



Universidade Federal do Estado do Rio de Janeiro
Centro de Ciências Biológicas e da Saúde
Instituto Biomédico
Programa de Pós-Graduação em Biologia Molecular e Celular

**Dinâmica Molecular e Análise *In Silico* da Proteína Triptofano Hidroxilase 2
(TPH2)**

Marta Costa de Freitas

Orientadora: Prof^a. Dr^a. Joelma Freire de Mesquita

Rio de Janeiro

2017

Marta Costa de Freitas

**Dinâmica Molecular e Análise *In Silico* da Proteína Triptofano Hidroxilase 2
(TPH2)**

Dissertação de mestrado apresentada ao Programa de Pós-Graduação em Biologia Molecular e Celular da Universidade Federal do Estado do Rio de Janeiro como parte dos requisitos necessários à obtenção do grau de Mestre em Biologia Molecular e Celular.

Orientadora: Prof^a. Dr^a Joelma Freire de Mesquita

Rio de Janeiro

2017

Catálogo informatizada pelo(a) autor(a)

F862 Freitas, Marta Costa de
Dinâmica Molecular e Análise In Silico da Proteína
Triptofano Hidroxilase 2 (TPH2)/ Marta Costa de
Freitas. -- Rio de Janeiro, 2017.
80 f.

Orientadora: Joelma Freire de Mesquita.
Dissertação (Mestrado) - Universidade Federal do
Estado do Rio de Janeiro, Programa de Pós-Graduação
em Genética e Biologia Molecular, 2017.

1. Triptofano Hidroxilase 2. 2. Desordens
Mentais. 3. Dinâmica Molecular. 4. Bioinformática.
I. Mesquita, Joelma Freire de, orient. II. Título.

Marta Costa de Freitas

**Dinâmica Molecular e Análise *In Silico* da Proteína Triptofano Hidroxilase 2
(TPH2)**

Dissertação de mestrado apresentada ao Programa de Pós-Graduação em Biologia Molecular e Celular da Universidade Federal do Estado do Rio de Janeiro como parte dos requisitos necessários à obtenção do grau de Mestre em Biologia Molecular e Celular.

Banca examinadora:

Dra. Carmen Lúcia Antão Paiva (Doutora em Ciências) UNIRIO

Dr. Rafael Braga Gonçalves (Doutor em Ciências) UNIRIO

Dra. Bárbara de Azevedo Abraham Vieira (Doutora em Ciências) UFRJ

Dr. Marcos Dias Pereira (Doutor em Ciências) UFRJ

*Ao meu Deus, família e amor.
Ainda que eu tivesse tudo,
Sem eles eu nada seria.*

AGRADECIMENTOS

Primeiramente a Deus, por me dar vida, propósito, oportunidades e capacidade. Por ser o único capaz de fazer por mim muito além do que sou capaz de pedir e imaginar. Somando-se tal fato ao seu amor e vontades perfeitas, não há como eu resistir a simplesmente fechar meus olhos e confiar! Em 2016, nos meus agradecimentos da monografia eu O agradei pela bela vista da Urca, da qual me despedi saudosamente quando entreguei a versão corrigida do trabalho e achei que nunca mais fosse apreciar do ângulo do CCBS. Ledo engano. Nessa hora Deus deve ter rido de mim. Obrigada também pelo Seu senso de humor!

Agradeço à segunda base que me sustenta, àquela ao qual foi designada a me aturar todos os dias e apoiar quando necessário. Entre “Você é doida de ficar indo pro Rio!” e “Você sabia que todo morador do Rio está sujeito a levar um tiro de bala perdida de fuzil?”, minha mãe também consegue orgulhosamente dizer que “Ela já está no mestrado. É bióloga!”. Mas ninguém conta meus feitos melhor do que minha irmã. Nunca sei onde me esconder quando começam! E entre os extremos do “não vai, é perigoso e desgastante” e o “vai, você é inteligente e precisa investir em ti”, a sobriedade me veio com lindos olhos verdes, pau-para-toda-obra, que há 8 anos me aturam por escolha. Obrigada, meu amor, porque mesmo quando quer dizer “fica”, me diz “vai” se for o melhor para mim; além de toda confiança. Obrigada a vocês, que apesar de exercerem papéis tão diferentes na minha vida, em comum acreditam que sou capaz!

Não poderia deixar de agradecer pela paciência do meu pequeno sobrinho Cadu, que por tantas vezes quis brincar comigo, mas ouviu “Não posso, estou estudando.”. Espero que isso tenha algum efeito positivo futuramente, porque sinceramente todas as vezes que preciso falar é de partir o coração.

E é claro que nada disso seria possível sem o BIG! Mesmo fazendo parte do grupo desde a graduação, confesso que só os conheci durante o mestrado. De tantas pessoas, o primeiro agradecimento não poderia ser para outrem que não a Prof^a. Joelma, que me deu a ideia de seguir para o mestrado e me acolheu acreditando que eu daria conta de estudar e trabalhar. Obrigada por investir nessa conquista, que é nossa! E confesso que eu teria demorado o dobro de tempo para chegar até aqui, se não possuísse ao meu lado colegas tão dispostos a compartilhar. Pensei que seria mais sensato não citar nomes, porque todos me ajudaram em algum momento e nem

sempre só com conhecimento acadêmico, mas brincadeiras, risos, histórias e comemorações. Só que isso seria injusto com as duas pessoas que mais perturbei durante esse ano e meio: Aloma e Gabriel, muito obrigada! Vocês são demais!

Chegando aos finalmentes (mesmo!), agradeço aos queridos professores da banca examinadora por aceitarem fazer parte dessa etapa de conclusão; e aqueles docentes do PPGBMC. Professores, vocês merecem todo o reconhecimento do mundo por terem escolhido compartilhar o que sabem e investirem em uma área que infelizmente é tão negligenciada no nosso país. Eu os admiro imensamente e quero ser como vocês quando eu crescer!

Por último, mas não menos importante, agradeço a UNIRIO por fornecer a oportunidade da pós-graduação. Nesta ocasião, não haveria professores a agradecer, laboratório do qual participar, pesquisa a qual realizar e dissertação para defender se nossa querida instituição não fosse forte o suficiente para resistir. E quem a faz forte?

Enfim, registro meu obrigada a todos que de um jeito ou de outro me incentivaram a chegar até aqui.

RESUMO

A bioinformática é uma ciência multidisciplinar que oferece ferramentas computacionais diversas de baixo custo, curto tempo de realização e resultados de alta qualidade. Por essas características tem se tornado essencial em diversos estudos de biomoléculas, como para aqueles relacionados a proteínas. A dinâmica molecular (DM) é uma técnica computacional amplamente utilizada no estudo de biosistemas complexos, e que simula o comportamento das moléculas em função do tempo. Assim, se torna próxima dos métodos experimentais e resulta em modelos próximos à realidade biológica. A Triptofano Hidroxilase 2 (TPH2) faz parte da família de hidroxilases de aminoácidos aromáticos (AAAH), caracterizadas por dependerem de tetrahydrobiopterina, ferro e oxigênio para sua ativação. A TPH2 catalisa a 5-hidroxilação do L-triptofano, que é o precursor da serotonina em sua biossíntese. Devido sua participação na via de síntese desse neurotransmissor, mutações pontuais não sinônimas na TPH2 vêm sendo associadas a transtornos psiquiátricos como a depressão unipolar, o comportamento bipolar e o transtorno de déficit de atenção e hiperatividade (TDAH). Sendo assim, esse trabalho visou a avaliação das variações da TPH2 por meio da dinâmica molecular para melhor entendimento da relação entre mutações, função e psicopatologias; assim como a melhoria de qualidade do modelo experimental existente. Por simulações de dinâmica molecular curtas foram avaliadas as estruturas de 25 mutantes, das quais R156E, R225Q, P277L, R303W, S383F, E430G, R441H e D473N mostraram maior comportamento dinâmico atípico em comparação com a estrutura selvagem. Para realização das simulações de dinâmica molecular longas, foram selecionadas apenas as mutantes previamente associadas a desordens mentais. Assim, utilizando o pacote GROMACS do Linux realizou-se as simulações das variantes P206S, R303W e R441H. As análises realizadas para avaliação do comportamento das estruturas mostraram que todas as mutações levam ao aumento do valor de RMSD e RMSF médio, quando comparadas à estrutura selvagem. P206S e R441H mostram menores raios de giro e área superfície de acesso ao solvente, enquanto R303W apresenta comportamento contrário para essas características. Quanto as ligações de hidrogênios, observou-se que P206S estabelece mais ligações inter e intramoleculares, enquanto R441H leva à diminuição das mesmas. Já R303W apresenta número médio de ligações intermolecular maior do que o observado

para WT e menor no estabelecimento de ligações intramoleculares. As análises de flexibilidade atentaram para a maior flexibilidade do terminal COOH nas variantes. Como um todo, tais resultados sugerem que todas as variantes mostram comportamento estrutural dinâmico atípico quando comparadas à estrutura selvagem, especialmente R303W que mostrou mudanças mais relevantes. Os resultados da análise *in silicio* e estrutura aprimorada estarão disponíveis em banco de dados.

ABSTRACT

Bioinformatics is a multidisciplinary science that offers several computational tools with low cost and execution time beside high quality results. Because of these features, it has becoming essential in many bimolecular studies, as those related to proteins. Molecular Dynamics (MD) is a computational technique widely used in complex biosystem studies and that simulates the target molecule behavior as a function of time. Thus, it is similar to experimental methodologies and generates model close to the biological reality. The Tryptophan Hydroxylases 2 (TPH2) is part of the hydroxylase of aromatic amino acid's family (AAAH), characterized by depending on tetrahydrobiopterin, iron and oxygen to their activation. TPH2 catalyzes the 5-hydroxylation of L-tryptophan that is the precursor of serotonin on its biosynthesis. Due to its participation on this neurotransmitter's synthesis via, non-synonymous point mutations in TPH2 has been being associated to psychiatric disorders as unipolar depression, bipolar disorder (BD) and attention deficit hyperactivity disorder (ADHD). Therefore, this study aimed to evaluate TPH2 variations by molecular dynamics to understand better the relationship between mutations, function and psychopathologies; as to improve the quality of the unique experimental structure existent for TPH2. Short MD were carried out to 25 mutations, from which R156E, R225Q, P277L, R303W, S383F, E430G, R441H, and D473N shown most atypical dynamic behavior when compared to the WT structure. To perform long MD, only those mutants previously associated to mental disorders were chosen. Thus, using GROMACS package for Linux, MD were carried out to P206S, R303W, and R441H variations and WT structures. The analysis done to evaluate the structures behavior shown that all mutations increase the average RMSD and RMSF values when compared to the WT structure. P206S and R441H show minor radius of gyration and solvent accessible surface area, while R303W presents contrary behavior to these features. Regarding to the number of hydrogen bonds it was observed that P206S establish more inter and intramolecular hydrogen bonds than the WT, while R441H establish less of them. Differently, R303W presents average number of intermolecular hydrogen bonds bigger than the showed for WT, and minor when it comes to intramolecular hydrogen bonds. Flexibility analysis attempted to the major C-terminal flexibility in the variants structures. All things considered, these results suggests that all variants showed atypical

dynamic behavior when compared to the WT structure, specially R303W that shows the most relevant alterations. The *in silico* analysis results and the optimized structure of TPH2 will be available in an online database.

LISTA DE FIGURAS

Figura 1 - Modelo experimental da Triptofano Hidroxilase 2 obtido no Protein Data Bank e colorido segundo sua estrutura secundária pelo Chimera.....	22
Figura 2 - Resultados gráficos retornados pelos validadores ERRAT e Verify 3D para validação da estrutura 4V06.....	33
Figura 3 - Gráficos de Ramachandran gerados pelo PROCHECK e Rampage como resultado da validação estrutural do modelo 4V06.....	35
Figura 4 - Resultados fornecidos pelo QMEAN Server e ProSa para a validação estrutural do modelo 4V06.....	36
Figura 5 - Gráfico contendo os valores de RMSD em nanômetro por snapshot das estruturas proteicas selvagem e mutantes obtidos no MDWeb para estudo de estabilidade.....	40
Figura 6 - Gráfico contendo as mutações para quais se observou valores diferentes de RMSD quando comparadas aos valores obtidos pela estrutura selvagem.....	40
Figura 7 - Gráfico resultante da análise de Raio de Giro em nanômetros por snapshot obtidos no MDWeb para estudo de compactação das estruturas proteicas durante a simulação da dinâmica.....	42
Figura 8 - Gráfico comparativo contendo as 5 mutações relacionadas a causar maior aumento na compactação da estrutura em comparação com a estrutura selvagem....	42
Figura 9 - Gráfico resultante da análise de RMSD em angstrom por resíduos obtidos no MDWeb para estudo de flexibilidade das estruturas proteicas durante a simulação da dinâmica.....	44
Figura 10 - Gráfico resultante da análise de B-factor em angstrom quadrado por resíduo obtidos no MDWeb para estudo de mobilidade termal das estruturas proteicas durante a simulação da dinâmica.....	44
Figura 11 - Gráfico comparativo entre os valores de RMSD obtidos para a estrutura da proteína selvagem e as mutantes em função do tempo de 80 ns decorridos de simulação.....	47
Figura 12 - Gráfico comparativo constando os valores obtidos na análise de Raio de Giro em nanômetros para as estruturas selvagem e as mutantes estudadas em função do tempo de 80 ns de simulação.....	48
Figura 13 - Gráficos comparativos entre os valores obtidos a partir da análise da área de superfície acessível ao solvente (SASA) em nanômetros quadrados para as estruturas selvagem e mutantes durante os 80 ns de simulação.....	50
Figura 14 - Gráficos constando o número de ligações de hidrogênio intramoleculares estabelecidas em cada estrutura proteica analisada durante a trajetória.....	51

Figura 15 - Gráfico constando o número de ligações de hidrogênio intermoleculares estabelecidas entre as estruturas proteicas analisadas e a água durante a trajetória.	52
Figura 16 - Gráficos constando os valores de RMSF em nanômetros para cada resíduo das estruturas selvagem e mutantes, durante o período de 80 ns da dinâmica molecular.....	54
Figura 17 - Gráficos constando os valores de RMSF em nanômetros para cada resíduo das estruturas selvagem e mutantes, durante o período de 62 ns a 80 ns da dinâmica molecular.....	56
Figura 18 - Figuras obtidas a partir do resultado da análise de B-factor para cada estrutura estudada.....	57
Figura 19 - Análise de Componentes Principais das trajetórias obtidas através da simulação de dinâmica molecular das estruturas selvagem e as variantes estudadas.....	60
Figura 20 - Modelo estrutural do domínio catalítico da TPH2, visualizado pelo Chimera.....	63
Figura 21 - Modelos estruturais obtidos como resultado das simulações de dinâmica molecular para a selvagem e variantes, visualizados e coloridos segundo suas estruturas secundárias pelo Chimera.....	65
Figura 22 - Alinhamento das sequências referentes aos modelos obtidos após a simulação da dinâmica molecular para as estruturas selvagem e variantes, realizado pelo Chimera.....	66

LISTA DE TABELAS

Tabela 1- Mutações compiladas dos bancos de dados Uniprot, OMIM, Pubmed e dbSNP.....	31
Tabela 2 - Resultados dos algoritmos para estudo de validação das estruturas refinadas.....	38
Tabela 3 - Comparação entre os resultados das análises obtidas pelo MDWeb e da predição funcional anteriormente realizada.....	45

LISTA DE ABREVIATURAS E SIGLAS

5-HT – 5-hidroxitriptofano (Serotonina)

5-HTP – 5-hidroxi-L-triptofano

AAAH – Hidroxilase de Aminoácidos Aromáticos (do inglês, *Aromatic Amino Acid Hydroxylase*)

BH4 – Tetrahidrobiopterina

DM – Dinâmica Molecular

DNA – Ácido Desoxirribonucleico (do inglês, *Deoxyribonucleic Acid*)

ED – Dinâmica Essencial (do inglês, *Essential Dynamics*)

GROMACS – Máquina Groningen para Simulações Química (do inglês, *Groningen Machine for Chemical Simulations*)

Hb – Ligações de Hidrogênio (do inglês, *Hydrogen bond*)

KoBa^{MIN} – Refinamento Potencial para Proteínas Baseado no Conhecimento (do inglês, *Knowledge-based Potential Refinement for Proteins*)

LINCS – Solucionador de Condições Lineares (do inglês, *Linear Constraint Solver*)

mRNA – Ácido Ribonucleico mensageiro (do inglês, *Messenger Ribonucleic Acid*)

NCBI – Centro Nacional para Informação de Biotecnologia (do inglês, *National Center for Biotechnology Information*)

NHB – Número de Ligações de Hidrogênio (do inglês, *Number of Hydrogen Bonds*)

nm – Nanômetros

NPT – Número de partículas, Pressão e Temperatura

nsSNPs – Polimorfismo de Nucleotídeo único não-sinônimo (do inglês, *Non-synonymous Single Nucleotide Polymorphism*)

nsSNVs – Variação de Nucleotídeo único não-sinônimo (do inglês, *Non-synonymous Single Nucleotide Variation*)

NVT – Número de partículas, Volume e Temperatura

OMIM – Herança Mendeliana no Homem Online (do inglês, *Online Mendelian Inheritance in Man*)

PAH – Fenilalanina Hidroxilase (do inglês, *Phenylalanine Hydroxylase*)

PCA – Análise de Componente Principal (do inglês, *Principal Component Analysis*)

PME – Malha de Partículas Ewald (do inglês, *Particle Mesh Ewald*)

ps - Picosegundo

RCSB PDB – Banco de Dados de Proteínas Colaborador de Pesquisa para Bioinformática Estrutural (do inglês, *Research Collaboratory for Structural Bioinformatics Protein Data Bank*)

Rg – Raio de giro (do inglês, *Radio of gyration*)

RMSD – Desvio Quadrático Médio (do inglês, *Root Mean Square Deviation*)

RMSF – Flutuação Quadrática Média (do inglês, *Root Mean Square Fluctuation*)

SASA – Área de Superfície Acessível ao Solvente (do inglês, *Solvent Accessible Surface Area*)

SNPs – Polimorfismo de Nucleotídeo Único (do inglês, *Single Nucleotide Polymorphism*)

SNVs – Variação de Nucleotídeo Único (do inglês, *Single Nucleotide Variation*)

TDAH – Transtorno Déficit de Atenção e Hiperatividade

TH – Tirosina Hidroxilase (do inglês, *Tyrosine Hydroxylase*)

TPH – Triptofano Hidroxilase (do inglês, *Tryptophan Hydroxylase*)

TPH1 – Triptofano Hidroxilase 1

TPH2 – Triptofano Hidroxilase 2

VMD – Dinâmica Molecular Visual (do inglês, *Visual Molecular Dynamics*)

WT – Selvagem (do inglês, *Wild Type*)

SUMÁRIO

1. Introdução.....	18
1.1 Bioinformática e proteínas.....	18
1.2 Triptofano Hidroxilase 2 (TPH2).....	20
1.3 Mutações não sinônimas de nucleotídeo único.....	24
2. Objetivos.....	26
3. Metodologia.....	26
3.1 Obtenção da estrutura proteica experimental.....	26
3.2 Compilação das variações não sinônimas.....	26
3.3 Refinamento e validação estrutural.....	27
3.4 Dinâmica molecular curta.....	27
3.5 Dinâmica molecular longa.....	28
3.5.1 Análise de componente principal.....	29
4. Resultados e Discussão.....	30
4.1 Obtenção da estrutura proteica experimental.....	30
4.2 Compilação das variações não sinônimas.....	30
4.3 Refinamento e validação estrutural.....	32
4.4 Dinâmica molecular.....	37
4.4.1 Dinâmica molecular curta.....	37
4.4.2 Dinâmica molecular longa.....	46

4.4.2.1 Análise de componente principal.....	58
5. Conclusão.....	67
Referências Bibliográficas.....	69
Anexo A.....	77
Anexo B.....	79

1. INTRODUÇÃO

1.1 Bioinformática e proteínas

A bioinformática pode ser entendida como uma ciência multidisciplinar, a qual oferece métodos computacionais capazes de analisar e resolver questões biológicas e mostra ser uma grande e efetiva aliada no estudo de proteínas e outras biomoléculas. Os meios *in silico* apresentam ser de relativo baixo custo, tempo de realização e fornece resultados de alta qualidade, tornando-se cada vez mais eficientes e essenciais (Verli, 2014c).

Por meio da bioinformática pode-se realizar análises em relação à sequência nucleotídica, como a predição desta; e em relação às estruturas proteicas, como a predição de energia livre, enovelamento, modelagem de estruturas teóricas, dentre outras possibilidades (Verli, 2014c). Assim, as ferramentas computacionais permitem estudos úteis em muitas áreas, como na farmacoterapia e biotecnologia, uma vez que podem prever detalhes de interações entre proteína-ligante e proteína-proteína importantes para o desenho racional de fármacos, melhor entendimento dos processos biológicos e no desenho de novas proteínas com funções específicas, como utilizado pela nanotecnologia (Capriles, et al., 2010).

Uma técnica computacional amplamente utilizada e útil na avaliação de sistemas complexos é a dinâmica molecular (DM), que consiste na simulação do comportamento molecular em função do tempo, através da análise dos movimentos dos átomos individuais ou em uma molécula. Por não ser uma avaliação estática, a DM se aproxima muito de métodos experimentais, resultando em modelos próximos à realidade biológica (Verli, 2014a). Desse modo, a simulação de dinâmica molecular vem sendo utilizadas em diversas aplicações, como para entender os processos envolvidos no enovelamento e estabilidade de proteínas, reconhecimento molecular, desenho de fármacos, reações enzimáticas, para determinação e refinamento de modelos estruturais e também no estudo de ácidos nucleicos (Beveridge; McConnell, 2000; Fan; Mark, 2003; Scheraga; Khalili; Liwo, 2007; Khan, et al., 2016).

Tipicamente, a simulação da DM requer a montagem e preparação do sistema a ser estudado e que contém a biomolécula alvo. Um espaço chamado caixa de

simulação é delimitado, onde a molécula de interesse é inserida, solvatada e neutralizada. Geralmente, essa caixa tem forma geométrica que permite o estabelecimento de condições periódicas de contorno, utilizadas para que o sistema não se encontre no vácuo e assim não perca suas partículas constituintes. Isso é possível a partir da replicação da caixa original em todas as direções do espaço. A forma mais utilizada é a cúbica (Namba; Da Silva; Da Silva, 2008; Verli, 2014a; Khan, et al., 2016).

Uma vez montado, o sistema passa por passos de minimização de energia para que maus contatos entre as moléculas sejam eliminados; assim como para que o sistema alcance seu menor nível de energia potencial livre, i.e., sua conformação mais estável (Soares, 1997; Namba; Da Silva; Da Silva, 2008; Verli, 2014a). Parte da preparação também consiste no estabelecimento das condições do sistema, as quais representam o seu estado e propriedades constantes. Para isso, é comum que primeiramente sejam estabelecidas as condições nas quais o número de partícula, volume, e temperatura sejam constantes (NVT), seguido das condições nas quais a pressão também se estabiliza (NPT). Após tais etapas, a simulação pode ser realizada (Allen; Tildesley, 1987; Verli, 2014a).

Durante o processo, o sistema é gradativamente aquecido até a temperatura desejada e os movimentos atômicos são calculados segundo as equações de movimento de Newton, fazendo-se uso de um campo de força que descreve a energia potencial do sistema a partir das contribuições inter-atômicas e inter-moleculares (Van Gunsteren; Berendsen, 1990; Soares, 1997; Verli, 2014a; Khan, et al., 2016). Inicialmente, as condições da simulação distam do equilíbrio termodinâmico, o que significa que suas propriedades não se mantêm constantes. Tal período é chamado de período de equilibração, e o tempo necessário para que se atinja o equilíbrio varia dependendo do sistema. Sendo assim, a simulação deve ser suficientemente longa para que tal estado seja alcançado (Namba; Da Silva; Da Silva, 2008; Verli, 2014a).

Após a simulação, inúmeras avaliações da trajetória gerada podem ser realizadas (Kuzmanic; Zagrovic, 2010; Yang, et al., 2014; Martínez, 2015) e assim tal técnica quando aplicada para comparação da dinâmica de diversos sistemas pode oferecer informações sobre as diferenças dos mesmos, como é de interesse na avaliação do efeito de mutações no comportamento proteico (Krebs; De Mesquita,

2016; Ndagi; Mhlongo; Soliman, 2017; Narang, et al, 2017).

As proteínas são biomoléculas tridimensionais produto da expressão gênica cujo correto funcionamento depende da sua estrutura (Pal; Eisenberg, 2005). Há vários níveis estruturais nos quais as proteínas podem ser organizadas. A começar pela estrutura primária, que consiste na sequência linear e unidimensional de aminoácidos que formam a molécula. Embora seja a mais simples, esse primeiro nível organizacional contém diversas informações sobre a conformação nativa da proteína e conseqüentemente sobre sua função (Verli, 2014b).

Em um segundo nível temos a organização da molécula em sua estrutura secundária, que diz respeito às formas assumidas de acordo com as ligações estabelecidas entre as cadeias laterais dos aminoácidos e entre esses e o solvente. As estruturas secundárias mais comumente citadas são as alças, hélices e folhas beta (Verli, 2014b; Khan, et al., 2016). Por sua vez, a estrutura terciária é estabelecida segundo sua sequência de aminoácidos, modificações pós-traducionais, interação com chaperonas, e aspectos ambientais e localização nos organismos (Godzik, 2011). É o correto enovelamento da proteína em sua estrutura tridimensional nativa o que permite a realização de sua função (Khan, et al., 2016).

Com o aprimoramento das técnicas de sequenciamento de DNA e conclusão do projeto Genoma, diversas proteínas foram sequenciadas e estão disponíveis para serem estudadas (Godzik, 2011). No entanto, fazendo-se uso apenas dos métodos experimentais tal demanda pode demorar a ser atendida. Desse modo a bioinformática mostra ser uma opção segura para o estudo efetivo dessas proteínas.

