



Universidade Federal do Estado do Rio de Janeiro – UNIRIO
Centro de Ciências Biológicas e da Saúde – CCBS
Programa de Pós-Graduação em Biologia Molecular e Celular – PPGBMC

**Modelagem estrutural e análises *in silico* das mutações da proteína
VCP humana relacionadas à Esclerose Lateral Amiotrófica**

Lorena Giannini Alves Moreira

Rio de Janeiro, 2016

Lorena Giannini Alves Moreira

Modelagem estrutural e análises *in silico* das mutações da proteína VCP humana relacionadas à Esclerose Lateral Amiotrófica

Dissertação apresentada ao Programa de Pós-Graduação em Biologia Molecular e Celular (PPGBMC) da Universidade Federal do Estado do Rio de Janeiro (UNIRIO) como parte dos requisitos para a obtenção do título de mestre em ciências.

Área de concentração: Biologia Molecular e Celular

Orientadora: Profa. Dra. Joelma Freire de Mesquita

UNIRIO

2016

Resumo

A esclerose lateral amiotrófica é uma desordem neurodegenerativa letal, caracterizada pela degeneração dos neurônios motores. Mutações no gene que codifica a proteína VCP (uma proteína da superfamília das AAA-ATPases) estão associadas com o fenótipo da esclerose lateral amiotrófica familiar (fALS). Neste estudo, pretendemos avaliar as consequências na estrutura da proteína VCP causadas por treze mutações relacionadas à ALS. Para obter as mutações relacionadas à ALS, buscamos nos bancos de dados PubMed, OMIM e UniProt. A estrutura tridimensional da proteína nativa da VCP foi obtida computacionalmente por modelagem ab initio (I-Tasser e Rosetta) e comparativa (Mholline, Swiss Model, HHpred, IntFOLD, Phyre2, and RaptorX). Para determinar o melhor modelo para gerar as estruturas mutantes realizamos alinhamentos estruturais de cada modelo com a estrutura da VCP determinada experimentalmente disponível no Protein Data Bank. Os valores de RMSD (root mean square deviation) e de TM-score foram analisados para selecionar o melhor modelo. Para as treze estruturas mutantes foi realizado a mutagênese in silico, e a simulação de dinâmica molecular (DM) foi realizado utilizando-se o programa GROMACS. Estas estruturas foram alinhadas com a estrutura nativa obtida por modelagem depois da simulação de dinâmica. O melhor modelo foi gerado por modelagem comparativa com o algoritmo I-Tasser. Os valores de RMSD dos alinhamentos entre as estruturas mutantes e nativas da proteína VCP variaram de 1.04 (R191G) a 1.12 (R191Q). Os resultados das simulações sugerem que a mutação R191Q pode resultar em alterações estruturais na proteína VCP no desenvolvimento da ALS.

Abstract

Amyotrophic lateral sclerosis (ALS) is a progressive, lethal, neurodegenerative disorder, characterized by the degeneration of motor neurons. Mutations in the gene encoding VCP protein (a protein of the superfamily of AAA-ATPase) are associated with a phenotype of familial ALS (fALS). In the present study, we aim to evaluate the consequences caused by thirteen mutations related to ALS in the VCP protein structure. Materials and methods: To obtain the mutations related to ALS, we performed data mining in the PubMed, OMIM and UniProt databases. The three-dimensional structure of the native VCP was obtained computationally by *ab initio* (I-Tasser and Rosetta) and comparative (Mholline, Swiss Model, HHpred, IntFOLD, Phyre2, and RaptorX) modeling. To determine the best model to create the mutant structures, structural alignments were performed using TM-align of each model with experimentally determined VCP structures available in the Protein Data Bank. RMSD (root mean square deviation) and the TM-score were used to select the best model. *In silico* mutagenesis was used to create the thirteen mutants, and molecular dynamics (MD) simulation was performed using the GROMACS software. These structures were aligned with the wild type structure obtained by modeling after the MD simulations. The best model was generated by *ab initio* modeling using the I-Tasser algorithm. The RMSD values of the alignments between the mutants and wild type structures of the modeled VCP varied between 1.04 (R191G) and 1.12 (R191Q). The simulation results suggest that the R191Q mutation may result in structural changes in VCP protein in ALS development.

Sumário

1 INTRODUÇÃO	11
ESCLEROSE LATERAL AMIOTRÓFICA.....	11
VCP.....	13
POLIMORFISMOS DE NUCLEOTÍDEO ÚNICO	15
ANÁLISES <i>IN SILICO</i>	16
OBJETIVO	17
2 MATERIAL E MÉTODOS	18
OBTENÇÃO DA SEQUÊNCIA PROTEICA.....	18
OBTENÇÃO DAS VARIANTES NATURAIS DA VCP	18
PREDIÇÃO FUNCIONAL DAS VARIANTES DA VCP HUMANA	18
PREDIÇÃO ESTRUTURAL DA VCP HUMANA SELVAGEM E DE SUAS VARIANTES.....	19
ANÁLISE DA CONSERVAÇÃO EVOLUCIONÁRIA	20
CRIAÇÃO DO BANCO DE DADOS DA VCP.....	20
3 RESULTADOS E DISCUSSÃO	20
A. PREDIÇÃO FUNCIONAL DAS VARIANTES DA VCP HUMANA	21
B. PREDIÇÃO ESTRUTURAL DA VCP HUMANA SELVAGEM E DE SUAS VARIANTES	24
C. ANÁLISE DA CONSERVAÇÃO EVOLUCIONÁRIA	32
4. CONCLUSÃO	34
5. REFERÊNCIAS BIBLIOGRÁFICAS	35

Agradecimentos

À D'us, em primeiro lugar, pela presença onipotente e incomensurável auxílio em todo meu caminho. Não tenho dúvidas de que, por muitas vezes, foi visto apenas um par de nossas pegadas na areia.

À minha amada família, Leonardo, Leticia, Larissa, papai Cláudio, mamãe Paola e vovó Lêda, pelos abraços de saudade e por entender a distância que jamais nos separou com seu amor inenarrável. Meus amores, vocês foram e sempre serão minha maior inspiração nos dias difíceis.

À minha família escolhida, Fernando Abreu, Heloisa Rocha, Elias Garcia, Osório Vieira, Clara Abreu, Camila Nogueira, Danielli Lopes, Ana Maseca, Rodolfo Galhardo, Fernanda Mendonça, Lucas Bastos, Bruno Pacífico, Rafael Tomé, Taiana Souza, e tantos outros, pelos ombros que me foram consolo, e por serem a família que eu pude escolher. Pela paciência em ouvir e compreender meus medos e anseios (que sei não serem poucos), pelo amor incondicional que apenas os verdadeiros e bons amigos sabem dar (ainda que venha em forma de bronca em alguns momentos.. rs). Se eu consegui continuar neste Rio de Janeiro “sozinha” por tanto tempo, tenham certeza de que a culpa é de vocês!

Aos nobres professores doutores de minha Universidade, principalmente, Ana Dumans, Ana Machado, Carlos de Maria, Andreia Ayres, Cládice Diniz e Amâncio Júnior, não somente pelo conhecimento teórico tão bem ministrado por vossas senhorias, mas pela experiência de vida e exemplo profissional no qual quero me espelhar num futuro próximo.

Ao amado amigo, professor e diretor de Políticas, Normatização e Registros Acadêmicos de Graduação, Angelo Telesforo Malaquias, pela paciência, compreensão, auxílio imensurável, conselhos e amizade dedicados nestes anos que

se passaram. Por ter ministrado uma das disciplinas em que mais aprendi em toda minha carreira acadêmica, ser o verdadeiro responsável pelo meu “11”, e por ser o Biomédico exemplo de minha formação.

À minha amada amiga, professora e orientadora Profa. Dra. Joelma Freire de Mesquita. Por ter me aceitado como sua orientanda antes mesmo do início das aulas. Pela confiança e ensinamentos sempre dedicados ao longo de toda a minha estada por esta Universidade como graduanda e agora como mestranda, pela paciência e compreensão de todos os dias. Por nossas publicações e prêmios conquistados! Minha querida Joelma, todas as minhas vitórias são nossas e espero que ainda venham muitas mais desta parceria!

Aos servidores da UNIRIO, principalmente, ao magnífico, e muito querido, reitor Luiz Pedro Jutuca, e à querida amiga Rose da secretaria do Departamento de Biologia Molecular e Celular (PPGBMC) e servidores da Pró-reitoria de Extensão (PROEXC), Graduação (PROGRAD) e Pesquisa (PROPG), pelo auxílio e solicitude de sempre.

Lista de Ilustrações

Figura 1 - Porcentagem de ALS explicada por mutações genéticas desde 1992.	12
Figura 2 - Funções Múltiplas da VCP e Co-fatores n-Terminais Envolvidos.	14
Figura 3 - Número de mutações da VCP classificadas como deletérias pelos algoritmos de predição. As barras azuis indicam o total de mutações classificadas como patogênicas (eixo y) em relação aos algoritmos utilizados (eixo x).....	22
Figura 4 - Número de classificações como deletérias por mutação da VCP.....	22
Figura 5 - Resultados obtidos através da análise do algoritmo VarMod.....	24
Figura 6 - Alinhamento tridimensional entre o modelo da VCP selvagem (wt) gerado pelo Mholline e a estrutura de PDBID 3QQ7.....	25
Figura 7 - Alinhamento tridimensional entre o modelo da VCP wt gerado pelo SwissModel e a estrutura de PDBID 3QQ7.....	26
Figura 8-Alinhamento tridimensional entre o modelo da VCP wt gerado pelo HHPred e a estrutura de PDBID 3QQ7.Em vermelho, observa-se o modelo e, em azul, a estrutura 3QQ7.	26
Figura 9 - Alinhamento tridimensional entre o modelo da VCP wt gerado pelo IntFOLDTS e a estrutura de PDBID 3QQ7.	27
Figura 10 - Alinhamento tridimensional entre o modelo da VCP wt gerado pelo Phyre 2 e a estrutura de PDBID 3QQ7.	27
Figura 11 - Alinhamento tridimensional entre o modelo da VCP wt gerado pelo RaptorX e a estrutura de PDBID 3QQ7.	28
Figura 12 - Alinhamento tridimensional entre o modelo 1 da VCP wt gerado pelo I-tasser e a estrutura de PDBID 3QQ7.....	28
Figura 13 - Alinhamento tridimensional entre o modelo 1 da VCP wt gerado pelo Rosetta e a estrutura de PDBID 3QQ7.....	29
Figura 14 - Alinhamento Estrutural entre a VCP nativa e as mutações R191Q (Direita) e R191G (Esquerda).....	32
Figura 15 - Resultados obtidos pelo algoritmo Consurf.....	33

