



UNIRIO

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

**Modelagem estrutural e dinâmica molecular de variantes
genéticas da proteína FUS humana na esclerose lateral
amiotrófica do tipo 6**

Luiz Felipe Sarmiento Bonet

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

Rio de Janeiro

2021

Luiz Felipe Sarmiento Bonet

Modelagem estrutural e dinâmica molecular de variantes genéticas da proteína FUS humana na esclerose lateral amiotrófica do tipo 6

Dissertação de mestrado apresentada ao Programa de Pós-Graduação em Genética e Biologia Molecular 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

2021

B712

Bonet, Luiz Felipe Sarmiento

Modelagem estrutural e dinâmica molecular de variantes genéticas da proteína FUS humana na esclerose lateral amiotrófica do tipo 6 / Luiz Felipe Sarmiento Bonet – Rio de Janeiro, 2021.

Orientadora: Joelma Freire de Mesquita.

Coorientador: Tiago Fleming Outeiro.

Dissertação (Mestrado) – Universidade Federal do Estado do Rio de Janeiro, Programa de Pós-Graduação em Biologia Molecular e Celular, 2021.

1. Esclerose lateral amiotrófica. 2. SNP. 3. Fused in sarcoma. 4. Biologia computacional. I. De Mesquita, Joelma Freire, orient. II. Outeiro, Tiago Fleming, coorient. III. Título.

Luiz Felipe Sarmiento Bonet

Modelagem estrutural e dinâmica molecular de variantes genéticas da proteína FUS humana na esclerose lateral amiotrófica do tipo 6

Dissertação de mestrado apresentada ao Programa de Pós-Graduação em Genética e Biologia Molecular 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

Dr^a. Joelma Freire de Mesquita

Doutora em Bioquímica – UNIRIO

Dr^a. Claudia Alessandra Fortes Aiub

Doutora em Biociências Nucleares – UNIRIO

Dr^a. Carmen Lucia Antão Paiva

Doutora em Genética – UNIRIO

Dr^a. Nicole de Miranda Scherer

Doutora em Bioinformática – INCA

Dr. Andre Luis dos Santos Figueiredo

Doutor em Neurologia – UNIFASE

Dedico este trabalho à memória de
minha avó, Maria Zilma.

AGRADECIMENTOS

À natureza, por nos ter dado vida e por ainda nos nutrir. E acima de tudo, por permitir que, apesar de tudo, evoluíssemos até o ponto em que chegamos. Que seja, hoje e sempre, objeto de incomparável fascínio e respeito.

À Professora Joelma Freire de Mesquita, cujas palavras me colocaram de volta no caminho da ciência, e que me guiou pelos caminhos tortuosos da ciência com zelo e maestria. Mais que uma orientadora, uma verdadeira mestra. Obrigado pelos ensinamentos, e por todo o carinho, dedicação e muita paciência. Agradeço também ao Professor Tiago Fleming, por aceitar ser meu coorientador.

Aos membros da banca avaliadora por tirarem tempo de suas ocupadas agendas e aceitarem participar dessa importante etapa da minha vida acadêmica. Obrigado pela paciência.

À minha mãe, Maria Antonieta. Mais que mãe: uma amiga, um ícone, minha âncora e meu esteio. À minha avó, Maria Zilma, que sempre nos apoiou. Eu devo tudo o que tenho e tudo o que sou a vocês. Se cheguei até aqui, é porque vocês estavam sempre junto a mim, guiando meus passos e me apoiando. Amo vocês, minha gratidão é eterna.

Ao meu pai, Luiz Cláudio, por encher minha infância de carinho e lembranças felizes, e por todo o apoio até hoje. E à minha prima e irmã de coração, Andressa, que enche meus dias de música, brilho e sorrisos. Amo vocês e espero tê-los sempre ao meu lado.

Às minhas avós, Zelda e Déa, pelo amor e pelos mimos. E à minha tia Bil, que desde neném me ensinava a ser uma pessoa melhor. Cozinheiras e arteiras de alma e de coração. Vocês foram incríveis e fazem falta na minha vida. Com vocês foi um pouco do brilho desse mundo.

À Aloma. Minha colega de laboratório, companheira de viagem, confidente e cúmplice. Muitas vezes a palavra 'amiga' parece não ser suficiente para tudo o que ela significa para mim. Obrigado por ainda estar ao meu lado, espero poder estar ao seu lado até o sempre.

Aos demais amigos, Tamires, Fernanda e Lima. Vocês ajudaram a manter minha sanidade e meu sorriso em momentos de dúvida e de escuridão. Obrigado por me aceitarem em suas vidas.

A Lauriza, Alessandra e Elisângela, tivemos a sorte de cairmos nessa aventura juntos. E, com parceria e apoio mútuo, chegamos juntos até o fim. Obrigado pelos momentos felizes.

A meus tios, Itália e Marcos, e a todos da minha família que de alguma forma me apoiaram nos caminhos da vida. Vocês foram imprescindíveis, eu agradeço infinitamente o amor e carinho.

Ao Programa de Pós-Graduação em Biologia Molecular e Celular da UNIRIO, que se tornou quase um lar para mim. Como representante discente tive a oportunidade de conhecer a maior parte dos membros do corpo docente e administrativo. Com isso, minha admiração e respeito por todos os profissionais envolvidos cresceu exponencialmente. Toda a dedicação de vocês não passa despercebida, obrigado.

A todos os demais professores e instrutores que me guiaram nessa longa jornada e marcaram minha vida profissional. Sempre lembrarei de suas lições e ensinamentos, obrigado pela dedicação.

Aos demais amigos e companheiros, de laboratório, de curso e da vida, que tive o prazer de conhecer nessa minha incrível aventura na ciência. Sempre levarei os momentos e lembranças alegres comigo, obrigado por tudo.

Agradeço também ao apoio financeiro das instituições de fomento que tornaram este trabalho possível: Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Financiadora de Estudos e Projetos (FINEP), and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

*Você jamais estará
excessivamente bem-vestido ou
será excessivamente educado.*

Oscar Wilde

RESUMO

Título: Modelagem estrutural e dinâmica molecular de variantes genéticas da proteína FUS humana associadas à esclerose lateral amiotrófica do tipo 6

A esclerose lateral amiotrófica (ALS) é a doença neurodegenerativa de neurônios motores com início na fase adulta mais frequente. A doença é caracterizada por degeneração dos neurônios motores superiores e inferiores, geralmente levando à morte em até 5 anos após o início dos sintomas. Enquanto a maior parte dos casos é esporádica, 5% a 10% dos casos podem ser associados a herança familiar, incluindo ALS tipo 6, que é associada a mutações no gene Fused in Sarcoma (FUS). O objetivo deste trabalho foi avaliar como as mutações associadas a ALS tipo 6 mais frequentes, R521C, R521H e P525L, afetam a estrutura e a função da proteína. Utilizamos, então, algoritmos de predição funcional para analisarmos os efeitos de variantes de nucleotídeo único (SNV) não-sinônimas, bem como realizamos análises da conservação estrutural e dos níveis de frustração da proteína, e dinâmicas moleculares. A maior parte dos algoritmos classificou as três mutações como deletérias. A análise apontou que as três mutações diminuem a estabilidade da proteína, especialmente a variante R521C, que também aumenta a tendência de ligação de chaperonas. A análise da frustração da proteína mostrou um aumento na frustração entre interações envolvendo o resíduo mutado R521C. A análise de conservação estrutural mostrou que os sítios 521 e 525 são altamente conservados. Os resultados das dinâmicas moleculares indicam que a estabilidade da proteína é comprometida por todas as mutações estudadas. Estas também afetaram a área de superfície e a compactidade da proteína. As mutações estudadas também mostraram elevada flexibilidade na maior parte dos resíduos, notadamente nos sítios de interação com a proteína de importe nuclear da FUS.

Palavras-chave: Esclerose lateral amiotrófica, SNV, Fused in sarcoma, dinâmica molecular, biologia computacional

ABSTRACT

Title: Structural modelling and molecular dynamics of genetic variants of human FUS protein associated with type 6 amyotrophic lateral sclerosis

Amyotrophic lateral sclerosis (ALS) is the most frequent adult-onset motor neuron disorder. The disease is characterized by degeneration of upper and lower motor neurons, leading to death usually within 5 years after the onset of symptoms. While most cases are sporadic, 5%-10% of cases can be linked to familial inheritance, including ALS type 6, which is associated with mutations in the Fused in Sarcoma (FUS) gene. The goal of this work was to evaluate how the most frequent ALS-related mutations in FUS, R521C, R521H, and P525L, affect the protein structure and function. We then used prediction algorithms to analyze the effects of the aforementioned non-synonymous single nucleotide variations, as well as performed evolutionary conservation analysis, protein frustration analysis, and molecular dynamics simulations. Most of the prediction algorithms classified the three mutations as deleterious to the human organism. All three mutations were predicted to reduce protein stability, especially the mutation R521C, which was also predicted to increase chaperone binding tendency. The protein frustration analysis showed an increase in frustration in the interactions involving the mutated residue R521C. Evolutionary conservation analysis showed that residues 521 and 525 of human FUS are highly conserved sites. The molecular dynamics results indicate that protein stability could be compromised in all three mutations. They also affected the exposed surface area and protein compactness. The analyzed mutations also displayed elevated flexibility in most residues in all variants, most notably in the interaction site with the nuclear import protein of FUS.

Keywords: Amyotrophic lateral sclerosis, SNV, Fused in sarcoma, molecular dynamics, computational biology

LISTA DE ILUSTRAÇÕES

Figura 1. Domínios da proteína FUS humana	20
Figura 2. Ligação entre FUS <i>wild-type</i> e Kap β 2	22
Figura 3. FUS protein domains.....	39
Figura 4. Mutational frustration analysis	41
Figura 5. Evolutionary conservation analysis	42
Figura 6. RMSD of the backbone atoms of the protein.....	44
Figura 7. RMSF of the backbone atoms of the protein	45
Figura 8. Rg of the backbone atoms of the protein.....	46
Figura 9. SASA of the protein.....	49

LISTA DE TABELAS

Tabela 1. Tipos de ALS	17
Tabela 2. Functional effect prediction of FUS mutations	38
Tabela 3. SNPeffect and FoldX effect predictions of FUS mutations.....	39

LISTA DE ABREVIATURAS E SIGLAS

ALS	do inglês, <i>Amyotrophic Lateral Sclerosis</i>
fALS	do inglês, <i>Familial Amyotrophic Lateral Sclerosis</i>
sALS	do inglês, <i>Sporadic Amyotrophic Lateral Sclerosis</i>
AMP	do francês, <i>Atrophie Musculaire Progressive</i>
AMS	Alinhamento Múltiplo de Sequências
CPU	do inglês, <i>Central Processing Unit</i>
FUS	do inglês, <i>Fused in Sarcoma</i>
GPU	do inglês, <i>Graphical Processing Unit</i>
Kap β 2	do inglês, <i>Karyopherin-β2</i>
NLS	do inglês, Nuclear Localization Signal
NPT	do inglês, <i>Number, Pressure and Temperature</i>
NVT	do inglês, <i>Number, Volume and Temperature</i>
R _g	do inglês, <i>Radius of Gyration</i>
RGG	da sequência de aminoácidos arginina-glicina-glicina
RMSD	do inglês, <i>Root Mean Square Deviation</i>
RMSF	do inglês, <i>Root Mean Square Fluctuation</i>
SASA	do inglês, <i>Solvent-accessible Surface Area</i>
SNV	do inglês, <i>Single-nucleotide Variant</i>
TLS	do inglês, <i>Translocated in Liposarcoma</i>
VMD	do inglês, <i>Visual Molecular Dynamics</i>

SUMÁRIO

1. INTRODUÇÃO	14
1.1. Esclerose Lateral Amiotrófica.....	14
1.1.1. ALS Tipo-6	18
1.2. Proteína Fused in Sarcoma.....	19
1.3. Mecanismos Propostos de Patogênese	22
1.4. Bioinformática	24
2. OBJETIVOS	27
2.1. Objetivo Geral	27
2.2. Objetivos Específicos	27
3. MATERIAL E MÉTODOS	28
3.1. Sequência e Estrutura Proteica.....	28
3.2. Predição Funcional de SNVs.....	28
3.3. Análise de Frustração da Proteína	29
3.4. Análise de Conservação Estrutural	30
3.5. Dinâmica Molecular	30
4. RESULTADOS E DISCUSSÃO.....	31
5. CONCLUSÃO.....	57
REFERÊNCIAS	58
ANEXO 1.....	68

1. INTRODUÇÃO

1.1. Esclerose Lateral Amiotrófica

A esclerose lateral amiotrófica (ALS) é uma doença neurodegenerativa que afeta os neurônios motores primários e secundários. As palavras *amiotrófica* e *esclerose* derivam do grego, significando respectivamente *falta de nutrição muscular* e *endurecimento*. Estes termos são sugestivos da atrofia e degeneração muscular acarretados pela doença. Lateral por sua vez, faz referência a área da coluna espinhal onde parte dos neurônios afetados, os neurônios motores, se localizam (MORRIS, 2015).

Apesar de ter sido descrita pelo neurologista francês Jean-Martin Charcot em 1869 (CHARCOT; JOFFROY, 1869), existem casos registrados de pacientes apresentando fraqueza muscular progressiva e atrofia muscular desde 1824 (ROWLAND, 2001). Nesse período prévio à descrição da doença por Charcot, os neurologistas chamaram este quadro de *atrofia muscular progressiva* (AMP) (VISSER; DE JONG; VISSER, 2008). Jean-Martin Charcot foi o primeiro a conectar a atrofia muscular à degeneração de neurônios motores, cunhando o termo esclerose lateral amiotrófica e diferenciando ALS de AMP (ROWLAND, 2001).

A ALS pode ser categorizada em dois tipos, ALS de início espinhal e ALS de início bulbar. A ALS espinhal possui o perfil sintomático mais clássico, onde degeneração neuronal se inicia pelos neurônios motores secundários, gerando fraqueza nos membros superiores e inferiores. Já na ALS bulbar, os primeiros neurônios motores a se degenerarem são os primários, apresentando sintomas como disfagia e disartria (WIJESKERA; LEIGH, 2009). Dois terços dos casos de ALS são de início espinhal, enquanto um terço dos casos são de início bulbar (WIJESKERA; LEIGH, 2009). Em ambos os subtipos, a idade em que os sintomas começam a se apresentar é, em média, 60 anos (VANCE *et al.*, 2009), tendo uma incidência de 1 a 2 a cada 100.000 habitantes por ano no ocidente (VAN DAMME; ROBBERECHT; VAN DEN BOSCH, 2017).

O quadro sintomático da ALS espinhal se inicia com fraqueza nos membros inferiores e superiores, causando problemas na marcha e controle fino dos movimentos. Fasciculações e câimbras também costumam ocorrer

(WIJESEKERA; LEIGH, 2009). Com a fraqueza muscular, os músculos começam a atrofiar, levando a espasticidade muscular. Conforme o quadro se agrava, os neurônios motores primários também são afetados, causando disfagia e disartria progressiva (MORRIS, 2015). Diferentemente do *onset* espinhal, onde disartria e disfagia são sintomas tardios, o primeiro sintoma da ALS bulbar costuma ser disartria, seguida por disfagia. Sialorreia também é comum em casos de ALS bulbar. Sintomas nos membros superiores e inferiores também são observados, surgindo juntamente com os sintomas supracitados (WIJESEKERA; LEIGH, 2009).

Em ambos os tipos de ALS, a degeneração de neurônios motores eventualmente começa a impossibilitar funções como a deglutição e a fala. Ao atingir os músculos respiratórios, a respiração também é prejudicada, sendo complicação respiratória a principal causa de morte em pacientes de ALS (MORRIS, 2015). Devido ao período mais curto entre o surgimento dos sintomas e o comprometimento da respiração, a ALS de início bulbar está associada a um diagnóstico pior (NAUMANN *et al.*, 2019). Dessa forma, a ALS é uma doença fatal, com uma expectativa de vida média de 1 a 5 anos após o surgimento dos sintomas (BENNION CALLISTER; PICKERING-BROWN, 2014).

Visto que a doença afeta os neurônios motores, as funções cognitivas, audição e visão não são afetadas nos estágios iniciais. Entretanto, com a progressão da doença, pode haver comprometimento de funções cognitivas envolvendo os lobos frontal e temporal (MORRIS, 2015). A musculatura oculomotora é a última a ser afetada pela doença, motivo pelo qual muitos pacientes utilizam o movimento dos olhos como forma de comunicação (AEBISCHER; KATO, 2007).

