.

Key-words: Bioinformatics; Modeling by homology; Molecular docking; Mucopolysaccharidoses.

¹ **Graduating in Biological Sciences**

² **Effective professor at the State University of Paraíba, PhD in Genetics and Molecular Biology, supervisor of the Course Completion Work.**

LISTA DE ILUSTRAÇÕES

Figura 1 - Processo de modelagem por homologia convencional.....	24
Figura 2 - Processo de modelagem por homologia para as mutações da proteína GALNS presentes nos pacientes do estado da Paraíba.....	25
Figura 3 - Variantes da proteína GALNS geradas pelo Modeller.....	30
Figura 4 - Variantes da proteína GALNS geradas pelo I-Tasser.....	31
Figura 5 - Alinhamento entre os modelos da GALNS selvagem (em verde) e da variante p.S341R (em azul).....	34
Figura 6 - Alinhamento entre os modelos da GALNS selvagem (em verde) e da variante p.V239F (em azul).....	34
Figura 7 - Alinhamento entre os modelos da GALNS selvagem (em verde) e da variante p.S341R (em azul) a nível de aminoácido.....	35
Figura 8 - <i>Dockings</i> moleculares dos modelos gerados pelo I-Tasser com o ligante sulfato de queratano.....	36
Figura 9 - <i>Dockings</i> moleculares dos modelos gerados pelo Modeller com o ligante sulfato de queratano.....	36
Figura 10 - <i>Docking</i> molecular da variante D45G gerado pelo I-tasser com o sulfato de queratano.....	37
Figura 11 - <i>Docking</i> molecular da variante D45G gerado pelo Modeller com o sulfato de queratano.....	38
Figura 12 - <i>Docking</i> molecular da variante D45G gerados pelo I-tasser e pelo Modeller, respectivamente, com o sulfato de condroitina.....	39
Figura 13 - <i>Docking</i> molecular da variante G301C gerada pelo I-tasser com o sulfato de queratano.....	39
Figura 14 - <i>Docking</i> molecular da variante G301C gerada pelo Modeller com o sulfato de queratano.....	40

Figura 15 - <i>Docking</i> molecular da variante G301C gerada pelo I-tasser com o sulfato de condroitina.....	40
Figura 16 - <i>Docking</i> molecular da variante G301C gerada pelo Modeller com o sulfato de condroitina.....	41
Figura 17 - <i>Docking</i> molecular da variante R386C gerada pelo I-tasser com o sulfato de queratano.....	42
Figura 18 - <i>Docking</i> molecular da variante R386C gerado pelo Modeller com o sulfato de queratano.....	42
Figura 19 - <i>Docking</i> molecular da variante R386C gerados pelo I-tasser e pelo Modeller, respectivamente, com o sulfato de condroitina.....	43
Figura 20 - <i>Docking</i> molecular da variante S341R gerado pelo I-tasser com o sulfato de queratano.....	44
Figura 21 - <i>Docking</i> molecular da variante S341R gerado pelo Modeller com o sulfato de queratano.....	44
Figura 22 - <i>Docking</i> molecular da variante S341R gerados pelo I-tasser e Modeller, respectivamente, com o sulfato de condroitina.....	45
Figura 23 - <i>Docking</i> molecular da variante V239F gerado pelo I-tasser com o sulfato de queratano.....	46
Figura 24 - <i>Docking</i> molecular da variante V239F gerado pelo Modeller com o sulfato de queratano.....	46
Figura 25 - <i>Docking</i> molecular da variante V239F gerada pelo I-tasser com o sulfato de condroitina.....	47
Figura 26 - <i>Docking</i> molecular da variante V239F gerada pelo Modeller com o sulfato de condroitina.....	47
Figura 27 - Região da mutação da variante D45G gerada pelo Modeller.....	49

LISTA DE TABELAS

Tabela 1 - Mucopolissacaridoses.....	17
Tabela 2 - Mutações do gene GALNS encontradas nos pacientes do estado da Paraíba.....	29
Tabela 3 - Comparação dos resultados da fase de análise da qualidade dos modelos do Modeller e do I-Tasser.....	32
Tabela 4 - Valores de RMSD e <i>TM-Score</i> gerados pelos alinhamentos dos modelos selvagem com as proteínas-alvo.....	32
Tabela 5 - Modelos em que o substrato se ligaram ao sítio ativo.....	48
Tabela 6 - Mutações do GALNS identificadas em pacientes do estado da Paraíba e seu efeito na proteína.....	49

LISTA DE ABREVIATURAS E SIGLAS

1D	Unidimensional
3D	Tridimensional
C6S	Condroitin-6-sulfato
DHA	<i>Dihydroxyalanine</i>
DNA	<i>Deoxyribonucleic Acid</i>
DOPE	<i>Discrete Optimized Protein Energy Score</i>
EC	<i>Enzyme Commission</i>
ERT	<i>Enzyme Replacement Therapy</i>
GAG	Glicosaminoglicano
GALNS	N-acetilgalactosamina-6-sulfatase
HGMD	<i>Human Gene Mutation Database</i>
HMDB	<i>Human Metabolome Database</i>
HSCT	<i>Hematopoietic Stem Cell Transplantation</i>
HTS	<i>High-Throughput Sequencing</i>
INAGEMP	Instituto Nacional de Genética Médica e Populacional
kb	<i>Kilobase</i>
kDa	<i>Kilodalton</i>
LSD	<i>Lysosomal storage diseases</i>
MPS	Mucopolissacaridose
mRNA	<i>Messenger Ribonucleic Acid</i>
NCBI	<i>National Center for Biotechnology Information</i>
OMIM	<i>Online Mendelian Inheritance in Man</i>
pb	Pares de Base
PDB	<i>Protein Data Bank</i>
QS	Sulfato de Queratano
RMN	Ressonância Magnética Nuclear
RMSD	<i>Root Mean Square Deviation</i>
RNA	<i>Ribonucleic Acid</i>
TM-Score	<i>Template Modeling Score</i>

LISTA DE SÍMBOLOS

Å Angstrom

SUMÁRIO

1 INTRODUÇÃO.....	14
2 REFERENCIAL TEÓRICO.....	17
2.1. Mucopolissacaridoses	17
2.2. Mucopolissacaridose IV-A (MPS IVA).....	18
2.3. Dados Epidemiológicos da MPS IV-A.....	19
2.4. Aspectos moleculares da MPS IV-A e a enzima GALNS	19
2.5. Mutações no gene GALNS	20
2.6. Modelagem de proteínas.....	21
2.7. Docking molecular	23
3 METODOLOGIA.....	24
3.1. Identificação e seleção da proteína-molde	26
3.2. Alinhamento das sequências.....	26
3.3. Desenvolvimento de software para inserção de mutações.....	26
3.4. Construção do modelo.....	27
3.4. Validação do modelo.....	27
3.5. Docking molecular	28
4 RESULTADOS	29
4.1. Identificação e seleção da proteína molde.....	29
4.2. Construção e validação dos modelos.....	29
4.3. Dockings moleculares.....	35
5 DISCUSSÃO	49
6 CONCLUSÃO	52
REFERÊNCIAS	53
APÊNDICE A - GRÁFICO DE RAMACHANDRAN DA VARIANTE p.D45G GERADA PELO MODELLER	60
APÊNDICE B - GRÁFICO DE RAMACHANDRAN DA VARIANTE p.G301C GERADA PELO MODELLER.....	61
APÊNDICE C - GRÁFICO DE RAMACHANDRAN DA VARIANTE p.R386C GERADA PELO MODELLER.....	62
APÊNDICE D - GRÁFICO DE RAMACHANDRAN DA VARIANTE p.S341R GERADA PELO MODELLER.....	63
APÊNDICE E - GRÁFICO DE RAMACHANDRAN DA VARIANTE p.V239F GERADA PELO MODELLER.....	64
APÊNDICE F - GRÁFICO DO VERIFY3D DA VARIANTE p.D45G GERADA PELO MODELLER.....	65
APÊNDICE G - GRÁFICO DO VERIFY3D DA VARIANTE p.G301C GERADA PELO MODELLER.....	65

APÊNDICE H - GRÁFICO DO VERIFY3D DA VARIANTE p.R386C GERADA PELO MODELLER.....	66
APÊNDICE I - GRÁFICO DO VERIFY3D DA VARIANTE p.S341R GERADA PELO MODELLER.....	66
APÊNDICE J - GRÁFICO DO VERIFY3D DA VARIANTE p.V239F GERADA PELO MODELLER.....	67
APÊNDICE K - GRÁFICO DE RAMACHANDRAN DA VARIANTE p.D45G GERADA PELO I-TASSER.....	68
APÊNDICE L - GRÁFICO DE RAMACHANDRAN DA VARIANTE p.G301C GERADA PELO I-TASSER.....	69
APÊNDICE M - GRÁFICO DE RAMACHANDRAN DA VARIANTE p.R386C GERADA PELO I-TASSER.....	70
APÊNDICE N - GRÁFICO DE RAMACHANDRAN DA VARIANTE p.S341R GERADA PELO I-TASSER.....	71
APÊNDICE O - GRÁFICO DE RAMACHANDRAN DA VARIANTE p.V239F GERADA PELO I-TASSER.....	72
APÊNDICE P - GRÁFICO DO VERIFY3D DA VARIANTE p.D45G GERADA PELO I-TASSER.....	73
APÊNDICE Q - GRÁFICO DO VERIFY3D DA VARIANTE p.G301C GERADA PELO I-TASSER.....	73
APÊNDICE R - GRÁFICO DO VERIFY3D DA VARIANTE p.R386C GERADA PELO I-TASSER.....	74
APÊNDICE S - GRÁFICO DO VERIFY3D DA VARIANTE p.S341R GERADA PELO I-TASSER.....	74
APÊNDICE T - GRÁFICO DO VERIFY3D DA VARIANTE p.V239F GERADA PELO I-TASSER.....	75

1 INTRODUÇÃO

A estrutura tridimensional da enzima GALNS que, quando defeituosa, é responsável por causar a Mucopolissacaridose do tipo IV-A (MPS IVA), é conhecida desde 2012 por conta do trabalho de Rivera-Colón *et al.* Muitas mutações já foram identificadas como causadoras desse distúrbio, no entanto, poucas tiveram suas estruturas tridimensionais desvendadas e analisadas.

Mutações na proteína afetam não apenas a estrutura da proteína, mas também sua função e estabilidade. A previsão da estabilidade da proteína mutante com precisão é desejada para descobrir os aspectos moleculares das doenças e o design de novas proteínas (KULSHRESHTHA, 2016). Desta forma, é fácil de identificar porquê do interesse em compreender as estruturas tridimensionais das variantes produzidas por mutações na enzima GALNS. Ao comparar essas variantes com a GALNS selvagem, é possível determinar quais graves são as alterações na estrutura da proteína como resultado dessas mutações (TORRIERI, 2015). Esse tipo de informação é de extrema importância para um melhor entendimento das relações entre genótipo e fenótipo associados à MPS IVA (TORRIERI, 2015).

Computadores e softwares especializados tornaram-se uma parte essencial do *kit* de ferramentas do biólogo (GAUTHIER, 2019). Seja para análise de rotina de DNA ou sequência de proteínas ou para analisar informações significativas em conjuntos de dados biológicos de tamanho gigantesco de gigabytes, praticamente todos os projetos de pesquisa modernos em biologia exigem, até certo ponto, o uso de computadores (GAUTHIER, 2019). É tentador traçar as origens da bioinformática até a recente convergência do sequenciamento de DNA, projetos de genoma em larga escala, internet e supercomputadores (HAGEN, 2000). No entanto, alguns cientistas que afirmam que a bioinformática está em sua infância reconhecem que os computadores eram ferramentas importantes na biologia molecular uma década antes que o sequenciamento de DNA se tornasse viável (HAGEN, 2000). Momentos históricos que podem ser destacados para o seu desenvolvimento são a publicação da estrutura do DNA por Watson e Crick em 1953, além do acúmulo de dados e conhecimentos da bioquímica e da estrutura proteica com os estudos de Pauling, Corey e Ramachandran na década de 1960 (DINIZ, 2017). Segundo Hagen (2000),

Três fatores importantes facilitaram o surgimento da biologia computacional durante o início dos anos 1960. Primeiro, uma coleção em expansão de sequências de aminoácidos fornecia uma fonte de dados e um conjunto de problemas interessantes que eram inviáveis de resolver sem o poder de processamento

de números dos computadores. Em segundo lugar, a ideia de que macromoléculas carregam informações tornou-se uma parte central da estrutura conceitual da biologia molecular. Embora alguns historiadores e filósofos tenham questionado o significado teórico dessa ideia para a biologia molecular moderna, parece provável que pensar em termos de informação macromolecular forneceu um importante elo conceitual entre a biologia molecular e a ciência da computação, da qual surgiu a teoria da informação formal. Terceiro, os computadores digitais de alta velocidade, que se desenvolveram a partir de programas de pesquisa de armas durante a Segunda Guerra Mundial, enfim se tornaram amplamente disponíveis para biólogos acadêmicos. Nem todos os biólogos tinham — ou desejavam ter — acesso a essas máquinas, mas, em 1960, a escassez de computadores não era mais um obstáculo sério para o desenvolvimento da biologia computacional.

Pioneira na sistematização do conhecimento da organização tridimensional da proteína, Margaret O. Dayhoff é considerada a mãe da bioinformática (DINIZ, 2017). Esse fato se deve ao seu papel no desenvolvimento de computadores capazes de determinar a sequência peptídica, programas para reconhecer e exibir estruturas para uso em cristalografia de raios-X e métodos computacionais para comparação de cadeias de proteínas, permitindo inferir as conexões evolutivas entre reinos (DINIZ, 2017). Uma das realizações mais significativas da Dra. Dayhoff foi a derivação de uma árvore filogenética abrangente, baseada em informações e sequências proteicas e de ácidos nucleicos, que forneceu uma estrutura que descreve a história evolutiva primitiva e ajudou a desvendar o mistério das origens simbióticas dos eucariotos (HUNT, 1984).

Na década de 1980, as discussões sobre o genoma humano emergiram naturalmente com os avanços no sequenciamento de DNA, devido aos custos cada dia mais acessíveis. Ao longo dos anos 90, o Projeto Genoma Humano conduziu o sequenciamento do DNA humano, que em 2001 culminou nas publicações dos dois concorrentes nessa tarefa (DA SILVA, 2018). Além disso, nos anos de 1990, surgiu a Internet moderna e, na segunda metade da década de 90, o mundo experimentou o fenômeno das bolhas na Internet. Os esforços dos projetos do genoma promoveram novas tecnologias para o sequenciamento, como as tecnologias de sequenciamento de alto rendimento (HTS — *High-Throughput Sequencing*), usadas em laboratórios em todo o mundo (DA SILVA, 2018).

A tecnologia de sequenciamento de DNA vem desempenhando um papel fundamental no avanço da biologia molecular (ZHOU, 2010). Na década que se seguiu às publicações do projeto genoma humano, testemunhamos uma tremenda transformação neste campo. A brusca velocidade dessa mudança é comparável à rápida evolução da indústria de semicondutores sob a lei de Moore (ZHOU, 2010). Com o desenvolvimento da tecnologia de sequenciamento de nova geração, a escala do banco de dados biológico de DNA, RNA e

proteínas tem aumentado de forma drástica, podendo ser considerada “Uma era do *big data* biológico”. Como analisar eficientemente o *big data* biológico se torna um grande desafio, e a bioinformática é um meio importante para lidar com esse desafio (LIANG, 2019). Segundo Lee (2015),

Até o final de 2015, cerca de 90 milhões de sequências de proteínas foram depositadas no banco de dados UniProtKB. No entanto, o número correspondente de estruturas de proteínas no Protein Data Bank (PDB) é de apenas em torno de 100.000. A lacuna está aumentando rapidamente [...], onde a proporção de sequências sobre a estrutura aumentou de menos de 1 magnitude para cerca de 3 magnitudes nas últimas duas décadas. Assim, o desenvolvimento de algoritmos baseados em computador eficientes que podem gerar previsões de estrutura 3D de alta resolução se torna provavelmente o único caminho para preencher a lacuna.

Muitas abordagens computacionais avançadas foram desenvolvidas ao longo dos anos, para prever a estabilidade e a função de uma proteína mutada (KULSHRESHTHA, 2016). Essas abordagens baseadas na estrutura, características de sequência e características combinadas fornecem estimativa razoavelmente precisa do impacto da substituição de aminoácidos na estabilidade e função da proteína (KULSHRESHTHA, 2016).

Assim, o objetivo deste trabalho é se utilizar dessas abordagens computacionais para tentar prever a estrutura tridimensional das variantes da enzima GALNS identificadas nos pacientes com MPS IVA do estado da Paraíba, e avaliar a forma que cada uma dessas variantes se relaciona com os substratos da enzima. Os objetivos específicos são: prever a estrutura tridimensional da enzima GALNS através do método de modelagem por homologia; desenvolver um software que facilite o processo de inserção das mutações na sequência da enzima GALNS a fim de facilitar a obtenção de resultados; utilizar a metodologia de *docking* molecular para entender como os modelos das variantes gerados interagem com os substratos.

2 REFERENCIAL TEÓRICO

2.1. Mucopolissacaridoses

Os distúrbios de armazenamento lisossomal (LSD, *Lysosomal storage diseases*) são distúrbios hereditários raros com mais de 50 tipos distintos. De forma individual, essas doenças são raras, mas, em coletivo, são relativamente comuns, com uma prevalência que varia de 12 a 25 por 100.000 nascidos vivos (AL-JASMI, 2012). As mucopolissacaridoses (MPS) são um grupo de distúrbios de armazenamento lisossomal causados pela deficiência de enzimas que catalisam a degradação gradativa de glicosaminoglicanos (GAGs). São caracterizadas pelo acúmulo intralisossomal e aumento da excreção urinária de GAGs não totalmente degradados, o que resulta em disfunção celular, tecidual e orgânica (COUTINHO, 2011). Sete tipos distintos de MPS — I, II, III, IV, VI, VII e IX — com 11 deficiências de enzimas lisossômicas específicas já foram descritos (Tabela 1) (LIN, 2018).

Tabela 1. Mucopolissacaridoses. Adaptado de Bochernitsan *et al.* (2015).

Tipo	Epônimo	Enzima	Gene	Metabólitos urinários
MPS I	Hurlers, Hurler–Scheie, Scheie	α -L-Iduronidase	IDUA	Sulfato de dermatano, sulfato de heparano
MPS II	Hunter	Iduronato-2-sulfatase	IDS	Sulfato de dermatano, sulfato de heparano
MPS III A	Sanfilippo A	Heparan-N-sulfatase	SGSH	Sulfato de heparano
MPS III B	Sanfilippo B	α -N-acetilglucosaminidase	NAGLU	Sulfato de heparano
MPS III C	Sanfilippo C	AcetilCoA: α -glucosamina acetil transferase	HGSNAT	Sulfato de heparano
MPS III D	Sanfilippo D	N-acetilglucosamina-6-sulfatase	GNS	Sulfato de heparano
MPS IV A	Morquio A	N-acetilgalactosamina-6-sulfatase	GALNS	Sulfato de queratano, sulfato de condroitina
MPS IV B	Morquio B	β -Galactosidase	GLB1	Sulfato de queratano
MPS VI	Maroteaux-Lamy	N-acetilgalactosamina-4-sulfatase	ARSB	Sulfato de dermatano
MPS VII	Sly	β -Glucuronidase	GUSB	Sulfato de dermatano, sulfato de heparano
MPS IX	Natowicz	Hialuronidase	HYAL1	Ácido hialurônico

Como em todos os distúrbios de mucopolissacaridose, o padrão de distribuição tecidual dos GAGs específicos acumulados determina as manifestações clínicas do distúrbio (WOOD, 2013).

2.2. Mucopolissacaridose IV-A (MPS IVA)

A mucopolissacaridose IV-A (MPS IVA; OMIM 253000) ou doença de Morquio A é uma doença autossômica recessiva causada pela deficiência de N-acetilgalactosamina-6-sulfatase (GALNS; EC 3.1.6.4), a enzima lisossômica responsável pela degradação hidrolítica do sulfato de queratano (QS) e do condroitin-6-sulfato (C6S) (CACIOTTI, 2018). Na MPS IVA, essas manifestações podem incluir baixa estatura, anormalidades esqueléticas, instabilidade cervical, resistência limitada, comprometimento visual e auditivo, problemas de saúde bucal, anomalias cardiovasculares e comprometimento significativo do sistema respiratório (WOOD, 2013). O cérebro e a medula espinhal não são afetados de forma direta pelo acúmulo de GAGs, e a inteligência é preservada, embora as oportunidades de educação possam ser reduzidas por limitações físicas (CHARROW, 2015). Os sintomas em geral se tornam aparentes no final do primeiro ou segundo ano de vida, com uma diminuição progressiva da velocidade de crescimento (CAMELIER, 2011). O paciente normalmente atinge uma estatura final entre 113 e 122 cm, apresentando cifose, *pectus carinatum*, cifoesciose e fâcies característica (CAMELIER, 2011).

Não há cura definitiva para a síndrome de Morquio. O tratamento dessa síndrome consiste no gerenciamento médico e cirúrgico dos sistemas envolvidos, com o objetivo de palição, prevenção e desaceleração da progressão das complicações (ALGAHIM, 2013). Uma vez diagnosticado, o tratamento da síndrome de Morquio requer uma abordagem sistêmica e multidisciplinar, geralmente envolvendo médicos de atenção primária, ortopedistas, pneumologista, cardiologista e anesthesiologista (ALGAHIM, 2013). Outras formas de tratamento foram desenvolvidas para MPS nas últimas duas décadas, experimental e clinicamente, incluindo Terapia de Reposição Enzimática (ERT, *Enzyme Replacement Therapy*), terapia gênica, cirurgia ortopédica, manejo da obstrução traqueal e transplante de células-tronco hematopoiéticas (HSCT, *hematopoietic stem cell transplantation*), levando à melhora parcial dos fenótipos clínicos (CHINEN, 2014; SAWAMOTO, 2020).

2.3. Dados Epidemiológicos da MPS IV-A

O MPS IVA é um distúrbio raro, sua prevalência varia bastante entre as populações. Estima-se que a incidência seja de 1/216.000 nascidos vivos na Colúmbia Britânica (DIETER, 2007); 1/76.000 nascidos vivos na Irlanda do Norte com base em casos notificados durante 30 anos; 1/450.000 nascidos vivos na Holanda ao longo de 26 anos; 1/201.000 na Austrália; 1/640.000 nascidos vivos na Austrália Ocidental no decorrer de 27 anos; e 1/450.000 nascidos vivos em Portugal durante 20 anos (DIETER, 2007).

No Brasil, foram identificados 153 casos de MPS IVA no período de 34 anos de coleta de dados, correlacionados com elevado grau de consanguinidade em cidades do interior (GIUGLIANI, 2017; SANTOS, 2014; WELLER, 2012; BOCHERNITSAN, 2018). Segundo Bochernitsan (2018), o estado com maior número de pacientes é a Paraíba, com a maioria portando a mesma mutação, a c.1023C>G (p.S341R), e compartilhando o mesmo haplótipo (abhcd). A mutação p.S341R se localiza no éxon 10 do gene GALNS e possui fenótipo severo (DÜNG, 2013). Provavelmente, essa mutação foi derivada de um ancestral comum no nordeste do Brasil, e no estado da Paraíba, onde a maioria dos pacientes foi identificada. Pode-se inferir que indivíduos de um ancestral comum tenham herdado ambas as cópias do gene mutado e, assim, tenham mantido o genótipo raro, de maneira que a manutenção deste alelos podem estar relacionados com a alta frequência de consanguinidade observada em pessoas com deficiência no estado da Paraíba (WELLER, 2012). De acordo com a história local da ocupação populacional, as famílias portuguesas com forte tradição de uniões consanguíneas foram as principais a colonizar a região (WELLER, 2012). A alta frequência da mesma mutação (com o mesmo haplótipo) e a alta taxa de consanguinidade dos pais são fortes evidências para a presença de um efeito fundador para os pacientes com MPS IV-A da Paraíba (SANTOS-LOPES, 2021).

2.4. Aspectos moleculares da MPS IV-A e a enzima GALNS

O gene GALNS está localizado no braço do cromossomo 16, (16q24.3) e abrange aproximadamente 50 kb contendo 13 introns e 14 exons. Um produto de *splicing* simples, mRNA GALNS de 1566 pb, codifica um precursor de 522 aminoácidos (incluindo um peptídeo sinal de 26 resíduos) que é processado em um polipeptídeo de 60 kDa (TOMATSU, 1995; CATARZI, 2012; OLARTE-AVELLANEDA, 2014). Até Junho de 2022, 378 mutações no gene GALNS foram relatadas (HGMD, 2022); mutações missense/nonsense representam um total de 249 (74,5%), deleções 41 (12,3%), mutações no sítio de *splicing* 32 (9,5%) e

inserções 5 (1,5%) (HGMD, 2022). As mutações GALNS que causam Morquio A são muito heterogêneas e são detectadas em todo o gene. Mesmo as mutações detectadas com mais frequência são relativamente incomuns; no entanto, os efeitos fundadores podem alterar muito as frequências dos alelos GALNS em populações individuais (MORRONE, 2014).