1.2 Triptofano Hidroxilase 2 (TPH2)

Um neurotransmissor responsável pela regulação de muitos sistemas homeostáticos ligados ao sistema nervoso central, como os relacionados ao apetite, desejo sexual, sono e humor é a serotonina (Walther; Bader, 2003). A alteração dos níveis da serotonina no organismo tem sido relacionada a doenças psiquiátricas como a depressão, desordens de ansiedade, desordem compulsiva obsessiva, psicose, distúrbios alimentares e abuso e dependência de substâncias (Veenstra-Vanderweele; Anderson; Cook, 2000).

A Triptofano Hidroxilase (TPH) está envolvida na biossíntese de serotonina (5-HT), sendo a enzima limitante da taxa de síntese uma vez que é responsável pelo primeiro passo dessa via. A TPH catalisa a 5-hidroxilação do aminoácido precursor da serotonina, o L-triptofano, obtendo-se como produto o 5-hidroxitriptofano (5-HTP). Posteriormente, este intermediário é descarboxilado por outra enzima e então o neurotransmissor é obtido (Mosienko, et al., 2015).

A Triptofano Hidroxilase 2 (TPH2) pertence à superfamília das Hidroxilases de Aminoácidos Aromáticos (AAAH), também composta pela Fenilalanina Hidroxilase (PAH), Tirosina Hidroxilase (TH) e Triptofano Hidroxilase 1 (TPH1). Quando descobertas, as enzimas desta família se distinguiram das demais conhecidas, pois dependiam do cofator não proteico chamado tetrahidrobiopterina (BH4) para exercerem suas funções (Kaufman, 1987; Hufton; Jennings; Colton, 1995).

Além da BH4, as hidroxilases também requerem ferro e oxigênio para alcançar sua forma ativa e realizarem a hidroxilação de seus substratos. Segundo o mecanismo atualmente proposto, a BH4 é utilizada como fonte de dois elétrons necessários para reduzir o ferro a sua forma ferrosa e o oxigênio em água, possibilitando assim a formação de uma enzima oxi-ferrosa intermediária altamente reativa. Essa ativação por redução ocorre em sítios específicos de ligação ao ferro, composto por duas histidinas e um glutamato (Figura 1) (Hufton; Jennings; Cotton, 1995; Fitzpatrick, 2003; Carkaci-Salli, et al., 2006). Uma vez ativa, as hidroxilases são capazes de realizarem suas funções, através da incorporação de um O₂ no anel aromático da cadeia lateral de seus respectivos aminoácidos, sendo proposto que isso ocorra por substituição eletrofílica e com a participação da pterina (Fitzpatrick, 2003).

As hidroxilases possuem similaridade de estrutura primária relativamente alta, sendo assim identificados três domínios funcionais em todas elas: um domínio regulatório na extremidade amino, menos semelhante entre elas; um domínio catalítico central altamente conservado; e um de oligomerização na extremidade carboxila (Mosienko, et al., 2015; Mckinney, et al., 2009). Devido suas diferentes atividades, é sugerido que a regulação de cada uma dessas enzimas ocorra de modo distinto (Kaufman, 1987), embora todas possuam sítios de fosforilação neste domínio (Hufton; Jennings; Cotton, 1995). Já o domínio catalítico contém os sítios necessários para catálise e especificidade de substrato, enquanto o de oligomerização é crucial

para formação de tetrâmeros (Fitzpatrick, 2003). Nos eucariotos, essas enzimas são todas homotetrâmeros, porém também existindo dímeros nas células. No entanto, a forma tetramérica seria a mais abundante assim como apresentaria maior atividade (Kaufman, 1987; Hufton; Jennings; Cotton, 1995).

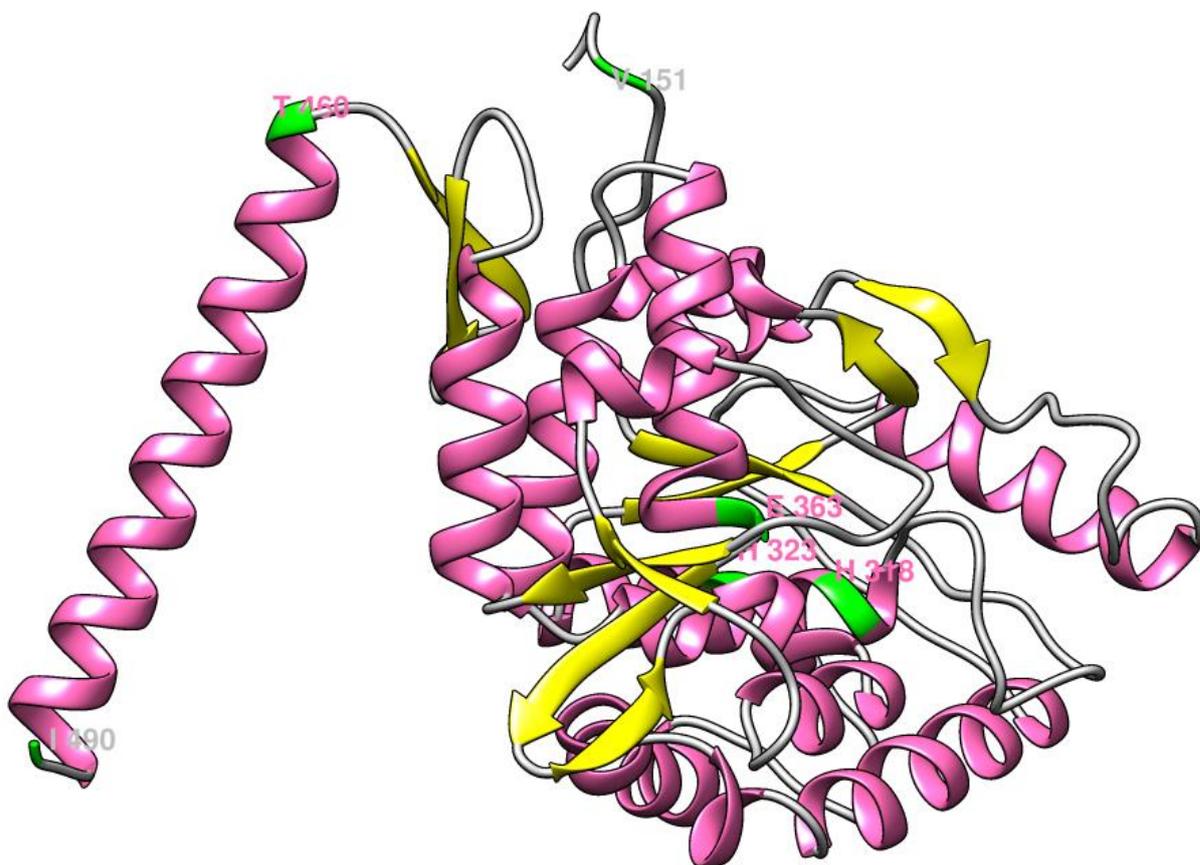


Figura 1 – Modelo experimental da Triptofano Hidroxilase 2 obtido no Protein Data Bank e colorido segundo sua estrutura secundária pelo Chimera. As alças estão coloridas de cinza, as hélices alfa de rosa e as folhas betas de amarelo. Destacados em verde estão o resíduo que marcar o começo do domínio catalítico (V151); os que participam da ligação ao ferro (H318, H323 e E363); e os resíduos que delimitam o domínio de oligomerização (T460 e I490).

Até o ano de 2003 havia sido caracterizada apenas uma isoforma da Triptofano Hidroxilase em vertebrados, responsável pela produção deste neurotransmissor tanto no sistema nervoso central quanto no periférico (Kaufman, 1987; Wang, et al, 2002; Fitzpatrick, 2003). Pouco era conhecido a respeito desta enzima comparado ao que se sabia para a PAH e a TH, uma vez que a baixa abundância de mRNA e as dificuldades de purificação da TPH impediam seu estudo (Cash, 1998). No entanto, complicações na interpretação da regulação desta proteína e seu comportamento

diferenciado no cérebro levantaram suspeitas sobre a existência de outra isoforma (Carkaci-Salli, et al, 2006). A partir de um estudo *knock-out* da única TPH até então conhecida, foi possível observar que apesar dos níveis de serotonina terem diminuído na periferia e glândula pineal, eles permaneceram normais no tronco encefálico. Tal resultado levou a descoberta da segunda isoforma, denominada TPH2, que era predominantemente expressa neste local do cérebro (Walther; Bader, 2003).

As TPHs são codificadas por genes diferentes, bem como apresentam regulações diferenciadas (Mosienko, et al., 2015). A TPH2 é codificada pelo gene de mesmo nome, que possui 93,5 kb de extensão com 11 éxons e está localizado no cromossomo 12q21.1 (Zill, et al., 2004; Xu, et al., 2014). Há evidências de que duas substâncias poderiam agir como ativadoras transcricionais da síntese de TPH2 sendo elas o hormônio vitamina D e o 17- β estradiol. O primeiro estaria associado ao autismo, sendo seu papel na regulação da TPH2 um mecanismo proposto para explicar a já existente relação entre tal hormônio e a serotonina (Patrick; Ames, 2014). Já a regulação por 17- β estradiol poderia ser utilizada como alvo de hormonioterapia em casos de desordens de ansiedade e depressão (Hiroi; Handa, 2013).

Atualmente sabe-se que após traduzida, a TPH2 tem sua localização subcelular no citosol de células serotoninérgicas principalmente da ponte, mas também é encontrada em menores níveis em diversas partes do cérebro como o córtex, núcleo estriado, hipocampo, putâmen, núcleo caudal e cerebelo. Além disso, também é encontrada no sistema nervoso periférico, apenas em neurônios entéricos do intestino. Já a TPH1 se mostra abundantemente expressa no citosol das células da glândula pineal, e nas células enterocromafins e mastócitos como previamente suposto (Windahl, et al, 2009; Mosienko, et al., 2015).

Sua estrutura monomérica possui 490 aminoácidos e aproximadamente 59kDa. Putativamente, os primeiros 150 aminoácidos da extremidade N-terminal compreendem o domínio regulatório; do 151 ao 459 o domínio catalítico; e em diante até o aminoácido final o domínio de tetramerização (Figura 1) (Carkaci-Salli, et a., 2006).

Embora a TPH1 e a TPH2 apresentem considerável homologia em sua sequência do domínio regulatório, a TPH2 contém 41 aminoácidos a mais, o que é

único desta enzima. Neste domínio, o resíduo S19 é candidato a fosforilação por proteína-cinase, além de mostrar interagir com proteínas 14-3-3 para aumentar a estabilidade. Em contrapartida, o domínio regulatório também contém sítios que reduzem a expressão enzimática por alteração da estabilidade e síntese da proteína. (Murphy, et al, 2008). Já seu domínio de tetramerização consiste em uma alfa-hélice que interagem entre si nos monômeros semelhantemente à interação que ocorre no motivo zíper de leucina. Dessa forma a estrutura quaternária é estabelecida, o que é observado em todas as hidroxilases. Tanto o domínio de regulação quanto o de tetramerização possuem importante papel na estabilidade e solubilidade da enzima. (Carkaci-Salli, et al., 2006).

1.3 Mutações Não Sinônimas de Nucleotídeo Único

Sendo assim, acredita-se que alterações na transcrição e funcionamento da proteína TPH2 poderiam estar relacionadas a psicopatologias devido a sua importante função em relação à serotonina (Chen; Miller, 2012). Mutações que ocorrem na sequência de nucleotídeo de um gene codificante são a causa que levam à alterações na funcionalidade proteica. Isso ocorre porque quando um aminoácido diferente é codificado, tal substituição pode implicar em mudança estrutural uma vez que a estrutura da proteína depende diretamente da sua estrutura primária. Conseqüentemente alterações na estrutura proteica podem afetar sua funcionalidade, já que estrutura e função são intrinsecamente relacionadas (Verli, 2014c; Ng; Henikoff, 2006).

A maioria das mutações genéticas ocorre por causa de variações ou polimorfismos de nucleotídeo único (SNVs, do inglês *Single Nucleotide Variations*; e SNPs, do inglês *Single Nucleotide Polimorphisms*). O termo “polimorfismo” é usado por definição quando uma mutação é encontrada em pelo menos 1% da população, enquanto que as variações possuem ocorrência abaixo dessa porcentagem (Bromberg; Rost, 2007). As mutações pontuais podem ser inserções, deleções ou substituições de uma única base do DNA que resultam na troca do aminoácido codificado naquela posição ou mudança do quadro de leitura, alterando assim o fenótipo (Crawford, et al., 2012). Quando essa troca leva à tradução de um aminoácido com características totalmente diferentes do original, a troca é dita não-sinônima. Porém, devido a redundância do código genético, pode ser que isso não ocorra, logo

a mutação é dita sinônima o (Bromberg; Rost, 2007). Assim, nsSNPs e nsSNVs podem originar proteínas cuja a função está alterada, devido a mudanças em sua estrutura e estabilidade.

Estudos prévios sugerem associação entre variações no gene e na proteína TPH2. À nível do gene, mas não causando substituições de aminoácido, mutações vem sendo associadas ao comportamento suicida (Sadkowski, et al., 2013), esquizofrenia paranoide (Xu, et al., 2014), dentre outros (Bragatti, et al., 2014; Wang, et al., 2015; Wang, et al. 2016). Já em relação a mutações não sinônimas de nucleotídeo único, a TPH2 vem sendo associada à desordem bipolar (Lin, et al., 2007; Cichon, et al., 2008), transtorno do déficit de atenção e hiperatividade (McKinney, et al., 2008) e depressão unipolar (Zhang, et al., 2005; Ma, et al., (2015).

A substituição de uma prolina para uma serina na posição 206 (P206S) da proteína TPH2 está associada à desordem bipolar (Cichon, et al., 2008), caracterizada por mudanças anormais no humor, energia e nível de atividade. Tais mudanças variam entre episódios de mania, quando o paciente se encontra extremamente disposto e com humor elevado e expansivo; e de depressão, quando se sente muito triste e sem esperança (NIH, 2013). Além de hereditária (Edvardsen, et al.; 2008), tal doença apresenta considerável taxa de suicídio entre seus pacientes (Guze; Robins, 1970; Cichon, et al., 2008).

Já a substituição de uma arginina por um triptofano na posição 303 (R303W) foi associada ao transtorno déficit de atenção e hiperatividade (TDAH) (McKinney, et al., 2008). O TDAH é uma desordem hereditária de alta prevalência na população, sendo caracterizada pela impulsividade, hiperatividade e incapacidade de manter a concentração (Hechtman, 1993). Tais sintomas interferem no aprendizado, e causam impactos à longo prazo, tanto acadêmico e profissional quanto em relacionamentos, além de estar associada ao uso abusivo de drogas e atividade criminal (Eloyan, et al., 2012).

Outra associação foi reportada a partir de estudo em que 9 pacientes com depressão unipolar que apresentavam a substituição de uma arginina por uma histidina na posição 441 (R441H) (Zhang, et al., 2005). A depressão unipolar também é conhecida como depressão maior. Seus sintomas são relacionados a distúrbios no

peso, apetite e sono, o que leva a morbidade e até mesmo a mortalidade, uma vez que é comum haver a presença de outras doenças com as cardiovasculares, obesidade e diabetes (Sullivan; Neale; Kendler, 2000; Ferrari, et al., 2013).

Diante da função da TPH2 e as implicações decorrente dos transtornos associados a mutações na estrutura desta proteína, torna-se evidente a importância da análise de sua estrutura selvagem e variantes. Visto isto, este trabalho aplicou métodos *in silico* e a dinâmica molecular para realização de tal análise, a fim de se obter mais informações sobre as possíveis consequências em sua estrutura e função causadas pelas substituições de aminoácidos. Portanto, a compreensão das bases moleculares é essencial para o auxílio do tratamento de transtornos psiquiátricos, uma vez que a TPH2 mostra ser um bom alvo em potencial para criação de novos fármacos e também marcador genético.

2. OBJETIVOS

Analisar por dinâmica molecular computacional os efeitos das variações de nucleotídeos únicos não sinônimas compiladas da proteína TPH2 pela comparação entre o comportamento dinâmico da estrutura da proteína selvagem e as mutantes a partir das suas trajetórias resultantes.

Avaliar a qualidade estrutural do modelo experimental usando algoritmos de validação, e obter uma estrutura tridimensional refinada pela DM a partir desta estrutura experimental, validando-a do mesmo modo.

3. METODOLOGIA

3.1 Obtenção da estrutura proteica experimental

A estrutura experimental da TPH2 selvagem (WT, do inglês *Wild Type*) foi obtida no banco de dados online Protein Data Bank (PDB) (Berman, et al., 2000), sendo seu código de identificação 4V06.

3.2 Compilação das variações não sinônimas

Em estudo anterior (De Freitas, 2016), as mutações pontuais não sinônimas da TPH2 conhecidas foram retiradas dos sites UniProt Knowledgebase (UniProtKB)

(Pundir; Martin; O'donovan, 2017), Online Mendelian Inheritance in Man (OMIM) (Online Mendelian Inheritance in Man, 2015) e do National Center for Biotechnology Information (NCBI) a partir dos bancos de pequenas variações genéticas (dbSNP) (Sherry, et al., 2001) e de periódicos científicos. Destes bancos de dados, o UniProtKB disponibiliza dados de todas as proteínas descritas; o OMIM contém informações detalhadas sobre genes e fenótipos humanos que são diariamente atualizadas; e o dbSNP contém anotações de variações genômicas.

3.3 Refinamento e Validação estrutural

A estrutura experimental obtida (PDB ID, 4V06) foi submetida a três servidores de refinamento: KoBaMIN (Rodrigues; Levitt; Chopra, 2012), GalaxyRefine (Heo; Park; Seok, 2013) e ModRefiner (Xu; Zhang, 2011). O primeiro realiza essa tarefa a partir da minimização de energia e de correções estereoquímicas da estrutura alvo (Rodrigues; Levitt; Chopra, 2012). O segundo usa uma técnica de perturbações estruturais repetidas, seguidas de relaxamento de toda a estrutura, pela dinâmica molecular (Heo; Park; Seok, 2013). Já o ModRefiner gera um modelo refinado a partir de traços de carbono alfa e de um campo de força que melhora as qualidades globais e locais da estrutura (Xu; Zhang, 2011).

Os modelos refinados resultantes foram posteriormente validados, o que consistiu na análise de sua qualidade, pelos algoritmos ERRAT (Colovos; Yeates 1993), Verify 3D (Eisenberg; Lüthy; Bowie, 1997), PROCHECK (Laskowski, et al., 1993), Rampage (Lovell, et al., 2003), QMEAN4 disponibilizado pelo QMEAN Server (Benkert; Kunzli; Schwede, 2009) e ProSA web (Wiederstein; Sippl, 2007). Todos os servidores e algoritmos foram utilizados em suas configurações padrões e gratuitamente.

3.4 Dinâmica Molecular curta

A princípio, a dinâmica molecular foi realizada no servidor MDWeb (Hospital, et al. 2012) que disponibiliza várias ferramentas e pacotes de dinâmica molecular tornando possível o estabelecimento do sistema de simulação, como também a realização e análise da mesma. O arquivo PDB da estrutura experimental foi submetido no MDWeb como *input*, servindo como a estrutura base de todas as

dinâmicas realizadas. Para a realização da dinâmica com as estruturas mutantes, a opção “*Mutate residue*” na estrutura base foi selecionada e a substituição de resíduo foi realizada pelo software Visual Molecular Dynamics (VMD) (Humphrey; Dalke; Schulten, 1996). Posteriormente, todos os passos foram realizados igualmente para selvagem e mutantes.

A opção “*Gromacs FULL MD Setup*” em “lista de operações” das estruturas foi selecionada a fim de se estabelecer os padrões da dinâmica molecular. O campo de força utilizado foi o AMBER-99SB*, uma vez que descreve satisfatoriamente os aspectos do comportamento molecular de proteínas. O próximo passo foi escolher o conjunto de condições NPT (número de partículas, pressão e temperatura constantes) para ser usado na dinâmica, com a opção “*Simple Box Solvent Molecular Dynamics (NPT)*”. Então, as simulações ocorreram por 2,5 ps e com tempo de integração de 2 fs, produzindo 25 snapshots de trajetória.

Algumas análises foram realizadas após a simulação da dinâmica molecular para a trajetória de todas as estruturas. Para isso, primeiramente foi escolhida a opção “*Remove water molecules and ions from trajectory*” a fim de se retirar o solvente e os íons necessários no passo anterior gerando-se uma trajetória seca. Para avaliar desvios e variações das estruturas moleculares durante a simulação, análises do RMSD (do inglês, *Root Mean Square Deviation*) da trajetória e do Raio de Giro (Rg) foram feitas. Já para avaliar a flexibilidade das estruturas durante a trajetória, foram selecionadas as análises de RMSD por resíduo e B-factor.

3.5 Dinâmica Molecular longa

O pacote do GROMACS do Linux versão 5.0.7 (Van Der Spoel, et al., 2005; Abraham, et al., 2015) foi usado para realizar a dinâmica molecular longa, tanto da proteína selvagem quanto para as mutantes selecionadas. Para isso, a estrutura cristalográfica foi utilizada como estrutura selvagem e também foi modificada para obtenção das estruturas variantes usando o software VMD (Humphrey; Dalke; Schulten, 1996). Para descrever a estrutura e dinâmica da proteína corretamente, o campo de força Amber99SB-ILDN (Lindorff-Larsen, et al., 2010) foi escolhido. As estruturas foram solvatadas em caixa cúbica com água explícita TIP3P, e íons de Na⁺Cl⁻ foram adicionados para neutralizar o sistema. Dois tempos de minimização de

energia foram realizados pelo método *steepest descent* por 1000 e 5000 passos, respectivamente, com objetivo de eliminar maus contatos nas estruturas, antes e depois da adição de água. Dois períodos de equilíbrio das estruturas também foram realizados, primeiro para estabelecer as condições NVT (número de partícula, volume e temperatura constantes) com temperatura de 300K por 100ps; e então as condições NPT (número de partícula, pressão e temperatura constantes) foram estabelecidas a 1 atm, pelo mesmo tempo. As simulações foram então feitas a 300K por 80ns, fazendo-se uso do algoritmo *LINCS* (do inglês, *Linear Constraint Solver*) (Hess, et al., 1997) para restringir interações específicas, e o *PME* (do inglês, *Particle Mesh Ewald*) (Essman, et al., 1995) para evitar o truncamento abrupto do potencial de interação devido ao aumento da distância entre os átomos.

Para estudar as diferenças estruturais e funcionais entre a proteína selvagem e as mutantes, cada uma delas foi submetida a análises de estabilidade e flexibilidade obtidas pelos valores do RMSD, Raio de Giro, RMSF, B-factor e número de ligações de hidrogênio (NHB, do inglês *Number of Hydrogen Bonds*), assim como à análise da Área de Superfície Acessível ao Solvente (SASA, do inglês *Solvent Accessible Surface Area*). A partir dos resultados obtidos foram criados gráficos pela ferramenta XMGrace (Turner, 2005), enquanto que os *outputs* em formato .pdb foram visualizados com o programa Chimera (Pettersen, et al., 2004).

3.5.1 Análise de Componente Principal (PCA)

Para estabelecer as principais mudanças conformacionais dos sistemas durante a simulação, a Análise de Componente Principal (PCA, do inglês *Principal Component Analysis*), também provida pelo pacote GROMCAS, foi utilizada. Tal técnica estatística permite a exibição de grande parte dos movimentos da proteína extraídos da decomposição de movimentos correlacionados. Uma vez que os movimentos essenciais são obtidos pela dissecação de um conjunto de conformações, a aplicação do PCA é chamada de Dinâmica Essencial (ED, do inglês *Essential Dynamics*) (David; Jacobs, 2014). Assim, a PCA permite a obtenção de uma imagem geral dos movimentos atômicos da estrutura, logo a determinação da flexibilidade e dinâmica conformacional (Ndagi; Mhlongo; Soliman, 2017). A técnica foi realizada para os átomos do carbono- α durante o período de 80ns da simulação de dinâmica. As coordenadas cartesianas de tais átomos foram utilizadas para

geração da matriz de covariância e os dois primeiros componentes foram utilizados para geração do resultado gráfico.

4. RESULTADOS E DISCUSSÃO

4.1 Obtenção da estrutura proteica experimental

Apenas uma estrutura experimental foi encontrada, sendo o seu código de identificação no Protein Data Bank o PDB ID: 4V06. Este modelo foi obtido por difração de Raio X como um dímero, além de conter uma pequena sequência da posição 142 a 147 correspondente ao seu *tag* de expressão e resolução de 2,63 Å. O monômero experimental possui 342 aminoácidos, pois abrange apenas o domínio catalítico e o de oligomerização. O domínio catalítico começa no aminoácido 151, caracterizado por um pentapeptídeo conservado que é observado em todas as enzimas dessa família (Carkaci-Salli, et al. 2006). Já o domínio de oligomerização compreende as posições 460 a 490 (Tideman, et al., 2016), sendo formado por uma alfa-hélice que possibilita a formação de dímeros e tetrâmeros.

Sendo assim, para alcançar os propósitos do estudo, o arquivo foi editado e apenas a estrutura da cadeia A monomérica sem o *tag* de expressão foi utilizada, isto é, do aminoácido de posição 148 ao 490.

4.2 Compilação das variações não sinônimas

A compilação realizada obteve um total de 46 mutações distribuídas por toda a extensão da proteína (Tabela 1), obtidas do UniProtKB, OMIM, dbSNP e literatura.

No UniProtKB, a TPH2 humana está registrada com o código Q8IWU9 e possui informações revisadas com evidência experimental no nível da proteína. A partir dessa fonte foram selecionadas todas as variações classificadas como “variantes naturais”, totalizando 11 variantes. Destas, apenas três também constavam no OMIM (TPH2 MIM number: *607478) como substituições sabidamente relacionadas a patologias segundo a literatura, à saber a P206S, R303W e R441H.

Já no dbSNP, que retornou o maior número de variações, foi necessário refinar a busca fazendo-se uso dos filtros: TPH2; Organismo: *Homo sapiens*; classe de

variação: snp; classe funcional: *missense*; e validação: por 1000 *Genomes*, *cluster* e *frequency*'.

