Universidade de São Paulo
Faculdade de Medicina de Ribeirão Preto
Programa de Pós Graduação em Genética

Análise estrutural de mutações na enzima GALNS associadas à Mucopolissacaridose IVA
utilizando a técnica de modelagem comparativa.

Érico Torrieri

Ribeirão Preto
2015

Universidade de São Paulo
Faculdade de Medicina de Ribeirão Preto
Programa de Pós Graduação em Genética

Análise estrutural de mutações na enzima GALNS associadas à Mucopolissacaridose IVA utilizando a técnica de modelagem comparativa.

Dissertação apresentada ao programa de Pós Graduação em Genética, da Faculdade de Medicina de Ribeirão Preto, da Universidade de São Paulo para a obtenção do título de Mestre em Ciências.

Área de concentração: Genética

Orientadora: Prof.^a Dr^a. Silvana Giuliatti

Ribeirão Preto
2015

FOLHA DE APROVAÇÃO

Análise estrutural de mutações na enzima GALNS associadas à Mucopolissacaridose IVA utilizando a técnica de modelagem comparativa.

Dissertação apresentada ao programa de Pós Graduação em Genética, da Faculdade de Medicina de Ribeirão Preto, da Universidade de São Paulo para a obtenção do título de Mestre em Ciências.

Área de concentração: Genética

Aprovada em:

Banca examinadora

Prof. Dr.

Instituição: _____

Julgamento: _____ Assinatura: _____

Prof. Dr.

Instituição: _____

Julgamento: _____ Assinatura: _____

Prof. Dr.

Instituição: _____

Julgamento: _____ Assinatura: _____

Autorizo a reprodução e divulgação total ou parcial deste trabalho, por qualquer meio convencional ou eletrônico, para fins de estudo e pesquisa desde que seja citada a fonte.

Catálogo de Publicação Preparada pela Biblioteca do Serviço de
Biblioteca de Documentação da Faculdade de Medicina de Ribeirão Preto da
Universidade de Ribeirão Preto

Torrieri, Érico

Análise estrutural de mutações na enzima GALNS associadas a
Mucopolissacaridose IVA utilizando a técnica de modelagem comparativa.

Érico Torrieri – Ribeirão Preto, 2015

73 p.

Dissertação de mestrado apresentada ao Departamento de Genética da Faculdade de Medicina
de Ribeirão Preto, Universidade de São Paulo - USP.

Orientadora: Prof.^a Dr^a. Silvana Giuliatti

MPS IVA, GALNS, mutações *missense*, modelagem por homologia

Dedicatória

Aos meus pais que sempre me incentivaram
e fizeram diversos sacrifícios para que
eu pudesse chegar onde cheguei.

Agradecimentos

Agradeço em especial a Professora Doutora **Silvana Giuliatti** pela orientação, compreensão, confiança, amizade, paciência, direcionamento e sabedoria.

Aos meus pais Sirley Carmem Ferreira da Silva Torrieri e Silvio Torrieri, pelo amor incondicional, compreensão e incentivo.

Ao meu irmão Raul Torrieri e sua esposa Patrícia Ruy, pelo incentivo e amizade.

À minha avó Maria das Dores Costa Val, pelo amor, carinho e incentivo.

À Priscilla, pela amizade, carinho, companheirismo e paciência.

Aos amigos que fiz durante o período em que estive desenvolvendo meu trabalho, pelos momentos divididos juntos, pelas risadas, apoio e pela amizade que será para a vida toda.

Aos colegas de laboratório Elvira R. Tamarozzi, Evandro Semighini Fabiano Santos, Iuliana Rodrigues, Maria Gabriela Rodrigues e Thaís C. Arns pelo companheirismo, apoio, amizade e pelas longas discussões filosóficas.

À secretária do programa de Pós-Graduação em Genética, Susie Nalon, pela disponibilidade, suporte técnico e acadêmico.

Aos membros da banca pelas contribuições, críticas e sugestões.

Ao programa de Pós-graduação em Genética em especial ao então coordenador Dr. Ademilson Espencer Racci Soares.

À USP e a Faculdade de Medicina, Ribeirão Preto pela oportunidade da realização do meu curso de Mestrado.

Aos órgãos de fomento CAPES e FAEPA pelo apoio financeiro durante o desenvolvimento do projeto

“A dúvida é o principio da sabedoria.”
Aristóteles

Resumo

As Mucopolissacaridoses (MPS) são um grupo de doenças de armazenamento lisossômico causadas por deficiência de enzimas que catalisam a degradação gradual das glicosaminoglicanas (GAGs). GAGs (anteriormente chamadas de mucopolissacarídeos) são produtos de degradação das proteoglicanas que existem na matriz extracelular e tem efeito proteolítico. A classificação das MPS é baseada na deficiência enzimática específica. A MPS IVA é causada por mutações no gene que codifica a enzima GALNS (N-acetilgalactosamina-6-sulfatase), a qual desempenha um papel crucial na degradação do sulfato de queratano e *condroitina-6-sulfatase*. As mutações na enzima se resumem em três categorias: interrupção do sítio ativo, alterações no núcleo hidrofóbico e exposição da superfície, onde mutações *missense* na estrutura podem afetar gravemente a atividade da proteína GALNS, alterando seu núcleo hidrofóbico ou modificando seu enovelamento (*folding*). Com a falta de tratamentos efetivos, sendo em sua maioria paliativos, e tendo como base a estrutura já resolvida da GLANS selvagem, este trabalho teve como objetivo modelar 3 variantes da enzima GALNs, sendo uma mutação no sítio ativo, uma no núcleo hidrofóbico e uma na superfície. Foi usado o software MODELLER 9.12 para a modelagem comparativa, os *softwares* Prochek, PROSA II, ERRATv2, Verify3d, ProQ para a avaliação dos modelos, o *software* NAND 2.10, para simulação de dinâmica molecular e o *software* Chimera 1.10.1 para cálculo de superfícies eletrostáticas e hidrofobicidade da superfície. Os modelos apresentaram bons resultados segundo os *softwares* de avaliação e análise visual. Apresentaram poucas diferenças estruturais em relação à estrutura da GALNS selvagem, demonstraram estabilidade em simulação de dinâmica molecular. Entretanto, algumas diferenças foram observadas com relação à distribuição de cargas e hidrofobicidade no sítio ativo do modelo da variante com mutação no sítio ativo. Pôde ser concluído que as 3 mutações analisadas não causaram alterações estruturais significativas, não interferiram na estabilidade estrutural em simulação de dinâmica molecular, entretanto, foi demonstrado que mutações na região do sítio ativo podem interferir na função da enzima.

Palavras Chave: MPS IVA, GALNS, mutações *missense*, modelagem por homologia

Abstract

The Mucopolysaccharidoses (MPS) are a group of lysosomal storage diseases caused by deficiencies in enzymes that catalyze the gradual glycosaminoglycans (GAGs) degradation. GAGs (formerly called mucopolysaccharides) are products of proteoglycan degradation that exist in the extracellular matrix and have proteolytic effect. The classification of MPS is based on the specific enzyme deficiency. MPS IVA is caused by mutations in the gene that encodes the GALNS enzyme (N-acetylglucosaminase-6-sulfatase), which plays a crucial role in the degradation of keratan sulfate and chondroitin-6-sulfatase. Mutations in the enzyme can be summarized in three categories: interruption of the active site, changes in the hydrophobic core and display surface, where missense mutations in the structure can seriously affect the activity of GALNS protein, changing its hydrophobic core or modifying its folding. With the lack of effective treatments, in its most palliative, and based on the wild GALNS structure already determined, this study aimed to model 3 variants of GALNS enzyme, a mutation in the active site, one in the hydrophobic core and a on the surface. 9.12 MODELLER was used for comparative modeling software, the software Procheck, Prose II, ERRATv2, Verify3d, ProQ models for the evaluation of the NAND 2.10 software, for molecular dynamics simulation and software Chimera 1.10.1 calculates electrostatic and hydrophobic surface. The models showed good results according to the evaluation software and visual analysis. Presented few structural differences from the wild GALNS structure and showed stability in molecular dynamics simulation. However, some differences were observed with respect to the charge distribution and hydrophobicity in the active site of the variants of the model with a mutation in the active site. It might be concluded that the three mutations analyzed did not cause significant structural changes and did not affect the structural stability in molecular dynamics simulation, however, it has been shown that mutations in the active site region may interfere with the function of this enzyme.

Keywords: MPS IVA, GALNS, Missense mutations, Homology modeling.

Lista de Figuras

Figura 1. Representação da estrutura tridimensional da enzima GALNS destacando as estruturas secundárias.....	24
Figura 2. Fluxograma que apresenta cada passo da metodologia utilizada no presente trabalho.....	28
Figura 3. Representação química das mutações <i>missense</i> selecionadas para o estudo.....	31
Figura 4. Gráfico de Ramachandran de um dos modelos gerados a partir da variante A107T.....	36
Figura 5. Representação tridimensional dos modelos selecionados para cada uma das variantes.....	37
Figura 6. Gráfico comparando a posição do Z-score da variante A107T em relação a Z-scores de preteínas determinadas experimentalmente de tamanho semelhante.....	39
Figura 7. A esquerda é demonstrado o gráfico de energia para a GALNS selvagem. A direita é demonstrado o gráfico de energia da variante A107T.....	40
Figura 8. Representação tridimensional da sobreposição da GALNS selvagem em azul com os modelos das variantes em verde.....	42
Figura 9. Demonstração da maior distância encontrada entre carbonos alfa da estrutura selvagem e uma variante.....	43
Figura 10. Demonstração das distâncias entre os aminoácidos I.....	44
Figura 11. Demonstração das distâncias entre os aminoácidos II.....	45
Figura 12. Demonstração das distâncias entre os aminoácidos III.....	46
Figura 13. Comparação entre as posições no espaço tridimensional dos aminoácidos da estrutura da GALNS selvagem e as variantes.....	47
Figura 14. Comparação entre as posições no espaço tridimensional dos aminoácidos da estrutura da GALNS selvagem e as variantes, demonstrando que não houve alterações nas estruturas secundárias.....	48
Figura 15. Representação da variante C79Y em uma caixa d'água, para a realização da simulação de dinâmica molecular.....	49
Figura 16. Gráfico do RMSD da trajetória dos átomos das variantes ao longo de 25 ns..	50
Figura 17. Comparação entre os sítios ativos GALNS selvagem com as variantes com relação à superfície eletrostática.....	52

Figura 18. Comparação entre os sítios ativos GALNS selvagem com as variantes com relação a polaridade da superfície.....	53
---	----

Lista de Tabelas

Tabela 1. Relação entre os tipos de MPS, enzimas deficientes e manifestações clínicas específicas.....	17
Tabela 2. Apresentação das mutações <i>missense</i> na enzima GALNS divididas em 3 grupos com 78 mutações na região do núcleo hidrofóbico, 32 mutações na superfície e 6 mutações no sítio ativo.....	30
Tabela 3. Resultados dos softwares de avaliação, com relação aos seus <i>scores</i> . Os resultados dos programas para cada variante são comparados com os <i>scores</i> considerados bons para cada programa.....	38
Tabela 4. Quantidade de desvios estruturais dos modelos das variantes em comparação com a estrutura tridimensional da GALNS selvagem agrupados em 3 categorias: presentes no núcleo hidrofóbico, presentes na superfície e presentes no sítio ativo.....	43

Lista de Siglas

C6S – *Condroitina-6-sulfatase*

GAG – Glicosaminoglicana

GALNS – N-acetilgalactosamina 6-sulfatase

MPS – Mucopolissacaridoses

MPS IVA – Mucopolissacaridose IVA

PDB – *Protein Data Bank*

RMSD – *Root mean square deviation*

RNM – Ressonância Magnética Nuclear

SC – Sulfato de condroitina

SD – Sulfato de dermatano

SE – Superfície eletrostática

SH – Sulfato de heparano

SQ – Sulfato de queratano

SM – Síndrome de Morquio

Lista de Símbolos

Å – Angstroms

fs – femto segundos

ns – nano segundos

pb – pares de bases

Sumário

1. Introdução.....	16
1.1. Mucopolissacaridoses	16
1.2. Mucopolissacaridose IVA	20
2. Justificativa	26
3. Objetivo	27
3.1. Objetivo Geral.....	27
3.2. Objetivo Específico	27
4. Materiais e Métodos	28
4.1. Obtenção do Molde	29
4.2. Obtenção de três variantes da GALNS	29
4.3. Modelagem por homologia	31
4.4. Avaliação dos modelos obtidos	32
4.5. Inspeção Visual	33
4.6. Simulações de Dinâmica Molecular	33
4.7. <i>Cálculo da superfície eletroestática</i>	33
4.8. Cálculo da hidrofobicidade da superfície.....	34
5. Resultados e Discussão	35
5.1. Modelagem Molecular das 3 variantes da enzima GALNS.....	35
5.2. Avaliação dos Modelos.....	37
5.3. Inspeção visual	40
5.4. Análise de Dinâmica molecular	49
5.5. Cálculo da superfície eletrostática	51
5.6. Cálculo hidrofobicidade da superfície	52
6. Conclusão	56
7. Referências.....	57
8. Anexo	63

1. Introdução

1.1. *Mucopolissacaridoses*

As mucopolissacaridoses (MPS) são doenças genéticas raras causadas pela deficiência de enzimas lisossômicas que afetam o catabolismo de glicosaminoglicanas (GAGs) também chamadas de mucopolissacarídeos, levando ao armazenamento lisossômico (Palmucci et al., 2013). As GAGs são componentes dos tecidos conjuntivos, ligadas às proteínas, formadas por cadeias de dissacarídeos que se repetem sequencialmente e se unem em uma proteína central, constituindo moléculas mais complexas denominadas proteoglicanas (Byers et al., 1998). Proteoglicanas são proteínas glicosiladas constituintes da matriz extracelular cuja função é controlar o fluxo de água e nutrientes para as células. Por sua vez, existem cinco diferentes tipos de GAGs que sofrem degradação lisossomal: sulfato de condroitina (SC), sulfato de dermatano (SD), sulfato de queratano (SQ), sulfato de heparano (SH) e ácido hialurônico. Estas GAGs compõem as proteoglicanas, sendo que cada uma tem uma única estrutura constituída de unidades repetidas de dissacarídeos (Leone et al., 2014).

Segundo revisão de Leone e colaboradores (2014), o primeiro caso de MPS foi descrito em 1917 por Hunter. Porém, somente em meados do século XX que estas doenças foram classificadas como MPS, devido ao isolamento de mucopolissacarídeos de tecidos e urina de pacientes (Leone et al., 2014).

A classificação das MPS é baseada na deficiência enzimática específica, onde são encontrados mutações em 11 enzimas diferentes responsáveis pela degradação das GAGs que causam 7 tipos diferentes de MPS (I, II, III, IV, VI, VII e IX) (Coutinho et al., 2012). Neste sentido, existem diferentes fenótipos clínicos para a mesma deficiência enzimática, bem como, fenótipos semelhantes para deficiências enzimáticas diferentes (Cimaz & La Torre, 2014). Esta relação entre as enzimas e os subtipos de MPS é apresentada na Tabela 1.