Lista de Tabelas

Tabela 1 - Resultados obtidos através dos alinhamentos estruturais - dos modelos obtidos vs estruturas experimentais - realizados pelo TM-align.....	29
Tabela 2 - Resultados obtidos através dos alinhamentos estruturais - VCP wt pós dinâmica vs mutantes - realizados pelo TM-align.....	31

Lista de Abreviaturas e Siglas

- AAA - Adenosina Trifosfatada ATPases
- ALS – Esclerose Lateral Amiotrófica (do inglês, *Amyotrophic Lateral Sclerosis*)
- DNA – Ácido Desoxirribonucleico
- fALS – Esclerose Lateral Amiotrófica Familiar (do inglês, *Familial Amyotrophic Lateral Sclerosis*)
- GWAS - Estudos de associação de genômica em larga escala (do inglês, Genome-wide Association Study)
- NGS – Sequenciamento de Nova Geração
- MND – Doença do Neurônio Motor (do inglês, *Motor Neuron Disease*)
- RE – Retículo Endoplasmático
- RMSD – Do inglês, *Root Mean Square Deviation*
- sALS – Esclerose Lateral Amiotrófica Esporádica (do inglês, *Sporadic Amyotrophic Lateral Sclerosis*)
- UPS - Sistema Ubiquitina-Proteassoma
- VCP – Proteína Contendo Valosina (do inglês, *Valosin-Containing Protein*)

1 Introdução

Esclerose Lateral Amiotrófica

A esclerose lateral amiotrófica (do inglês “amyotrophic lateral sclerosis”, ALS), também conhecida como doença de Lou Gehrig’s, é uma doença neurodegenerativa que afeta os neurônios motores superiores e inferiores, levando ao óbito em cerca de 3 a 5 anos devido a um quadro clínico de insuficiência respiratória (BEGHI et al., 2006). Tendo sido descrita pela primeira vez em 1869 por Jean-Martin Charcot, a ALS possui uma incidência de 1 a 5 casos a cada 100.000 indivíduos e é considerada a forma mais comum de Doenças do Neurônio Motor (do inglês “motor neuron diseases”, MND) em adultos (MANCUSO; NAVARRO, 2015).

Apesar da causa direta da ALS ainda ser uma incógnita, sabe-se que aproximadamente 5 a 10% dos pacientes possui uma forma hereditária da doença chamada esclerose lateral amiotrófica familiar (do inglês “familial amyotrophic lateral sclerosis”, fALS) que é associada a mutações em pelo menos 24 genes diferentes e pode ser de origem autossômica dominante (maioria dos casos de manifestação na idade adulta), recessiva (frequente nos casos de início juvenil) ou ligada ao cromossomo X (em famílias nas quais os pacientes masculinos apresentam sintomatologia mais severa) (YAMASHITA; ANDO, 2015). Entretanto, algumas destas alterações genéticas também foram encontradas em pacientes que possuem a esclerose lateral amiotrófica esporádica (do inglês “sporadic amyotrophic lateral sclerosis”, sALS), sem origens hereditárias (YAMASHITA; ANDO, 2015). Dentre os genes que possuem mutações relacionadas à ALS, estão: TARDBP, FUS, FIG4, OPTN, SPG11, ATXN2, DAO, UBQLN2, VCP, PFN1, SIGMAR1 (IGUCHI et al., 2013). A Figura 1 ilustra a porcentagem de ALS que foi relacionada a mutações desde 1992. No Brasil, observa-se uma predominância masculina e de caucasianos entre os afetados e mortalidade entre 0,61 e 0,89 por 100000 habitantes dentre os maiores de 20 anos e entre 1,77 e 2,3 por 100000 habitantes dentre maiores de 45 anos (MOURA; CASULARI; CARVALHO GARBINHOVAES, 2016).

Embora existam essas duas categorias de ALS, os pacientes de ambas as formas apresentam fenótipos patológicos similares de fraqueza e atrofia muscular

progressiva, com manifestações de espasticidade e contraturas (MANCUSO; NAVARRO, 2015). Clinicamente, os primeiros sintomas observados são comumente a disartria seguida de disfagia (podendo causar desnutrição) e, posteriormente, dor e falta de equilíbrio motor (GORDON, 2013). Entretanto, as funções sensoriais e o controle do esfíncter, assim como os movimentos oculares, são normalmente preservados (GORDON, 2013). Não existe uma cura para a ALS e os tratamentos existentes são paliativos, baseando-se no alívio da sintomática clínica e melhora da qualidade de vida dos pacientes com o medicamento Riluzol (ZHOU; YI; BONEWALD, 2015). Segundo Al-Chalabi e Lewis (2011), a ALS é a causa mais comum de busca à eutanásia (AL-CHALABI; LEWIS, 2011).

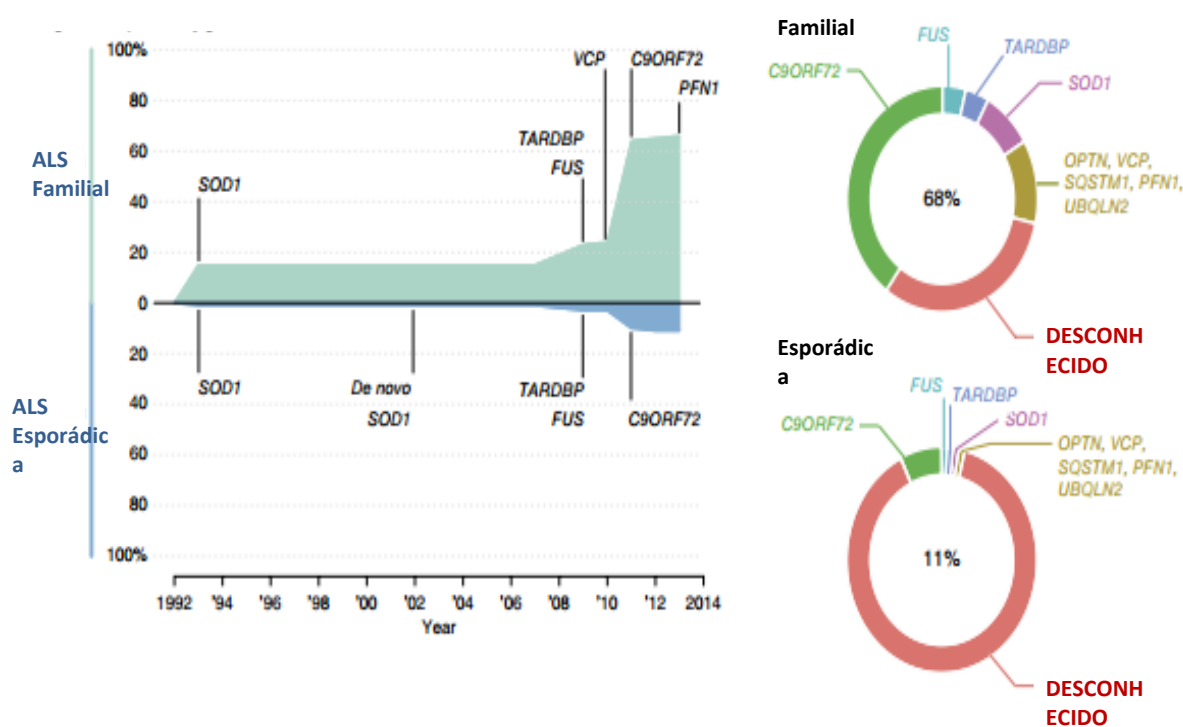


Figura 1 - Porcentagem de ALS explicada por mutações genéticas desde 1992. Linha do tempo das descobertas genéticas na ALS familiar e esporádica. Os valores representam a porcentagem de ALS explicada por cada gene em populações de ascendência Europeia. Fonte: Adaptado de Alan E Renton, Adriano Chiò & Bryan J Traynor. State of play in amyotrophic lateral sclerosis genetics. Nature Neuroscience.(RENTON; CHIÒ; TRAYNOR, 2014).

Desde a década passada, vários consórcios internacionais a níveis genéticos e epigenéticos vêm sendo desenvolvidos e alguns deles pretendem também obter informações sobre estilos de vida e fatores de risco ambientais que possam

demonstrar ter relação com a patogênese da ALS (INGRE et al., 2015). Dessa forma, as tecnologias de Sequenciamento de Nova Geração (NGS) são capazes de gerar grandes quantidades de dados brutos para serem processados a partir da utilização de ferramentas de bioinformática, o que faz com que projetos de ensaios clínicos possam ser facilitados devido aos conhecimentos genéticos advindos destas e de outras novas tecnologias baseadas na análise genética da patologia (SU et al., 2014).

No decorrer dos anos, o entendimento da base genética de doenças complexas, assim como a identificação de centenas de loci genéticos envolvidos nas mesmas, foram amplificados graças à utilização de estudos de associação de genômica em larga escala (GWAS) (TAN et al., 2014). Estes estudos possuem grande relevância, pois a identificação de substituições de aminoácidos responsáveis por alterar a função proteica possui importância significativa para a pesquisa em proteínas e suas implicações no desenvolvimento de patologias (NG; HENIKOFF, 2001).

VCP

A Proteína VCP (do inglês, *Valosin-containing Protein*), também conhecida como p97, pertence à família das AAA ATPases (Adenosina Trifosfatada ATPases associadas à diversas atividades celulares) do tipo II e é composta de um domínio N-terminal de ligação ao cofator e ao substrato, de dois domínios AAA ATPases chamados D1 e D2, e um domínio curto C-terminal de ligação a alguns fatores de modificação de substrato, formando uma estrutura homohexamérica de anel (CHAPMAN; FRY; KANG, 2011; YAMANAKA; SASAGAWA; OGURA, 2012). Apesar da localização celular da VCP ser principalmente o citosol, é possível encontrar uma fração da mesma associada a membranas de organelas, como o Complexo de Golgi e o Retículo Endoplasmático, e uma pequena quantidade no núcleo (YE, 2006).

