

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

Programa de Pós-Graduação em Genética e Biologia Molecular
- PPGBM

**Comparação entre métodos de modelagem molecular
aplicados à enzima α -L-Iduronidase e análise dos
mutantes R89W e R89Q**

Aluna: Danieli Forgiarini Figueiredo

Dissertação submetida ao Programa de Pós-Graduação em Genética e Biologia Molecular da UFRGS como requisito parcial para a obtenção do grau de **Mestre em Genética e Biologia Molecular**.

Orientadora: Dra. Marialva Sinigaglia.

Co-orientador: Dr. Gustavo Fioravanti Vieira

Porto Alegre, agosto de 2013

AGRADECIMENTOS

Em primeiro momento agradeço pela minha orientadora por ter me incentivado nos momentos mais difíceis que passei tanto pessoal quanto acadêmico. Pelo perfeccionismo nas evidências do trabalho e por mostrar que por mais difícil que seja um projeto de pesquisa sempre temos a capacidade de extrair o máximo de informações deste

Ao meu co-orientador Gustavo por toda ajuda no desenvolvimento do trabalho e na geração das imagens. Agradeço também por transformar o ambiente de trabalho mais leve e divertido, por ser direto e preciso em relação as reuniões da IDUA e por mostrar sempre o “foco” do trabalho.

Fico grata também pela segunda família, da qual pude escolher, durante esse tempo, a Equipe NBLI, por me acolher e ser tão unida tanto nos momento acadêmicos quando pessoais, em especial aos colegas Dinler e Maurício, por me ajudarem no fechamento através de discussões, opiniões e ajudas com os gráficos e imagens. Além de me ajudarem no desenvolvimento da dinâmica molecular da IDUA. Ao Marcus por me auxiliar nos gráficos da Dinâmica e no programa Chimera.

Agradeço ao secretário do PPGBM Elmo, por ser sempre cordial e atencioso com dúvidas frequentes que apresentava sobre o PPGBM.

A minha querida avó paterna Odete, pelo incentivo aos estudos e dedicação a vida acadêmica e também pelo seu carinho incondicional.

Aos meus padrinhos Marta e Milton que sempre me incentivaram e me acolheram como pais.

Aos meus irmãos, pela parceria e respeito.

As minhas amigas Mayara, Carla, Karina e Temis por entender minha ausência na época do mestrado

RESUMO

A Mucopolissacaridose do tipo I é uma doença autossômica recessiva, causada por mutações na enzima α -L-iduronidase, a qual é responsável pela degradação dos glicosaminoglicanos heparan sulfato e dermatan sulfato. Pacientes MPS I apresentam um amplo espectro de manifestações clínicas que vão desde a forma mais grave da doença (Síndrome de Hurler) até a forma mais branda (síndrome de Scheie). A gravidade da doença está intimamente relacionada a mutações na enzima α -L-iduronidase, as quais podem ter um impacto tanto na estabilidade quanto na atividade enzimática. O objetivo principal deste trabalho foi realizar a modelagem molecular da enzima α -L-iduronidase humana utilizando diferentes metodologias. A partir dessas modelagens realizamos comparações com a estrutura cristalografada dessa proteína (PDB ID: 4JXP - *publicado após a obtenção dos modelos*), com o intuito de detectar o método mais apropriado para modelagens que apresentam moldes de baixa identidade. Adicionalmente, avaliamos através da dinâmica molecular duas mutações - R89W e R89Q - que ocorrem no mesmo resíduo da IDUA e apresentam impactos diferenciados sobre a função da proteína. Dentre as técnicas de modelagem utilizadas nesse trabalho as que apresentaram os melhores desempenhos foram a modelagem mista (utilizando a ferramenta I-tasser) bem como a modelagem por homologia (através da ferramenta Modeller 9v9) utilizando um alinhamento baseado em reconhecimento de padrões de enovelamento. Em relação à validação dos modelos destacamos a importância da utilização de diferentes ferramentas de validação, bem como a inclusão da simulação de dinâmica molecular como passo final na avaliação dos modelos. Com relação aos dados de dinâmica molecular, a substituição de uma arginina (R) na posição 89 por um triptofano (W) apresentou um impacto maior no sítio ativo da IDUA do que a substituição por uma glutamina (Q). Esse impacto se refletiu tanto no número de ligações de hidrogênio realizadas pelo resíduo na posição 89 e nas distâncias dos dois glutamatos catalíticos (E182 e E299) quanto na desestruturação da hélice α 6. A metodologia utilizada neste trabalho pode ser

aplicada para a modelagem das outras 5 enzimas da via de degradação de glicosaminoglicanos (responsáveis pelas mucopolissacaridoses II, IIIA, IIIB, IIIC e IIID) que ainda não apresentam sua estrutura cristalografada.