TABELA 1 - MUTAÇÕES COMPILADAS DOS BANCOS DE DADOS UNIPROT, OMIM, PUBMED E dbSNP.

Posições		Mutações	Posições		Mutações
1	22	S → L	23	225	R → Q
2	25	P → L	24	276	R → S
3	32	G → D	25	277	P → L
4	36	L → V e L → P	26	303	R → W
5	41	S → Y	27	328	A → V
6	53	S → R	28	339	I → M
7	55	R → C	29	345	G → E
8	64	T → I	30	348	D → H
9	65	A → T	31	363	E → K
1	78	V → I	32	378	E → T
11	82	R → S	33	383	S → F
12	83	L → V	34	396	C → R
13	86	E → Q	35	404	T → K
14	88	R → H	36	430	E → G
15	91	M → I	37	432	M → I
16	100	R → Q	38	436	A → E
17	101	R → Q	39	441	R → H
18	124	Q → R	40	471	R → H e R → K
19	134	T → M	41	473	D → N
20	145	E → Q	42	474	L → V
21	156	R → Q	43	479	D → E
22	206	P → S	44	486	Q → K

Fonte: Dados de pesquisa, (2015). A cor roxa destaca a posição que consiste num sítio de ligação ao ferro.

Ao todo, 46 mutações foram compiladas dessas fontes, sendo elas localizadas por todos os três domínios da proteína. Porém, apenas as que se encontravam no domínio catalítico e de tetramerização foram analisadas por dinâmica molecular curta

(um total de 25 mutações), uma vez que tais mutações previamente mostraram ser mais danosas à funcionalidade da proteína segundo análise de predição funcional (De Freitas, 2016); e somente as mutações P206S, R303W e R441H, já associadas a desordens psiquiátricas pelo OMIM, foram submetidas à dinâmica molecular longa, visto que tal técnica demanda mais tempo e desempenho computacional. Para essas três mutantes, constatou-se que foram classificadas como danosas por pelo menos 50% dos preditores utilizados (De Freitas, 2016).

4.3 Refinamento e Validação estrutural

Antes da realização das simulações de dinâmica longa pelo pacote GROMACS, a qualidade estrutural do modelo 4V06 foi avaliada, uma vez que foi observado que sua resolução era de 2,63 Å. Desse modo, buscou-se tanto avaliar sua qualidade estrutural, como se obter uma estrutura refinada de qualidade superior por refinamento. A necessidade da avaliação de qualidade estrutural deve-se ao fato de que, eventualmente, modelos obtidos por cristalografia de Raio X podem conter artefatos, i.e., proteínas que não a desejada elucidadas juntamente com a proteína alvo. Isso ocorre uma vez que o processo para obtenção da estrutura pode não ser bem-sucedido por diversas razões, tal quando outra proteína diferente do alvo é obtida durante a purificação; ou a amostra está contaminada ou até mesmo rotulada erroneamente (Niedzialkowska, et al., 2016). Assim, o arquivo pdb do modelo experimental da TPH2 foi primeiramente submetido a algoritmos de refinamento e posteriormente de validação para se assegurar que o modelo 4V06 foi elucidado corretamente, e possibilitar a comparação e avaliação entre os modelos aprimorados gerados e o original.

Os algoritmos KoBaMIN e ModRefiner retornaram como resultado apenas um modelo refinado cada, ambos nomeados de Modelo 1. Já o GalaxyRefine resultou em cinco modelos, diferenciados dos números de 1 a 5.

Os resultados obtidos pelos algoritmos de validação (Tabela 2) foram comparados a fim de se observar melhorias em comparação à estrutura experimental, assim como escolher o melhor modelo a ser utilizado como estrutura selvagem e base para obtenção das mutantes.

O ERRAT gera resultado a partir de uma função de erro baseada na estatística de interações entre átomos não ligados. Ele é capaz de diferenciar regiões corretas e incorretas, uma vez que uma distribuição de átomos randômica foge à teoria de que os mesmos estão organizados de modo não-aleatório na proteína. Estruturas consideradas boas apresentam menos de 5% de regiões com um valor de erro abaixo de 95%. A partir da pontuação do valor de erro, uma porcentagem para qualidade geral da proteína é obtida, sendo essa chamada de fator de qualidade. Assim, é considerado que estruturas de alta resolução apresentem mais de 95% de qualidade geral, enquanto que as de resolução mais baixa apresentem qualidade geral média de 91%. No entanto, pontuações acima de 50% são consideradas aceitáveis (Colovos; Yeates 1993) (Figura 2a).

A estrutura experimental mostrou ter resolução abaixo da considerada alta, com fator de qualidade geral igual a 92,49%. Considerando os resultados do ERRAT para todos os modelos refinados, o Modelo 1 do KobaMIN se destacou dos demais, com 98,51% de qualidade geral, embora o modelo 2 também tenha obtido um valor alto (95,49%). Os demais modelos apresentaram pontuação abaixo de 95%, porém sendo a menor pontuação obtida de 89,49%.

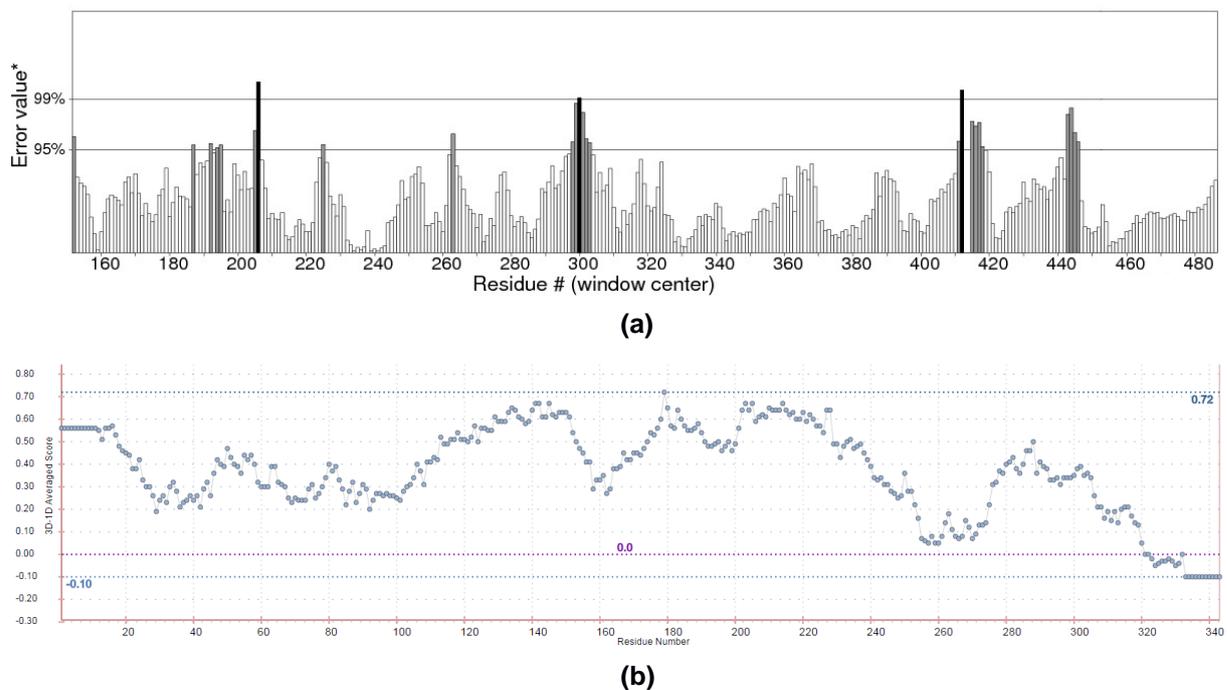


Figura 2 - Resultados gráficos retornados pelos validadores ERRAT e Verify 3D para validação da estrutura 4V06. (a) Gráfico provido pelo ERRAT constando a pontuação de erro atribuída a cada resíduos. Dois limiares facilitam a identificação dos resíduos que obtiveram valores de erro acima de 95% e 99%. (b) Gráfico disponibilizado pelo Verify 3D relacionando cada resíduo à pontuação média

de comparação entre a estrutura tridimensional e a primária.

O Verify 3D checa a acurácia do modelo pela comparação de compatibilidade entre a estrutura tridimensional (3D) do modelo e a estrutura primária da proteína (1D). De modo semelhante, essa acurácia é analisada segundo uma percentagem geral, neste caso obtida segundo a avaliação dos resíduos que obtiveram a média 3D-1D maior ou igual à 0,2 (Eisenberg; Lüthy; Bowie, 1997) (Figura 2b). Assim, quanto maior a porcentagem geral mais preciso é o modelo, sendo aceitável validações acima de 80%. Foi observado que todos os modelos obtiveram mais de 80% dos resíduos na média estipulada. O destaque foi para o Modelo 2 do GalaxyRefine (88,34%), e também para o Modelo 5 que obteve porcentagem um pouco inferior (85,71%).

A análise estereoquímica das estruturas feitas pelo PROCHECK e Rampage verifica a estabilidade conformacional da estrutura levando em consideração as combinações de angulação *phi* (Φ) e *psi* (Ψ) que ocorrem nos aminoácidos, mais especificamente nas ligações entre o carbono alfa e suas cadeias amino e carboxila do resíduo. A partir disso, aminoácidos fora das áreas energeticamente favoráveis são identificados, o que causa dúvida sobre a validade estereoquímica do modelo, pois combinações impróprias são relacionadas à erro. Tais resultados são mostrados no gráfico de Ramachandran, o qual é dividido em 4 regiões para o PROCHECK e 3 para o Rampage, desde a área mais favorável à não permitida ou *outlier* (Laskowski, et al., 1993; Lovell, et al., 2003) (Figura 3).

Para análise do gráfico e valores gerados pelo PROCHECK, foi desejado que a estrutura obtivesse mais de 90% de resíduos na área energeticamente mais favorável e 0% na área proibida (Figura 3a). Para análise dos resultados do Rampage, considerou-se que houvesse mais de 98% dos resíduos na área favorável e igualmente 0% na *outlier* (Figura 3b). Todos os modelos gerados obtiveram os valores dentro dos padrões considerados. Porém, destacaram os Modelos 1, 2 e 3 obtidos pelo GalaxyRefine e também o único modelo obtido pelo ModRefiner. Apesar desses serem os modelos com maior número de resíduos em área favorável, os Modelos 4 e 5 também obtiveram bons resultados, embora ligeiramente inferiores.

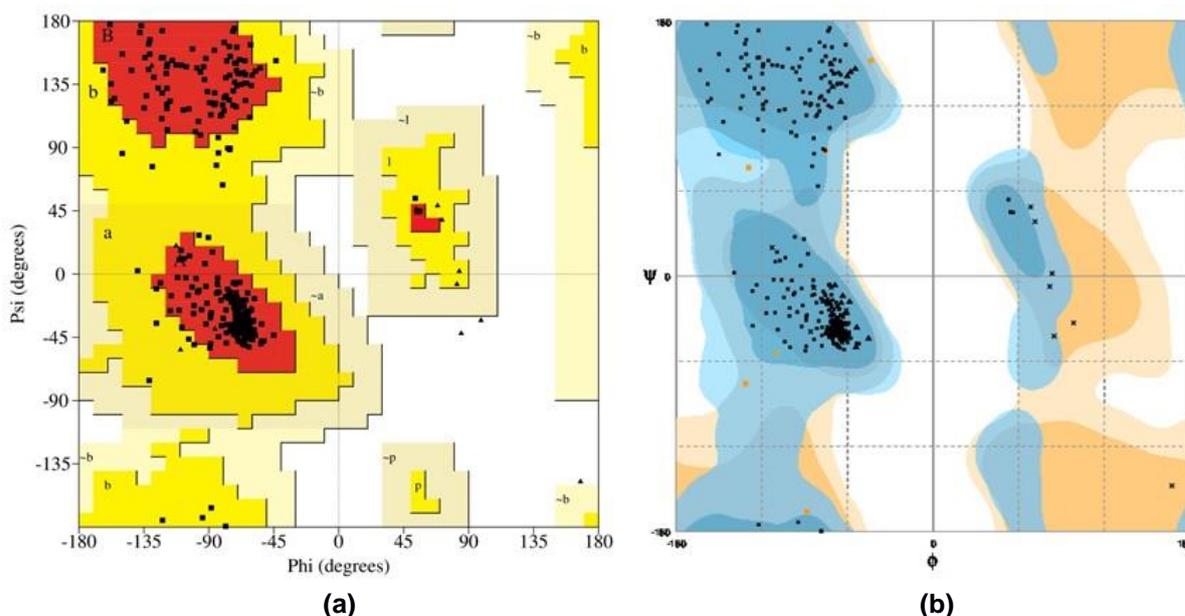


Figura 3 – Gráficos de Ramachandran gerados pelo PROCHECK e Rampage como resultado da validação estrutural do modelo 4V06. (a) Gráfico gerado pelo PROCHECK. Os quadrados representam as combinações dos ângulos de torção phi e psi para resíduos que não glicina ou prolina. Os triângulos representam as combinações para glicina e prolina. A região vermelha, também chamada de A, B e L, consiste na região mais favorável. A cor amarela (a, b, l e p) consiste na região adicionalmente permitida. A cor bege (a, b, l e p) consiste na região liberalmente permitida. A cor branca compreende a região proibida para resíduos que não prolina e glicina. (b) Gráfico gerado pelo Rampage. Os quadrados representam as combinações dos ângulos de torção phi e psi para resíduos que não prolina e glicina. Os triângulos representam a plotagem da combinação dos ângulos para prolinas, e o X para glicinas. A região azul escura representa a mais favorável para os resíduos em geral e prolinas, enquanto a região azul clara representa a região permitida. As regiões laranja representam a área favorável apenas para glicinas, assim como a região laranja claro é permitida apenas para esses mesmos resíduos.

Já o QMEAN Server disponibiliza o QMEAN4, uma função de pontuação da qual deriva valores para mensuração da qualidade absoluta do modelo. Ele é calculado a partir de quatro termos estatísticos, e por isso é dito composto (Figura 4a). A partir dele provém uma pontuação normalizada chamada Z-score. O QMEAN4 Z-score que avalia o quanto a qualidade da estrutura desvia do padrão observado para estruturas de alta qualidade e mesmo tamanho usadas como referência (Figura 4b). Assim, valores mais próximos de 0 são desejados (Benkert; Kunzli; Schwede, 2009; Benkert; Biasini; Schwede, 2011). Baseado nas pontuações resultantes para os modelos validados, os Modelos 1, 3 e 4 do GalaxyRefine se destacaram obtendo os valores mais próximos de 0 (-0,57, -0,57 e -0,50 respectivamente) significando que tais estruturas desviaram menos que 1 desvio padrão dos Z-scores observados para as proteínas de referência.

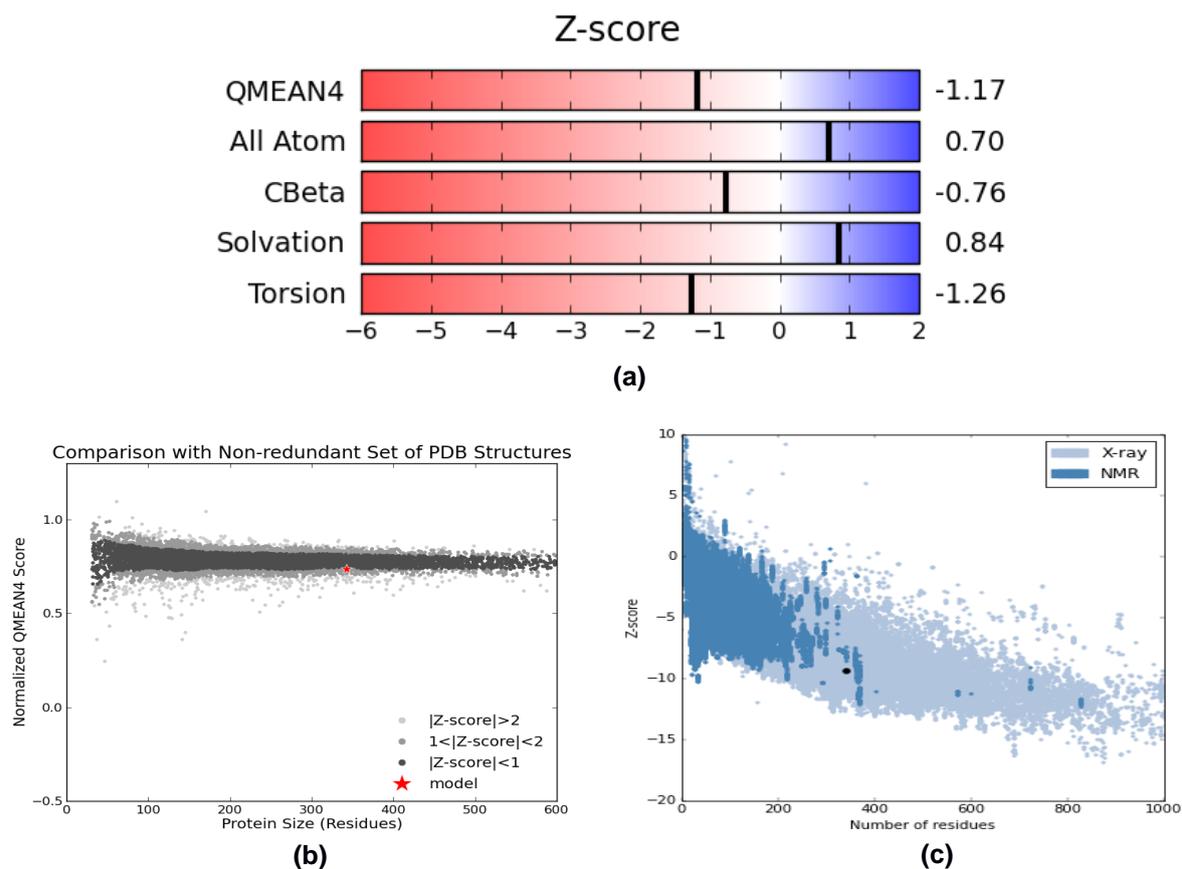


Figura 4 – Resultados fornecidos pelo QMEAN Server e ProSa para a validação estrutural do modelo 4V06. (a) Pontuações normalizadas (Z-score) obtidas para o QMEAN4 e os quatro termos que o compõe. (b) Gráfico comparativo do QMEAN4 Z-score para 4V06 em comparação com a pontuação de proteínas de referência e mesmo tamanho. (c) Gráfico comparativo entre o Z-score resultado do ProSa para 4V06 e a pontuação obtida por proteínas de referência e mesmo tamanho, obtida por Raio X (área azul clara) ou Ressonância Magnética Nuclear (azul escuro).

A análise realizada pelo ProSA é similar à feita pelo QMEAN Z-score. É desejável que o valor da pontuação esteja dentro da extensão observada para proteínas já elucidadas por Ressonância Nuclear Magnética ou Raio-X e referência em qualidade, intervalo esse mostrado em forma gráfica (Wiederstein; Sippl, 2007) (Figura 4c). Todos os modelos apresentaram, segundo o ProSA, qualidade de estrutura experimental obtida por Raio-X, como de fato a estrutura original foi obtida. Como os valores variaram pouco entre si, nenhuma estrutura apresentou resultado de Z-score a ser destacado.

Sendo assim, considerando todos os resultados e buscando o modelo que obteve os valores ótimos na maioria dos algoritmos de validação, concluiu-se que o Modelo 2 do GalaxyRefine possuiu a melhor qualidade observada. No entanto, foi decidido que a estrutura experimental original deveria ser utilizada, uma vez que mostrou possuir boa qualidade.

4.4 Dinâmica Molecular

A fim de se analisar o comportamento da proteína selvagem e as suas variantes selecionadas para posterior comparação, duas dinâmicas foram realizadas: a primeira sendo por um servidor e em intervalo de tempo pequeno, com objetivo de definir os parâmetros; e a segunda, pelo pacote GROMACS e com duração maior. A estrutura experimental 4V06 foi utilizada como estrutura selvagem, após edição de seu arquivo PDB, enquanto que as variantes foram obtidas por mutagênese *in silico* pelo software VMD.

Após as simulações realizadas tanto pelo MD-Web quanto pelo GROMACS, os valores do RMSD foram utilizados para analisar o equilíbrio e estabilização da estrutura proteica e seu sistema durante a trajetória. O Raio de Giro resultou em valores referentes à compactação estrutural, e assim a dimensão geral da proteína. Já os valores do RMSF (análogo ao RMSD por resíduo) e B-factor foram utilizados para analisar a flexibilidade local dos resíduos das estruturas.

Quatro análises foram realizadas com a utilização do pacote GROMACS, a quantificação da área de superfície acessível ao solvente (SASA) e do número de pontes de hidrogênio estabelecidas tanto dentro da própria proteína (intramoleculares), quanto entre a proteína e o solvente (intermoleculares), além da Análise de Componente Principal (PCA). Esta última avalia a conformação e flexibilidade das estruturas fazendo-se uso de uma metodologia estatística, diferente das análises fornecidas pelo RMSF e B-factor.

4.4.1 Dinâmica Molecular curta

Esta dinâmica consistiu em simulações curtas de 2,5 ps, sendo importante ressaltar que os resultados por ela gerados são preliminares, uma vez que certas análises necessitam de maior intervalo de tempo para obter resultados válidos.

TABELA 2– RESULTADOS DOS ALGORÍTMOS PARA ESTUDO DE VALIDAÇÃO DAS ESTRUTURAS REFINADAS.

VALIDAÇÃO									
		PROCHECK			Rampage		QMEAN4	ProSA	
	Estrutura	ERRAT	Verify 3D	Mais favorável	Menos favorável	Mais favorável	Menos favorável	z-score	z-score
Experimental	4V06	92,49%	84,26%	92,3%	0%	98,2%	0%	-1,17	-9,41
GalaxyRefiner	Modelo 1	89,49%	84,26%	97,4%	0%	98,8%	0,3%	-0,57	-9,84
	Modelo 2	95,49%	88,34%	97%	0%	98,8%	0,3%	-0,72	-9,84
	Modelo 3	93,35%	84,26%	96,4%	0%	99,1%	0,3%	-0,57	-9,84
	Modelo 4	93,09%	80,76%	97%	0%	98,2%	0,3%	-0,50	-9,95
	Modelo5	93,35%	85,71%	96,4%	0%	98,8%	0,3%	-0,65	-9,77
ModRefiner	Modelo 1	*	83,09%	95,1%	0%	99,1%	0%	-1,23	*
KobaMIN	Modelo 1	98,51%	85,13%	96,4%	0%	98,02%	0,3%	-0,78	-9,78
Dinâmica Molecular	4V06 DM	*	90,8%	96,1%	0%	98,5%	0,3%	-0,80	-8,90

Fonte: Dados de Pesquisa, (2016, 2017). * Sem resultado devido a erro.

O RMSD mede a distância média entre átomos de estruturas sobrepostas, assim sendo comumente utilizado para comparação de similaridade (Verli, 2014a). Além disso, os valores de RMSD podem também trazer informações sobre a estabilização do sistema, isto é, o momento no qual a estrutura passa a convergir sobre sua conformação média mais estável (Yang, et al., 2014; Martínez, 2015; Joshi, et al., 2017). De modo geral, no início da simulação os valores tendem a aumentar bruscamente enquanto as estruturas buscam equilibrar-se até alcançarem um platô que sugere que as tais atingiram o equilíbrio. Como os valores de RMSD são dados em função do tempo, essa análise permite observação do período de tempo que as estruturas levam para se estabilizar. Neste caso, a diferença foi medida entre os átomos do *backbone* estrutura de partida (o modelo cristalográfico) e cada estrutura obtida durante os *frames* subsequentes da trajetória de simulação.

Assim, os resultados das análises de RMSD da trajetória (Figura 5) para as estruturas mutantes foram comparados aos da proteína selvagem a fim de se notar as diferenças de comportamento no equilíbrio e estabilidade das estruturas.

Observou-se para a estrutura selvagem que os valores do RMSD foram crescentes durante a trajetória, como esperado. Os valores variaram entre 0,080nm e 0,090nm a partir do 8° snapshot. Após o 19° snapshot, o valor se manteve relativamente estável a 0,092nm até o final da trajetória.

A partir disso, foi observado que apenas as mutações R225Q, I339M e Q486K (Figura 6a) apresentaram RMSD médio abaixo do observado para a selvagem. Já as mutações R156E, P206S, D348H, S383F, T404K, A436E e R441H (Figura 6b) mostraram valores de RMSD acima dos observados para a selvagem. As demais mutações não citadas obtiveram valores similares. Sendo assim, essas 10 mutações mostraram alterar o valor de RMSD quando comparadas a WT, distinguindo-se mais (RMSD maiores) ou menos (RMSD menores) da estrutura inicial do que o esperado (observado para a selvagem).