A enzima GALNS é um oligômero com massa molecular de 120 kDa, sendo composta por polipeptídios de 40 e 15 kDa que são ligados por pontes dissulfeto (Bochernitsan, 2015). Como outras enzimas lisossomais, a GALNS possui um peptídeo sinal que conduz a proteína nascente ao retículo endoplasmático e sistema de aparelho de Golgi, no qual a proteína é modificada por N-glicosilação e processamento proteolítico (OLARTE-AVELLANEDA, 2014). A proteína GALNS madura tem um domínio arilsulfatase que é altamente conservado dentro da família de proteínas sulfatases humanas e entre ortólogos de diferentes espécies (CATARZI, 2012). Ela é responsável por clivar as ligações éster de sulfato, removendo os grupos sulfatos da porção terminal do QS e do C6S. Através desta hidrólise, os grupos sulfatos são removidos a partir de uma N-acetilgalactosamina-6-sulfato ou galactose-6-sulfato de C6S e QS, respectivamente (BOCHERNITSAN, 2015). O sítio ativo de GALNS é encontrado em uma bolsa no centro do domínio α/β N-terminal. Os resíduos do sítio ativo primário incluem Asp39, Asp40, Arg83, Tyr108, Lys140, His142, His236, Asp288, Asn289, Lys310 e DHA 79 (RIVERA-COLÓN, 2012). A modificação pós-traducional mais importante da GALNS, comum a outras sulfatases, é a ativação do sítio ativo por meio da conversão da cisteína 79 em formilglicina (FGly), em um processo mediado pela enzima geradora de formilglicina (OLARTE-AVELLANEDA, 2014).

2.5. Mutações no gene GALNS

Uma mutação é uma alteração na sequência de um ácido nucleico que resulta em uma alteração na estrutura ou função do ácido nucleico ou de uma proteína resultante (HERBERTSON, 2017). Existem muitos tipos de mutações no DNA, e algumas mudanças afetam apenas um nucleotídeo. Haverá substituição de um nucleotídeo por outro ou este pode ser completamente perdido. Outras modificações de DNA incluem rearranjos, deleções ou duplicatas de longas seções da sua sequência (THANKACHAN, 2021). Uma mutação *missense* é uma alteração na região de codificação que leva a uma substituição de aminoácidos; uma mutação que cria um códon de parada é uma mutação *nonsense* (KOPP, 2007). Pequenas deleções ou inserções de nucleotídeos causam um deslocamento do quadro

de leitura (mutações de deslocamento de quadro), e isso geralmente leva a um produto proteico truncado (KOPP, 2007).

Na MPS IVA, várias mutações são comuns; as mutações recorrentes mais prevalentes no gene GALNS são: c.1156C4T (p.R386C), c.901G4T (p.G301C), c.337A4T (p.I113F), c.1A4G (p.M1 β), c.757C4T (p.R253W), c.871G4A (p.A291T), c.935C4G (p.T312S) e c.1171A4G (p.M391V), representando 8,9%, 6,8%, 5,7%, 2,3%, 2,1%, 1,8%, 1,8%, 1,8% e 1,8%, respectivamente (KHAN, 2017).

No estudo realizado por Bochernitsan (2018) com 65 pacientes brasileiros e 3 de outros países da América do Sul, as cinco mutações mais frequentes identificadas foram: p.Ser341Arg, p.Gly301Cys, p.Arg386Cys, p.Arg94Leu e p.Gly116Ser. Todos os pacientes com a p.S341R em homozigose incluídos nesse estudo (n = 11) apresentaram um haplótipo único (abhcd), sendo seis heterozigotos para essa mutação, e 5 apresentaram esse haplótipo em heterozigose. Essa mutação foi detectada no Nordeste entre 15 pacientes (11 em homozigose e 4 em heterozigose), a maioria do estado da Paraíba (n = 12). Além disso, dois pacientes heterozigotos para a mutação p.S341R foram detectados em duas outras regiões, Sudeste e Sul (BOCHERNITSAN, 2018), além desta mutação foi descrita apenas em dois pacientes do Sri Lanka (SANTOS-LOPES, 2022).

2.6. Modelagem de proteínas

As interações de proteínas com demais macromoléculas, mais em específico com outras proteínas, desempenham um papel crucial na maioria dos processos celulares (LENSINK, 2017). Caracterizar as propriedades bioquímicas, biofísicas e estruturais dessas interações e os conjuntos multicomponentes que elas formam é, portanto, necessário para entender a função celular (LENSINK, 2017). A função de um polipeptídeo é determinada por sua sequência e sua estrutura tridimensional (3D). Projetos de sequenciamento genômico em larga escala fornecem aos pesquisadores milhões de sequências de proteínas, de vários organismos, em um ritmo sem precedentes (WEBB, 2017). No entanto, a taxa de caracterização estrutural experimental dessas sequências é limitada pelo custo, tempo e desafios experimentais. Na ausência de estruturas determinadas experimentalmente, os modelos de estrutura de proteínas derivadas da computação são valiosos para gerar hipóteses testáveis. (WEBB, 2017). O conhecimento da estrutura terciária de um polipeptídeo constitui uma informação valiosa para determinação de sua finalidade, pois pode permitir a

identificação de domínios conhecidos, como sítios catalíticos, sítios de modificação alostérica e outros (PROSDOCIMI, 2002).

Para Fiser (2017),

O estudo dos princípios que ditam a estrutura tridimensional das proteínas naturais pode ser abordado tanto pelas leis da física quanto pela teoria da evolução, e cada uma dessas abordagens fornece bases para uma classe de métodos de previsão de estruturas proteicas. A primeira abordagem, métodos de modelagem *ab initio*, ou sem modelo, prediz a estrutura somente a partir da sequência. Os métodos *ab initio* assumem que a estrutura nativa corresponde ao mínimo de energia livre global acessível durante a vida útil da proteína e tentam encontrar esse mínimo através da exploração de muitas conformações proteicas concebíveis. A segunda classe de métodos, chamada modelagem baseada em modelo, inclui ambas as técnicas de encadeamento que retornam uma descrição tridimensional completa do alvo. Essa classe depende da similaridade detectável que abrange a maior parte da sequência modelada e pelo menos uma estrutura conhecida. A modelagem comparativa refere-se aos casos de modelagem baseados em modelo quando não apenas a dobra é determinada a partir de um conjunto possível de modelos disponíveis, mas um modelo de átomo completo é construído.

A modelagem comparativa da estrutura da proteína tem sido usada para produzir modelos de estrutura confiáveis para pelo menos um domínio em mais da metade de todas as sequências conhecidas (WEBB, 2017). Portanto, as abordagens computacionais podem fornecer informações estruturais para duas ordens de magnitude mais sequências do que os métodos experimentais, e espera-se que cada vez mais se dependa disso, pois a diferença entre o número de sequências conhecidas e o número de estruturas determinadas experimentalmente continua a aumentar (WEBB, 2017). Se os modelos estão ausentes da biblioteca Protein Data Bank, eles precisam ser construídos do zero, ou seja, a partir de um método *ab initio*. Esta é a categoria mais difícil de previsão da estrutura da proteína (WU, 2007). Ambas as abordagens para a previsão da estrutura têm suas vantagens e limitações. Em princípio, a abordagem *ab initio* pode ser aplicada para modelar qualquer sequência. No entanto, devido à complexidade e ao nosso conhecimento limitado do problema de dobramento de proteínas, os métodos *ab initio* geralmente resultam em modelos de resolução relativamente baixa (FISER, 2017). Em geral, os métodos *ab initio* são restritos a proteínas relativamente pequenas e não são tão bons quanto os que são baseados em modelos estruturais relacionados (BAKER, 2006). A maior parte da modelagem *ab initio* pode ser pensada como um problema de otimização que tenta identificar a estrutura nativa de uma proteína individual, encontrando o modelo de menor energia em uma grande galeria de modelos possíveis (BAKER, 2006). Em contraste com as técnicas *ab initio*, a modelagem comparativa da estrutura da proteína geralmente fornece modelos que são comparáveis à

cristalografia de raios X de baixa resolução ou às estruturas de solução de Ressonância Magnética Nuclear (RMN) de média resolução. No entanto, sua aplicabilidade é limitada às sequências que podem ser mapeadas com segurança para estruturas conhecidas (FISER, 2017).

2.7. *Docking* molecular

O campo de *docking* (docagem ou encaixe) molecular surgiu durante as últimas três décadas impulsionado pelas necessidades da biologia molecular estrutural e da descoberta de drogas baseadas em estrutura (MORRIS, 2008). Ele é um método de design de drogas baseado em estrutura que simula a interação molecular e prediz o modo de ligação e a afinidade entre receptores e ligantes (FAN, 2019). Foi muito facilitado pelo crescimento dramático na disponibilidade e poder dos computadores, e a crescente facilidade de acesso a bancos de dados de pequenas moléculas e proteínas (MORRIS, 2008).

Segundo Dias (2008),

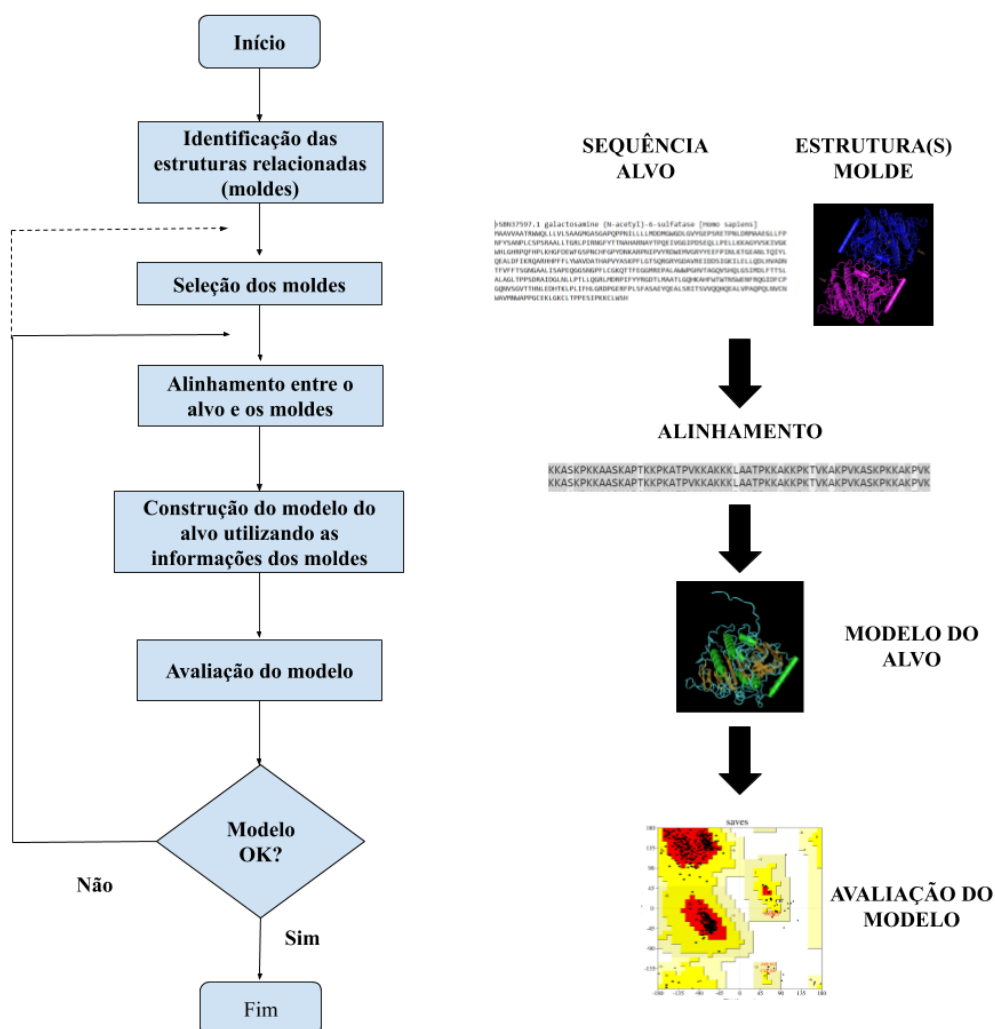
Simulações de *docking* molecular podem ser utilizadas para reproduzir dados experimentais através de algoritmos de validação de *docking*, onde as conformações proteína-ligante ou proteína-proteína são obtidas *in silico* e comparadas com estruturas obtidas por cristalografia de raios X ou ressonância magnética nuclear. Além disso, o *docking* é uma das principais ferramentas para procedimentos de triagem virtual, onde uma biblioteca de vários compostos é “encaixada” contra um alvo de droga e retorna o melhor resultado.

Uma metodologia de encaixe bem-sucedida deve ser capaz de prever corretamente a posição do ligante nativo dentro do sítio de ligação do receptor (ou seja, encontrar a geometria experimental do ligante dentro de um certo limite de tolerância) e as interações moleculares físico-químicas associadas (GUEDES, 2014).

3 METODOLOGIA

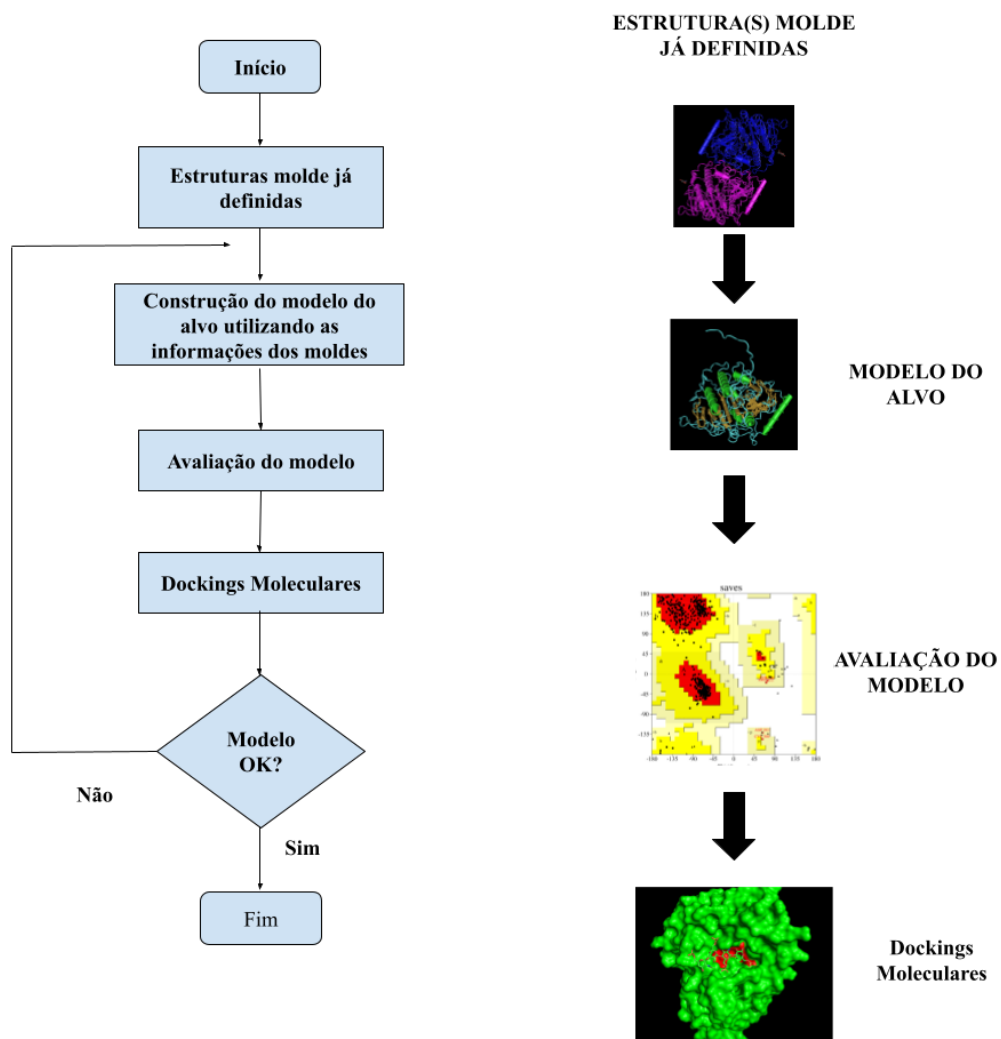
A modelagem comparativa geralmente consiste em quatro etapas principais: (a) atribuição de dobras que identifica similaridade geral entre a sequência de destino e pelo menos uma estrutura conhecida (modelo); (b) alinhamento da sequência alvo e do(s) modelo(s); (c) construir um modelo baseado no alinhamento com o(s) modelo(s) escolhido(s); e (d) prever a precisão do modelo (WEBB, 2017). O processo está esquematizado na figura 1.

Figura 1: Passo a passo do processo de modelagem por homologia convencional. Adaptado de MARTÍ-RENOM *et al.*, 2000.



No entanto, pelo fato deste trabalho se tratar de um processo de modelagem de variantes de uma proteína específica, e não de proteínas diferentes com a mesma origem evolutiva, haverá alterações em algumas etapas do processo (Figura 2).

Figura 2: Passo a passo do processo de modelagem por homologia para as mutações da proteína GALNS presentes nos pacientes do estado da Paraíba. Adaptado de MARTÍ-RENOM *et al.*, 2000.



As diferenças entre as duas metodologias serão explicadas a seguir:

3.1. Identificação e seleção da proteína-molde

Na primeira etapa, diferentemente do processo convencional de modelagem, já foi determinado que a enzima GALNS selvagem foi utilizada como molde para a construção dos modelos das enzimas variantes, sendo necessário apenas escolher quais modelos disponíveis serão utilizados na modelagem. Além disso, a sequência da GALNS selvagem foi adquirida no banco de dados do *National Center for Biotechnology Information* (NCBI), e a partir dela foram geradas as sequências das proteínas-alvo. Em geral, para a modelagem de proteínas mutadas, a geração das sequências alvo a partir da sequência molde se faz de forma manual, no entanto, por questão de praticidade e a fim de evitar possíveis erros durante o processo, um software que adiciona as mutações na sequência em seus locais exatos de forma automática foi desenvolvido.

3.2. Alinhamento das sequências

Na segunda etapa da modelagem por homologia, métodos de alinhamento global entre as sequências geralmente são aplicados a fim de analisar-se se a proteína escolhida como molde e as proteínas alvo têm similaridade suficiente para que os modelos sejam gerados. No entanto, neste trabalho, as proteínas alvo diferem da proteína molde apenas por mutações de ponto, e assim, o alinhamento não é necessário.

3.3. Desenvolvimento de software para inserção de mutações

O desenvolvimento desse programa inicialmente possuía o objetivo de agilizar e automatizar o processo de inserção das mutações, o que diminuiria a chance de erro humano, mas ele também acabou por ser útil em manter a sequência completa da proteína, uma vez que foi observado que os modelos obtidos por cristalografia de raios X nos trabalhos de Rivera-Colón e colaboradores (2012) apresentam aminoácidos a menos nas suas extremidades (27 aminoácidos a menos na região N-Terminal, e 2 a menos na C-Terminal, no caso da 4FDJ, e 28 aminoácidos a menos na região N-Terminal, no caso da 4FDI), assim, a modificação direta dessas estruturas para a adição das mutações faria com que os modelos gerados apresentassem a mesma característica.

Inicialmente, um protótipo do programa foi desenvolvido em Python, para que fosse aferido se de fato esse tipo de aplicação facilitaria de fato a obtenção de resultados. Uma vez que isso foi confirmado, a versão final foi desenvolvida, uma aplicação web denominada

“Gerador de Mutações de Ponto”, que está disponível em <<https://kellorran.github.io/>>. Essa versão final foi desenvolvida em JavaScript.

Neste software, o usuário é capaz de inserir uma mutação de ponto em uma sequência usando a função “modificar a sequência”. Para isso, é necessário apenas que o usuário informe a posição da mutação em questão e a modificação em si que ela causa.

Além de “modificar sequência”, o usuário também pode utilizar a função “analisar quantidades”. Com ela é possível tanto descobrir o número de monômeros naquela sequência que está sendo analisada quanto saber a quantidade de um monômero específico. Isto é, em uma sequência de exemplo: “AAACCT”, o número de monômeros da sequência retornado seria 6, a quantidade de monômeros “A” seria 3, do monômero “C” seria 2 e do monômero “T” apenas 1. Essa funcionalidade foi adicionada para que o usuário possa ter certeza que a adição da mutação de ponto feita de forma automática pela aplicação não mudou a quantidade de monômeros na sequência, e que a quantidade de monômeros iguais ao que foi adicionado à sequência subiu em 1, ou do que foi substituído pela adição da mutação diminuiu em 1, por exemplo.

3.4. Construção do modelo

A terceira etapa da modelagem é a construção do modelo, a qual é baseada na informação contida no alinhamento gerado entre as sequências. É a modelagem que utiliza a satisfação das restrições espaciais, empregando distâncias geométricas e técnicas de otimização para satisfazer as restrições espaciais das proteínas molde dadas por meio do alinhamento. A construção dos modelos tridimensionais de todas as variantes foi efetuada tanto com o programa MODELLER (ŠALI, 1993) quanto com o I-Tasser (ROY, 2010; YANG, 2015a; YANG, 2015b), os resultados obtidos através desses dois programas foram comparados entre si posteriormente, a fim de se analisar qual das duas metodologias gerou melhores modelos. Para realizar o refinamento dos modelos gerados, o 3DRefine foi utilizado (BHATTACHARYA, 2012; BHATTACHARYA, 2013; BHATTACHARYA, 2016). A visualização da estrutura foi realizada através do software PyMol.

3.4. Validação do modelo

A qualidade dos modelos 3D gerados foi validada pelo gráfico de Ramachandran (RAMACHANDRAN, 1963) com o programa PROCHECK (LASKOWSKI, 1998), que avalia a estrutura tridimensional da proteína indicando a sua qualidade estereoquímica. A

qualidade do modelo está ligada à escolha do molde e à construção do alinhamento que servirá de referência para a construção dos átomos que compõem a proteína. O modelo gerado também foi avaliado por um programa que realiza comparações entre os modelos obtidos e proteínas resolvidas com alta resolução, o VERIFY 3D (LUTHY, 1992).

Os modelos tridimensionais das variantes foram posteriormente alinhados à GALNS selvagem com o uso da ferramenta TM-Align (ZHANG, 2005). O software PyMol foi utilizado para uma inspeção visual sobre a conformação das estruturas geradas em comparação com a proteína não mutada a fim de identificar possíveis diferenças estruturais.

3.5. Docking molecular

Após todas as etapas anteriores, todos os modelos gerados foram submetidos a um *docking* molecular, a fim de tentar prever como as variantes se relacionam com os ligantes. O programa utilizado nesta etapa será o Hex, na sua versão 8.0.0. A visualização dos modelos será feita através do PyMol.

4 RESULTADOS

4.1. Identificação e seleção da proteína molde

A genotipagem da mutação no gene GALNS dos pacientes do estado da Paraíba foi realizada pelo Instituto Nacional de Genética Médica e Populacional (INAGEMP). Além da p.S341R, outras cinco mutações também estão presentes nos pacientes desse estado. Todas elas estão apresentadas na tabela 2. As variantes selecionadas para a modelagem foram: p.D45G, p.G301C, p.R386C, p.S341R e p.V239F. Para a inserção de cada uma das mutações na sequência do gene GALNS, a aplicação web Gerador de Mutações de Ponto, anteriormente citada, foi utilizada.

Tabela 2. Mutações do gene GALNS encontradas nos pacientes do estado da Paraíba.