A afecção dos lobos frontal e temporal nos estágios tardios da ALS faz uma conexão com outra doença neurodegenerativa, a demência frontotemporal (FTD). Esta é a segunda forma mais frequente de demência de *onset* juvenil, apresentando degeneração dos neurônios nos lobos frontal e temporal, afetando comportamento, fala, e, em estágios avançados, a memória (NOLAN; TALBOT; ANSORGE, 2016). De fato, 50% dos pacientes de ALS apresentam sintomas de FTD no decorrer do quadro clínico e 15% dos pacientes de ALS são diagnosticados com FTD. Estudos genéticos recentes apontam que ALS e FTD

clássicas são pontos extremos de um contínuo, o que explicaria esta sobreposição de sintomas no quadro clínico dos pacientes (JI *et al.*, 2017).

Apesar desta e de outras descobertas, e da ampla literatura científica versando sobre a ALS, pouco é conhecido sobre as bases moleculares da doença. Essa falta de conhecimento contribui para a falta de tratamentos efetivos contra a doença (PETROV *et al.*, 2017). Com isso, existem apenas dois fármacos aprovados para o tratamento da ALS, edaravone e riluzole. O edaravone age como *scavenger* de radicais livres, enquanto as especificidades do mecanismo de ação do riluzole não são claras. Previamente, acreditava-se que este funcionava como modulador glutamatérgico. A hipótese mais aceita atualmente é que o mecanismo de ação do fármaco é altamente complexo, envolvendo modulação de canais de íons (PETROV *et al.*, 2017). Tanto o edaravone quanto o riluzole aumentam a sobrevivência do paciente em apenas 3 a 5 meses (BREINER; ZINMAN; BOURQUE, 2020).

Embora a etiologia exata da ALS seja ainda desconhecida, sabe-se que 90% a 95% dos casos são esporádicos, sendo considerada uma doença de origem multifatorial (HOLLINGER; OKOSUN; MITCHELL, 2016). Os demais casos, 5% a 10%, podem ser associados a herança familiar (CORCIA *et al.*, 2017). Desse modo, os dois tipos da doença são chamados de ALS esporádica (sALS) e ALS familiar (fALS), respectivamente.

Casos de fALS estão associados a mutações em genes específicos, tanto de maneira causativa, quanto como fator de suscetibilidade (CORCIA *et al.*, 2017). Mais de vinte genes já foram associados ao desenvolvimento de fALS (VAN DAMME; ROBBERECHT; VAN DEN BOSCH, 2017), como mostra a tabela 1. Com isso, a fALS é classificada em tipos, variando de 1 a 26 dependendo do gene onde a mutação ocorre, com o fenótipo da doença podendo também variar entre cada tipo de fALS (CORCIA *et al.*, 2017). Vale destacar que mutações nestes genes também estão presentes em 5% a 10% dos casos de sALS, mesmo não havendo histórico familiar, o que corrobora a ideia de que sALS é uma doença de origem complexa e multifatorial (AL-CHALABI; VAN DEN BERG; VELDINK, 2017).

Tabela 1 – Listagem dos tipos de ALS. AD representa herança autossômica dominante, enquanto AR representa herança autossômica recessiva.

Tipo de ALS	Gene	Locus	Herança
ALS1	Superoxide dismutase 1 (<i>SOD1</i>)	21q22.11	AD, AR
ALS2	Alsin (<i>ALS2</i>)	2q33.1	AR
ALS3	Desconhecido	18q21	AD
ALS4	Senataxin (<i>SETX</i>)	9q34.13	AD
ALS5	Spatacsin (<i>SPG11</i>)	15q21.1	AR
ALS6	Fused in sarcoma (<i>FUS</i>)	16p11.2	AD, AR
ALS7	Desconhecido	20p13	AD
ALS8	Vesicle-associated membrane protein-associated protein B/C (<i>VAPB</i>)	20q13.32	AD
ALS9	Angiogenin (<i>ANG</i>)	14q11.2	AD
ALS10	TAR DNA-binding protein 43 (<i>TARDBP</i>)	1p36.22	AD
ALS11	Polyphosphoinositide phosphatase (<i>FIG4</i>)	6q21	AD
ALS12	Optineurin (<i>OPTN</i>)	10p13	AD, AR
ALS13	Ataxin-2 (<i>ATXN2</i>)	12q24.12	AD
ALS14	Valosin-containing protein (<i>VCP</i>)	9p13.3	AD
ALS15	Ubiquilin 2 (<i>UBQLN2</i>)	Xp11.21	Ligada ao X
ALS16	Sigma-1 Receptor (<i>SIGMAR1</i>)	9p13.3	AR
ALS17	Charged multivesicular body protein 2B (<i>CHMP2B</i>)	3p11.2	AD
ALS18	Profilin 1 (<i>PFN1</i>)	17p13.2	AD
ALS19	Chorion protein gene ErB.4 (<i>ERBB4</i>)	2q34	AD
ALS20	Heterogeneous nuclear ribonuclearprotein A1 (<i>HNRNPA1</i>)	12q13.13	AD
ALS21	Matrin 3 (<i>MATR3</i>)	5q31.2	AD
ALS22	Tubulin alpha-4A (<i>TUBA4A</i>)	2q35	AD

ALS23	Annexin A11 (<i>ANXA11</i>)	10q22.3	AD
ALS24	NIMA-related kinase (<i>NEK1</i>)	4q33	AD
ALS25	Kinesin family member 5 (<i>KIF5A</i>)	12q13.3	AD
ALS26	TIA1 Cytotoxic granule-associated RNA-binding protein (<i>TIA1</i>)	2p13.3	AD
ALS-FTD1	Chromosome 9 open reading frame 72 (<i>C9ORF72</i>)	9p21.2	AD
ALS-FTD2	Coiled-coil-helix-coiled-coil-helix domain containing protein 10 (<i>CHCHD10</i>)	22q11.23	AD
ALS-FTD3	Sequestosome (<i>SQSM1</i>)	5q35.3	AD
ALS-FTD4	TANK-binding kinase (<i>TBK1</i>)	12q14.2	AD

Fonte: Adaptado de CORCIA, *et al.*, 2017.

1.1.1. ALS Tipo-6

Especificamente, a ALS do tipo 6 está associada a mutações na proteína fused in sarcoma. Este tipo de ALS corresponde a cerca de 5% dos casos de fALS (KWIATKOWSKI *et al.*, 2009) e a menos de 1% dos casos de sALS (YAMASHITA *et al.*, 2012) globalmente. Entretanto, estudos mostram que estes valores podem variar com grupos étnicos. Por exemplo, a ALS tipo 6 é o segundo tipo mais frequente em pacientes de fALS asiáticos, ao passo que ALS tipo 6 é apenas o quarto tipo mais frequente em pacientes de fALS europeus (ZOU *et al.*, 2017).

Certas mutações nesta proteína impedem sua localização adequada no núcleo celular, fazendo com que a FUS mutante se acumule no citosol. Com isso, análises *post-mortem* de tecido neuronal de pacientes de ALS apresentam agregados citosólicos FUS-positivos, tanto em neurônios, quanto em células da glia (DENG, H.; GAO; JANKOVIC, 2014). Estes agregados citosólicos são uma marca distintiva de alguns tipos de fALS e sALS, e tendem a crescer em número ao longo do tempo (SUZUKI *et al.*, 2012).

Apesar dos agregados citosólicos conterem moléculas de FUS mutada em seu corpo, outras proteínas também já foram observadas nestes agregados,

como é o caso da TDP-43 (DORMANN; HAASS, 2011). De fato, moléculas de TDP-43 *wild-type* são observadas em uma grande parcela de casos de ALS, tanto familiar, quanto esporádica, com a exceção de casos de fALS do tipo 1 (PARAKH; ATKIN, 2016), associada a mutações na superóxido dismutase 1 (SOD1). A própria FUS, assim como a TDP-43, é comumente encontrada em casos de sALS (PARAKH; ATKIN, 2016), atribui-se essa ocorrência a propriedades de domínios príon-like presentes tanto na FUS, quanto na TDP-43 (BURATTI *et al.*, 2005).

Entre si, os tipos de ALS podem apresentar fenótipos de doença bastante diferentes. Por exemplo, casos de ALS tipo 6 estão associados a uma progressão mais severa do que ALS dos tipos 1 e 10 (NAUMANN *et al.*, 2019). Mas mesmo entre as diferentes mutações na proteína FUS associadas ao desenvolvimento de ALS tipo 6, também se observa uma grande variação no fenótipo da doença. Um estudo mostrou que pacientes com mutações nos resíduos R524 e G507 apresentaram os primeiros sintomas, em média, 36 anos mais tarde que pacientes com a mutação P525L, que apresentaram os primeiros sintomas na faixa de 19 anos de idade (NAUMANN *et al.*, 2019). De fato, o paciente de ALS tipo 6 mais jovem já registrado, apresentando sintomas com apenas 11 anos de idade, era portador da mutação P525L (CONTE *et al.*, 2012). Esta mutação também está associada a uma maior proporção de pacientes que apresentam ALS de *onset* bulbar, quase 60%, especialmente se comparado à mutações no resíduo R521 da FUS, onde mais de 90% dos casos de ALS são de *onset* espinhal (NAUMANN *et al.*, 2019), que tem uma progressão considerada menos agressiva que ALS de *onset* bulbar (CHIÒ *et al.*, 2002).

1.2. Proteína Fused in Sarcoma

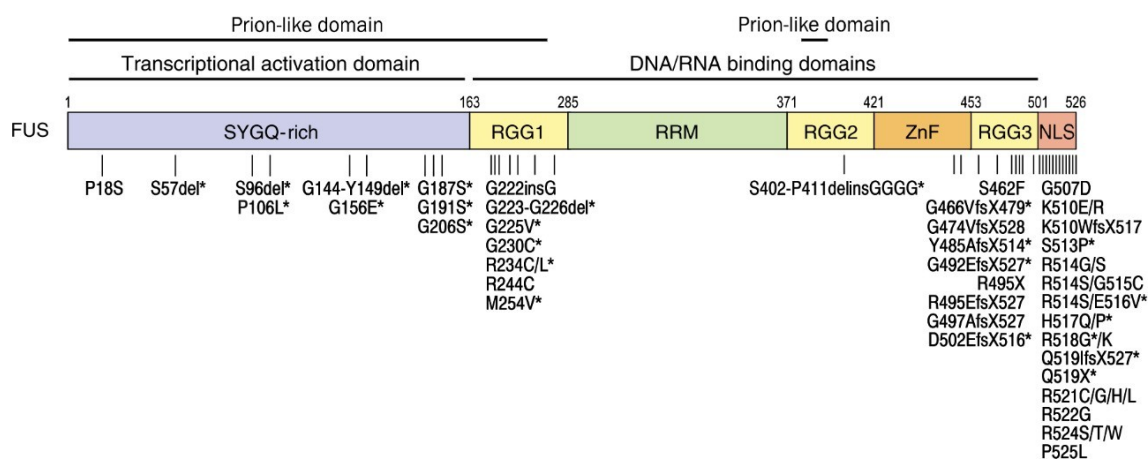
A proteína *Fused in Sarcoma*, também conhecida como *Translocated in Liposarcoma* (TLS), foi inicialmente identificada em 1993 como um oncogene em lipossarcomas, criado pela translocação cromossômica entre a região aminoterminal da FUS e um fator de transcrição (CROZAT *et al.*, 1993; RABBITTS *et al.*, 1993). Em 2009 foi descoberta a presença de agregados

citossólicos de FUS em casos de fALS (VANCE *et al.*, 2009) e FTD (MUNOZ *et al.*, 2009), e a proteína voltou a ser alvo de estudos.

Hoje se sabe que a FUS é uma proteína codificada pelo gene FUS, localizado no cromossomo 16, em 16p11.2 (SWETHA; RAMAIAH; ANBARASU, 2017). Trata-se de uma proteína da família TET, composta pela FUS, Ewing's sarcoma (EWS), e TATA-binding protein-associated factor 15 (TAF15) (TAN; MANLEY, 2009).

Composta por 526 aminoácidos, a FUS é uma ribonucleoproteína composta por sete domínios, como mostra a figura 1. Da região aminoterminal à região carboxiterminal, estes são: um domínio príon-like de baixa complexidade rico em Gly, Ser, Tyr e Gln; um domínio repleto de padrões de Arg-Gly-Gly (RGG); um sítio de reconhecimento de RNA; uma segunda região com padrões RGG; um domínio dedo-de-zinco do tipo RanBP2; uma terceira região com padrões RGG; um sítio de localização nuclear do tipo PY (YOSHIZAWA *et al.*, 2018).

Figura 1. Domínios da proteína FUS humana. A figura mostra os domínios que compõem a FUS, bem como mostra as funções de algumas regiões e alguns exemplos das mutações descritas.

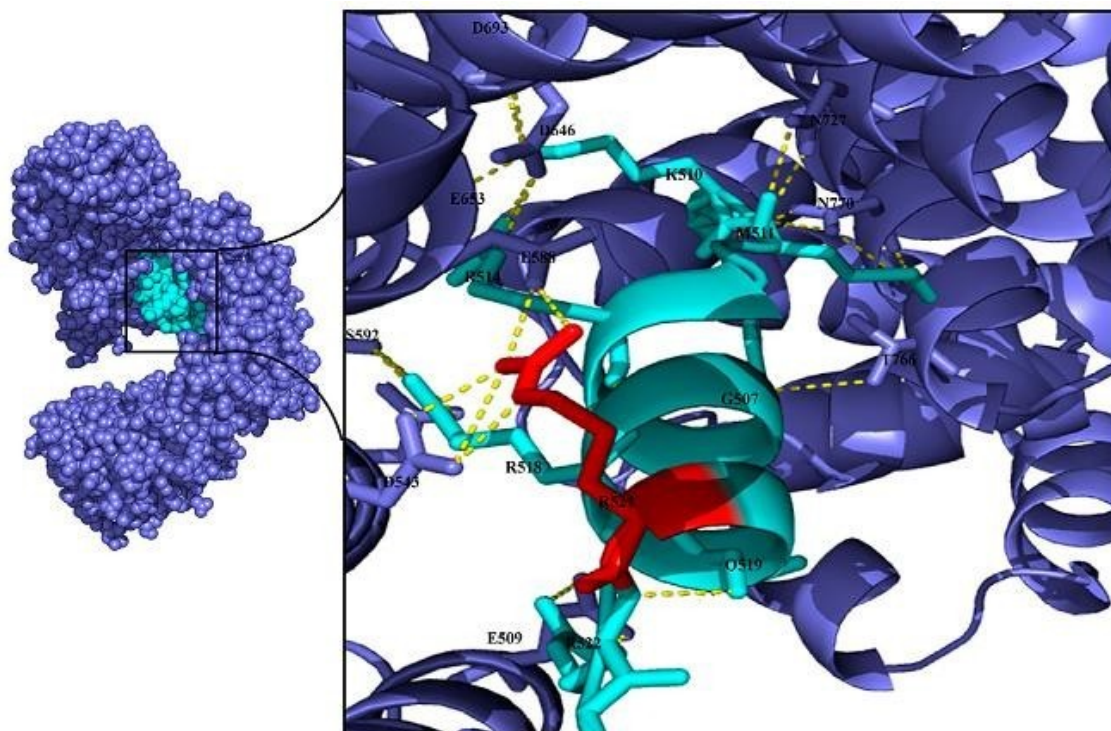


Fonte: Adptado de DORMANN; HAASS, 2013.

Várias funções são associadas à FUS, especialmente no metabolismo do DNA/RNA, como a regulação de *splicing* e de transcrição (DORMANN; HAASS, 2013), mas também na formação de grânulos de estresse (NAUMANN *et al.*, 2019). Outros estudos também apontam que a proteína pode ter um papel importante no reparo do DNA, visto que a proteína é rapidamente recrutada para regiões onde há quebra da dupla-fita (SHANG; HUANG, 2016). Dessa forma, defeitos nesta proteína podem impedi-la de exercer seu papel no reparo do DNA, gerando uma acumulação de danos no DNA. Sendo esse um dos fatores essenciais para a carcinogênese, imagina-se que poderia haver alguma relação entre defeitos na FUS e tumores malignos. Entretanto, não foi encontrada uma relação direta entre câncer e FALS (FANG *et al.*, 2013), mas sim uma relação entre mutações na FUS e morte celular dos neurônios (NAUMANN *et al.*, 2019).

Como proteína nuclear, a FUS deve ser importada para o núcleo, onde exerce sua função. Para isso, a molécula é carregada para dentro do núcleo por uma proteína chamada Carioferina- $\beta 2$ (Kap $\beta 2$), que se liga ao sítio de localização nuclear PY na extremidade carboxiterminal da FUS (SWETHA; RAMAIAH; ANBARASU, 2017), como mostra a figura 2. Estudos mostram que a Kap $\beta 2$ forma várias pontes salinas e de hidrogênio com resíduos do sítio de localização nuclear da FUS (ZHANG; CHOOK, 2012). Curiosamente, entre as mutações na FUS associadas à ALS, um grande número ocorre neste domínio da proteína, especialmente considerando que este é o menor domínio da proteína (DORMANN; HAASS, 2013).