Como o nome da família AAA ATPases indica, a proteína VCP possui uma vasta gama de funções celulares, sendo a maioria delas relacionada à sinalização por ubiquitina ou para o Sistema Ubiquitina-Proteassoma (UPS), no qual acredita-se que a VCP atue como uma chaperona ubiquitina seletiva, enviando os alvos proteicos poliubiquitinados para o proteassoma 26S que os degradará (KLOPPSTECK et al., 2012). Ainda, a VCP possui funções relacionadas à ativação

da transcrição, mitose, apoptose, via de reparo do DNA (Ácido Desoxirribonucleico), degradação associada ao Retículo Endoplasmático (RE), fusão membrana homotípica entre outras (Figura 2)(KLOPPSTECK et al., 2012). Esta grande diversidade funcional apresentada pela VCP deve-se principalmente ao diferencial de ligação dos distintos cofatores e adaptadores proteicos (YAMANAKA; SASAGAWA; OGURA, 2012).

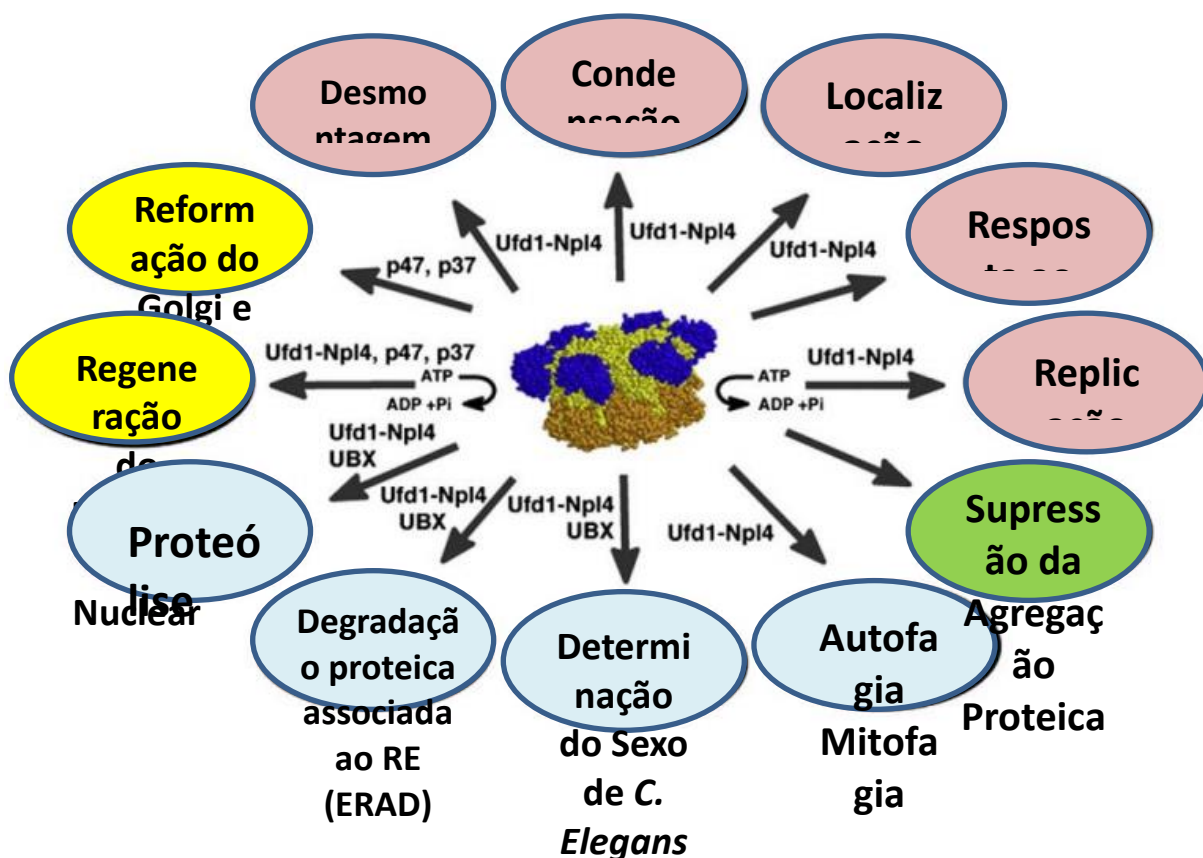


Figura 2 - Funções Múltiplas da VCP e Co-fatores n-Terminais Envolvidos.

Amarelo (Função relacionada à reformação de membrana); Azul (Degradação proteica); Verde (Agregação proteica); Rosa (Ciclo celular e DNA). Fonte: Adaptado de Kunitoshi Yamanaka, Yohei Sasagawa, Teru Ogura. Recent advances in p97/VCP/Cdc48 cellular functions. Biochimica et Biophysica Acta (Yamanaka, Sasagawa et al. 2012).

Alterações na proteína VCP foram relacionadas a várias doenças neurodegenerativas como Alzheimer, Parkinson, esclerose lateral amiotrófica e doenças de repetição de poliglutamina (ISHIKAWA et al., 2012). Além disso, é possível observar o envolvimento da ubiquitinação na patogênese de muitas doenças neurodegenerativas, incluindo a ALS, tendo em vista que essas patologias

apresentam inclusões formadas pelo acúmulo anormal de proteínas imunoreativas para a ubiquitina (YING; WANG; WANG, 2013). Sendo assim, a ligação existente entre a função de genes envolvidos com a ALS e o UPS pode auxiliar no desenvolvimento de novos tratamentos para a ALS e outras doenças do neurônio motor (YING; WANG; WANG, 2013).

Polimorfismos de Nucleotídeo Único

Polimorfismos de nucleotídeo único (SNPs) são polimorfismos que ocorrem em posições de um único par de bases em DNA genômico, o SNP não sinônimo (nsSNP) produz uma proteína diferente e pode afetar a função da proteína (Cooper, Ball et al. 1998). Um dos principais interesses em pesquisas do genoma humano é descobrir se nsSNPs específicos afetam a saúde humana. Os projetos de sequenciamento de alto desempenho de genomas humanos e estudos genômicos de larga escala (GWAS) revelam milhões de, até então desconhecidas variações de nucleotídeos únicos (SNPs) (ALKAN; COE; EICHLER, 2011).

Como marcadores genéticos, SNPs podem ser utilizados para seguir os padrões de herança de regiões cromossômicas de geração em geração e são ferramentas poderosas no estudo de fatores genéticos associados com doenças humanas (JOHNSON; TODD, 2000; RISCH, 2000). Técnicas experimentais para a identificação de SNPs em larga escala na população humana têm permitido o aumento exponencial da base de dados dbSNP NCBI que contém mais de 10 milhões de casos validados (SHERRY et al., 2001).

Uma abordagem de genômica funcional de alto rendimento permite uma melhor compreensão do papel desempenhado pelas etapas de vias biológicas envolvidas em uma grande variedade de doenças (CAPRILES et al., 2010), e um dos principais

interesses na pesquisa do genoma humano é descobrir se mutações específicas afetam a saúde humana (CALABRESE et al., 2009).

Análises in silico

Com o avanço das técnicas de biologia molecular no sequenciamento de genomas, tem sido produzido um grande volume de dados biológicos, implicando na necessidade da computação, através da bioinformática e seus métodos *in silico*, a fim de realizar-se uma análise mais completa desta informação, de forma a determinar a função e a estrutura das sequências proteicas que são obtidas via sequenciamento.

Existem vários métodos computacionais disponíveis para predizer quando uma mutação é relacionada à doença ou não, a partir da sequência da proteína e/ou alinhamentos múltiplos de sequências de proteínas (CALABRESE et al., 2009). A informação evolucionária codificada no perfil da sequência proteica é a parte mais importante para a melhoria do desempenho preditivo, como indicado pelos resultados de diversos preditores descritos na literatura (CALABRESE et al., 2009).

Neste estudo, foram obtidas 13 variantes naturais da VCP para a realização de análises *in silico*, que podem determinar se estas mutações influenciam na estrutura ou na estabilidade tridimensional da proteína. Os possíveis efeitos das variantes não sinônimas, que promovem mudanças de aminoácidos sobre a função das proteínas, foram inferidos a partir da utilização de ferramentas de bioinformática concebidos especificamente para estes tipos de interpretação, tais como PolyPhen-2 (ADZHUBEI; JORDAN; SUNYAEV, 2013). Devido à importância de se compreender se as variantes são ou não relacionadas a doenças, programas como o SnpEff

(DE BAETS et al., 2011), PhD-SNP (CAPRIOTTI; ALTMAN, 2011), PMUT (FERRER-COSTA et al., 2005), SIFT (SIM et al., 2012), SNAP (JOHNSON et al., 2008), SNPs & GO (CAPRIOTTI et al., 2013) e nsSNPAnalyzer (BAO; ZHOU; CUI, 2005) foram utilizados para prever se um determinado SNP afeta a função da proteína.

Este trabalho, através da utilização de ferramentas de bioinformática, tem como finalidade realizar a análise *in silico* das mutações da VCP relacionadas à Esclerose Lateral Amiotrófica, objetivando determinar se estas mutações afetam a estrutura, atividade ou estabilidade tridimensional da proteína. Dessa forma, a partir de nossos resultados, será possível obter uma maior compreensão das bases moleculares destas mutações e seus envoltórios com a patogênese da doença, além deles permitirem a investigação de possíveis vias que possam servir de alvo para um futuro tratamento da ALS.

Objetivo

O objetivo deste estudo é a análise *in silico* de 13 variantes naturais da proteína VCP descritas na literatura e nos bancos de dados de SNPs com a finalidade de determinar se estas mutações influenciam positiva ou negativamente a estrutura ou estabilidade tridimensional da proteína, e o desenvolvimento de um banco de dados online da VCP com nossos resultados.