Alteração nucleotídica	Efeito no aminoácido	Éxon	Alelos detectados	Referências
c.1023C>G	p.S341R	10	4	TOMATSU, 2004
c.1156C>T	p.R386C	11	34	FUKUDA, 1996
c.901G>T	p.G301C	9	26	KATO, 1997
c.134A > G	p.D45G	2	3	BOCHERNITSAN, 2018
c.120+1G>A	-	1	3	LARADI, 2006
c.715G>T	p.V239F	7	2	TOMATSU, 2005

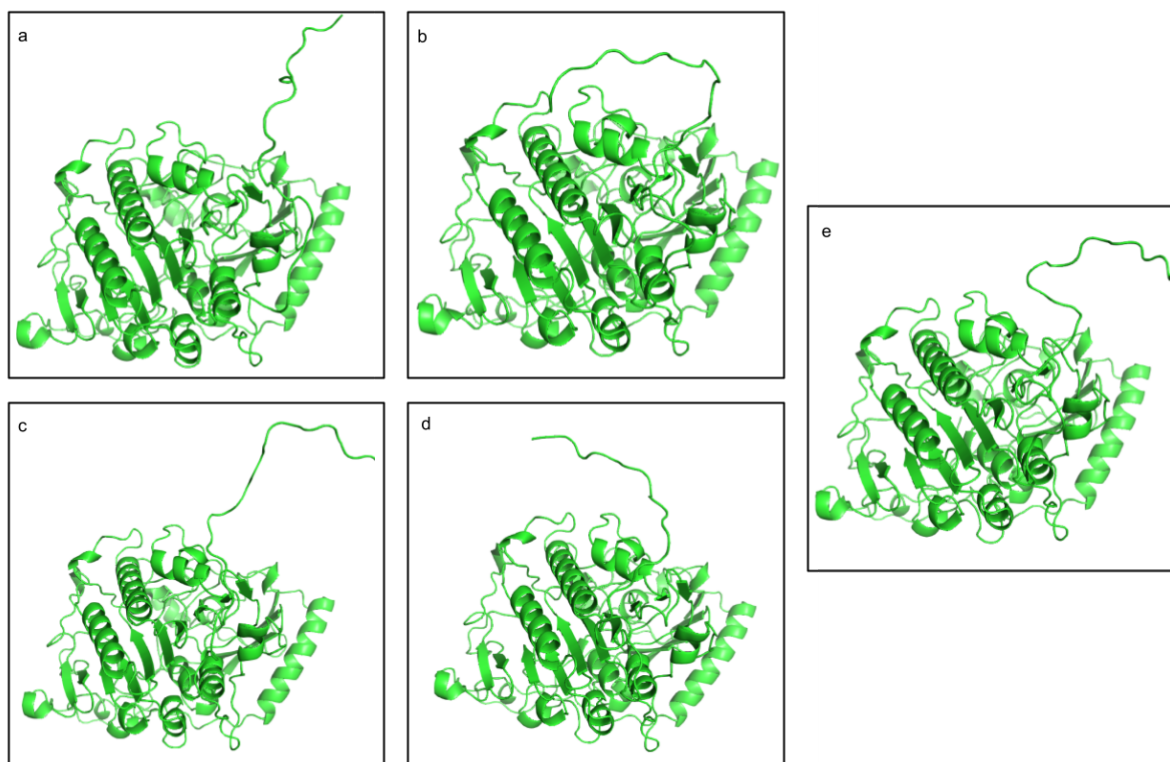
4.2. Construção e validação dos modelos

Após as sequências das variantes estarem todas realizadas e apuradas, iniciou-se o processo de modelagem. O programa Modeller na sua versão 9.25 foi utilizado primariamente. Os modelos 4FDJ e 4FDI da GALNS selvagem desvendados por Rivera-Colón *et al.* (2012), disponibilizados no *Protein Data Bank* (PDB) e já citados anteriormente foram utilizados como moldes para a construção das estruturas terciárias das proteínas mutadas. Ao final da execução, o Modeller apresenta 5 modelos para cada proteína, e a mais provável de se assemelhar à estrutura real é determinada pela sua pontuação de energia protéica discreta otimizada, ou *DOPE score*. Aquelas selecionadas nesse estágio passaram para a etapa de refinamento.

O refinamento dos modelos foi realizado através do software 3d Refine. As estruturas tridimensionais finais das variantes geradas pelo Modeller podem ser vistas na Figura 3.

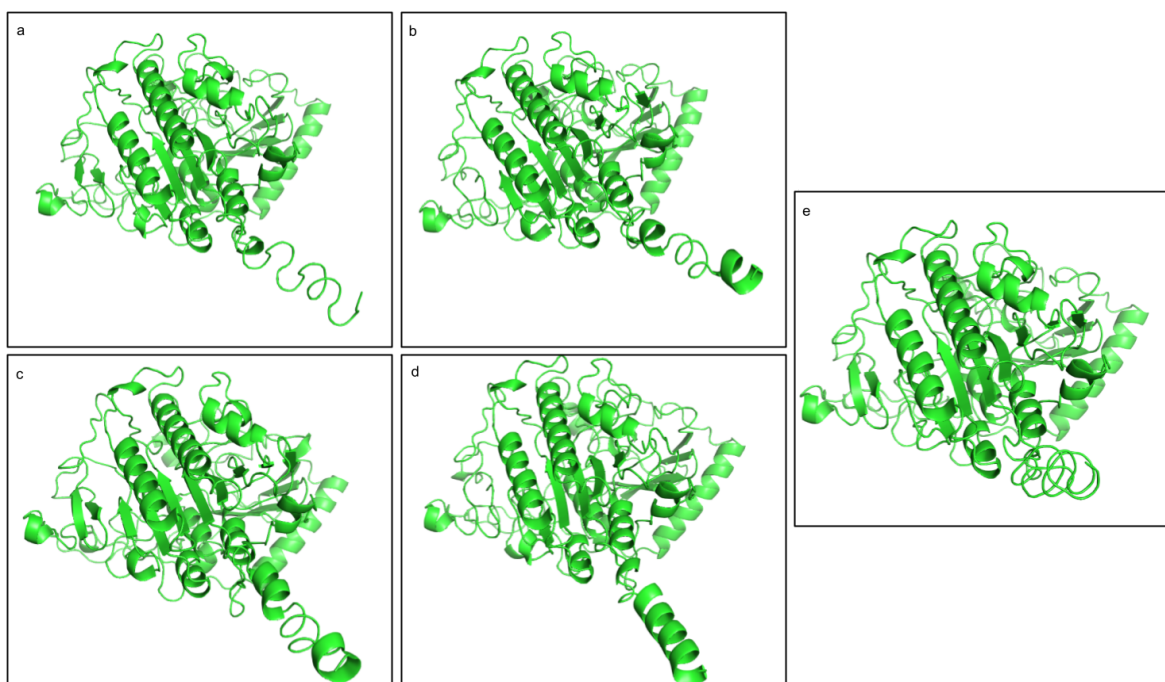
Para avaliar a qualidade estereoquímica dos modelos produzidos o Procheck foi utilizado. Esse programa analisa a geometria resíduo a resíduo e a geometria geral da estrutura (LASKOWSKI, 1996) para gerar um gráfico de Ramachandran, que indica as regiões favoráveis, as regiões adicionalmente permitidas, as generosamente permitidas e as não permitidas para os graus de todos os ângulos Phi e Psi dos aminoácidos (TORRIERI, 2015). O Verify3D foi utilizado para determinar a compatibilidade dos modelos atômicos (3D) com suas próprias sequências de aminoácidos (1D) atribuindo uma classe estrutural com base em sua localização e ambiente (alfa, beta, loop, polar, apolar etc) e comparando os resultados com boas estruturas (BOWIE, 1991; LÜTHY, 1992). Os modelos considerados bons pelo gráfico de Ramachandran são aqueles que alcançam 90% ou mais de seus resíduos localizados em regiões favoráveis (A, B, L) desconsiderando os resíduos de glicina, por estes não possuírem cadeia lateral (TORRIERI, 2015). No Verify3D os considerados bons são aqueles em que 80% ou mais dos aminoácidos tiveram pontuação $\geq 0,2$ no perfil 3D/1D. Os gráficos gerados por esses programas estão todos disponibilizados na lista de Apêndices.

Figura 3. Variantes da proteína GALNS geradas pelo Modeller. a) Variante p.D45G b) Variante p.G301C c) Variante R386C d) Variante p.S341R e) Variante p.V239F. Imagens geradas pelo PyMol. Fonte: Próprio autor.



Essas mesmas etapas se repetiram no segundo processo de modelagem que foi realizado através do I-Tasser. A Figura 4 mostra as estruturas tridimensionais finais que foram geradas através desse programa. Os gráficos gerados na análise da qualidade dos modelos do I-Tasser também estão disponíveis nos Apêndices.

Figura 4. Variantes da proteína GALNS geradas pelo I-Tasser. a) Variante p.D45G b) Variante p.G301C c) Variante R386C d) Variante p.S341R e) Variante p.V239F. Imagens geradas pelo PyMol. Fonte: Próprio autor.



Após a geração desses modelos, foi realizada a comparação dos resultados da análise de qualidade das proteínas-alvo produzidas tanto pelo Modeller quanto pelo I-Tasser (Tabela 3). Os valores adquiridos pela produção dos gráficos de Ramachandran das estruturas geradas pelo Modeller foram maiores que os do I-Tasser em todas as variantes, mas apesar disso, apenas os modelos p.S341R e p.V239F obtiveram resultados acima dos 90%. Já em relação aos valores retornados pelo Verify3D, todas as variantes, tanto as do Modeller quanto as do I-Tasser, passaram na análise, isto é, obtiveram resultados acima de 80%. As que foram geradas pelo Modeller, entretanto, obtiveram valores maiores em três das cinco variantes (p.G301C, p.R386C e p.V239F). Esses resultados indicam que os modelos produzidos pelo Modeller têm uma maior qualidade que os modelos gerados pelo I-Tasser.

Tabela 3. Comparação dos resultados da fase de análise da qualidade dos modelos do Modeller e do I-Tasser.

Variante	Modeller		I-Tasser	
	Ramachandran (A, B, L)	Verify3D	Ramachandran (A, B, L)	Verify3D
p.D45G	89,80%	92,53%	78,40%	95,40%
p.G301C	89,60%	92,34%	82,00%	86,21%
p.R386C	89,10%	93,10%	83,80%	91,76%
p.S341R	90,70%	91,38%	82,40%	93,30%
p.V239F	91,00%	92,91%	78,70%	92,15%

O *TM-Align* foi usado para a comparação entre a GALNS selvagem e as variantes. Ele gera um alinhamento resíduo a resíduo entre dois modelos tridimensionais com base na similaridade estrutural usando iterações de programação dinâmica heurística (ZHANG, 2005). A estrutura que foi utilizada como modelo da proteína selvagem para essa comparação foi recuperada no formato PDB, disponível no *Protein Data Bank* com o identificador 4FDJ. Trata-se da mesma estrutura que foi utilizada como um dos moldes na construção das estruturas das variantes pelo Modeller, resolvida pela técnica de difração de Raios-X e com uma resolução de 2.81Å. Como já citado anteriormente, foi observado que essa estrutura possui 27 aminoácidos a menos na região N-Terminal, e 2 a menos na C-Terminal em comparação com o que foi descrito por Hendriksz *et al.* (2013), assim, as análises foram feitas levando-se em consideração a GALNS selvagem com 493 aminoácidos (TORRIERI, 2015). Os valores do Desvio Quadrático Médio da Raiz (RMSD, *Root Mean Square Deviation*) e o *TM-Score* obtidos pelos alinhamentos e gerados pelo *TM-Align* podem ser vistos na tabela 4.

Tabela 4. Valores de RMSD e *TM-Score* gerados pelos alinhamentos dos modelos selvagem com as proteínas-alvo.

Variante	Modeller		I-Tasser	
	RMSD	<i>TM-Score</i>	RMSD	<i>TM-Score</i>
p.D45G	0,23	0,99916	0,31	0,99850
p.G301C	0,22	0,99923	0,30	0,99857
p.R386C	0,19	0,99940	0,29	0,99864
p.S341R	0,20	0,99936	0,31	0,99844
p.V239F	0,21	0,99930	0,31	0,99849

TM-score e RMSD são padrões utilizados para medir a similaridade estrutural entre dois modelos tridimensionais. Sempre que surge a questão da semelhança conformacional, particularmente para proteínas globulares, a medição padrão é o RMSD em coordenadas atômicas após as estruturas terem sido superpostas de forma otimizada pela translação e rotação do corpo rígido (MAIOROV, 1995). O RMSD é calculado entre átomos equivalentes em duas estruturas, definido como

$$rmsd = \sqrt{\frac{\sum d^2 i}{n}}$$

onde d é a distância entre cada um dos n pares de átomos equivalentes em duas estruturas superpostas de forma otimizada. O RMSD é 0 para estruturas idênticas e seu valor aumenta à medida que as estruturas se tornam mais diferentes (CARUGO, 2001).

Porém, um problema que se encontra ao usar RMSD para comparar estruturas de proteínas é que a sua distribuição depende do tamanho da proteína. Isso se torna importante quando os modelos de várias proteínas de extensões diferentes são avaliados em comparação uns com os outros. A dependência de RMSD no tamanho da proteína pode ser eliminada pelo cálculo do TM-score (KUFAREVA, 2011). O TM-Score tem resultados que variam entre 0 e 1, isto é, um valor de $0,0 < \text{TM-score} < 0,30$ indica uma similaridade estrutural aleatória, enquanto $0,5 < \text{TM-Score} < 1,00$ pode significar que as moléculas estão enoveladas mais ou menos da mesma forma (ZHANG, 2005).

Nessa etapa de avaliação não foi possível identificar diferenças estruturais muito acentuadas entre a GALNS selvagem e as variantes, como pode ser ilustrado pelos exemplos das Figuras 5 e 6. De fato, em uma análise visual, parece haver quase uma sobreposição entre a proteína selvagem e as mutadas, inclusive na própria região em que se encontra a mutação (Figura 7), o que pode ser corroborado pelos valores de RMSD e TM-Score gerados no alinhamento, que indicam uma alta similaridade entre as estruturas. Os arquivos em formato PDB dos modelos 3D produzidos por homologia das variantes da proteína GALNS, assim como os alinhamentos dessas mutações com a proteína selvagem gerados pelo TM-Align estão disponíveis no repositório <<https://github.com/Kellorran/VariantesGALNS.git>>

Figura 5. Alinhamento entre os modelos da GALNS selvagem (em verde) e da variante p.S341R (em azul). A posição da mutação está em vermelho e indicada por uma seta. Imagem gerada pelo PyMol. Fonte: Próprio autor.



Figura 6. Alinhamento entre os modelos da GALNS selvagem (em verde) e da variante p.V239F (em azul). A posição da mutação está em vermelho e indicada por uma seta. Imagem gerada pelo PyMol. Fonte: Próprio autor.

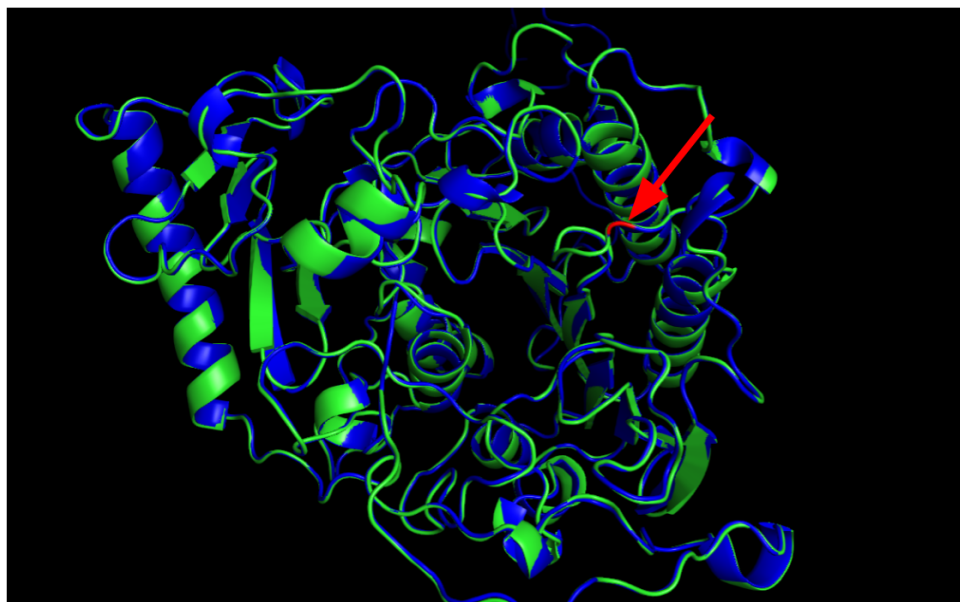
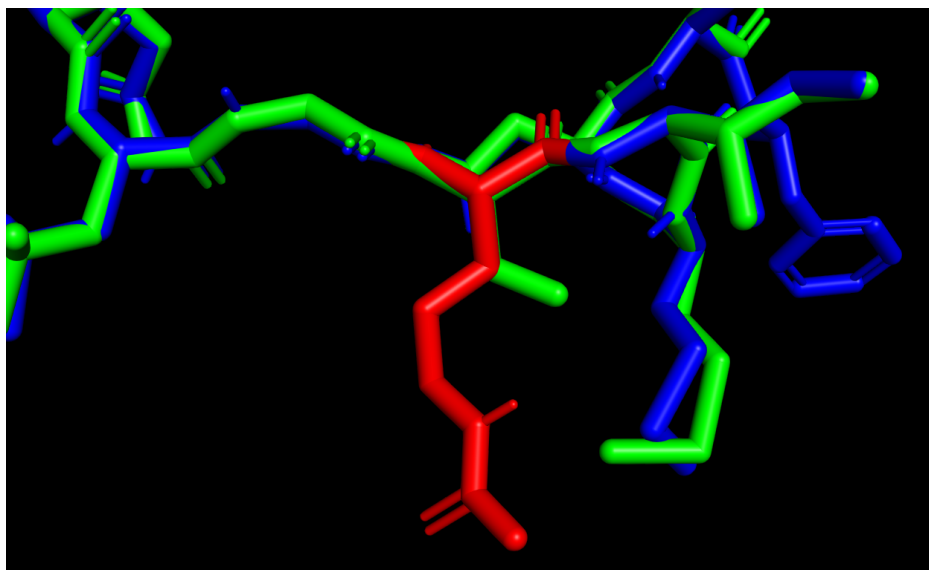


Figura 7. Alinhamento entre os modelos da GALNS selvagem (em verde) e da variante p.S341R (em azul) a nível de aminoácido. O aminoácido modificado pela mutação está representado em vermelho. Imagem gerada pelo PyMol. Fonte: Próprio autor.



4.3. Dockings moleculares

Os *Dockings* molecular, tanto da proteína selvagem quanto das variantes com os seus ligantes, foram realizados através do Hex na versão 8.0.0 (MACINDOE, 2010). Os modelos em formato pdb do sulfato de queratano e do condroitin-6-sulfato foram adquiridos no *Protein Data Bank* (PDB) e no *Human Metabolome Database* (HMDB), respectivamente. O modelo da proteína selvagem utilizado foi o 4FDJ, já citado anteriormente, porém modificado para apresentar apenas um dos monômeros da enzima GALNS a fim de facilitar a comparação com os modelos das variantes, que também se apresentam dessa forma. Os *dockings* realizados podem ser encontrados no repositório <<https://github.com/Kellorran/DockingsGALNS.git>>.

Diferentemente do que era esperado pelos resultados dos alinhamentos pelo *TM-Align*, as variantes apresentaram resultados diferentes entre si em relação ao *docking*, e foi possível perceber diferenças estruturais tanto entre elas quanto em relação ao modelo selvagem quando as superfícies das proteínas foram visualizadas (Figuras 8 e 9). Esses resultados indicam que as diferenças estruturais dos modelos, embora sejam baixas, são o suficiente para alterar suas relações com os ligantes em pelo menos alguns dos casos.

Figura 8. *Dockings* moleculares dos modelos gerados pelo I-Tasser com o ligante sulfato de queratano. a) Proteína GALNS selvagem b) Variante p.D45G c) Variante p.G301C d) Variante R386C e) Variante p.S341R f) Variante p.V239F. Imagens geradas pelo PyMol. Fonte: Próprio autor.

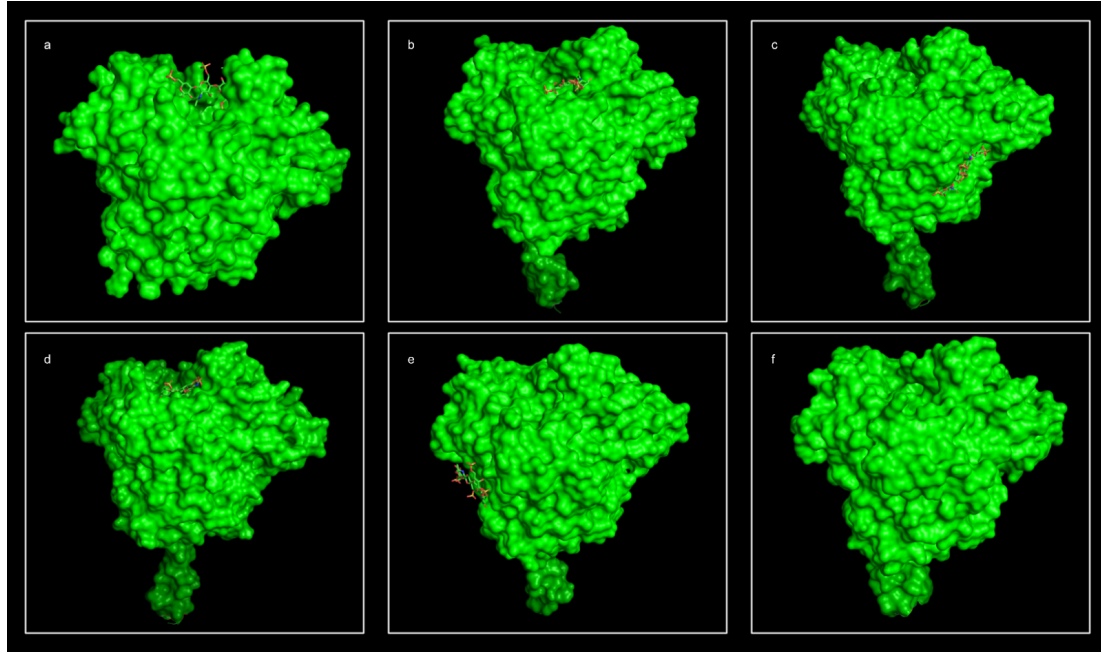
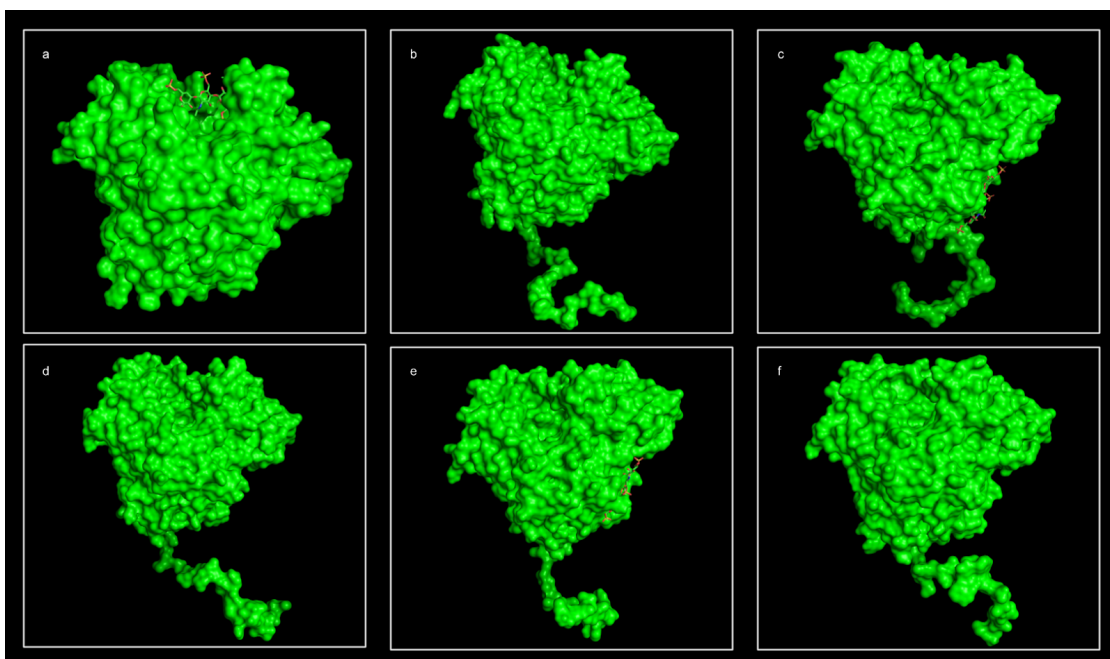


Figura 9. *Dockings* moleculares dos modelos gerados pelo Modeller com o ligante sulfato de queratano. a) Proteína GALNS selvagem b) Variante p.D45G c) Variante p.G301C d) Variante R386C e) Variante p.S341R f) Variante p.V239F. Imagens geradas pelo PyMol. Fonte: Próprio autor.



Quanto aos resultados individuais do *docking* molecular de cada uma das variantes, foi possível observar o seguinte:

1. Em relação à variante p.D45G:
 - 1.1. Os modelos obtidos no Modeller e no I-tasser tiveram resultados diferentes nos *dockings* moleculares do sulfato de queratano como ligante, mas os resultados foram iguais com o sulfato de condroitina (Figuras 10, 11 e 12);
 - 1.2. É possível constatar que no modelo gerado pelo I-tasser, o sulfato de queratano se liga ao sítio ativo da proteína, mas no que foi gerado pelo Modeller, a molécula se liga à face oposta à do sítio ativo (Figuras 10 e 11);
 - 1.3. O sulfato de condroitina não se ligou ao sítio ativo nem ao modelo do I-tasser, nem ao do Modeller, mas se ligaram à mesma região da proteína, longe do sítio ativo, em ambos (Figura 12).