Figura 2 – Ligação entre FUS *wild-type* e Kap $\beta 2$. O fragmento de FUS é mostrado em ciano, enquanto a Kap $\beta 2$ é mostrada em azul. As pontes de hidrogênio são mostradas em amarelo e o resíduo 521 está destacado em vermelho.



Fonte: Swetha (2017).

Acredita-se também que a FUS possui funções fora do núcleo celular, possivelmente tendo um papel na exportação de mRNA (DORMANN; HAASS, 2013). Em neurônios, a proteína está presente em grânulos e espinhas dendríticas, onde pode participar da síntese local de proteínas e da plasticidade sináptica (AOKI *et al.*, 2012). Também é possível que a FUS extranuclear neuronal tenha alguma função no desenvolvimento das espinhas dendríticas, visto que camundongos com *knockout* de FUS apresentam espinhas dendríticas anormais (FUJII; TAKUMI, 2005).

1.3. Mecanismos Propostos de Patogênese

Em virtude das diferenças no fenótipo da doença entre fALS do tipo 6 e outros tipos, como tipos 1 e 10, existe a hipótese que o desenvolvimento da fALS tipo 6 ocorra de maneira distinta dos demais tipos (SHANG; HUANG, 2016). Em contrapartida, alguns estudos também mostram possíveis paralelos entre efeitos de mutações na FUS com efeitos de mutações na SOD1, por exemplo (FARG *et al.*, 2012).

Uma das hipóteses sobre a patogênese da fALS tipo 6 envolve os agregados citosólicos presentes em neurônios ou células da glia afetados, um sinal típico da ALS. Estudos mostram que o domínio prion-like de baixa complexidade da FUS e suas regiões RGG tendem a interagir com regiões irmãs em outras moléculas de FUS, ou mesmo de moléculas com regiões similarmente desordenadas, como a TDP-43 (YOSHIZAWA *et al.*, 2018). Dessa forma, mutações no sítio de localização nuclear da proteína parecem não ter papel na associação entre moléculas de FUS (JU *et al.*, 2011). Isso sugere que a concentração citoplasmática de FUS tem maior influência na agregação de moléculas de FUS (DORMANN; HAASS, 2011). De fato, a ligação da FUS de alta afinidade com a Kap β 2 diminui sua interatividade com outras moléculas de FUS (YOSHIZAWA *et al.*, 2018), dessa forma o bom funcionamento das duas proteínas consegue manter a FUS em níveis que impeçam que a proteína se agregue em níveis anormais. Mutações que causem efeitos prejudiciais à afinidade da FUS com a Kap β 2 podem, conseqüentemente, perturbar este equilíbrio (YOSHIZAWA *et al.*, 2018).

A questão seguinte, então, passa a ser se a doença é causada pela ausência da proteína no núcleo para exercer sua função nuclear ou se há um ganho de função tóxico da proteína nos agregados citosólicos, ou mesmo se o que ocorre é um conjunto dos dois cenários. Alguns estudos em modelos não-humanos sugerem que o que ocorre é um ganho de função tóxico (BATEMAN *et al.*, 2015; HO; LING, 2019; JU *et al.*, 2011; SUN *et al.*, 2011). Estes estudos sugerem que os níveis elevados de FUS citosólicos causam um desbalanço em grande escala da proteostase e problemas no processamento normal do RNA em decorrência de inibição da autofagia, que causa estresse proteotóxico (HO; LING, 2019). De fato, a FUS está envolvida no metabolismo de inúmeros mRNAs (JI *et al.*, 2017), de modo que problemas no processamento do RNA podem desencadear graves conseqüências para a célula, especialmente ao se tratar da afecção de proteínas essenciais para a função e integridade neuronal (HO; LING, 2019). Alguns exemplos de alvos da FUS que são de grande importância são os mRNAs das proteínas TAU e ND1-L, proteínas essenciais para neurônios motores, que são altamente dependentes dos processos de transporte celular (DORMANN; HAASS, 2013). Outros autores, entretanto, apresentam isso como

uma perda da função normal da proteína, destacando as centenas de genes que tem seu *splicing* afetado pela disfunção na proteína, muitas que podem estar envolvidas na patogênese da ALS (JI *et al.*, 2017).

Existem dados que apontam alguns dos alvos dessa cascata de disfunções. A localização errônea da FUS para o citoplasma faz com que moléculas da proteína entrem ou interajam com o retículo endoplasmático, possivelmente ativando a resposta a proteínas mal enoveladas da organela, podendo até elevar os níveis de fatores de transcrição pro-apoptóticos como o CHOP (FARG *et al.*, 2012). Esta proteína está associada a morte celular de neurônios motores por estresse no retículo endoplasmático (SAXENA; CABUY; CARONI, 2009). Estes estudos também sugerem que o estresse no retículo endoplasmático pode se iniciar antes mesmo da acumulação de FUS disfuncional em agregados, visto que níveis elevados de CHOP foram encontrados em células neuronais com FUS mutante, mas ainda sem a formação de agregados citosólicos (FARG *et al.*, 2012). É possível também que o estresse no retículo endoplasmático se inicie também por outros fatores, como a perturbação na homeostase celular previamente mencionada como consequência de mutações na FUS (SALMINEN *et al.*, 2010).

Outra organela que também pode ser afetada por mutações na FUS é a mitocôndria. Estudos mostram que a proteína, exógena à mitocôndria, se acumula dentro da organela, onde a proteína afeta diretamente sua funcionalidade, diminuindo a produção de ATP e ativando a resposta a proteínas mal enoveladas da mitocôndria, de modo similar ao que é observado no retículo endoplasmático (DENG, *et al.*, 2018). Em modelo não-humano, algumas mutações da FUS, como R521C, R521H, e P525L, entre outras, são capazes até mesmo de encurtamento e fragmentação de mitocôndrias (NAKAYA; MARAGKAKIS, 2018). Estas são algumas pistas sobre a patogênese da fALS que, com as ferramentas certas, podem ajudar a desvendar as bases molecular da doença.

1.4. Bioinformática

Desde a primeira aplicação da informática no estudo de proteínas há mais de três décadas atrás, o campo da bioinformática evoluiu juntamente com a tecnologia de computadores e revolucionou a pesquisa biológica, auxiliando no estudo de proteômica, genômica, transcriptômica, entre outros (CHEN; HUANG; WU, 2017). O crescente poder de processamento das máquinas permite que abordemos problemas cada vez mais complexos, e o desenvolvimento de tecnologias mais avançadas e baratas permite uma crescente acessibilidade do campo para pesquisadores. Com isso, o conhecimento científico também cresce cada vez mais, por exemplo: em conjunção com técnicas avançadas de sequenciamento genético, o número de sequências depositadas em bancos de dados cresce exponencialmente (CHEN; HUANG; WU, 2017).

Com a crescente complexidade dos sistemas biológicos, a biologia computacional teve que avançar além do processamento serial das unidades de processamento central (CPU) (SHI *et al.*, 2015). Como alternativa, a biologia computacional passou a fazer uso do imenso poder de processamento das unidades de processamento gráfico (GPU) e do paralelismo que sua estrutura oferece. Entre as vantagens do uso de GPUs estão o custo relativamente baixo e a alta eficiência, que, dependendo da tarefa, pode ser competitiva com infraestruturas de alta-performance com dezenas de CPUs (NOBILE *et al.*, 2017). Em contrapartida, a arquitetura diferenciada que dá esse paralelismo às GPUs as torna relativamente mais difíceis de se usar do que CPUs, requerendo o uso de softwares que sejam especificamente programados para fazer uso da arquitetura das GPUs (DEMATTÉ; PRANDI, 2010).

Em termos de custo-benefício, o uso de métodos computacionais também traz inúmeras vantagens. Talvez o exemplo mais claro disso seja no desenvolvimento de novos fármacos, um processo que leva treze anos e pode custar US\$1,8 bilhão segundo estimativas (PAUL *et al.*, 2010). Grande parte desse tempo e do dinheiro são gastos em tentativas que acabam por nunca chegar ao mercado. Neste caso, técnicas computacionais podem ser empregadas na triagem e validação dos genes e proteínas alvo e na otimização de potenciais fármacos, poupando muito tempo e dinheiro (BANEGAS-LUNA *et al.*, 2019). Assim, a bioinformática cimenta-se como uma ferramenta rápida, confiável e até mais barata, visto que várias técnicas podem ser realizadas sem

uma grande infraestrutura, ou até mesmo por computação voluntária (GUERRERO *et al.*, 2014). Se aplicada em conjunção às técnicas experimentais, a bioinformática pode e vem revolucionando o desenho racional de fármacos (WOOLLER *et al.*, 2017).

Com a riqueza de informação sendo depositada diariamente em bancos de dados, a bioinformática também nos permite fazer uma análise integrativa de dados complexos, a fim de gerar informações de utilidade clínica (BECKMANN; LEW, 2016). Dessa forma, a bioinformática é uma das bases da medicina de precisão, permitindo a construção de melhores diagnósticos e prognósticos (KÖNIG *et al.*, 2017). Mas também nos permite um *insight* em muitos processos biológicos que envolvem sistemas altamente dinâmicos e interativos, como por exemplo a patogênese de doenças (AGRAWAL; ZITNIK; LESKOVEC, 2018).

2. OBJETIVOS

2.1. Objetivo Geral

Verificar, por meio de técnicas computacionais, alterações estruturais e funcionais ocasionadas por mutações genéticas não sinônimas na proteína FUS e analisar o comprometimento de estabilidade, flexibilidade, e atividade da proteína. Bem como relacionar as alterações observadas ao desenvolvimento da esclerose lateral amiotrófica.

2.2. Objetivos Específicos

- I. Avaliar a predição funcional dos efeitos das mutações sobre a proteína FUS humana;
- II. Analisar e comparar os níveis de frustração proteica da proteína FUS e as mutações estudadas;
- III. Analisar a conservação estrutural da proteína FUS;
- IV. Realizar dinâmicas moleculares da proteína FUS, *wild-type* e mutações estudadas, e analisar seus comportamentos em ambiente celular simulado.

3. MATERIAL E MÉTODOS

3.1. Sequência e Estrutura Proteica

A sequência selvagem da proteína FUS humana foi obtida em formato FASTA do banco de dados UNIPROT (BATEMAN *et al.*, 2015), ID:P35637. Foi realizado um levantamento das mutações da FUS já descritas na literatura científica. Este levantamento foi feito com informações dos bancos de dados UNIPROT, *Online Mendelian Inheritance in Man* (OMIM) (AMBERGER *et al.*, 2015) e *Amyotrophic Lateral Sclerosis Online Genetics Database* (ALSoD) (ABEL *et al.*, 2012). As mutações encontradas neste levantamento foram filtradas, usando informações do banco de dados UNIPROT, por associação com ALS de acordo com a literatura científica. A seleção final das mutações a serem estudadas foi feita pela avaliação de fatores como, por exemplo, frequência das mutações, e agressividade do fenótipo de doença associado a cada mutação.

O modelo tridimensional da FUS *wild-type* foi obtido do banco de dados Protein Data Bank (PDB) (BURLEY *et al.*, 2017). Não sendo encontrado um modelo completo da estrutura tridimensional da proteína, foi selecionado para análise computacional um fragmento contendo o sítio de localização nuclear da proteína (YOSHIZAWA *et al.*, 2018), PDB ID:4FDD.

3.2. Predição Funcional de SNVs

Dez algoritmos foram utilizados a fim de prever se as mutações estudadas causariam alterações deletérias ao funcionamento da proteína. Os algoritmos utilizados foram: MAPP (STONE; SIDOW, 2005), PhD-SNP (CAPRIOTTI; CALABRESE; CASADIO, 2006), Polyphen-1 (RAMENSKY; BORK; SUNYAEV, 2002), Polyphen-2 (ADZHUBEI *et al.*, 2010), SIFT (NG; HENIKOFF, 2001), SNAP (BROMBERG; YACHDAV; ROST, 2008), SNPs&GO (CALABRESE *et al.*, 2009), PMut (LÓPEZ-FERRANDO *et al.*, 2017), PROVEAN (CHOI; CHAN, 2015) e PredictSNP (BENDL *et al.*, 2014).

Os algoritmos supracitados classificam as mutações como deletérias ou neutras através de cálculos utilizando parâmetros respectivos a cada algoritmo. Em termos gerais, algoritmos como PhD-SNP, Polyphen-2, SNAP, SNPs&GO e

PMut fazem uso de aprendizado de máquina ou redes neurais (ADZHUBEI *et al.*, 2010; BROMBERG; YACHDAV; ROST, 2008; CALABRESE *et al.*, 2009; CAPRIOTTI; CALABRESE; CASADIO, 2006; LÓPEZ-FERRANDO *et al.*, 2017). Outros utilizam abordagens diferentes, como Polyphen-1, SIFT e PROVEAN que analisam informações evolutivas, por meio de alinhamento múltiplo de sequências (AMS) (CHOI; CHAN, 2015; NG; HENIKOFF, 2001; RAMENSKY; BORK; SUNYAEV, 2002). O algoritmo MAPP também utiliza informações evolutivas em conjunto com as propriedades físico-químicas da proteína (STONE; SIDOW, 2005). Outros ainda fazem uso de outros algoritmos, como é o caso do PredictSNP (BENDL *et al.*, 2014). Este é um classificador consenso, utilizando os resultados de oito algoritmos a fim de gerar um *score* e, a partir desse *score*, classificar se uma mutação é deletéria ou não.

Fatores adicionais decorrentes das mutações foram analisados na plataforma SNPEffect (DE BAETS *et al.*, 2012). Este é composto por cinco algoritmos: TANGO, que avalia a tendência de agregação; WALTZ, que avalia a propensão amiloide; LIMBO, que avalia a tendência de ligação a chaperonas; e FoldX, que avalia a estabilidade da proteína. Apesar de não classificar as mutações diretamente como deletérias ou neutras, estes fatores ajudam a identificar se elas podem ter alterações potencialmente deletérias ou não (DE BAETS *et al.*, 2012).

3.3. Análise de Frustração da Proteína

A análise de frustração de proteína foi realizada no servidor Protein Frustratometer 2 (PARRA *et al.*, 2016). Para essa análise, foi selecionado o método de cálculo mutacional, sem aplicação de eletrostática. Os demais parâmetros utilizados foram os padrões.

A estrutura selvagem e as variantes foram analisadas individualmente. Enquanto a estrutura selvagem foi obtida do PDB, foi necessário criar a estrutura tridimensional das variantes. Para tal, a estrutura selvagem foi importada para o software Visual Molecular Dynamics (VMD) (HUMPHREY; DALKE; SCHULTEN, 1996), versão 1.9.3. As estruturas variantes foram então geradas utilizando o

Mutator Plugin 1.3 (GAJULA *et al.*, 2014), uma extensão de mutagênese *in silico* do software.

3.4. Análise de Conservação Estrutural

Para a análise de conservação estrutural, foi utilizado o servidor ConSurf (ASHKENAZY *et al.*, 2016). O ConSurf é uma ferramenta que realiza alinhamento de sequências da sequência *input* no banco de dado selecionado pelo usuário e realiza o alinhamento múltiplo das sequências obtidas a partir dos resultados da busca no banco de dados. A ferramenta cria, então, uma árvore filogenética e calcula scores de conservação para cada resíduo de aminoácido da sequência *input*.

O UNIPROT foi selecionado como banco de dados para a busca de homólogos, sendo selecionado o HMMER como algoritmo de busca com 1 iteração e valor E de corte de 0.0001. Os valores máximo e mínimo de identidade para homólogos foram definidos em 95% e 35%, respectivamente. O método de alinhamento escolhido para a construção do AMS foi MAFFT-L-INS-i, enquanto o método de cálculo selecionado foi o método Bayesiano. O modelo de substituição evolutivo selecionado foi o padrão, “melhor modelo”.

3.5. Dinâmica Molecular

O Mutator Plugin 1.3 (GAJULA *et al.*, 2014), disponível no software VMD 1.9.1 (HUMPHREY; DALKE; SCHULTEN, 1996) foi usado para induzir as substituições R521C, R521H e P525L na estrutura da FUS *wild-type* nativa, que foi obtida no banco de dados PDB (ID: 4FDD). As estruturas selvagem e variantes foram submetidas a dinâmica molecular utilizando o pacote GROMACS (VAN DER SPOEL *et al.*, 2005), versão 5.0.7.

Utilizando o campo de força Amber99SB-ILDN (LINDORFF-LARSEN *et al.*, 2010), foi criada uma caixa triclnica com dimensões 56Å, 34Å e 31Å, onde a estrutura proteica foi solvatada em moléculas de água TIP3P (HARRACH; DROSSEL, 2014). O sistema foi neutralizado pela adição de íons Na⁺ e Cl⁻ e em seguida foi feita a minimização de energia pelo método *steepest descent*. Após

a minimização foram realizadas duas etapas de equilibração, a primeira com número, volume e temperatura constantes (NVT), e a segunda com número, pressão e temperatura constantes (NPT). Ambas as etapas de equilibração foram realizadas por 100ps com temperatura de 300K. O algoritmo Linear Constraint Solver (LINCS) (HESS *et al.*, 1997) foi utilizado na contenção das ligações covalentes, enquanto as interações eletrostáticas foram computadas através do método Particle-Mesh Ewald (PME) (DARDEN; YORK; PEDERSEN, 1993). As dinâmicas de produção foram executadas a 300K e 1atm por 300ns.