2 Material e Métodos

Obtenção da Sequência Proteica

A sequência de aminoácidos da proteína VCP (UNIPROT ID: P55072) foi obtida a partir do banco de dados *Protein Knowledge Base* (THE UNIPROT CONSORTIUM, 2015).

Obtenção das Variantes Naturais da VCP

As 13 variantes naturais da proteína VCP humana que levam à Esclerose Lateral Amiotrófica foram obtidas dos bancos de dados OMIM (AMBERGER; BOCCHINI; HAMOSH, 2011) e UNIPROT (THE UNIPROT CONSORTIUM, 2015).

Seguindo a metodologia já estabelecida em nosso grupo (CARVALHO; DE MESQUITA, 2013; MOREIRA et al., 2013), a predição funcional das variantes relacionadas à ALS (esporádica e familiar), obtidas através de revisão da literatura, foi realizada a partir da utilização de ferramentas de bioinformática projetadas para qualificar enovelamentos incorretos da proteína, agregação, patogenicidade e interação com chaperonas. Ainda, a estrutura tridimensional das mutações da VCP envolvidas na ALS foi obtida a partir de modelagem computacional, dinâmica molecular e validação por alinhamentos tridimensionais.

Predição Funcional das Variantes da VCP Humana

Para predizer os possíveis efeitos das mutações na função da proteína VCP foram utilizados os programas: PolyPhen-2 (ADZHUBEI et al., 2010), SnpEff (DE BAETS et al., 2011), PhD-SNP (CAPRIOTTI; ALTMAN, 2011), PMUT (FERRER-COSTA et al., 2005), SIFT (SIM et al., 2012), SNAP (BROMBERG; YACHDAV; ROST, 2008), SNPs&GO (CAPRIOTTI et al., 2013), nsSNPAnalyzer (BAO; ZHOU; CUI, 2005), I-Mutant (<http://folding.uib.es/i-mutant/i-mutant2.0.html>), VarMod (PAPPALARDO; WASS, 2014) e MutPred (LI et al., 2009).

Predição Estrutural da VCP humana selvagem e de suas variantes

A estrutura tridimensional da proteína VCP humana selvagem foi determinada computacionalmente através de modelagem comparativa e *ab initio*. Na modelagem comparativa, são empregadas as coordenadas de estruturas tridimensionais determinadas experimentalmente por cristalografia de raios-X ou ressonância magnética nuclear para a construção do modelo teórico. Em contrapartida, a modelagem *ab initio* cria modelos teóricos a partir dos parâmetros termodinâmicos da proteína.

Para tanto, foram usados os seguintes algoritmos de modelagem comparativa: Mholline (CAPRILES et al., 2010), Swiss Model (BIASINI et al., 2014), HHpred (ROSE et al., 2009), IntFOLD (ROCHE et al., 2011), Phyre2 (KELLEY et al., 2015), RaptorX (KÄLLBERG et al., 2014), e M4T (RYKUNOV et al., 2009). E dois algoritmos de modelagem *ab initio*: I-Tasser (YANG et al., 2014) e Rosetta (SONG et al., 2013).

Visando à determinação do melhor modelo a ser utilizado para a etapa de Dinâmica Molecular, foram realizados alinhamentos estruturais de cada um dos modelos obtidos por modelagem com três fragmentos de estruturas experimentais da VCP contidos no banco de dados Protein Data Bank (WESTBROOK et al., 2002), sendo dois deles de *Mus musculus* (PDB ID: 1R7R (HUYTON et al., 2003) e 3CF3 (DAVIES; BRUNGER; WEIS, 2008)) e um humano, PDB ID: 3QQ7 (HÄNZELMANN; BUCHBERGER; SCHINDELIN, 2011). Com o intuito de se investigar a similaridade estrutural, foi utilizado o algoritmo de alinhamento tridimensional TM-align (ZHANG; SKOLNICK, 2005). Os valores de RMSD (do inglês, “Root Mean Square Deviation”) e TM-score, que indicam a similaridade estrutural entre duas estruturas proteicas, fornecidos pelo TM-align foram os critérios utilizados para definir o modelo a ser empregado para criar as estruturas mutantes por simulação de dinâmica molecular.

Após a definição do modelo a ser utilizado, foi realizada a simulação de dinâmica molecular do mesmo através do pacote GROMACS 4.5 (PRONK et al., 2013), disponível no servidor MDWeb (HOSPITAL et al., 2012). Ainda através deste mesmo algoritmo, por mutagênese *in silico*, foram obtidas as estruturas tridimensionais das 13 variantes. Cada uma das estruturas foi então alinhada à estrutura selvagem após a dinâmica, pelo TM-align, a fim de se comparar as alterações estruturais entre as mutantes e a proteína nativa.

Análise da Conservação Evolucionária

A análise da conservação evolucionária dos resíduos que compõem a VCP foi realizada pelo algoritmo ConSurf (ASHKENAZY et al., 2010). Este algoritmo calcula o escore de conservação evolucionária, através do método Bayesiano (por inferência estatística), para cada uma das posições e gera um mapa colorimétrico da estrutura proteica com seus resultados (ASHKENAZY et al., 2010)(Landau, Mayrose et al. 2005). A estrutura proteica submetida para a construção do mapa pelo algoritmo foi o melhor modelo gerado na etapa de modelagem molecular.

Criação do Banco de Dados da VCP

Foi desenvolvido um banco de dados manualmente curado a partir da utilização de um “server side include” em Java, o JMOL. A base de dados no <http://bioinfogroup.com/database> contém todos os resultados obtidos neste trabalho, tendo sido os mesmos obtidos individual e manualmente em cada algoritmo citado.

3 Resultados e Discussão

A motivação para a escolha da proteína VCP é a relação de suas variantes com o desenvolvimento de fELA e o número crescente de estudos que têm sido realizados para uma maior compreensão da doença e dos algoritmos de predição de SNPs.

Neste trabalho, foram analisadas 13 variantes naturais da proteína VCP obtidas através de dois bancos de dados diferentes (OMIM e Uniprot). Até o momento, nenhum banco de dados continha um conjunto completo de análises *in silico* por mutação, revelando a falta de um banco de dados mais atualizado e completo sobre a VCP. O banco de dados da VCP, <http://bioinfogroup.com/database>, foi criado para suprir estas demandas e formar uma base de dados dos resultados encontrados durante a execução deste projeto e de outros no futuro, permitindo que

pesquisadores e leigos possam explorar a VCP e suas variantes. O usuário pode conhecer e analisar os algoritmos que são mais frequentemente utilizados para prever os efeitos dos SNPs, a partir do acesso aos resultados individuais de aplicação de cada uma das 13 mutações aqui descritas.

Através de revisão da literatura foram obtidas 13 mutações não sinônimas da VCP que foram descritas em pacientes diagnosticados com a ALS esporádica ou familiar (seguindo o padrão de representação aminoácido selvagem/posição/aminoácido mutante): R95G (MEHTA et al., 2013), R155C (MEHTA et al., 2013), R155H (JOHNSON et al., 2010; MEHTA et al., 2013), R159G (JOHNSON et al., 2010), R159H (KOPPERS et al., 2012), R191G (GONZÁLEZ-PÉREZ et al., 2012), R191Q (JOHNSON et al., 2010), A232E (FERRER-COSTA; OROZCO; DE LA CRUZ, 2002), D592N (JOHNSON et al., 2010), R159C (ABRAMZON et al., 2012), N387T (ABRAMZON et al., 2012), R662C (ABRAMZON et al., 2012), I114V (KOPPERS et al., 2012).

Predição Funcional das Variantes da VCP Humana

Em relação aos algoritmos que qualificam as mutações como patogênicas ou neutras, foram obtidos os seguintes resultados (ilustrados graficamente nas Figuras 3 e 4):

- nsSNPAnalyzer: Classificou 4 mutações como patogênicas (R159G, R191G, R159C e R662C) e as demais como neutras;
- PhD-SNP: Classificou 3 mutações como neutras (R95G, R191Q e I114V) e as restantes como patogênicas;
- Pmut: Classificou 2 mutações como neutras (N387T e I114V) e as demais como patogênicas;
- Polyphen-2: Classificou 4 mutações como neutras (R159H, A232E, N387T e I114V) e as demais como patogênicas;
- SIFT: Classificou 6 mutações como neutras (R95G, R159G, R159H, A232E, D592N e I114V) e as demais como patogênicas;

- SNAP: Classificou 6 mutações como neutras (R95G, R155H, R159H, A232E, N387T e I114V) e as restantes como patogênicas;
- SNPs&GO: Classificou 2 mutações como neutras (A232E e I114V) e as demais como patogênicas.

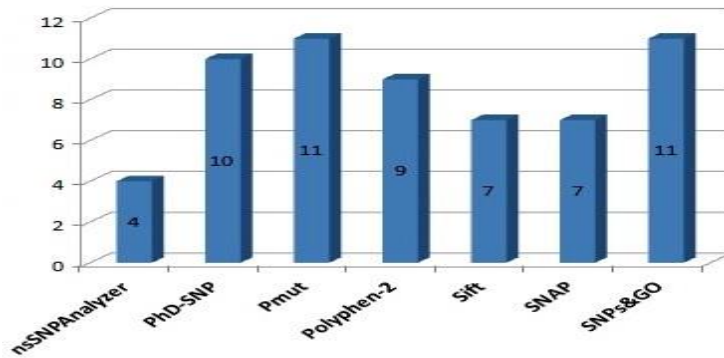


Figura 3 - Número de mutações da VCP classificadas como deletérias pelos algoritmos de predição. As barras azuis indicam o total de mutações classificadas como patogênicas (eixo y) em relação aos algoritmos utilizados (eixo x).