Figura 10. *Docking* molecular da variante D45G gerado pelo I-tasser com o sulfato de queratano. a) Imagem com a superfície da proteína. Os resíduos do sítio ativo estão em vermelho. b) Imagem com a proteína em “cartoon”. O substrato está em roxo, os resíduos do sítio ativo estão coloridos. Imagem gerada pelo PyMol. Fonte: Próprio autor.

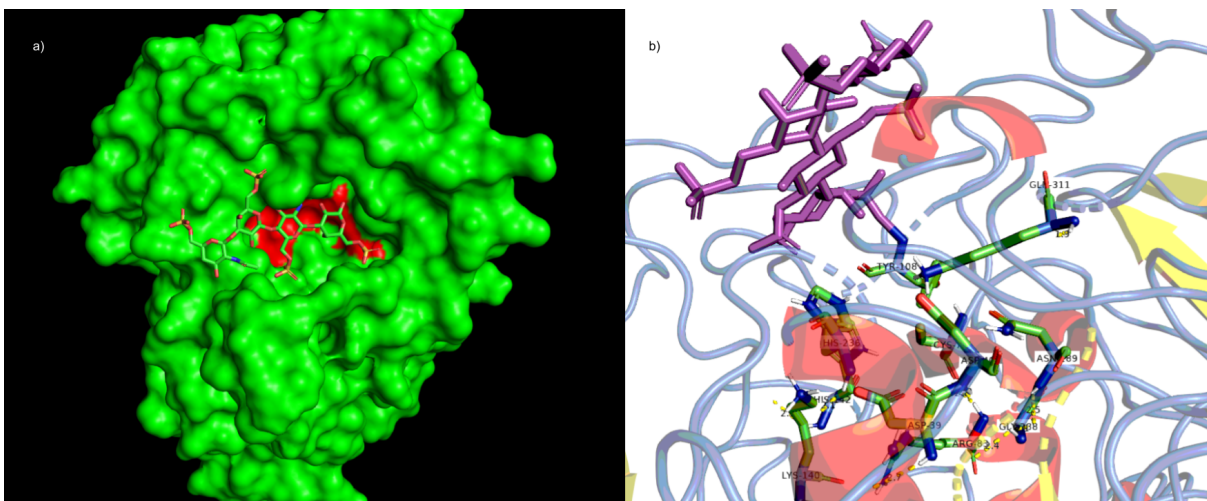


Figura 11. *Docking* molecular da variante D45G gerado pelo Modeller com o sulfato de queratano a) Imagem com a superfície da proteína. Os resíduos do sítio ativo estão em vermelho. b) Imagem com a proteína em “*cartoon*”. O substrato está em roxo, os resíduos do sítio ativo estão coloridos. Imagem gerada pelo PyMol. Fonte: Próprio autor.

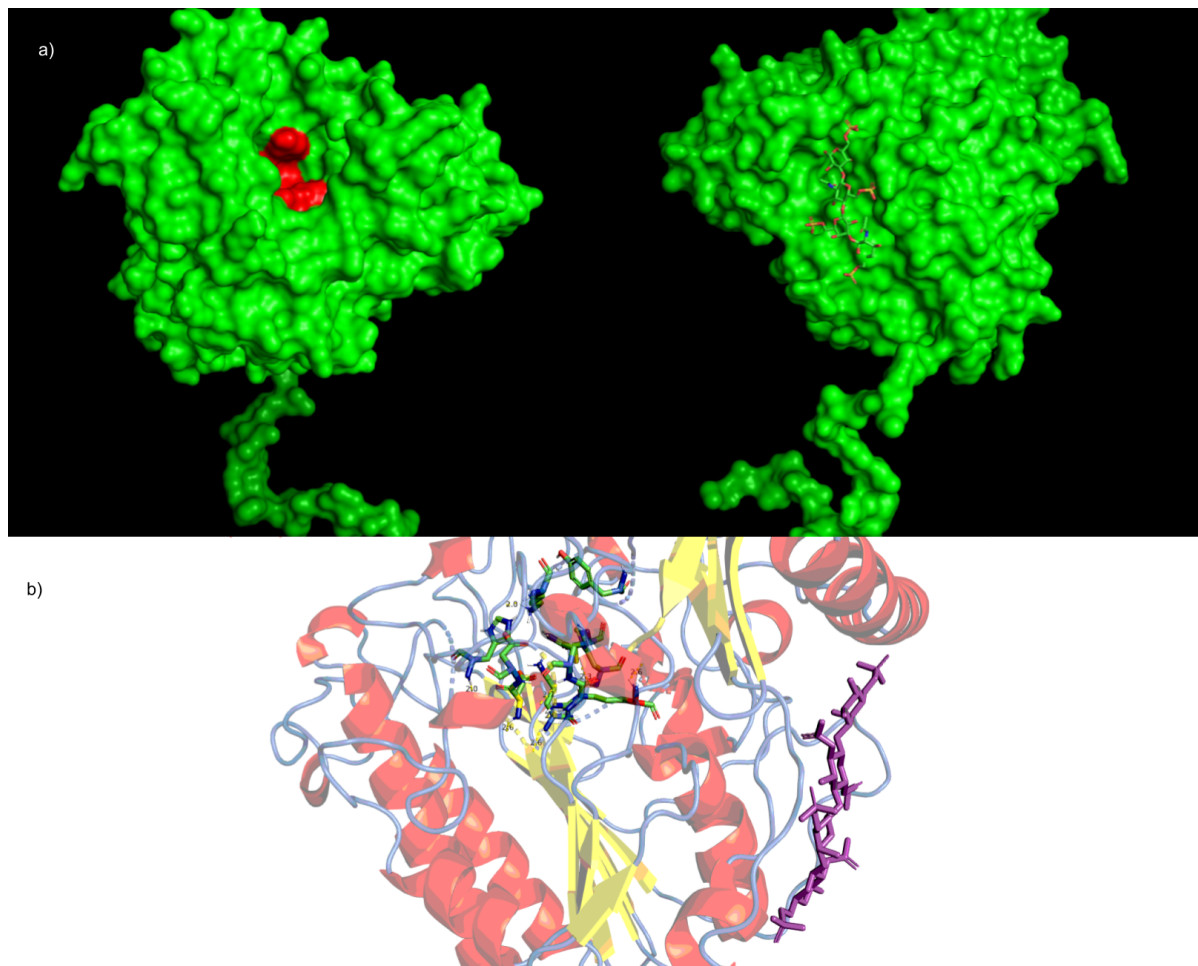
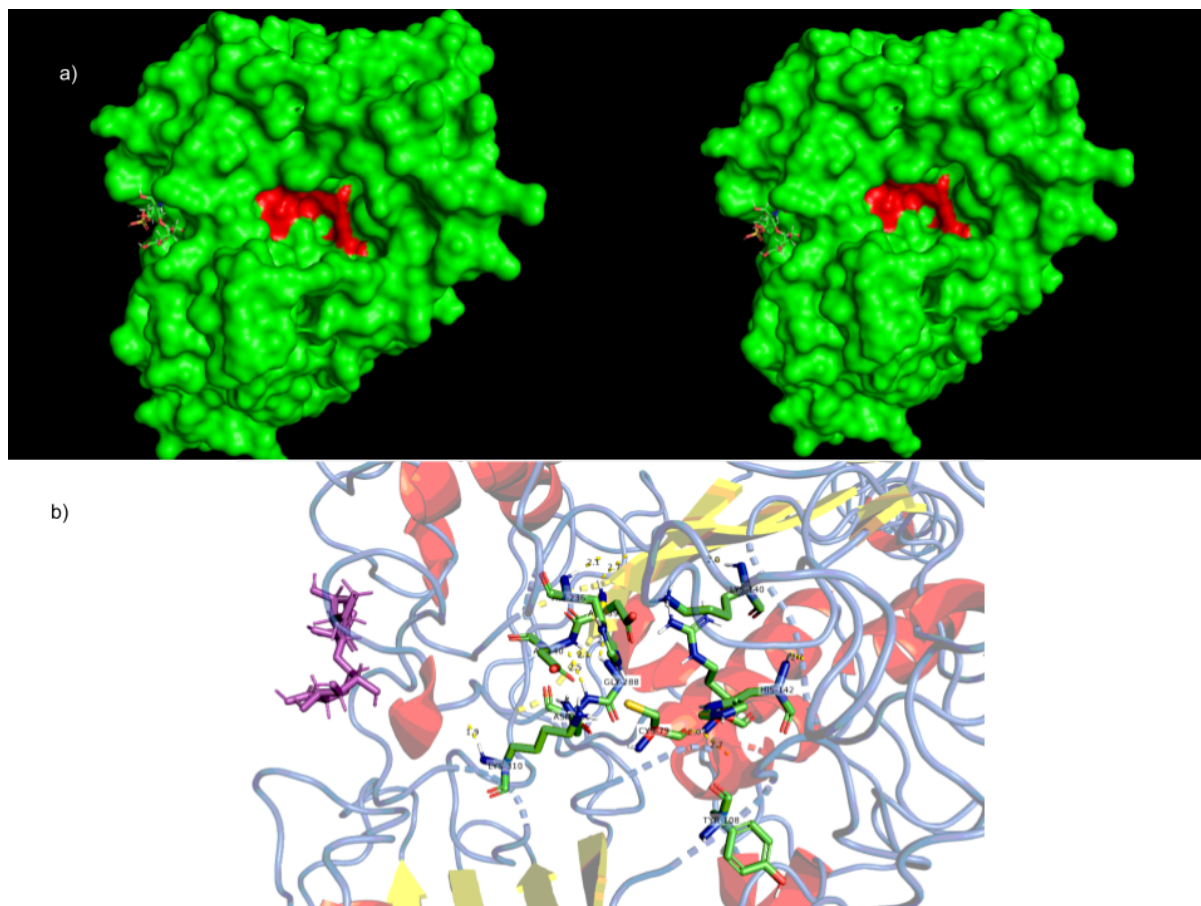


Figura 12. *Docking* molecular da variante D45G gerados pelo I-tasser e pelo Modeller, respectivamente, com o sulfato de condroitina. a) Imagem com a superfície da proteína. Os resíduos do sítio ativo estão em vermelho. b) Imagem com a proteína gerada pelo I-Tasser em “*cartoon*”. O substrato está em roxo, os resíduos do sítio ativo estão coloridos. Imagem gerada pelo PyMol. Fonte: Próprio autor.



2. Em relação à variante p.G301C:

- 2.1. O sulfato de queratano não se ligou ao sítio ativo da proteína em nenhum dos modelos, mas o sulfato de condroitina se ligou ao sítio ativo de ambos (Figuras 13, 14, 15 e 16).

Figura 13. *Docking* molecular da variante G301C gerada pelo I-tasser com o sulfato de queratano. a) Imagem com a superfície da proteína. Os resíduos do sítio ativo estão em vermelho. b) Imagem com a proteína em “*cartoon*”. O substrato está em roxo, os resíduos do sítio ativo estão coloridos. Imagem gerada pelo PyMol. Fonte: Próprio autor.

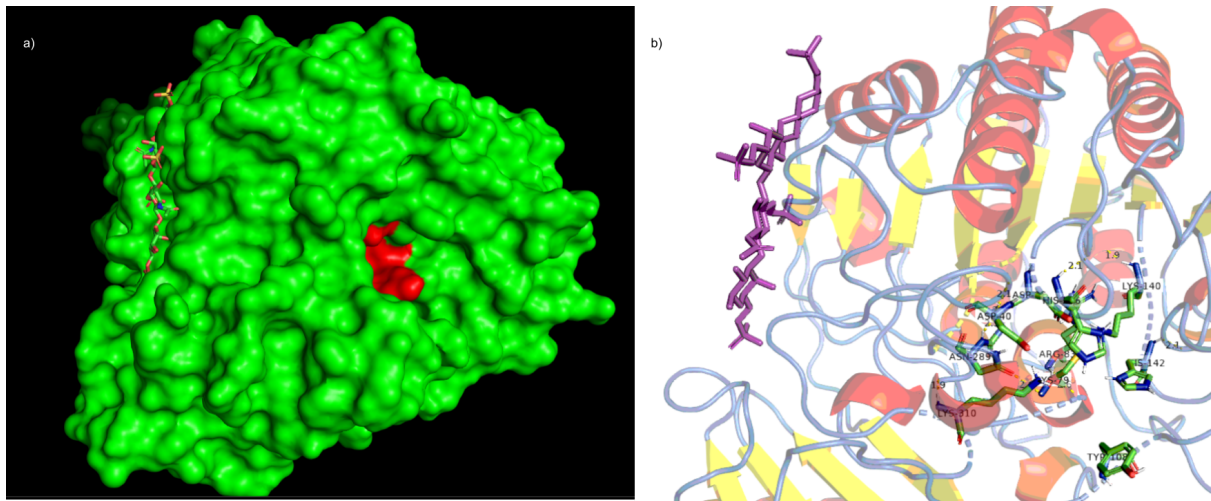


Figura 14. *Docking* molecular da variante G301C gerada pelo Modeller com o sulfato de queratano. a) Imagem com a superfície da proteína. Os resíduos do sítio ativo estão em vermelho. b) Imagem com a proteína em “cartoon”. O substrato está em roxo, os resíduos do sítio ativo estão coloridos. Imagem gerada pelo PyMol. Fonte: Próprio autor.

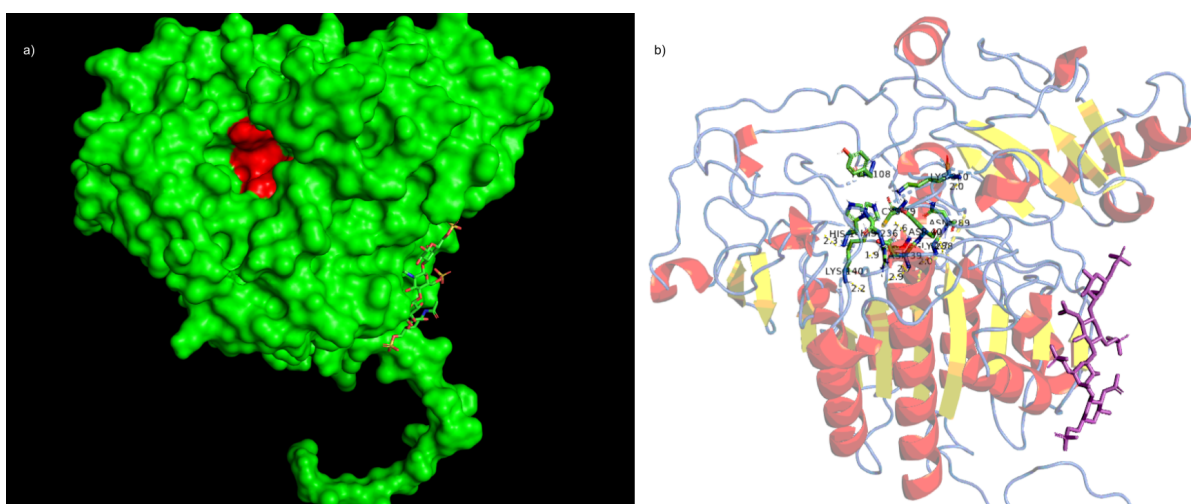


Figura 15. *Docking* molecular da variante G301C gerada pelo I-tasser com o sulfato de condroitina. a) Imagem com a superfície da proteína. Os resíduos do sítio ativo estão em vermelho. b) Imagem com a proteína em “cartoon”. O substrato está em roxo, os resíduos do sítio ativo estão coloridos. Imagem gerada pelo PyMol. Fonte: Próprio autor.

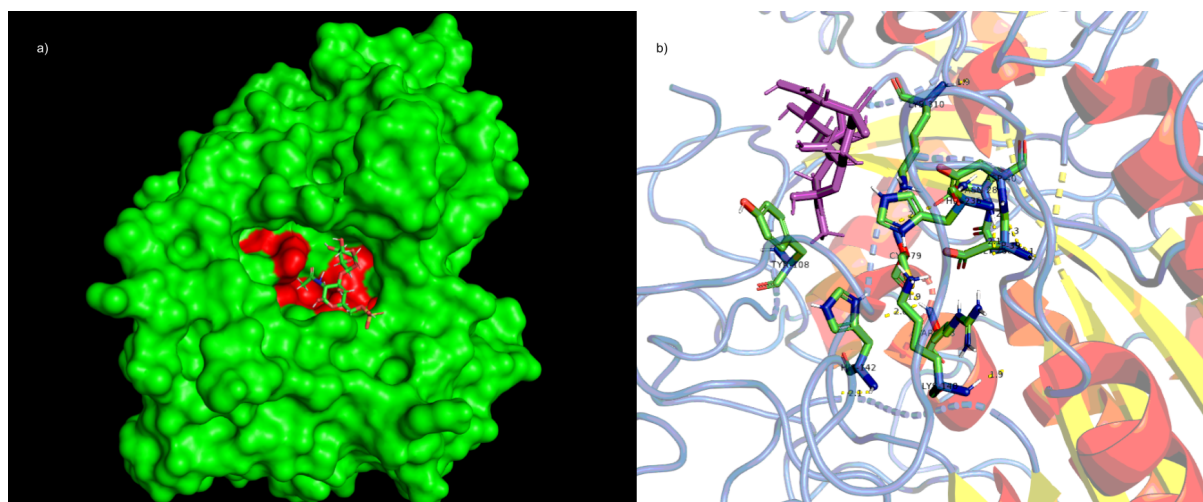
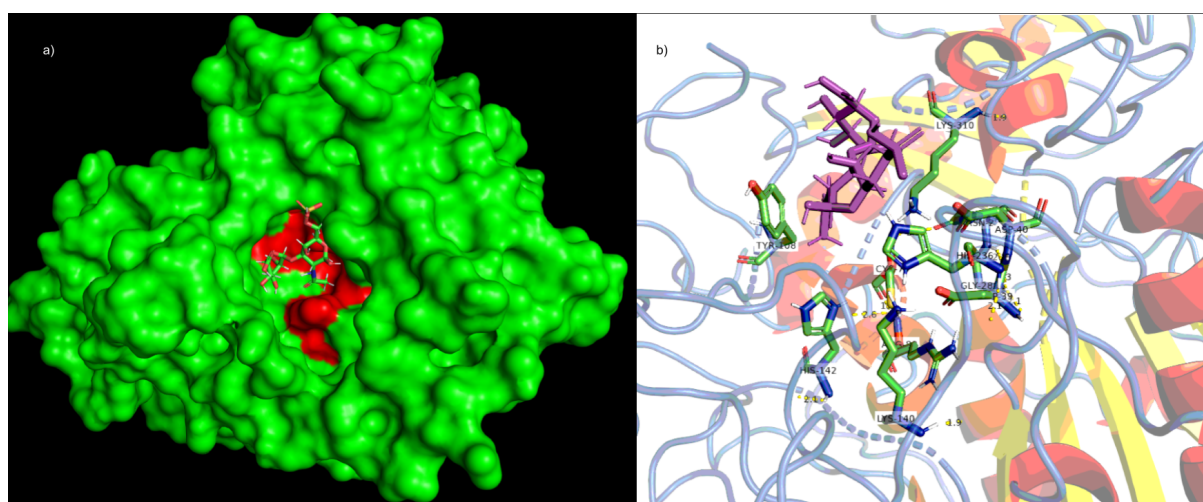


Figura 16. *Docking* molecular da variante G301C gerada pelo Modeller com o sulfato de condroitina. a) Imagem com a superfície da proteína. Os resíduos do sítio ativo estão em vermelho. b) Imagem com a proteína em “cartoon”. O substrato está em roxo, os resíduos do sítio ativo estão coloridos. Imagem gerada pelo PyMol. Fonte: Próprio autor.



3. Em relação à variante p.R386C:

- 3.1. É possível constatar que no modelo gerado pelo I-tasser, o sulfato de queratano se liga ao sítio ativo da proteína, mas no que foi gerado pelo Modeller, a molécula se liga à face oposta à do sítio ativo, assim como aconteceu com a variante p.D45G (Figuras 17 e 18);
- 3.2. O sulfato de condroitina se ligou ao sítio ativo tanto do modelo gerado pelo I-tasser quanto pelo Modeller (Figura 19).

Figura 17. *Docking* molecular da variante R386C gerada pelo I-tasser com o sulfato de queratano. a) Imagem com a superfície da proteína. Os resíduos do sítio ativo estão em vermelho. b) Imagem com a proteína em “*cartoon*”. O substrato está em roxo, os resíduos do sítio ativo estão coloridos. Imagem gerada pelo PyMol. Fonte: Próprio autor.

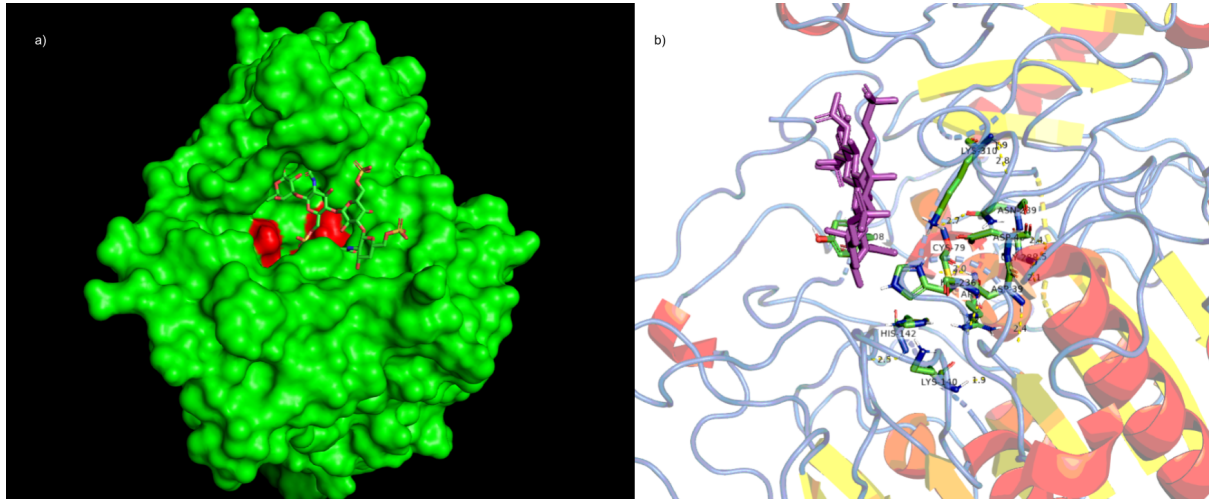


Figura 18. *Docking* molecular da variante R386C gerado pelo Modeller com o sulfato de queratano. a) Imagem com a superfície da proteína. Os resíduos do sítio ativo estão em vermelho. b) Imagem com a proteína em “*cartoon*”. O substrato está em roxo, os resíduos do sítio ativo estão coloridos. Imagem gerada pelo PyMol. Fonte: Próprio autor.