Três experimentos independentes foram realizados para a proteína selvagem e cada uma das variantes estudadas. As trajetórias resultantes foram analisadas usando os seguintes programas do pacote GROMACS: *gmx rms*, *gmx rmsf*, *gmx gyrate* e *gmx sasa*. Estes geraram correspondendo ao desvio médio quadrático (RMSD), flutuação média quadrática (RMSF), raio de giro (Rg) e área de superfície acessível ao solvente (SASA). O pacote ggplot2 do software R foi utilizado para gerar os calcular os valores médios de cada triplicata e para gerar os respectivos gráficos.

4. RESULTADOS E DISCUSSÃO

Molecular dynamics and protein frustration Analysis of Human Fused in Sarcoma Protein Variants in Amyotrophic Lateral Sclerosis Type 6: an In Silico approach

Molecular dynamics and protein frustration Analysis of Human Fused in Sarcoma Protein Variants in Amyotrophic Lateral Sclerosis Type 6: an In Silico approach

L.F.S. Bonet¹; J.P. Loureiro¹; G.R.C. Pereira¹; A.N.R. Da Silva¹; J.F. De Mesquita^{1*}

1 Laboratory of Bioinformatics and Computational Biology, Department of Genetics and Molecular Biology, Federal University of the State of Rio de Janeiro, Rio de Janeiro, Rio de Janeiro, Brazil.

* Corresponding author

E-mail: joelma.mesquita@unirio.br (JFM)

Abstract

Amyotrophic lateral sclerosis (ALS) is the most frequent adult-onset motor neuron disorder. The disease is characterized by degeneration of upper and lower motor neurons, leading to death usually within five years after the onset of symptoms. While most cases are sporadic, 5%-10% of cases can be associated to familial inheritance, including ALS type 6, which is associated with mutations in the Fused in Sarcoma (FUS) gene. This work aimed to evaluate how the most frequent ALS-related mutations in FUS, R521C, R521H, and P525L affect the protein structure and function. We used prediction algorithms to analyze the effects of the non-synonymous single nucleotide polymorphisms and performed evolutionary conservation analysis, protein frustration analysis, and molecular dynamics simulations. Most of the prediction algorithms classified the three mutations as deleterious. All three mutations were predicted to reduce protein stability, especially the mutation R521C, which was also predicted to increase chaperone binding tendency. The protein frustration analysis showed an increase in frustration in the interactions involving the mutated residue 521C. Evolutionary conservation analysis showed that residues 521 and 525 of human FUS are highly conserved sites. The molecular dynamics results indicate that protein stability could be compromised in all three mutations. They also affected the exposed surface area and protein compactness. The analyzed mutations also displayed high flexibility in most residues in all variants, most notably in the interaction site with the nuclear import protein of FUS.

Introduction

Amyotrophic Lateral Sclerosis (ALS) is a neurodegenerative disease characterized by damage to upper and lower motor neurons, resulting in progressive muscle atrophy, paralysis, and usually death due to respiratory failure within 1 to 5 years after the onset of symptoms (BENNION CALLISTER; PICKERING-BROWN, 2014). With an annual incidence of 2 per 100000 people worldwide, ALS is the most common adult-onset motor neuron disorder (REDLER; DOKHOLYAN, 2012). Despite the vast literature on the disease, its

physiological basis is still poorly understood, being considered a multifactorial condition in about 90-95% of cases (AL-CHALABI; HARDIMAN, 2013). The remaining 5-10% can be associated with familial inheritance, termed familial ALS (fALS), with several known genes relating to ALS either causatively or as susceptibility factors (CORCIA *et al.*, 2017).

Among these genes, the Fused in Sarcoma (FUS) is associated with approximately 5% of fALS cases (KWIATKOWSKI *et al.*, 2009). FUS is a DNA- and RNA-binding protein associated with splicing regulation, stress granule formation, mRNA transport, and DNA repair (NAUMANN *et al.*, 2019). FUS is a nuclear protein composed of 526 amino acid residues and seven domains (YOSHIZAWA *et al.*, 2018). Notably, most of the FUS mutations associated with ALS development occur in the protein's C-terminal region (PARAKH; ATKIN, 2016), specifically in its nuclear localization signal (NLS) domain. Among these non-synonymous single nucleotide polymorphisms (nsSNP), variant R521C is the most frequent mutation in ALS type-6 patients (KWIATKOWSKI *et al.*, 2009; VANCE *et al.*, 2009; YAMASHITA *et al.*, 2012). Variants R521H and P525L follow closely as second and third most common (NAUMANN *et al.*, 2019) mutations in ALS type 6 cases. Additionally, P525L has been associated with a more aggressive disease phenotype (NAUMANN *et al.*, 2019).

There are very few treatment options for ALS patients, with merely two approved drugs, both of which increase patient lifetime by 3-5 months (BREINER; ZINMAN; BOURQUE, 2020). The lack of knowledge regarding the pathogenesis of the disease contributes to this absence of effective treatments (PETROV *et al.*, 2017). Nevertheless, advances in computational technologies and applied bioinformatics have allowed scientists a new perspective on the molecular basis of diseases, which is an essential piece in rational drug design and precision medicine. Here, we performed molecular dynamics to better understand how ALS-related mutations in FUS behave inside a simulated cellular environment and how this behavior can be associated with fALS development. Furthermore, we used several computational algorithms to predict the effects of mutations R521C, R521H, and P525L and analyze the evolutionary conservation of amino acids in the human FUS protein.

Materials and Methods

We used the methodology previously described by our group (DA SILVA *et al.*, 2019; PEREIRA *et al.*, 2020; PEREIRA; VIEIRA; DE MESQUITA, 2021), to analyze the effects that mutations R521C, R521H, and P525L have over the human FUS protein.

Ethics Statement

This research does not involve human participants or human tissues. All data used as a starting point for our study were obtained from freely available databases, such as Protein Data Bank and UniProt. No participant consent or ethics committee approval is needed in this case.

Sequence and Structure Retrieval

The wild-type sequence of the human FUS protein, as well as a list of its naturally occurring variants, were obtained from the UNIPROT (UNIPROT CONSORTIUM, 2021) database [ID:P35637]. The structural model for the human FUS protein was obtained from the Protein Data Bank (ROSE *et al.*, 2017). As there is no model for the complete protein structure, a fragment [ID=4FDD] containing the nuclear localization signal (YOSHIZAWA *et al.*, 2018) was selected for computational analysis after meticulous research showed that molecular dynamics simulations (MD) involving protein fragments were viable (SUGIHARA; SUWA; BONDAR, 2014; SWETHA; RAMAIAH; ANBARASU, 2017). This fragment ranged from position 507 to 526 of the human FUS protein.

nsSNP Analysis

Ten algorithms were used to predict whether the variants could cause deleterious functional alterations on the protein: PredictSNP (BENDL *et al.*, 2014), MAPP (STONE; SIDOW, 2005), PhD-SNP (CAPRIOTTI; CALABRESE; CASADIO, 2006), Polyphen-1 (RAMENSKY; BORK; SUNYAEV, 2002), Polyphen-2

(ADZHUBEI; JORDAN; SUNYAEV, 2013), SIFT (NG; HENIKOFF, 2001), SNAP (BROMBERG; YACHDAV; ROST, 2008), SNPs&GO (CALABRESE *et al.*, 2009), PMut (LÓPEZ-FERRANDO *et al.*, 2017) and PROVEAN (CHOI; CHAN, 2015). Additional factors, such as aggregation tendency (TANGO), amyloid propensity (WALTZ), chaperone binding tendency (LIMBO), and protein stability were analyzed using the SNPeffect (DE BAETS *et al.*, 2012) and FoldX (SCHYMKOWITZ *et al.*, 2005) algorithms.

Protein Frustration Analysis

To investigate if the studied mutations caused changes in protein frustration levels, the wild-type structure and the variant structures were analyzed in the Protein Frustratometer 2 server (PARRA *et al.*, 2016). The variant structures (R521C, R521H, and P525L) were generated using the model obtained from the PDB and the *in silico* mutagenesis extension of the Visual Molecular Dynamics software (HUMPHREY *et al.*, 1996), version 1.9.3.

Evolutionary Conservation Analysis

The CONSURF algorithm was used to evaluate the evolutionary conservation of amino acid residues in the FUS protein (ASHKENAZY *et al.*, 2016). UNIPROT was selected as the database, and the homolog search algorithm used was HMMER, in 1 iteration and an E-value cutoff of 0.0001. Maximum and minimum identities for homologs were set as 95% and 35%, respectively. The alignment method used to build the Multiple Sequence Alignment was MAFFT-L-INS-i. The calculation method used was the Bayesian method, and the evolutionary substitution model was set to default.

Molecular Dynamics

The wild-type and variant models of the FUS protein were submitted to molecular dynamics using the Linux GROMACS (VAN DER SPOEL *et al.*, 2005) package, version 5.0.7. Using the Amber99SB-ILDN (LINDORFF-LARSEN *et al.*, 2010)

force field, the protein fragment was solvated in TIP3P (HARRACH; DROSSEL, 2014) water molecules inside a triclinic box with dimensions of 56Å, 34Å, and 31Å. The dynamics were neutralized by the addition of Na⁺ and Cl⁻ ions. The systems underwent energy minimization using the steepest descent method. After minimization, NVT (constant number of particles, volume, and temperature) and NPT (constant number of particles, pressure, and temperature) ensembles were performed, both at a pressure of 1atm and temperature of 300K for 100ps. The Linear Constraint Solver (LINCS) algorithm (HESS *et al.*, 1997) was utilized to contain covalent bonds, while the electrostatic interactions were computed using the Particle Mesh Ewald (PME) method (DARDEN; YORK; PEDERSEN, 1993). Production simulations were performed at 300K and 1atm for 300ns for the wild-type FUS and variants.

Three independent experiments were performed for the wild-type protein and its variants. The resulting trajectories were analyzed using the following GROMACS distribution programs: gmx rms, gmx rmsf, gmx gyrate and gmx sasa, which generated parameter values for Root-mean-Square Deviation (RMSD), Root-mean-Square Fluctuation (RMSF), radius of gyration (Rg), and solvent-accessible surface area (SASA). The ggplot2 package from the R software was used to generate the MD simulation graphs and calculate the mean values for the wild-type FUS and variants among the triplicates.

Results and Discussion

nsSNP Analysis

Most of the algorithms accurately predicted the mutations R521C, R521H, and P525L to be deleterious, as shown in Table 2. PredictSNP, MAPP, Polyphen-1, Polyphen-2, SIFT, SNAP, and Pmut were the most accurate predictors, classifying all three mutations as deleterious, while PHD-SNP and SNPs&GO classified all three as neutral. Interestingly, PROVEAN classified variants R521C and P525L as deleterious but R521H as neutral.

Table 2. Functional effect prediction of FUS mutations by ten nsSNP prediction algorithms.

Prediction algorithms	Protein Variants		
	R521C	R521H	P525L
PredictSNP	Deleterious	Deleterious	Deleterious
MAPP	Deleterious	Deleterious	Deleterious
PHD-SNP	Neutral	Neutral	Neutral
Polyphen-1	Deleterious	Deleterious	Deleterious
Polyphen-2	Deleterious	Deleterious	Deleterious
SIFT	Deleterious	Deleterious	Deleterious
SNAP	Deleterious	Deleterious	Deleterious
SNPs&GO	Neutral	Neutral	Neutral
PMut	Deleterious	Deleterious	Deleterious
PROVEAN	Deleterious	Neutral	Deleterious

Despite affecting the same amino acid position, the R521C mutation exchanges a positively charged and basic amino acid for a non-charged and polar amino acid. On the other hand, variant R521H swaps a positively charged and basic amino acid for another positively charged and basic amino acid, which would possibly have less impact on protein structure and function. Particularly regarding that alteration in the protein net charge can affect its ability to interact with other molecules (MOLINIER-FRENKEL; MESTIVIER; CASTELLANO, 2016).

Moreover, two algorithms classified as neutral three mutations that are known to be disease-related, reinforcing the importance of using several prediction algorithms. Since these algorithms use different parameters to make predictions, using several algorithms can yield more reliable results, as previously shown by our group (DE CARVALHO; DE MESQUITA, 2013).

Table 3 shows that none of the mutations were predicted to affect the aggregation tendency (TANGO) or amyloid propensity (WALTZ) of FUS. The human FUS protein is composed of seven domains (Fig 3), a prion-like region in its N-terminus that is rich in low-complexity (LC) amino acids, three regions with Arg-Gly-Gly motifs separating an RNA recognition motif, and a RanBP2-type zinc-finger domain (NGUYEN *et al.*, 2011), and a Pro-Tyr nuclear localization signal (PY-NLS) in its C-terminus (YOSHIZAWA *et al.*, 2018). Studies have shown that both the prion-like LC region (MATSUMOTO *et al.*, 2018) and the RGG motif regions (PATEL *et al.*, 2015) contribute to the self-association tendencies of the protein.

These SNPeffect results corroborate this idea; since these mutations occur in the NLS, they would likely have no direct effect on aggregation tendency, according to the algorithm. Similarly, amyloid fibrils have been observed only in the LC region (MURRAY *et al.*, 2017) and thusly are also unlikely to be directly affected by the studied mutations.

Table 3. SNPeffect and FoldX effect predictions of FUS mutations.

Predictive algorithms	Protein Variants		
	R521C	R521H	P525L
TANGO	No effect	No effect	No effect
WALTZ	No effect	No effect	No effect
LIMBO	Increases	No effect	No effect
FoldX	Greatly reduces	Reduces	Reduces

TANGO evaluates aggregation tendency. WALTZ evaluates amyloid propensity. LIMBO evaluates chaperone binding tendency. FoldX evaluates protein stability.

FUS Protein Domains

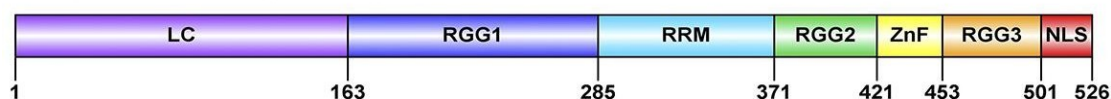


Fig 3. FUS protein domains. LC represents the low complexity, prion-like domain; RGG1 represents the first R-G-G pattern-rich domain; RRM represents the RNA-recognition motif; RGG2 represents the second R-G-G pattern-rich domain; ZnF represents the zinc-finger domain; RGG3 represents the third R-G-G pattern-rich domain, and NLS represents the nuclear localization signal.

A hallmark of ALS is the formation of cytosolic aggregates. The composition of these aggregates depends mainly on the genetic mutation that is observed in the patient (CIRYAM *et al.*, 2017). Aside from the mutated protein, other components can be found in these inclusions. For instance, misfolded wild-type TDP-43 is present in 97% of ALS cases (MAEKAWA *et al.*, 2009). Chaperones, specifically protein disulfide-isomerase (PDI), have also been observed as components in some FUS-positive inclusions (FARG *et al.*, 2012), although the role of

chaperones in ALS is still poorly understood. SNPeffect predicted an increase in chaperone binding tendency in mutation R521C, while R521H and P525L were unaffected.

Protein Frustration Analysis

The results obtained from SNPeffect led us to investigate protein frustration in FUS since chaperones can mitigate protein frustration in the energy landscape of protein folding (BUCCI, 2017). Protein frustration was best conceptualized by Ferreiro (2014), where the author gives the example of a network of magnets that attract and repel each other according to their individual spins and the resulting energy state between pairs of magnets is termed frustration. Similarly, the amino acid residues that compose a protein have electrostatic charges and, in some cases, polarity, so the same concept can be applied.

As shown in Fig 4, there is little difference in mutational frustration in the variants compared to the wild-type protein, except for variant R521C. In this case, there is a noticeable increase in frustration in interactions involving the mutated residue. This indicates that other amino acids would be more favorable in that position than the mutated cysteine (C). This change in frustration could also corroborate the SNPeffect results, as the elevated level of protein frustration could increase the need for chaperones to mitigate the frustration (BUCCI, 2017).