ABSTRACT

The Mucopolysaccharidosis type I is an autosomal recessive disorder caused by mutations in the alpha-L-iduronidase enzyme, which is responsible for degradation of heparan and dermatan sulfate glycosaminoglycans. MPS I patients have a wide spectrum of clinical manifestations ranging from the most severe (Hurler syndrome) to the mildest (Scheie syndrome) form of the disease. The severity of the disease is closely related to mutations in the α -L-iduronidase, which may have an impact both in stability and activity of the enzyme. The main objective of this work was to perform a molecular modeling of the human α -L-iduronidase enzyme through different methodologies. From these, we compare the models with the crystal structure of IDUA (PDB ID: 4JXP - released after the models construction), in order to detect the most suitable method for modeling presenting templates with low identity (below 25%). In addition, molecular dynamics was applied to evaluate two mutations - R89W and R89Q - which occur at the same residue and have different impacts on IDUA protein function. Among the modeling techniques used in this work, the threading approach (using the I-Tasser) as well the homology modeling (through the Modeller 9v9) using a folding pattern recognition-based alignment showed the best performances. Regarding the models validation, we highlight the importance of different tools as well the inclusion of molecular dynamics simulation as the final step in the evaluation of the models. The molecular dynamics results demonstrated that the substitution of an arginine (R) at position 89 by a tryptophane (W) had a greater impact on the IDUA active site structure than a replacement for a glutamine (Q). This is reflected both in the number of hydrogen bonds made by the residue 89 and the distances between the two catalytic glutamates (E299 and e182), but also by the disruption of helix α 6. The methodology applied in this work can be used to model other enzymes involved in the pathway of glycosaminoglycans degradation (responsible for mucopolysaccharidosis II, IIIA, IIIB, IIIC and IIID) that have not yet crystal

structures available.

LISTA DE ABREVIATURAS

- 4JXO- PDB do cristal da enzima α -L- iduronidase humana
- 4JXP- PDB do cristal da enzima α -L- iduronidase humana
- CHO-células de ovário de hamster chinês
- D298- Ácido Aspártico 298
- DM- Dinâmica molecular
- DS- Dermatan Sulfato
- E131- Glutamato 131
- E178- Glutamato 178
- E182 -Glutamato 182
- E299 -Glutamato 299
- EIM- Erros Inatos do Metabolismo
- F46- Fenilalanina 46
- F48 – Fenilalanina 48
- GAGS- Glicosaminoglicanos
- HGMD -The Human Gene Mutation Database
- HS- Heparan Sulfato
- IDUA – RM- Modelo IDUA (alinhamento Rempel- MODELLER)
- IDUA- α -L-iduronidase
- IDUA –PM- Modelo IDUA (alinhamento Phyre- MODELLER)
- LSD- Doenças de depósito lisossômico
- M-6-P- Manose 6 fosfato
- M6PR - Receptor de Manose-6-fosfato
- MPS - Mucopolissacaridose
- MPS IH – Síndrome de Hurler