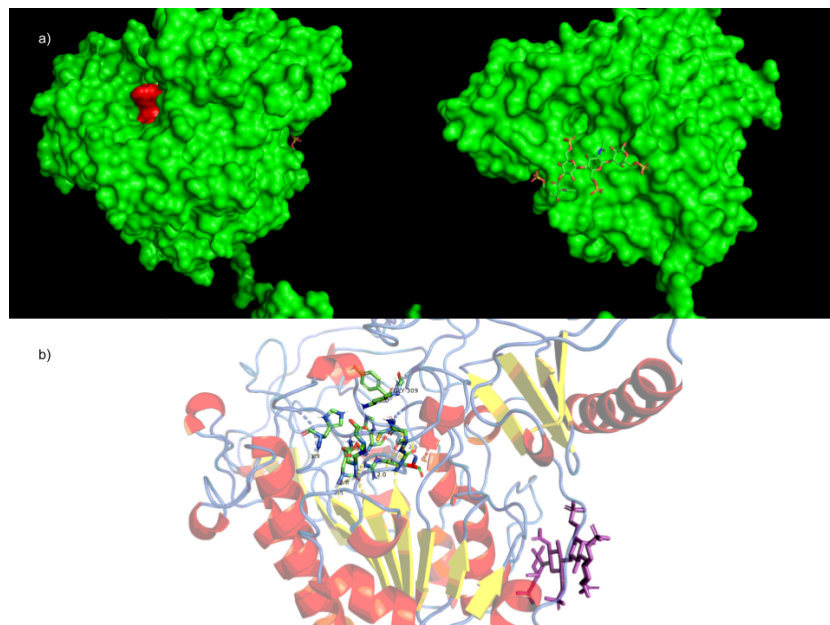
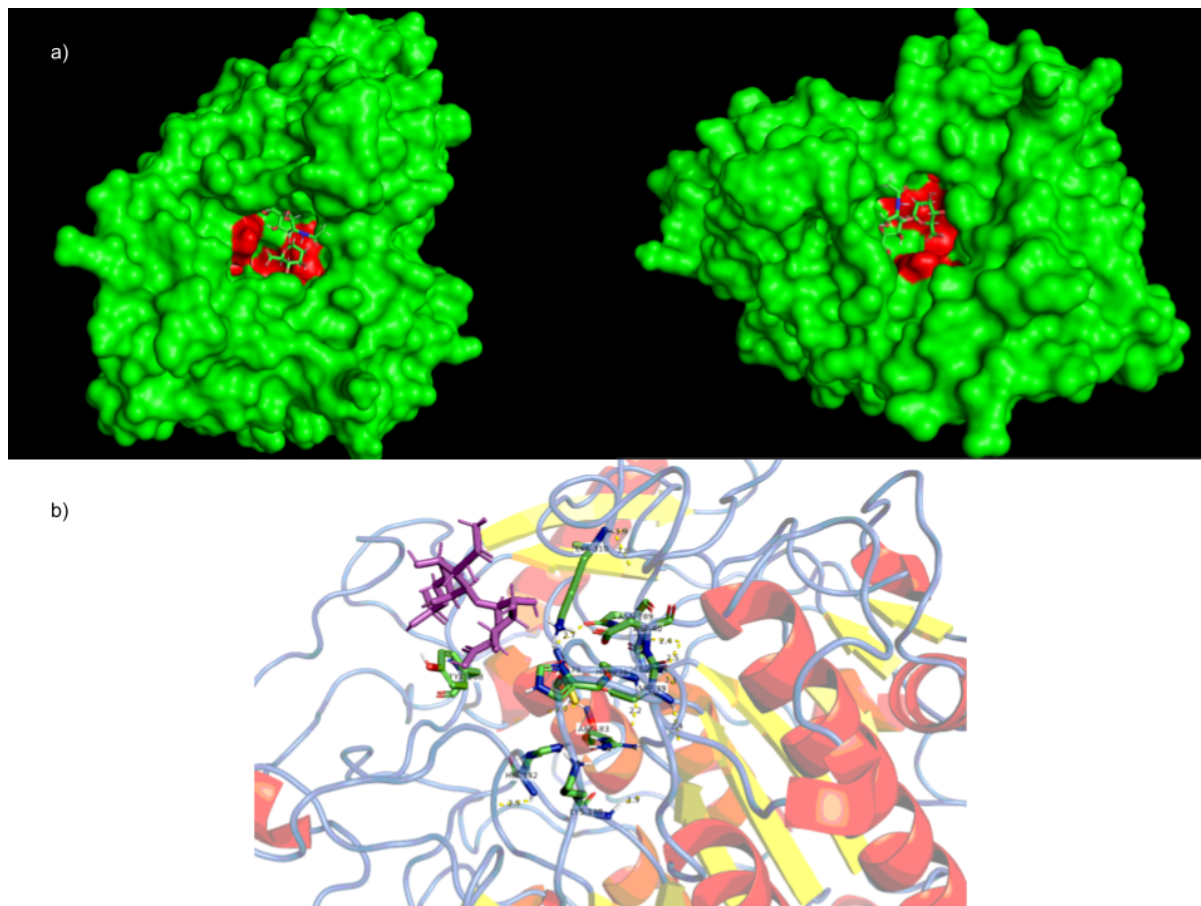


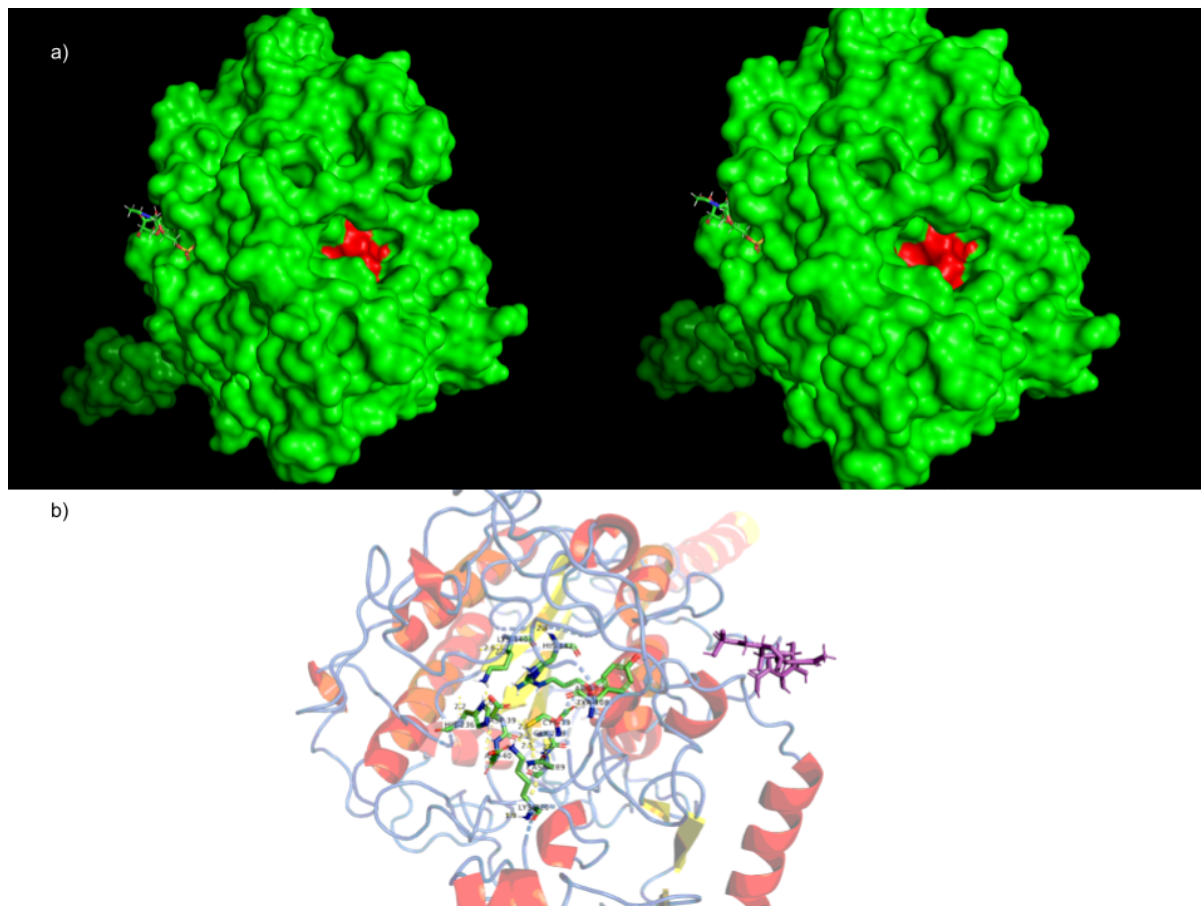
Figura 19. *Docking* molecular da variante R386C gerados pelo I-tasser e pelo Modeller, respectivamente, com o sulfato de condroitina. a) Imagem com a superfície da proteína. Os resíduos do sítio ativo estão em vermelho. b) Imagem com a proteína gerada pelo I-Tasser em “*cartoon*”. O substrato está em roxo, os resíduos do sítio ativo estão coloridos. Imagem gerada pelo PyMol. Fonte: Próprio autor.



4. Em relação à variante p.S341R:

- 4.1. Nem o sulfato de queratano nem o sulfato de condroitina se ligaram ao sítio ativo de nenhum dos modelos. O sulfato de condroitina se ligou à mesma região da proteína, fora do sítio ativo, em ambos os modelos (Figuras 20, 21 e 22).

Figura 22. *Docking* molecular da variante S341R gerados pelo I-tasser e Modeller, respectivamente, com o sulfato de condroitina. a) Imagem com a superfície da proteína. Os resíduos do sítio ativo estão em vermelho. b) Imagem com a proteína gerada pelo I-Tasser em “*cartoon*”. O substrato está em roxo, os resíduos do sítio ativo estão coloridos. Imagem gerada pelo PyMol. Fonte: Próprio autor.



5. Em relação à variante p.V239F:

5.1. O sulfato de queratano não se ligou ao sítio ativo da proteína em nenhum dos modelos, mas o sulfato de condroitina se ligou ao sítio ativo de ambos, assim como aconteceu com a variante p.G301C (Figuras 23, 24, 25 e 26).

Figura 23. *Docking* molecular da variante V239F gerado pelo I-tasser com o sulfato de queratano. a) Imagem com a superfície da proteína. Os resíduos do sítio ativo estão em vermelho. b) Imagem com a proteína em “cartoon”. O substrato está em roxo, os resíduos do sítio ativo estão coloridos. Imagem gerada pelo PyMol. Fonte: Próprio autor.

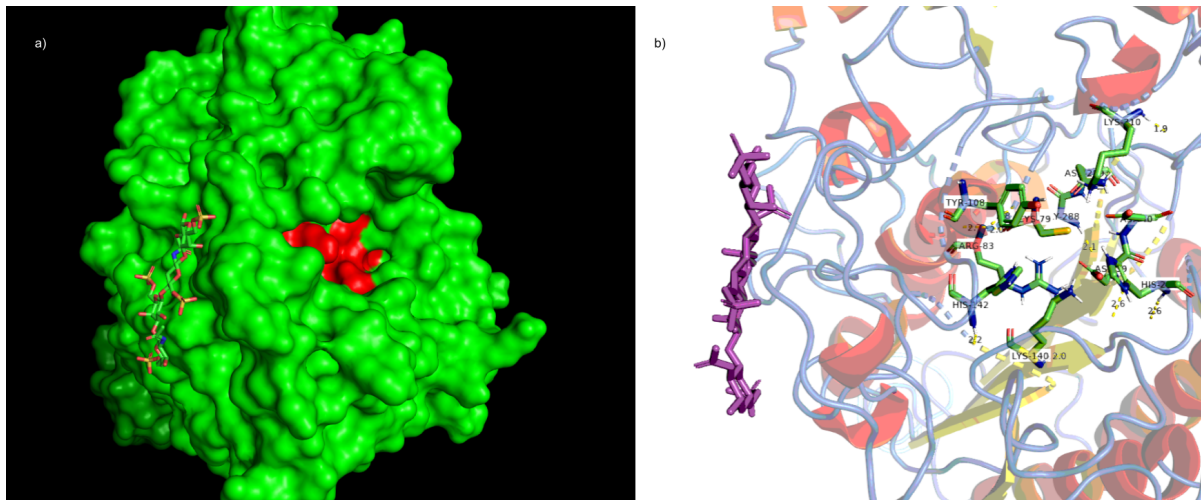


Figura 24. *Docking* molecular da variante V239F gerado pelo Modeller com o sulfato de queratano. a) Imagem com a superfície da proteína. Os resíduos do sítio ativo estão em vermelho. b) Imagem com a proteína em “cartoon”. O substrato está em roxo, os resíduos do sítio ativo estão coloridos. Imagem gerada pelo PyMol. Fonte: Próprio autor.

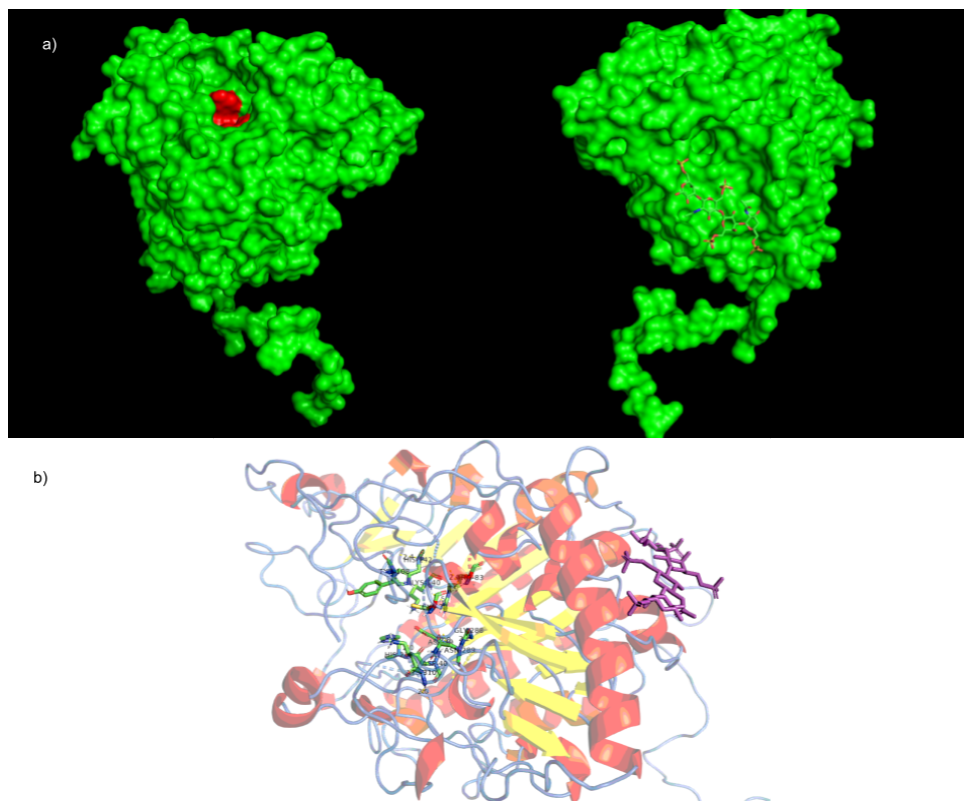


Figura 25. *Docking* molecular da variante V239F gerada pelo I-tasser com o sulfato de condroitina. a) Imagem com a superfície da proteína. Os resíduos do sítio ativo estão em vermelho. b) Imagem com a proteína em “*cartoon*”. O substrato está em roxo, os resíduos do sítio ativo estão coloridos. Imagem gerada pelo PyMol. Fonte: Próprio autor.

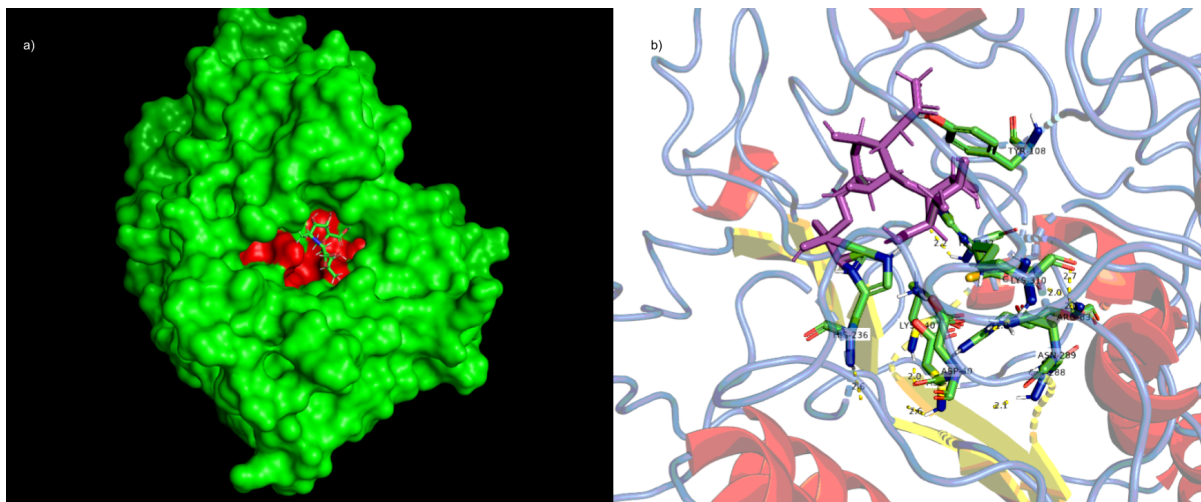
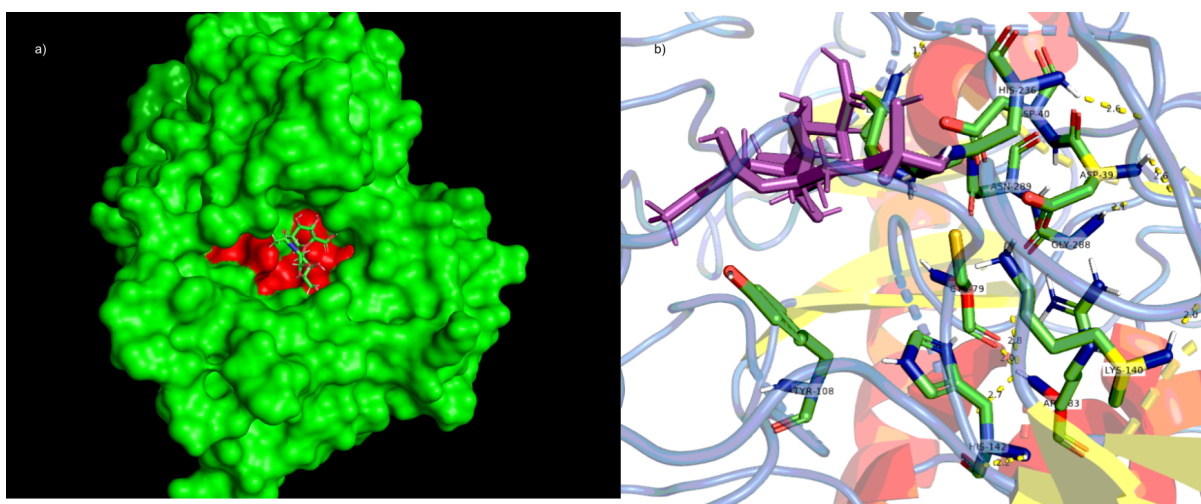


Figura 26. *Docking* molecular da variante V239F gerada pelo Modeller com o sulfato de condroitina. a) Imagem com a superfície da proteína. Os resíduos do sítio ativo estão em vermelho. b) Imagem com a proteína em “*cartoon*”. O substrato está em roxo, os resíduos do sítio ativo estão coloridos. Imagem gerada pelo PyMol. Fonte: Próprio autor.



As informações supracitadas podem ser sumarizadas da seguinte forma:

Tabela 5. Modelos em que os substratos se ligaram ao sítio ativo.

Variante	Sulfato de Queratano		Sulfato de Condroitina	
	I-Tasser	Modeller	I-Tasser	Modeller
p.D45G	X			
p.G301C			X	X
p.R386C	X		X	X
p.S341R				
p.V239F			X	X

5 DISCUSSÃO

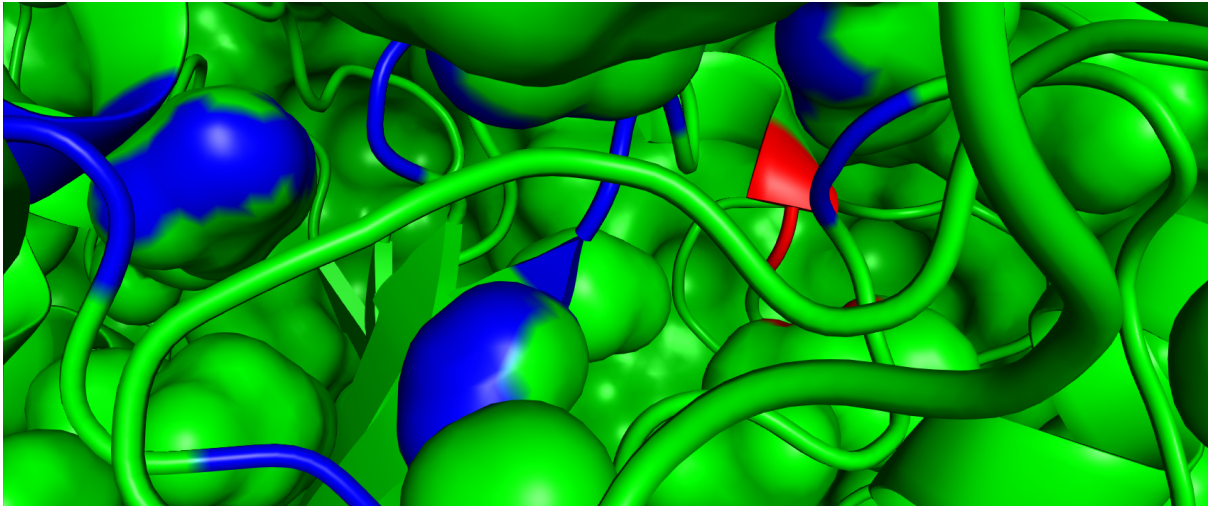
Rivera-Colón *et al.* (2012) apresentou em seu trabalho uma lista com diversas mutações do tipo Missense causadoras de doenças do gene GALNS, identificando desde as suas categorias (enterradas, na superfície da proteína ou no sítio ativo) até seus efeitos na proteína. Dentre essas mutações estão: G301C, R386C, S341R e V239F (Tabela 4). Assim, a mutação p.D45G é a única presente nos pacientes do estado da Paraíba não apresentada neste trabalho.

Tabela 6. Mutações do GALNS identificadas em pacientes do estado da Paraíba e seu efeito na proteína. Adaptado de Rivera-Colón *et al.* (2012).

Posição	Superfície atômica acessível, Å ²	Efeito na GALNS	Fenótipo da doença	Categoria da mutação
V239F	0,0	Enterrada; sem espaço para uma cadeia lateral maior	Severo	Enterrada
G301C	3,7	Enterrada; pouco espaço para uma cadeia lateral maior	Severo	Enterrada
S341R	0,8	Pouco espaço para uma cadeia lateral maior	Severo	Enterrada
R386C	1,1	Perturbação do núcleo hidrofóbico	Severo	Enterrada

A mutação D45G foi relatada pela primeira vez no Brasil, mais especificamente no Paciente 12 apresentado na Tabela 2 (DOS SANTOS-LOPES, 2021). Pela análise visual dos modelos da variante D45G, é possível observar que a região da mutação se encontra abaixo da superfície da proteína, só sendo possível visualizá-la por dentro da estrutura (Figura 27). Isso indica que, assim como todas as outras mutações abordadas nesse trabalho, ela está dentro da categoria “Enterrada”.

Figura 27. Região da mutação da variante D45G gerada pelo Modeller. Região da mutação em vermelho. Resíduos do sítio ativo em azul. Imagem gerada pelo PyMol. Fonte: Próprio autor.



Em relação aos *dockings* moleculares, não é possível correlacionar a forma que os substratos se ligam ou não às variantes com o fenótipo da doença, uma vez que todas elas possuem um fenótipo severo, no entanto, a realização de novos estudos com essa metodologia em variantes com fenótipos mais leves pode tornar possível essa comparação e, assim, se tornam uma perspectiva.

É importante perceber também que, mesmo nas variantes em que os substratos se ligam ao sítio ativo da enzima, essa ligação se dá de forma diferente de como ocorre no tipo selvagem (Figuras 8 e 9), isto é, a posição dos substratos em relação ao sítio ativo é diferente em todas essas variantes quando comparadas com a posição dos substratos na GALNS selvagem. Esses resultados podem indicar que, mesmo que o sítio ativo ainda possua afinidade com os substratos nessas variantes, a atividade enzimática ainda pode ser afetada. Esse fato pode se relacionar, inclusive, com o fenótipo severo dessas mutações, uma vez que, estando a atividade catalítica comprometida para ambos os substratos para todas essas variantes, o acúmulo de GAGs é, logicamente, maior. Mais estudos que usem essa metodologia em variantes com fenótipos mais leves também podem tornar possível uma comparação com esses resultados, pois com esses dados, é possível se analisar se nas mutações que possuem esses fenótipos mais leves os substratos se ligam de forma diferente com as variantes do que com a enzima selvagem ou não, por exemplo, ou se elas sequer chegam a se ligar.

Por fim, existem diversas formas em que mutações podem afetar a atividade enzimática, seja pelo aumento da instabilidade da proteína, pela criação de novos sítios de ligação, pela interrupção da transcrição através da geração de um códon de parada, entre outros. Os efeitos na proteína GALNS das mutações V239F, G301C, S341R e R386C foram descritos por Rivera-Colón *et al.* (2012) e estão apresentados na Tabela 6. Mais estudos também são necessários para se avaliar se esses efeitos das mutações são existentes nos modelos gerados no presente trabalho, e se esses efeitos são os responsáveis pelas alterações das formas de ligação com os substratos com as variantes em relação à proteína selvagem, além de também serem essenciais em determinar quais foram os efeitos gerados pela mutação na variante D45G, que não está presente no estudo de Rivera-Colón *et al.* (2012).

6 CONCLUSÃO

Em síntese, esse trabalho forneceu novos *insights* acerca das variantes mutantes da enzima GALNS presentes nos pacientes com MPS IVA do estado da Paraíba. Ao realizar a modelagem por homologia dessas variantes, foi possível estudar os efeitos dessas mutações na estrutura e na função da proteína. Os *dockings* moleculares permitiram a investigação das interações das proteínas mutantes com os substratos da proteína GALNS, fornecendo informações valiosas sobre os efeitos das mutações na capacidade da enzima de se ligar a essas moléculas. Esse conhecimento pode ser aplicado no desenvolvimento de novas terapias para a MPS IVA. Os resultados trazem novas informações e também destacam a importância potencial da mutação D45G, relatada pela primeira vez no Brasil e pouco estudada na literatura. Apesar dos valiosos *insights* obtidos nesse trabalho, também foram encontrados diversos desafios e limitações que precisam ser abordados em estudos futuros. Por exemplo, o uso de dois programas diferentes de modelagem por homologia rendeu resultados diferentes em termos de interações das proteínas mutantes com substratos. Isso implica que mais pesquisas são necessárias para um total entendimento dos fatores que podem afetar a precisão da modelagem por homologia, e que métodos mais robustos de previsão da estrutura das proteínas devem ser testados. No geral, essa pesquisa contribuiu para um crescente corpo de conhecimento sobre a enzima GALNS e suas variantes mutantes e prepara o terreno para uma investigação mais aprofundada sobre sua estrutura e função.