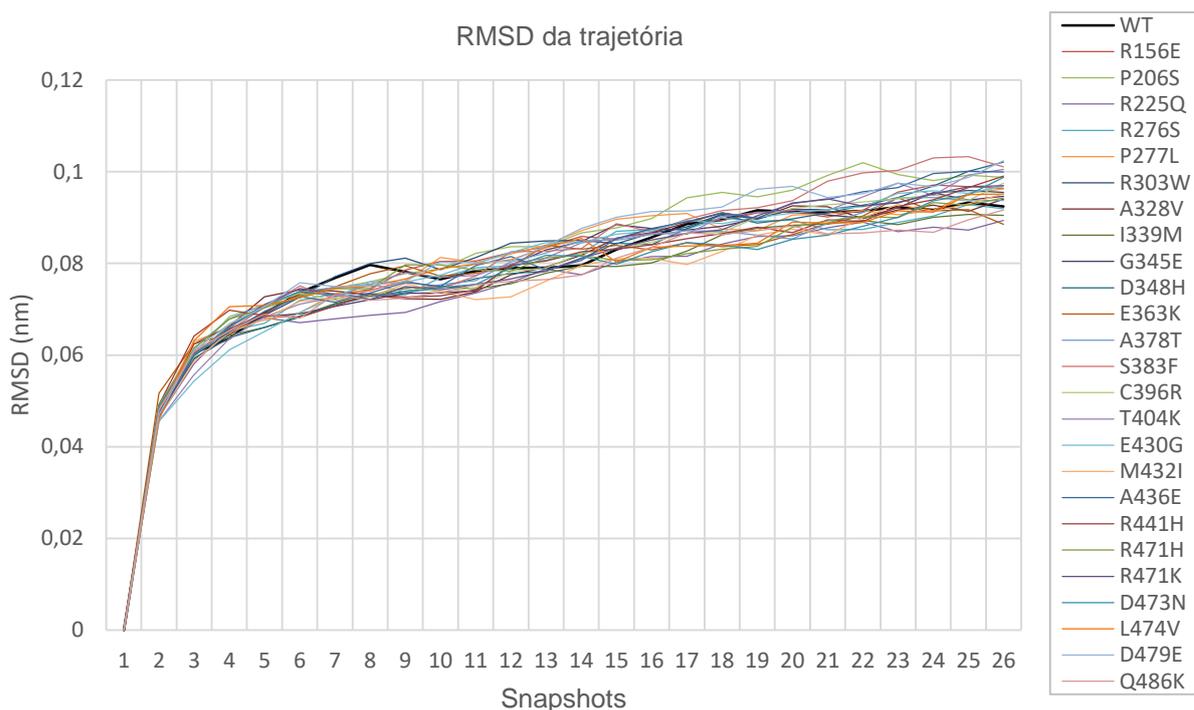


Figura 5 – Gráfico contendo os valores de RMSD em nanômetro por snapshot das estruturas proteicas selvagem e mutantes obtidos no MDWeb para estudo de estabilidade.

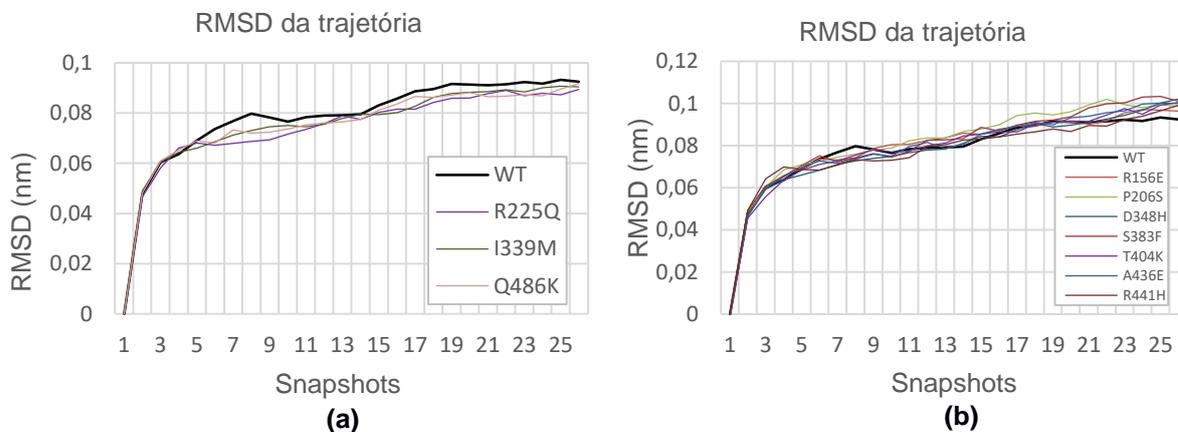


Figura 6 - Gráficos contendo as mutações para quais se observou valores diferentes de RMSD quando comparadas aos valores obtidos pela estrutura selvagem. (a) Gráfico contendo as 3 mutações cujos valores de RMSD foram menores do que os obtidos pela selvagem. (b) Gráfico contendo as 7 mutações cujos valores de RMSD foram maiores quando comparados com a estrutura selvagem.

Quanto a estabilização do sistema e o momento no qual isso acontece, tal simulação em curto período de tempo não forneceu tempo suficiente para que os sistemas entrassem em equilíbrio. Tal resultado foi atingido durante a simulação longa discutido à frente.

Outra análise a respeito da conformação das estruturas foi realizada a partir dos valores de Raio de Giro, que foi utilizado para analisar a dimensão estrutural da proteína, ou seja, sua compactação e também enovelamento e desnaturação (Ishak, et al., 2017). Tais valores são obtidos a partir da medição da distância entre os átomos de carbono alfa (considerando sua massa e posição) e o centro de massa da molécula. Ao comparar os valores obtidos entre mutantes e selvagem, sugere-se que valores abaixo ao da estrutura selvagem representem um aumento da compactação geral da estrutura mutante, ou seja, uma diminuição na dimensão geral da molécula. Já valores de Rg acima dos observados para a selvagem sugerem o contrário. Valores constantes sugerem que a proteína está estavelmente enovelada.

Sendo assim, para a estrutura selvagem, observou-se que o Rg se manteve entre os valores de 2,150nm e 2,160nm a maior parte da trajetória. Porém, no 23º snapshot esse valor diminuiu, chegando a 2,147nm no tempo final. Já para as mutantes, foi observado de modo geral que os valores de Rg para todas as mutações foram abaixo dos valores obtidos para a selvagem, indicando que as mutações levam a um aumento da compactação da estrutura durante a trajetória (Figura 7). As estruturas com as mutações A328V, E363K, E430G, R471K e D473N apresentaram os maiores aumentos (Figura 8). A pequena variação de valores sugeriu estruturas estáveis quanto ao seu empacotamento.

As análises de flexibilidade foram comparadas de modo semelhante utilizando os resultados do RMSD por resíduo e B-factor. O RMSD por resíduo (ou RMSF) mede a distância entre átomos de estruturas sobrepostas (Ishak, et al., 2017) semelhante à análise de RMSD da cadeia principal, mas neste caso entre o carbono alfa de cada resíduo sobreposto. Desse modo, é usado para analisar a flexibilidade, sendo interpretado como a mobilidade desses átomos durante a simulação. Assim, valores altos de RMSD por resíduo caracterizam alta flexibilidade do mesmo.

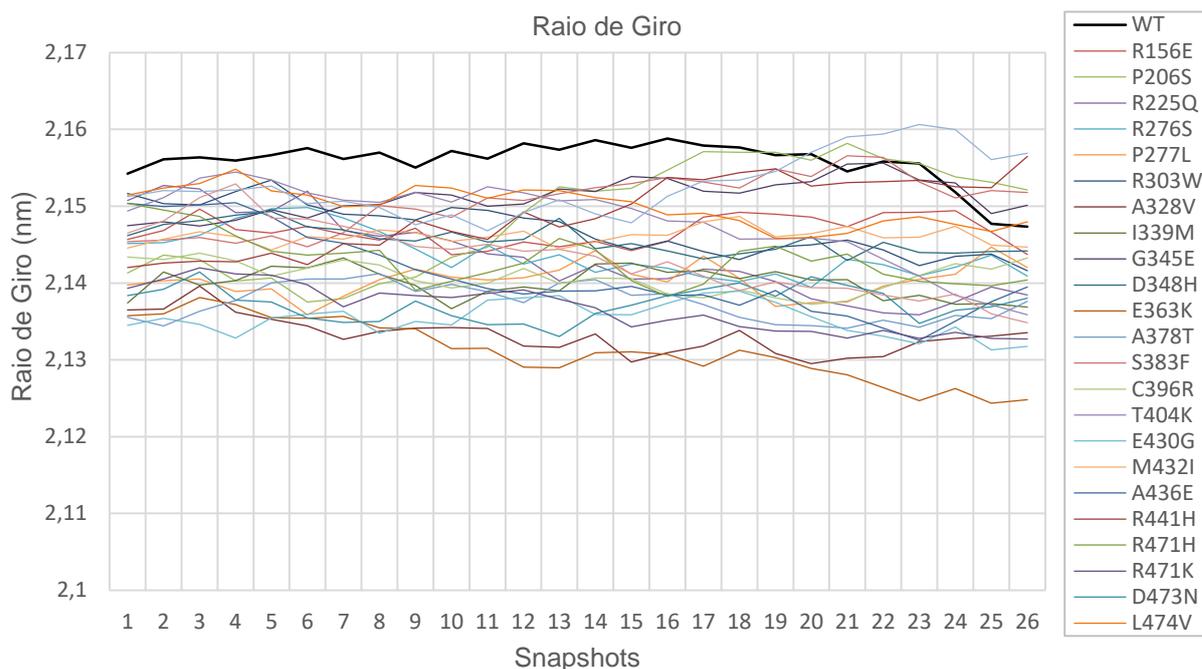


Figura 7 – Gráfico resultante da análise de Raio de Giro em nanômetros por snapshot obtidos no MDWeb para estudo de compactação das estruturas proteicas durante a simulação da dinâmica. Gráfico comparativo dos valores de raio de giro observados para a proteína selvagem e suas mutantes.

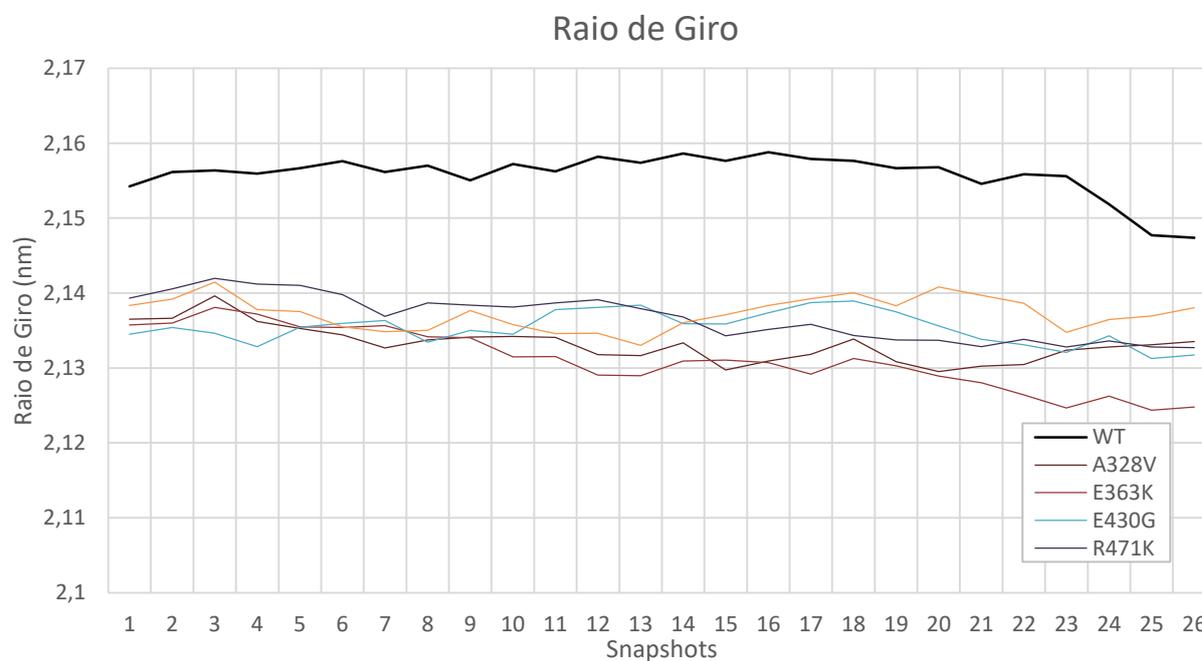


Figura 8 - Gráfico comparativo contendo as 5 mutações relacionadas a causar maior aumento na compactação da estrutura em comparação com a estrutura selvagem.

A estrutura selvagem mostrou RMSF variante entre um máximo de 0,889Å e mínimo de 0,249Å, mas a maioria dos valores se mantiveram entre 0,3 e 0,6Å (Figura 9). Foi observado que algumas mutações se diferenciaram notavelmente da selvagem por apresentarem picos nos valores de RMSF maiores que 1Å, sendo elas: R156E, P277L, R303W, S383F, C396R, E430G, R441H, R471K, D473N e D479E.

Já o B-factor, ou fator de temperatura, indica a mobilidade termal inerente dos átomos da proteína e assim fornece esclarecimentos sobre as mudanças de flexibilidade da estrutura durante a trajetória (Kuzmanic; Zagrovic, 2010). Mais uma vez, o aumento dos valores analisados e comparados à selvagem caracteriza um aumento regional da flexibilidade (Figura 10). Para a estrutura selvagem, foi observado B-factor entre um mínimo de 1,631Å² e 20,800 Å², sendo que a média obtida para os valores foi de 4,876 Å².

As mutações R156E, R225Q, P277L, R303W, G345E, S383F, C396R, E430G, R441H, R471K, D473N e D479E apresentaram valores de B-factor notáveis acima de 25Å², mostrando que essas variações levam a um aumento considerável da mobilidade desses resíduos e regiões adjacentes em relação à proteína selvagem.

Assim, considerando todas as análises de relevância feitas por simulações de dinâmica molecular curta, e somando-se a elas os resultados prévios de predição funcional dos nsSNVs e nsSNPs (Tabela 3), pode-se concluir que as mutações R156E, R225Q, P277L, R303W, S383F, E430G, R441H e D473N são as mutações que mais mostraram afetar a estrutura (estabilidade e flexibilidade) e função da proteína, sendo excelentes alvo para estudos posteriores. Tal resultado, apesar de prévios, estão de acordo como esperado, uma vez que destacou as mutantes R303W e R441H já associadas à psicopatologias.

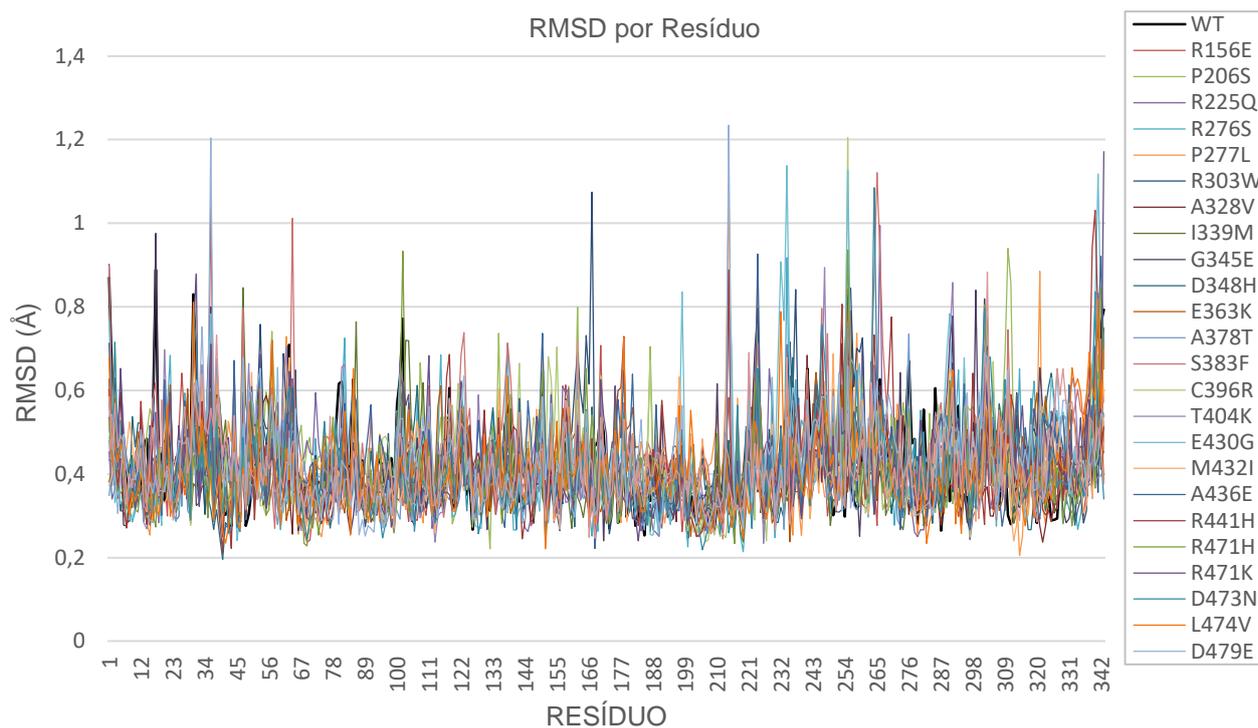


Figura 9 – Gráfico resultante da análise de RMSD em angstrom por resíduo obtidos no MDWeb para estudo de flexibilidade das estruturas proteicas durante a simulação da dinâmica. Gráfico comparativo dos valores de RMSD por resíduo observados para a proteína selvagem e suas mutantes.

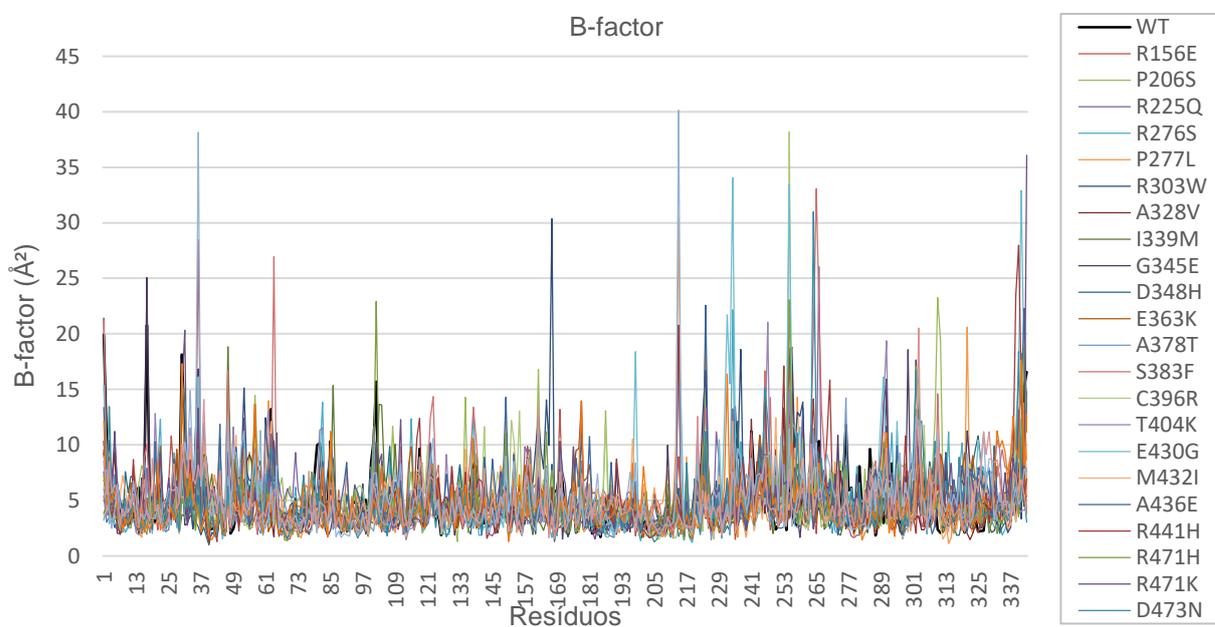


Figura 10 – Gráfico resultante da análise de B-factor em angstrom quadrado por resíduo obtidos no MDWeb para estudo de mobilidade termal das estruturas proteicas durante a simulação da dinâmica. Gráfico comparativo dos valores de B-factor observados para a proteína selvagem e suas mutantes.

TABELA 3- COMPARAÇÃO ENTRE OS RESULTADOS DAS ANÁLISES OBTIDAS PELO MDWEB E DA PREDIÇÃO FUNCIONAL* ANTERIORMENTE REALIZADA.

Mutações	RMSD da trajetória	Raio de Giro	RMSF	B-factor	Predição Funcional*	Clas. Final
R156E	Reduz	Aumenta	Aumenta	Aumenta	Deletéria	
P206S	Reduz	Aumenta/Reduz	Similar	Similar	-	
R225Q	Aumenta	Aumenta	Similar	Aumenta	Deletéria	
R276S	Similar	Aumenta	Similar	Similar	Deletéria	
P277L	Similar	Aumenta	Aumenta	Aumenta	Deletéria	
R303W	Similar	Aumenta	Aumenta	Aumenta	Deletéria	
A328V	Similar	Aumenta	Similar	Similar	-	
I339M	Reduz	Aumenta	Similar	Similar	-	
G345E	Similar	Aumenta/Reduz	Similar	Aumenta	Deletéria	
D348H	Reduz	Aumenta	Similar	Similar	Deletéria	
E363K	Similar	Aumenta	Similar	Similar	Deletéria	
A378T	Similar	Aumenta	Similar	Similar	Deletéria	
S383F	Reduz	Aumenta/Reduz	Aumenta	Aumenta	Deletéria	
C396R	Similar	Aumenta	Aumenta	Aumenta	-	
T404K	Reduz	Aumenta	Similar	Similar	-	
E430G	Similar	Aumenta	Aumenta	Aumenta	Deletéria	
M432I	Similar	Aumenta	Similar	Similar	-	
A436E	Reduz	Aumenta	Similar	Similar	Deletéria	
R441H	Reduz	Aumenta/Reduz	Aumenta	Aumenta	Deletéria	
R471H	Similar	Aumenta	Similar	Similar	Deletéria	
R471K	Similar	Aumenta	Aumenta	Aumenta	-	
D473N	Similar	Aumenta	Aumenta	Aumenta	Deletéria	
L474V	Similar	Aumenta/similar	Similar	Similar	-	
D479E	Reduz	Aumenta/Reduz	Aumenta	Aumenta	-	
Q486K	Reduz	Aumenta/Reduz	Similar	Similar	-	

Fonte: Dados de pesquisa, (2015* e 2016). O caractere – diz respeito a mutações classificadas como neutras ou inconclusivas segundo a análise de predição funcional.

4.4.2 Dinâmica Molecular longa

O pacote GROMACS é um software grátis que oferece técnicas computacionais avançadas para a realização de simulações de dinâmica molecular. A partir das diversas opções de cálculos, preparação e ferramentas de análises o usuário pode escolher configurações iniciais do sistema, descrever as interações atômicas e modelos físicos, realizar a simulação e fazer observações sobre a trajetória (Van Der Spoel, et al., 2005; Abraham, et al., 2015).

De modo semelhante à dinâmica curta, a estrutura experimental foi utilizada como selvagem e a partir dela as mutantes foram obtidas por mutagênese *in silico* fazendo-se uso do software VMD. Apenas as mutações P206S, R303W e R441H foram escolhidas para realização da dinâmica longa, uma vez que a maioria dos preditores classificaram-nas como danosas de acordo com a predição funcional realizada em estudo anterior (50%, 90% e 90%, respectivamente); e devido ao fato de estarem associadas à desordens de grande interesse social.

Após a realização das simulações, cada trajetória foi analisada quanto a estabilidade conformacional e flexibilidade das estruturas mutantes em relação à selvagem a fim de se comparar alterações no comportamento dinâmico da proteína. Para isso, os seguintes resultados foram obtidos e analisados: RMSD, Raio de Giro, Área de Superfície Acessível ao Solvente (SASA), Número de Ligações de Hidrogênio (NHB) intramoleculares e intermoleculares, RMSF, B-factor e a Análise de Componente Principal (PCA).

Novamente o RMSD foi usado para analisar o equilíbrio e estabilização da proteína durante a trajetória. De modo geral, os valores de RMSD das estruturas variaram entre o mínimo e o máximo de 0,08nm e 0,49nm, como mostrado na Figura 11. A proteína selvagem obteve valor de RMSD geral médio de 0,396nm. Todas as variantes mostraram valores de RMSD médio acima dos registrados para a WT. P206S (0,410nm) obteve a média mais próxima da observada para a proteína selvagem. Por outro lado, R303W e R441H obtiveram médias mais discrepantes, de 0,419nm e 0,430nm, respectivamente. Apesar da imagem gráfica sugerir que P206S apresentaria o maior RMSD médio entre as variantes, isso não ocorreu, pois seus valores apresentaram-se acima dos da WT apenas após 20ns, enquanto R303W e

R441H já haviam superado a selvagem logo no início da trajetória.

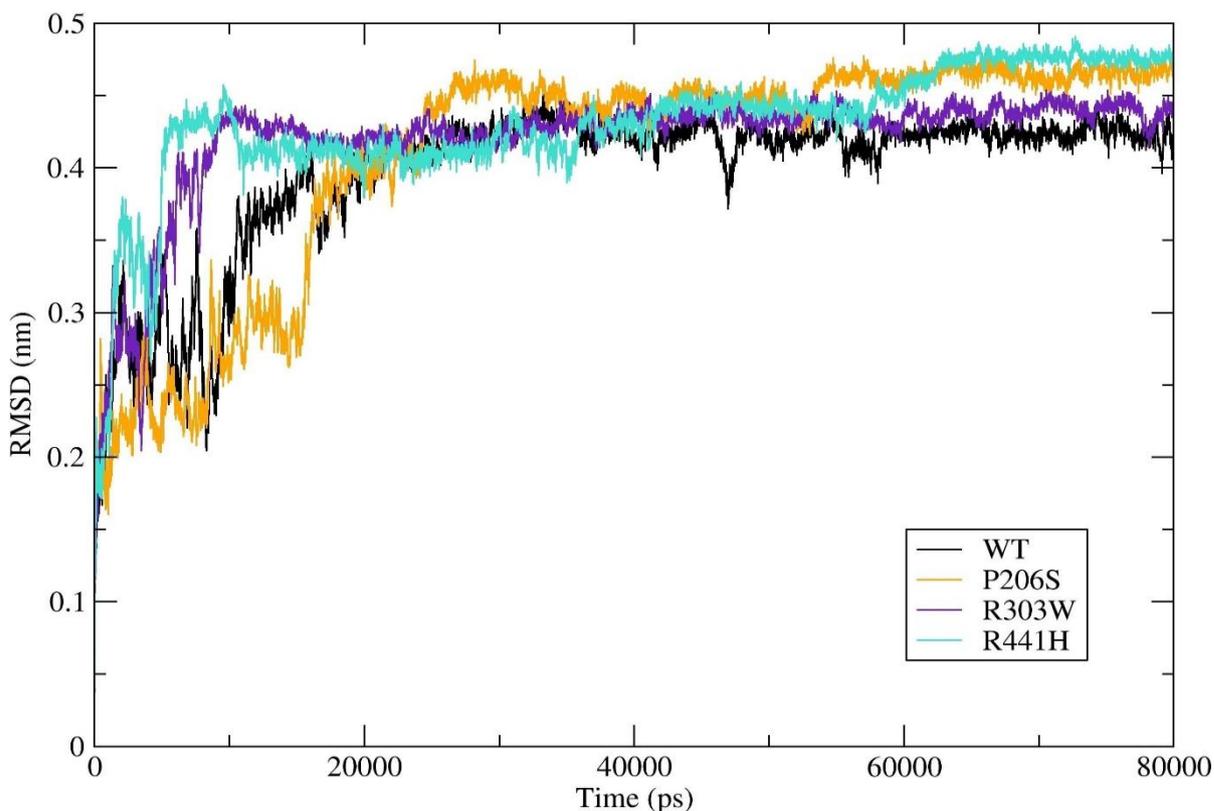


Figura 11 - Gráfico comparativo entre os valores de RMSD obtidos para a estrutura da proteína selvagem e as mutantes em função do tempo de 80 ns decorridos de simulação. Os valores em preto são referentes à estrutura selvagem; em laranja à P206S; em roxo à R303W; e em ciano à R441H.