MPS S- Síndrome de Scheie

MPSII- Mucopolissacaridose tipo II

MPSIII- Mucopolissacaridose tipo III

MPSI- Mucopolissacaridose tipo I

MPSIV- Mucopolissacaridose tipo IV

N181- Asparagina 181

N336 - Asparagina 336

N372- Asparagina 372

N451- Asparagina 451

PDB- Banco de dados de Proteínas

R89- Arginina 89

R89Q- Substituição Arginina 89 Glutamina

R89W- Substituição Arginina 89 Triptofano

RMSD- Desvio Médio Quadrático

RMSF- Desvio médio quadrático Flutuante

T50- Treonina 50

TRE- terapia de reposição enzimática

W89- Triptofano 89

LISTA DE FIGURAS

- Figura 1:** Via de degradação do heparan (direita) e do dermatan sulfato (esquerda). Adaptado de (Neufeld & Muenzer, 2001). 16
- Figura 2:** Transporte das enzimas lisossômicas em uma célula normal. O resíduo de M6P é adicionado à enzima lisossômica no Complexo de Golgi que por sua vez se liga aos receptores de M6P. O complexo enzima-receptor é entregue aos endossomos onde o Ph baixo promove a dissociação do mesmo. A enzima migra ao lisossomo maduro e o receptor M6P é reciclado ao Aparato de Golgi. Uma pequena quantidade de M6P é fosforilado e escapa da captura pelos receptores de M6P intracelulares sendo liberada para o exterior da célula. Esta enzima pode ser recapturada pela ligação ao receptor de M6P localizado na superfície da vesícula, recoberta por clatrina. Adaptado de (Muenzer & Fisher, 2004). 19
- Figura 3:** Domínio catalítico da α -L-iduronidase. Sobreposição dos modelos com a estrutura cristalográfica da IDUA (4JXP). O cristal está representado em cinza claro. A seta em vermelho destaca os dois grampos- β que protraem do barril.39
- Figura 4:** Comparação entre as estruturas cristalográficas de representantes da família 39 de Glicosil hidrolases. (A) beta-D-xylosidase de *T. saccharolyticum* (1UHV) e (B) α -L-iduronidase humana (4JXP). Em vermelho o domínio Barril ($\alpha\beta$)₈, em lilás o domínio β -sanduíche e em verde o domínio que se assemelha ao da fibronectina. Em amarelo são destacados os grampos- β que protraem do Barril.41
- Figura 5:** Avaliação ao longo de 50ns de simulação da estrutura secundária do cristal (4JXP) e dos modelos de IDUA gerados.....42
- Figura 6:** Comparação entre o desvio médio quadrático (RMSD) dos modelos de IDUA gerados através de diferentes metodologias e o Modelo-4JXP, ao longo de 50 ns.43
- Figura 7:** Desvio médio quadrático (RMSD) do domínio catalítico da IDUA

(excluindo as regiões de alça) ponderado pela dinâmica do modelo-4JXP.....	44
Figura 8: RMSF do domínio catalítico dos modelos de IDUA ao longo de 50 ns de simulação	46
Figura 9: Desvio médio quadrático do domínio catalítico da IDUA selvagem e mutante ao longo de 50 ns de simulação.....	51
Figura 10: Raio de giro do domínio catalítico da IDUA selvagem e mutante ao longo de 50 ns.....	52
Figura 11: Flutuação quadrática média dos resíduos do domínio catalítico da IDUA.....	53
Figura 12: Variação da distância entre os resíduos catalíticos ao longo de 50 ns de simulação. (A) Distância absoluta, linha vermelha representa a região abaixo de 5,5 Å. (B) Distância média. (C) sobreposição da estrutura inicial após minimização de energia (tonalidade clara) com a estrutura no final da simulação (50ns).....	55
Figura 13: Impacto das mutações R89W e R89Q em relação ao número de ligações de hidrogênio realizadas pelo resíduo 89. É possível observar uma redução no número de ligações da mutante R89W, quando comparada com a proteína selvagem. Gráfico superior apresenta o número absoluto de ligações de hidrogênio ao longo do tempo, enquanto o gráfico inferior apresenta uma média dos pontos do gráfico superior.	59
Figura 14: Comparação do domínio catalítico da IDUA selvagem e dos mutantes no início e ao final de 50 ns de simulação. Tonalidades mais claras representam a estrutura inicial, tonalidades mais escuras a estrutura final aos 50 ns. A hélice $\alpha 6$ é indicada.....	62

LISTA DE TABELAS

Tabela 1: Gravidade e comprometimento neurológico de pacientes MPS I de acordo com o acúmulo de GAGs.17

Tabela 2: Avaliação dos modelos de IDUA através do Gráfico de Ramachandran.36

Tabela 3: Avaliação dos modelos de IDUA nos programas ModFOLD e Verify 3D.36

Tabela 4: Comparação entre os modelos de IDUA e a estrutura cristalográfica 4JXP considerando o domínio catalítico e a proteína total.....40