REFERÊNCIAS

ALGAHIM, Mohamed F.; ALMASSI, G. Hossein. Current and emerging management options for patients with Morquio A syndrome. **Therapeutics and clinical risk management**, v. 9, p. 45, 2013.

AL-JASMI, Fatma A. et al. Prevalence and novel mutations of lysosomal storage disorders in United Arab Emirates. In: **JIMD Reports-Volume 10**. Springer, Berlin, Heidelberg, 2012. p. 1-9.

BAKER, Matthew L. et al. Ab initio modeling of the herpesvirus VP26 core domain assessed by CryoEM density. **PLoS computational biology**, v. 2, n. 10, p. e146, 2006.

BHATTACHARYA, Debswapna; CHENG, Jianlin. 3Drefine: Consistent protein structure refinement by optimizing hydrogen bonding network and atomic-level energy minimization. **Proteins: Structure, Function, and Bioinformatics**, v. 81, n. 1, p. 119-131, 2013a.

BHATTACHARYA, Debswapna; CHENG, Jianlin. i3Drefine software for protein 3D structure refinement and its assessment in CASP10. **PloS one**, v. 8, n. 7, p. e69648, 2013b.

BHATTACHARYA, Debswapna et al. 3Drefine: an interactive web server for efficient protein structure refinement. **Nucleic acids research**, v. 44, n. W1, p. W406-W409, 2016.

BOCHERNITSAN, Aline Nemetz. Mucopolissacaridose IVA: análise molecular e caracterização de haplótipos intragênicos no gene Galns. 2015.

BOCHERNITSAN, Aline Nemetz et al. Spectrum of GALNS mutations and haplotype study in Brazilian patients with Mucopolysaccharidosis type IVA. **Meta Gene**, v. 16, p. 77-84, 2018.

BOWIE, James U.; LUTHY, Roland; EISENBERG, David. A method to identify protein sequences that fold into a known three-dimensional structure. **Science**, v. 253, n. 5016, p. 164-170, 1991.

CACIOTTI, Anna et al. Mis-splicing of the GALNS gene resulting from deep intronic mutations as a cause of Morquio a disease. **BMC medical genetics**, v. 19, n. 1, p. 183, 2018.

CAMELIER, Marli V. et al. Practical and reliable enzyme test for the detection of mucopolysaccharidosis IVA (Morquio Syndrome type A) in dried blood samples. **Clinica chimica acta**, v. 412, n. 19-20, p. 1805-1808, 2011.

CATARZI, S. et al. Morquio A syndrome due to maternal uniparental isodisomy of the telomeric end of chromosome 16. **Molecular genetics and metabolism**, v. 105, n. 3, p. 438-442, 2012.

CHARROW, Joel et al. Diagnostic evaluation, monitoring, and perioperative management of spinal cord compression in patients with Morquio syndrome. **Molecular genetics and metabolism**, v. 114, n. 1, p. 11-18, 2015.

CHINEN, Yasutsugu et al. Long-term therapeutic efficacy of allogenic bone marrow transplantation in a patient with mucopolysaccharidosis IVA. **Molecular genetics and metabolism reports**, v. 1, p. 31-41, 2014.

COUTINHO, Maria Francisca; LACERDA, Lucia; ALVES, Sandra. Glycosaminoglycan storage disorders: a review. **Biochemistry research international**, v. 2012, 2012.

DA SILVA, Waldeyr MC et al. Graph databases in molecular biology. In: **Brazilian Symposium on Bioinformatics**. Springer, Cham, 2018. p. 50-57.

DIAS, Raquel; DE AZEVEDO, Jr; WALTER, F. Molecular docking algorithms. **Current drug targets**, v. 9, n. 12, p. 1040-1047, 2008.

DIETER, Tatiana et al. Common N-acetylgalactosamine-6-sulfate sulfatase (GALNS) exon mutations in Brazilian patients with mucopolysaccharidosis IVA (MPS IVA). **Genetics and Molecular Biology**, v. 30, n. 3, p. 524-528, 2007.

DINIZ, W.; CANDURI, Fernanda. Bioinformatics: an overview and its applications. **Genet Mol Res**, v. 16, n. 1, 2017.

DOS SANTOS-LOPES, Simone Silva et al. Demographic, clinical, and ancestry characterization of a large cluster of mucopolysaccharidosis IV A in the Brazilian Northeast region. **American Journal of Medical Genetics Part A**, 2021.

DÛNG, Vũ Chí et al. Mucopolysaccharidosis IVA: correlation between genotype, phenotype and keratan sulfate levels. **Molecular genetics and metabolism**, v. 110, n. 1-2, p. 129-138, 2013.

FAN, Jiyu; FU, Ailing; ZHANG, Le. Progress in molecular docking. **Quantitative Biology**, v. 7, n. 2, p. 83-89, 2019.

FISER, András. Comparative protein structure modelling. In: **From protein structure to function with bioinformatics**. Springer, Dordrecht, 2017. p. 91-134.

FUKUDA, Seiji et al. Mucopolysaccharidosis IVA: submicroscopic deletion of 16q24.3 and a novel R386C mutation of N-acetylgalactosamine-6-sulfate sulfatase gene in a classical Morquio disease. **Human mutation**, v. 7, n. 2, p. 123-134, 1996.

GAUTHIER, Jeff et al. A brief history of bioinformatics. **Briefings in bioinformatics**, v. 20, n. 6, p. 1981-1996, 2019.

GIUGLIANI, Roberto et al. Relative frequency and estimated minimal frequency of Lysosomal Storage Diseases in Brazil: Report from a Reference Laboratory. **Genetics and molecular biology**, v. 40, n. 1, p. 31-39, 2017.

GUEDES, Isabella A.; DE MAGALHÃES, Camila S.; DARDENNE, Laurent E. Receptor–ligand molecular docking. **Biophysical reviews**, v. 6, n. 1, p. 75-87, 2014.

HAGEN, Joel B. The origins of bioinformatics. **Nature Reviews Genetics**, v. 1, n. 3, p. 231-236, 2000.

HENDRIKSZ, Christian J. et al. Clinical overview and treatment options for non-skeletal manifestations of mucopolysaccharidosis type IVA. **Journal of inherited metabolic disease**, v. 36, n. 2, p. 309-322, 2013.

HERBERTSON, E. C. et al. Third-line antiretroviral therapy in a Nigerian clinic: case series. **Archives of Medicine**, v. 9, n. 4, p. 0-0, 2017.

HUNT, Lois T. Margaret Oakley Dayhoff 1925-1983. **Bulletin of Mathematical Biology**, v. 46, n. 4, p. 467-472, 1984.

KATO, Zenichiro et al. A novel common missense mutation G301C in the N-acetylgalactosamine-6-sulfate sulfatase gene in mucopolysaccharidosis IVA. **Human genetics**, v. 101, n. 1, p. 97-101, 1997.

KHAN, Shaikat A. et al. Epidemiology of mucopolysaccharidoses. **Molecular genetics and metabolism**, v. 121, n. 3, p. 227-240, 2017.

KOPP, Peter. Application of Molecular Biology and Genetics in Endocrinology. **Endocrine Practice**, v. 13, n. 5, p. 534-541, 2007.

KUFAREVA, Irina; ABAGYAN, Ruben. Methods of protein structure comparison. In: **Homology modeling**. Humana Press, 2011. p. 231-257.

KULSHRESHTHA, Shweta et al. Computational approaches for predicting mutant protein stability. **Journal of computer-aided molecular design**, v. 30, n. 5, p. 401-412, 2016.

LARADI, S. et al. Mucopolysaccharidosis type IV: N-acetylgalactosamine-6-sulfatase mutations in Tunisian patients. **Molecular Genetics and metabolism**, v. 87, n. 3, p. 213-218, 2006.

LASKOWSKI, R.A.; MACARTHUR, M.W.; MOSS D.S.; THORNTON, J.M. PROCHECK: A program to check the stereochemical quality of protein structures. **Journal of Applied Crystallography**, v.26, p.283-291, 1998.

LASKOWSKI, Roman A. et al. AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. **Journal of biomolecular NMR**, v. 8, n. 4, p. 477-486, 1996.

LEE, Jooyoung; FREDDOLINO, Peter L.; ZHANG, Yang. Ab initio protein structure prediction. In: **From protein structure to function with bioinformatics**. Springer, Dordrecht, 2017. p. 3-35.

LENSINK, Marc F.; VELANKAR, Sameer; WODAK, Shoshana J. Modeling protein–protein and protein–peptide complexes: CAPRI 6th edition. **Proteins: Structure, Function, and Bioinformatics**, v. 85, n. 3, p. 359-377, 2017.

LIANG, Xin et al. Molecular Computing and Bioinformatics. **Molecules**, v. 24, n. 13, 2019.

LIN, Hsiang-Yu et al. Cardiac features and effects of enzyme replacement therapy in Taiwanese patients with Mucopolysaccharidosis IVA. **Orphanet journal of rare diseases**, v. 13, n. 1, p. 148, 2018.

LUTHY, R.; JAMES, U. B.; EISENBERG, D. Assessment of protein models with three-dimensional profiles. **Nature**, v.356, p.83-85, 1992.

LÜTHY, Roland; BOWIE, James U.; EISENBERG, David. Assessment of protein models with three-dimensional profiles. **Nature**, v. 356, n. 6364, p. 83-85, 1992.

MACINDOE, Gary et al. HexServer: an FFT-based protein docking server powered by graphics processors. **Nucleic acids research**, v. 38, n. suppl_2, p. W445-W449, 2010.

MAIOROV, Vladimir N.; CRIPPEN, Gordon M. Size-independent comparison of protein three-dimensional structures. **Proteins: Structure, Function, and Bioinformatics**, v. 22, n. 3, p. 273-283, 1995.

MARTÍ-RENOM, Marc A. et al. Comparative protein structure modeling of genes and genomes. **Annual review of biophysics and biomolecular structure**, v. 29, n. 1, p. 291-325, 2000.

MORRIS, Garrett M.; LIM-WILBY, Marguerita. Molecular docking. In: **Molecular modeling of proteins**. Humana Press, 2008. p. 365-382.

MORRONE, A. et al. Molecular testing of 163 patients with Morquio A (Mucopolysaccharidosis IVA) identifies 39 novel GALNS mutations. **Molecular genetics and metabolism**, v. 112, n. 2, p. 160-170, 2014.

OLARTE-AVELLANEDA, Sergio et al. Computational analysis of human N-acetylgalactosamine-6-sulfate sulfatase enzyme: an update in genotype–phenotype correlation for Morquio A. **Molecular biology reports**, v. 41, n. 11, p. 7073-7088, 2014.

PROSDOCIMI, Francisco et al. Bioinformática: manual do usuário. **Biotecnologia Ciência & Desenvolvimento**, v. 29, p. 12-25, 2002.

RAMACHANDRAN, G.N.; RAMAKRISHNAN C.; SASISEKHARAN, V. Stereochemistry of Polypeptide Chain Conformations. **Journal of Molecular Biology**, v.7, p.95-99, 1963.

RIVERA-COLÓN, Yadilette et al. The structure of human GALNS reveals the molecular basis for mucopolysaccharidosis IV A. **Journal of molecular biology**, v. 423, n. 5, p. 736-751, 2012.

ROY, Ambrish; KUCUKURAL, Alper; ZHANG, Yang. I-TASSER: a unified platform for automated protein structure and function prediction. **Nature protocols**, v. 5, n. 4, p. 725-738, 2010.

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

SANTOS, Silvana et al. As causas da deficiência física em municípios do nordeste brasileiro e estimativa de custos de serviços especializados. **Ciência & Saúde Coletiva**, v. 19, p. 559-568, 2014.

SAWAMOTO, Kazuki et al. Mucopolysaccharidosis IVA: diagnosis, treatment, and management. **International Journal of Molecular Sciences**, v. 21, n. 4, p. 1517, 2020.

THANKACHAN, Shyny; KULKARNI, Bhagavan G. Exploration of Ancient Wisdom of Clinical Genetics in Ayurvedic Classics. **Current Aspects in Pharmaceutical Research and Development Vol. 4**, p. 69-78, 2021.

The Human Gene Mutation Database (HGMD®). Disponível em: <<http://www.hgmd.cf.ac.uk/ac/gene.php?gene=GALNS>>. Acesso em: 23 de Setembro de 2021.

TOMATSU, Shunji et al. Identification of a common mutation in mucopolysaccharidosis IVA: correlation among genotype, phenotype, and keratan sulfate. **Journal of human genetics**, v. 49, n. 9, p. 490-494, 2004.

TOMATSU, Shunji et al. Mucopolysaccharidosis IVA: identification of a common missense mutation I113F in the N-acetylgalactosamine-6-sulfate sulfatase gene. **American journal of human genetics**, v. 57, n. 3, p. 556, 1995.

TOMATSU, Shunji et al. Mutation and polymorphism spectrum of the GALNS gene in mucopolysaccharidosis IVA (Morquio A). **Human mutation**, v. 26, n. 6, p. 500-512, 2005.

TORRIERI, Érico. **Análise Estrutural de Mutações na Enzima GALNS associadas à Mucopolissacaridose IVA utilizando a Técnica de Modelagem Comparativa**. Tese de Doutorado. Universidade de São Paulo.

VRIEND, G. WHAT IF: A molecular modelling and drug design program. **Journal of Molecular Graphics**, v.8, p.52-56, 1990.

WEBB, Benjamin; SALI, Andrej. Protein structure modeling with MODELLER. In: **Functional Genomics**. Humana Press, New York, NY, 2017. p. 39-54.

WELLER, Mathias et al. Consanguineous unions and the burden of disability: A population-based study in communities of Northeastern Brazil. **American Journal of Human Biology**, v. 24, n. 6, p. 835-840, 2012.

WOOD, Timothy C. et al. Diagnosing mucopolysaccharidosis IVA. **Journal of inherited metabolic disease**, v. 36, n. 2, p. 293-307, 2013.

WU, Sitao; SKOLNICK, Jeffrey; ZHANG, Yang. Ab initio modeling of small proteins by iterative TASSER simulations. **BMC biology**, v. 5, n. 1, p. 1-10, 2007.

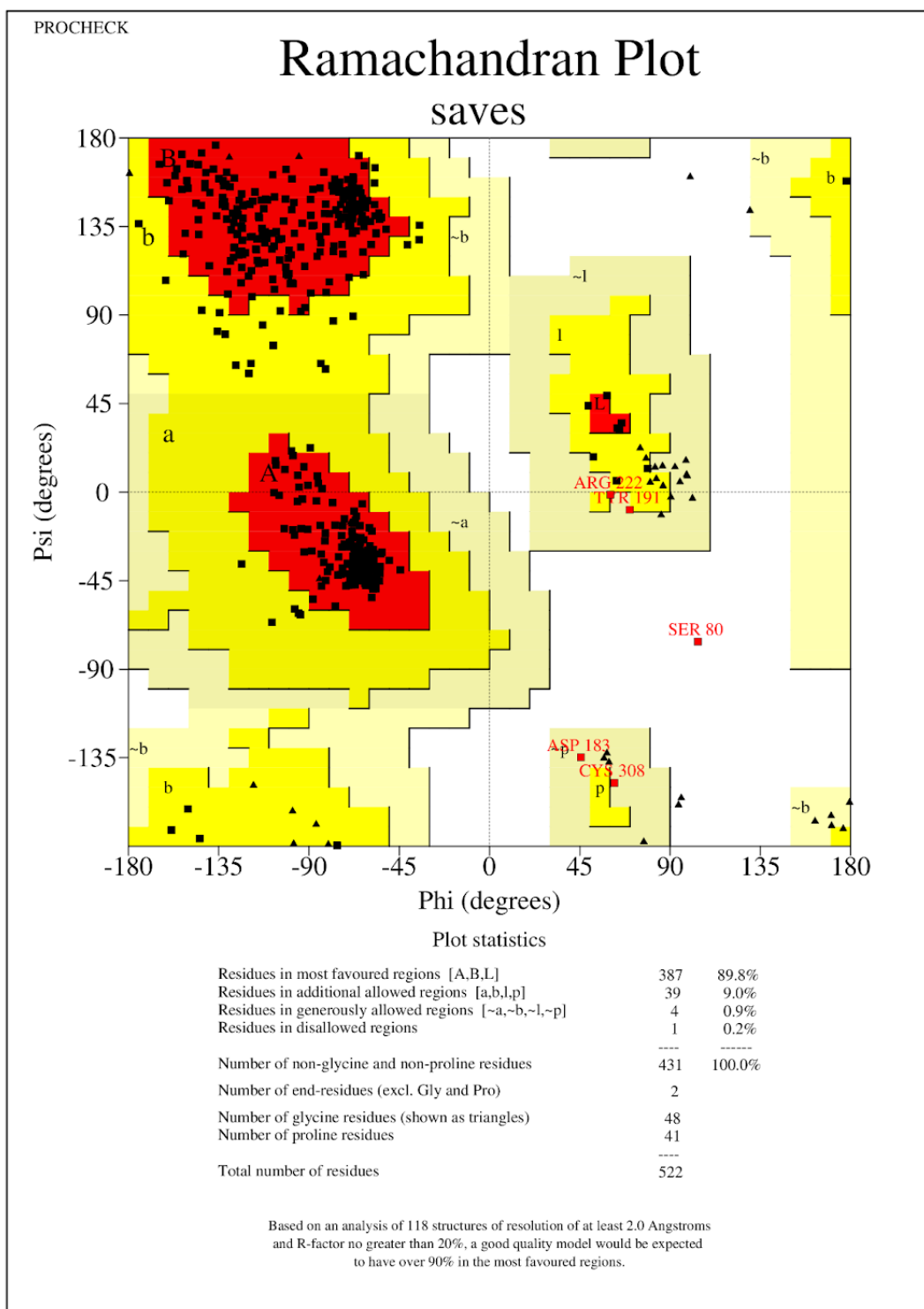
YANG, Jianyi; ZHANG, Yang. I-TASSER server: new development for protein structure and function predictions. **Nucleic acids research**, v. 43, n. W1, p. W174-W181, 2015a.

YANG, Jianyi et al. The I-TASSER Suite: protein structure and function prediction. **Nature methods**, v. 12, n. 1, p. 7-8, 2015b.

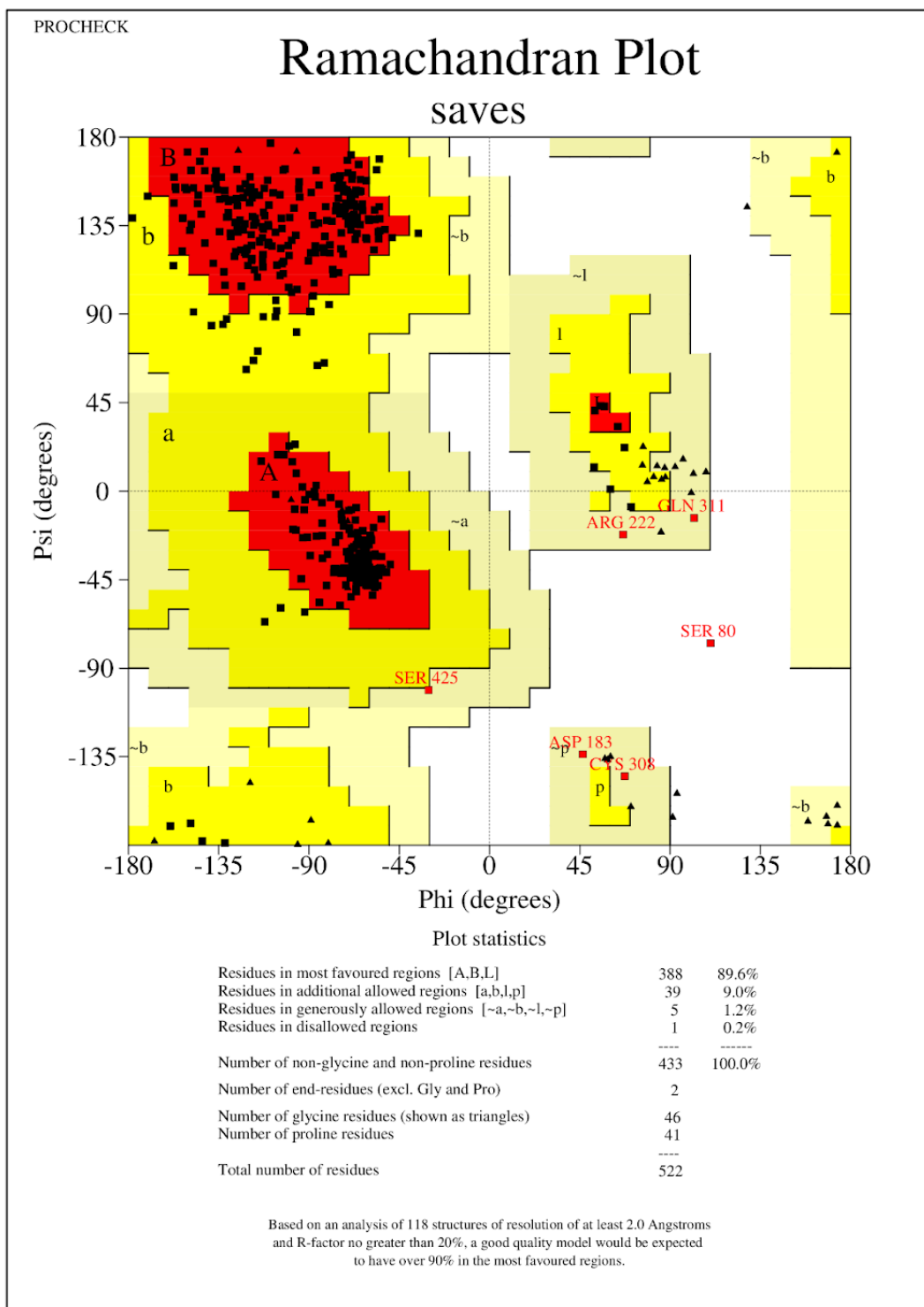
ZHANG, Yang; SKOLNICK, Jeffrey. TM-align: a protein structure alignment algorithm based on the TM-score. **Nucleic acids research**, v. 33, n. 7, p. 2302-2309, 2005.

ZHOU, Xiaoguang et al. The next-generation sequencing technology and application. **Protein & cell**, v. 1, n. 6, p. 520-536, 2010.