Tabela 1. Relação entre os tipos de MPS, enzimas deficientes e manifestações clínicas específicas. Adaptado de Leone et al., 2014.

Tipo	Epônimo	Enzima deficiente	Manifestação clínica
I H	Hurler	α -L-idrunidase	Distúrbio mental, hidrocefalia, dimorfismo facial, viceromegalia, rigidez articular, opacificação da
I H/s	Hurler-Schie	α -L-idrunidase	córnea (aumento da severidade em casos de Hurler-Shey)
I S	Schie	α -L-idrunidase	Rigidez e contraturas articulares
II	Hunter	Iduronatesulfatase	Quadro clínico da Hurler, regressão psicomotora, rigidez articular.
III A	Sanfilippo A	Heparano N-sulfatase	
III B	Sanfilippo B	α -N-acetilglucosaminide Aciltransferase	
III C	Sanfilippo C	Acetil-CoA α -glucosaminide aciltransferase	Comprometimento cognitivo grave, sérias modificações ósseas com envolvimento articular
III D	Sanfilippo D	N-Acetylglucosamine-6-sulphatase	
IV A	Morquio A	N-Acetylglucosamine-6-sulphatase	Displasia esquelética grave e distintiva, frouxidão ligamentar, baixa estatura, doença valvular cardíaca, habilidades mentais mantidas
IV B	Morquio B	β -Galactosidase	
VI	Maroteux-Lamy	N-Acetylglucosamine-4-sulfatase (arryl-sulfatase B)	Displasia esquelética grave, baixa estatura, valvulopatia cardíaca, leve a severa opacificação da córnea, cognição preservada
VII	Sly	β -Glucuronidase	Hidropisia fetal ao nascimento, displasia esquelética, baixa estatura, opacificação da córnea, atraso no desenvolvimento, hepatomegalia
IX	Natowicz	Hyaluronidase	Nódulos macios, rigidez articular, baixa estatura, dimorfismo facial leve, cognição preservada

As MPS são clinicamente heterogêneas, caracterizadas pela deficiência em uma das várias enzimas lisossômicas necessárias para a quebra intracelular das GAGs (Byers et al., 1998), causando a deterioração progressiva das funções viscerais, esqueléticas e

neurológicas no paciente (Noh & Lee, 2014). Além das características enzimáticas, outras como as patológicas são encontradas em portadores da doença. Essas características são causadas pelo acúmulo intralisossomal das GAGs não degradadas em vários órgãos e tecidos nos pacientes afetados, incluindo: coração, fígado, baço, rim, sistema respiratório, ossos, articulações, sistema nervoso e sistema nervoso central (Tomatsu et al., 2014). O aumento de GAGs na urina, sangue (Tomatsu et al., 2005a) e líquido cefalorraquidiano (Munoz-Rojas et al., 2008; Federico et al., 1981), resulta em uma série de sinais e sintomas integrantes de um quadro clínico multissistêmico que, se não tratado, leva à disfunções celulares, à lesão progressiva dos tecidos afetados e à falência múltipla dos órgãos, com consequente redução de expectativa de vida (Cimaz & La Torre, 2014; Muenzer & Fisher, 2004).

A transmissão das MPS é feita de forma autossômica e recessiva, exceto para a MPS do tipo II (Síndrome de Hunter) na qual a herança é ligada ao X (Muenzer, 2004; Muenzer & Fisher, 2004), sua incidência na população é rara, com 1 caso à cada 25000 nascidos vivos (Palmucci et al., 2013) afetando igualmente homens e mulheres (Cimaz & La Torre, 2014), mas esse dado é provavelmente subestimado, porque as formas mais leves de MPS muitas vezes passam despercebidas.

SINTOMAS

As MPS ocasionam uma grande variedade de apresentações clínicas. A maioria dos pacientes com MPS é assintomático quando recém-nascidos, podendo posteriormente, apresentar sintomas clínicos que incluem hérnia inguinal ou umbilical, desenvolvimento anormal dos ossos, baixa estatura, hepatoesplenomegalia, e distúrbios mentais (Tabela 1). Essas manifestações clínicas podem progredir ao longo de vários anos causando o comprometimento de órgãos e tecidos. Dessa forma, pacientes com MPS frequentemente necessitam de intervenções cirúrgicas que podem levar a um alto índice de complicações (Hendriksz et al., 2015; Ard et al., 2005). Na maioria dos pacientes com o fenótipo grave, os sintomas surgem após o nascimento e progridem rapidamente, enquanto nos fenótipos atenuados, se manifestam na infância ou adolescência (Lehman et al., 2011).

Os sintomas e a gravidade das MPS variam de acordo com cada paciente e seu subtipo de MPS. Por isso, um diagnóstico inicial da doença através dos sintomas se faz necessário, uma vez que a média de vida na maioria dos pacientes sem tratamento é apenas de uma a duas décadas (Tomatsu et al., 2014).

DIAGNÓSTICO

A determinação qualitativa e quantitativa das GAGs urinárias ocorre por um teste de triagem simples para os pacientes com sinais clínicos sugestivos de MPS. Geralmente é realizada uma combinação do quadro clínico e análise das GAGs urinárias para alcançar o diagnóstico de MPS. Um resultado positivo é muito sugestivo para MPS, mas os resultados falso-negativos são também comuns. Eles ocorrem por causa da baixa sensibilidade dos diferentes ensaios e devido as amostras serem muito diluídas. Assim, uma análise de GAG urinária negativa não exclui a doença (Mahalingam et al., 2004). Diagnósticos definitivos geralmente são possíveis por meio da medição de atividade enzimática em cultura de fibroblastos ou leucócitos (Coutinho et al., 2012). Portanto, a identificação de biomarcadores da doença é importante no diagnóstico, identificação da gravidade clínica, prognóstico, patogênese e monitoramento de terapias (Tomatsu et al., 2014). Neste sentido, a identificação do genótipo pode ser importante para prever o fenótipo, tomar decisões terapêuticas, para possibilitar o aconselhamento genético para a família e para auxiliar no diagnóstico pré-natal (Giugliani et al., 2010). Complementarmente, a análise de imagens radiológicas também pode facilitar o diagnóstico precoce e seleção de um tratamento eficaz (Leone et al., 2014).

TRATAMENTO

As MPS são doenças de natureza progressiva, onde pacientes afetados necessitam de suporte regular multidisciplinar para resolver uma série de complicações sistêmicas. Atualmente, os tratamentos indicados são: terapia enzimática de substituição (ERT) (Muenzer, 2014), o transplante de células-tronco hematopoiéticas (TCTH) (Prasad & Kurtzberg, 2010), a terapia de redução de substrato (SRT) (Jakóbkiewicz-Banecka et al., 2011), a terapia gênica (Baldo et al., 2014) e drogas anti-inflamatórias (Simonaro et al., 2010). O início do tratamento logo após o nascimento ou durante os primeiros estágios da doença fornece a maioria dos benefícios na melhora clínica da doença. Portanto, o sucesso do tratamento destas doenças depende do diagnóstico precoce.

1.2. *Mucopolissacaridose IVA*

A MPS IV é uma doença autossômica recessiva, adquirida de pais sadios heterozigotos portadores de um gene mutado. Existem duas formas reconhecidas: tipo A, caracterizada por uma deficiência na atividade da enzima N-acetilgalactosamina 6-sulfato-sulfatase (GALNS) que leva a degradação insuficiente de SQ e condroitina-6-sulfatase (C6S), e tipo B, caracterizada pela deficiência da enzima β -galactosidase (β - GAL) levando a degradação insuficiente de SQ (Rivera-Colón et al., 2012; Kubaski et al., 2013; Northover et al., 1996; Caciotti et al., 2011). Este trabalho está focado no estudo da MPS IVA, também conhecida como Síndrome de Morquio (SM).

Inicialmente, em 1929 a SM foi descrita pelo pediatra uruguaio Morquio e pelo radiologista inglês Ulfrich como nanismo com malformações ósseas (Hendriksz et al., 2013; Chaves et al., 2003). Atualmente dados epidemiológicos mostram a incidência na população mundial 1: 201.000 nascidos vivos (Hendriksz et al., 2013). Porém, a SM ainda é confundida com nanismo, ocasionando dados epidemiológicos incertos e inexatos. Neste sentido, são necessários mais estudos em populações diversificadas sobre a epidemiologia da MPS IVA.

A existência heterogenia da MPS IVA com relação a genótipos e fenótipos, proporciona uma gravidade variável da doença. As manifestações clínicas da doença dependem da atividade residual da GALNS, taxa de acúmulo e localização do material armazenado nos tecidos. Estudos mostraram semelhança nas manifestações clínicas entre irmãos com MPS IVA, fornecendo evidências para correlações genótipo-fenótipo (Hendriksz et al., 2015). Deste modo, pacientes são aparentemente normais ao nascer, em seguida, começam a desenvolver sintomas a taxas diferentes, dependentes da gravidade da sua doença (Montano et al., 2007). Pacientes severamente afetados, apresentaram sinais e sintomas durante o primeiro ano de vida e geralmente não sobrevivem após a segunda ou terceira década de vida. Já os pacientes com doença atenuada, podem não desenvolver sintomas iniciais até o fim da infância ou adolescência (Montano et al., 2007), podendo sobreviver até a sétima década de vida (Dũng et al., 2013).

Crianças com MPS IVA costumam crescer normalmente nos primeiros meses de suas vidas, mas o crescimento começa a desacelerar por volta dos 18 meses. Aqueles que são gravemente afetados geralmente param de crescer por volta dos oito anos de idade. O tronco é relativamente mais curto que os membros, a boca alargada, a mandíbula quadrada e a ponte nasal achatada, o pescoço é muito curto, mas a textura dos cabelos

não é afetada como acontece nas outras MPS. Na MPS IVA, o crescimento da coluna fica prejudicado nesses indivíduos. O osso esterno se desenvolve normalmente, porém é pressionado na junção com as costelas, deformando e estendendo-se para fora do peito. Esta condição recebe o nome de “peito de pombo” ou do latim *pectuscarinatum*. Outras alterações comuns são cifose, hiperlordose lombar, escoliose, deformidades ovoides em todas as vértebras, valgismo de joelho, pé plano e valgo bilateralmente, hiperextensão das articulações (principalmente de punhos) e hipoplasiaodontogênica. Apresentam ainda alterações extraesqueléticas, sendo as principais: resistência limitada, deficiência visual e auditiva, desafios de saúde bucal, alterações cardiovasculares, e significativo comprometimento do sistema respiratório (Wood et al., 2013) e ao contrário de outras MPS, não existe nenhum envolvimento do sistema nervoso central na MPS IVA, assim, as funções cognitivas são preservadas (Hendriksz et al., 2013; Tomatsu et al., 2011; Couprie et al., 2010; Leal et al., 2010; Montano et al., 2007).

Historicamente, exames de GAGs totais são realizados na urina, no entanto, como o exame urinário é considerado como uma ferramenta de rastreio, ele pode apresentar resultados falsos negativos. Esses falsos negativos podem surgir porque as avaliações são feitas tanto para a quantidade total de GAG excretada quanto para tipos específicos. A realização desses dois testes simultâneos evita resultados falsos negativos como no caso de níveis altos de KS estarem presentes sem elevar os limites da quantidade total de GAGs para indivíduos não afetados (Hendriksz et al., 2013). Outro fator que pode levar a resultados falsos negativos é que a excreção urinária de GAGs em pacientes afetados é maior em recém nascidos e crianças jovens. Já em pacientes próximos da puberdade, a taxa de acumulação diminui podendo até mesmo desaparecer, mantendo-se constante até a idade adulta (Hendriksz et al., 2013; Wood et al., 2013). Portanto, esses resultados falso-negativos podem ocorrer conforme o avanço da idade do paciente. Assim, recomenda-se que os testes enzimáticos ou moleculares sejam usados para confirmar ou descartar o diagnóstico de MPS IVA.

O teste de atividade enzimática exige a demonstração da deficiência na atividade da enzima GALNS, assim, sua medição é uma avaliação laboratorial comum para o diagnóstico da doença. A enzima é ativa em leucócitos e fibroblastos, portanto, qualquer tipo de amostra é aceitável para diagnóstico. A avaliação da atividade enzimática é normalmente realizada utilizando radioativos artificiais ou o substrato fluorogênico (Hendriksz et al., 2013).

A análise molecular, pode ser utilizada para confirmar os resultados da atividade enzimática e facilitar o aconselhamento genético da família. Portanto, objetivo dessa técnica é identificar mutações genéticas que resultam em atividade enzimática diminuída ou ausente da GALNS (Wood et al., 2013).

Para chegar a um diagnóstico preciso, é fundamental que os resultados de ensaios de diferentes tipos de amostras corroborem com o diagnóstico. Qualquer discrepância deve ser investigada e resolvida. O processo de diagnóstico completo envolve a triagem seguida por análise de atividade enzimática em fibroblastos ou leucócitos e por fim, confirmação por análise molecular. O método de rastreio selecionado muitas vezes depende da disponibilidade de testes e de amostras.

O gene humano que codifica a GALNS, foi mapeado no cromossomo 16q24.3, que abrange aproximadamente 50000 pares de bases, sendo constituído por 13 íntrons e 14 éxons que codificam uma proteína de 522 aminoácidos (Olarte-Avellaneda et al., 2014; Hendriksz et al., 2013). Até 2012, foi relatado um total de 157 mutações, das quais 126 eram mutações *missense* (Rivera-Colón et al., 2012), sendo assim, as mais prevalentes entre as mutações na enzima GALNS (Tomatsu et al., 2005b). Anteriormente, foi demonstrado que estas mutações afetam gravemente a atividade da proteína GALNS, alterando seu núcleo hidrofóbico ou modificando sua dobragem, mas não a sua estrutura secundária (Tomatsu et al., 2005b).

A MPS IVA é uma doença pouco estudada, sendo que, escassos trabalhos tentaram elucidar a associação entre fenótipo e genótipo. Outros trabalhos abordaram estudos na tentativa de entender a associação entre mutações *missense* na enzima GALNS e fenótipos graves da MPS IVA. Neste contexto, Sukegawa e colaboradores (2000), avaliaram a atividade da enzima GALNS para 32 mutações *missense* encontradas em pacientes, onde o fenótipo grave da doença foi correlacionado com a baixa ou inexistente atividade da enzima. Em pacientes com fenótipo moderado, a atividade também foi baixa. Por outro lado, pacientes com fenótipo leve apresentaram atividade enzimática moderada. O estudo de Tomatsu e colaboradores (2006) também pode ser destacado, uma vez que estudaram as associações entre os fenótipos clínicos causados por mutações *missense* na enzima GALNS e os níveis de conservação evolutiva. Foram avaliadas 110 mutações *missense* diferentes, identificadas em 190 pacientes com MPS IVA que foram classificados de acordo com sua altura: grave se a altura fosse menor que 1,25 metros e atenuado caso a altura fosse maior que 1,25 metros. A sequência da GALNS humana foi comparada a outras de diversos animais (cão, vaca, porco, rato, peixe-zebra, baiacu, sapo, galinha, e

ouriço-do-mar) a fim de identificar níveis de conservação evolutiva. Assim, foi possível associar mutações em regiões conservadas com fenótipos graves (Tomatsu et al., 2006).