Tabela 5: Ligações de hidrogênio formadas pelo resíduo 89 em diferentes pontos da simulação. Arginina na proteína selvagem, W e Q nos mutantes.60

Sumário

1	Introdução	14
1.1	Epidemiologia e História.....	14
1.1.1	Doenças de Depósito lisossômico	14
1.1.2	Mucopolissacaridose	14
1.1.3	Mucopolissacaridose tipo I	15
1.2	A Enzima Iduronidase, Biossíntese e Tráfego.....	17
1.2.1	Caracterização da enzima.....	19
1.3	O Impacto de Diferentes Mutações na IDUA em Relação ao Fenótipo da MPSI.....	20
1.4	Diagnóstico	20
1.5	Tratamentos Para a Mucopolissacaridose Tipo I.....	21
1.5.1	Terapia de Reposição Enzimática.....	21
1.5.2	Transplante de Células Tronco Hematopoiéticas	22
1.5.3	Terapia utilizando Chaperonas Farmacológicas	22
1.6	Modelagem Molecular de Proteínas.....	23
1.7	Justificativa do Trabalho.....	27
2	Objetivos	29
2.1	Objetivo Geral	29
2.2	Objetivos Específicos.....	29
3	Metodologia.....	31
3.1	Predição da estrutura tridimensional da proteína α -L-Iduronidase.....	31
3.1.1	Modelo IDUA I-tasser.....	31
3.1.2	Modelo IDUA - Phyre 2.0/Modeller (IDUA - PM)	31

3.1.3	Modelo IDUA - Rempel e cols /Modeller (IDUA - RM).....	32
3.2	Simulação de Dinâmica Molecular	32
3.2.1	Simulações envolvendo análise dos modelos gerados por homologia a partir do molde 1UHV e pela técnica de "threading"	33
3.2.2	Simulações de modelos gerados a partir do cristal 4JXP para avaliação do impacto das mutações R89W e R89Q na estrutura da Idua	34
4	Resultados e Discussão.....	35
5	Conclusões	68
6	Perspectivas.....	69
7	Referências.....	67

1 Introdução

1.1 *Epidemiologia e História*

1.1.1 **Doenças de Depósito lisossômico**

As doenças de depósito lisossômico (DLs), também conhecidas como doenças de acúmulo, englobam um grupo de aproximadamente 50 desordens genéticas causadas por alterações na atividade, modificações pós-traducionais e no tráfego de hidrolases lisossomais, bem como em proteínas de membrana e de transporte (Parenti, 2009). Estas alterações ocasionam um acúmulo gradual de glicosaminoglicanos, gangliosídeos, glicogênio, glicoesfingolípídeos e glicoproteínas nos lisossomos (Hartung et al., 2004). Estima-se que em cada 7.000 nascimentos vivos uma criança seja portadora de doença lisossômica. Estes dados são preocupantes, principalmente se considerarmos que de todas as doenças de erros inatos do metabolismo (EIM) detectadas no Laboratório de Referência para EIM do Hospital de Clínicas de Porto Alegre, as DLs correspondem a 59,8% (Mayer, 2011). Estas doenças são conhecidas e tradicionalmente classificadas de acordo com as propriedades químicas dos componentes acumulados.

As DLs afetam vários tecidos e órgãos ocasionando fenótipos complexos, com variações nas manifestações neurológicas, digestivas e esqueléticas (Parenti, 2009). De maneira geral, as doenças relacionadas ao sistema nervoso central apresentam um caráter progressivo promovendo alta debilidade cognitiva ao paciente.