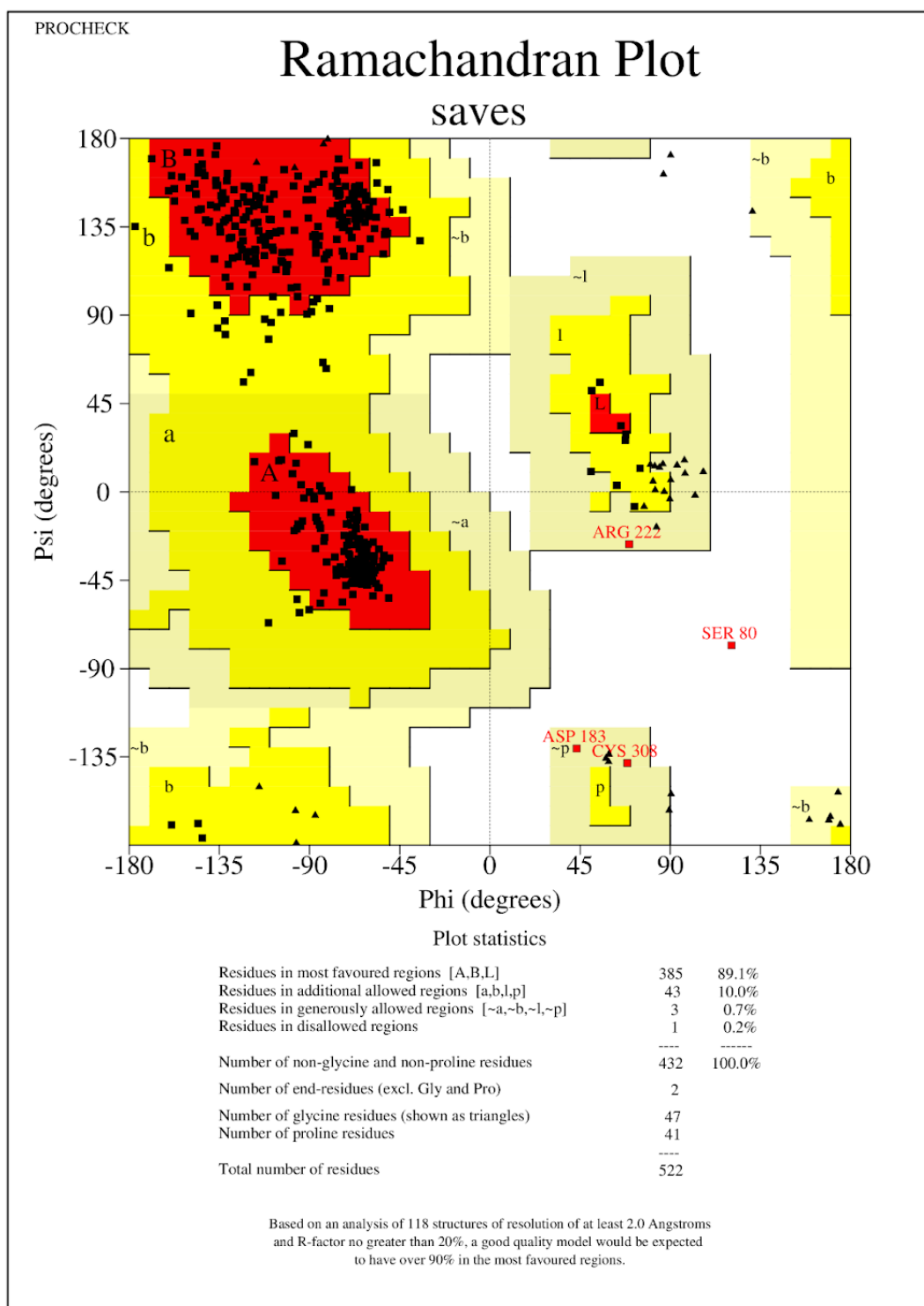
APÊNDICE A - GRÁFICO DE RAMACHANDRAN DA VARIANTE p.D45G
GERADA PELO MODELLER



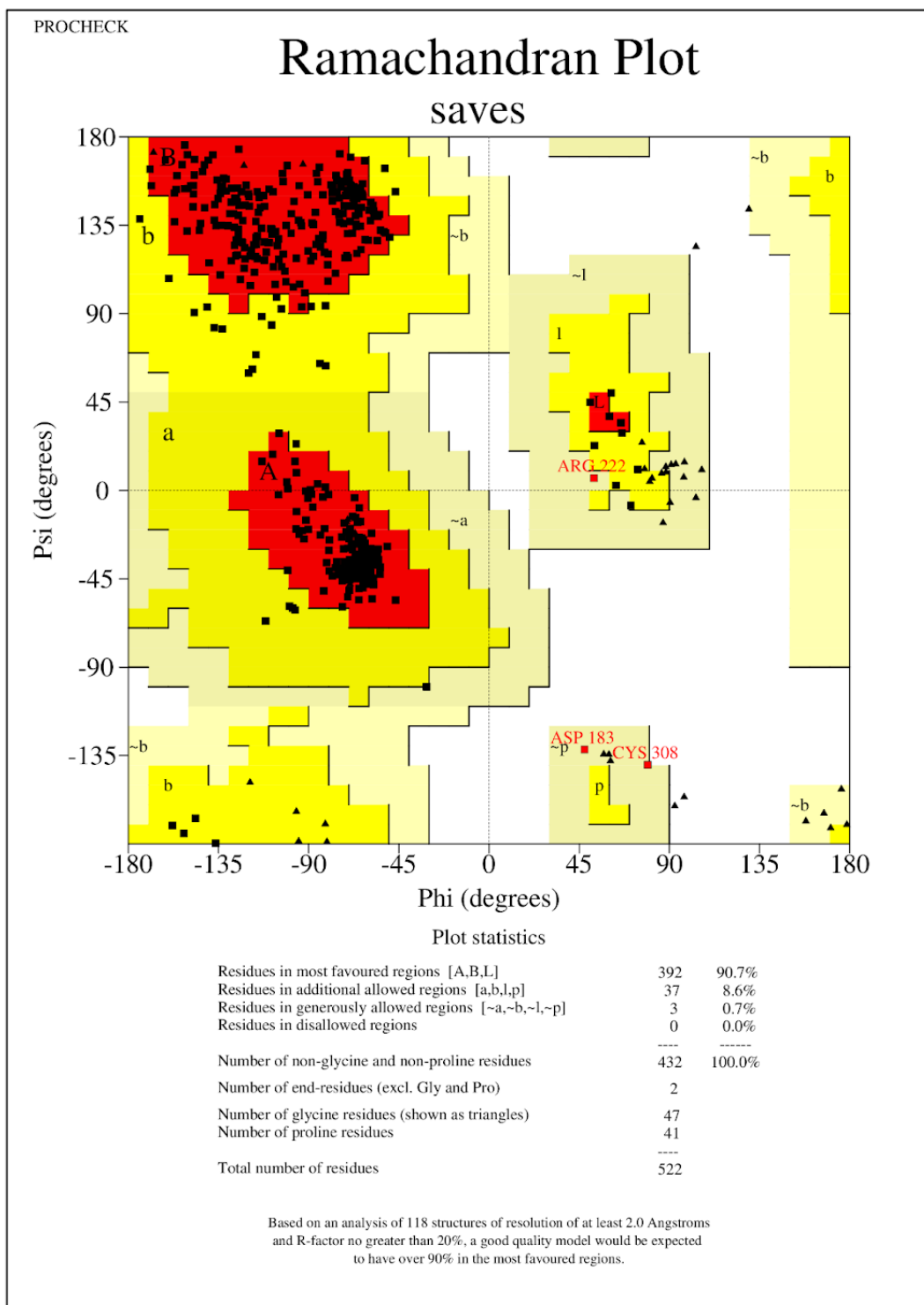
APÊNDICE B - GRÁFICO DE RAMACHANDRAN DA VARIANTE p.G301C
GERADA PELO MODELLER



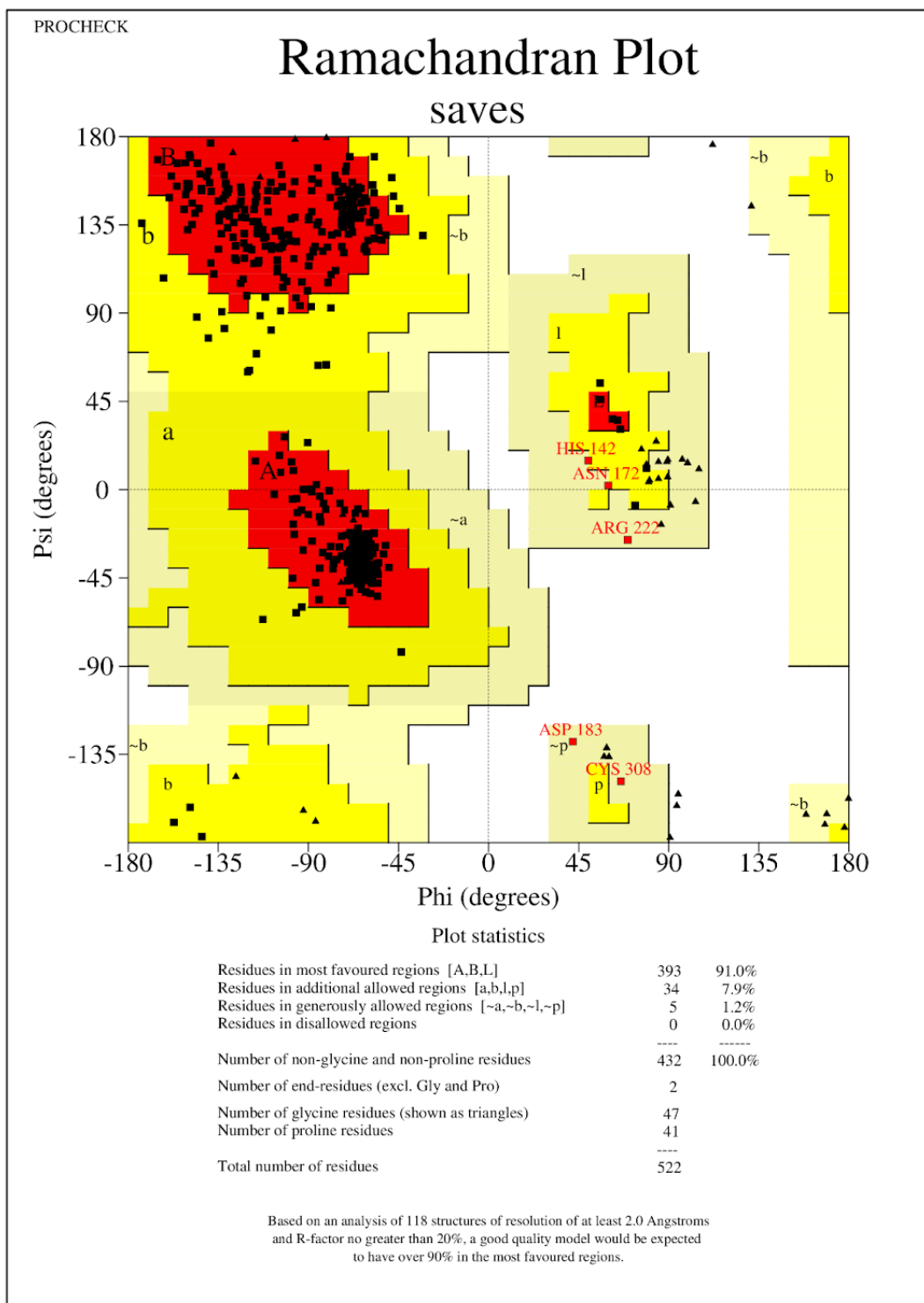
APÊNDICE C - GRÁFICO DE RAMACHANDRAN DA VARIANTE p.R386C
GERADA PELO MODELLER



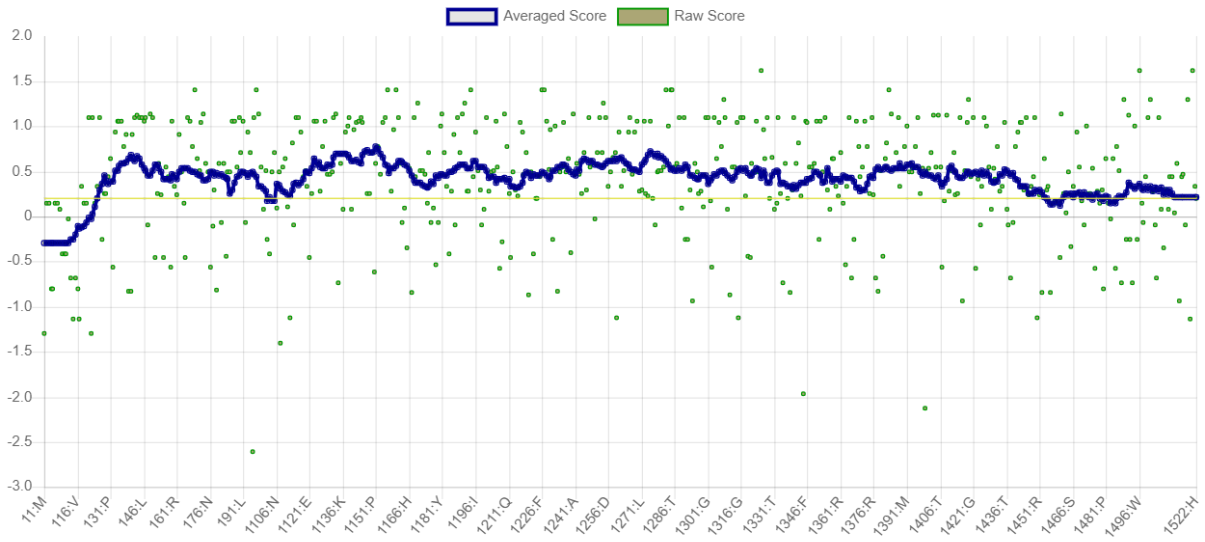
APÊNDICE D - GRÁFICO DE RAMACHANDRAN DA VARIANTE p.S341R
GERADA PELO MODELLER



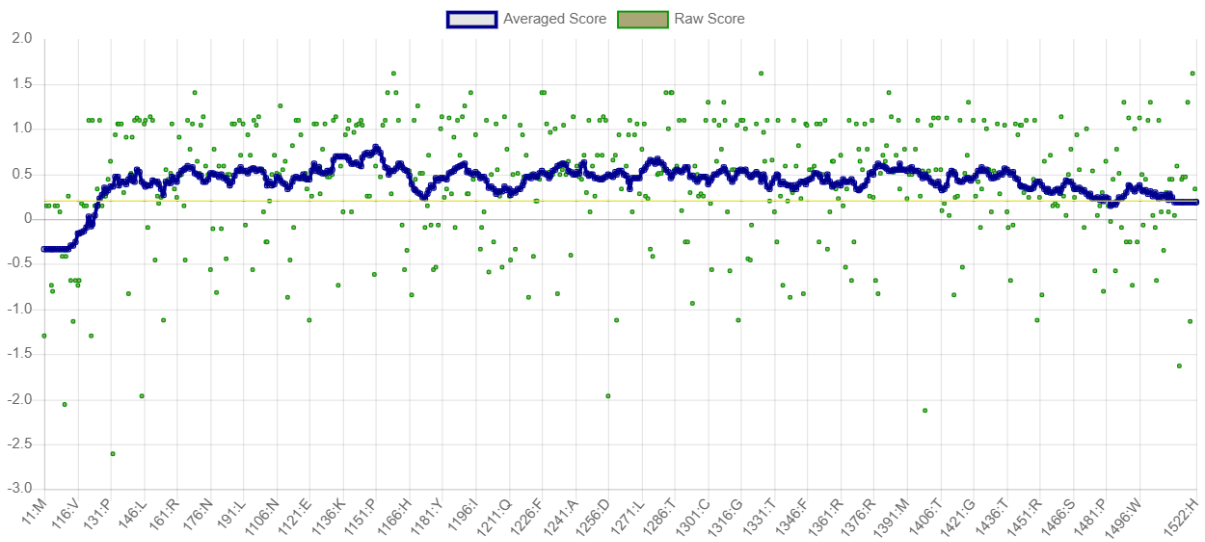
APÊNDICE E - GRÁFICO DE RAMACHANDRAN DA VARIANTE p.V239F
GERADA PELO MODELLER



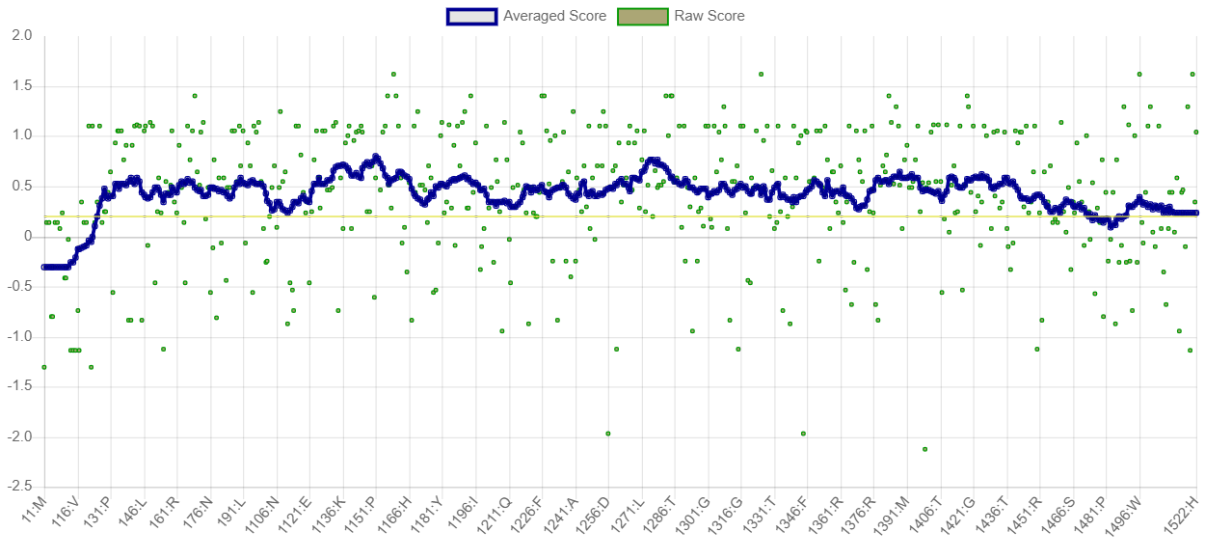
APÊNDICE F - GRÁFICO DO VERIFY3D DA VARIANTE p.D45G GERADA PELO MODELLER