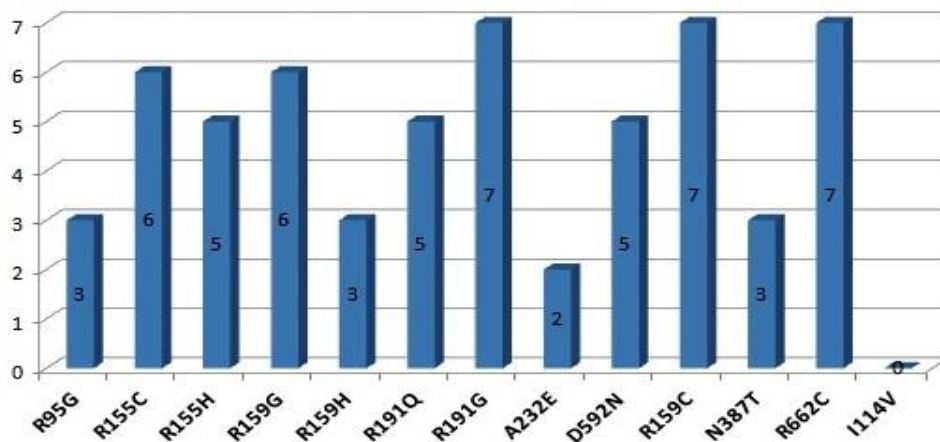


Figura 4 - Número de classificações como deletérias por mutação da VCP. As barras azuis indicam o total de algoritmos que classificaram as mutações da VCP como patogênicas (eixo y) por mutação (eixo x).

A partir da análise dos resultados e sabendo-se que se trata de mutações já classificadas como patogênicas, pode-se verificar que os algoritmos Pmut e SNPs&GO foram os que apresentaram maior acurácia na classificação de variantes da VCP. Em contrapartida, o nsSNPAnalyzer foi o que apresentou a menor acurácia.

O SNPeffect contém em sua composição 4 algoritmos que fazem diferentes análises: TANGO – Analisa a tendência à agregação (FERNANDEZ-ESCAMILLA et

al., 2004), WALTZ – Avalia a propensão à formação amiloide (MAURER-STROH et al., 2010), LIMBO – Analisa a tendência de ligação à chaperonas (DE BAETS et al., 2011) e FoldX – Avalia alterações na estabilidade da estrutura proteica (FERNANDEZ-ESCAMILLA et al., 2004). De acordo com as análises destes algoritmos, as mutações R159G, R159H e R159C foram classificadas como responsáveis por diminuir a tendência à agregação; a mutação R155C por aumentar a tendência à agregação; a mutação N387T por diminuir a tendência de ligação à chaperonas; e, por fim, as mutações R95G, R191G e I114V por reduzirem a estabilidade proteica. Nenhuma das mutações foi classificada como responsável por causar alterações na propensão amiloide.

O algoritmo I-Mutant classifica as mutações quanto às alterações de estabilidade proteica em relação à proteína nativa, assim como o FoldX. Sua avaliação classificou todas as mutantes como responsáveis por causar alterações de estabilidade na proteína, de forma que as mutantes A232E e N387T foram classificadas como mais estáveis do que a proteína nativa, e as demais como menos estáveis.

O algoritmo VarMod realiza uma série de predições acerca das variantes submetidas. Seus resultados (Figura 5) seguem um padrão colorimétrico onde azul indica que a variante provavelmente não afeta a função proteica e vermelho indica o oposto. O item "VarMod Pred" indica a predição global realizada pelo VarMod; o "Interface" sugere que a variante encontra-se em um local de interface proteína-proteína ou não; o "Binding site", se a variação encontra-se num sítio de ligação da proteína; o item "Conservation", o grau de conservação da posição da mutação; o "size", se há uma diferença entre as cadeias laterais dos dois aminoácidos (mutante/nativo); o "Charge", se há diferença entre as cargas dos aminoácidos (mutante/nativo); e o "Functional group", se há diferença entre os grupamentos químicos funcionais dos aminoácidos (mutante/nativo) (Pappalardo and Wass 2014).

Através da análise dos resultados da Figura 5, observamos que as mutações que possuíram maior escore geral do VarMod foram a R662C e a R159C, respectivamente, sendo portanto estas as mutações consideradas como responsáveis pelas maiores alterações na função da VCP.

Variant	VarMod	Probability	Interface	Binding site	Conservation	size	Charge	Func grp
R95G	0.588	1.000	-	0.557				
I114V	0.221	1.000	-	0.515				
R155C	0.677	0.000	-	0.505				
R155H	0.214	0.000	-	0.505				
R159H	0.301	1.000	-	0.435				
R159C	0.811	1.000	-	0.435				
R159G	0.563	1.000	-	0.435				
R191G	0.697	1.000	-	0.609				
R191Q	0.672	1.000	-	0.609				
A232E	0.546	1.000	-	0.632				
N387T	0.698	1.000	-	0.872				
D592N	0.632	1.000	-	0.505				
R662C	0.870	1.000	-	0.669				

Figura 5 - Resultados obtidos através da análise do algoritmo VarMod

O MutPred também analisa alterações nas características estruturais e funcionais entre as sequências nativa e mutante na forma de probabilidades de ganho ou perda de estrutura e função. Seu resultado fornece um escore geral (g-score), que é a probabilidade de uma mutação ser associada à doença (Li, Krishnan et al. 2009). Um g-score maior do que 0,5 sugere que a mutação seja deletéria e um g-score maior do que 0,75 indica que a proteína provavelmente é deletéria com alto grau de confiabilidade. Em nossas análises, as variantes R95G, R155C, R155H, R159G, R159H, A232E, R159C, N387T, R662C e I114V apresentaram resultados de g-score entre 0,5 e 0,75, sugerindo que essas provavelmente são variantes deletérias. Enquanto as demais mutações apresentaram resultados de g-score inferiores a 0,5.

a. Predição Estrutural da VCP humana selvagem e de suas variantes

Para elaboração do modelo estrutural teórico da proteína VCP selvagem, foram realizadas técnicas de modelagem comparativa e *ab initio*. Após a modelagem, os modelos gerados foram alinhados, pelo algoritmo TM-align, um a um com fragmentos de estruturas experimentais da VCP obtidas do *Protein Data Bank* (duas murinas e uma humana, respectivamente 3CF3, 1R7R e 3QQ7). O resultado da modelagem/alinhamento com o fragmento humano realizado será apresentado

separadamente por algoritmo a seguir e de forma resumida com os valores de TM-score (escore de similaridade estrutural fornecido pelo TM-align, quanto mais próximo de 1, maior a similaridade e, quanto mais próximo de 0, menor) e RMSD (distância interatômica, em angstroms, entre duas estruturas proteicas) na Tabela 1.

Mholline: O modelo gerado por modelagem comparativa possui 806 resíduos de aminoácido (número de aminoácidos da estrutura completa da VCP), mas apresentou um grande fragmento não resolvido. A representação de seu alinhamento com o fragmento da estrutura da VCP humana (PDBID: 3QQ7), obtido do *Protein Data Bank*, encontra-se na Figura 6. Em vermelho, observa-se o modelo e, em azul, a estrutura 3QQ7.

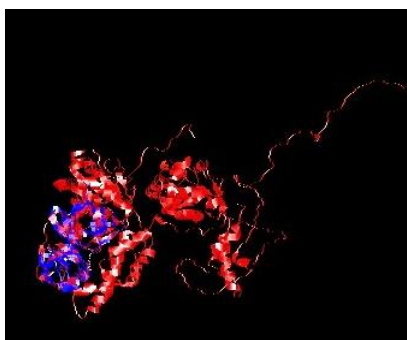


Figura 6 - Alinhamento tridimensional entre o modelo da VCP selvagem (wt) gerado pelo Mholline e a estrutura de PDBID 3QQ7.

Em vermelho, observa-se o modelo e, em azul, a estrutura 3QQ7

SwissModel: O modelo gerado por modelagem comparativa possui apenas 743 resíduos de aminoácido, referentes às posições 21 à 763. A representação de seu alinhamento com o fragmento da estrutura da VCP humana (PDBID: 3QQ7), obtido do *Protein Data Bank*, encontra-se na Figura 7. Em vermelho, observa-se o modelo e, em azul, a estrutura 3QQ7.

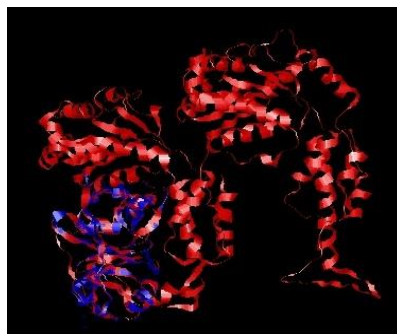


Figura 7 - Alinhamento tridimensional entre o modelo da VCP wt gerado pelo SwissModel e a estrutura de PDBID 3QQ7.

Em vermelho, observa-se o modelo e, em azul, a estrutura 3QQ7

HHPred: O modelo gerado por modelagem comparativa possui 806 resíduos de aminoácido, mas é possível observar grandes fragmentos não resolvidos em sua estrutura em ambas as extremidades. A representação de seu alinhamento com o fragmento da estrutura da VCP humana (PDBID: 3QQ7), obtido do *Protein Data Bank*, encontra-se na Figura 8. Em vermelho, observa-se o modelo e, em azul, a estrutura 3QQ7.

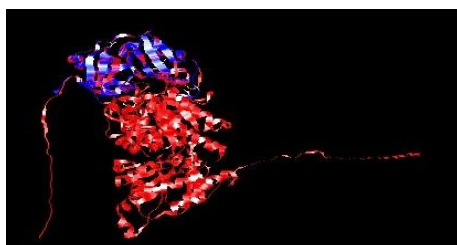


Figura 8-Alinhamento tridimensional entre o modelo da VCP wt gerado pelo HHPred e a estrutura de PDBID 3QQ7.Em vermelho, observa-se o modelo e, em azul, a estrutura 3QQ7.

IntFOLDTS: O modelo gerado por modelagem comparativa também possui 806 resíduos de aminoácido, mas é possível observar grandes fragmentos não resolvidos em sua estrutura. A representação de seu alinhamento com o fragmento da estrutura da VCP humana (PDBID: 3QQ7), obtido do *Protein Data Bank*, encontra-se na Figura 9. Em vermelho, observa-se o modelo e, em azul, a estrutura 3QQ7.