1.1.2 **Mucopolissacaridose**

A Mucopolissacaridose (MPS) é uma doença genética rara causada pela deficiência de enzimas lisossômicas específicas que afetam o catabolismo de glicosaminoglicanos (GAGs). A via de degradação das GAGs é composta por 11 enzimas que quando deficientes ocasionam sete tipos diferentes de MPS (Neufeld & Muenzer, 2001). Os glicosaminoglicanos compõem a matriz extracelular constituindo-se no seu componente estrutural fundamental. São formados por longas cadeias polissacarídicas não ramificadas compostas por unidades dissacarídicas repetitivas de ácido urônico e de N-acetilglicosamina ou N-

acetilgalactosamina. Muitas destas moléculas são sulfatadas em toda a sua extensão conferindo a molécula uma carga negativa (Winchester, 2005). Outra característica das GAGs é que elas são extremamente hidrofílicas e desta forma adotam uma conformação estendida em forma de gel. Essa característica é essencial para a difusão de nutrientes, sinalização celular, comunicação célula-célula bem como para a formação da cartilagem (fornecendo grande resistência a compressão). As GAGs podem ser encontradas no organismo em uma forma livre ou apresentar-se ligados a um núcleo proteico (por ligações covalentes), constituindo assim os chamados proteoglicanos. É importante salientar que os glicosaminoglicanos são produtos da clivagem dos proteoglicanos. No meio intracelular a degradação ocorre por ação das enzimas lisossômicas gerando monossacarídeos e sulfato inorgânico (Hopwood & Morris, 1990; Sánchez & Fuentes, 2001).

Pacientes que possuem deficiência de qualquer enzima da via de degradação lisossômica de GAGs apresentam um acúmulo deste substrato no interior dos lisossomos em vários órgãos e tecidos. A MPS resulta em uma série de sinais e sintomas, integrantes de um quadro clínico multissistêmico. As gravidades das MPS estão correlacionadas com o quadro genótipo-fenótipo sendo que dentre as manifestações mais frequentes destacam-se: face característica, opacificação da córnea e/ou outros problemas oculares, macroglossia, perda auditiva, hidrocefalia, cardiopatia, problemas respiratórios, hepatoesplenomegalia, hérnia inguinal e umbilical, disostose múltipla, limitação da mobilidade articular e dano cognitivo (em parte dos casos). O paciente com MPS está sujeito a várias cirurgias durante o transcorrer de sua vida uma vez que muitos órgãos e tecidos encontram-se comprometidos (Parenti, 2009; Sánchez & Fuentes, 2001).

1.1.3 Mucopolissacaridose tipo I

A mucopolissacaridose tipo I (MPS I) é uma doença lisossomal que apresenta uma herança autossômica recessiva, causada por alterações na atividade da enzima α -L-iduronidase (IDUA). Como demonstrado na (**Figura 1**), esta é a segunda enzima da rota de degradação das GAGs e é responsável pela

clivagem dos resíduos terminais do ácido idurônico provenientes de dermatan sulfato e heparan sulfato (Neufeld & Muenzer, 2001). A ausência ou mau funcionamento desta enzima ocasiona acúmulo de GAGs parcialmente degradadas no lisossomo.