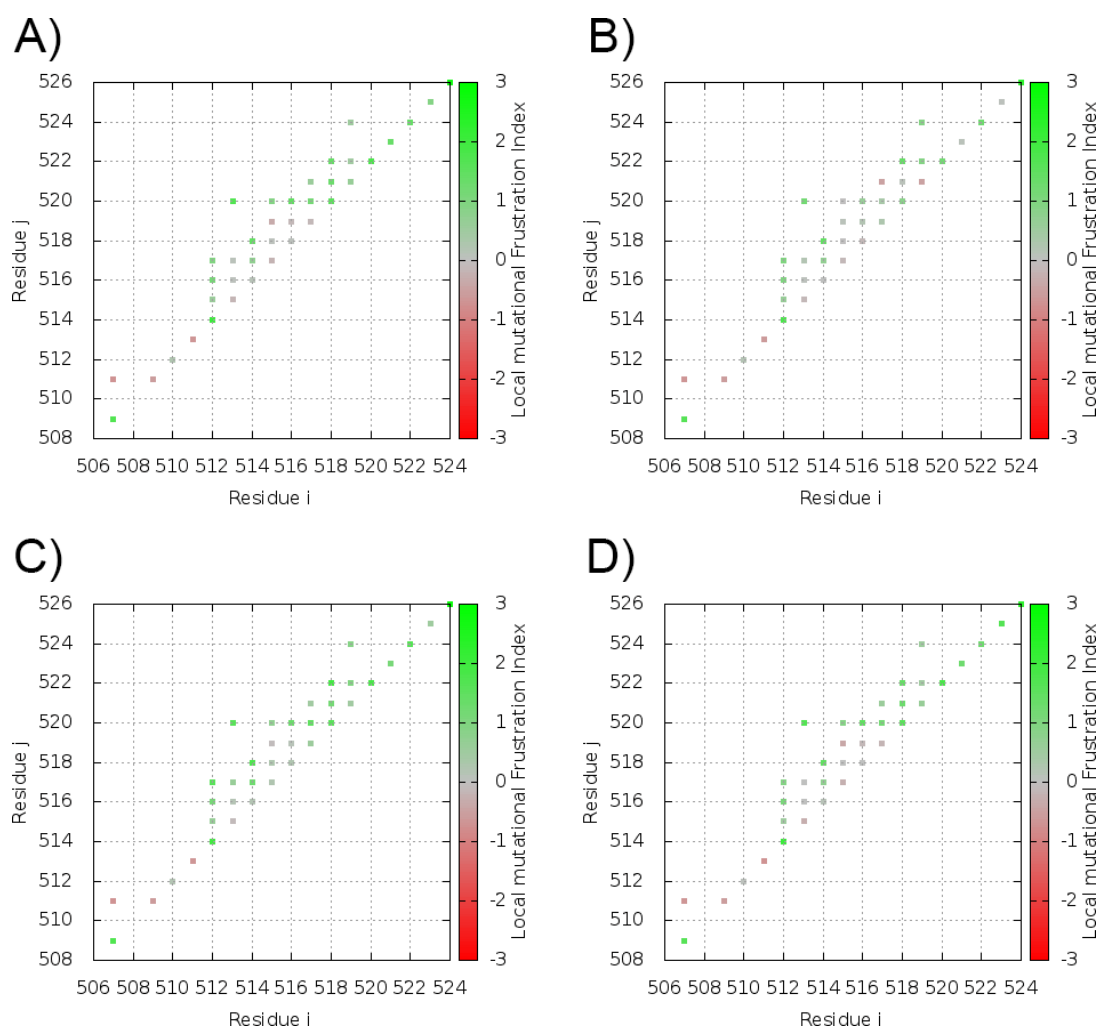


Fig 4. Mutational frustration analysis. Levels of mutational frustration by each residue-residue interaction. The colored scale represents the levels of frustration, ranging from highly frustrated (red), to neutral (gray) to minimally frustrated (green). A) Wild-type FUS. B) R521C variant. C) R521H variant. D) P525L variant.

Interestingly, these results show that residue 521 interacts with residue 517, but also with residues 518, 519, and 523, which are binding residues for Karyopherin- β 2 (Kap β 2), the importin that transports FUS, and other proteins, into the nucleus (SWETHA *et al.*, 2017). While it is unlikely that such electrostatic interactions would, by themselves, dramatically alter the structural stability of a protein (PACE; LAURENTS; THOMSON, 1990), the significance of the affected residues could make them delicate to change. Especially since hydrophobicity is a hallmark of PY-NLSs (SONIAT; CHOOK, 2016) like the one present in FUS, and the amino acid cysteine is more hydrophobic.

Evolutionary Conservation Analysis

CONSURF server analyses the evolutionary dynamics of amino acid residues between homologous sequences and gives each residue a score ranging from 1, more variable, to 9, more conserved (ASHKENAZY *et al.*, 2016). Results showed that the amino acid residues in the NLS of the protein tend to be more conserved than variable, as shown in Fig 5. This is expected due to the relevance of the domain, since areas that have an important role in protein function or structure tend to be more conserved (ASHKENAZY *et al.*, 2016). Indeed, seven out of the ten residues in the NLS that interact with Kap β 2 were given conservations scores of seven or higher. Mutations in the NLS of FUS have been shown to weaken the affinity between FUS protein and the importin, and this loss of affinity causes the mislocalization of the protein to the cytosol, an effect that has been linked to ALS (TYZACK *et al.*, 2019).

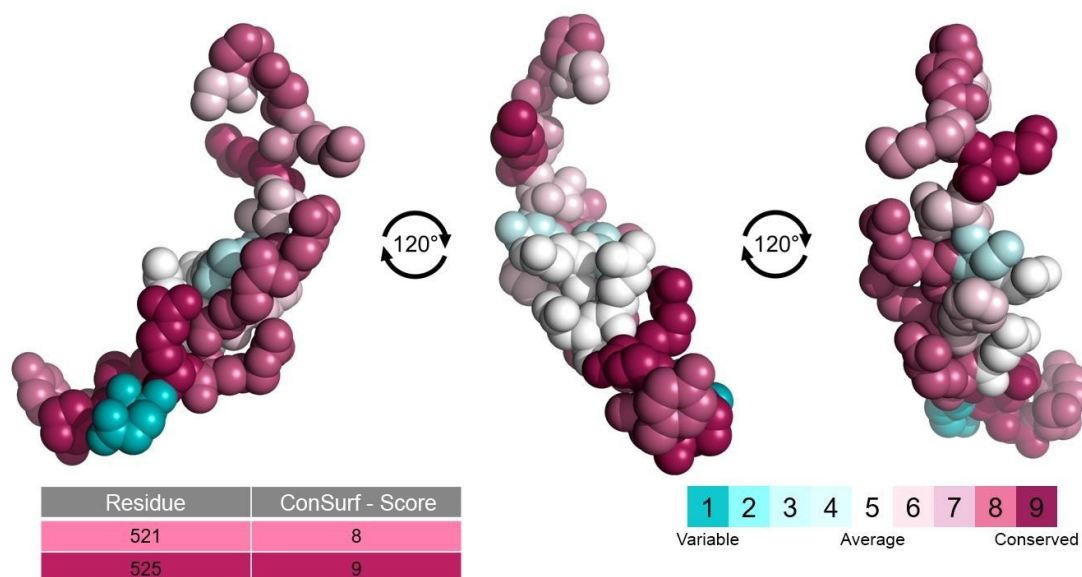


Fig 5. Evolutionary conservation analysis. Amino acid conservation analysis of wild-type FUS protein fragment (residue 507 to 526) represented in a space-filling model and colored according to conservation scores. Conservation scores of the residues where R521C, R521H, and P525L mutations occur are highlighted in the table. CONSURF color scheme representing conservation levels. Cyan represents more variable positions, while bordeaux represents more conserved positions.

The amino acid residues where the studied mutations occur were also scored with high conservation levels by the algorithm. Residue 521 received a score of 8, while residue 525 received a score of 9. This corroborates that mutations occurring in more conserved sites tend to be more deleterious (SUBRAMANIAN, 2011), seeing that variant P525L is frequently associated with a more aggressive disease phenotype (NAUMANN *et al.*, 2019). Bulbar onset is more frequent in P525L than in other FUS mutations (NAUMANN *et al.*, 2019), which has been shown to have a smaller time frame between disease onset and death than spinal onset (CHIÒ *et al.*, 2002). Additionally, the youngest case ever registered of ALS type 6 reported a carrier of the P525L FUS mutation with disease onset at merely 11 years of age (CONTE *et al.*, 2012), 28 years sooner than the median onset age for the condition (NAUMANN *et al.*, 2019).

Molecular Dynamics

During MD simulations, the trajectories of atoms and molecules are obtained by computational calculation of Newtonian equations of motion and application of different force fields (LIU *et al.*, 2018). Thusly, MDs can provide relevant information that would not be so easily discovered via experimental approaches (KARPLUS; MCCAMMON, 2002), which tend to be much more costly. This type of information is also fundamental to improve efficiency and reduce the blindness of drug discovery (LIU *et al.*, 2018).

RMSD is the average root-mean-square displacement of atoms between a reference structure and a dynamic structure, calculated at every frame of the simulation (VAN DER SPOEL *et al.*, 2005). The reference structure is usually the first frame of the structure in the simulation, which was the case in this study. RMSD, thus, is a good measure of structural similarity throughout the simulation (MARTÍNEZ, 2015). When a molecular system approaches equilibrium, the protein structure generally floats around an average stable conformation, and consequently, the RMSD values tend to form a plateau (DE OLIVEIRA *et al.*, 2019).

Wild-type FUS protein seemed to reach an equilibrium at around 125ns into the simulation when the RMSD value started fluctuating around 1nm, as shown in Fig 6. Variant P525L was the only mutation to come close to an equilibrated state, which can be observed from the onwards of 160ns. R521C and R521H stand out by generally having lower and higher RMSD values, respectively, and seemingly did not achieve equilibrium during the simulations.

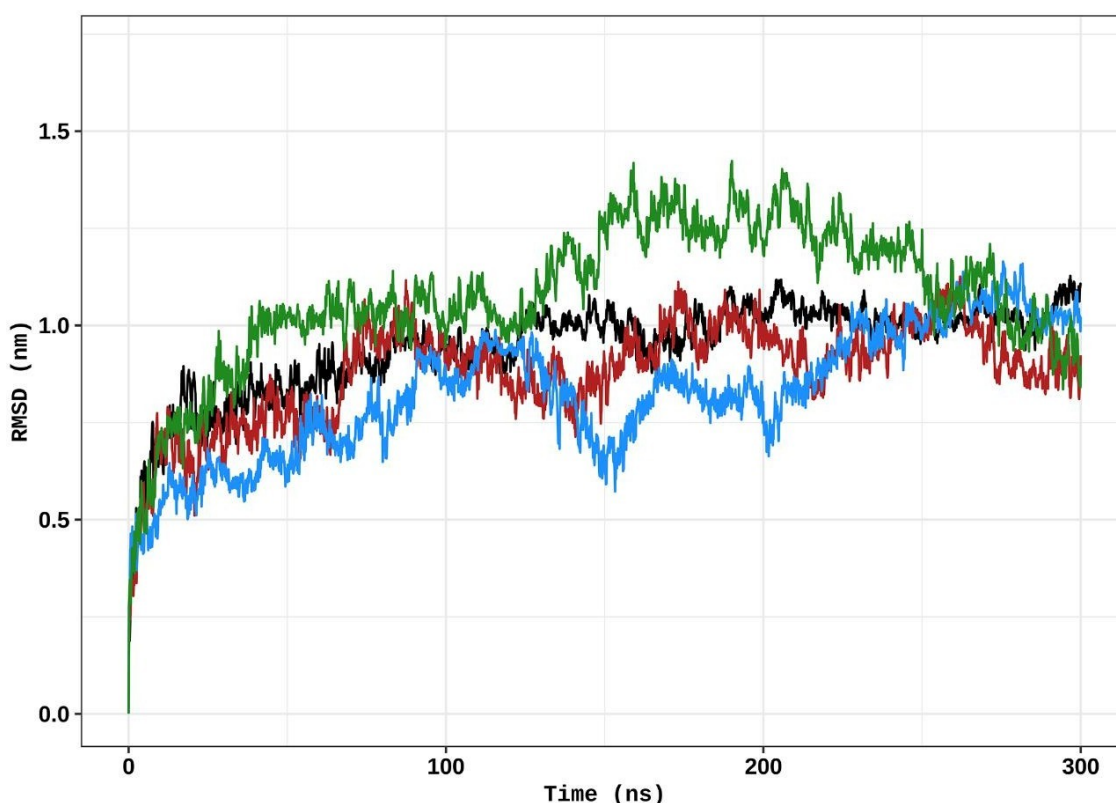


Figure 6. Root mean square deviation (RMSD) of the backbone atoms of the protein as a function of time. The wild-type FUS protein is represented in black, variant R521C is represented in blue, variant R521H is represented in green, and variant P525L is represented in red.

Root mean square fluctuation (RMSF) is similar to RMSD in that it measures the average displacement of atoms, but it does so relatively to individual residues throughout the simulation. Simply put, RMSF analyses flexibility differences for the amino acid residues (KUMAR, C. V. *et al.*, 2014).

Overall, variant P525L was shown to have residues slightly more flexible than the other variants (Fig 7). Residue 507 was shown to be the most different between

the variants and wild-type. While in the simulations of the P525L and R521C variants, residue 507 was approximately 33% and 60% more flexible than the wild-type, respectively, this residue was nearly three times as flexible as the wild-type in the R521H simulation. Conversely, residue 513 showed the least variation of flexibility among all proteins, with all four receiving an RMSF value of approximately 0.5nm. Curiously, residue 507 is one of the sites that interact with Kap β 2 (SWETHA; RAMAIAH; ANBARASU, 2017).

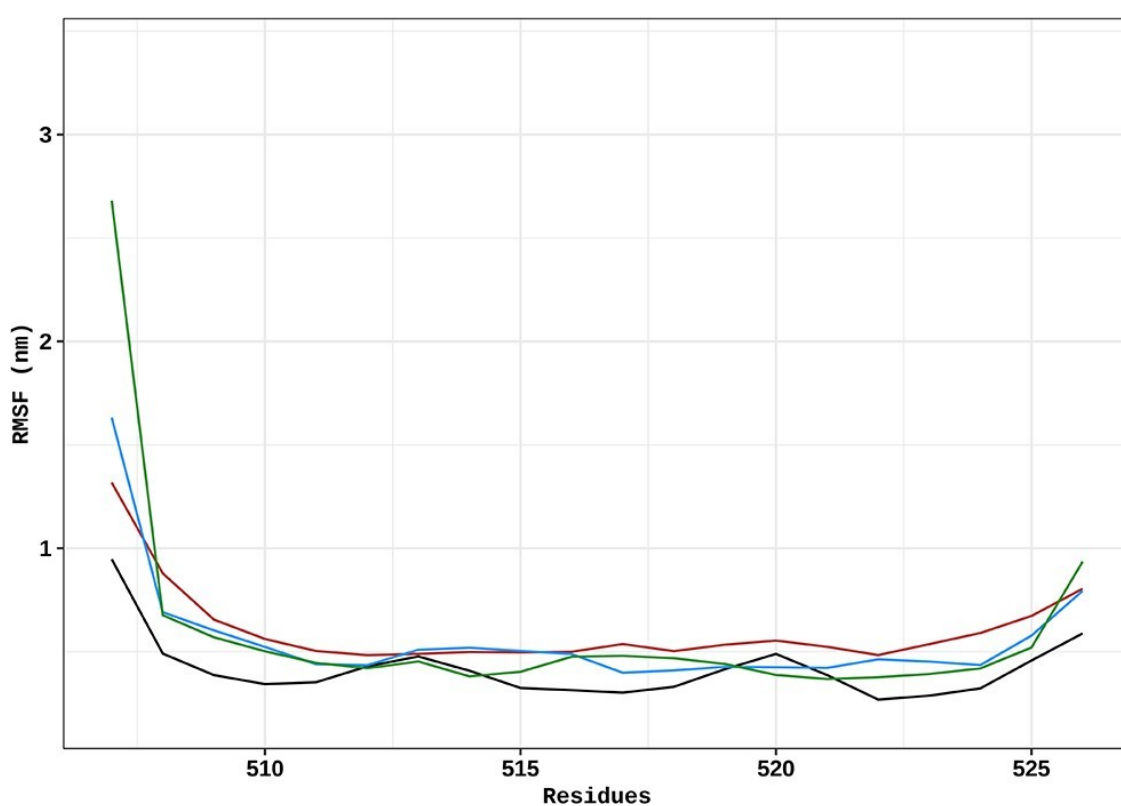


Figure 7. Root mean square fluctuation (RMSF) of the backbone atoms of the protein as a function of each amino acid residue. The wild-type FUS protein is represented in black, variant R521C is represented in blue, variant 521H is represented in green and variant P525L is represented in red.

Aside from residue 507, none of the other amino acid residues that interact with Kap β 2 showed any distinct behavior other than that flexibility was generally higher in the variants than in the wild-type. Notably, alterations in flexibility can be very disruptive to proteins, putting at risk protein binding affinity and specificity

(FORREY; DOUGLAS; GILSON, 2012), stability (KUMAR, S.; WOLFSON; NUSSINOV, 2001), and even misfolding (TEILUM; OLSEN; KRAGELUND, 2009), which is, by itself, linked to many diseases.