Alguns estudos *in silico* com o mesmo objetivo também podem ser destacados, como é o caso do trabalho já citado de Sukegawa e colaboradores (2000), que além da medição de atividade enzimática, modelaram a enzima GALNS através da metodologia de modelagem por homologia tendo como moldes a *N*-acetilgalactosamina-4-sulfatase e arilsulfatase A, que tem, respectivamente, 24 e 29% de identidade com a GALNS. Assim, foi possível encontrar as localizações das mutações na estrutura tridimensional, correlacionando-as com os tipos de fenótipos. Foi encontrado que a maior parte das mutações responsáveis por causar fenótipos graves leva à destruição do núcleo hidrofóbico ou modificação no sítio ativo. Neste sentido, outro estudo *in silico* realizado por Sudhakar e colaboradores (2011) utilizou diversas ferramentas de bioinformática, para encontrar a estrutura 3D da GALNS. Utilizando o método de modelagem comparativa tendo como molde a *N*-acetilgalactosamina-4-sulfatase, foi possível constatar que mutações *missense* podem causar diversas alterações estruturais e funcionais na GALNS. Recentemente, um estudo *in silico* publicado em 2014, teve como princípio elucidar o efeito das mutações na enzima GALNS. A proteína foi modelada computacionalmente usando o método de modelagem por homologia, tendo como molde a enzima arilsulfatase A humana, que é uma enzima lisossomal com mecanismo de ação semelhante a GALNS. Também, foi avaliado qual seria o efeito da ligação das enzimas mutadas com o substrato, através da introdução manual das mutações na estrutura tridimensional da GALNS. Com isso, constataram que as mutações que geram os fenótipos mais graves se ligam com menor efetividade ao substrato (Olarte-Avellaneda et al., 2014).

Porém, Rivera Colón e colaboradores (2012) resolveram a estrutura terciária da enzima GALNS através da técnica de cristalografia de raios-X (Figura 1). Através da localização das mutações na estrutura, foi demonstrado que as mutações *missense* são resumidas em três categorias: alterações no sítio ativo, alterações no núcleo hidrofóbico e alterações na superfície da enzima. Neste mesmo estudo foram descritas as mutações como sendo apenas 6 (5%) que afetam o sítio ativo, enquanto 78 (65%) afetam o núcleo hidrofóbico e 32 mutações (27%) causam alterações da superfície da enzima.

Figura 1. Representação da estrutura tridimensional da enzima GALNS destacando as estruturas secundárias. Figura gerada pelo software Pymol (The PyMOL Molecular Graphics System, 2010)

Dessa forma, pode-se identificar o interesse no conhecimento das estruturas tridimensionais das variantes geradas por mutações *missense* na enzima GALNS, para que através de comparações com a GALNS selvagem, seja possível avaliar o quão severas são as alterações causadas por essas mutações na estrutura tridimensional da enzima. Esse tipo de informação é de extrema importância para um melhor entendimento das relações entre genótipo e fenótipo associados à MPS IVA.

Atualmente, existem principalmente dois tipos de abordagem experimental para obtenção de estruturas tridimensionais: método de cristalografia de raio X e ressonância nuclear magnética (RMN) (Baxevanis & Ouellette, 2001). No entanto, esses métodos possuem algumas limitações, como por exemplo altos custos e tempo. No caso da cristalografia por raio X, existe a necessidade de obtenção de um cristal, que pode levar de alguns meses, até vários anos para ser gerado, além da possibilidade da geração de artefatos. Já no caso da RMN existem limitações com relação ao tamanho da proteína (DePristo et al., 2004; Baxevanis & Ouellette, 2001). Portanto, uma possível abordagem que reduziria os custos e o tempo total do estudo seria a modelagem *in silico*. Para esse tipo de modelagem são utilizadas principalmente duas abordagens, modelagem *ab-initio* e modelagem por homologia. A modelagem *ab-initio* parte da inexistência de um molde,

dessa forma, o modelo é construído sob a orientação de uma função de energia gerando múltiplos modelos possíveis onde um ou mais modelos poderão ser selecionados (Baker et al., 2001). Por outro lado, a modelagem por homologia também chamada de modelagem comparativa, utiliza estruturas de proteínas determinadas experimentalmente (moldes) para prever a conformação de outra proteína que tem uma sequência de aminoácidos semelhante (alvo) (Johnson et al., 1994). Os passos envolvidos nesse método são: seleção de uma ou mais estruturas molde, alinhamento entre as sequências da estrutura molde e da proteína alvo, a modelagem, refinamento e avaliação dos modelos gerados (Eswar et al., 2006). Sendo assim a GALNS selvagem resolvida por cristalografia de raio X pode ser usada como molde em uma abordagem utilizando modelagem por homologia.

2. Justificativa

O conhecimento das alterações causadas pelas mutações *missense* na estrutura da GALNS é uma informação extremamente valiosa, pois poderá guiar novas pesquisas que buscam o desenvolvimento de diferentes enzimas recombinantes para o tratamento da SM. Tendo em vista que até o momento a maioria dos tratamentos é apenas paliativo, este tipo de estratégia ganhou destaque. Além disso, tratamentos de reposição enzimática se tornaram bastantes eficientes em outros tipos de MPS como MPS I, II e IV (Giugliani et al., 2010), sendo que uma abordagem foi recentemente descrita na própria MPS IVA (Sanford & Lo, 2014). Porém, a terapia com reposição enzimática é ainda muito questionada na pesquisa clínica, principalmente, por comprometer a resposta imunológica desses pacientes. Desta forma, o presente estudo foi delineado para elucidar as alterações causadas pelas mutações *missense* na enzima GALNS e, posteriormente, encontrar uma forma de avaliar o quão severas são as alterações causadas por mutações *missense* nesta enzima.

A abordagem da bioinformática estrutural foi escolhida, utilizando-se ferramentas computacionais modernas e robustas. Por se tratar do primeiro estudo usando como molde a estrutura da GALNS selvagem resolvida por cristalografia para modelar mutações *missense*, acredita-se que os modelos gerados estarão complacentes com as estruturas originais. Portanto, o uso desta abordagem torna possível o desenvolvimento de experimentos práticos mais acertados, reduzindo os custos e os prazos na pesquisa científica da doença.

3. Objetivo

3.1. *Objetivo Geral*

O presente estudo teve como objetivo obter:

- Modelar 3 mutações na proteína GALNS, sendo uma mutação no sítio ativo, uma no núcleo hidrofóbico e uma na superfície.

3.2. *Objetivo Específico*

- Analisar e entender as diferenças estruturais entre a GALNS selvagem e as mutantes selecionadas.

4. Materiais e Métodos

A metodologia desenvolvida neste trabalho está resumida no fluxograma abaixo (Figura 2)

Figura 2. Fluxograma que apresenta cada passo da metodologia utilizada no presente trabalho.

4.1. Obtenção do Molde

O molde foi obtido através do repositório público *Protein Data Bank* (PDB) (Berman et al., 2000), conhecido por conter informações de estruturas tridimensionais de grandes moléculas biológicas como proteínas e ácidos nucleicos. A estrutura escolhida como molde, foi recuperada no formato pdb cujo identificador é 4FDI. Trata-se da estrutura da GALNS selvagem, resolvida por Rivera-Colón e colaboradores (2012) pela técnica de difração de raio x, com resolução de 2.2 Å. Esta mesma estrutura foi observada e verificou-se que continha 28 aminoácidos a menos na região N-Terminal em comparação com o que foi descrito por Hendriksz e colaboradores (2013). Assim, às análises foram realizadas levando-se em consideração a GALNS com 494 aminoácidos.

4.2. Obtenção de três variantes da GALNS

As mutações na enzima GALNS foram obtidas através do trabalho de Rivera-Colón e colaboradores (2012), onde, em 157 mutações, somente 126 eram *missense*. Dessas, 10 mutações foram excluídas das análises por tratarem-se de mutações indefinidas, ou seja, com posição da mutação conhecida, mas sem informação das trocas dos aminoácidos. Sendo assim, restaram apenas 116 mutações *missense* (Tabela 2). Deste total, 3 mutações foram selecionadas para o prosseguimento das análises, uma mutação presente no núcleo hidrofóbico, uma presente na superfície e uma presente no sítio ativo. Para a região do núcleo hidrofóbico foi selecionada a mutação de troca de uma alanina por uma treonina (Figura 3a) na posição 107, localizada em uma região conservada cujo fenótipo é indefinido (Tomatsu et al., 2006). Para a região da superfície foi escolhida uma mutação de troca de uma alanina por uma valina (Figura 3b) na posição 351, que está posicionada em uma região não conservada, responsável por gerar um fenótipo grave (Olarte-Avellaneda et al., 2014). No sítio ativo foi selecionada a mutação de troca de uma cisteína por uma tirosina (Figura 3c) na posição 79, que está posicionada em uma região conservada, responsável por gerar um fenótipo grave (Tomatsu et al., 2006).

Tabela 2. Apresentação das mutações *missense* na enzima GALNS divididas em 3 grupos com 78 mutações na região do núcleo hidrofóbico, 32 mutações na superfície e 6 mutações no sítio ativo.

Núcleo Hidrofóbico			Superfície	Sítio Ativo
L36P	P163H	D344E	S53F	C79Y
M41L	N164T	L345P	D60N	S80L
G42E	H166Q	F346L	R61W	H142R
G47R	G168R	L352P	G155R	H236D
G66R	P179S	R380T	G155E	K310N
F69V	P179H	R380S	F167V	N289S
P77R	P179L	R386C	D171A	
T88I	A203V	R386H	E185G	
R90W	W230G	M391V	N204K	
R94G	D233N	A392V	G247D	
R94C	V239F	H401Y	R253W	
R94L	A257T	N407H	R259Q	
G96C	E260D	E450V	S295F	
G96V	F284V	M494V	A351V	
F97V	S287L		P357L	
A107T	G290S		R361G	
Q111R	A291T		L366F	
I113F	A291D		L369P	
G116S	G301C		R376Q	
P125L	L307P		D388N	
S135R	G309R		L395V	
V138A	T312A		L395P	
G139S	T312S		H398D	
W141R	G316V		W409S	
W141C	M318R		G421E	
H150Y	A324E		Q422K	
P151S	W325C		F452I	
P151L	G340D		F452L	
F156S	S341R		S470P	
F156C	M343L		P484S	
W159C	M343R		N487S	
S162F	D344N		V488M	

Adaptado de Rivera Colón et al., 2012

Figura 3. Representação química das mutações *missense* selecionadas para o estudo. a) Substituição de uma alanina por uma treonina ocorrida no núcleo hidrofóbico. b) Substituição de uma alanina por uma valina ocorrida na superfície. c) Substituição de uma cisteína por uma tirosina ocorrida no sítio ativo.

As mutações foram manualmente introduzidas na estrutura primária da GALNS, gerando 3 variantes, uma com mutação presente no núcleo hidrofóbico (variante A107T), uma com mutação na superfície (variante 351V) e uma presente no sítio ativo (variante C79Y).

4.3. Modelagem por homologia

O *software* MODELLER 9.10 (Eswar et al., 2006) foi utilizado para a realização da modelagem por homologia das 3 variantes da GALNS. Foram gerados 5 modelos para cada uma das variantes, sendo que o MODELLER baseou-se em suas energias para indicar qual o melhor. Entretanto, com o objetivo de identificar possíveis diferenças

estruturais entre os modelos gerados para cada uma das variantes, suas estruturas foram alinhadas através da ferramenta *on-line* IPBA (www.dsimb.inserm.fr/dsimb_tools/ipba), para realização do cálculo de *root mean square deviation* (RMSD) entre cada um dos modelos. O RMSD é um método matemático que calcula o desvio médio quadrático para todos os átomos de uma estrutura, ou seja, esse método estima os desvios médios de uma estrutura proteica em relação a uma estrutura de referência (Maiorov et al, 1994). O gráfico de Ramachandran construído através da ferramenta Procheck (Laskowski et al., 1993), também foi utilizado com o objetivo de avaliar as possíveis diferenças estruturais entre os modelos. Esse gráfico indica regiões favoráveis, regiões adicionalmente permitidas, generosamente permitidas e não permitidas para os graus de todos os ângulos Phi e Psi dos aminoácidos.

4.4. Avaliação dos modelos obtidos

Vários programas foram utilizados para avaliar a qualidade dos modelos selecionados na etapa de modelagem por homologia. A qualidade estereoquímica do modelo foi avaliada com o programa Procheck que avalia os comprimentos das ligações, os ângulos planos, a planaridade dos anéis de cadeias laterais, a planaridade de cadeias peptídicas, os ângulos torcionais da cadeia principal das cadeias laterais, o impedimento estérico entre pares de átomos não ligados e a qualidade do gráfico de Ramachandran. O programa PROSA II (Wiederstein & Sippl, 1993) foi utilizado para avaliar o ambiente de cada aminoácido da proteína modelada, tendo como referência o ambiente esperado em proteínas análogas em alta resolução. O programa ERRATv2 (Colovos & Yates, 1993), foi utilizado para avaliar o progresso do refinamento proteico através da interação entre átomos não ligados. O programa Verify3d (Bowie et al., 1991) foi utilizado para verificar a compatibilidade entre a sequência de aminoácidos e a estrutura terciária. O programa ProQ é um método baseado em redes neurais para predizer a qualidade do modelo proteico. Ele extrai características estruturais, tal como a frequência de contato átomo-átomo, usando dois tipos de medidas de qualidade chamadas de LGscore que geralmente apresentam melhores resultados para proteínas de cadeia longa (Cristobal et al., 2001) e MaxSub que geralmente apresentam melhores resultados para proteínas de cadeia curta (Siew et al., 2000).

4.5. Inspeção Visual

Além do cálculo da distância entre os modelos e a GALNS selvagem, os *softwares* de visualização Pymol e DS Visualizer 4.0 (Accelrys Software Inc., 2013) foram utilizados para análises visuais sobre a conformação das estruturas geradas. Além disso, foi realizada a sobreposição dos modelos com a estrutura selvagem para identificação e medições de possíveis diferenças estruturais.

4.6. Simulações de Dinâmica Molecular

Foram realizadas 3 simulações de dinâmica molecular, uma para cada modelo gerado. Foi utilizado o software NAMD 2.10 (Phillips et al., 2005) que implementa o campo de força CHARMM27 (Mackerell et al., 1998). Alguns passos anteriores como: criação do arquivo .psf (armazena informações essenciais ao cálculo de dinâmica molecular sobre a estrutura da proteína tais como vários tipos de interações de ligação), adição das coordenadas dos hidrogênios que garantirão que suas posições sejam razoáveis e a solvatação da proteína em uma caixa d'água foram realizadas usando o software VMD 1.9 (Humphrey et al., 1996). As simulações foram realizadas a uma temperatura de 310 K, a pressão de 1 atm. As caixas d'água utilizadas continham distâncias de ao menos 2 Å para os átomos mais distantes do centro de cada modelo. Antes de cada simulação, todo sistema foi minimizado por 100 passos para reduzir interações desfavoráveis entre a proteína e solvente. Não foi adicionado nenhum tipo de íon para balanceamento de cargas. Os demais parâmetros foram deixados em *default*, pois não interfeririam no cálculo proposto. O *software* VMD foi novamente utilizado ao fim da simulação para calcular RMSD das trajetórias dos modelos.