Nestes casos foi possível observar o platô que marca o estado de equilíbrio do sistema, i.e., quando a estrutura passa a convergir sobre sua posição mais estável, mostrando assim valores de RMSD pouco variáveis. Para a WT, seus valores ao longo da trajetória mostraram se estabilizar logo após 60ns entre 0,400nm e 0,450nm, não apresentando mudanças relevantes até o final da simulação. Já ambas as estruturas com as variantes P206S e R303W atingiram tal equilíbrio antes da WT, poucos nanossegundos antes dos tempos 50ns e 20ns, respectivamente. Por outro lado, R441H mostrou convergir para sua posição mais estável em aproximadamente 62ns. É válido ressaltar neste ponto que para ser observado o platô remetente à estabilização da estrutura foram necessários pelo menos 20ns, o que significa que o observável na dinâmica curta representa apenas um momento ínfimo ainda da fase de equilibração e por esse motivo tais resultados foram considerados preliminares.

Nota-se então que a presença das mutações R303W e R441H afetou o comportamento estrutural da proteína durante a simulação, aumentando o valor médio do RMSD mais notavelmente do que a variante P206S, porém concluiu-se que todas elas mostraram alterar o equilíbrio e estabilidade da estrutura. Além disso, todas as mutantes mostraram alteração no tempo necessário para atingirem estabilidade conformacional no meio ao qual foram inseridas.

Mais uma vez a análise do Raio de Giro foi utilizada para obter-se dados sobre a compactação da estrutura proteica, e assim a dimensão geral do polipeptídeo. Nos resultados mostrados na Figura 12 foi observado que a proteína selvagem obteve um valor médio geral de Rg igual a 2,018nm.

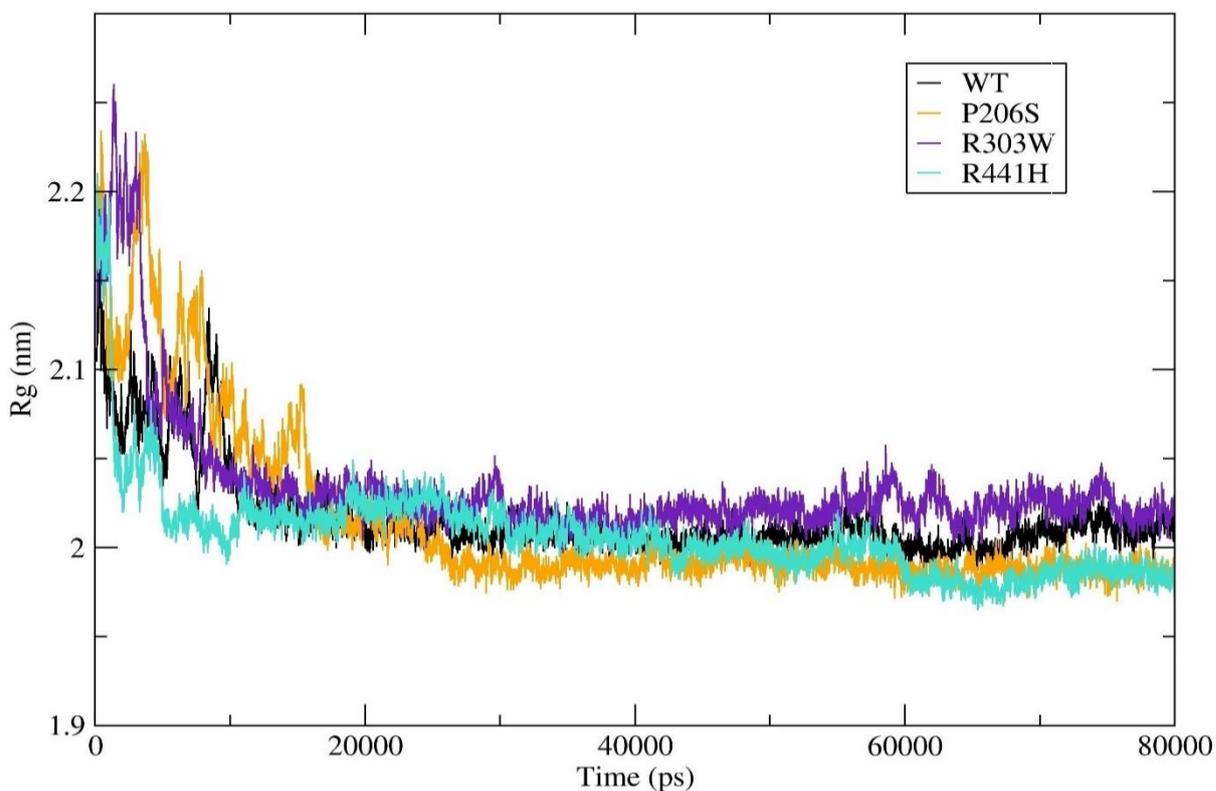


Figura 12 - Gráfico comparativo constando os valores obtidos na análise de Raio de Giro em nanômetros para as estruturas selvagem e as mutantes estudadas em função do tempo de 80 ns de simulação. Os valores em preto são referentes à estrutura selvagem; em laranja à P206S; em roxo à R303W; e em ciano à R441H.

Ambas mutantes P206S e R441H apresentaram Rg médio ligeiramente menor do que a proteínas selvagem (2,014nm e 2,007nm), sugerindo então que elas podem causar um leve aumento na compactação, ou seja, uma diminuição da dimensão geral da estrutura. O contrário foi observado para a mutante R303W, cujos valores se

mantiveram acima dos observados para a selvagem resultando no valor médio de 2,034nm, sugerindo aumento da dimensão geral. Todas variações mostraram causar alterações na compactação da estrutura, embora nenhuma delas tenha atingido valores variáveis o bastante que pudessem indicar desenovelamento ou desnaturação, como proposto.

Com o pacote GROMACS foi possível a realização de outros tipos de análises as quais permitiram inferir novas informações a respeito das estruturas da proteína. A análise de SASA permite estabelecer a área da superfície da proteína que está acessível ao solvente (Khan, et al., 2016; Joshi, et al., 2017). Durante a trajetória, a proteína selvagem obteve valores de SASA em torno de 171,704nm², enquanto que as mutantes P206S, R303W e R441H apresentaram valores médios de 167,722nm², 172,955nm² e 167,906nm², respectivamente. Tais resultados mostraram que R303W aumentou ligeiramente a área de superfície acessível ao solvente, enquanto as outras duas mutações a diminuíram com mais significância (Figura 13).

As interações dependentes do tempo também foram avaliadas pela quantificação do número de ligações de hidrogênio (NHB) estabelecidas tanto entre segmentos da própria proteína (intramoleculares), quanto entre o soluto e solvente (intermoleculares). Na análise das médias do NHB intramolecular (Figura 14) foi observado que tanto a proteína selvagem (281,103 números de ligações) quanto as variantes P206S (278,872 números de ligações) e R441H (275,903 números de ligações) obtiveram valores semelhantes dos números desse tipo de ligação. Mais uma vez a proteína com a variante R303W apresentou a maior diferença entre seu valor médio (274,148 número de ligações) e o observado para a proteína selvagem. Isso mostra que essa variante também afeta a quantidade de pontes de hidrogênio intramoleculares estabelecidas, apresentando em média 9 ligações a menos.

A análise do NHB intramolecular é importante uma vez que as ligações de hidrogênio que ocorrem na proteína são essenciais para a formação α -hélice e folhas- β . Por sua vez, a correta funcionalidade da proteína depende dessas estruturas secundárias (Ishak, et al., 2017). Assim, os resultados expostos sugerem que a R303W pode estar modificando a estrutura secundária em algum nível mais do que as outras variantes. Isso poderia gerar consequências na função da enzima.

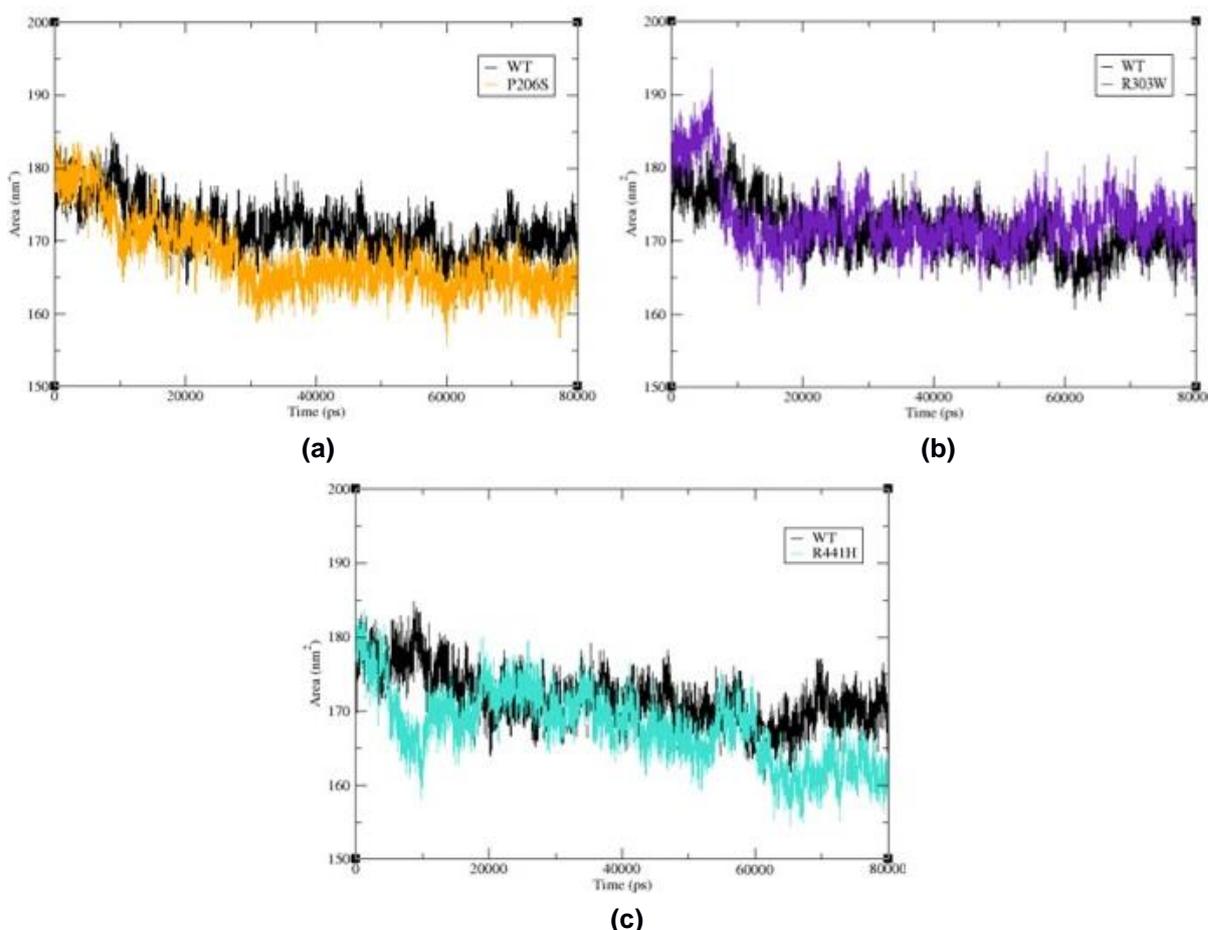


Figura 13 - Gráficos comparativos entre os valores obtidos a partir da análise da área de superfície acessível ao solvente (SASA) em nanômetros quadrados para as estruturas selvagem e mutantes durante os 80 ns de simulação. (a) Valores referentes às estruturas selvagem (preto) e a variante P206S (laranja). (b) Valores referentes às estruturas selvagem (preto) e a variante R303W (roxo). (c) Valores referentes às estruturas selvagem (preto) e a variante R441H (ciano).

Já os resultados referentes ao NHB intermolecular mostraram um padrão similar de valores entre a WT e as mutantes P206S e R441H (Figura 15), o que foi reforçado pelas médias gerais de 706,750, 709,608 e 706,684, respectivamente. Por outro lado, R303W apresentou NHB médio de 718,905 ligações. Mais uma vez a estrutura R303W mostrou poder causar as maiores mudanças dessas características, uma vez que estabeleceu em média 12 ligações intermoleculares a mais que a WT e maior diferença entre o balanço de interações intra-intermoleculares.

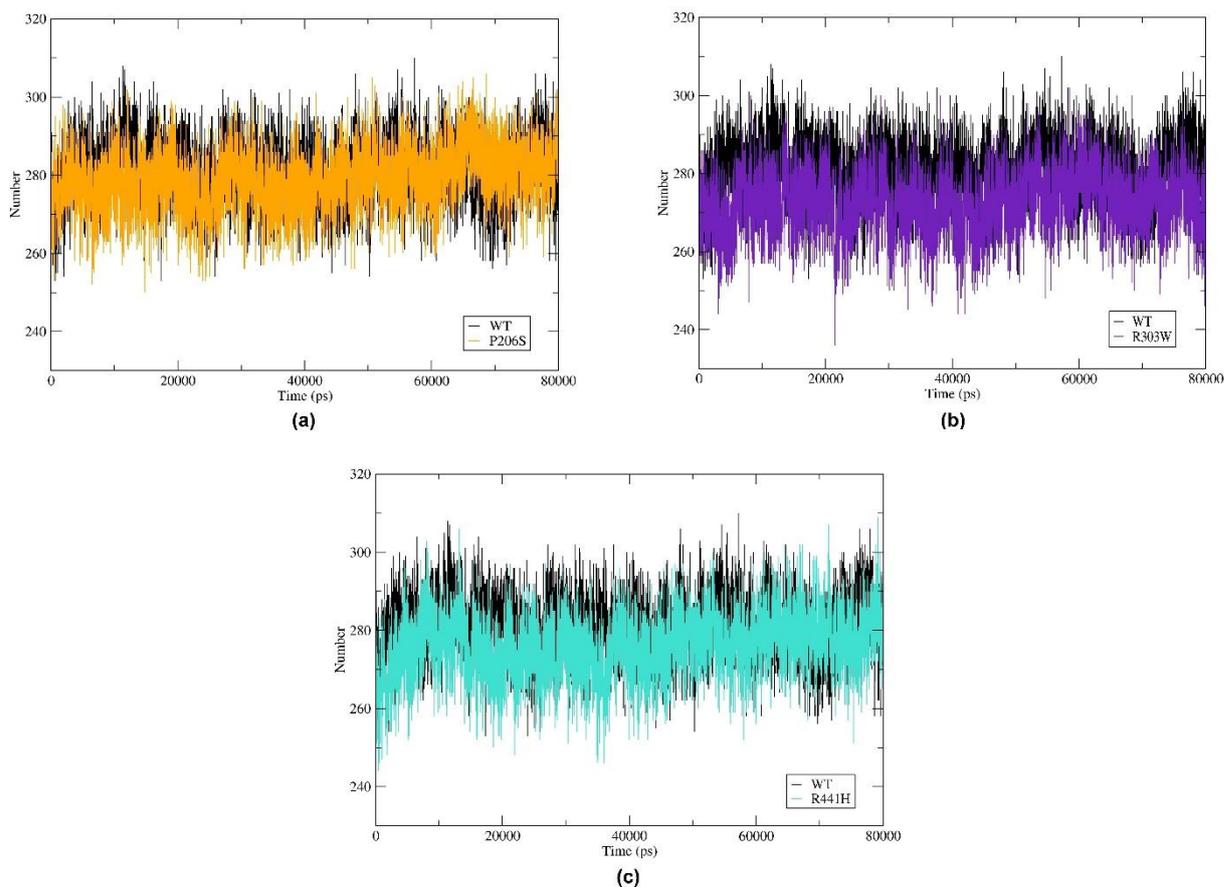


Figura 14 - Gráficos constando os números de ligações de hidrogênio intramoleculares estabelecidas em cada estrutura proteica analisada durante a trajetória. Valores referentes às estruturas selvagem e todas as mutantes analisadas. (a) Valores referentes às estruturas selvagem (preto) e a variante P206S (laranja). (b) Valores referentes às estruturas selvagem (preto) e a variante R303W (roxo). (c) Valores referentes às estruturas selvagem (preto) e a variante R441H (ciano).

A relação entre o solvente e sua influência sobre o soluto é fundamental para o comportamento da proteína, uma vez que sua resistência a se desenovelar e a realização de suas relações químicas dependem disso (De Marco, et al., 2014). Além disso, a estabilidade e enovelamento da estrutura proteica dependem do balanço entre interações intramoleculares e intermoleculares. A mutante R303W mostrou alterar ambas características com maior significância, logo podendo causar desequilíbrio entre essas interações e conseqüentemente alterações na estrutura da proteína. P206S também mostrou alterar tais interações de modo semelhante, ligeiramente diminuindo o número de ligações de hidrogênio intramoleculares e aumentando o de intermoleculares, porém não com tal relevância.

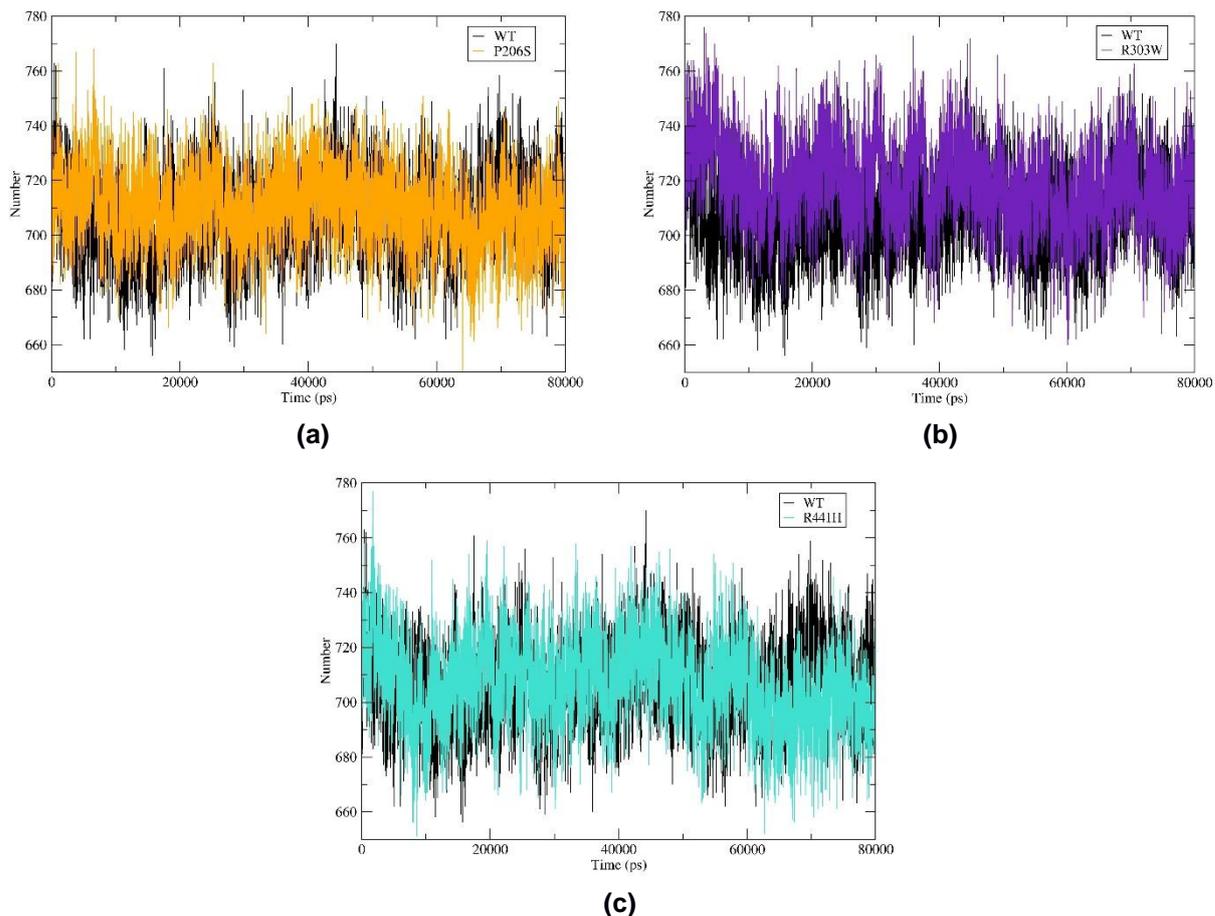


Figura 15 - Gráfico constando o número de ligações de hidrogênio intermoleculares estabelecidas entre as estruturas proteicas analisadas e a água durante a trajetória. (a) Valores referentes às estruturas selvagem (preto) e a variante P206S (laranja). (b) Valores referentes às estruturas selvagem (preto) e a variante R303W (roxo). (c) Valores referentes às estruturas selvagem (preto) e a variante R441H (ciano).

Já R441H apesar de ter apresentado a segundo menor diminuição do NHB intramolecular, seu valor médio geral para a análise das interações intermoleculares foi muito próximo do observado para a WT, não mostrando afetar tanto o balanço dessas interações como as outras variantes.

Em suma, os resultados das análises de Rg, SASA e NHB mostram que as diferenças mais relevantes no estabelecimento das ligações de hidrogênio competem a estrutura que apresentou perda de compactação e ganho de superfície acessível ao solvente. Ou seja, supõem-se que a presença da mutação R303W altere o estabelecimento das ligações de hidrogênio, levando à diminuição daquelas intramoleculares e conseqüentemente ao ganho de dimensão geral e área de superfície acessível ao solvente. Tais mudança por sua vez possibilitariam o aumento das interações intermoleculares.

Essa suposição se baseia no princípio de que as interações que ocorrem entre os segmentos da estrutura colaboram consideravelmente para sua resistência ao desenovelamento (Khan, et al., 2016). No entanto, as mesmas correlações não foram notadas para as outras duas mutantes, possivelmente porque a diferença entre os valores do número de ligações de hidrogênio intramoleculares foi pequena em relação à WT, não tão relevantes quanto a observada para R303W.

A análise dos valores de RMSF foram utilizados para descrever a flexibilidade local dos resíduos componentes das estruturas durante a trajetória, assim como a análise de RMSD por resíduo. No entanto, neste caso a medida da distância média é calculada entre todos os átomos dos resíduos de aminoácidos da estrutura de partida e cada estrutura subsequente durante a simulação, e não apenas dos átomos de carbono alfa. Ao final, o gráfico gerado contém o valor médio de RMSF observado para cada resíduo. Quanto maior esses valores, maior a flexibilidade naquela posição.

A partir das médias gerais dos valores RMSF para cada estrutura, foi observado que a mutante R303W possui estrutura com uma maior flexibilidade geral (0,190nm), seguindo pela mutação P206S (0,171nm) e R441H (0,157nm) quando comparadas à WT (0,150nm) (Figura 16). Localmente, foi observado algumas diferenças relevantes: apenas em duas posições as mutantes apresentaram perda de flexibilidade considerável, sendo a primeira na posição 167, onde na estrutura selvagem apresenta RMSF igual a 0,281nm, enquanto nas mutantes apresenta RMSF menores de 0,134nm (P206S), 0,199nm (R303W) e 0,097nm (R441H); e a segunda na posição 425, onde a WT apresenta RMSF igual a 0,266nm, enquanto as mutantes 0,109nm (P206S), 0,128nm (R303W) e 0,153 (R441H).

Analisando-se ganho local de flexibilidade, P206S mostrou um ligeiro aumento acima de 0,26nm nos resíduos 200, 202 e 205, enquanto a WT apresentou valores de 0,103nm, 0,174nm e 0,129nm nas respectivas posições. Não coincidentemente, tais posições se encontram próximas ao local da mutação, e podem ter ocorrido devido as mudanças das propriedade físico-químicas causadas pela substituição de um aminoácido médio e hidrofóbico (prolina) por um pequeno e polar (serina). Quando observado o valor obtido na posição onde ocorreu a substituição, notou-se que S206 (0,211nm) era mais flexível do que P206 (0,114nm).