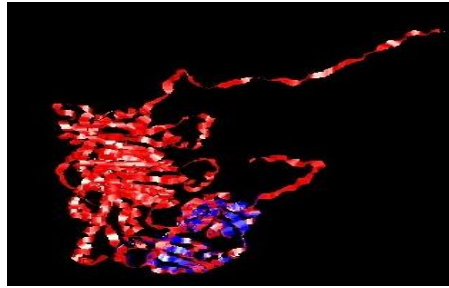


Figura 9 - Alinhamento tridimensional entre o modelo da VCP wt gerado pelo IntFOLDTS e a estrutura de PDBID 3QQ7.

Em vermelho, observa-se o modelo e, em azul, a estrutura 3QQ7.

Phyre 2: O modelo gerado por modelagem comparativa também possui 806 resíduos de aminoácido. A representação de seu alinhamento com o fragmento da estrutura da VCP humana (PDBID: 3QQ7), obtido do *Protein Data Bank*, encontra-se na Figura 10. Em vermelho, observa-se o modelo e, em azul, a estrutura 3QQ7.

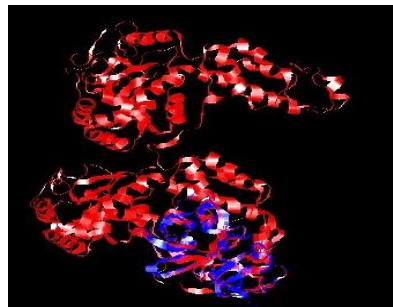


Figura 10 - Alinhamento tridimensional entre o modelo da VCP wt gerado pelo Phyre 2 e a estrutura de PDBID 3QQ7.

Em vermelho, observa-se o modelo e, em azul, a estrutura 3QQ7

RaptorX: O modelo gerado por modelagem comparativa possui 778 resíduos de aminoácido, referentes às posições 1 à 778. A representação de seu alinhamento com o fragmento da estrutura da VCP humana (PDBID: 3QQ7), obtido do *Protein Data Bank*, encontra-se na Figura 11. Em vermelho, observa-se o modelo e, em azul, a estrutura 3QQ7.

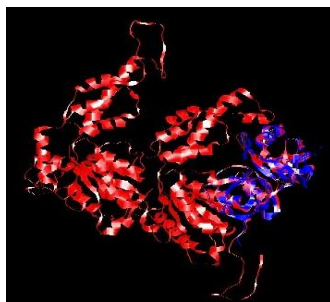


Figura 11 - Alinhamento tridimensional entre o modelo da VCP wt gerado pelo RaptorX e a estrutura de PDBID 3QQ7.

Em vermelho, observa-se o modelo e, em azul, a estrutura 3QQ7.

I-tasser: Os cinco modelos gerados por modelagem *ab initio* para a proteína VCP Humana pelo I-tasser possuem 806 aminoácidos cada um. A representação do alinhamento do modelo 1 (melhor modelo dentre os cinco) com o fragmento da estrutura da VCP humana (PDBID: 3QQ7), obtido do *Protein Data Bank*, encontra-se na Figura 12. Em vermelho, observa-se o modelo e, em azul, a estrutura 3QQ7.

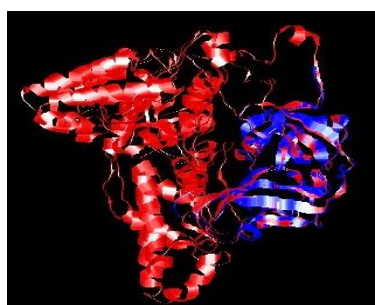


Figura 12 - Alinhamento tridimensional entre o modelo 1 da VCP wt gerado pelo I-tasser e a estrutura de PDBID 3QQ7

Alinhamento tridimensional entre o modelo 1 da VCP wt gerado pelo I-tasser e a estrutura de PDBID 3QQ7

Rosetta: Assim como o Itasser, o Rosetta gerou 5 modelos para a VCP humana por modelagem *ab initio*. Entretanto, o melhor modelo gerado (modelo 1) possui apenas 367 resíduos de aminoácido, referentes às posições 1 à 367. A representação de seu alinhamento com o fragmento da estrutura da VCP humana (PDBID: 3QQ7), obtido do *Protein Data Bank*, encontra-se na Figura 13. Em vermelho, observa-se o modelo e, em azul, a estrutura 3QQ7.

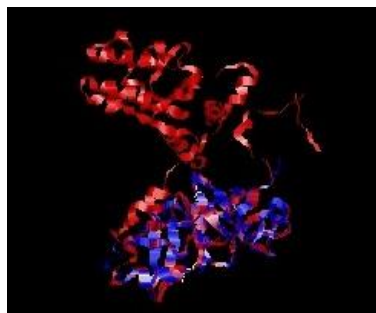


Figura 13 - Alinhamento tridimensional entre o modelo 1 da VCP wt gerado pelo Rosetta e a estrutura de PDBID 3QQ7

Em vermelho, observa-se o modelo e, em azul, a estrutura 3QQ7

A tabela a seguir traz os resultados quantitativos para os alinhamentos estruturais realizados pelo TM-align. A primeira coluna refere-se às estruturas que foram alinhadas; a segunda e terceira coluna aos valores de TM-score obtidos (quanto mais próximo de 1, mais similar são as estruturas); a quarta coluna aos valores de RMSD (quanto mais próximo de 0, mais similares são as estruturas); e, por fim, a quinta coluna refere-se ao número de aminoácidos alinhados.

Tabela 1 - Resultados obtidos através dos alinhamentos estruturais - dos modelos obtidos vs estruturas experimentais - realizados pelo TM-align.

Modelos	TM-score (Normalizado pela cadeia 1)	TM-score (Normalizado pela cadeia 2)	RMSD (Å)	Comprimento Alinhado
Murina 1R7R + Modelo VCPwt 1 I-tasser	0.93515	0.79753	2.55	679
Murina 1R7R + Modelo VCPwt Mholline	0.90920	0.77625	3.19	674
Murina 1R7R + Modelo VCPwt Raptor X	0.92806	0.81946	2.69	679
Murina 1R7R + Modelo VCPwt Hhpred	0.52703	0.45719	6.32	480
Murina 1R7R + Modelo VCPwt Phyre2	0.93994	0.80139	2.60	682

Murina 1R7R + Modelo VCPwt Intfoldts	0.93170	0.79492	2.70	680
Murina 1R7R + Modelo VCPwt 1 Rosetta	0.31744	0.54430	5.29	270
Murina 1R7R + Modelo VCPwt SwissModel	0.92270	0.85139	2.70	675
Homo sapiens 3QQ7 + Modelo VCPwt 1 I-tasser	0.95573	0.20520	1.36	168
Homo sapiens 3QQ7 + Modelo VCPwt Mholline	0.95605	0.20499	1.45	168
Homo sapiens 3QQ7 + Modelo VCP wt 1 Rosetta	0.95853	0.44597	1.44	168
Homo sapiens 3QQ7 + Modelo VCPwt RaptorX	0.95290	0.21234	1.39	168
Homo sapiens 3QQ7 + Modelo VCPwt IntFOLDTS	0.95438	0.20510	1.38	168
Homo sapiens 3QQ7 + Modelo VCPwt phyre2	0.95899	0.20527	1.37	168
Homo sapiens 3QQ7 + Modelo VCPwt Hhpred	0.96107	0.20533	1.39	168
Homo sapiens 3QQ7 + Modelo VCPwt SwissModel	0.95668	0.22153	1.29	167
Murina 3CF3 + Modelo VCPwt Mholline	0.85696	0.77445	3.91	699
Murina 3CF3 + Modelo VCPwt 1 I-tasser	0.98681	0.88615	1.08	723
Murina 3CF3 + Modelo VCPwt Raptor X	0.98777	0.91856	1.05	723
Murina 3CF3 + Modelo VCPwt Hhpred	0.49107	0.44598	6.15	453
Murina 3CF3 + Modelo VCPwt Phyre2	0.89834	0.81231	3.21	722
Murina 3CF3 + Modelo VCPwt Intfoldts	0.97904	0.87968	1.43	723
Murina 3CF3 + Modelo	0.30093	0.54177	5.24	268

VCPwt 1 Rosetta				
Murina 3CF3 + Modelo	0.99520	0.96852	0.64	723
VCPwt SwissModel				

A partir dos dados representados na Tabela 1, pela comparação dos valores de RMSD e TM-score, e das análises estruturais individuais, foi possível definir o melhor modelo teórico-computacional da VCP Humana: Modelo 1 do I-tasser. Este modelo foi então utilizado para a simulação de dinâmica molecular através do algoritmo GROMACS e criação das 13 variantes da VCP por mutagênese *in silico*, no servidor MDWeb.

Após esta etapa, as mutações geradas foram alinhadas uma a uma, através do TM-align, com a VCP selvagem resultante da dinâmica molecular do modelo 1 do I-tasser. Os resultados destes alinhamentos encontram-se na Tabela 2 e seguem o mesmo padrão da tabela 1.

Tabela 2 - Resultados obtidos através dos alinhamentos estruturais - VCP wt pós dinâmica vs mutantes - realizados pelo TM-align.

Modelos	TM-score (Normalizado pela cadeia 1)	TM-score (Normalizado pela cadeia 2)	RMSD (Å)	Comprimento Alinhado
VCP wt (Model 1 Itasser) + VCP R95G	0.98786	0.98786	1.08	806
VCP wt (Model 1 Itasser) + VCP R155C	0.98740	0.98740	1.10	806
VCP wt (Model 1 Itasser) + VCP R155H	0.98654	0.98654	1.14	806
VCP wt (Model 1 Itasser) + VCP R159G	0.98844	0.98844	1.05	806
VCP wt (Model 1 Itasser) + VCP R159H	0.98759	0.98759	1.09	806
VCP wt (Model 1 Itasser) + VCP R191G	0.98865	0.98865	1.04	806
VCP wt (Model 1 Itasser) + VCP R191Q	0.98694	0.98694	1.12	806
VCP wt (Model 1 Itasser) + VCP A232E	0.98782	0.98782	1.08	806
VCP wt (Model 1 Itasser) + VCP D592N	0.98734	0.98734	1.10	806
VCP wt (Model 1 Itasser) + VCP R159C	0.98714	0.98714	1.11	806
VCP wt (Model 1 Itasser) + VCP N387T	0.98847	0.98847	1.05	806
VCP wt (Model 1 Itasser) + VCP	0.98803	0.98803	1.07	806

R662C				
VCP wt (Model 1 Itasser) + VCP I114V	0.98800	0.98800	1.07	806

Tendo em vista os resultados da tabela 2, é possível concluir que a mutante que sofreu mais alterações estruturais em relação à VCP nativa foi a R191Q (alinhamento representado na Figura 14). Enquanto que a variante que apresentou menos alterações foi a R191G (alinhamento representado na Figura 14).