4.7. Cálculo da superfície eletroestática

O *software* Chimera 1.10.1 (Pettersen et al., 2004) foi utilizado para realização do cálculo da distribuição eletroestática da superfície de acordo com a lei de Coulomb. O principal objetivo foi identificar possíveis alterações na distribuição de cargas ao longo da superfície dos modelos em comparação com a estrutura selvagem da GALNS.

4.8. Cálculo da hidrofobicidade da superfície

O *software* Chimera 1.10.1 (Pettersen et al., 2004) foi utilizado também para realizar o cálculo das possíveis alterações na hidrofobicidade ao longo da superfície dos modelos em relação à estrutura da GALNS selvagem.

5. Resultados e Discussão

5.1. Modelagem Molecular das 3 variantes da enzima GALNS

O *software* MODELLER foi utilizado para realizar as modelagens das estruturas tridimensionais. Para cada uma das três variantes (variante A107T, variante A351V e variante C79Y) foram gerados 5 modelos diferentes. Algumas análises foram realizadas a fim de identificar o quão diferentes eram os modelos.

A primeira análise foi o cálculo de RMSD entre os modelos gerados. O cálculo foi realizado separadamente, onde todos os modelos de cada variante foram comparados entre si. O RMSD entre os modelos da variante A107T variou entre 0,10 e 0,15 Å. Para os modelos gerados a partir da variante A351V, o RMSD dos modelos variaram entre 0,09 e 0,13 Å; enquanto que, para a variante C79Y, o RMSD entre os modelos variam entre 0,09 e 0,14 Å.

Outro experimento de comparação entre os modelos gerados para cada variante foi a construção do gráfico de Ramachandran. Neste sentido, sabe-se que um modelo para ser considerado bom, deve ter no mínimo 90% de seus resíduos localizados em regiões favoráveis (A, B, L), desconsiderando os resíduos de glicina (não possuem cadeia lateral) e prolina (o C α está ligado a cadeia lateral) (Figura 4) (Laskowski et al., 1993). Através dessa análise foi demonstrado que todos os modelos apresentaram valores acima de 90%, com variação de no máximo 1% entre os modelos. Desta forma, diante das diferenças estruturais mínimas observadas entre os modelos, foram selecionados os modelos mais adequados através do programa Modeller, utilizando como critério a energia mínima (Figura 5).

Figura 4. Gráfico de Ramachandran representativo para o modelo da variante A107T. Foi demonstrando que 90,5% dos aminoácidos estão presentes em regiões favoráveis (vermelho), 8,6% em regiões adicionalmente favoráveis (amarelo), 0,7% em regiões generosamente favoráveis (bege) e 0,2% regiões não permitidas (em branco). Os quadrados pretos representam os aminoácidos e os triângulos pretos representam glicinas e prolina.

Figura 5. Representação tridimensional dos modelos selecionados para cada uma das variantes, demonstrando suas estruturas secundárias. a) Variante A107T b) Variante A351V c) Variante C79Y. Essas imagens foram geradas pelo programa Pymol.

5.2. Avaliação dos Modelos

Para este trabalho foram utilizados os *softwares* Procheck, PROSA II, ERRATv2, Verify 3d, ProQ. Cada uma das ferramentas indica quais são os melhores valores de *score*, que representam a qualidade do modelo. A Tabela 3 apresenta os resultados obtidos para cada modelo comparando-os com os melhores scores indicados por cada ferramenta, exceto no caso do programa Prosa II que apresenta seus resultados em formato gráfico.

Tabela 3. Resultados dos *softwares* de avaliação, com relação aos seus *scores*. Os resultados dos programas para cada variante são comparados com os *scores* considerados bons para cada programa.

	Procheck	ERRATv2	Verify 3D	LG score	MaxSub
	%	%	%		
Scores considerados bons	>= 90	>=95	>80	>2,5	>0,5
Variante A107T	90,5	84,47	98,58	6,4	0,527
Variante A351V	90,2	81,28	90,49	6,36	0,522
Variante C79Y	90	71,8	94,74	6,63	

O programa PROSAll exibe dois gráficos como resultado, um demonstrando a posição do Z-score do modelo em relação a outras proteínas (Figura 6) e outro demonstrando a qualidade local do modelo com relação a energia em função da sequência de aminoácidos (Figura 7). O Z-score indica a qualidade global dos modelos. O ponto preto no gráfico indica se o Z-score obtido para o modelo analisado está de acordo com os Z-scores de proteínas determinadas experimentalmente de tamanhos semelhantes obtidos através do banco de dados PDB. Em azul escuro estão representados os Z-scores de proteínas determinadas por ressonância magnética nuclear (RNM) e em azul claro estão representados os Z-scores de proteínas determinadas por difração de raio x.

Todos os modelos tiveram resultados bastantes semelhantes para as duas análises realizadas pelo programa. A figura 6 representa a qualidade global para o modelo da variante A107T.

Figura 6. Gráfico comparando a posição do Z-score da variante A107T com os Z-scores de preteínas determinadas experimentalmente de tamanho semelhante. Em azul escuro os Z-scores de proteínas resolvidas por RNM e em azul claro os Z-scores de proteínas determinadas por cristalografia de raio X. O ponto preto indica o Z-score da variante A107T. Gráfico gerado pelo programa Prosa II.

Através do cálculo de energia dos três modelos, foram vistos alguns valores positivos (Figura 7 b), que indicariam partes problemáticas da proteína. Porém, valores muito semelhantes foram observados quando a GALNS selvagem foi analisada pelo mesmo programa (Figura 7 a). Portanto, acredita-se que esses resultados não tenham sido causados por algum erro no processo de modelagem, tampouco que as mutações tenham causado alguma alteração para que resultados diferentes com relação à energia fossem observados. A Figura 7 apresenta um comparativo entre os resultados do cálculo de energia para a estrutura da GALNS selvagem e os resultados do cálculo de energia para a variante A107T.

Figura 7. A esquerda é demonstrado o gráfico de energia para a GALNS selvagem. Através do programa Prosa II foi obtido o gráfico de energia da variante A107T. As linhas em verde claro (fina) correspondem ao cálculo da energia levando-se em consideração uma janela de 10 aminoácidos. As linhas em verde escuro corresponde ao cálculo de energia levando-se em consideração uma janela de 40 aminoácidos. Gráfico gerado pelo programa Prosa II.

Desta forma, os resultados dos programas utilizados demonstram que os três modelos selecionados na etapa anterior foram bem avaliados por todas as ferramentas.

5.3. Inspeção visual

A inspeção visual foi realizada com o objetivo de avaliar as posições dos resíduos de cada modelo e fazer comparações com a GALNS selvagem. Utilizando o *software* Pymol foi possível sobrepor a estrutura tridimensional de cada variante com a GALNS selvagem. Primeiramente, foi verificado que o RMSD entre os modelos das variantes e a estrutura selvagem foi muito pequeno. O RMSD entre a estrutura selvagem com o modelo da variante A107T foi de 0,10 Å, já o RMSD entre estrutura selvagem e o modelo da variante A351V foi de 0,09 Å, enquanto que o RMSD entre a estrutura selvagem e o modelo gerado a partir do modelo da variante C79Y foi de 0,11 Å, levando-se em

consideração que o RMSD entre os moldes e a estrutura selvagem contou com o alinhamento de 494 resíduos.

Esses valores são menores do que os encontrados por Sudhakar e colaboradores (2011), onde foram modeladas a estrutura da GALNS selvagem e mais 103 variantes geradas por mutações *missense*, divididas entre fenótipos graves e atenuados. O cálculo de RMSD entre o modelo da GALNS selvagem e os modelos das variantes com mutações em regiões conservadas responsáveis por gerar fenótipos graves variou entre 0,21 e 0,23 Å. Por outro lado, esses valores foram menores do que observados em mutações em regiões não conservadas que levam aos fenótipos graves (0,23 a 0,31 Å) (Sudhakar et al., 2011). Neste sentido, é interessante notar que os valores observados em nosso trabalho são menores tanto para alterações em regiões conservadas quanto para regiões não conservadas. Estes resultados podem ser explicados, pelo fato do molde utilizado por Sudhakar (Sudhakar et al., 2011) conter somente 29% de identidade com a GALNS selvagem, o que em teoria, poderia gerar uma distância maior entre as estruturas.

O próximo passo foi realizar a inspeção visual, a fim de identificar possíveis diferenças estruturais entre os modelos e a estrutura selvagem. A GALNS selvagem está representada na cor azul, enquanto que as variantes estão representadas na cor verde (Figura 8).

Figura 8. Representação tridimensional da sobreposição da GALNS selvagem em azul com os modelos das variantes em verde. a) Sobreposição da GALNS selvagem com a variante A107T b) Sobreposição da GALNS Selvagem com a variante A351V c) Sobreposição da GALNS Selvagem com a variante C79Y. Figuras geradas pelo programa Pymol.

A inspeção visual foi realizada a partir das estruturas secundárias e terciárias, levando-se em consideração somente desvios na cadeia principal das estruturas com distância mínima de 0,1 Å. A Figura 8 mostra que as alterações entre variantes e GALNS selvagem foram discretas. Entretanto, foi possível encontrar alguns ligeiros desvios estruturais dos modelos em relação à estrutura da GALNS selvagem. Os desvios encontrados foram divididos em três categorias: regiões de *looping*, alfa hélices e folhas beta, como mostrado na Tabela 4. As alterações encontram-se distribuídas de forma aleatória pela estrutura dos modelos, não existindo nenhum indício de padrão ou preferência por alguma região em específico. Em sua maioria trata-se de pequenos desvios do carbono alfa das variantes em relação à estrutura selvagem. O maior desvio

observado entre carbonos alfa foi de 0,9 Å entre duas glicinas localizadas na posição 327 da sequência de aminoácidos em comparação da variante A351V com a estrutura selvagem (Figura 9). O modelo que mais apresentou alterações na comparação com a estrutura da GALNS selvagem foi o gerado a partir da variante C79Y, somando um total de nove alterações ao longo da estrutura.

Tabela 4. Quantidade de desvios estruturais dos modelos das variantes em comparação com a estrutura tridimensional da GALNS selvagem agrupados em 3 categorias: presentes no núcleo hidrofóbico, presentes na superfície e presentes no sítio ativo.

	Região das mutações	<i>Looping</i>	Alfa Hélice	Folha Beta	Total
Variante A107T	Núcleo Hidrofóbico	4	2	0	6
Variante A351V	Superfície	3	1	0	4
Variante C79Y	Sítio Ativo	4	4	1	9

Figura 9. Demonstração da maior distância encontrada entre carbonos alfa da estrutura selvagem e uma variante. Foram observados 0,9 Å de distância entre os carbonos alfas de duas glicinas. GALNS selvagem na cor azul e variante com mutação na superfície na cor verde. Os hidrogênios não estão representados nesta figura. Figura gerada pelo programa Pymol.

Também, foram avaliadas alterações locais introduzidas pelas substituições dos aminoácidos na região da mutação. Na estrutura da GALNS selvagem, a alanina (posição

107) está posicionada entre uma leucina (posição 486), uma treonina (posição 99) e um triptofano (posição 520), as distâncias com esses resíduos podem ser vistas na Figura 11a. Outras distâncias foram observadas no modelo da variante A107Y, onde na mesma posição teve uma menor distância entre a treonina (que substitui a alanina) e o triptofano, enquanto, que as distâncias para os outros resíduos aumentaram (Figura 10b). É possível observar também, uma troca de posição entre o oxigênio e o carbono localizados no radical da treonina (posição 99).

Figura 10. Demonstração das distâncias entre os aminoácidos I. a) Medição das distâncias entre a alanina (posição 107) e os aminoácidos mais próximos na estrutura da GALNS selvagem. b) Medição das distâncias entra a Treonina e os aminoácidos mais próximos na variante A107T. Imagem gerada pelo programa DS Visualizer.

Através da Figura 11a, pode ser visto que na estrutura selvagem a alanina está próxima de uma leucina (posição 369) e uma asparagina (posição 367). No modelo da variante A351V a substituição de uma alanina por uma valina não alterou as distâncias entre os resíduos (Figura 11 b), somente foi observada um pequena alteração no radical da leucina.

Figura 11. Demonstração das distâncias entre os aminoácidos II. a) Medição das distâncias entre a alanina (posição 351) e os aminoácidos mais próximos na estrutura da GALNS selvagem b) Medição das distâncias entre a Valina e os aminoácidos mais próximos na variante A351V. Imagem gerada pelo programa DS Visualizer.

No centro do sítio ativo na posição 79 é encontrada uma cisteína, que tem um papel crucial para a funcionalidade da enzima já que uma modificação pós-traducional mediada pela enzima formadora de formilglicina a torna cataliticamente ativa. A conversão de uma cisteína na posição 79 por um aldeído formilglicina (trocando um átomo de enxofre e um de hidrogênio por um átomo de oxigênio com ligação dupla) torna possível a reação química com o substrato (Olarte-Avellaneda et al., 2014; Rivera-Colón et al., 2012). Portanto, a troca de uma cisteína por uma tirosina na posição 79 faz com que a enzima perca sua função. Na estrutura selvagem, a cisteína está posicionada com o enxofre apontado para fora do sítio, isso faz com que ela esteja mais próxima dos resíduos: lisina (posição 140), histidina (posição 236) e lisina (posição 310) (Figura 12 a). Quando a variante C79Y foi analisada, foi visto que o radical da tirosina está em uma posição diferente, voltado para dentro do sítio ativo, estando dessa forma mais próximo de outros resíduos: metionina (posição 38), arginina (posição 83) e ácido aspártico (posição 288) (Figura 12 b).

Figura 12. Demonstração das distâncias entre os aminoácidos III. a) Medição das distâncias entre a citosina (posição 79) e os aminoácidos mais próximos na estrutura da GALNS selvagem b) Medição das distâncias entre a tirosina e os aminoácidos mais próximos na variante C79Y. Imagem gerada pelo programa DS Visualizer.

Analisando a posição dos aminoácidos responsáveis pelas mutações, foi possível constatar que os desvios observados em relação à estrutura da GALNS selvagem são mínimos. A distância entre os carbonos alfa entre a alanina (GALNS selvagem) e a treonina (variante A107T) foi de apenas 0,1 Å, sendo que seus radicais estão a 0,2 Å de distância (Figura 13 a). Uma situação semelhante foi encontrada quando a GALNS selvagem e variante A351V foram comparadas na exata região da mutação. Foi observada uma distância de 0,1 Å entre carbonos alfa e uma distancia de 0,3 Å entre seus radicais (Figura 13 c). Diferentemente ao observado na exata região da mutação quando comparadas GALNS selvagem e variante C79Y. Apesar da distância entre os carbonos alfa ser também 0,1 Å, os radicais sofreram uma torção assumindo posições diferentes (Figura 13 b).