Radius of gyration (RG) is the mass-weighted root-mean-square distance of the atoms in a structure to their center of mass (SWETHA; RAMAIAH; ANBARASU, 2017). Thus, the radius of gyration can indicate the overall dimensions of a protein throughout an MD (KUMAR, C. V. *et al.*, 2014).

Wild-type FUS shows steady RG values, especially after 50ns, where the RG values fluctuate around 0.6nm until the end of the trajectory (Fig 8). Variant R521C presented steady RG values around 150ns, with values similar to those of the wild-type. Interestingly, variants R521H and P525L have seemingly opposite behavior between 160ns and 290ns, but they do not present a steady behavior throughout the simulation. In general, R521H and P525L presented less compactness than the wild-type fragment over time. Variant R521C showed similarities to the wild-type at certain points but was generally less compact, as well.

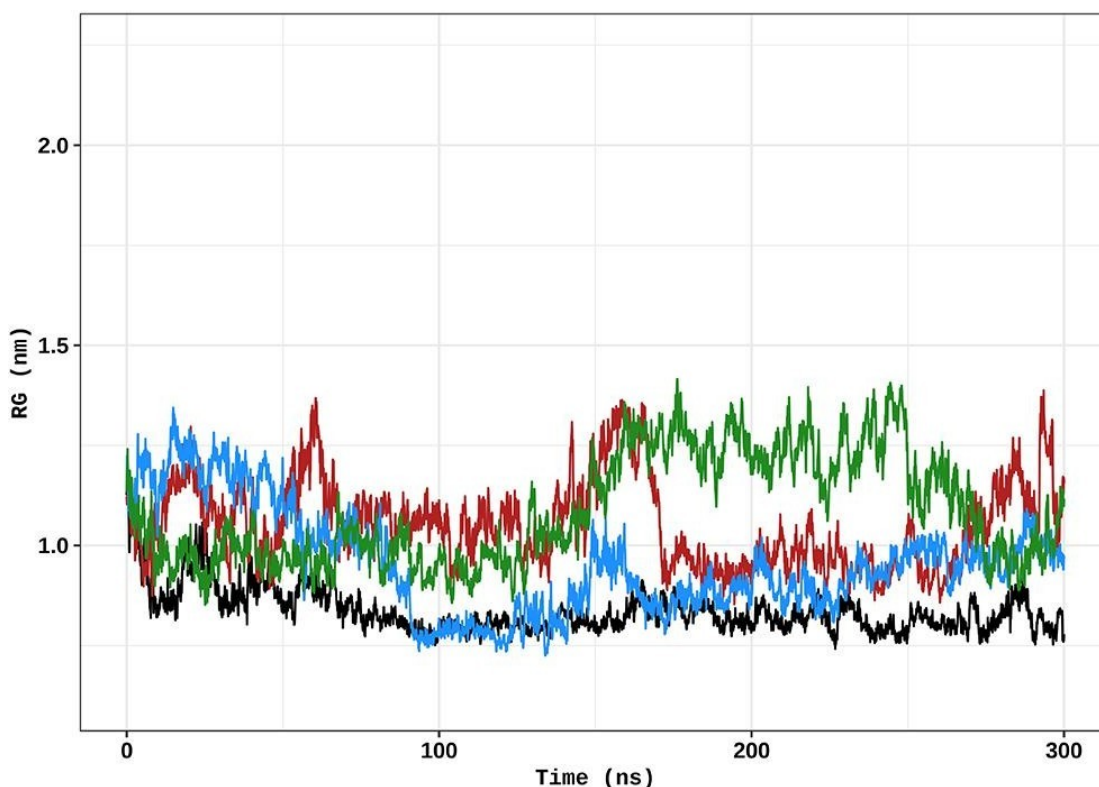


Figure 8. Radius of gyration (RG) of the backbone atoms of the protein as a function of time. The wild-type FUS protein is represented in black, variant R521C is represented in blue, variant 521H is represented in green, and variant P525L is represented in red.

Solvent-accessible surface area (SASA) is the area of the protein that is exposed to solvent (PEREIRA *et al.*, 2019). Thus, SASA can provide information about the ability of the protein to interact (KUMAR, C. V. *et al.*, 2014). Interestingly, FUS tends to interact with other FUS proteins and self-assemble, forming the cytosolic inclusions that are believed to contribute to neurodegeneration in ALS (MATSUMOTO *et al.*, 2018).

Overall, all variant fragments were shown to have higher surface areas, which corroborates the RG analysis results that showed that the mutant fragments were generally less compact than the wild-type protein fragment, as seen in Fig 9. Moreover, similarly to the observed in the RMSD, and RG analyses, the SASA values for the wild-type were considerably steadier than those of the analyzed variants, which could further suggest that the variants R521C, R521H, and P525L could reduce protein stability (PEREIRA *et al.*, 2020). This could also corroborate the results obtained from SNPeffect, which showed that all three mutations decrease protein stability.

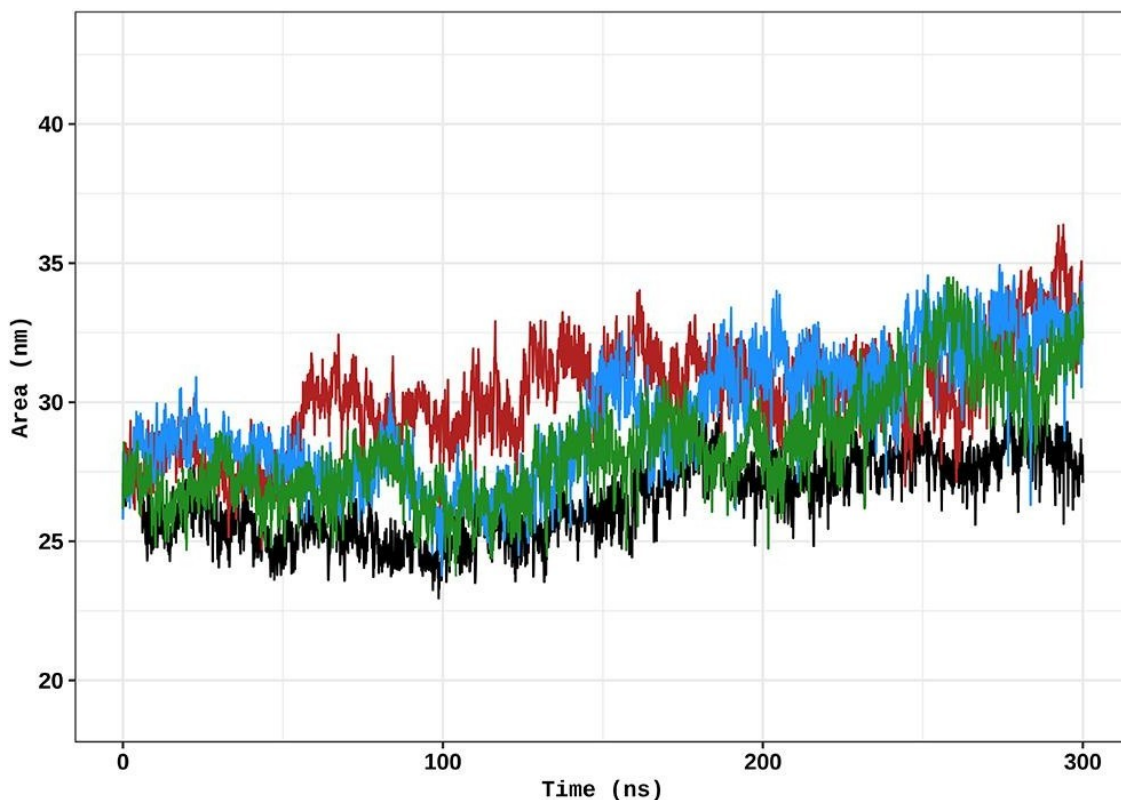


Figure 9. Solvent-accessible surface area (SASA) of the protein as a function of time. The wild-type FUS protein is represented in black, variant R521C is represented in blue, variant 521H is represented in green, and variant P525L is represented in red.

The results obtained in this study showed that mutations R521C, R521H, and P525L could alter stability and flexibility for the FUS protein, which are factors essential to the function of a protein (KUMAR, S.; WOLFSON; NUSSINOV, 2001). Flexibility, as previously mentioned, is an important factor in binding affinity and specificity (FORREY; DOUGLAS; GILSON, 2012). Interestingly, the increase in flexibility observed in the studied mutations occur on several residues that interact with Kap β 2, the protein that carries transports FUS into the nucleus. Thusly, this increase in flexibility provides a possible link to the decrease in binding affinity between mutated FUS proteins and Kap β 2 (SWETHA; RAMAIAH; ANBARASU, 2017). By lowering the binding affinity between these two proteins, mutated FUS would be more likely to remain in the cytoplasm, as is observed in type 6 ALS patients (CIRYAM *et al.*, 2017).

Conclusions

Through computational techniques, we were able to investigate how mutations R521C, R521H, and P525L affect the nuclear localization signal of FUS. Most of the functional prediction algorithms were capable of accurately predicting the three mutations as deleterious to the human organism. They also showed that the mutations decrease protein stability and, in the case of R521C, increase chaperone binding tendency. Relatedly, the protein frustration analysis showed an increase in frustration between the mutated R521C residue and the ones it interacts with. The evolutionary conservation analysis showed that the three mutations occur in highly conserved and probable important protein sites. The molecular dynamics results indicate that protein stability could be compromised in all three mutations. They also indicated that flexibility is elevated in most residues in all variants, most notably in the interaction site with the nuclear import protein of FUS. The analyzed mutations also affect the exposed surface area and protein compactness. These alterations in protein stability and flexibility could negatively impact the affinity between mutated FUS proteins and Kap β 2, leading to the formation of the FUS-positive cytoplasmic aggregates associated with type 6 ALS.

Acknowledgments

We would like to thank the financial support provided by Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) (Grant number: 01), Financiadora de Estudos e Projetos (FINEP), and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). We gratefully acknowledge the support of NVIDIA Corporation by donating the Titan X Pascal GPU used in this study.

References

ADZHUBEI, I.; JORDAN, D. M.; SUNYAEV, S. R. Predicting functional effect of human missense mutations using PolyPhen-2. **Current Protocols in Human Genetics**, v. 7, n. 76, p. Unit7.20, 2013.

AL-CHALABI, A.; HARDIMAN, O. The epidemiology of ALS: A conspiracy of genes, environment and time. **Nature Reviews Neurology**, v. 9, n. 11, p. 617–628, 2013.

ASHKENAZY, H. *et al.* ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. **Nucleic Acids Research**, v. 44, n. W1, p. W344–W350, 2016.

BENDL, J. *et al.* PredictSNP: Robust and Accurate Consensus Classifier for Prediction of Disease-Related Mutations. **PLoS Computational Biology**, v. 10, n. 1, p. 1–11, 2014.

BENNION CALLISTER, J.; PICKERING-BROWN, S. M. Pathogenesis/genetics of frontotemporal dementia and how it relates to ALS. **Experimental Neurology**, v. 262 Pt B, p. 84–90, 2014.

BREINER, A.; ZINMAN, L.; BOURQUE, P. R. Edaravone for amyotrophic lateral sclerosis: barriers to access and lifeboat ethics. **Canadian Medical Association Journal**, v. 192, n. 12, p. E319–E320, 2020.

BROMBERG, Y.; YACHDAV, G.; ROST, B. SNAP predicts effect of mutations on protein function. **Bioinformatics**, v. 24, n. 20, p. 2397–2398, 2008.

BUCCI, M. Minimizing frustration. **Nature Chemical Biology**, v. 13, n. 1, p. 1–1, 2017.

CALABRESE, R. *et al.* Functional annotations improve the predictive score of human disease-related mutations in proteins. **Human Mutation**, v. 30, n. 8, p. 1237–1244, 2009.

CAPRIOTTI, E.; CALABRESE, R.; CASADIO, R. Predicting the insurgence of human genetic diseases associated to single point protein mutations with support

vector machines and evolutionary information. **Bioinformatics**, v. 22, n. 22, p. 2729–2734, 2006.

CHIÒ, A. *et al.* Early symptom progression rate is related to ALS outcome: a prospective population-based study. **Neurology**, v. 59, n. 1, p. 99–103, 2002.

CHOI, Y.; CHAN, A. P. PROVEAN web server: A tool to predict the functional effect of amino acid substitutions and indels. **Bioinformatics**, v. 31, n. 16, p. 2745–2747, 2015.

CIRYAM, P. *et al.* Spinal motor neuron protein supersaturation patterns are associated with inclusion body formation in ALS. **Proceedings of the National Academy of Sciences of the United States of America**, v. 114, n. 20, p. E3935–E3943, 2017.

CONTE, A. *et al.* P525L FUS mutation is consistently associated with a severe form of juvenile amyotrophic lateral sclerosis. **Neuromuscular Disorders : NMD**, v. 22, n. 1, p. 73–75, 2012.

CORCIA, P. *et al.* Genetics of amyotrophic lateral sclerosis. **Revue Neurologique**, v. 173, n. 5, p. 254–262, 2017.

DA SILVA, A. N. R. *et al.* SOD1 in amyotrophic lateral sclerosis development – in silico analysis and molecular dynamics of A4F and A4V variants. **Journal of Cellular Biochemistry**, v. 120, n. 10, p. 17822–17830, 2019.

DARDEN, T.; YORK, D.; PEDERSEN, L. Particle mesh Ewald: An $N \cdot \log(N)$ method for Ewald sums in large systems. **The Journal of Chemical Physics**, v. 98, n. 12, p. 10089–10092, 1993.

DE BAETS, G. *et al.* SNPeffect 4.0: On-line prediction of molecular and structural effects of protein-coding variants. **Nucleic Acids Research**, v. 40, n. D1, 2012.

DE CARVALHO, M. D. C.; DE MESQUITA, J. F. Structural Modeling and In Silico Analysis of Human Superoxide Dismutase 2. **PLoS ONE**, v. 8, n. 6, 2013.

DE OLIVEIRA, C. C. S. *et al.* In silico analysis of the V66M variant of human BDNF in psychiatric disorders: An approach to precision medicine. **PLoS ONE**, v. 14, n. 4, p. 1–28, 2019.

FARG, M. A. *et al.* Mutant FUS induces endoplasmic reticulum stress in amyotrophic lateral sclerosis and interacts with protein disulfide-isomerase. **Neurobiology of Aging**, v. 33, n. 12, p. 2855–2868, 2012.

FERREIRO, D. U.; KOMIVES, E. A.; WOLYNES, P. G. Frustration in Biomolecules. **Quarterly Reviews of Biophysics**, v. 47, n. 4, p. 285–363, 2014.

FORREY, C.; DOUGLAS, J. F.; GILSON, M. K. The Fundamental Role of Flexibility on the Strength of Molecular Binding. **Soft Matter**, v. 8, n. 23, p. 6385–6392, 2012.

HARRACH, M. F.; DROSSEL, B. Structure and dynamics of TIP3P, TIP4P, and TIP5P water near smooth and atomistic walls of different hydroaffinity. **Journal of Chemical Physics**, v. 140, n. 17, 2014.

HESS, B. *et al.* LINCS: A linear constraint solver for molecular simulations. **Journal of Computational Chemistry**, v. 18, n. 12, p. 1463–1472, 1997.

HUMPHREY, W.; DALKE, A.; SCHULTEN, K. Visual Molecular Dynamics. **Journal of Molecular Graphics**, v. 14, p. 33–38, 1996.

KARPLUS, M.; MCCAMMON, J. A. Molecular dynamics simulations of biomolecules. **Nature Structural Biology**, v. 9, n. 9, p. 646–652, 2002.

KUMAR, C. V. *et al.* Computational Analysis Reveals the Association of Threonine 118 Methionine Mutation in PMP22 Resulting in CMT-1A. **Advances in Bioinformatics**, v. 2014, p. 502618, 2014.

KUMAR, S.; WOLFSON, H. J.; NUSSINOV, R. Protein flexibility and electrostatic interactions. **IBM Journal of Research and Development**, v. 45, n. 3.4, p. 499–512, 2001.

KWIATKOWSKI, T. J. *et al.* Mutations in the FUS/TLS Gene on Chromosome 16 Cause Familial Amyotrophic Lateral Sclerosis. **Science**, v. 323, n. 5918, p. 1205–1208, 2009.

LINDORFF-LARSEN, K. *et al.* Improved side-chain torsion potentials for the Amber ff99SB protein force field. **Proteins: Structure, Function and Bioinformatics**, v. 78, n. 8, p. 1950–1958, 2010.

LIU, X. *et al.* Molecular dynamics simulations and novel drug discovery. **Expert Opinion on Drug Discovery**, v. 13, n. 1, p. 23–37, 2018.

LÓPEZ-FERRANDO, V. *et al.* PMut: A web-based tool for the annotation of pathological variants on proteins, 2017 update. **Nucleic Acids Research**, v. 45, n. W1, p. W222–W228, 2017.