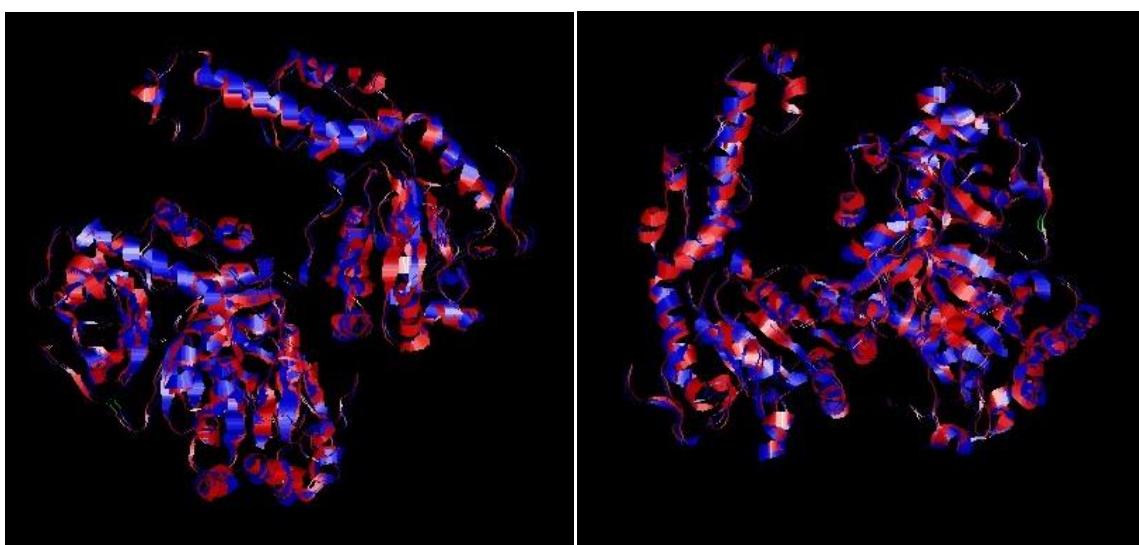


Figura 14 - Alinhamento Estrutural entre a VCP nativa e as mutações R191Q (Direita) e R191G (Esquerda)

A posição 191 encontra-se realçada em verde.

c. Análise da Conservação Evolucionária

Por fim, os resultados do algoritmo Consurf (Figura 15) baseiam-se em uma representação tridimensional da estrutura proteica indicando o grau de conservação de cada aminoácido através de um padrão colorimétrico. Nesta pontuação de conservação colorimétrica, resíduos muito conservados são representados em bordô, enquanto os pouco conservados apresentam-se em azul turquesa (Ashkenazy, Erez et al. 2010). Para que o Consurf pudesse fornecer a estrutura tridimensional, foi submetida a melhor estrutura obtida por modelagem molecular (Modelo 1 do algoritmo Itasser).

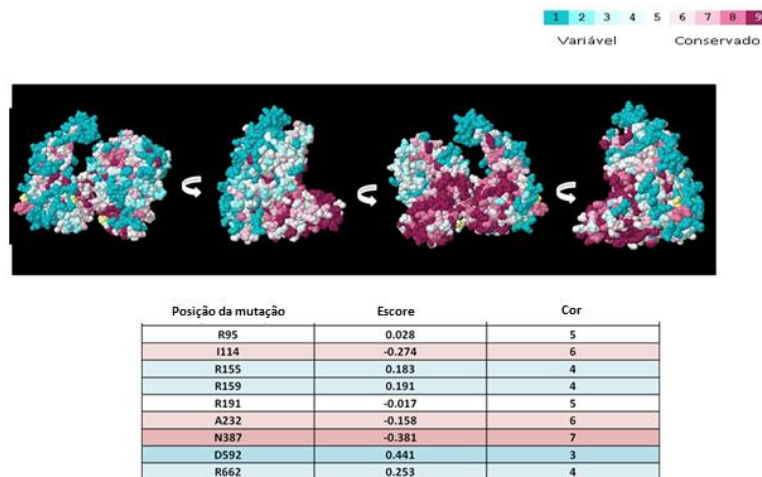


Figura 15 - Resultados obtidos pelo algoritmo ConSurf

Na figura, observa-se a estrutura do Modelo VCP 1 gerado pelo I-tasser com seus aminoácidos representados segundo o padrão colorimétrico de conservação. Na tabela inferior, encontram-se as pontuações de conservação e o escore do ConSurf para cada posição em que existe mutações relacionadas à ALS.

A partir da análise dos resultados do ConSurf, observa-se que as mutações que ocorrem em locais mais conservados são a I114V, A232E e N387T. Enquanto as variantes localizadas nas posições 155, 159, 592 e 662 são as que ocorrem nos aminoácidos menos conservados. Tendo em vista que a conservação estrutural de resíduos de aminoácidos é diretamente relacionada à função que estes exercem na proteína, as mutações que ocorrem nas posições mais conservadas tendem a ser mais prejudiciais à função proteica do que as variações que ocorrem em posições mais variáveis.

4. Conclusão

Todas as variantes relacionadas à ALS foram classificadas como patogênicas por pelo menos dois algoritmos de predição de patogenicidade. As mutantes R662C e R159C foram classificadas como deletérias por todos os algoritmos utilizados e os algoritmos que apresentaram maior acurácia da análise de patogenicidade dessas variantes foram o Pmut e o SNPs&GO. Portanto, caso surjam novas mutações da VCP não descritas na literatura, estes dois últimos algoritmos são os mais recomendados para análise. Ainda, de acordo com os resultados obtidos pelo SNPeffect, as mutações R159G, R159H e R159C foram classificadas como responsáveis por diminuir a tendência à agregação; a mutação R155C por aumentar a tendência à agregação; a mutação N387T por diminuir a tendência de ligação à chaperonas; e, por fim, as mutações R95G, R191G e I114V por reduzirem a estabilidade proteica. Diferentemente deste último resultado, o algoritmo I-mutant classificou todas as variantes como responsáveis por alterar a estabilidade proteica.

No que concerne à análise de conservação estrutural, as variantes que ocorrem em posição mais conservada são as I114V, A232E e N387T. Enquanto as variantes localizadas nas posições 155, 159, 592 e 662 são as que ocorrem nos aminoácidos menos conservados.

Em relação à análise estrutural, o melhor modelo gerado computacionalmente para a VCP selvagem foi obtido do I-tasser, tendo sido esse utilizado para criar as variantes por mutagênese *in silico*. Dentre as variantes estudadas, a que possui maior diferença estrutural em relação à VCP selvagem foi a R191Q e a que possui menor diferença foi a R191G.

5. Referências Bibliográficas

- ABRAMZON, Y. et al. Valosin-containing protein (VCP) mutations in sporadic amyotrophic lateral sclerosis. **Neurobiology of aging**, v. 33, n. 9, p. 2231.e1–2231.e6, set. 2012.
- ADZHUBEI, I. A. et al. A method and server for predicting damaging missense mutations. **Nature methods**, v. 7, n. 4, p. 248–9, abr. 2010.
- ADZHUBEI, I.; JORDAN, D. M.; SUNYAEV, S. R. Predicting functional effect of human missense mutations using PolyPhen-2. **Current Protocols in Human Genetics**, v. Chapter 7, n. SUPPL.76, 2013.
- AL-CHALABI, A.; LEWIS, C. M. Modelling the effects of penetrance and family size on rates of sporadic and familial disease. **Human heredity**, v. 71, n. 4, p. 281–8, jan. 2011.
- ALKAN, C.; COE, B. P.; EICHLER, E. E. **Genome structural variation discovery and genotyping.**, 2011. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/21358748>>
- AMBERGER, J.; BOCCHINI, C.; HAMOSH, A. A new face and new challenges for Online Mendelian Inheritance in Man (OMIM®). **Human mutation**, v. 32, n. 5, p. 564–567, 2011.
- ASHKENAZY, H. et al. **ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids.** **Nucleic acids research**, 2010. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/20478830>>
- BAO, L.; ZHOU, M.; CUI, Y. nsSNPAnalyzer: identifying disease-associated nonsynonymous single nucleotide polymorphisms. **Nucleic acids research**, v. 33, n. Web Server issue, p. W480–2, jul. 2005.
- BEGHI, E. et al. The epidemiology of ALS and the role of population-based registries. **Biochimica et biophysica acta**, v. 1762, n. 11-12, p. 1150–7, jan. 2006.
- BIASINI, M. et al. SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. **Nucleic acids research**, v. 42, n. Web Server issue, p. W252–8, 1 jul. 2014.
- BROMBERG, Y.; YACHDAV, G.; ROST, B. SNAP predicts effect of mutations on protein function. **Bioinformatics**, v. 24, n. 20, p. 2397–2398, 2008.

CALABRESE, R. et al. Functional annotations improve the predictive score of human disease-related mutations in proteins. **Human mutation**, v. 30, n. 8, p. 1237–44, ago. 2009.

CAPRILES, P. V. S. Z. et al. 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. 610, jan. 2010.

CAPRIOTTI, E. et al. WS-SNPs&GO: a web server for predicting the deleterious effect of human protein variants using functional annotation. **BMC genomics**, v. 14 Suppl 3, n. Suppl 3, p. S6, 2013.

CAPRIOTTI, E.; ALTMAN, R. B. **Improving the prediction of disease-related variants using protein three-dimensional structure.****BMC bioinformatics**, 2011.

Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/21992054>>

CARVALHO, M. D. C.; DE MESQUITA, J. F. Structural Modelling and In Silico Analysis of Human Superoxide Dismutase 2. **PLOS ONE**, p. in press, 2013.