Figura 13. Comparação entre as posições no espaço tridimensional dos aminoácidos da estrutura da GALNS selvagem e as variantes. Em azul aminoácidos presentes na estrutura selvagem, em verde aminoácidos substituídos. a) Comparação entre as posições da alanina e da treonina. b) Comparação entre as posições da cisteína e da tirosina. c) Comparação entre as posições da alanina e da valina. Figura gerada pelo programa Pymol.

Entretanto, esses pequenos desvios não foram suficientes para causar alterações relevantes nas estruturas secundárias das variantes, tanto na exata posição das mutações (Figura 14) quanto nas estruturas como um todo (Figura 8). Esses dados corroboram com os resultados obtidos por Tamarozzi e colaboradores (2014), onde em um estudo de predição da estrutura secundária de 161 variantes geradas por mutações na enzima GALNS, ficou constatado, que as mutações *missense* não causam alterações relevantes na estrutura secundária da GALNS.

Figura 14. Comparação entre as posições no espaço tridimensional dos aminoácidos da estrutura da GALNS selvagem e as variantes, demonstrando que não houveram alterações nas estruturas secundárias. Em azul está representada a estrutura selvagem, em verde estão representadas as variantes. a) A troca de uma alanina por uma treonina não causou nenhuma alteração na região de *looping*. b) A troca de uma cisteína por uma tirosina não causou nenhum desvio na região de *looping*. c) A troca de um alanina por uma valina não causou nenhuma alteração na alfa hélice. Figura gerada pelo programa Pymol.

5.4. Análise de Dinâmica molecular

De forma geral esses pequenos desvios observados nas três variantes em relação à estrutura da GALNS selvagem, não foram suficientes para causar alterações severas na sua estrutura tridimensional. Posteriormente, os modelos das variantes da GALNS foram submetidos à simulação de 25 ns de dinâmica molecular para avaliar sua estabilidade estrutural em ambiente aquoso com temperatura e pressão próximos aos encontrados em ambiente celular (Figura 15).

Figura 15. Representação da variante C79Y em uma caixa d'água, para a realização da simulação de dinâmica molecular. No centro da figura em azul claro, está representada a proteína. Os pontos vermelhos representam as moléculas de água. Figura gerada pelo programa VMD.

A cada 250 fs da simulação a posição dos átomos era medida para que ao fim da simulação fosse possível calcular os valores de RMSD da trajetória dos modelos ao longo de toda a simulação. Os resultados dos cálculos de RMSD para a trajetória dos modelos mostraram que todos estabilizaram a partir dos 5 ns apresentando variação menor do que 0,5 Å. Não foram observadas grandes distâncias entre o início e o fim da simulação (figura 16).

A simulação realizada com o modelo da variante C79Y, mostra, que o modelo se mantém estável durante toda simulação. A simulação realizada com o modelo da variante A351V, mostra que entre 11 e 12 ns uma variação maior é observada, mas logo após a estrutura volta a se estabilizar. A simulação realizada com o modelo da variante A107T passa a ter intervalos de variação um pouco maiores a partir de 12 ns.

Esses resultados de estabilidade das variantes em simulação de dinâmica molecular, são reforçados pela segunda parte do trabalho realizado por Tamarozzi colaboradores (2014). Onde foram comparados os resultados de cinco programas diferentes com o objetivo de avaliar a desordem intrínseca dos aminoácidos em 161 mutações, sendo que, 129 eram mutações *missense*. Os resultados mostraram que as mutações *missense* não causaram alterações relevantes no padrão de desordem intrínseca da GALNS. Constatando que a estabilidade das variantes era esperada, fato este, que está de acordo com o que foi discutido por Sudhakar e colaboradores (2011), que, a gravidade dos fenótipos pode não estar diretamente relacionada às alterações na estrutura da enzima, e sim, à localização dessas mutações.

Figura 16. Gráfico do RMSD da trajetória dos átomos das variantes ao longo de 25 ns. As linhas coloridas no gráfico indicam o quanto as estruturas se distanciaram de seu estado inicial. Em azul está representada a variante A107T, em vermelho a variante A351V e em verde a variante C79Y.

5.5. Cálculo da superfície eletrostática

O cálculo da superfície eletrostática (SE) possibilita a informação sobre a distribuição de cargas na superfície da molécula. O maior interesse estava em analisar se alguma mutação causou uma diferença na distribuição de cargas ao longo da superfície dos modelos em relação a estrutura da GALNS selvagem. Os resultados mostraram que não existem diferenças consideráveis no padrão de distribuição de cargas ao longo das superfícies dos modelos em relação à estrutura da GALNS selvagem, exceto para a variante C79Y que apresentou uma diferença no seu sítio ativo (Figura 17 d). A troca de uma cisteína por uma tirosina na posição 79, que se encontra no centro do sítio ativo, fez com que uma nova porção positivamente carregada surgisse. O sítio ativo tem carga parcialmente positiva que se relaciona com cargas predominantemente negativas de seu substrato (Funderburgh, 2002). Portanto, essa alteração pode atrapalhar o encaixe do substrato com o sítio ativo.

A Figura 17 apresenta um comparativo da distribuição eletrostática do sítio ativo entre a GALNS selvagem e as variantes. Deste comparativo, somente foi observado alteração na variante C79Y. As proteínas são coloridas de acordo com o potencial eletrostático, ou seja, vermelho representa as regiões com potencial negativo, azul representa as regiões com potencial positivo e branco representa as regiões neutras.

Figura 17. Comparação entre os sítios ativos GALNS selvagem com as variantes. Vermelho representa regiões negativamente carregadas; Azul representa regiões positivamente carregadas; Branco representa as regiões neutras. A seta vermelha indica o lugar onde foi observada a alteração. **a)** Região do sítio ativo da GALNS selvagem. **b)** Região do sítio ativo da variante com mutação no núcleo hidrofóbico. **c)** Região do sítio ativo da variante com mutação na superfície. **d)** Região do sítio ativo da variante com mutação no sítio ativo. Imagens geradas pelo programa Chimera 1.10.1.

5.6. Cálculo hidrofobicidade da superfície

O cálculo da hidrofobicidade da superfície dos modelos foi realizado para identificar possíveis alterações na polaridade da superfície causada pelas mutações. A Figura 19 apresenta um comparativo da hidrofobicidade da superfície do sítio ativo entre GALNS selvagem e as variantes destacando a alteração na variante C79Y. Os resultados mostraram que a única alteração observada foi na variante C79Y, na região do sítio ativo. A troca de uma cisteína por uma tirosina introduziu uma região hidrofóbica (Figura 18 d)

onde originalmente na GALNS selvagem existia uma região hidrofílica (Figura 18 a). Este fato pode contribuir para que o substrato não se encaixe nos sítio ativo. A superfície foi colorida em azul que representa as regiões hidrofílicas, em laranja representando as regiões hidrofóbicas e regiões neutras em branco.

Figura 18. Comparação entre os sítios ativos GALNS selvagem com as variantes com relação à polaridade da superfície. Azul representa regiões hidrofílicas; Laranja representa regiões hidrofóbicas; Branco representa as regiões neutras. a) região do sítio ativo da GALNS selvagem. A seta vermelha indica o lugar onde foi observada a alteração. b) região do sítio ativo da variante com mutação no núcleo hidrofóbico. c) região do sítio ativo da variante com mutação na superfície. d) região do sítio ativo da variante com mutação no sítio ativo. Imagens geradas pelo programa Chimera 1.10.1.

Essas análises mostraram que mutações no sítio ativo têm uma chance maior de estarem relacionadas a fenótipos graves, uma vez que a função da enzima pode ser comprometida. Esses dados são reforçados por um estudo *in silico* publicado em 2014,

que teve como princípio entender melhor o efeito das mutações na enzima GALNS (Olarte-Avellaneda et al., 2014). A proteína GALNS foi modelada computacionalmente usando o método de modelagem por homologia, tendo como molde a enzima arilsulfatase A humana, que é uma enzima lisossomal com mecanismo de ação semelhante a GALNS. Para avaliar o efeito da ligação das enzimas mutadas com o substrato, os autores introduziram, manualmente, as mutações no modelo da GALNS gerado, realizando o *docking* com substratos naturais e artificiais. Os resultados mostraram que na maioria dos casos, mutações no sítio eram responsáveis por ligações com menor afinidade com o substrato (Olarte-Avellaneda et al., 2014).

Este foi o primeiro estudo realizando modelagem comparativa com a GALNS selvagem resolvida por cristalografia de raio X, para modelagem de variantes geradas por mutações *missense*. Para tal, uma nova abordagem será necessária para que a região N-terminal seja modelada. Mesmo assim, foi possível constatar que não foram observadas alterações estruturais relevantes tanto em estruturas secundárias quanto em estruturas terciárias em nenhuma das três variantes analisadas, o que proporcionou estabilidade aos modelos em ensaios de dinâmica molecular.

Para a variante A351V cuja mutação se encontra na superfície, os resultados do presente estudo contrariam o diagnóstico de mutação causadora de fenótipo grave. Isso pode ser explicado pelo fato dos critérios utilizados por diversos autores (Dũng et al., 2013; Rivera-Colón et al., 2012; Tomatsu et al., 2006) para classificação da gravidade dos fenótipos levar em consideração somente a altura dos pacientes. Entretanto, a altura do paciente pode ser influenciada por outros fatores que não uma deficiência na atividade da GALNS. Existem outros fatores como alterações não esqueléticas incluindo anormalidades no sistema visual, auditivo, digestivo, cardiovascular e no sistema respiratório, que também devem ser levadas em consideração na definição da gravidade do fenótipo tendo em vista que esse tipo alteração tem um impacto significativo para a qualidade de vida do paciente (Olarte-Avellaneda et al., 2014; Hendriksz et al., 2013). Assim, sobre um outro ponto de vista com relação à caracterização da gravidade dos fenótipos da MPS IVA, algumas mutações podem não ser caracterizadas como causadoras de fenótipos graves.

O conhecimento de variantes da GALNS estáveis e potencialmente funcionais proporciona o desenvolvimento de novas pesquisas como de enzimas recombinantes para o tratamento de reposição enzimática. O único tratamento deste tipo disponível no mercado foi desenvolvido pela empresa BioMarin. Uma forma recombinante da enzima GALNS fabricada a partir de células de ovário de hamster chinês, chamada Elosulfase

alfa (VimizimTM) em 2014, (Sanford & Lo, 2014). Sobre este tratamento, alguns estudos apontaram que Elosulfase alfa se mostrou eficaz após 24 semanas da introdução da reposição enzimática. Foram descritos efeitos adversos com o tratamento enzimático, como: febre, vômito, dor de cabeça, náuseas, dor abdominal, arrepios e fadiga foram observados em alguns casos. Também foi registrado que após 30 minutos a 3 horas do tratamento, alguns pacientes sofreram anafilaxia, o que levou a inclusão do uso de anti-histamínicos com o tratamento de reposição enzimática com Elosulfase alfa (Hendriks et al., 2014; Sanford & Lo, 2014).

Os resultados obtidos nesse estudo permitem sugerir que a identificação da severidade das alterações que cada mutação *missense* causa na proteína pode guiar a forma como os tratamentos serão realizados. Estes achados representam novas perspectivas para o tratamento da MPS. Porém, novos estudos devem ser realizados para validar estes resultados, como por exemplo, o teste de *docking* com o substrato, para posteriormente avançar na pesquisa de novos alvos terapêuticos com a reposição enzimática. Essas pesquisas são importantes para garantir uma melhor qualidade de vida dessas pessoas e melhorar a aceitação pelo paciente ao tratamento, sem que haja necessidade de outros fármacos para diminuir os efeitos adversos da reposição enzimática.

6. Conclusão

Pode ser concluído que as mutações A107T, A351V e C79Y não causaram alterações estruturais significativas, e também não interferiram na estabilidade estrutural em simulação de dinâmica molecular. Além disso, foi demonstrado que mutações no sítio ativo possuem maior probabilidade de interferir na função da enzima. Outros estudos podem ser necessários com relação às alterações epigenéticas (Tomatsu et al, 2004) e nas vias metabólicas da GALNS para auxiliar a elucidar a grande heterogeneidade fenotípica da doença.

Em suma, esta metodologia associada a estudos de *docking* com os substratos naturais, mostrou-se bastante efetiva na identificação de enzimas estáveis e funcionais, podendo ser reproduzida para todas as mutações *missense*. Portanto, este trabalho poderá possibilitar o surgimento de novas pesquisas quanto ao desenvolvimento de enzimas recombinantes para o tratamento de reposição enzimática.

7. Referências

1. Accelrys Software Inc., *Discovery Studio Modeling Environment, Release 4.0*, San Diego: Accelrys Software Inc., 2013.
2. Ard, J. L., Bekker, A., & Frempong-Boadu, A. K. (2005). Anesthesia for an adult with mucopolysaccharidosis I. *Journal of clinical anesthesia*, 17(8), 624-626.
3. Baker, D., & Sali, A. (2001). Protein structure prediction and structural genomics. *Science*, 294(5540), 93-96.
4. Baldo, G., Giugliani, R., & Matte, U. (2014). Gene delivery strategies for the treatment of mucopolysaccharidoses. *Expert opinion on drug delivery*, 11(3), 449-459.
5. Baxevanis, A. D., & Ouellette, B. F. (2004). *Bioinformatics: a practical guide to the analysis of genes and proteins* (Vol. 43). John Wiley & Sons.
6. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., & Bourne, P. E. (2000). The protein data bank. *Nucleic acids research*, 28(1), 235-242.
7. Byers, S., Rozaklis, T., Brumfield, L. K., Ranieri, E., & Hopwood, J. J. (1998). Glycosaminoglycan accumulation and excretion in the mucopolysaccharidoses: characterization and basis of a diagnostic test for MPS. *Molecular genetics and metabolism*, 65(4), 282-290.
8. Bowie, J. U., Luthy, R., & Eisenberg, D. (1991). A method to identify protein sequences that fold into a known three-dimensional structure. *Science*, 253(5016), 164-170.
9. Caciotti, A., Garman, S. C., Rivera-Colón, Y., Procopio, E., Catarzi, S., Ferri, L., & Morrone, A. (2011). GM1 gangliosidosis and Morquio B disease: an update on genetic alterations and clinical findings. *Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease*, 1812(7), 782-790.
10. Chaves, A. G., Tavares, K. B., Val, J. R., Matsuyama, C., & Riskalla, P. E. (2003). Síndrome de Morquio: relato de caso e revisão da literatura. *Rev Bras Otorrinolaringol*, 69(2), 267-71.
11. Cimaz, R., & La Torre, F. (2014). Mucopolysaccharidoses. *Current rheumatology reports*, 16(1), 1-9.