MAEKAWA, S. *et al.* TDP-43 is consistently co-localized with ubiquitinated inclusions in sporadic and Guam amyotrophic lateral sclerosis but not in familial amyotrophic lateral sclerosis with and without SOD1 mutations. **Neuropathology**, v. 29, n. 6, p. 672–683, 2009.

MARTÍNEZ, L. Automatic identification of mobile and rigid substructures in molecular dynamics simulations and fractional structural fluctuation analysis. **PLoS ONE**, v. 10, n. 3, p. e0119264, 2015.

MATSUMOTO, T. *et al.* Self-assembly of FUS through its low-complexity domain contributes to neurodegeneration. **Human Molecular Genetics**, v. 27, n. 8, p. 1353–1365, 2018.

MOLINIER-FRENKEL, V.; MESTIVIER, D.; CASTELLANO, F. Alterations of the immunosuppressive IL411 enzyme activity induced by naturally occurring SNP/mutations. **Genes and Immunity**, v. 17, n. 2, p. 148–152, 2016.

MURRAY, D. T. *et al.* Structure of FUS Protein Fibrils and Its Relevance to Self-Assembly and Phase Separation of Low-Complexity Domains. **Cell**, v. 171, n. 3, 2017.

NAUMANN, M. *et al.* Phenotypes and malignancy risk of different FUS mutations in genetic amyotrophic lateral sclerosis. **Annals of Clinical and Translational Neurology**, v. 6, n. 12, p. 2384–2394, 2019.

NG, P. C.; HENIKOFF, S. Predicting deleterious amino acid substitutions. **Genome Research**, v. 11, n. 5, p. 863–874, 2001.

NGUYEN, C. D. *et al.* Characterization of a Family of RanBP2-Type Zinc Fingers that Can Recognize Single-Stranded RNA. **Journal of Molecular Biology**, v. 407, n. 2, p. 273–283, 2011.

PACE, C. N.; LAURENTS, D. V.; THOMSON, J. A. pH dependence of the urea and guanidine hydrochloride denaturation of ribonuclease A and ribonuclease T1. **Biochemistry**, v. 29, n. 10, p. 2564–2572, 1990.

PARAKH, S.; ATKIN, J. D. Protein folding alterations in amyotrophic lateral sclerosis. **Brain Research**, v. 1648, Pt.B , p. 633–649, 2016.

PARRA, R. G. *et al.* Protein Frustratometer 2: a tool to localize energetic frustration in protein molecules, now with electrostatics. **Nucleic Acids Research**, v. 44, n. W1, p. W356–W360, 2016.

PATEL, A. *et al.* A Liquid-to-Solid Phase Transition of the ALS Protein FUS Accelerated by Disease Mutation. **Cell**, v. 162, n. 5, p. 1066–1077, 2015.

PEREIRA, G. R. C. *et al.* In silico analysis and molecular dynamics simulation of human superoxide dismutase 3 (SOD3) genetic variants. **Journal of Cellular Biochemistry**, v. 120, n. 3, p. 3583–3598, 2019.

PEREIRA, G. R. C. *et al.* In silico analysis of the tryptophan hydroxylase 2 (TPH2) protein variants related to psychiatric disorders. **PLoS ONE**, v. 15, n. 3, p. e0229730, 2020.

PEREIRA, G. R. C.; VIEIRA, B. de A. A.; DE MESQUITA, J. F. Comprehensive in silico analysis and molecular dynamics of the superoxide dismutase 1 (SOD1) variants related to amyotrophic lateral sclerosis. **PLoS ONE**, v. 16, n. 2, p. e0247841, 2021.

PETROV, D. *et al.* ALS clinical trials review: 20 years of failure. Are we any closer to registering a new treatment? **Frontiers in Aging Neuroscience**, v. 9, n. MAR 2017.

RAMENSKY, V.; BORK, P.; SUNYAEV, S. Human non-synonymous SNPs: server and survey. **Nucleic Acids Research**, v. 30, n. 17, p. 3894–3900, 2002.

REDLER, R. L.; DOKHOLYAN, N. V. The complex molecular biology of amyotrophic lateral sclerosis (ALS). **Progress in Molecular Biology and Translational Science**, v. 107, p. 215–262, 2012.

ROSE, P. W. *et al.* The RCSB protein data bank: integrative view of protein, gene and 3D structural information. **Nucleic Acids Research**, v. 45, n. D1, p. D271–D281, 2017.

SCHYMKOWITZ, J. W. H. *et al.* Prediction of water and metal binding sites and their affinities by using the Fold-X force field. **Proceedings of the National Academy of Sciences of the United States of America**, v. 102, n. 29, p. 10147–10152, 2005.

SONIAT, M.; CHOOK, Y. M. Karyopherin- β 2 Recognition of a PY-NLS Variant that Lacks the Proline-Tyrosine Motif. **Structure**, v. 24, n. 10, p. 1802–1809, 2016.

STONE, E. A.; SIDOW, A. Physicochemical constraint violation by missense substitutions mediates impairment of protein function and disease severity. **Genome Research**, v. 15, n. 7, p. 978–986, 2005.

SUBRAMANIAN, S. Fixation of deleterious mutations at critical positions in human proteins. **Molecular Biology and Evolution**, v. 28, n. 9, p. 2687–2693, 2011.

SUGIHARA, M.; SUWA, M.; BONDAR, A.-N. Dynamics of bovine opsin bound to G-protein fragments. **Journal of Structural Biology**, v. 188, n. 1, p. 79–86, 2014.

SWETHA, R. G.; RAMAIAH, S.; ANBARASU, A. R521C and R521H mutations in FUS result in weak binding with Karyopherin β 2 leading to Amyotrophic lateral sclerosis: a molecular docking and dynamics study. **Journal of Biomolecular Structure and Dynamics**, v. 35, n. 10, 2017.

TEILUM, K.; OLSEN, J. G.; KRAGELUND, B. B. Functional aspects of protein flexibility. **Cellular and Molecular Life Sciences**, v. 66, n. 14, p. 2231–2247, 2009.

TYZACK, G. E. *et al.* Widespread FUS mislocalization is a molecular hallmark of amyotrophic lateral sclerosis. **Brain**, v. 142, n. 9, p. 2572–2580, 2019.

UNIPROT CONSORTIUM. UniProt: the universal protein knowledgebase in 2021. **Nucleic Acids Research**, v. 49, n. D1, p. D480–D489, 2021.

VAN DER SPOEL, D. *et al.* GROMACS: Fast, flexible, and free. **Journal of Computational Chemistry**, v. 26, n. 16, p. 1701–1718, 2005.

VANCE, C. *et al.* Mutations in FUS, an RNA processing protein, cause familial amyotrophic lateral sclerosis type 6. **Science**, v. 323, n. 5918, p. 1208–1211, 2009.

YAMASHITA, S. *et al.* Sporadic juvenile amyotrophic lateral sclerosis caused by mutant FUS/TLS: possible association of mental retardation with this mutation. **Journal of Neurology**, v. 259, n. 6, p. 1039–1044, 2012.

YOSHIZAWA, T. *et al.* Nuclear Import Receptor Inhibits Phase Separation of FUS through Binding to Multiple Sites. **Cell**, v. 173, n. 3, p. 693-705.e22, 2018.

5. CONCLUSÃO

A maior parte dos algoritmos de predição funcional foi capaz de classificar as três mutações estudadas, conhecidamente associadas à ALS, como deletérias. Os resultados do SNPEffect mostraram que todas as três mutações diminuem a estabilidade da proteína, sendo que a mutação R521C apresentou uma maior diminuição da estabilidade da proteína e um aumento na tendência a ligação de chaperonas, diferente das demais mutações. A mutação R521C também mostrou alterações na análise da frustração proteica, onde foram observados níveis maiores de frustração entre o resíduo mutado (521) e os resíduos com que este interage. A análise da conservação estrutural mostrou que o sítio de localização nuclear é uma região conservada da proteína FUS, e que as três mutações ocorrem em sítios altamente conservados. Os dados obtidos através das dinâmicas moleculares indicaram que as mutações podem afetar a estabilidade, compactação e área de superfície da proteína, bem como sua flexibilidade.

Os dados obtidos através de técnicas computacionais demonstraram que as mutações podem causar alterações em vários aspectos da proteína, como estabilidade, compactação e flexibilidade. Isso, em conjunto com o elevado nível de conservação estrutural dos sítios onde ocorrem as mutações estudadas, corrobora com a hipótese de que estas mutações estão associadas ao desenvolvimento de ALS, visto que estas alterações podem impactar negativamente a interação entre a FUS mutada e Kap β 2, possivelmente levando à formação dos agregados citosólicos que são relacionados à ALS.

REFERÊNCIAS

- ABEL, O. *et al.* ALSod: A user-friendly online bioinformatics tool for amyotrophic lateral sclerosis genetics. **Human Mutation**, [s. l.], v. 33, n. 9, p. 1345–1351, 2012. Available at: <https://doi.org/10.1002/humu.22157>
- ADZHUBEI, I. A. *et al.* A method and server for predicting damaging missense mutations. **Nature Methods**, [s. l.], v. 7, n. 4, p. 248–249, 2010. Available at: <https://doi.org/10.1038/nmeth0410-248>
- AEBISCHER, P.; KATO, A. C. Playing Defense Against Lou Gehrig's Disease. **Scientific American**, [s. l.], v. 297, n. 5, p. 86–93, 2007. Available at: <https://doi.org/10.1038/scientificamerican1107-86>
- AGRAWAL, M.; ZITNIK, M.; LESKOVEC, J. Large-scale analysis of disease pathways in the human interactome. **Pacific Symposium on Biocomputing**, [s. l.], v. 23, p. 111–122, 2018. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/29218874>
- AL-CHALABI, A.; VAN DEN BERG, L. H.; VELDINK, J. Gene discovery in amyotrophic lateral sclerosis: implications for clinical management. **Nature Reviews Neurology**, [s. l.], v. 13, n. 2, p. 96–104, 2017. Available at: <https://doi.org/10.1038/nrneurol.2016.182>
- AMBERGER, J. S. *et al.* OMIM.org: Online Mendelian Inheritance in Man (OMIM®), an online catalog of human genes and genetic disorders. **Nucleic Acids Research**, [s. l.], v. 43, n. Database issue, p. D789-98, 2015. Available at: <https://doi.org/10.1093/nar/gku1205>
- AOKI, N. *et al.* Localization of fused in sarcoma (FUS) protein to the post-synaptic density in the brain. **Acta neuropathologica**, [s. l.], v. 124, n. 3, p. 383–394, 2012. Available at: <https://doi.org/10.1007/s00401-012-0984-6>
- ASHKENAZY, H. *et al.* ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. **Nucleic Acids Research**, [s. l.], v. 44, n. W1, p. W344–W350, 2016. Available at: <https://doi.org/10.1093/nar/gkw408>
- BANEGAS-LUNA, A. J. *et al.* Advances in distributed computing with modern

drug discovery. **Expert Opinion on Drug Discovery**, [s. l.], v. 14, n. 1, p. 9–22, 2019. Available at: <https://doi.org/10.1080/17460441.2019.1552936>

BATEMAN, A. *et al.* UniProt: A hub for protein information. **Nucleic Acids Research**, [s. l.], v. 43, n. D1, p. D204–D212, 2015. Available at: <https://doi.org/10.1093/nar/gku989>

BECKMANN, J. S.; LEW, D. Reconciling evidence-based medicine and precision medicine in the era of big data: challenges and opportunities. **Genome Medicine**, [s. l.], v. 8, n. 1, p. 134, 2016. Available at: <https://doi.org/10.1186/s13073-016-0388-7>

BENDL, J. *et al.* PredictSNP: Robust and Accurate Consensus Classifier for Prediction of Disease-Related Mutations. **PLoS Computational Biology**, [s. l.], v. 10, n. 1, p. 1–11, 2014. Available at: <https://doi.org/10.1371/journal.pcbi.1003440>

BENNION CALLISTER, J.; PICKERING-BROWN, S. M. Pathogenesis/genetics of frontotemporal dementia and how it relates to ALS. **Experimental Neurology**, [s. l.], v. 262 Pt B, p. 84–90, 2014. Available at: <https://doi.org/10.1016/j.expneurol.2014.06.001>

BREINER, A.; ZINMAN, L.; BOURQUE, P. R. Edaravone for amyotrophic lateral sclerosis: barriers to access and lifeboat ethics. **Canadian Medical Association Journal**, [s. l.], v. 192, n. 12, p. E319–E320, 2020. Available at: <https://doi.org/10.1503/cmaj.191236>

BROMBERG, Y.; YACHDAV, G.; ROST, B. SNAP predicts effect of mutations on protein function. **Bioinformatics**, [s. l.], v. 24, n. 20, p. 2397–2398, 2008. Available at: <https://doi.org/10.1093/bioinformatics/btn435>

BURATTI, E. *et al.* TDP-43 binds heterogeneous nuclear ribonucleoprotein A/B through its C-terminal tail: an important region for the inhibition of cystic fibrosis transmembrane conductance regulator exon 9 splicing. **The Journal of Biological Chemistry**, [s. l.], v. 280, n. 45, p. 37572–37584, 2005. Available at: <https://doi.org/10.1074/jbc.M505557200>

BURLEY, S. K. *et al.* Protein Data Bank (PDB): The Single Global Macromolecular Structure Archive. **Methods in Molecular Biology**, [s. l.], v.

1607, p. 627–641, 2017. Available at: https://doi.org/10.1007/978-1-4939-7000-1_26

CALABRESE, R. *et al.* Functional annotations improve the predictive score of human disease-related mutations in proteins. **Human Mutation**, [s. l.], v. 30, n. 8, p. 1237–1244, 2009. Available at: <https://doi.org/10.1002/humu.21047>

CAPRIOTTI, E.; CALABRESE, R.; CASADIO, R. Predicting the insurgence of human genetic diseases associated to single point protein mutations with support vector machines and evolutionary information. **Bioinformatics**, [s. l.], v. 22, n. 22, p. 2729–2734, 2006. Available at: <https://doi.org/10.1093/bioinformatics/btl423>

CHARCOT, J. M.; JOFFROY, A. Deux cas d'atrophie musculaire progressive avec lésions de la substance grise et des faisceaux antérolatéraux de la moelle épinière. **Archives de Physiologie Normale et Pathologique**, [s. l.], v. 2, p. 354–367, 745–760, 1869.