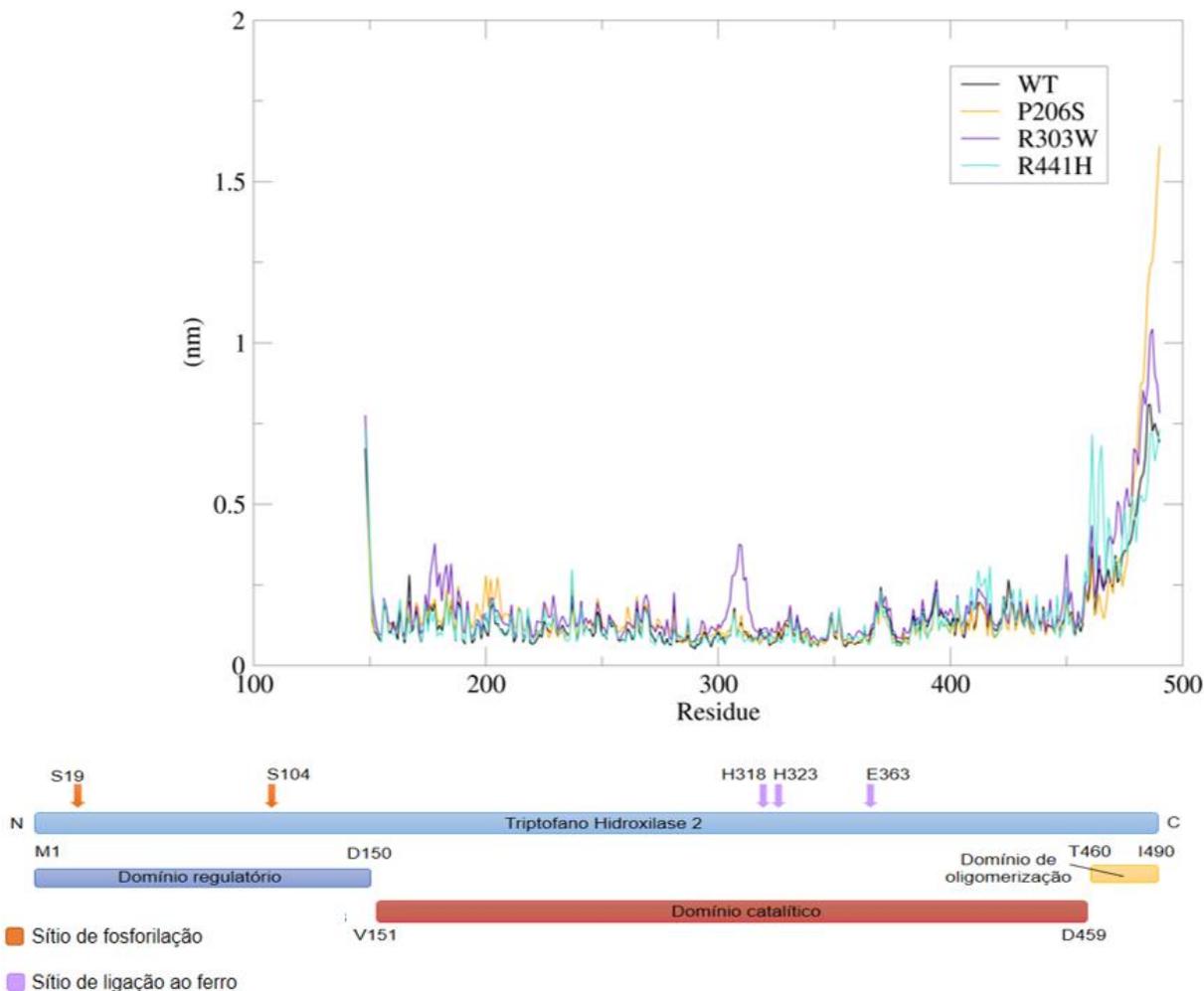


Figura 16 – Gráfico constando os valores de RMSF médio em nanômetros para cada resíduo das estruturas selvagem e mutantes, obtido durante o período de 80 ns da dinâmica molecular. Os valores em preto são referentes à estrutura selvagem; em laranja à P206S; em roxo à R303W; e em ciano à R441H. Abaixo, esquema da estrutura dos domínios da TPH2 e seus sítios de fosforilação e de ligação ao ferro, para fins comparativos.

R303W mostrou notável ganho de flexibilidade acima de 0,279nm nos resíduos 177, 178, 180, 182, 183 e 185, os quais na WT não excederam o valor de 0,233nm. Outra região de ganho de flexibilidade nesta mesma mutante foi entre os resíduos 306 e 312 ($0,264\text{nm} \leq \text{RMSF} \leq 0,377\text{nm}$), locais onde na WT foi observado o valor máximo de 0,179nm na posição 307. Esse aumento pode estar associado à presença da mutação na posição 303, uma vez que as interações nessa região podem ter sido alteradas devido à troca de uma arginina para um triptofano, o que leva à mudança das propriedades físico-químicas, que nesse caso é de um aminoácido grande, básico e hidrofílico para um grande, aromático e hidrofóbico. Inclusive, no local da substituição também foi notado uma maior flexibilidade para W303 (0,143nm) do que

para R303 (0,064nm). Uma última posição de ganho notável para R303W foi no aminoácido 450, onde uma diferença de mais de 10nm pode ser observada entre esta variante (0,346nm) e a WT (0,240nm).

Já a estrutura com a substituição R441H apresentou ganho acima de 0,244nm de flexibilidade nos resíduos 237, 411-415 e 417, enquanto que essas posições na WT não apresentaram valores maiores do que 0,196nm. Na posição da variação, mais uma vez o resíduo mutante (0,170nm) apresentou maior flexibilidade do que o resíduo selvagem (0,104nm). Neste caso, a falta de associação à perda ou ganho de resíduos próximo ao mutante pode estar relacionada ao fato de que tal mutação é a menos diferenciada quando comparada às outras, na qual um aminoácido grande, básico e polar (arginina) é trocado por um médio, básico e polar (histidina). Também foi observado o ganho de flexibilidade dos resíduos subsequentes às mutações em todos os casos, porém não de forma relevante.

Quanto as extremidades, foi observado que para todas as estruturas variantes houve um ganho de flexibilidade no terminal COO-. A média dos valores de RMSF para os últimos 30 resíduos (domínio de oligomerização) foram obtidas, mostrando que de fato P206S (0,576nm) mostrou maior ganho do que a R303W (0,557nm), e a WT e R441H que obtiveram a mesma média de 0,486nm. Porém, o intrigante maior aumento nos últimos 5 resíduos de P206S levou à análise adicional para ser explicada.

Partindo do princípio de que durante o período de equilibração os valores analisados tendem a ser mais variáveis, a análise do RMSF também foi realizada para os sistemas apenas no intervalo de tempo os quais todos se encontravam em equilíbrio, ou seja, de 62ns até 80ns. Como mostrado na Figura 17, esses resultados revelaram que o ganho exagerado de flexibilidade de fato ocorreu antes do sistema atingir seu equilíbrio. Neste caso, a média geral do RMSF reduziu para todas as estruturas, mas R303W (0,115nm) permaneceu mostrando maior valor que a WT (0,098nm). Por outro lado, P206S e R441H passaram a obter menor média em relação a WT, de 0,089nm e 0,093nm, respectivamente. Novamente o RMSF apenas para o domínio de oligomerização foi calculado. Assim, foi observado que P206S apresenta menor flexibilidade do que a WT no C-terminal quando em seu estado equilibrado (0,111nm e 0,174nm, respectivamente), enquanto que R303W (0,215nm) continuou

mais flexível e R441H (0,118nm) similar a estrutura nativa.

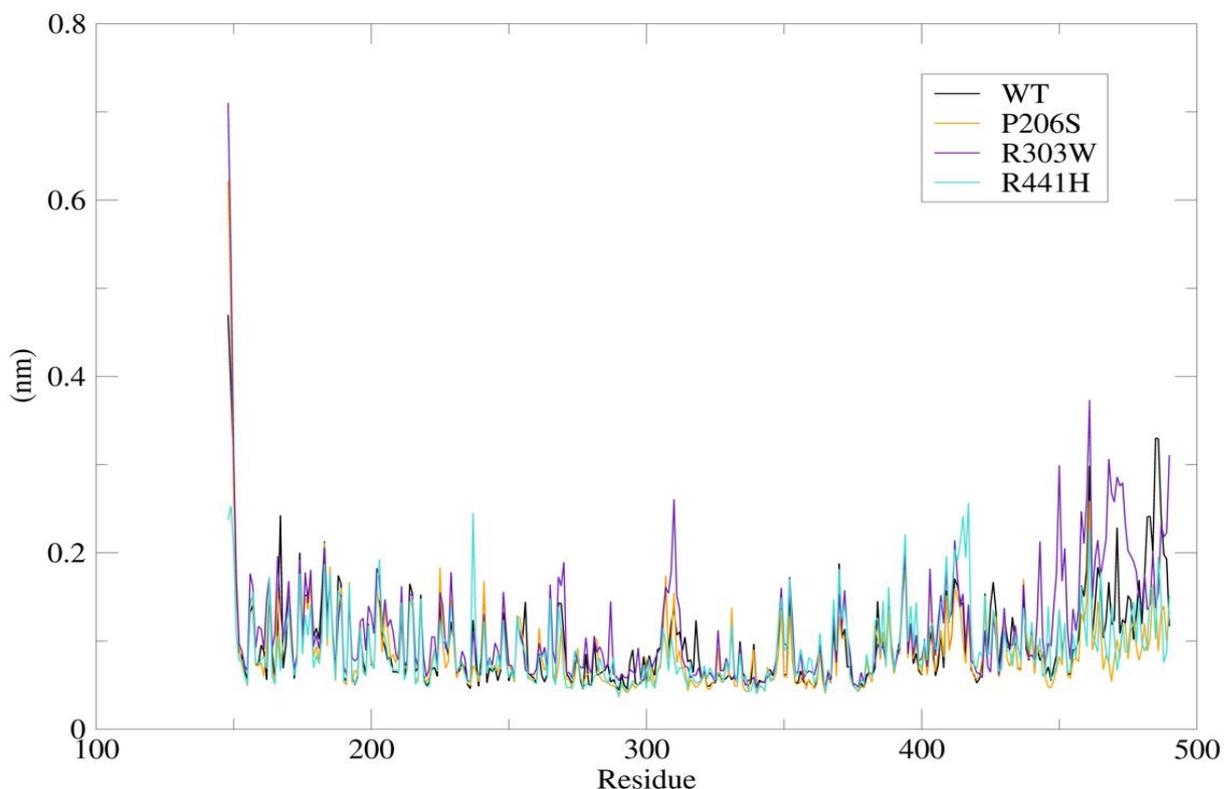


Figura 17 - Gráficos constando os valores de RMSF em nanômetros para cada resíduo das estruturas selvagem e mutantes, durante o período de 62 ns a 80 ns da dinâmica molecular. Os valores em preto são referentes à estrutura selvagem; em laranja à P206S; em roxo à R303W; e em ciano à R441H.

Embora similarmente todas as estruturas apresentaram uma alta flexibilidade no terminal NH₃⁺, os primeiros 10 resíduos não foram considerados nessa análise. Isso deve-se ao fato de que na célula a TPH2 também é composta por um domínio regulatório, o qual não foi elucidado no modelo experimental utilizado. A falta deste domínio poderia estar causando essa alta flexibilidade, observada inclusive para a WT, e afetando o entendimento da análise.

Os resultados do B-factor, diferentemente do resultado gráfico obtido pelo MDWeb, foram obtidos por meio de imagens (Figura 18) nas quais as estruturas são coloridas e espessadas segundo a mobilidade dos resíduos. Sendo assim, aqueles aminoácidos mais rígidos são representados pela cor azul escura, e gradativamente uma escala de cores das mais frias para as mais quentes segue mostrando o aumento da flexibilidade até o extremo oposto que destaca alta flexibilidade na cor vermelha.

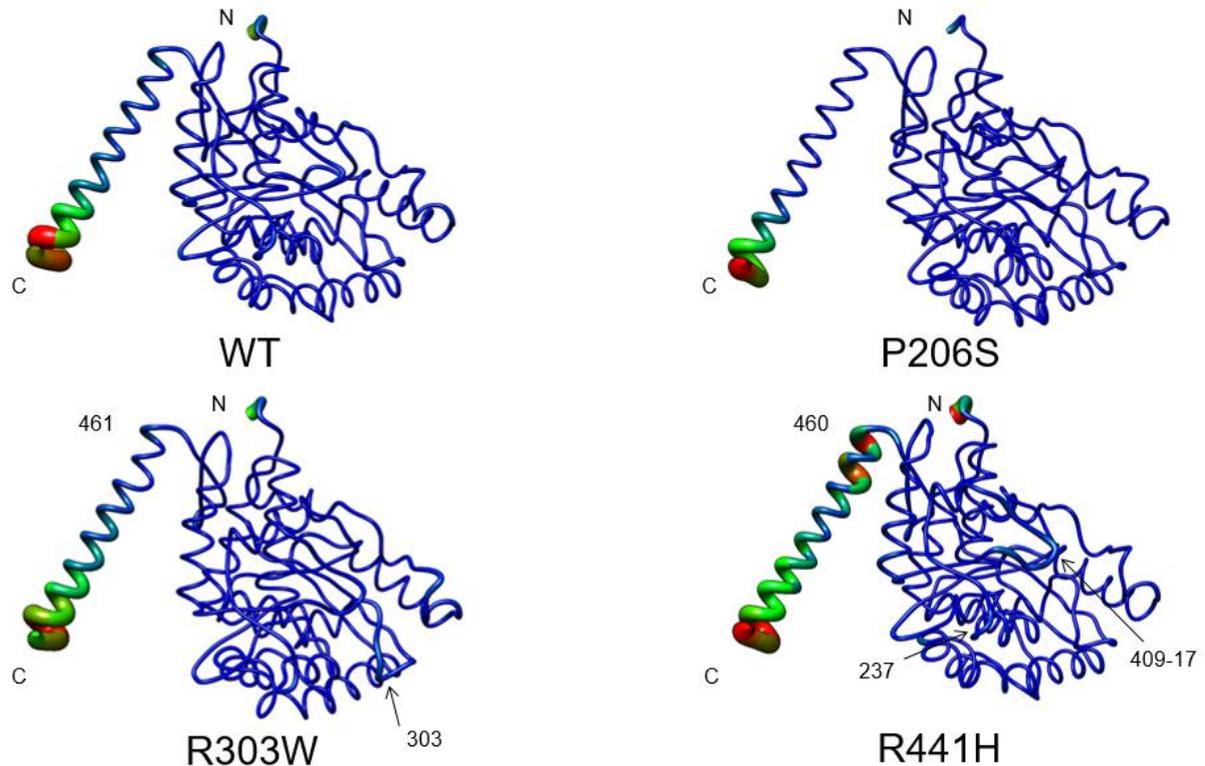


Figura 18 - Figuras obtidas a partir do resultado da análise de B-factor para cada estrutura estudada. (a) Estrutura selvagem. (b) Estrutura variante P206S (c) Estrutura variante R303W (d) Estrutura variante R441H. A cor azul escuro e filamento fino representam maior rigidez nos resíduos, enquanto em extremo oposto a cor vermelha e filamento espesso representam maior flexibilidade. O gradiente de cor e espessuras entre ambos extremos representam diversos níveis intermediário de flexibilidade. Os resíduos destacados nas imagens referentes as mutantes R303W e R441H mostram resíduos que mostraram ganho de flexibilidade tanto nesta análise quanto na realizada por RMSF. Imagem obtidas com o software Chimera.

Corroborando com o observado no RMSF, as extremidades NH₃⁺ e COO⁻ mostraram ser mais flexíveis do que o resto da estrutura em todos os casos. Alguns resíduos das estruturas mutantes se destacaram no ganho de flexibilidade quando comparados aos da WT: os resíduos 461 e 303 da estrutura R303W, e os 237 e 409-417 da estrutura R441H, assim corroborando o observado no RMSF. Essa análise atentou para um notável aumento na flexibilidade no início do domínio de oligomerização para a estrutura com a mutação R441H, não tão evidente segundo as análises do RMSF. De fato, segundo gráfico da Figura 16, tal porção se mostrou ser mais flexível do que as outras estruturas a partir da posição 460 até assemelhar-se aos valores obtidos para WT, o que resultou em médias parecidas para ambas as estruturas como observado. Também mostrou que, apesar P206S tenha obtido a maior média para a extremidade C-terminal, nessa região a variante possui padrão similar a WT, exceto mesmo nos resíduos finais que se apresentaram mais flexíveis e

colaboraram para o elevado RMSF médio observado. Para R303W evidenciou uma maior quantidade de resíduos flexíveis no terminal C, em acordo com o RMSF.

4.4.2.1 Análise de Componente Principal (PCA)

Outra análise adicional realizada pelo GROMACS foi a de Análise de Componente Principal (PCA). Esta é uma técnica estatística utilizada para descrição da dinâmica conformacional da proteína, descrevendo-a por meio da significância dos movimentos coletivos da estrutura que são convertidos em movimentos principais (Chandrasekaran; Rajasekaran, 2017). Ou seja, o número de dimensões necessárias para descrever a dinâmica da conformação é reduzido por decomposição de seus movimentos de maior escala espacial até os de menor (Ndagi; Mhlongo; Soliman, 2017).

Para isso, primeiramente uma matriz de covariância das posições atômicas durante a trajetória é obtida, a partir das coordenadas cartesianas desses átomos. Pela diagonalização da matriz, autovetores e autovalores são obtidos. O primeiro representa um único deslocamento correlacionado de um grupo de átomos num espaço multidimensional. Já o segundo são a amplitude da movimentação ao longo dos autovetores. Para cada autovetor há um autovalor correspondente (Chandrasekaran; Rajasekaran, 2017).

A análise desses autovetores também é realizada pelo GROMCAS, e para isso eles são organizados de acordo com seus autovalores em ordem decrescente. Então, os primeiros autovetores (que são aqueles que possuem os maiores autovalores) são utilizados, pois definem satisfatoriamente um subespaço conformacional essencial dentro do qual ocorrem os movimentos de grande escala mais significativos da proteína (Liu et al, 2010). Isto é suficiente para descrição da movimentação geral da mesma (Narang, et al., 2017). Assim, os primeiros autovalores descrevem os componentes principais da movimentação e podem ser interpretados como a direção que representa a variação máxima das amostras (Chandrasekaran; Rajasekaran, 2017).

Desde modo foi possível avaliar a dinâmica conformacional geral das estruturas durante a dinâmica, fazendo-se uso dos dois primeiros componentes principais. A

Figura 19 mostra os clusters de conformações terciárias ao longo dos planos PC1 e PC2. Foi possível observar que a estrutura WT apresentou ampla área de dispersão ao longo dos componentes, mostrando considerável variação conformacional e de flexibilidade.

A estrutura com a mutação P206S apresentou área de dispersão semelhante à WT, porém ligeiramente maior ao longo do PC1, devido a maior flexibilidade de seus resíduos, o que pode ser observado de modo geral na Figura 19b; e semelhante ao longo do PC2 (Figura 19c). Tais resultados possibilitam inferir que esta estrutura variante se mostrou tão variável e flexível quanto a WT, porém com comportamento diferenciado.

Notavelmente a variante R303W mostrou maior área de dispersão em relação à WT, principalmente ao longo do PC1. Tal observação ocorreu devido à maior flexibilidade dos seus resíduos ao longo dos PC1 e PC2 (Figuras 19b e c). Sendo assim, tais resultados sugerem que esta estrutura seja mais variável e flexível do que a WT, mostrando comportamento atípico da proteína.

Já R441H apresentou valores menos dispersos, porém variando em extensão semelhante à WT ao longo dos componentes principais. Tal comportamento pode ser relacionado aos baixos valores de RMSF obtidos ao longo do PC1, mas similares ao longo do PC2 (Figuras 19b e c). Assim, pode-se sugerir que R441H mostrou comportamento mais similar a WT quanto sua variabilidade conformacional, porém todas as mutações mostram afetar tal aspecto.

A análise dos componentes principais mostrou estar em concordância com o observado para os resultados do RMSF e B-factor, assim como corroboram com as outras características avaliadas mostrando que tais proteína afetam a dinâmica molecular da proteína.

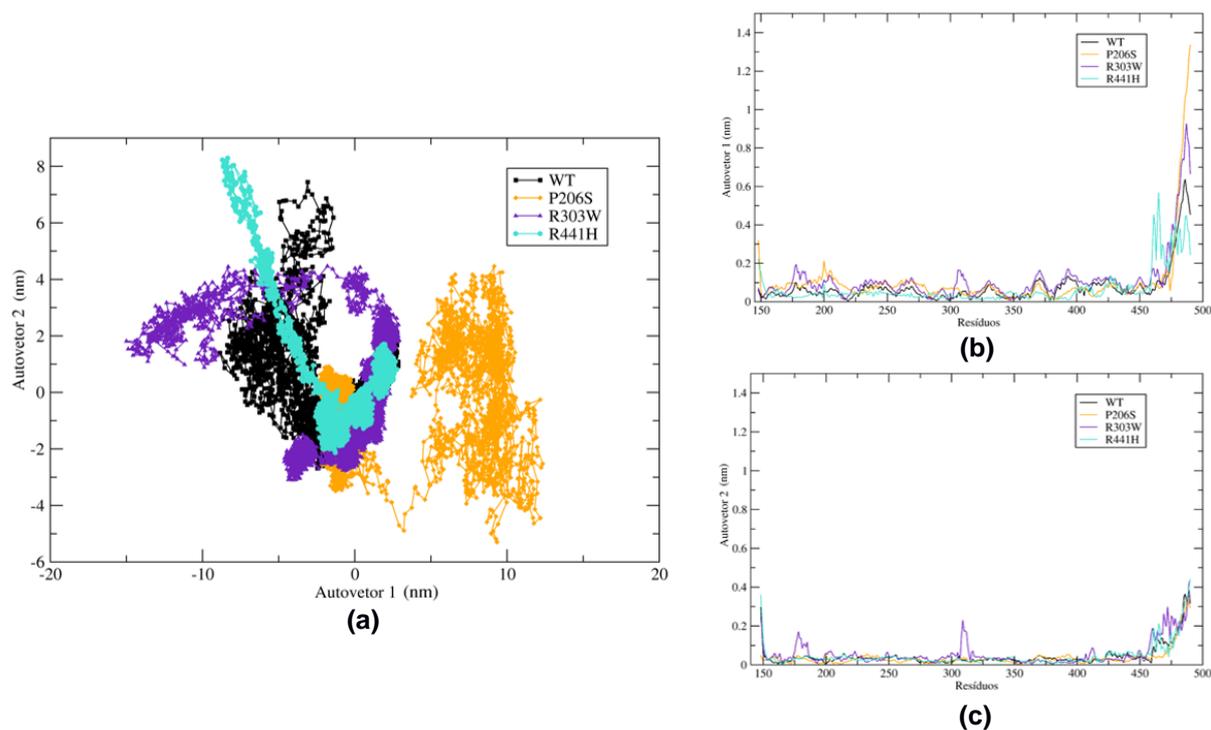


Figura 19 – Análise de Componentes Principais das trajetórias obtidas na simulação de dinâmica molecular das estruturas selvagem e as variantes estudadas. (a) Gráfico de dispersão comparativo entre os autovalores obtidos para os componentes principais 1 e 2 das estruturas. (b) Gráfico dos valores de RMSF obtidos pelas estruturas ao longo do autovetor 1. (c) Gráfico dos valores de RMSF obtidos pelas estruturas ao longo do autovetor 2.

Do exposto com os resultados do RMSF, B-factor e PCA, todas as mutações mostraram poder alterar em algum nível a flexibilidade da proteína, especialmente R303W. De fato, substituições de aminoácidos podem causar mudanças nas interações entre resíduos e arredores que afetam a entropia do sistema e consequentemente a flexibilidade da estrutura (Khan, 2016). A flexibilidade reflete na dinâmica proteica e em sua função, permitindo que as interações necessárias ocorram (Craveur, et al., 2015). Rigidez e flexibilidade necessitam estar balanceadas para desempenho correto da função (Khan, 2016). No caso da família AAAH, o mecanismo de ativação dessas enzimas depende da redução do ferro pela tetrahydrobiopterina (Walther; Bader, 2003). Para que isso ocorra, TPH2 possui três sítios de ligação ao ferro nas posições 318, 323 e 363 (UniProtKB ID Q8IWU9) (RSCB PDB 4V06). Apesar de R303W ocorrer a 14,4Å de distância do sítio mais próximo (McKinney, et al. 2009), essa variante ainda assim poderia alterar interações ao redor de si e mais próximas aos sítios e consequentemente a ativação da enzima. Os resultados do RMSF mostram como W303 afeta os resíduos que a cercam.

Além do mais, todas as mutações poderiam afetar a interação que permite a

formação de homotetrâmeros da enzima TPH2. O domínio responsável pela oligomerização consiste em uma α -hélice a qual interagem entre os monômeros do mesmo modo como o proposto para o clássico motivo zíper de leucina. Tal domínio compreende os aminoácidos 460 ao 490 (Fitzpatrick, 2003; Carkaci-Salli, et al., 2006). As análises mostraram como as variações podem aumentar a flexibilidade desta extremidade. No mais, outras interações além daquelas necessárias para ativação e formação de tetrâmeros poderiam ser afetadas, como aquelas que ocorrem entre proteína-proteína. Embora o UniProt forneça informações sobre interações entre a TPH2 e a SYBU e DJC12, outros estudos são necessários para se entender o papel de outras proteínas na função da TPH2.

A conformação precisa de uma proteína depende da sua estrutura primária, que é descrita por uma sequência de aminoácidos específica, uma vez que o processo de enovelamento depende das interações estabelecidas pelas cadeias laterais de tais aminoácidos. A atividade e regulação da proteína são influenciadas pelo empacotamento, uma vez que neste estado a molécula se encontra mais estável (Khan, et al., 2016). Todas as análises realizadas quanto o equilíbrio, variações conformacionais, interações e flexibilidade mostram possíveis mudanças da conformação estrutural das variantes em relação ao tipo selvagem. Desse modo, as mutações pontuais não sinônimas analisadas podem estar relacionadas às diferenças notadas na estabilidade estrutural das mutantes, especialmente a R303W que mostrou alterações mais relevantes. Uma possível explicação pode ser o fato de que as substituições analisadas mostram possuir propriedades físico-químicas distintas dos aminoácidos selvagens em diferentes graus. No entanto, análises mais específicas seriam necessárias para mensurar o que tais mudanças conformacionais podem causar na funcionalidade da enzima.