12. Colovos, C., & Yeates, T. O. (1993). Verification of protein structures: patterns of nonbonded atomic interactions. *Protein Science*, 2(9), 1511-1519.
13. Couprie, J., Denis, P., Guffon, N., Reynes, N., Masset, H., & Beby, F. (2010). Manifestations ophtalmologiques de la maladie de Morquio. *Journal Français d'Ophtalmologie*, 33(9), 617-622.
14. Coutinho, M. F., Lacerda, L., & Alves, S. (2011). Glycosaminoglycan storage disorders: a review. *Biochemistry research international*, 2012.
15. Cristobal, S., Zemla, A., Fischer, D., Rychlewski, L., & Elofsson, A. (2001). A study of quality measures for protein threading models. *BMC bioinformatics*, 2(1), 5.
16. DePristo, M. A., de Bakker, P. I., & Blundell, T. L. (2004). Heterogeneity and inaccuracy in protein structures solved by X-ray crystallography. *Structure*, 12(5), 831-838.
17. Dũng, V. C., Tomatsu, S., Montaño, A. M., Gottesman, G., Bober, M. B., Mackenzie, W., ... & Orii, T. (2013). Mucopolysaccharidosis IVA: correlation between genotype, phenotype and keratan sulfate levels. *Molecular genetics and metabolism*, 110(1), 129-138.
18. Eswar, N., Eramian, D., Webb, B., Shen, M. Y., & Sali, A. (2008). Protein structure modeling with MODELLER. In *Structural Proteomics* (pp. 145-159). Humana Press.
19. Federico, A., Capece, G., Cecio, A., D'Auria, N., Di Iorio, G., Ronsisvalle, L., & Di Natale, P. (1981). Sanfilippo B syndrome (MPS III B): case report with analysis of CSF mucopolysaccharides and conjunctival biopsy. *Journal of neurology*, 225(2), 77-83.
20. Funderburgh, J. L. (2002). Keratan sulfate biosynthesis. *IUBMB life*, 54(4), 187-194.
21. Giugliani, R., Federhen, A., Muñoz Rojas, M. V., Vieira, T. A., Artigalas, O. A. P., PINTO, L. L. D. C., ... & Martins, A. M. (2010). Terapia de reposição enzimática para as mucopolissacaridoses I, II e VI: recomendações de um grupo de especialistas brasileiros.
22. Hendriksz, C. J., Berger, K. I., Giugliani, R., Harmatz, P., Kampmann, C., Mackenzie, W. G., & Savarirayan, R. (2015). International guidelines for the management and treatment of Morquio A syndrome. *American Journal of Medical Genetics Part A*, 167(1), 11-25.
23. Hendriksz, C. J., Burton, B., Fleming, T. R., Harmatz, P., Hughes, D., Jones, S. A., ... & STRIVE Investigators. (2014). Efficacy and safety of enzyme replacement therapy with BMN 110 (elosulfase alfa) for Morquio A syndrome

(mucopolysaccharidosis IVA): a phase 3 randomised placebo-controlled study. *Journal of inherited metabolic disease*, 37(6), 979-990.

24. Hendriksz, C. J., Harmatz, P., Beck, M., Jones, S., Wood, T., Lachman, R., & Tomatsu, S. (2013). Review of clinical presentation and diagnosis of mucopolysaccharidosis IVA. *Molecular genetics and metabolism*, 110(1), 54-64.
25. Humphrey, W., Dalke, A., & Schulten, K. (1996). VMD: visual molecular dynamics. *Journal of molecular graphics*, 14(1), 33-38.
26. Jakóbkiewicz-Banecka, J., Piotrowska, E., Gabig-Ciminska, M., Borysiewicz, E., Slominska-Wojewodzka, M., Narajczyk, M., ... & Wegrzyn, G. (2011). Substrate reduction therapies for mucopolysaccharidoses. *Current pharmaceutical biotechnology*, 12(11), 1860-1865.
27. Johnson, M. S., Srinivasan, N., Sowdhamini, R., & Blundell, T. L. (1994). Knowledge-based protein modeling. *Critical Reviews in Biochemistry and Molecular Biology*, 29(1), 1-68.
28. Kubaski, F., Brusius-Facchin, A. C., Palhares, H. M., Balarin, M. A. S., Viapiana-Camelier, M., Guidobono, R., ... & Leistner-Segal, S. (2013). Identification of a novel missense mutation in Brazilian patient with a severe form of mucopolysaccharidosis type IVA. *Gene*, 517(1), 112-115.
29. Laskowski, R. A., MacArthur, M. W., Moss, D. S., & Thornton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *Journal of applied crystallography*, 26(2), 283-291.
30. Leal, G. N., de Paula, A. C., Leone, C., & Kim, C. A. (2010). Echocardiographic study of paediatric patients with mucopolysaccharidosis. *Cardiology in the Young*, 20(03), 254-261.
31. Lehman, T. J., Miller, N., Norquist, B., Underhill, L., & Keutzer, J. (2011). Diagnosis of the mucopolysaccharidoses. *Rheumatology*, 50(suppl 5), v41-v48.
32. Leone, A., Rigante, D., Amato, D. Z., Casale, R., Pedone, L., Magarelli, N., & Colosimo, C. (2014). Spinal involvement in mucopolysaccharidoses: a review. *Child's Nervous System*, 1-10.
33. MacKerell, A. D., Bashford, D., Bellott, M. L. D. R., Dunbrack, R. L., Evanseck, J. D., Field, M. J., ... & Karplus, M. (1998). All-atom empirical potential for molecular modeling and dynamics studies of proteins. *The Journal of Physical Chemistry B*, 102(18), 3586-3616.

34. Mahalingam, K., Janani, S., Priya, S., Elango, E. M., & Sundari, R. M. (2004). Diagnosis of mucopolysaccharidoses: how to avoid false positives and false negatives. *The Indian Journal of Pediatrics*, 71(1), 29-32.
35. Maierov, V. N., & Crippen, G. M. (1994). Significance of root-mean-square deviation in comparing three-dimensional structures of globular proteins. *Journal of molecular biology*, 235(2), 625-634.
36. Montano, A. M., Tomatsu, S., Gottesman, G. S., Smith, M., & Orii, T. (2007). International Morquio A Registry: clinical manifestation and natural course of Morquio A disease. *Journal of inherited metabolic disease*, 30(2), 165-174.
37. Muenzer, J. (2014). Early initiation of enzyme replacement therapy for the mucopolysaccharidoses. *Molecular genetics and metabolism*, 111(2), 63-72.
38. Muenzer, J. (2004). The mucopolysaccharidoses: a heterogeneous group of disorders with variable pediatric presentations. *The Journal of pediatrics*, 144(5), S27-S34.
39. Muenzer, J., & Fisher, A. (2004). Advances in the treatment of mucopolysaccharidosis type I. *New England Journal of Medicine*, 350(19), 1932-1934.
40. Munoz-Rojas, M. V., Vieira, T., Costa, R., Fagundes, S., John, A., Jardim, L. B., & Giugliani, R. (2008). Intrathecal enzyme replacement therapy in a patient with mucopolysaccharidosis type I and symptomatic spinal cord compression. *American journal of medical genetics Part A*, 146(19), 2538-2544.
41. Nelson, J. (1997). Incidence of the mucopolysaccharidoses in Northern Ireland. *Human genetics*, 101(3), 355-358.
42. Noh, H., & Lee, J. I. (2014). Current and potential therapeutic strategies for mucopolysaccharidoses. *Journal of clinical pharmacy and therapeutics*, 39(3), 215-224.
43. Northover, H., Cowie, R. A., & Wraith, J. E. (1996). Mucopolysaccharidosis type IVA (Morquio syndrome): a clinical review. *Journal of inherited metabolic disease*, 19(3), 357-365.
44. Olarte-Avellaneda, S., Rodríguez-López, A., Alméciga-Díaz, C. J., & Barrera, L. A. (2014). Computational analysis of human N-acetylgalactosamine-6-sulfate sulfatase

- enzyme: an update in genotype–phenotype correlation for Morquio A. *Molecular biology reports*, 41(11), 7073-7088.
45. Palmucci, S., Attinà, G., Lanza, M. L., Belfiore, G., Cappello, G., Foti, P. V., & Ettore, G. C. (2013). Imaging findings of mucopolysaccharidoses: a pictorial review. *Insights into imaging*, 4(4), 443-459.
 46. Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., & Ferrin, T. E. (2004). UCSF Chimera—a visualization system for exploratory research and analysis. *Journal of computational chemistry*, 25(13), 1605-1612.
 47. Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., & Schulten, K. (2005). Scalable molecular dynamics with NAMD. *Journal of computational chemistry*, 26(16), 1781-1802.
 48. Prasad, V. K., & Kurtzberg, J. (2010, January). Transplant outcomes in mucopolysaccharidoses. In *Seminars in hematology* (Vol. 47, No. 1, pp. 59-69). WB Saunders.
 49. Rivera-Colón, Y., Schutsky, E. K., Kita, A. Z., & Garman, S. C. (2012). The structure of human GALNS reveals the molecular basis for mucopolysaccharidosis IV A. *Journal of molecular biology*, 423(5), 736-751.
 50. Sanford, M., & Lo, J. H. (2014). Elosulfase alfa: first global approval. *Drugs*, 74(6), 713-718.
 51. Siew, N., Elofsson, A., Rychlewski, L., & Fischer, D. (2000). MaxSub: an automated measure for the assessment of protein structure prediction quality. *Bioinformatics*, 16(9), 776-785.
 52. Simonaro, C. M., Ge, Y., Eliyahu, E., He, X., Jepsen, K. J., & Schuchman, E. H. (2010). Involvement of the Toll-like receptor 4 pathway and use of TNF- α antagonists for treatment of the mucopolysaccharidoses. *Proceedings of the National Academy of Sciences*, 107(1), 222-227.
 53. Sudhakar, S. C., & Mahalingam, K. (2011). Structural and functional analysis of N-acetyl galactosamine-6-sulfate sulfatase using bioinformatics tools: insight into Mucopolysaccharidosis IVA. *Journal of Pharmacy Research*, 4(11), 3958-3962.
 54. Sukegawa, K., Nakamura, H., Kato, Z., Tomatsu, S., Montañó, A. M., Fukao, T., & Kondo, N. (2000). Biochemical and structural analysis of missense mutations in N-acetylgalactosamine-6-sulfate sulfatase causing mucopolysaccharidosis IVA phenotypes. *Human molecular genetics*, 9(9), 1283-1290.
 55. Wood, T. C., Harvey, K., Beck, M., Burin, M. G., Chien, Y. H., Church, H. J. & Hendriksz, C. J. (2013). Diagnosing mucopolysaccharidosis IVA. *Journal of inherited metabolic disease*, 36(2), 293-307.

56. Tamarozzi, E. R., Torrieri, E., Semighini, E. P., & Giuliatti, S. (2013). In silico analysis of mutations occurring in the protein N-acetylgalactosamine-6-sulfatase (GALNS) and causing mucopolysaccharidosis IVA. *Genetics and molecular research: GMR*, 13(4), 10025-10034.
57. Tomatsu, S., Yasuda, E., Patel, P., Ruhnke, K., Shimada, T., Mackenzie, W. G., & Orii, T. (2014). Morquio A syndrome: diagnosis and current and future therapies. *Pediatric endocrinology reviews: PER*, 12(0 1), 141.
58. Tomatsu, S., M Montano, A., Oikawa, H., J Rowan, D., Smith, M., Barrera, L., & Orii, T. (2011). Mucopolysaccharidosis type IVA (morquio a disease): clinical review and current treatment: a special review. *Current pharmaceutical biotechnology*, 12(6), 931-945.
59. Tomatsu, S., Montañó, A. M., Lopez, P., Trandafirescu, G., Gutierrez, M. A., Oikawa, H., & Noguchi, A. (2006). Determinant factors of spectrum of missense variants in mucopolysaccharidosis IVA gene. *Molecular genetics and metabolism*, 89(1), 139-149.
60. Tomatsu, S., Okamura, K., Maeda, H., Taketani, T., Castrillon, S. V., Gutierrez, M. A., & Noguchi, A. (2005a). Keratan sulphate levels in mucopolysaccharidoses and mucopolipidoses. *Journal of inherited metabolic disease*, 28(2), 187-202.
61. Tomatsu, S., Montano, A. M., Nishioka, T., Gutierrez, M. A., Peña, O. M., Lopez, P., & Orii, T. (2005b). Mutation and polymorphism spectrum of the GALNS gene in mucopolysaccharidosis IVA (Morquio A). *Human mutation*, 26(6), 500-512.
62. Tomatsu, S., Nishioka, T., Montañó, A. M., Gutierrez, M. A., Pena, O. S., Orii, K. O., & Noguchi, A. (2004). Mucopolysaccharidosis IVA: identification of mutations and methylation study in GALNS gene. *Journal of medical genetics*, 41(7), e98-e98.
63. Wiederstein, M., & Sippl, M. J. (2007). ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic acids research*, 35(suppl 2), W407-W410.

8. Anexo

***In silico* analysis of mutations occurring in the protein N-acetylgalactosamine-6-sulfatase (GALNS) and causing mucopolysaccharidosis IVA**

E.R. Tamarozzi, E. Torrieri, E.P. Semighini and S. Giuliatti

Departamento de Genética, Faculdade de Medicina de Ribeirão Preto,
Universidade de São Paulo, SP, Brasil

Corresponding author: E.R. Tamarozzi
E-mail: Elvira.tamarozzi@gmail.com

Genet. Mol. Res. 13 (4): 10025-10034 (2014)
Received April 29, 2014
Accepted August 7, 2014
Published November 28, 2014
DOI <http://dx.doi.org/10.4238/2014.November.28.7>

ABSTRACT. The goals were to analyze and characterize the secondary structure, regions of intrinsic disorder and physicochemical characteristics of three classes of mutations described in the enzyme N-acetylgalactosamine-6-sulfatase that cause mucopolysaccharidosis IVA: missense mutations, insertions and deletions. All mutations were compared to wild-type enzyme, and the results showed that with 25 of 129 missense mutations secondary structure was maintained and that 104 mutations showed minor changes, such as an increase or decrease in the size of the elements. The secondary structure of all insertions and deletions introduced important changes, such as increase in the number and size of elements. The results obtained from intrinsic disorder analysis revealed that missense mutations caused no alterations. However, the insertions and deletions led to major regions of intrinsic disorder. The physicochemical characteristics of the amino acids found in missense mutations revealed unchanged characteristics in 32 of the 129 mutations. However, the remainder had changes that could lead to a modification of tertiary structure.

The results proved that it was feasible and necessary to obtain the three-dimensional structure of the enzyme with its mutants to better understand the change in function.

Key words: Morquio syndrome; Bioinformatics; Protein structure

INTRODUCTION

Intrinsically disordered proteins perform several important roles in cellular activities, and they are related to a large number of diseases (Mizianty et al., 2011). The traditional structure-function relationship approach has, as one of its principles, the concept that the biological function of a protein is critically dependent on a well-defined conformational structure (Dunker et al., 2002). The deficiencies in lysosomal enzymes responsible for the catabolism of biomolecules lead to the accumulation of undegraded substrates in a wide range of human tissues and, eventually, to the so-called deposit diseases (Neufeld and Muenzer, 2001). N-acetylgalactosamine-6-sulfatase (GALNS) is the enzyme responsible for the first step in the degradation process of the glycosaminoglycans (GAG) called keratan sulfate (KS) and chondroitin-6-sulfate (C6S). It hydrolyzes the bond between GAG and the sulfate group linked to the 6th carbon of the saccharides, and its deficiency leads to the accumulation of KS and C6S in lysosomes, causing a cellular dysfunction known as mucopolysaccharidosis (MPV) IV, or Morquio syndrome (MS) (Pajares et al., 2012; Rivera-Colón et al., 2012; Kubaski et al., 2013).