CHEN, C.; HUANG, H.; WU, C. H. Protein Bioinformatics Databases and Resources. **Methods in Molecular Biology**, [s. l.], v. 1558, p. 3–39, 2017. Available at: https://doi.org/10.1007/978-1-4939-6783-4_1

CHIÒ, A. *et al.* Early symptom progression rate is related to ALS outcome: a prospective population-based study. **Neurology**, [s. l.], v. 59, n. 1, p. 99–103, 2002. Available at: <https://doi.org/10.1212/wnl.59.1.99>

CHOI, Y.; CHAN, A. P. PROVEAN web server: A tool to predict the functional effect of amino acid substitutions and indels. **Bioinformatics**, [s. l.], v. 31, n. 16, p. 2745–2747, 2015. Available at: <https://doi.org/10.1093/bioinformatics/btv195>

CONTE, A. *et al.* P525L FUS mutation is consistently associated with a severe form of juvenile amyotrophic lateral sclerosis. **Neuromuscular Disorders : NMD**, [s. l.], v. 22, n. 1, p. 73–75, 2012. Available at: <https://doi.org/10.1016/j.nmd.2011.08.003>

CORCIA, P. *et al.* Genetics of amyotrophic lateral sclerosis. **Revue Neurologique**, [s. l.], v. 173, n. 5, p. 254–262, 2017. Available at: <https://doi.org/10.1016/j.neurol.2017.03.030>

CROZAT, A. *et al.* Fusion of CHOP to a novel RNA-binding protein in human myxoid liposarcoma. **Nature**, [s. l.], v. 363, n. 6430, p. 640–644, 1993. Available at: <https://doi.org/10.1038/363640a0>

DARDEN, T.; YORK, D.; PEDERSEN, L. Particle mesh Ewald: An $N \cdot \log(N)$ method for Ewald sums in large systems. **The Journal of Chemical Physics**, [s. l.], v. 98, n. 12, p. 10089–10092, 1993. Available at: <https://doi.org/10.1063/1.464397>

DE BAETS, G. *et al.* SNPeffect 4.0: On-line prediction of molecular and structural effects of protein-coding variants. **Nucleic Acids Research**, [s. l.], v. 40, n. D1, 2012. Available at: <https://doi.org/10.1093/nar/gkr996>

DEMATTE, L.; PRANDI, D. GPU computing for systems biology. **Briefings in Bioinformatics**, [s. l.], v. 11, n. 3, p. 323–333, 2010. Available at: <https://doi.org/10.1093/bib/bbq006>

DENG, H.; GAO, K.; JANKOVIC, J. The role of FUS gene variants in neurodegenerative diseases. **Nature Reviews Neurology**, [s. l.], v. 10, n. 6, p. 337–348, 2014. Available at: <https://doi.org/10.1038/nrneurol.2014.78>

DENG, J. *et al.* FUS interacts with ATP synthase beta subunit and induces mitochondrial unfolded protein response in cellular and animal models. **Proceedings of the National Academy of Sciences of the United States of America**, [s. l.], v. 115, n. 41, p. E9678–E9686, 2018. Available at: <https://doi.org/10.1073/pnas.1806655115>

DORMANN, D.; HAASS, C. Fused in sarcoma (FUS): an oncogene goes awry in neurodegeneration. **Molecular and Cellular Neurosciences**, [s. l.], v. 56, p. 475–486, 2013. Available at: <https://doi.org/10.1016/j.mcn.2013.03.006>

DORMANN, D.; HAASS, C. TDP-43 and FUS: A nuclear affair. **Trends in Neurosciences**, [s. l.], v. 34, n. 7, p. 339–348, 2011. Available at: <https://doi.org/10.1016/j.tins.2011.05.002>

FANG, F. *et al.* Amyotrophic lateral sclerosis and cancer: a register-based study in Sweden. **Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration**, [s. l.], v. 14, n. 5–6, p. 362–368, 2013. Available at: <https://doi.org/10.3109/21678421.2013.775309>

- FARG, M. A. *et al.* Mutant FUS induces endoplasmic reticulum stress in amyotrophic lateral sclerosis and interacts with protein disulfide-isomerase. **Neurobiology of Aging**, [s. l.], v. 33, n. 12, p. 2855–2868, 2012. Available at: <https://doi.org/10.1016/j.neurobiolaging.2012.02.009>
- FUJII, R.; TAKUMI, T. TLS facilitates transport of mRNA encoding an actin-stabilizing protein to dendritic spines. **Journal of Cell Science**, [s. l.], v. 118, n. Pt 24, p. 5755–5765, 2005. Available at: <https://doi.org/10.1242/jcs.02692>
- GAJULA, K. S. *et al.* High-throughput mutagenesis reveals functional determinants for DNA targeting by activation-induced deaminase. **Nucleic Acids Research**, [s. l.], v. 42, n. 15, p. 9964–9975, 2014. Available at: <https://doi.org/10.1093/nar/gku689>
- GUERRERO, G. D. *et al.* A performance/cost evaluation for a GPU-based drug discovery application on volunteer computing. **BioMed Research International**, [s. l.], v. 2014, p. 474219, 2014. Available at: <https://doi.org/10.1155/2014/474219>
- HARRACH, M. F.; DROSSEL, B. Structure and dynamics of TIP3P, TIP4P, and TIP5P water near smooth and atomistic walls of different hydroaffinity. **Journal of Chemical Physics**, [s. l.], v. 140, n. 17, 2014. Available at: <https://doi.org/10.1063/1.4872239>
- HESS, B. *et al.* LINCS: A linear constraint solver for molecular simulations. **Journal of Computational Chemistry**, [s. l.], v. 18, n. 12, p. 1463–1472, 1997. Available at: [https://doi.org/10.1002/\(SICI\)1096-987X\(199709\)18:12<1463::AID-JCC4>3.0.CO;2-H](https://doi.org/10.1002/(SICI)1096-987X(199709)18:12<1463::AID-JCC4>3.0.CO;2-H)
- HO, W. Y.; LING, S. C. Elevated FUS levels by overriding its autoregulation produce gain-of-toxicity properties that disrupt protein and RNA homeostasis. **Autophagy**, [s. l.], v. 15, n. 9, p. 1665–1667, 2019. Available at: <https://doi.org/10.1080/15548627.2019.1633162>
- HOLLINGER, S. K.; OKOSUN, I. S.; MITCHELL, C. S. Antecedent disease and amyotrophic lateral sclerosis: What is protecting whom? **Frontiers in Neurology**, [s. l.], v. 7, n. MAR, 2016. Available at: <https://doi.org/10.3389/fneur.2016.00047>

HUMPHREY, W.; DALKE, A.; SCHULTEN, K. Visual Molecular Dynamics.

Journal of Molecular Graphics, [s. l.], v. 14, p. 33–38, 1996.

Jl, A. L. *et al.* Genetics insight into the amyotrophic lateral

sclerosis/frontotemporal dementia spectrum. **Journal of Medical Genetics**, [s. l.], v. 54, n. 3, p. 145–154, 2017. Available at:

<https://doi.org/10.1136/jmedgenet-2016-104271>

JU, S. *et al.* A yeast model of FUS/TLS-dependent cytotoxicity. **PLoS Biology**,

[s. l.], v. 9, n. 4, p. e1001052, 2011. Available at:

<https://doi.org/10.1371/journal.pbio.1001052>

KÖNIG, I. R. *et al.* What is precision medicine? **The European Respiratory**

Journal, [s. l.], v. 50, n. 4, 2017. Available at:

<https://doi.org/10.1183/13993003.00391-2017>

KWIATKOWSKI, T. J. *et al.* Mutations in the FUS/TLS Gene on Chromosome

16 Cause Familial Amyotrophic Lateral Sclerosis. **Science**, [s. l.], v. 323, n.

5918, p. 1205–1208, 2009. Available at:

<https://doi.org/10.1126/science.1166066>

LINDORFF-LARSEN, K. *et al.* Improved side-chain torsion potentials for the

Amber ff99SB protein force field. **Proteins: Structure, Function and**

Bioinformatics, [s. l.], v. 78, n. 8, p. 1950–1958, 2010. Available at:

<https://doi.org/10.1002/prot.22711>

LÓPEZ-FERRANDO, V. *et al.* PMut: A web-based tool for the annotation of

pathological variants on proteins, 2017 update. **Nucleic Acids Research**, [s. l.],

v. 45, n. W1, p. W222–W228, 2017. Available at:

<https://doi.org/10.1093/nar/gkx313>

MORRIS, J. Amyotrophic Lateral Sclerosis (ALS) and Related Motor Neuron

Diseases: An Overview. **The Neurodiagnostic Journal**, [s. l.], v. 55, n. 3, p.

180–194, 2015. Available at: <https://doi.org/10.1080/21646821.2015.1075181>

MUNOZ, D. G. *et al.* FUS pathology in basophilic inclusion body disease. **Acta**

Neuropathologica, [s. l.], v. 118, n. 5, p. 617–627, 2009. Available at:

<https://doi.org/10.1007/s00401-009-0598-9>

NAKAYA, T.; MARAGKAKIS, M. Amyotrophic Lateral Sclerosis associated FUS mutation shortens mitochondria and induces neurotoxicity. **Scientific Reports**, [s. l.], v. 8, n. 1, 2018. Available at: <https://doi.org/10.1038/s41598-018-33964-0>

NAUMANN, M. *et al.* Phenotypes and malignancy risk of different FUS mutations in genetic amyotrophic lateral sclerosis. **Annals of Clinical and Translational Neurology**, [s. l.], v. 6, n. 12, p. 2384–2394, 2019. Available at: <https://doi.org/10.1002/acn3.50930>

NG, P. C.; HENIKOFF, S. Predicting deleterious amino acid substitutions. **Genome Research**, [s. l.], v. 11, n. 5, p. 863–874, 2001. Available at: <https://doi.org/10.1101/gr.176601>

NOBILE, M. S. *et al.* Graphics processing units in bioinformatics, computational biology and systems biology. **Briefings in Bioinformatics**, [s. l.], v. 18, n. 5, p. 870–885, 2017. Available at: <https://doi.org/10.1093/bib/bbw058>

NOLAN, M.; TALBOT, K.; ANSORGE, O. Pathogenesis of FUS-associated ALS and FTD: insights from rodent models. **Acta Neuropathologica Communications**, [s. l.], v. 4, n. 1, p. 99, 2016. Available at: <https://doi.org/10.1186/s40478-016-0358-8>

PARAKH, S.; ATKIN, J. D. Protein folding alterations in amyotrophic lateral sclerosis. **Brain Research**, [s. l.], v. 1648, n. Pt B, p. 633–649, 2016. Available at: <https://doi.org/10.1016/j.brainres.2016.04.010>

PARRA, R. G. *et al.* Protein Frustratometer 2: a tool to localize energetic frustration in protein molecules, now with electrostatics. **Nucleic Acids Research**, [s. l.], v. 44, n. W1, p. W356–W360, 2016. Available at: <https://doi.org/10.1093/nar/gkw304>

PAUL, S. M. *et al.* How to improve R&D productivity: the pharmaceutical industry's grand challenge. **Nature Reviews. Drug Discovery**, [s. l.], v. 9, n. 3, p. 203–214, 2010. Available at: <https://doi.org/10.1038/nrd3078>

PETROV, D. *et al.* ALS Clinical Trials Review: 20 Years of Failure. Are We Any Closer to Registering a New Treatment? **Frontiers in Aging Neuroscience**, [s. l.], v. 9, n. MAR, p. 68, 2017. Available at: <https://doi.org/10.3389/fnagi.2017.00068>

RABBITTS, T. H. *et al.* Fusion of the dominant negative transcription regulator CHOP with a novel gene FUS by translocation t(12;16) in malignant liposarcoma. **Nature Genetics**, [s. l.], v. 4, n. 2, p. 175–180, 1993. Available at: <https://doi.org/10.1038/ng0693-175>

RAMENSKY, V.; BORK, P.; SUNYAEV, S. Human non-synonymous SNPs: server and survey. **Nucleic Acids Research**, [s. l.], v. 30, n. 17, p. 3894–3900, 2002. Available at: <https://doi.org/10.1093/nar/gkf493>

ROWLAND, L. P. How amyotrophic lateral sclerosis got its name: the clinical-pathologic genius of Jean-Martin Charcot. **Archives of Neurology**, [s. l.], v. 58, n. 3, p. 512–515, 2001. Available at: <https://doi.org/10.1001/archneur.58.3.512>

SALMINEN, A. *et al.* Endoplasmic reticulum stress in age-related macular degeneration: trigger for neovascularization. **Molecular Medicine**, [s. l.], v. 16, n. 11–12, p. 535–542, 2010. Available at: <https://doi.org/10.2119/molmed.2010.00070>

SAXENA, S.; CABUY, E.; CARONI, P. A role for motoneuron subtype-selective ER stress in disease manifestations of FALS mice. **Nature Neuroscience**, [s. l.], v. 12, n. 5, p. 627–636, 2009. Available at: <https://doi.org/10.1038/nn.2297>

SHANG, Y.; HUANG, E. J. Mechanisms of FUS mutations in familial amyotrophic lateral sclerosis. **Brain Research**, [s. l.], v. 1647, p. 65–78, 2016. Available at: <https://doi.org/10.1016/j.brainres.2016.03.036>

SHI, Y. *et al.* Large-scale neural circuit mapping data analysis accelerated with the graphical processing unit (GPU). **Journal of Neuroscience Methods**, [s. l.], v. 239, p. 1–10, 2015. Available at: <https://doi.org/10.1016/j.jneumeth.2014.09.022>

STONE, E. A.; SIDOW, A. Physicochemical constraint violation by missense substitutions mediates impairment of protein function and disease severity. **Genome Research**, [s. l.], v. 15, n. 7, p. 978–986, 2005. Available at: <https://doi.org/10.1101/gr.3804205>

SUN, Z. *et al.* Molecular determinants and genetic modifiers of aggregation and toxicity for the ALS disease protein FUS/TLS. **PLoS Biology**, [s. l.], v. 9, n. 4, p. e1000614, 2011. Available at: <https://doi.org/10.1371/journal.pbio.1000614>

SUZUKI, N. *et al.* FUS/TLS-immunoreactive neuronal and glial cell inclusions increase with disease duration in familial amyotrophic lateral sclerosis with an R521C FUS/TLS mutation. **Journal of Neuropathology and Experimental Neurology**, [s. l.], v. 71, n. 9, p. 779–788, 2012. Available at: <https://doi.org/10.1097/NEN.0b013e318264f164>

SWETHA, R. G.; RAMAIAH, S.; ANBARASU, A. R521C and R521H mutations in FUS result in weak binding with Karyopherin β 2 leading to Amyotrophic lateral sclerosis: a molecular docking and dynamics study. **Journal of Biomolecular Structure and Dynamics**, [s. l.], v. 35, n. 10, 2017. Available at: <https://doi.org/10.1080/07391102.2016.1209130>

TAN, A. Y.; MANLEY, J. L. The TET family of proteins: functions and roles in disease. **Journal of molecular cell biology**, [s. l.], v. 1, n. 2, p. 82–92, 2009. Available at: <https://doi.org/10.1093/jmcb/mjp025>

VAN DAMME, P.; ROBBERECHT, W.; VAN DEN BOSCH, L. Modelling amyotrophic lateral sclerosis: Progress and possibilities. **DMM Disease Models and Mechanisms**, [s. l.], v. 10, n. 5, p. 537–549, 2017. Available at: <https://doi.org/10.1242/dmm.029058>

VAN DER SPOEL, D. *et al.* GROMACS: Fast, flexible, and free. **Journal of Computational Chemistry**, [s. l.], v. 26, n. 16, p. 1701–1718, 2005. Available at: <https://doi.org/10.1002/jcc.20291>

VANCE, C. *et al.* Mutations in FUS, an RNA processing protein, cause familial amyotrophic lateral sclerosis type 6. **Science**, [s. l.], v. 323, n. 5918, p. 1208–1211, 2009. Available at: <https://doi.org/10.1126/science.1165942>

VISSER, J.; DE JONG, J. M. B. V.; VISSER, M. d. The history of progressive muscular atrophy: Syndrome or disease? **Neurology**, [s. l.], v. 70, n. 9, p. 723–727, 2008. Available at: <https://doi.org/10.1212/01.wnl.0000302187.20239.93>

WIJESEKERA, L. C.; LEIGH, P. N. Amyotrophic lateral sclerosis. **Orphanet Journal of Rare Diseases**, [s. l.], v. 4, n. 1, p. 3, 2009. Available at: <https://doi.org/10.1186/1750-1172-4-3>

WOOLLER, S. K. *et al.* Bioinformatics in translational drug discovery. **Bioscience Reports**, [s. l.], v. 37, n. 4, 2017. Available at:

<https://doi.org/10.1042/BSR20160180>



YAMASHITA, S. *et al.* Sporadic juvenile amyotrophic lateral sclerosis caused by mutant FUS/TLS: possible association of mental retardation with this mutation. **Journal of Neurology**, [s. l.], v. 259, n. 6, p. 1039–1044, 2012. Available at: <https://doi.org/10.1007/s00415-011-6292-6>

YOSHIZAWA, T. *et al.* Nuclear Import Receptor Inhibits Phase Separation of FUS through Binding to Multiple Sites. **Cell**, [s. l.], v. 173, n. 3, p. 693-705.e22, 2018. Available at: <https://doi.org/10.1016/j.cell.2018.03.003>

ZHANG, Z. C.; CHOOK, Y. M. Structural and energetic basis of ALS-causing mutations in the atypical proline-tyrosine nuclear localization signal of the Fused in Sarcoma protein (FUS). **Proceedings of the National Academy of Sciences of the United States of America**, [s. l.], v. 109, n. 30, p. 12017–12021, 2012. Available at: <https://doi.org/10.1073/pnas.1207247109>

ZOU, Z.-Y. *et al.* Genetic epidemiology of amyotrophic lateral sclerosis: a systematic review and meta-analysis. **Journal of Neurology, Neurosurgery, and Psychiatry**, [s. l.], v. 88, n. 7, p. 540–549, 2017. Available at: <https://doi.org/10.1136/jnnp-2016-315018>

ANEXO 1 – Comprovante de submissão do artigo ao periódico PLoS ONE.

PLOS ONE  

HOME • LOGOUT • HELP • REGISTER • UPDATE MY INFORMATION • JOURNAL OVERVIEW • CONTACT US • SUBMIT A MANUSCRIPT • INSTRUCTIONS FOR AUTHOR: Role: **Author** Username: DeMesquita

Click 'File Inventory' to download the source files for the manuscript. Click 'Revise Submission' to submit a revision of the manuscript. If you Decline To Revise the manuscript, it will be moved to the Declined Revisions folder.

IMPORTANT: If your revised files are not ready to be submitted, do not click the 'Revise Submission' link.

Page: 1 of 1 (1 total submissions) Display 10 results per page.

# Action	Manuscript Number	Title	Initial Date Submitted	Date Revision Due	Current Status	View Decision
Action Links	PONE-D-21-22871	Molecular dynamics and protein frustration Analysis of Human Fused In Sarcoma Protein Variants in Amyotrophic Lateral Sclerosis Type 6: an In Silico approach	Jul 14 2021 2:03PM	Sep 9 2021 11:59PM	Revise	Major Revision

Page: 1 of 1 (1 total submissions) Display 10 results per page.