Estudos prévios realizados por McKinney et al. (2009) reportaram que todas essas variantes poderiam diminuir a estabilidade conformacional, atividade e solubilidade da proteína. R303W destacou-se como a variante que apresentou maior diminuição dessa característica, mostrando também desativar a proteína (atividade nula). Nossos dados mostraram estar de acordo com os achados prévios, exceto para a análise de solubilidade. Como mostrado, os valores mais altos de RMSD apresentados pelas estruturas mutantes quando comparadas à WT sugerem que elas

diferem mais na estrutura inicial, a qual se encontra em sua conformação mais estável, pois tem sua energia livre minimizada antes do início da simulação. As modificações dos valores de Rg e NHB corroboram para reforçar mudanças estruturais, indicando que tais mutantes afetam a estabilidade conformacional. A respeito da solubilidade de R303W, que no presente estudo mostrou-se ser maior do que a WT divergindo dos achados *in vivo*, essa diferença pode estar relacionada ao fato de que os sistemas usados por McKinney et al. consistiam na análise da proteína inteira, ou seja, inclusive o seu domínio de regulação. Este domínio tem influência notavelmente grande na solubilidade da proteína, no sentido de limitar tal propriedade (Carkaci-Salli, et al., 2006). Assim, uma comparação equivalente entre os achados não seria possível neste caso. No entanto, é provável que o aumento observado neste estudo desta característica em R303W pode estar relacionado à falta do domínio NH2 terminal no modelo usado.

Winge et al. (2007) achou resultados similares aos de McKinney, que reportaram como R441H poderia diminuir a atividade, solubilidade e estabilidade da proteína. As simulações de DM, de fato R441H mostraram levar a diminuição de muitos parâmetros em relação à WT, como discutido sobre a interpretação do RMSD, Rg, SASA, NHB intramolecular e intermolecular em menor nível. Logo, o presente trabalho reforça o observado previamente.

Estrutura refinada pela DM

A simulação de DM também pode ser aplicada para refinar modelos experimentais derivados de cristalografia de raio-X e NMR (Martínez, 2015), uma vez que durante a preparação do sistema dois passos de minimização do potencial de energia foram realizados, inclusive em água, o que permitiu a obtenção de uma estrutura proteica mais estável e próxima à realidade biológica (Benkert; Biasini; Schwede, 2011).

Para isso, o *output* necessário da simulação referente à estrutura selvagem foi selecionado (*average.pdb*) (Figura 20) e validado a fim de compará-lo ao modelo 4V06 original. Como mostrado na Tabela 3, o validador ERRAT não retornou resultados por causa de um erro. Para os outros algoritmos a estrutura refinada pela DM, nomeada 4V06 DM, mostrou valores melhores do que a estrutura experimental original. Houve

um aumento na porcentagem do Verify 3D de 84,26% para 90,80%. Os valores dos gráficos de Ramachandran do PROCHECK e Rampage também mostraram melhorias na porcentagem de combinação nas áreas mais favoráveis. O QMEAN4 Z-score retornou valor absoluto de 1,17 para 4V06, o que significa que a qualidade da estrutura original desvia mais do que 1 desvio padrão do esperado. Já a estrutura refinada obteve pontuação absoluta de 0,80. Esse valor menor que um e maior que 0 a classifica como de boa qualidade, e melhor do que 4V06. Para o ProSa ambos os modelos estão dentro da extensão considerada aceitável para estruturas obtidas por cristalografia de raio-X. Assim, o modelo refinado provou ter qualidade maior do que seu *input*. Tal modelo, assim como os resultados da análise funcional estarão disponíveis no bando de dados online do nosso grupo.

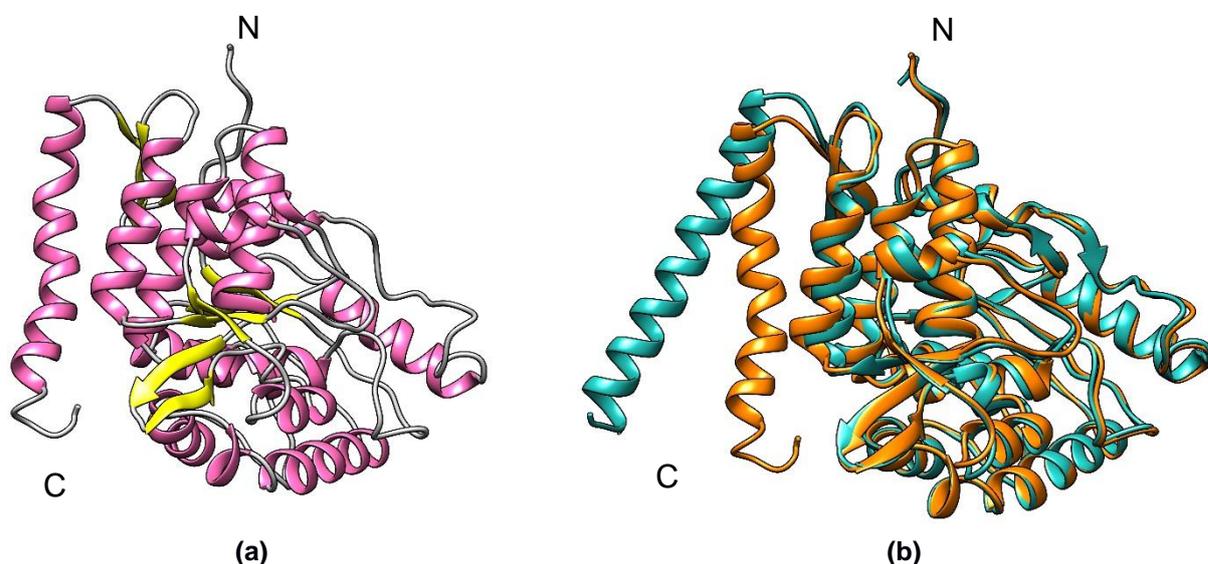


Figura 20 – Modelo estrutural do domínio catalítico da TPH2, visualizado pelo Chimera. (a) Modelo refinado obtido após a simulação de dinâmica molecular da proteína selvagem a partir do modelo estrutural 4V06, colorido de acordo com sua estrutura secundária. As alfa-hélices se encontram na cor rosa; as folhas-beta na cor amarela; voltas e alças estão coloridas de cinza. (b) Alinhamento estrutural entre o modelo 4V06 original (em verde) e o modelo refinado pela DM (em laranja).

A partir do average.pdb também foi possível ter uma primeira visão das modificações de estrutura secundária potencialmente causada pelas mutações. Para isso, este *output* foi obtido também para as mutantes (Figura 21) e comparado ao do modelo refinado. Todas as mutantes apresentarem valores possivelmente associáveis à perda de estrutura secundária segundo as análises de suas trajetórias (maior RMSD e RMSF, e menor NHB intramoleculares), notavelmente foi observado que a α -hélice de oligomerização modificou-se para alças em determinados pontos nas estruturas P206S e R303W.

Outras pequenas alterações também puderam ser observadas no alinhamento das sequências realizado pelo Chimera (Figura 22) que mostra destacado em amarelo os resíduos que fazem parte de alfa hélices e em verde de folhas beta. McKinney et al. (2009) havia proposto que a mutação P206S causaria uma perda de estrutura secundária, uma vez a substituição da prolina levaria à perda de restrições que tal aminoácido impõe à estrutura. Neste trabalho, o alinhamento da sequência desses *outputs* permitiu a conclusão de que tal fato não ocorre, não só no local da substituição P206S, como para R303W e R441H. Uma possível explicação para isso pode estar no fato de que todos os resíduos nativos na estrutura selvagem se encontram em voltas/alças, estruturas essas que mostram acomodar mutações mais facilmente no decorrer da evolução, mantendo a estrutura local intacta (Verli, 2014c). No entanto, uma análise mais precisa usando-se o GROMACS deve ser feita para contabilizar o número de resíduos por estrutura secundária em função do tempo para medir essas modificações com maior acurácia.

Considerando os resultados de todos os parâmetros analisados, pode-se concluir que as três mutações afetam a dinâmica estrutural da proteína. Notavelmente, a estrutura carregando a substituição R303W apresentou o maior número de mudanças relevantes, enquanto P206S mostrou aumentar o RMSD, compactação, NHB intra e intermoleculares e flexibilidade geral, mas diminuir a SASA. Já R441H mostrou diminuir a maioria das características analisadas, mas aumentar o RMSD e a média geral de RMSF. Todas estruturas mostraram padrão de flexibilidade distintos da WT como também mostrado pelas análises de B-factor e PCA. Assim, sugere-se que tais alterações gerais e locais podem interferir nas interações da proteína e conseqüentemente no seu funcionamento. Desse modo, as associações previamente realizadas entre os transtornos psiquiátricos e tais mutações podem ser embasadas no fato de que as variantes alteram o comportamento estrutural da proteína, e conseqüentemente podem alterar os níveis de serotonina devido à disfunção da enzima mutada, causando o desequilíbrio das vias reguladas por esse neurotransmissor. Porém outros estudos se tornam necessário para avaliar as implicações dessas alterações mais precisamente no funcionamento da TPH2.

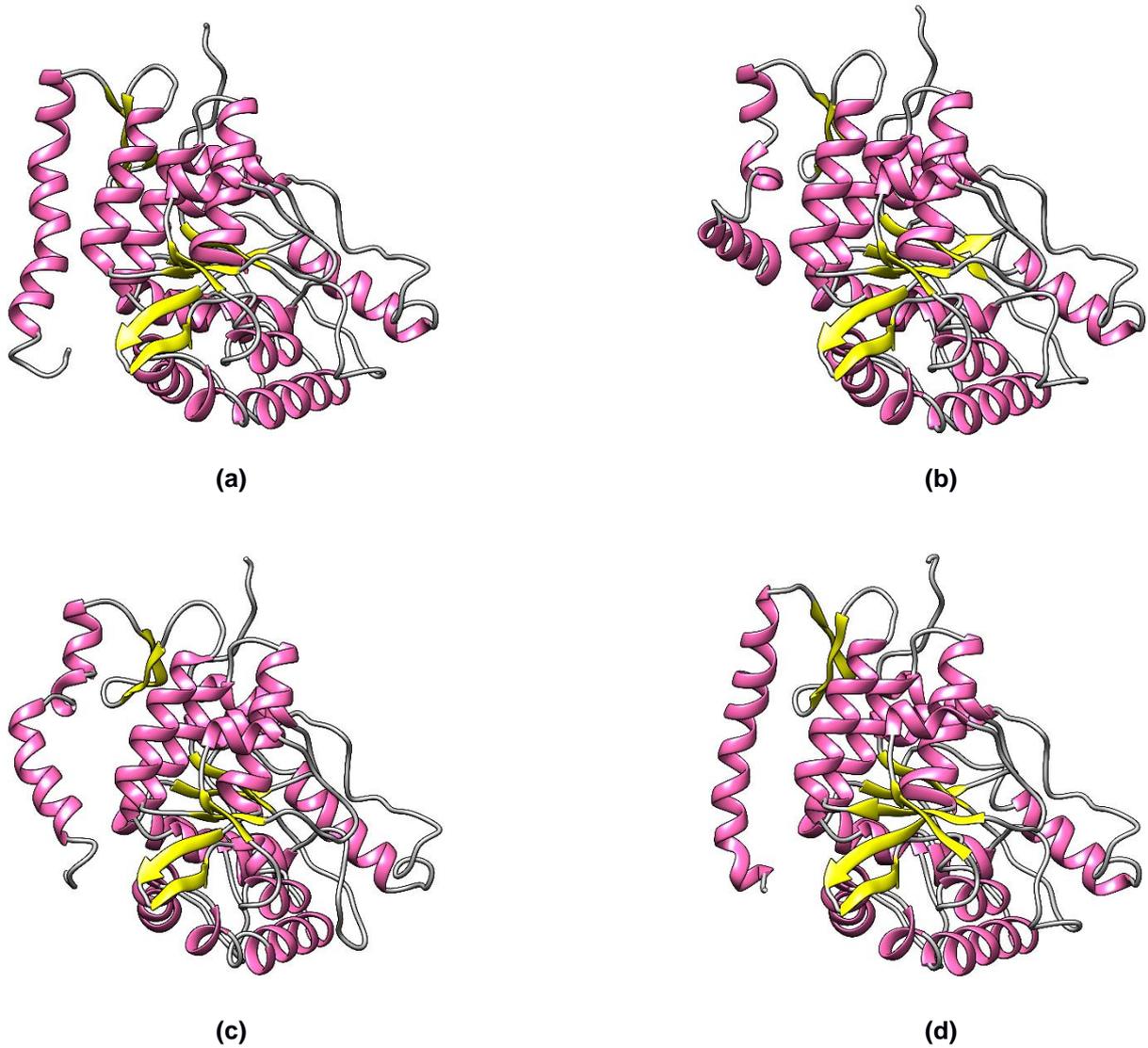


Figura 21 – Modelos estruturais obtidos como resultado das simulações de dinâmica molecular para a selvagem e variantes, visualizados e coloridos segundo suas estruturas secundárias pelo Chimera. (a) Estrutura 4V06. (b) Estrutura variante P206S. (c) Estrutura variante R303W. (d) Estrutura variante R441H. As alfa-hélices se encontram na cor rosa; as folhas-beta na cor amarela; voltas e alças estão coloridas de cinza.

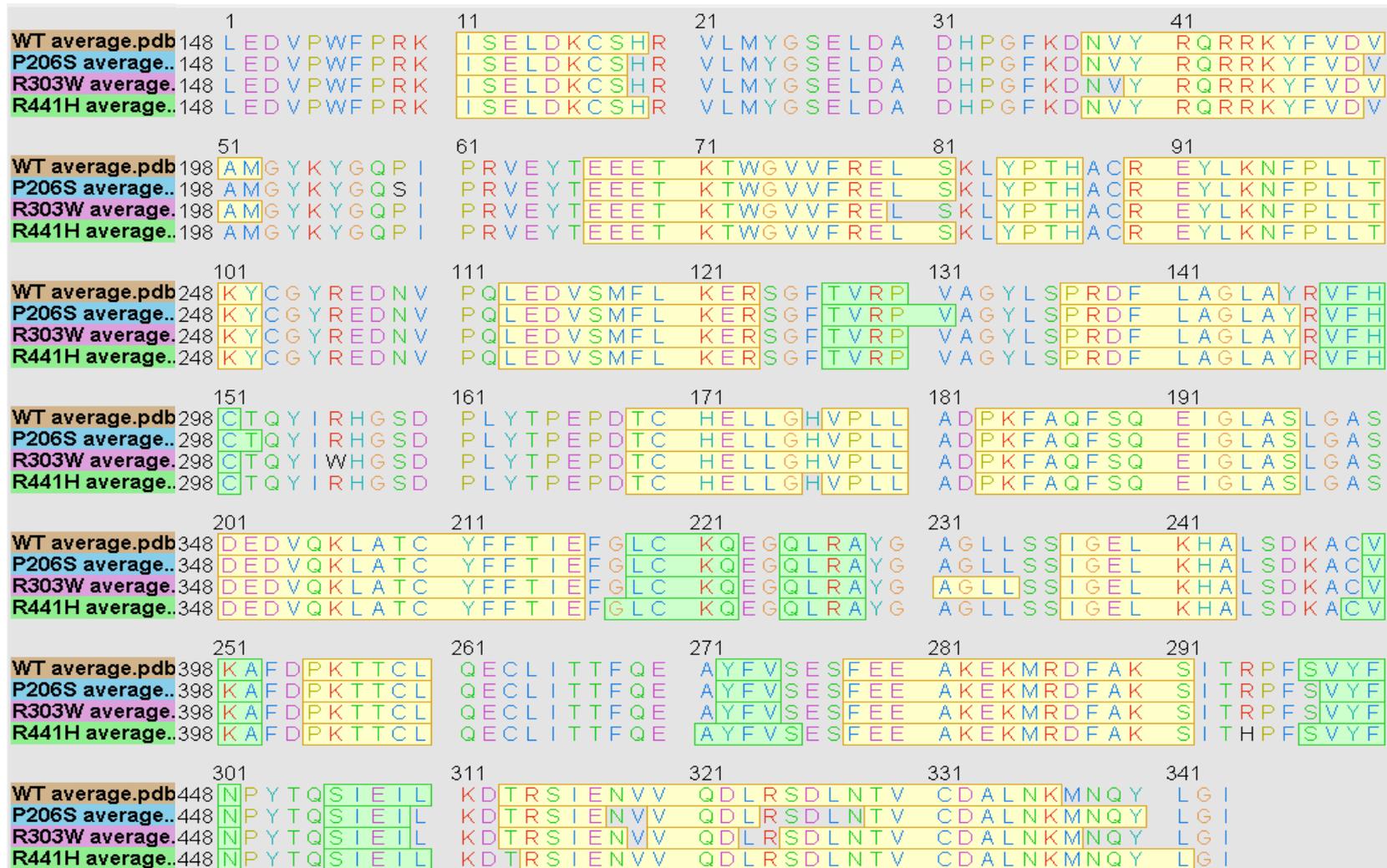


Figura 22 – Alinhamento das seqüências referentes aos modelos obtidos após a simulação da dinâmica molecular para as estruturas selvagem e variantes, realizado pelo Chimera. Os resíduos destacados com preenchimento amarelo fazem parte de alfa-hélice, enquanto aqueles destacados com preenchimento verde fazer parte de folhas-beta. A ausência de preenchimento indica regiões de voltas e alças.

5. CONCLUSÃO

Fazendo-se uso das análises *in silico* foi possível inferir as alterações causadas pelas mutações selecionadas, tanto pela análise funcional quanto de seu comportamento estrutural dinâmico pelas simulações de dinâmica molecular.

Das 46 variantes primeiramente compiladas, as três mutações previamente associadas a doenças psiquiátricas foram selecionadas para serem comparadas à proteínas selvagem por dinâmica molecular. Tais variantes haviam sido anteriormente classificadas por nosso grupo como danosas por pelo menos 50% dos preditores utilizados para predição funcional.

Os resultados prévios obtidos na DM curta realizada pelo MDWeb destacaram as mutações R156E, R225Q, P277L, R303W, S383FM, E430G, R441H e D473N como as que mais alteram a dinâmica estrutural da TPH2. Tal resultado previamente destaca o possível dano causado pelas mutações R303W e R441H, associadas ao Transtorno Déficit de Atenção e Hiperatividade e Depressão, respectivamente.

Já as simulações de dinâmica molecular longa com o GROMACS relevaram pela análise das trajetórias resultantes que todas as mutantes mostraram ter maior valor médio de RMSD do que a WT e afetar o empacotamento da proteína, como mostrado nos valores médios do Rg e do NHB intra e intermolecular. Sugere-se que P206S e R441H aumentam trivialmente a compactação da estrutura, enquanto R303W diminui tal característica. Esta mesma variante mostra estabelecer menos ligações de hidrogênio intramolecular e mais intermoleculares, além de apresentar a maior área de superfície acessível ao solvente quando comparada à WT e as outras mutantes.

Do exposto, conclui-se que P206S, R303W e R441H afetam o comportamento e dinâmica estrutural da TPH2, especialmente R303W. Desse modo, os resultados corroboram e dão suporte as associações entre essas variantes e suas respectivas desordens mentais. Mais estudos são necessários para aquisição de mais conhecimento sobre a TPH2 e como essa proteína e suas variantes podem ser exploradas como alvo de medicamentos, ajudando no tratamento das psicopatologias

associadas.

Além disso, foi possível realizar inferências prévias sobre alteração da estrutura secundárias das variantes a partir da análise de suas estruturas médias obtidas na dinâmica. Do mesmo modo, o modelo 4V06 refinado foi obtido com sucesso e estará disponível no banco de dados do nosso grupo (www.snpmol.org) para utilização de estudos, assim como os dados das análises *in silico*.

REFERÊNCIAS BIBLIOGRÁFICAS

- Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. (2015) GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. **SoftwareX**, v.1-2, p. 19-25.
- Allen, M. P.; Tildesley, D. J. (1987) Computer simulation of liquids, Oxford University Press, New York, p.39.
- Benkert, P.; Biasini, M.; Schwede, T. (2011) Toward the estimation of the absolute quality of individual protein structure models. **Bioinformatics**, v. 27, n. 3, p. 343-350.
- Benkert, P.; Kunzli, M.; Schwede, T. (2009) QMEAN server for protein model quality estimation. **Nucleic Acids Research**. v. 37, p. W510–W514.
- Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. M.; Weissig, H.; Shindyalov, N. N.; Bourne, P. E. (2000) The Protein Data Bank. **Nucleic acids research**, v. 28, n. 1, p. 235–242.
- Beveridge, D. L.; McConnell, K. J. (2000) Nucleic acids: theory and computer simulations. **Current Opinion in Structural Biology**, v.10, p. 182-196.
- Bragatti, J. A.; Bandeira, I. C.; De Carvalho, A. M.; Abujamra, A. L.; Leistner-Segal, S.; Bianchin, M. (2014) Tryptophan hydroxylase 2 (TPH2) gene polymorphism and psychiatric comorbidities in temporal lobe epilepsy. **Epilepsy and Behavior**, v. 32, p. 59-63.
- Bromberg, Y.; Rost, B. (2007) SNAP: predict effect of non-synonymous polymorphisms on function. **Nucleic Acids Research**, v. 35, n. 11, p. 3823–3835.
- Capriles, P. V. S. Z.; Guimarães, A. C. R.; D'otto, T., et al. (2010) Structural modelling and comparative analysis of homologous, analogous and specific proteins from *Trypanosoma cruzi* versus *Homo sapiens*: putative drug targets for chagas' disease treatment. **BMC Genomics**, v. 11, p. 1-10.
- Carkaci-Salli, N.; Flanagan, J. M.; Martz, M. K.; Salli, U.; Walther, D. J.; Bader, M.; Vrana, K. E. (2006) Functional domains of human tryptophan hydroxylase 2 (hTPH2). **Journal of Biological Chemistry**, v. 281, p. 28105-28112.
- Cash, C. D. (1998) Why tryptophan hydroxylase is difficult to purify: a reactive oxygen-derived species-mediated phenomenon that may be implicated in human pathology. **General Pharmacology: The Vascular System**, v. 30, n. 4, p. 569-574.
- Chandrasekaran, P.; Rajasekaran, R. (2017) Deciphering the molecular effects of mutations on ATRX cause ATRX syndrome: A molecular dynamics study. **Journal of Cellular Biochemistry**, v. 118, p. 3318-3327.
- Chen, G.-L.; Miller, G. (2012) M. Advances in tryptophan hydroxylase-2 gene expression regulation: new insights into serotonin-stress interaction and clinical

implications. **American journal of medical genetics B Neuropsychiatric genetics**, v. 159B, n. 2, p. 152–171.

Cichon, S.; Winge, I.; Mattheisen, M.; Georgi, A.; Karpushova, A.; Freudenberg, J.; Freudenberg-Hua, Y.; Babadjanova, G.; Van Den Bogaert, A.; Abramova, L. I.; Kapiletti, S.; Knappskog, P. M.; McKinney, J.; Maier, W.; Jamra, R. A.; Schulze, T.G.; Schumacher, J.; Propping P.; Rietschel, M.; Haavik, J.; Nöthen, M. M. (2008) Brain-specific tryptophan hydroxylase 2 (TPH2): A functional Pro206Ser substitution and variation in the 5'-region are associated with bipolar affective disorder. **Human Molecular Genetics**, v. 17, n. 1, p. 87–97.

Colovos, C.; Yeates, T. (1993) Verification of protein structures: Patterns of nonbonded atomic interactions. **Protein Science**, v. 9, p. 1511–1519.

Craveur, P.; Joseph, A. P.; Esque, J.; Narwani, T. J.; Noël, F.; Shinada, N.; Goguet, M.; Leonard, S.; Poulain, P.; Bertrand, O.; Faure, G.; Rebehmed, J.; Chozlane, A.; Swapna, L. S.; Bhaskara, R. M.; Barnoud, J.; Téletchéa, S.; Jallu, V.; Cern, J.; Schneider, B.; Etchebest, C.; Srivivasan, N.; Gelly, J. C.; Brevern, A. G. (2015) Protein flexibility in the light of structural alphabets. **Frontiers Molecular Biosciences**, v. 2, p. 1- 20.

Crawford, A.; Fassett, R. G.; Geraghty D. P.; Kunde, D. A.; Ball, M.J.; Robertson I. K.; Coombes, J. S. (2012) Relationships between single nucleotide polymorphisms of antioxidant enzymes and disease. **Gene**, v. 501, n. 2, p. 89–103.

David, C. C.; Jacobs, D. J. (2014) Principal Component Analysis: A Method for Determining the Essential Dynamics of Protein. **Methods in Molecular Biology**, v. 1084, p. 193-226.

De Freitas, M. C. (2016) Modelagem Estrutural e Análise In Silico dos Polimorfismos da Proteína Triptofano Hidroxilase 2 (TPH2). 2016. 64f. Monografia – Universidade Federal do Estado do Rio de Janeiro, Rio de Janeiro.

De Marco, L.; Thämer, M.; Reppert, M.; Tokmakoff, A. (2014) Direct observation of intermolecular interactions mediated by hydrogen bonding. **The Journal of Chemical Physics**, v. 141, n.3, p. 034502-1-034502-10.

Edvardsen, J.; Torgersen, S.; Roysamb, E.; Lygren, S.; Skre, I.; Onstad, A.; Oien, P. A. (2008) Heritability of bipolar spectrum disorders. Unity or heterogeneity? **Journal of Affective Disorders**, v. 106, p. 229-240.

Eisenberg, D.; Lüthy, R.; Bowie, J. U. (1997) VERIFY3D: Assessment of protein models with three-dimensional profiles. **Methods in Enzymology**, v. 277, p. 396–404.

Eloyan, A.; Muschell, J.; Nebel, M. B.; Liu, H.; Han, F.; Zhao, T.; Barber, A. D.; Joel, S.; Pekar, J. J.; Mostofsky, S. H.; Caffo, B. (2012) Automated diagnoses of attention deficit hyperactive disorder using magnetic resonance imaging. **Frontiers in Systems Neuroscience**, v. 6, p.1-9.

Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. (1995)

A smooth particle mesh Ewald method. **Journal of Chemical Physics**, v. 103, n. 19, p.31–34.

Fan, H.; Mark, A. E. (2003) Refinement of homology-based protein structures by molecular dynamics simulation techniques. **Protein Science**, v. 13, p. 211-220.

Ferrari A. J.; Charlson, F. J.; Norman, R. E.; Patten, S. B.; Freedman, G.; Murray, C. J. L.; Vos, T.; Whiteford, H. A.; (2013) Burden of Depressive Disorders by Country, Sex, Age, and Year: Findings from the Global Burden of Disease Study 2010. **PLoS Medicine**, v. 10, p.e1001547.