MS is an autosomal recessive disease, acquired from healthy heterozygous parents carrying a mutation in band q24.3 of chromosome 16, a mutation that was discovered by *in situ* hybridization assays (Tomatsu et al., 2005). MS is characterized by severe osseous deformities and extra skeletal changes, such as corneal opacity, hepatomegaly, cardiac valve lesions, prognatism, diastema, thinning tooth enamel, and auditory changes that vary from conductive to sensorineural dysacusis. (Montaño et al., 2007; Couprie et al., 2010; Leal et al., 2010; Tomatsu et al., 2011; Hendriksz et al., 2013).

MS has two phenotypes related to its severity: a mild form, which generally allows the patient to have a normal life, and a severe form, which usually leads to the patient's death before the second decade of life.

Genetic analysis in different populations has shown a heterogeneity in the q24.3 mutation. However, the correlation between genotypes and phenotypes is not yet clearly understood (Tomatsu et al., 2011; Rivera-Colón et al., 2012; Pajares et al., 2012; Kubaski et al., 2013).

It is difficult to estimate the prevalence of MS, due to shortages of basic population studies and epidemiological data, which is estimated to be between 1 in 40,000 and 1 in 50,000 births (Northover et al., 1996; Nelson, 1997; Rivera-Colón et al., 2012). In Brazil, only the state of Rio Grande do Sul has data on the incidence of MS, showing 1 in every 275,000 live births (Schwartz et al., 2008).

Up to the submission of this paper, there was no known cure for MS. Current treatment consists of a multidisciplinary management with psychological, physiotherapeutic, nutritional, medical, and surgical interventions. These interventions are palliative and preventive and aim to delay the progress of the disease (Algahim and Almassi, 2013).

The most often reported mutations of GALNS enzyme that cause MS are classified in three categories: changes in the catalytic site, changes in the hydrophobic core and changes in

enzyme surface (Tomatsu et al., 2004; Rivera-Colón et al., 2012).

Tomatsu et al. (2005) reported a total of 148 mutations, with 103 of them being missense mutations, a number recently revised by Rivera-Colón et al. (2012) to 129 missense mutations. Since the tertiary structure of the GALNS enzyme is known (Rivera-Colón et al., 2012), the search for possible regions of intrinsic disorder in the mutant protein has substantial importance to enable protein modeling and virtual screening assays of these mutants, allowing the discovery of molecules with theoretic binding potential to the enzyme, and eventual pharmacological reactivation of the affected enzyme.

The purpose of this study was to perform *in silico* analysis of possible regions of intrinsic disorder (ID) and to compare the secondary structure of the wild-type protein with the 161 known (129 missense, 5 insertions, and 27 deletions) mutations, to determine if the increase in ID is responsible for the protein's loss of function.

MATERIAL AND METHODS

The genetic mutation data of the GALNS gene were described by Terzioglu (2002), Tomatsu et al. (2004), Tomatsu et al. (2005) and Rivera-Colón et al. (2012).

The FASTA sequence of GALNS has 522 amino acids and was downloaded in FASTA format from the NCBI public database, available at <http://www.ncbi.nlm.nih.gov/protein/P34059>.

The comparative analysis of chemical characteristics of wild type GALNS and the missense mutants followed the technique described by Nelson and Cox (2009).

The ID regions of the wild type and mutant proteins were predicted by the softwares meta Protein DisOrder prediction System - metaPrDOS (Ishida and Kinoshita, 2008), Prediction of Intrinsically Unstructured Proteins - IUPred (Dosztányi et al., 2005), Disorder Content predictor - DisCon (Mizianty et al., 2011), POODLE-S (Shimizu et al., 2005) and Prediction of Protein Binding Regions in Disordered Proteins - ANCHOR (Dosztányi et al., 2009).

The PSIPRED software (Buchan et al., 2013) was used to predict the secondary structures of the mutants.

RESULTS

All 161 mutations (Terzioglu et al., 2002; Tomatsu et al., 2004, 2005; Rivera-Colón et al., 2012) were individually analyzed, and the results were compared to the data for wild-type GALNS.

The results of predicted secondary structures, described from PSIPRED, revealed that wild-type GALNS has 9 α -helices and 9 β -sheets. It also revealed small structural changes at α -helices and/or β -sheets in 104 of the 129 missense mutations when compared to wild-type GALNS (Table 1). On the other hand, the 27 deletions and 5 insertions showed substantial changes in the size and number of α -helices and β -sheets in comparison with wild-type GALNS (Table 2).

The next step was the prediction of the ID regions, where there was no pronounced discrepancy between the results of the five software used to predict the ID regions, considering 5% as the margin of error.

The results showed that wild-type GALNS has an average of less than 2.3% (about 10 aminoacid residues) of ID in its whole length. Besides the differences between the methods

Table 1. GALNS mutations and its effects on secondary structures and intrinsic disorder (ID).

Mutation type	Protein region	Amino acid changed and mutation place	Average ID of the mutation	Secondary structure change	Structure changed (sheet/helix)
Missense	Surface	E 185 G	<2.4%	Yes	Both
Missense	Surface	N 204 A	<2.4%	Yes	Both
Missense	Surface	L 395 V	<2.4%	Yes	Both
Missense	Surface	H 398 D	<2.4%	Yes	Both
Missense	Surface	G 421 E	<2.4%	Yes	Both
Missense	Surface	Q 422 K	<2.4%	Yes	Both
Missense	Surface	F 452 I	<2.4%	Yes	Both
Missense	Surface	F 452 L	<2.4%	Yes	Both
Missense	Surface	S 53 F	<2.4%	Yes	Sheet
Missense	Surface	D 60 N	<2.4%	Yes	Sheet
Missense	Surface	G 155 E	<2.4%	Yes	Sheet
Missense	Surface	G 155 R	<2.4%	Yes	Sheet
Missense	Surface	F 167 V	<2.4%	Yes	Sheet
Missense	Surface	D 171 A	<2.4%	Yes	Sheet
Missense	Surface	R 259 Q	<2.4%	Yes	Sheet
Missense	Surface	S 295 F	<2.4%	Yes	Sheet
Missense	Surface	L 369 P	<2.4%	Yes	Sheet
Missense	Surface	W 409 S	<2.4%	Yes	Sheet
Missense	Surface	P 484 S	<2.4%	Yes	Sheet
Missense	Surface	V 488 M	<2.4%	Yes	Sheet
Missense	Surface	R 253 W	<2.4%	Yes	Helix
Missense	Surface	N 487 S	<2.4%	Yes	Helix
Missense	Surface	R 61 W	<2.4%	No	
Missense	Surface	G 247 D	<2.4%	No	
Missense	Surface	A 351 V	<2.4%	No	
Missense	Surface	P 357 L	<2.4%	No	
Missense	Surface	R 361 G	<2.4%	No	
Missense	Surface	L 366 F	<2.4%	No	
Missense	Surface	R 376 Q	<2.4%	No	
Missense	Surface	D 388 N	<2.4%	No	
Missense	Surface	L 395 P	<2.4%	No	
Missense	Surface	S 470 P	<2.4%	No	
Missense	Active Site	C 79 Y	<2.4%	Yes	Both
Missense	Active Site	H 142 R	<2.4%	Yes	Both
Missense	Active Site	S 80 L	<2.4%	Yes	Sheet
Missense	Active Site	H 236 D	<2.4%	Yes	Sheet
Missense	Active Site	N 289 S	<2.4%	No	
Missense	Active Site	K 310 N	<2.4%	No	
Missense	Hydrophobic Core	G 96 C	<2.4%	Yes	Both
Missense	Hydrophobic Core	F 97 V	<2.4%	Yes	Both
Missense	Hydrophobic Core	A 107 T	<2.4%	Yes	Both
Missense	Hydrophobic Core	S 135 R	<2.4%	Yes	Both
Missense	Hydrophobic Core	V 138 A	<2.4%	Yes	Both
Missense	Hydrophobic Core	W 141 C	<2.4%	Yes	Both
Missense	Hydrophobic Core	W 141 R	<2.4%	Yes	Both
Missense	Hydrophobic Core	P 151 S	<2.4%	Yes	Both
Missense	Hydrophobic Core	P 163 H	<2.4%	Yes	Both
Missense	Hydrophobic Core	H 166 Q	<2.4%	Yes	Both
Missense	Hydrophobic Core	A 291 T	<2.4%	Yes	Both
Missense	Hydrophobic Core	L 36 P	<2.4%	Yes	Sheet
Missense	Hydrophobic Core	M 41 L	<2.4%	Yes	Sheet
Missense	Hydrophobic Core	G 42 E	<2.4%	Yes	Sheet
Missense	Hydrophobic Core	G 47 R	<2.4%	Yes	Sheet
Missense	Hydrophobic Core	G 66 R	<2.4%	Yes	Sheet
Missense	Hydrophobic Core	F 69 V	<2.4%	Yes	Sheet
Missense	Hydrophobic Core	P 77 R	<2.4%	Yes	Sheet
Missense	Hydrophobic Core	T 88 I	<2.4%	Yes	Sheet
Missense	Hydrophobic Core	R 94 C	<2.4%	Yes	Sheet
Missense	Hydrophobic Core	R 94 G	<2.4%	Yes	Sheet
Missense	Hydrophobic Core	G 96 V	<2.4%	Yes	Sheet
Missense	Hydrophobic Core	Q 111 R	<2.4%	Yes	Sheet
Missense	Hydrophobic Core	G 116 S	<2.4%	Yes	Sheet
Missense	Hydrophobic Core	P 125 L	<2.4%	Yes	Sheet

Continued on next page

Table 1. Continued.

Mutation type	Protein region	Amino acid changed and mutation place	Average ID of the mutation	Secondary structure change	Structure changed (sheet/helix)
Missense	Hydrophobic Core	H 150 Y	<2.4%	Yes	Sheet
Missense	Hydrophobic Core	P 151 L	<2.4%	Yes	Sheet
Missense	Hydrophobic Core	W 159 C	<2.4%	Yes	Sheet
Missense	Hydrophobic Core	S 162 F	<2.4%	Yes	Sheet
Missense	Hydrophobic Core	N 164 T	<2.4%	Yes	Sheet
Missense	Hydrophobic Core	G 168 R	<2.4%	Yes	Sheet
Missense	Hydrophobic Core	P 179 H	<2.4%	Yes	Sheet
Missense	Hydrophobic Core	P 179 L	<2.4%	Yes	Sheet
Missense	Hydrophobic Core	A 203 V	<2.4%	Yes	Sheet
Missense	Hydrophobic Core	W 230 G	<2.4%	Yes	Sheet
Missense	Hydrophobic Core	V 239 F	<2.4%	Yes	Sheet
Missense	Hydrophobic Core	E 260 D	<2.4%	Yes	Sheet
Missense	Hydrophobic Core	F 284 V	<2.4%	Yes	Sheet
Missense	Hydrophobic Core	S 287 L	<2.4%	Yes	Sheet
Missense	Hydrophobic Core	G 290 S	<2.4%	Yes	Sheet
Missense	Hydrophobic Core	A 291 D	<2.4%	Yes	Sheet
Missense	Hydrophobic Core	G 301 C	<2.4%	Yes	Sheet
Missense	Hydrophobic Core	L 307 P	<2.4%	Yes	Sheet
Missense	Hydrophobic Core	G 309 R	<2.4%	Yes	Sheet
Missense	Hydrophobic Core	T 312 S	<2.4%	Yes	Sheet
Missense	Hydrophobic Core	G 316 V	<2.4%	Yes	Sheet
Missense	Hydrophobic Core	M 318 R	<2.4%	Yes	Sheet
Missense	Hydrophobic Core	A 324 E	<2.4%	Yes	Sheet
Missense	Hydrophobic Core	W 325 C	<2.4%	Yes	Sheet
Missense	Hydrophobic Core	G 340 D	<2.4%	Yes	Sheet
Missense	Hydrophobic Core	S 341 R	<2.4%	Yes	Sheet
Missense	Hydrophobic Core	M 343 L	<2.4%	Yes	Sheet
Missense	Hydrophobic Core	M 343 R	<2.4%	Yes	Sheet
Missense	Hydrophobic Core	D 344 N	<2.4%	Yes	Sheet
Missense	Hydrophobic Core	F 346 L	<2.4%	Yes	Sheet
Missense	Hydrophobic Core	R 380 S	<2.4%	Yes	Sheet
Missense	Hydrophobic Core	R 380 T	<2.4%	Yes	Sheet
Missense	Hydrophobic Core	R 386 H	<2.4%	Yes	Sheet
Missense	Hydrophobic Core	M 391 V	<2.4%	Yes	Sheet
Missense	Hydrophobic Core	A 392 V	<2.4%	Yes	Sheet
Missense	Hydrophobic Core	H 401 Y	<2.4%	Yes	Sheet
Missense	Hydrophobic Core	E 450 V	<2.4%	Yes	Sheet
Missense	Hydrophobic Core	D 233 N	<2.4%	Yes	Helix
Missense	Hydrophobic Core	R 90 W	<2.4%	No	
Missense	Hydrophobic Core	I 113 F	<2.4%	No	
Missense	Hydrophobic Core	G 139 S	<2.4%	No	
Missense	Hydrophobic Core	F 156 S	<2.4%	No	
Missense	Hydrophobic Core	P 179 S	<2.4%	No	
Missense	Hydrophobic Core	A 257 T	<2.4%	No	
Missense	Hydrophobic Core	T 312 A	<2.4%	No	
Missense	Hydrophobic Core	D 344 E	<2.4%	No	
Missense	Hydrophobic Core	L 345 P	<2.4%	No	
Missense	Hydrophobic Core	L 352 P	<2.4%	No	
Missense	Hydrophobic Core	R 386 C	<2.4%	No	
Missense	Hydrophobic Core	N 407 H	<2.4%	No	
Missense	Hydrophobic Core	M 494 V	<2.4%	No	
Missense	Undescribed	M 1 V	<2.4%	Yes	Both
Missense	Undescribed	W 10 X	<2.4%	Yes	Both
Missense	Undescribed	L 15 M	<2.4%	Yes	Both
Missense	Undescribed	Q 111 X	<2.4%	Yes	Both
Missense	Undescribed	K 129 X	<2.4%	Yes	Both
Missense	Undescribed	Q 148 X	<2.4%	Yes	Both
Missense	Undescribed	Q 211 X	<2.4%	Yes	Both
Missense	Undescribed	W 230 X	<2.4%	Yes	Both
Missense	Undescribed	W 325 X	<2.4%	Yes	Both
Missense	Undescribed	Q 338 X	<2.4%	Yes	Both
Missense	Undescribed	Q 374 X	<2.4%	Yes	Both
Missense	Undescribed	Q 422 X	<2.4%	Yes	Both
Missense	Undescribed	Q 473 X	<2.4%	Yes	Both
Missense	Undescribed	G 23 R	<2.4%	Yes	Sheet

Table 2. Secondary structure changes and intrinsic disorder (ID) of the 27 deletions and 5 insertions known in GALNS.