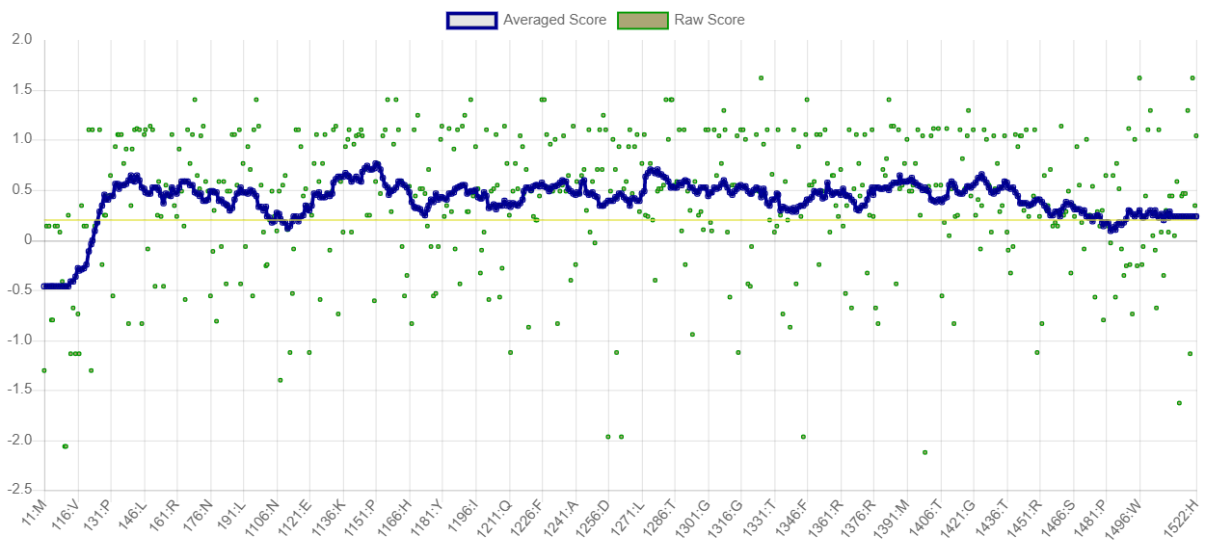
APÊNDICE G - GRÁFICO DO VERIFY3D DA VARIANTE p.G301C GERADA PELO MODELLER



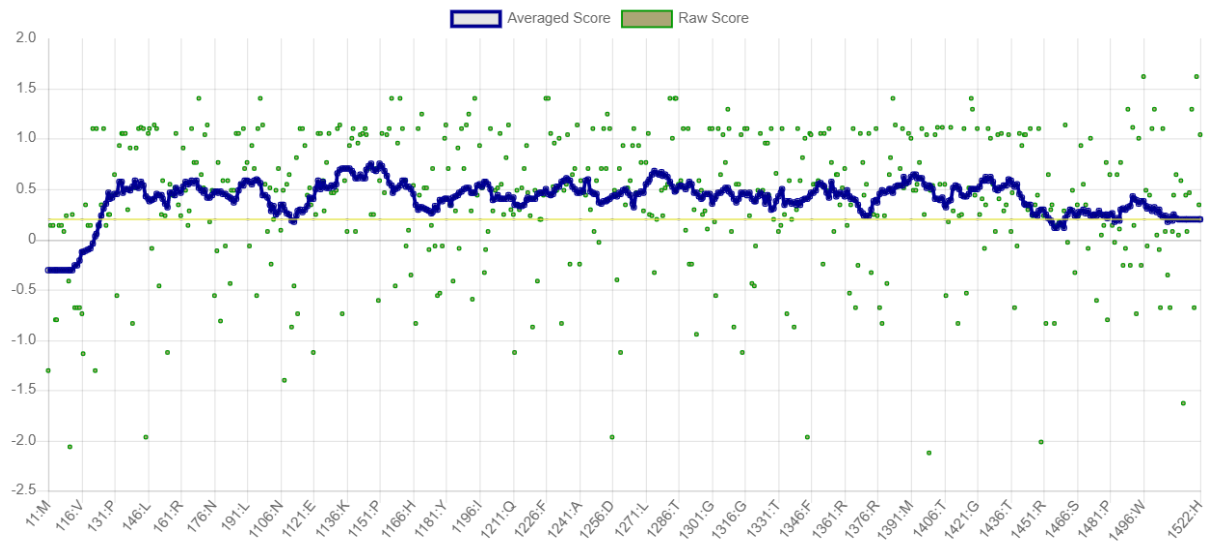
APÊNDICE H - GRÁFICO DO VERIFY3D DA VARIANTE p.R386C GERADA PELO MODELLER



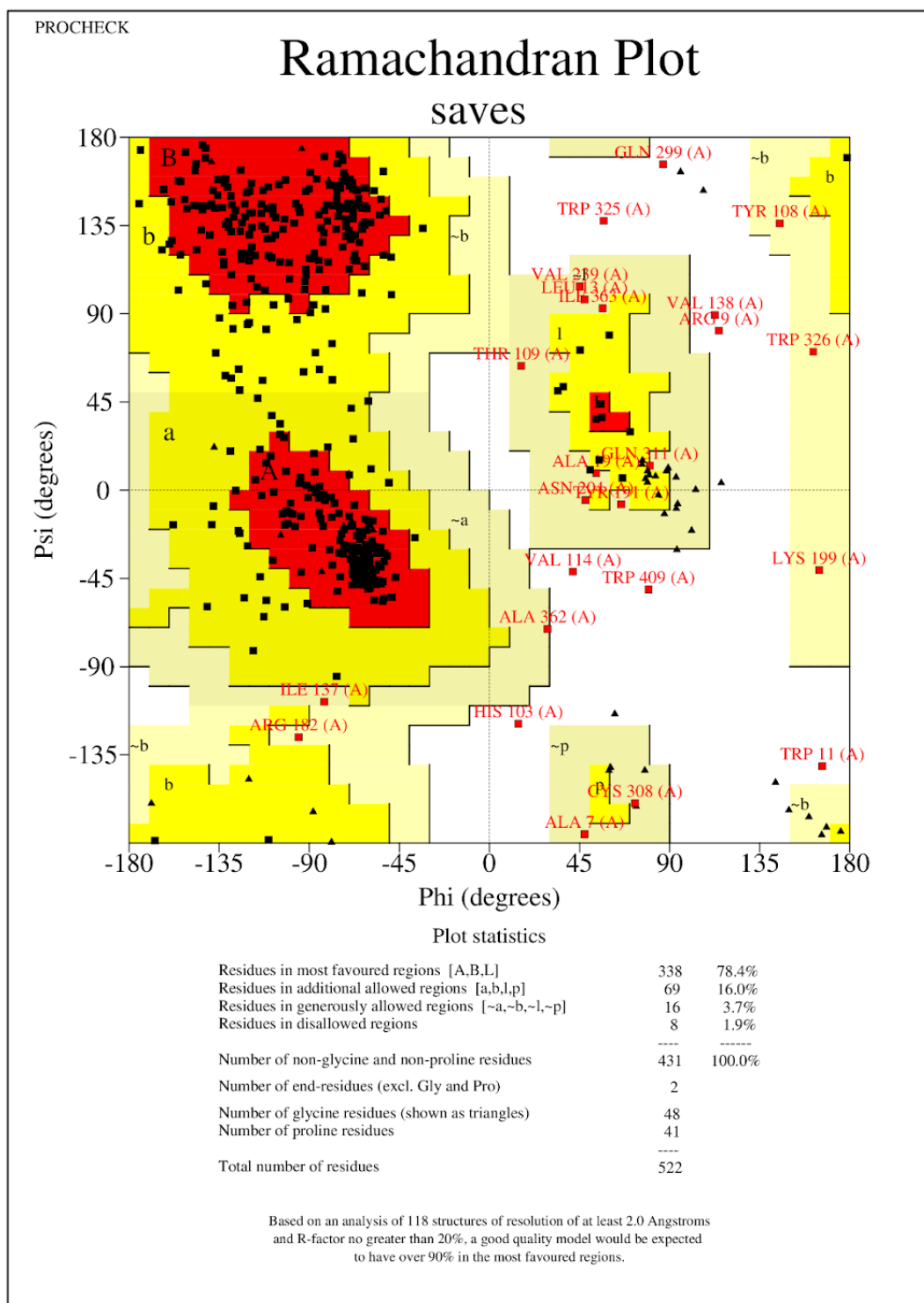
APÊNDICE I - GRÁFICO DO VERIFY3D DA VARIANTE p.S341R GERADA PELO MODELLER



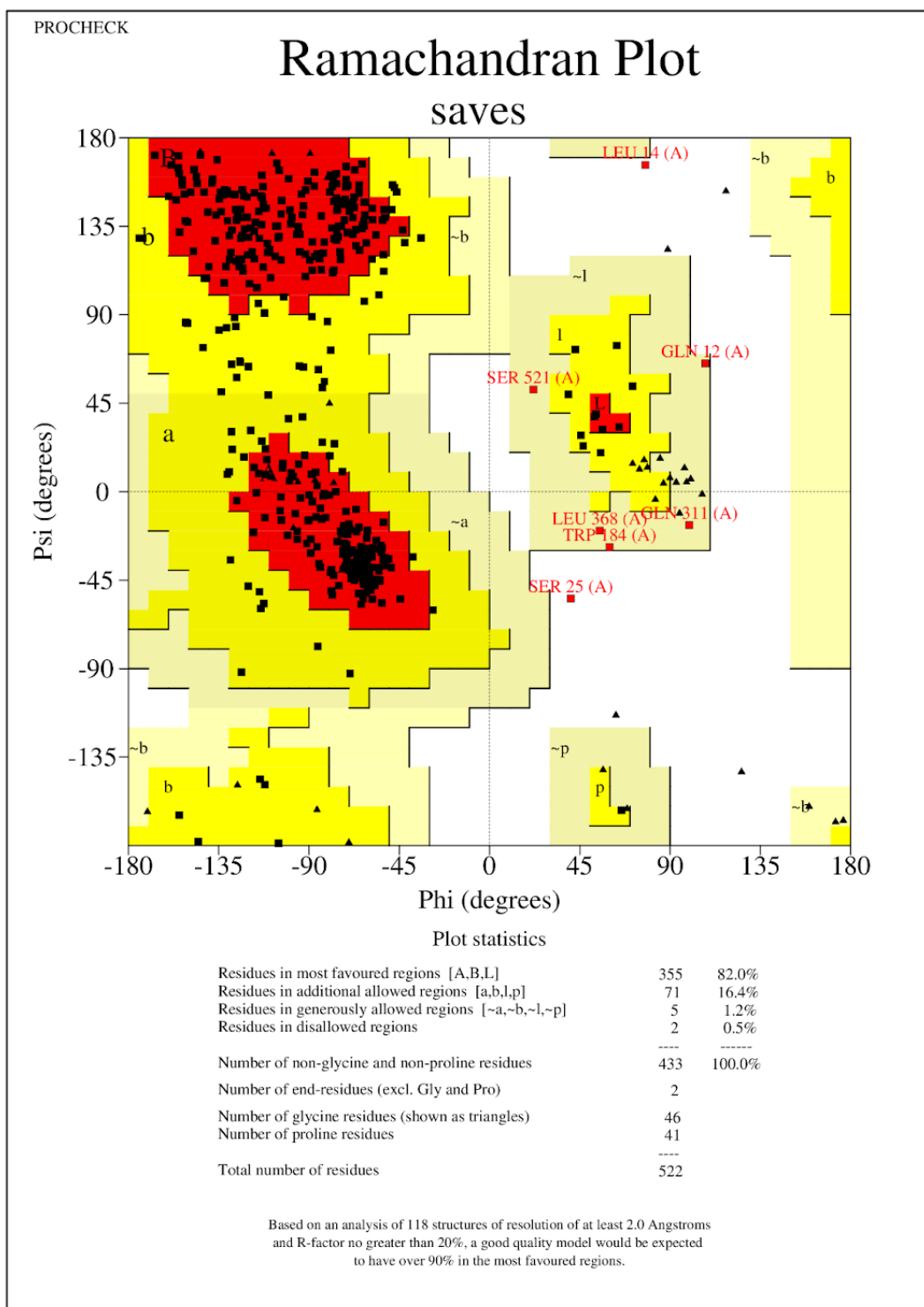
APÊNDICE J - GRÁFICO DO VERIFY3D DA VARIANTE p.V239F GERADA PELO MODELLER



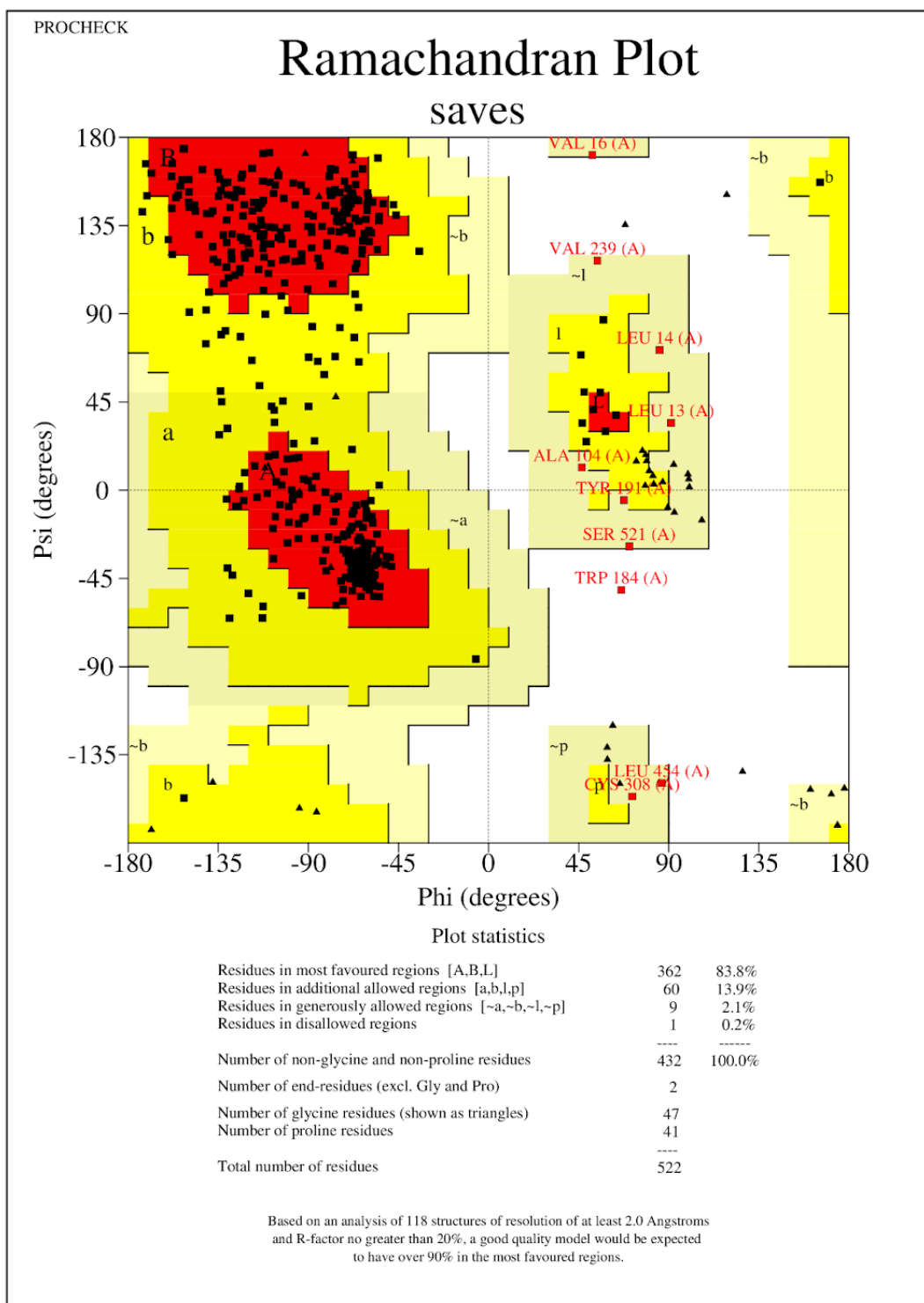
APÊNDICE K - GRÁFICO DE RAMACHANDRAN DA VARIANTE p.D45G
GERADA PELO I-TASSER



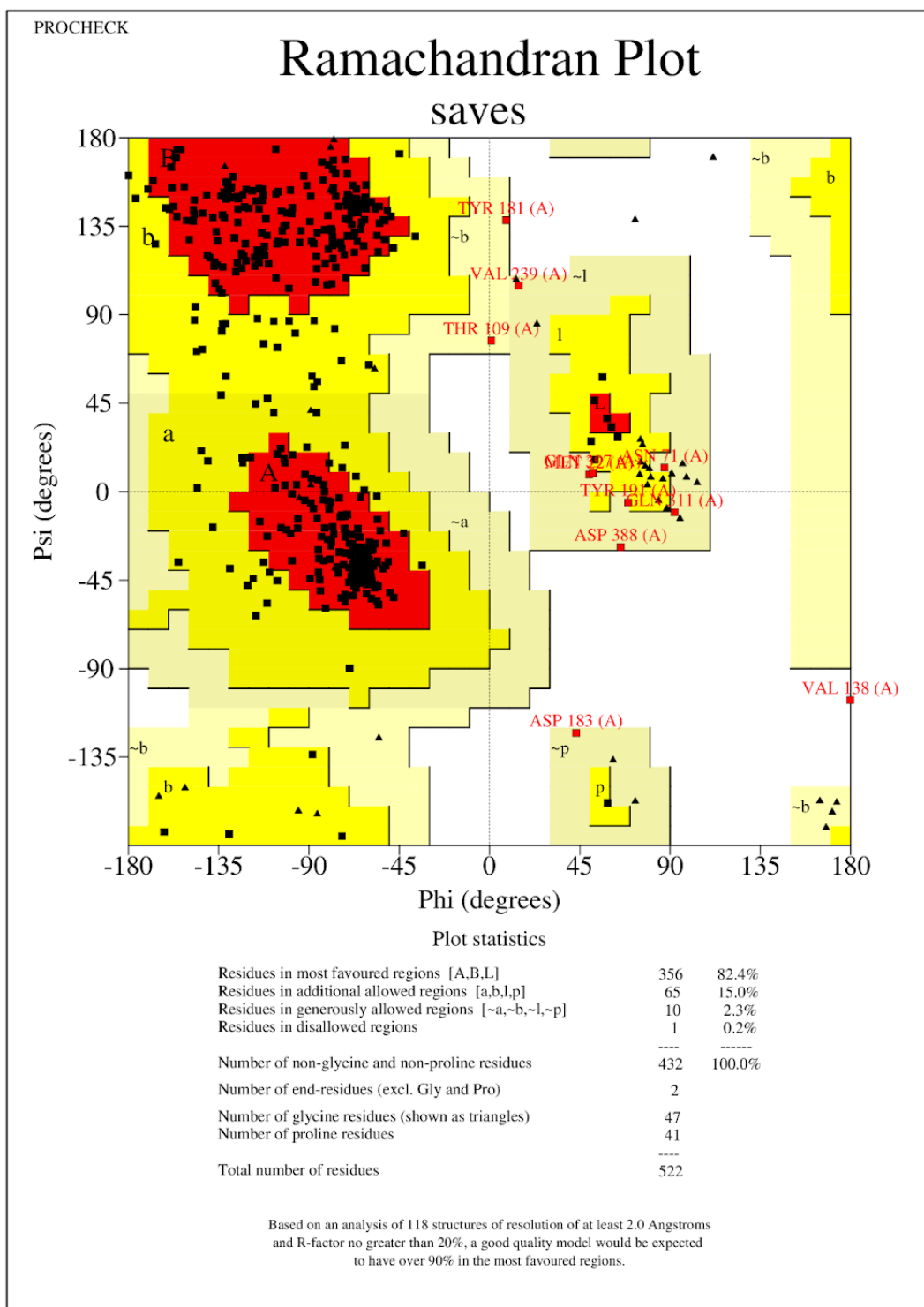
APÊNDICE L - GRÁFICO DE RAMACHANDRAN DA VARIANTE p.G301C
GERADA PELO I-TASSER



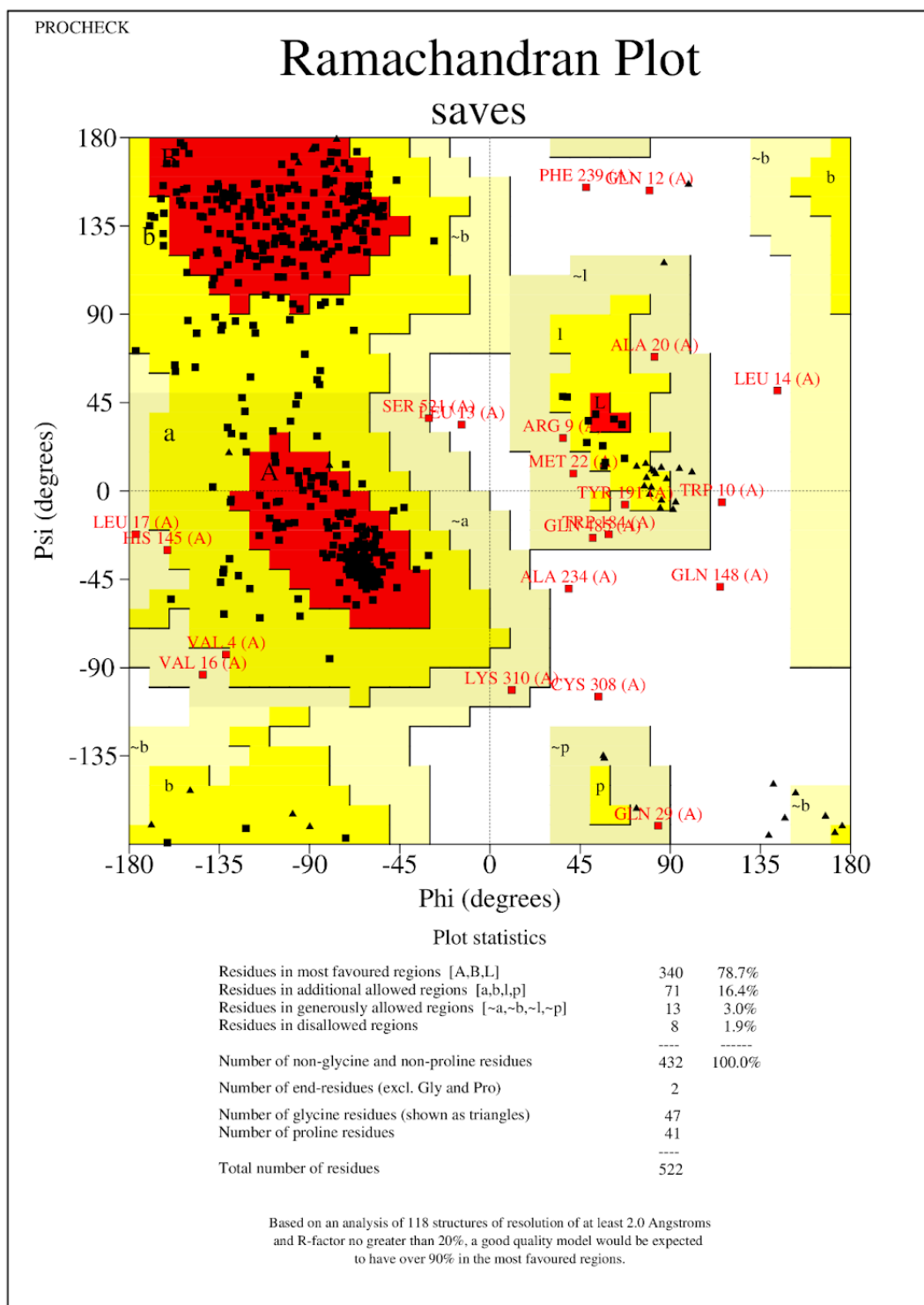
APÊNDICE M - GRÁFICO DE RAMACHANDRAN DA VARIANTE p.R386C
GERADA PELO I-TASSER



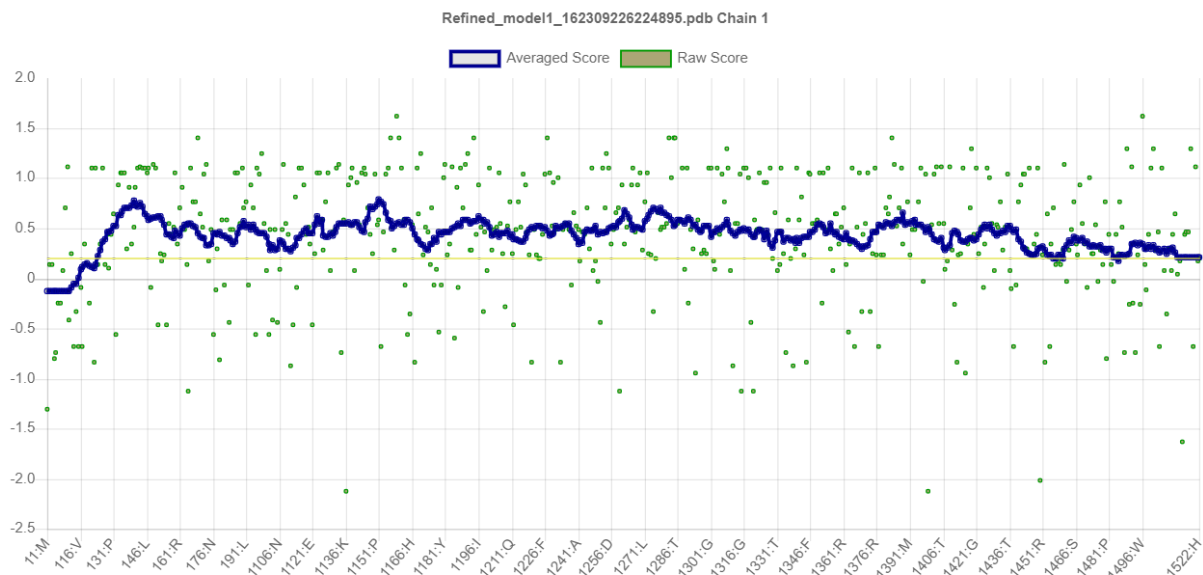
APÊNDICE N - GRÁFICO DE RAMACHANDRAN DA VARIANTE p.S341R
GERADA PELO I-TASSER



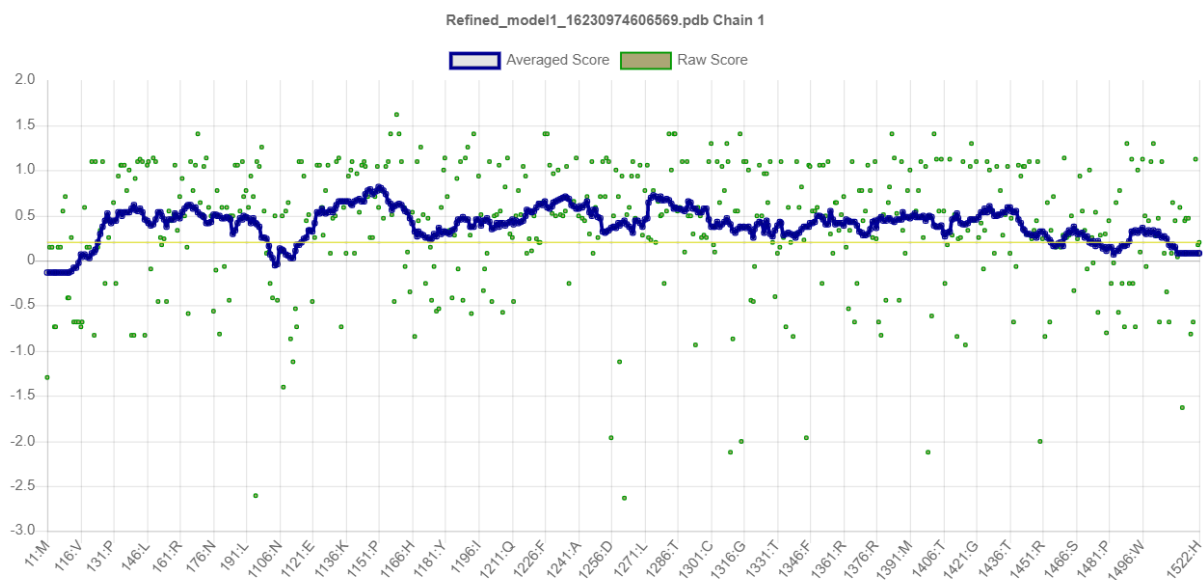
APÊNDICE O - GRÁFICO DE RAMACHANDRAN DA VARIANTE p.V239F
GERADA PELO I-TASSER



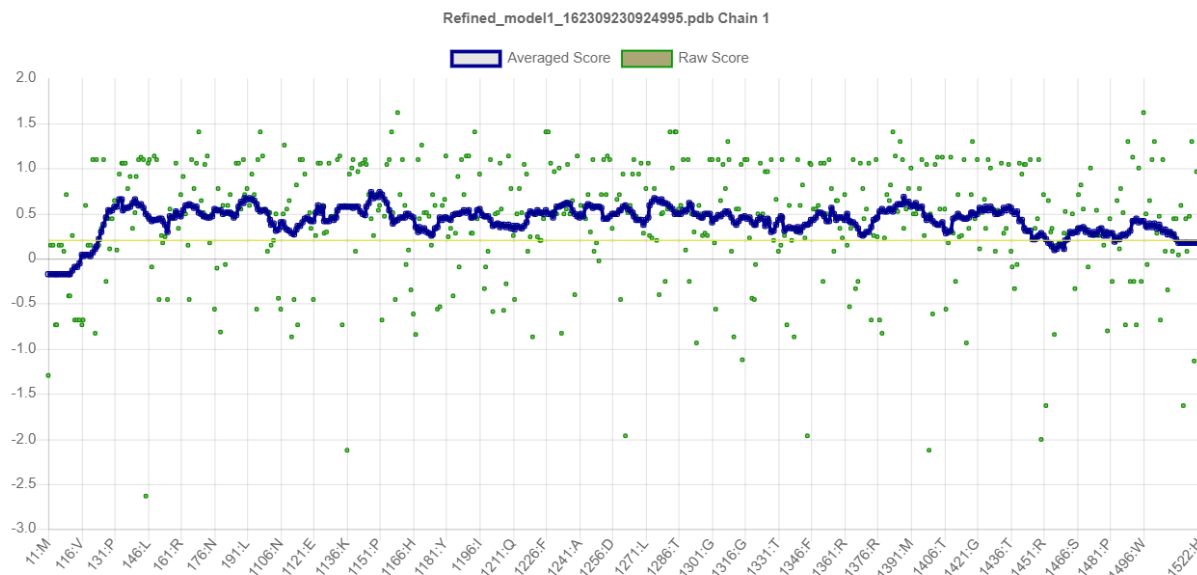
APÊNDICE P - GRÁFICO DO VERIFY3D DA VARIANTE p.D45G GERADA PELO I-TASSER



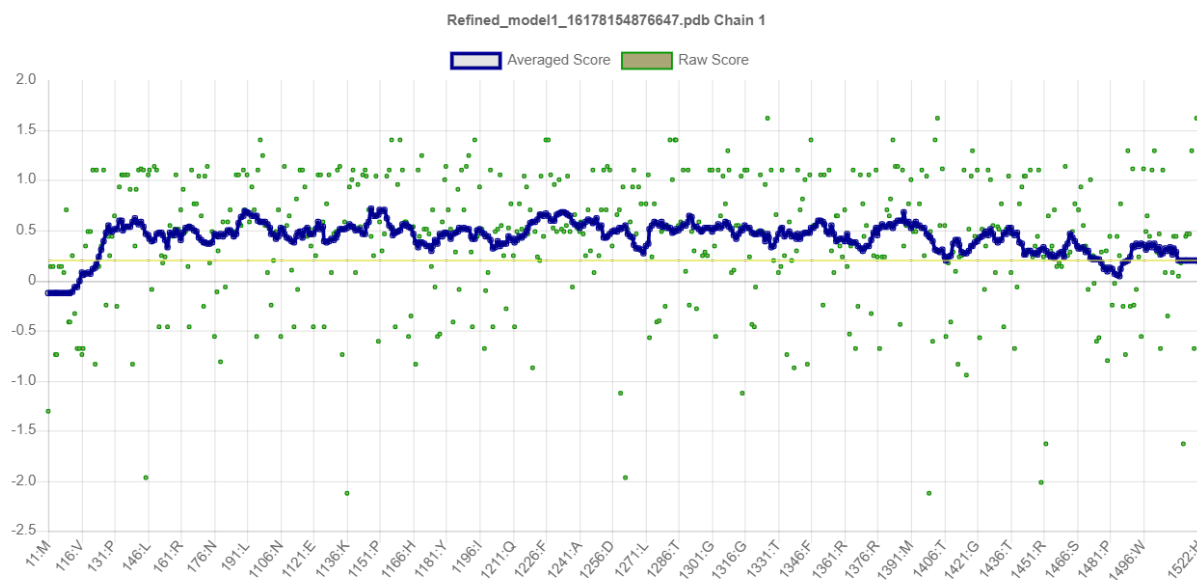
APÊNDICE Q - GRÁFICO DO VERIFY3D DA VARIANTE p.G301C GERADA PELO I-TASSER



APÊNDICE R - GRÁFICO DO VERIFY3D DA VARIANTE p.R386C GERADA PELO I-TASSER



APÊNDICE S - GRÁFICO DO VERIFY3D DA VARIANTE p.S341R GERADA PELO I-TASSER



APÊNDICE T - GRÁFICO DO VERIFY3D DA VARIANTE p.V239F GERADA PELO I-TASSER