CHAPMAN, E.; FRY, A. N.; KANG, M. The complexities of p97 function in health and disease. **Molecular bioSystems**, v. 7, n. 3, p. 700–10, mar. 2011.

DAVIES, J. M.; BRUNGER, A. T.; WEIS, W. I. Improved structures of full-length p97, an AAA ATPase: implications for mechanisms of nucleotide-dependent conformational change. **Structure (London, England : 1993)**, v. 16, n. 5, p. 715–26, maio 2008.

DE BAETS, G. et al. **SNPeffect 4.0: on-line prediction of molecular and structural effects of protein-coding variants.****Nucleic acids research**, 2011.

Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/22075996>>

FERNANDEZ-ESCAMILLA, A.-M. et al. Prediction of sequence-dependent and mutational effects on the aggregation of peptides and proteins. **Nat Biotech**, v. 22, n. 10, p. 1302–1306, 2004.

FERRER-COSTA, C. et al. **PMUT: a web-based tool for the annotation of pathological mutations on proteins****Bioinformatics**, 2005. Disponível em:

<<http://www.ncbi.nlm.nih.gov/pubmed/15879453>>

FERRER-COSTA, C.; OROZCO, M.; DE LA CRUZ, X. Characterization of disease-associated single amino acid polymorphisms in terms of sequence and structure properties. **Journal of molecular biology**, v. 315, n. 4, p. 771–86, jan. 2002.

GONZÁLEZ-PÉREZ, P. et al. Novel mutation in VCP gene causes atypical

amyotrophic lateral sclerosis. **Neurology**, v. 79, n. 22, p. 2201–2208, 2012.

GORDON, P. H. Amyotrophic Lateral Sclerosis: An update for 2013 Clinical Features, Pathophysiology, Management and Therapeutic Trials. **Aging and disease**, v. 4, n. 5, p. 295–310, jan. 2013.

HÄNZELMANN, P.; BUCHBERGER, A.; SCHINDELIN, H. Hierarchical binding of cofactors to the AAA ATPase p97. **Structure (London, England : 1993)**, v. 19, n. 6, p. 833–43, 8 jun. 2011.

HOSPITAL, A. et al. MDWeb and MDMoby: an integrated web-based platform for molecular dynamics simulations. **Bioinformatics (Oxford, England)**, v. 28, n. 9, p. 1278–9, 1 maio 2012.

HUYTON, T. et al. The crystal structure of murine p97/VCP at 3.6Å. **Journal of structural biology**, v. 144, n. 3, p. 337–48, dez. 2003.

IGUCHI, Y. et al. Amyotrophic lateral sclerosis: an update on recent genetic insights. **Journal of neurology**, v. 260, n. 11, p. 2917–27, nov. 2013.

INGRE, C. et al. Risk factors for amyotrophic lateral sclerosis. **Clinical epidemiology**, v. 7, p. 181–93, jan. 2015.

ISHIKAWA, H. et al. Increased expression of valosin-containing protein in the skin of patients with amyotrophic lateral sclerosis. **Journal of clinical neuroscience : official journal of the Neurosurgical Society of Australasia**, v. 19, n. 4, p. 522–6, abr. 2012.

JOHNSON, A. D. et al. SNAP: A web-based tool for identification and annotation of proxy SNPs using HapMap. **Bioinformatics**, v. 24, n. 24, p. 2938–2939, 2008.

JOHNSON, G. C.; TODD, J. A. Strategies in complex disease mapping. **Current opinion in genetics & development**, v. 10, n. 3, p. 330–4, jun. 2000.

JOHNSON, J. O. et al. Exome sequencing reveals VCP mutations as a cause of familial ALS. **Neuron**, v. 68, n. 5, p. 857–64, 9 dez. 2010.

KÄLLBERG, M. et al. RaptorX server: a resource for template-based protein structure modeling. **Methods in molecular biology (Clifton, N.J.)**, v. 1137, p. 17–27, jan. 2014.

KELLEY, L. A. et al. The Phyre2 web portal for protein modeling, prediction and analysis. **Nature protocols**, v. 10, n. 6, p. 845–858, jun. 2015.

KLOPPSTECK, P. et al. Regulation of p97 in the ubiquitin-proteasome system by the UBX protein-family. **Biochimica et biophysica acta**, v. 1823, n. 1, p. 125–9, jan. 2012.

KOPPERS, M. et al. VCP mutations in familial and sporadic amyotrophic lateral sclerosis. **Neurobiology of aging**, v. 33, n. 4, 2012.

LI, B. et al. Automated inference of molecular mechanisms of disease from amino acid substitutions. **Bioinformatics (Oxford, England)**, v. 25, n. 21, p. 2744–50, 1 nov. 2009.

MANCUSO, R.; NAVARRO, X. Amyotrophic lateral sclerosis: Current perspectives from basic research to the clinic. **Progress in neurobiology**, v. 133, p. 1–26, out. 2015.

MAURER-STROH, S. et al. Exploring the sequence determinants of amyloid structure using position-specific scoring matrices. **Nature methods**, v. 7, n. 3, p. 237–242, 2010.

MEHTA, S. G. et al. Genotype-phenotype studies of VCP-associated inclusion body myopathy with Paget disease of bone and/or frontotemporal dementia. **Clinical genetics**, v. 83, n. 5, p. 422–31, maio 2013.

MOREIRA, L. G. A. et al. Structural and functional analysis of human SOD1 in amyotrophic lateral sclerosis. **PloS one**, v. 8, n. 12, p. e81979, 2 jan. 2013.

MOURA, M. C.; CASULARI, L. A.; CARVALHO GARBI NOVAES, M. R. Ethnic and demographic incidence of amyotrophic lateral sclerosis (ALS) in Brazil: A population based study. **Amyotrophic lateral sclerosis & frontotemporal degeneration**, p. 1–7, 5 fev. 2016.

NG, P.; HENIKOFF, S. **Predicting deleterious amino acid substitutions. Genome Research**, 2001. Disponível em:
<<http://genome.cshlp.org/content/11/5/863.full.pdf+html>>

PAPPALARDO, M.; WASS, M. N. VarMod: modelling the functional effects of non-synonymous variants. **Nucleic Acids Research**, 6 jun. 2014.

PRONK, S. et al. GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. **Bioinformatics (Oxford, England)**, v. 29, n. 7, p. 845–54, 1 abr. 2013.

RENTON, A. E.; CHIÒ, A.; TRAYNOR, B. J. State of play in amyotrophic lateral sclerosis genetics. **Nature neuroscience**, v. 17, n. 1, p. 17–23, jan. 2014.

RISCH, N. J. Searching for genetic determinants in the new millennium. **Nature**, v. 405, n. 6788, p. 847–56, 15 jun. 2000.

ROCHE, D. B. et al. The IntFOLD server: an integrated web resource for protein fold recognition, 3D model quality assessment, intrinsic disorder prediction, domain

prediction and ligand binding site prediction. **Nucleic acids research**, v. 39, n. Web Server issue, p. W171–6, jul. 2011.

ROSE, A. et al. RHYTHM--a server to predict the orientation of transmembrane helices in channels and membrane-coils. **Nucleic acids research**, v. 37, n. Web Server issue, p. W575–80, jul. 2009.

RYKUNOV, D. et al. Improved scoring function for comparative modeling using the M4T method. **Journal of structural and functional genomics**, v. 10, n. 1, p. 95–9, mar. 2009.

SHERRY, S. T. et al. dbSNP: the NCBI database of genetic variation. **Nucleic acids research**, v. 29, n. 1, p. 308–11, 1 jan. 2001.

SIM, N.-L. et al. SIFT web server: predicting effects of amino acid substitutions on proteins. **Nucleic acids research**, v. 40, n. Web Server issue, p. W452–7, jul. 2012.

SONG, Y. et al. High-Resolution Comparative Modeling with RosettaCM. **Structure (London, England : 1993)**, v. 21, n. 10, p. 1735–42, 8 out. 2013.

SU, X. W. et al. Genetic heterogeneity of amyotrophic lateral sclerosis: implications for clinical practice and research. **Muscle & nerve**, v. 49, n. 6, p. 786–803, jun. 2014.

TAN, M.-S. et al. Genome-wide association studies in neurology. **Annals of translational medicine**, v. 2, n. 12, p. 124, dez. 2014.

THE UNIPROT CONSORTIUM. UniProt: a hub for protein information. **Nucleic Acids Research**, v. 43, n. D1, p. D204–212, 27 out. 2015.

WESTBROOK, J. et al. **The Protein Data Bank: unifying the archive.****Nucleic acids research**, 2002. Disponível em:
<<http://www.ncbi.nlm.nih.gov/pubmed/11752306?dopt=AbstractPlus&holding=f1000,f1000m,isrctn>>

YAMANAKA, K.; SASAGAWA, Y.; OGURA, T. Recent advances in p97/VCP/Cdc48 cellular functions. **Biochimica et biophysica acta**, v. 1823, n. 1, p. 130–7, jan. 2012.

YAMASHITA, S.; ANDO, Y. Genotype-phenotype relationship in hereditary amyotrophic lateral sclerosis. **Translational neurodegeneration**, v. 4, p. 13, jan. 2015.

YANG, J. et al. The I-TASSER Suite: protein structure and function prediction. **Nature methods**, v. 12, n. 1, p. 7–8, 30 dez. 2014.

YE, Y. Diverse functions with a common regulator: ubiquitin takes command of an AAA ATPase. **Journal of structural biology**, v. 156, n. 1, p. 29–40, out. 2006.

YING, Z.; WANG, H.; WANG, G. The ubiquitin proteasome system as a potential

target for the treatment of neurodegenerative diseases. **Current pharmaceutical design**, v. 19, n. 18, p. 3305–14, jan. 2013.

ZHANG, Y.; SKOLNICK, J. TM-align: a protein structure alignment algorithm based on the TM-score. **Nucleic acids research**, v. 33, n. 7, p. 2302–9, jan. 2005.

ZHOU, J.; YI, J.; BONEWALD, L. Muscle-Bone Crosstalk in Amyotrophic Lateral Sclerosis. **Current osteoporosis reports**, v. 13, n. 5, p. 274–9, out. 2015.