Mutation type	Mutation place	No. of deleted or inserted nucleotides	Rise greater than 2.5% of ID	ID percentage of the mutation	Major changes at secondary structures
Deletion	334	1	Yes	32.0%	Yes
Deletion	489	1	Yes	27.0%	Yes
Deletion	498	1	Yes	27.0%	Yes
Deletion	605	1	Yes	32.0%	Yes
Deletion	708	1	Yes	25.0%	Yes
Deletion	874	1	Yes	20.0%	Yes
Deletion	1168	1	Yes	15.0%	Yes
Deletion	1319	1	Yes	7.5%	Yes
Deletion	121 to 244	124	Yes	42.0%	Yes
Deletion	1213 to 1219	6	No	2.2%	Yes
Deletion	1290 to 1291	2	Yes	14.5%	Yes
Deletion	154 to 165	1	Yes	38.0%	Yes
Deletion	209 to 220	12	No	2.2%	Yes
Deletion	235 to 236	2	Yes	41.0%	Yes
Deletion	49 to 55	7	Yes	42.0%	Yes
Deletion	532 to 533	2	Yes	42.0%	Yes
Deletion	833 to 859	26	Yes	26.0%	Yes
Deletion	853 to 855	3	No	2.5%	Yes
Deletion	976 to 978	3	No	2.5%	Yes
Deletion	106 to 111	6	No	2.5%	Yes
Deletion	655	1	Yes	38.0%	Yes
Deletion	600	1	Yes	38.0%	Yes
Deletion	389	1	Yes	42.0%	Yes
Deletion	763	1	Yes	35.0%	Yes
Deletion	929	1	Yes	35.0%	Yes
Deletion	850 to 852	3	No	2.5%	Yes
Deletion	1343 to 1344	2	Yes	16.0%	Yes
Insertion	556	1	Yes	41%	Yes
Insertion	1232-1233	1	Yes	18%	Yes
Insertion	1177-1178	1	Yes	15.0%	Yes
Insertion	1355-1356	1	Yes	7.0%	Yes
Insertion	501-502	1	Yes	42.0%	Yes

of ID prediction, metaPrDOS, IUPred, DisCon, POODLE-S and ANCHOR, all the results of wild-type GALNS were consistent with each other.

The 129 missense mutations analyzed were grouped into four groups: 6 mutations at the catalytic site residues, 32 at the protein surface residues, 76 mutations at the protein hydrophobic core and 14 mutations at undescribed regions. All results were similar, without any considerable anomaly between the methods, with up to 2.4% of an average predicted ID for these mutations (Table 1), and also similar to those from the secondary structure predictions.

However, at the 27 deletions and 5 insertions, predicted ID varied from 2.2 to 42% (Table 2).

DISCUSSION

Tomatsu et al. (2005) and Rivera-Colón et al. (2012) showed that missense mutations [also called non synonymous substitution (Zhang et al., 2010, 2012)] are the most common cause of MPS IVA. It causes amino acid changes due to single nucleotide polymorphisms, which can drastically change protein structure and function. The specific mechanism that causes these mutations remains unknown. Yue and Moulton (2006) and Care et al. (2007) reported that non synonymous substitution mutations have an impact on the structural stability of the protein, altering its function. These data led us to analyze the secondary structures of the wild-type enzyme and its mutants, searching for the presence of regions with ID in the proteins.

According to Gnad et al. (2013), depending on its position and nature of the amino acid substitution of these mutations, it can cause a variety of changes of variable proportions in protein structure, affecting GALNS physiological function.

The missense mutations present in GALNS can cause changes in its active site, hydrophobic core and surface (Rivera-Colón et al., 2012). Zhang et al. (2011, 2012) reported that any change that affects the geometry of the active site (mutation that can be outside of the catalytic site) of a protein can, thus, affect its function. As our results showed, the changes in the secondary structures of missense mutant GALNS can explain the lower or inexistent enzymatic activity of these proteins. The analysis of the chemical characteristics of substituted amino acids has also showed that, in most of the missense mutations, they were modified, explaining then the predicted changes in the secondary structures.

Valdés et al. (2013) outlined that the ID, under physiological conditions, is common in a wide range of proteins. This phenomenon plays a key role in the biological activities of proteins that have a diverse range of molecular recognition patterns. To understand the role of ID in protein function, it is important to determine if the ID is part of its physiological role or, if the disorder is caused by mutations, leading to diseases such as synucleinopathies (Dill et al., 2007; Uversky et al., 2009).

It was expected to see ID caused by missense mutations in MPS IV-related GALNS. However, the mutant protein responsible for MPS IVA did not show ID regions, just like the wild-type protein, corroborating the findings of Rivera-Colón et al. (2012), showing GALNS as a very specific and conserved protein in vertebrates.

Protein stability is a key feature to protein function (Ye et al., 2006; Zhang et al., 2012), and a missense mutation may affect, besides the secondary structure, its stability, leading to misfolding (Dobson, 2003; Koukouritaki et al., 2007; Zhang et al., 2012). Our findings showed that, unexpectedly, even in missense mutations, without ID regions, when there are no changes in 19.4% of the secondary structures and in 25.6% of the chemical characteristics of

the amino acids, the function of GALNS is still suppressed in patients with MPS IVA. In the remaining part that showed minor changes in secondary structure (80.4%) and changes in the chemical characteristics of the amino acids (74.4%), no ID regions were observed. However, these changes may explain the suppression of protein function, since these variations can cause misfolding and also change the pattern of substrate recognition, which, in this case, is very specific to KS and C6S.

The missense mutations are not the only cause of MPS IVA. Maki (2002) pointed out that, in general terms, mutations involving deletions and insertions are less frequent than missense mutations. In analyzing the results of ID regions caused by insertions and deletions and the prediction of their secondary structures, we observed an increase in ID regions and wide changes in secondary structures, in contrast to that observed in missense mutations, but very expected, and that is probably the reason for the loss of enzyme function, since the whole amino acid sequence changes after a nucleotide deletion or insertion.

The prediction of secondary structures and the search for ID regions through *in silico* analysis of wild-type GALNS and its mutations led us to two different results. The results of the missense mutations analysis led us to infer, by observing their secondary structures conserved in more than 99% of its length, that it is possible to model their tertiary structures from wild-type GALNS. These results are reinforced by Uversky et al. (2009), who showed that proteins without ID have a more stable geometry at the active site in comparison to disordered proteins, increasing the chances of finding ligands to GALNS. On the other hand, the analysis of deletions and insertions showed it to be different from that obtained for missense mutations, due to major changes in the primary and secondary structure. These results do not discard the modeling of mutant GALNS and are very important in the understanding of MPS IVA and its phenotypic variation process.

CONCLUSION

Of all the 161 mutations analyzed, 80% were missense and did not show a significant percentage of ID increase when compared to wild-type GALNS. The 27 deletions and 5 insertions showed a high increase in ID in most of its structures. Secondary structure analysis revealed that 16% of the mutations, all missense, had no changes, 64% of the mutations, also all missense, showed small changes and 20% of the mutations (insertions and deletions) showed wide changes in secondary structures in comparison to wild-type GALNS. The analysis of the chemical features of the mutated amino acids showed that 25% of them were replaced by homologous amino acids.

On the basis of these data, it can be concluded that it is possible and viable to use wild-type GALNS to model the structures of 80.1% of the mutations and then to use these models to understand how these mutations affect protein catalytic activity.

REFERENCES

- Algham MF and Almassi GH (2013). Current and emerging management options for patients with Morquio A syndrome. *Ther. Clin. Risk Manag.* 9: 45-53.
- Buchan DW, Minneci F, Nugent TC, Bryson K, et al. (2013). Scalable web services for the PSIPRED protein analysis workbench. *Nucleic Acids Res.* 41: W349-W357.
- Care MA, Needham CJ, Bulpitt AJ and Westhead DR (2007). Deleterious SNP prediction: be mindful of your training data! *Bioinformatics* 23: 664-672.

- Couprie J, Denis P, Guffon N, Reynes N, et al. (2010). Manifestations ophtalmologiques de la maladie de Morquio. *J. Français d'Ophtalmol.* 33: 617-622.
- Dill KA, Ozkan SB, Weikl TR, Chodera JD, et al. (2007). The protein folding problem: when will it be solved? *Curr. Opin. Struct. Biol.* 17: 342-346.
- Dobson CM (2003). Protein folding and misfolding. *Nature* 426: 884-890.
- Dosztányi Z, Csizmók V, Tompa P and Simon I (2005). IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. *Bioinformatics* 21: 3433-3434.
- Dosztányi Z, Meszaros B and Simon I (2009). ANCHOR: web server for predicting protein binding regions in disordered proteins. *Bioinformatics* 25: 2745-2746.
- Dunker AK, Brown CJ, Lawson JD, Iakoucheva LM, et al. (2002). Intrinsic disorder and protein function. *Biochemistry* 41: 6573-6582.
- Gnad F, Baucom A, Mukhyala K, Manning G, et al. (2013). Assessment of computational methods for predicting the effects of missense mutations in human cancers. *BMC Genomics* 14 (Suppl 3): S7.
- Hendriksz CJ, Al-Jawad M, Berger KI, Hawley SM, et al. (2013). Clinical overview and treatment options for non-skeletal manifestations of mucopolysaccharidosis type IVA. *J. Inherit. Metab. Dis.* 36: 309-322.
- Ishida T and Kinoshita K (2008). Prediction of disordered regions in proteins based on the meta approach. *Bioinformatics* 24: 1344-1348.
- Koukouritaki SB, Poch MT, Henderson MC, Siddens LK, et al. (2007). Identification and functional analysis of common human flavin-containing monooxygenase 3 genetic variants. *J. Pharmacol. Exp. Ther.* 320: 266-273.
- Kubaski F, Brusius-Facchin AC, Palhares HM, Balarin MA, et al. (2013). Identification of a novel missense mutation in Brazilian patient with a severe form of mucopolysaccharidosis type IVA. *Gene* 517: 112-115.
- Leal GN, de Paula AC, Leone C and Kim CA (2010). Echocardiographic study of paediatric patients with mucopolysaccharidosis. *Cardiol. Young* 20: 254-261.
- Maki H (2002). Origins of spontaneous mutations: specificity and directionality of base-substitution, frame shift, and sequence-substitution mutageneses. *Annu. Rev. Genet.* 36: 279-303.
- Mizianty MJ, Zhang T, Xue B, Zhou Y, et al. (2011). In-silico prediction of disorder content using hybrid sequence representation. *BMC Bioinformatics* 12: 245.
- Montaño AM, Tomatsu S, Gottesman GS, Smith M, et al. (2007). International Morquio A Registry: clinical manifestation and natural course of Morquio A disease. *J. Inherit. Metab. Dis.* 30: 165-174.
- Nelson DL and Cox MM (2009). *Lehninger Principles of Biochemistry*. 5th edn. W.H. Freeman Publishers, New York.
- Nelson J (1997). Incidence of the mucopolysaccharidoses in Northern Ireland. *Hum. Genet.* 101: 355-358.
- Neufeld EF and Muenzer J (2001). The Mucopolysaccharidoses. In: *The Metabolic & Molecular Bases of Inherited Disease*. 8th edn. Vol. 3. (Scriver CR, Beaudet AL, Sly WS and Valle D, eds.). McGraw-Hill, New York.
- Northover H, Cowie RA and Wraith JE (1996). Mucopolysaccharidosis type IVA (Morquio syndrome): a clinical review. *J. Inherit. Metab. Dis.* 19: 357-365.
- Pajares S, Alcalde C, Couce ML, Del TM, et al. (2012). Molecular analysis of mucopolysaccharidosis IVA (Morquio A) in Spain. *Mol. Genet. Metab.* 106: 196-201.
- Rivera-Colón Y, Schutsky EK, Kita AZ and Garman SC (2012). The structure of human GALNS reveals the molecular basis for mucopolysaccharidosis IV A. *J. Mol. Biol.* 423: 736-751.
- Schwartz I, Martins TF and Federhen A (2008). A Rede MPS Faz Quatro Anos: Um Panorama Geral sobre a Epidemiologia das MPS no Brasil. In: *Anais do XX Congresso Brasileiro de Genética Médica*, Brasil.
- Shimizu K, Muraoka Y, Hirose S and Noguchi T (2005). Feature Selection Based on Physicochemical Properties of Redefined N-term Region and C-term Regions for Predicting Disorder, *Proc of IEEE CIBCB*, 262-267.
- Terzioglu M, Tokatli A, Coskun T and Emre S (2002). Molecular analysis of Turkish mucopolysaccharidosis IVA (Morquio A) patients: identification of novel mutations in the N-acetylgalactosamine-6-sulfate sulfatase (GALNS) gene. *Hum. Mutat.* 20: 477-478.
- Tomatsu S, Nishioka T, Montaño AM, Gutierrez MA, et al. (2004). Mucopolysaccharidosis IVA: identification of mutations and methylation study in GALNS gene. *J. Med. Genet.* 41: e98.
- Tomatsu S, Montaño AM, Nishioka T, Gutierrez MA, et al. (2005). Mutation and polymorphism spectrum of the GALNS gene in mucopolysaccharidosis IVA (Morquio A). *Hum. Mutat.* 26: 500-512.
- Tomatsu S, Montaño AM, Oikawa H, Smith M, et al. (2011). Mucopolysaccharidosis type IVA (Morquio A disease): clinical review and current treatment. *Curr. Pharm. Biotechnol.* 12: 931-945.
- Uversky VN, Oldfield CJ, Midic U, Xie H, et al. (2009). Unfoldomics of human diseases: linking protein intrinsic disorder with diseases. *BMC Genomics* 10 (Suppl 1): S7.
- Valdés JJ, Schwarz A, Cabeza de Vaca I and Calvo E (2013). Tryptogalinin is a tick kunitz serine protease inhibitor with a unique intrinsic disorder. *PLoS One* 8: 62562.

- Ye Y, Li Z and Godzik A (2006). Modeling and Analyzing Three-Dimensional Structures of Human Disease Proteins. Proceedings of the Pacific Symposium on Biocomputing, Hawaii, 439-450.
- Yue P and Moulton J (2006). Identification and analysis of deleterious human SNPs. *J. Mol. Biol.* 356: 1263-1274.
- Zhang Z, Teng S, Wang L, Schwartz CE, et al. (2010). Computational analysis of missense mutations causing Snyder-Robinson syndrome. *Hum. Mutat.* 31: 1043-1049.
- Zhang Z, Norris J, Schwartz C and Alexov E (2011). *In silico* and *in vitro* investigations of the mutability of disease-causing missense mutation sites in spermine synthase. *PLoS One* 6: e20373.
- Zhang Z, Miteva MA, Wang L and Alexov E (2012). Analyzing effects of naturally occurring missense mutations. *Comput. Math. Methods Med.* 2012: 1-15.