Fitzpatrick, P. F. (2003) Mechanism of aromatic amino acid hydroxylation. **Biochemistry**, v. 42, n.48, p. 14083-14091.

Godzik, A. (2011) Metagenomics and the protein universe. **Current Opinion in Structural Biology**, v. 21, n. 3, p. 398-403.

Guze, S. B.; Robins, E. (1970) Suicide and Primary Affective Disorders. **The British Journal of Psychiatry**, v. 117, p. 437-438.

Heo, L.; Park, H.; Seok, C. (2013) GalaxyRefine: protein structure refinement driven by side-chain repacking. **Nucleic Acids Research**, v.41, p. W384-W388.

Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. (1997) LINCS: A Linear Constraint Solver for Molecular Simulations. **Journal of Computational Chemistry**, v. 18, n. 12, p.1463–1472.

Hetchman, L. (1993) Genetic and Neurobiological Aspects of Attention Deficit Hyperactive Disorder: A Review. **Journal of Psychiatry & Neuroscience**, v.19, n. 3, p. 193-201.

Hiroi, R.; Handa, R. J. (2013) Estrogen receptor-B regulates human tryptophan hydroxylase-2 through an estrogen response element in the 5' untranslated region. **Journal of Neurochemistry**, v. 127, p. 487-495.

Hospital, A.; Andrio, P.; Fenollosa, C.; Cicin-Sain, D.; Orozco, M.; Gelpí, J. L. (2012) MDWeb and MDMoby: an integrated web-based platform for molecular dynamics simulations. **Bioinformatics**, v. 28, n.9, p. 1278-1279.

Huften, S. E.; Jennings, I. G.; Cotton, R. G. H. (1995) Structure and function of the aromatic amino acid hydroxylases. **Biochemistry Journal**, v. 331, p. 353-366.

Humphrey, W.; Dalke, A.; Schulten, K. (1996) VMD: visual molecular dynamics. **Journal of Molecular Graphics**. v.14, p. 33–38.

Ishak, S. N. H.; Aris, S. N. A. M.; Halim, K. B. A.; Ali, M. S. M.; Leow, T. C.; Kamarudin, N. H. A.; Masomian, M.; Rahman, R. N. Z. R. A. (2017) Molecular dynamic simulation of space and earth-grown crystal structures of thermostable T1 lipase *Geobacillus zalihae* revealed a better structure. **Molecules**, v. 22, n. 10, p. 1574-1586.

Joshi, H.; Bhatia, D.; Krishnan, Y.; Maiti, P. K. (2017) Probing the structure and *in silico* stability of cargo loaded. **Nanoscale**, v. 9, p. 4467-4477.

Kaufman, S. Aromatic Amino Acid Hydroxylases. (1987) In: Boyer, P. D.; Krebs, E. G.(Editores). The Enzymes, 3^a ed. Academic Press, Inc, v. XVIII, p. 217-317.

Khan, F. I.; Wei, D. Q.; Gu, K. R.; Hassan, M. I.; Tabrez, S. (2016) Current updates on computer aided protein modeling and designing. **International Journal of Biological Macromolecules**, v. 85, p. 48-62.

Krebs, B.B.; De Mesquita, J. F. (2016) Amyotrophic Lateral Sclerosis Type 20 – In Silico Analysis and Molecular Dynamcis Simulation of hnRNPA1. **PLoS ONE**, v.11, n. 7, p. 1-18.

Kuzmanic, A.; Zagrovic, B. (2010) Determination of ensemble-average pairwise root mean-square deviation from experimental b-factors. **Biophysical Journal**, v. 98, n. 5, p. 861-871.

Laskowski, R. A.; Macarthur, M. W.; Moss, D. S.; Thornton, J. M. (1993) PROCHECK: a program to check the stereochemical quality of protein structures. **Journal of Applied Crystallography**, v. 26, p. 283–291.

Lin, Y. M. J.; Chao, S. C.; Chen, T. M.; Lai, T. J.; Chen, J. S.; Sun, H. S. (2007) Association of Functional Polymorphisms of the Human Tryptophan Hydroxylase 2 Gene with Risk for Bipolar Disorder in Han Chinese. **Archives of General Psychiatry**, v. 64, n. 9, p. 1015–1024.

Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E. (2010). Improved side-chain torsion potentials for the Amber ff99SB protein force field. **Proteins**, v. 78, n.8, p.1950-1958.

Liu, S. Q.; Meng, Z. H.; Fu, Y. X.; Zhang, K. Q. (2010) Insights derived from molecular dynamics simulation into the molecular motions of serine protease proteinase K. **Journal of Molecular Modeling**, v.16, p. 17– 28.

Lovell, S. C.; Davis, I. W.; Arendall III, W. B.; De Bakker, P. I. W.; Word, J. M.; Prisant, M. G.; Richardson, J. S.; Richardson, D. C. (2003) Structure validation by Calpha geometry: phi,psi and Cbeta deviation. **Proteins**. v. 50, p. 437–450.

Ma, J.; Xiao, H.; Yang, Y.; Cao, D.; Wang, L.; Yang, X.; Qiu, X.; Qiao, Z.; Song, J.; Liu, Y.; Wang, P.; Zhou, J.; Zhu, X. (2015) Interaction of tryptophan hydroxylase 2 gene and life events in susceptibility to major depression in a Chinese Han population. *Journal of Affective Disorders*, v. 188, p. 304-309.

Martínez, L. (2015) Automatic identification of mobile and rigid substructures in molecular dynamics simulations and fractional structural fluctuation analysis. **PLoS One**, v. 20, n. 3, p. e0119264.

McKinney, J. A.; Turel, B.; Winge, I.; Knapskog, P. M.; Haavik, J. (2009) Functional properties of missense variants of human tryptophan hydroxylase 2. **Human Mutation**,

v. 30, n. 5, p. 787–794.

McKinney, J.; Johansson, S.; Halmoy, A.; Dramsdahl, M.; Winge, I.; Kanppskog, P. M.; Haavik, J. (2008) A loss-of-function mutation in tryptophan hydroxylase 2 segregating with attention-deficit/hyperactivity disorder. **Molecular Psychiatry**, v.13, p. 365-367.

Mosienko, V.; Bies, D.; Pasqualetti, M.; Waider, J.; Matthes, S.; Qadri, F.; Bader, M.; Alenina, N. (2015) Life without brain serotonin: Reevaluation of serotonin function with mice deficient in brain serotonin synthesis. **Behavioural Brain Research**, v. 277, p. 78-88.

Murphy, K. L.; Zhang, X.; Galnetdlnov, R.R.; Beaulieu, J. M.; Caron, M. G. A. (2008) Regulatory Domain in the N Terminus of Tryptophan Hydroxylase 2 Controls Enzyme Expression. **The Journal of Biological Chemistry**, v. 283, n. 19, p.13216-13224.

Namba, A. M.; Da Silva, V. B.; Da Silva, C. H. T. P. (2008) Dinâmica molecular: teoria e aplicações em planejamento de fármacos. **Eclética Química**, v. 33, n. 3, p. 13-23.

Narang, S. S.; Shuabi, S.; Goyal, D.; Gaoyal, B. (2017) Assessing the effect of D59P mutation in the DE loop region in amyloid aggregation propensity of β 2-microglobulin: A molecular dynamics simulation study. **Journal of Cellular Biochemistry**, p. 1-11.

National Institute of Mental Health (2013). *Bipolar Disorder*. Disponível em: https://www.nimh.nih.gov/health/topics/bipolar-disorder/index.shtml#part_152505 Acessado em: 29/10/2017

Ndagi, U.; Mhlongo, N. N.; Soliman, M. E. (2017) The impact of Thr91 mutation on c-Src resistance to UM-164: molecular dynamics study revealed a new opportunity for drug design. **Molecular BioSystems**, v. 13, p. 1157-1171.

Ng, P. C.; Henikoff, S. (2006) Predicting the Effects of Amino Acid Substitutions on Protein Function. **Annual Review Genomics and Human Genetics**, v. 7, p. 61- 80.

Niedzialkowska, E.; Gasiorowska, O.; Handing, K. B.; Majorek, K. A.; Porebski, P. J.; Shabalín, I. G.; Zasadzinska, E.; Cymborowski, M.; Minor, W. (2016) Protein purification and crystallization artifacts: The tale usually not told. **Protein Science**, v. 25, n. 3, p. 720-733.

Online Mendelian Inheritance in Man, OMIM®. (2015) McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University (Baltimore, MD). World Wide Web URL: <http://omim.org/>

Pal, D.; Eisenberg, D. (2005) Inference of Protein Function from Protein Structure. **Structure**, v. 13, p. 121-130.

Patrick, R. P.; Ames, B. N. (2014) Vitamin D hormone regulates serotonin synthesis. Part 1: relevance for autism. **The FASEB Journal**, v. 28, p. 2398- 2413.

Pettersen, E.F.; Goddard, T.D.; Huang, C.C.; Couch, G.S.; Greenblatt, D.M.; Meng, E.C.; Ferrin, T.E. (2004) UCSF Chimera - A Visualization System for Exploratory

- Research and Analysis. **Journal of Computational Chemistry**, v. 25, p. 1605-1612.
- Pundir, S.; Martin, M. J.; O'donovan, C. (2017) UniProt Protein Knowledgebase. **Protein Bioinformatics**, v. 1558, p. 41-55.
- Rodrigues, J. P. G. L. M.; Levitt, M.; Chopra, G. (2012) KoBaMIN: a knowledge-based minimization web server for protein structure refinement. **Nucleic Acids Research**, v. 40, p. W323–W328.
- Sadkowski, M.; Dennis, B., Clayden, R. C.; Elsheikh, W.; Rangarajan, S.; Dejesus, J.; Samaan, Z. (2013) The role of the serotonergic system in suicidal behavior. **Neuropsychiatric Disease and Treatment**, v. 9, p. 1699-1716.
- Scheraga, H. A.; Khalili, M.; Liwo, A. (2007) Protein-Folding Dynamics: Overview of Molecular Simulation Techniques. **Annual Review of Physical Chemistry**, v. 58, p. 57-83.
- Sherry, S. T.; Ward, M. H.; Kholodov, M.; Baker, J.; Phan, L.; Smigielski, E. M.; Sirotkin, K. (2001) dbSNP: the NCBI database of genetic variation. **Nucleic acids research**, v. 29, n. 1, p. 308–311.
- Soares, C. M. (1997) Simulação de proteínas usando métodos de mecânica/dinâmica molecular *Química*, v. 64, p. 33-44.
- Sullivan, P. F.; Neale, M. C.; Kendler, K. S. (2000) Genetic Epidemiology of Major Depression: Review and Meta-Analysis. **The American Journal of Psychiatry**, v. 157, p. 1552-1562.
- Tideman, K. D.; Christensen, H. E. M.; Hoeck, N.; Harris, P.; Boesen, J.; Peters, G. H. (2016) Stabilization of tryptophan hydroxylase 2 by l-phenylalanina-induced dimerization. **FEBS Open Bio**, v. 6, n. 10, p. 987-999.
- Turner, P. J. (2005) XMGrace, Version 5.1.19. Center for Coastal and Land-Margin Research, Oregon Graduate Institute of Science and Technology, Beaverton, OR. Disponível em: <<http://plasma-gate.weizmann.ac.il/Grace/>> Acessado em: 29/10/2017.
- Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. (2005) GROMACS: fast, flexible, and free. **Journal Computational Chemistry**, v.26, n. 16, p. 1701–1718.
- Van Gunsteren, H. J.; Berendsen, H. J. C. (1990) Computer Simulation of Molecular Dynamics: Methodology, Applications, and Perspectives in Chemistry. **Angewandte Chemie International Edition in English**, v. 29, p. 992-1023.
- Veenstra-Vanderweele, J.; Anderson, G. M.; Cook, E. H. (2000) Pharmacogenetics and the serotonin system: Initial studies and future directions. **European Journal of Pharmacology**, v. 410, p. 165–181.
- Verli, H. Dinâmica Molecular (2014a) In: Verli, H. (Org). *Bioinformática: da Biologia à*

Flexibilidade Molecular. 1ª ed. São Paulo: SBBq, p. 173-187.

Verli, H. Níveis de Informação Biológica (2014b) In: Verli, H. (Org). Bioinformática: da Biologia à Flexibilidade Molecular. 1ª ed. São Paulo: SBBq, p. 2-12.

Verli, H. O que é Bioinformática? (2014c) In: Verli, H. (Org). Bioinformática: da Biologia à Flexibilidade Molecular. 1ª ed. São Paulo: SBBq, p. 2-12.

Walther, D. J.; Bader, M. (2003) A unique central tryptophan hydroxylase isoform. **Biochemical Pharmacology**, v. 66, n. 9, p. 1673–1680.

Wang, C. H.; Liu, C.; Cong, E. Z.; Xu, G. L.; Lv, T. T.; Zhang, Y. I.; Ning, Q. F.; Wang, J. K.; Nie, H. Y.; Li, Y. (2016) Association of tryptophan hydroxylase-2 polymorphisms with oppositional defiant disorder in a Chinese Han population. **Behavioral and Brain Functions**, v. 12, n. 30, p.1-12.

Wang, L.; Erlandsen, H.; Haavik, J.; Knappskog, P. M.; Stevens, R. C. (2002) Three-Dimensional Structure of Human Tryptophan Hydroxylase and Its Implications for the Biosynthesis of the Neurotransmitters Serotonin and Melatonin. **Biochemistry**, v. 41, n. 42, p.12569-12574.

Wang, X.; Wang, Z.; Wu, Y.; Hou, Z.; Yuan, Y.; Hou G. (2015) Tryptophan hydroxylase 2 gene is associated with cognition in late-onset depression in a Chinese Han population. **Neuroscience Letters**, v. 600, p. 98-103.

Wiederstein, M.; Sippl, M. J. (2007) ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. **Nucleic Acids Research**, v. 35, p. W407–W410.

Windahl, M. S.; Boesen, J.; Karlsen, P. E.; Christensen, H. E. M. (2009) Expression, Purification and Enzymatic Characterization of the Catalytic Domains of Human Tryptophan Hydroxylase Isoforms. **Protein Journal**, v. 28, p. 400-406.

Winge, I.; McKinney, A.; Knappskog, P. M.; Haavik, J. (2007) Characterization of wild-type and mutant forms of human tryptophan hydroxylase 2. **Journal of Neurochemistry**, v. 100, n. 6, p. 1648-1657.

Wlodawer, A.; Minor, W.; Dauter, Z.; Jaskolski, M. (2008) Protein crystallography for non-crystallographers, or how to get the best (but not more) from published macromolecular structures. **FEBS Journal**, v. 275, n. 1, p. 1-21.

Xu, D.; Zhang, Y. (2011) Improving the Physical Realism and Structural Accuracy of Protein Models by a Two-Step Atomic-Level Energy Minimization. **Biophysical Journal**, v. 101, p. 2525-2534.

Xu, X. M.; Ding, M.; Pang, H.; Wang, B. J. (2014) TPH2 gene polymorphisms in the regulatory region are associated with paranoid schizophrenia in Northern Han Chinese. **Genetics and Molecular Research**, v. 13, n. 1, p. 1497-1507.

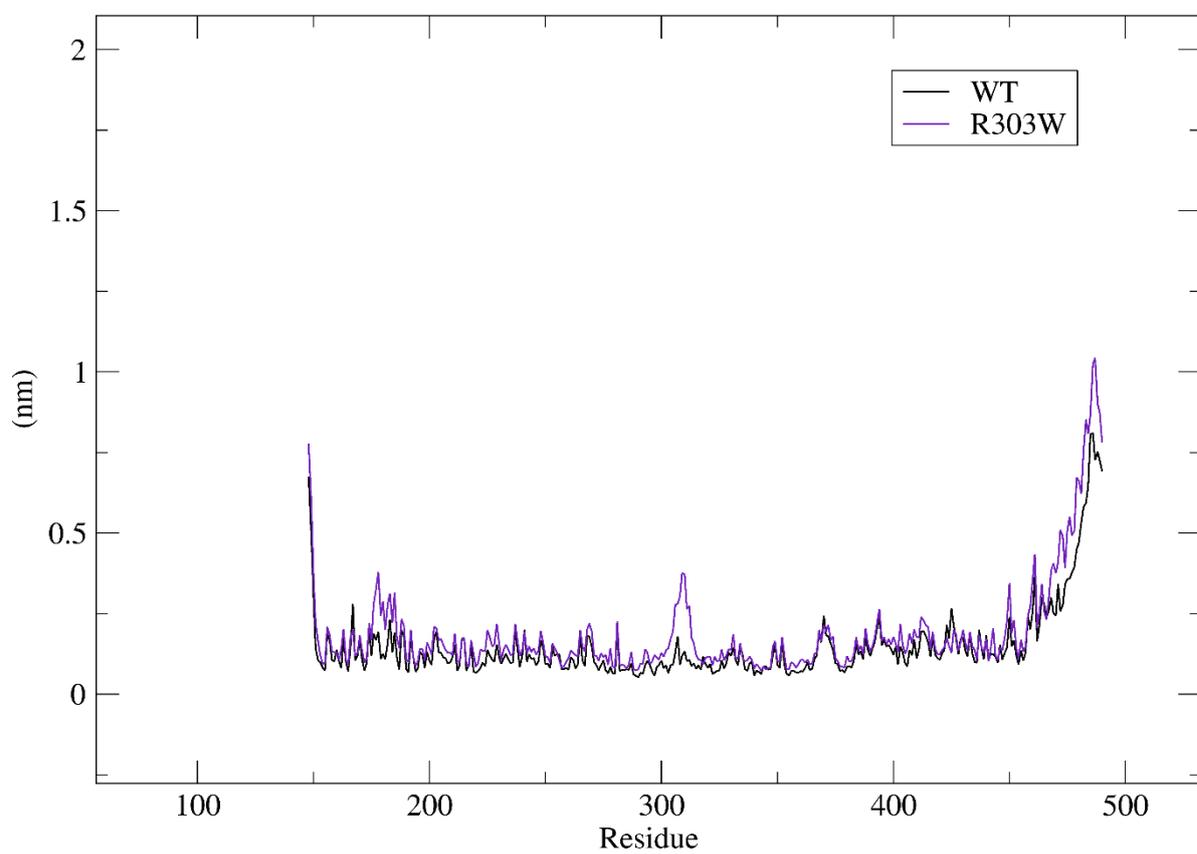
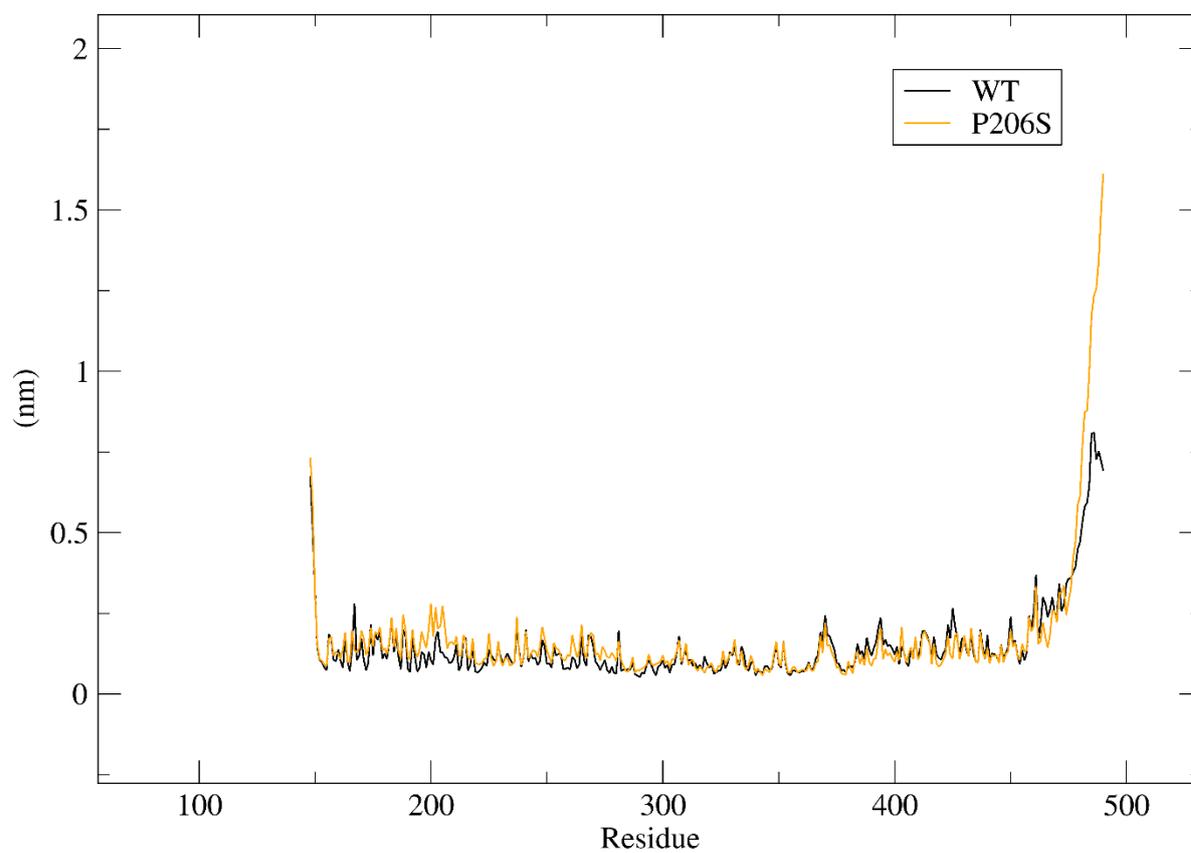
Yang, L. Q.; Sang, P.; Tao, Y.; Fu, Y. X.; Zhang, K. Q.; Xie, Y. H.; Liu, S. Q. (2014)

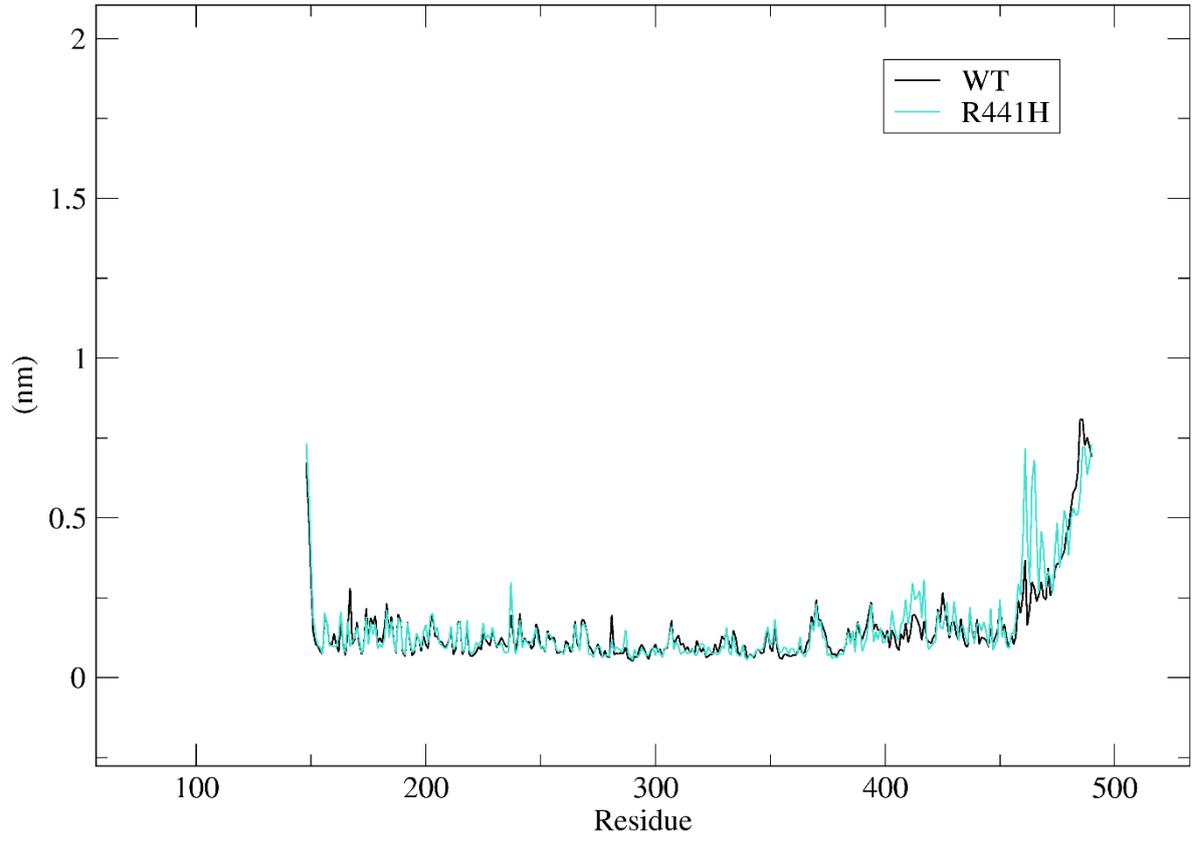
Protein dynamics and motions in relation to their functions: several case studies and the underlying mechanisms. **Journal of Biomolecular Structures and Dynamics**, v. 32, n. 3, p. 371-393.

Zhang, Y.; Gainetdinov, R. R.; Beaulieu, J. –M.; Sotnikova, T. D.; Burch, L. H.; Willians, R. B.; Schawartz, D. A.; Krishnan, K. R. R.; Garon, M. G. (2005) Loss-of-function Mutation in Tryptophan Hydroxylase-2 Identified in Unipolar Major Depression. **Neuron**, v. 45, n. 1, p. 11–16.

Zill, P.; Büttner, A.; Eisenmenger, W.; Bondy, B.; Ackenheil, M. (2004). Regional mRNA expression of a second tryptophan hydroxylase isoform in postmortem tissue samples of two human brains. **European Neuropsychopharmacology**, v. 14, p. 282-284.

Anexo A – Gráficos individuais constando os valores de RMSF para cada estrutura mutante em comparação com à selvagem, durante 80 ns de simulação.





Anexo B – Gráficos individuais constando os valores de RMSF para cada estrutura mutante em comparação com a selvagem, durante o período de 62 ns a 80 ns da simulação.