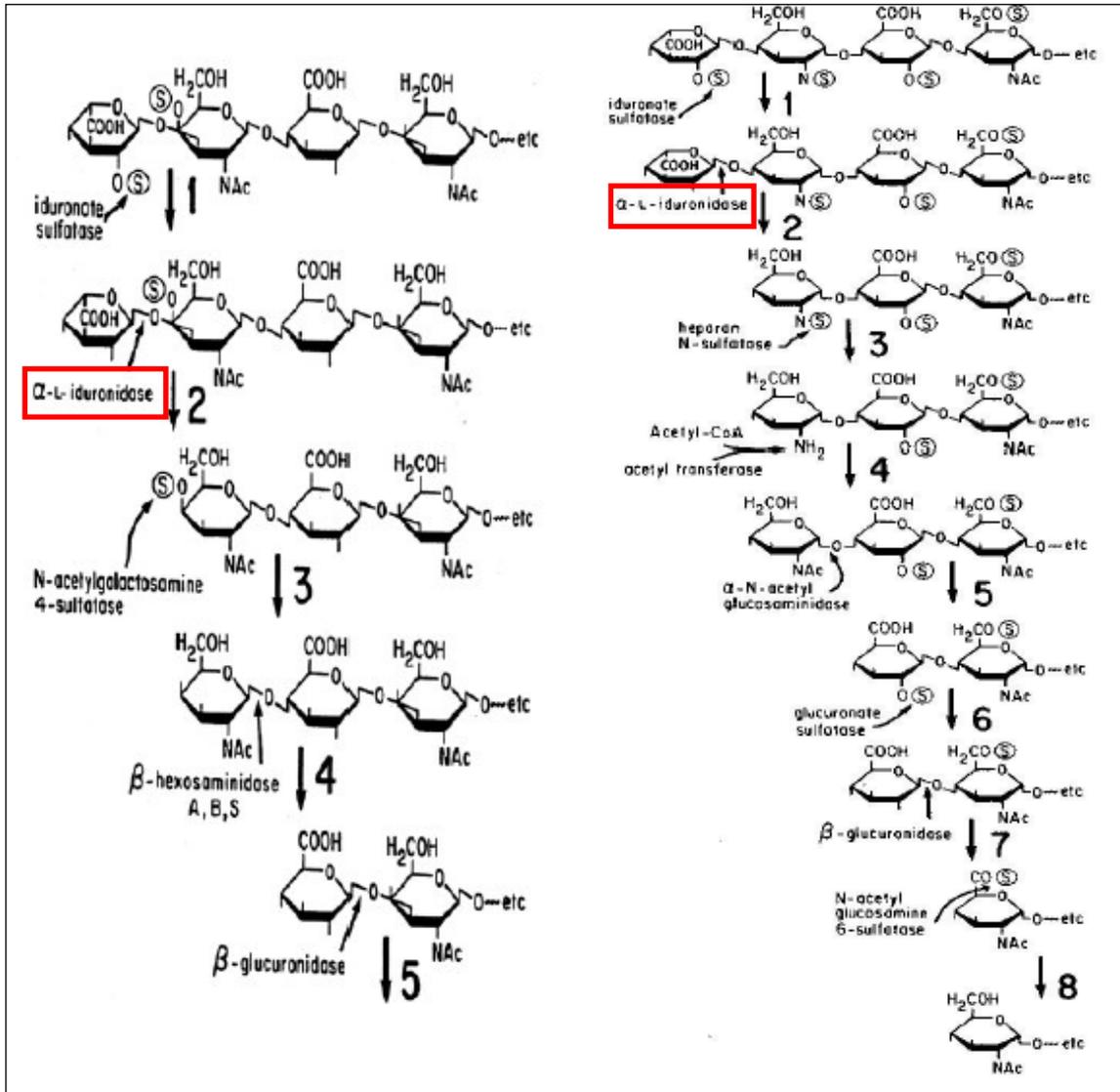


Figura 1: Via de degradação do heparan (direita) e do dermatan sulfato (esquerda). Adaptado de (Neufeld & Muenzer, 2001).

De acordo com a sua gravidade e comprometimento neurológico a MPS I é clinicamente dividida em 3 tipos: (i) Síndrome de Hurler, (ii) Síndrome de Hurler-Scheie e (iii) Síndrome de Scheie. Como se pode observar na **Tabela 1** a Síndrome de Hurler é a forma grave da doença, os sintomas aparecem cedo por

volta de dois anos de idade e a expectativa média de vida na ausência de tratamento é de aproximadamente 10 anos. Já a Síndrome de Hurler-Scheie constitui-se na forma intermediária onde o início dos sintomas ocorre mais tarde, por volta dos 3 a 8 anos de idade. Os indivíduos que apresentam esta síndrome possuem uma expectativa de vida maior, entre 20 e 30 anos. Por sua vez a Síndrome de Scheie refere-se à forma leve da doença e o indivíduo apresenta uma expectativa média de vida normal (Neufeld & Muenzer, 2001; Pastores *et al.*, 2007; Soliman *et al.*, 2007). Embora exista esta classificação clínica, o que se observa é que existe uma grande heterogeneidade e sobreposição dos sintomas descritos na tabela 1, sendo que alguns autores sugerem que é mais adequado classificar o paciente como apresentando a forma severa ou leve (Vijay & Wraith, 2005).

Tabela 1: Gravidade e comprometimento neurológico de pacientes MPS I de acordo com o acúmulo de GAGs.

	Síndrome de Hurler *	Síndrome de Hurler e Scheie *	Síndrome de Scheie *
Forma	<i>Grave</i>	<i>Moderada</i>	<i>Leve</i>
Início dos sintomas	Em torno de 2 anos de idade, sintomas progridem rapidamente.	3 a 8 anos de idade	5 a 15 anos de idade, progressão lenta dos sintomas
Expectativa média de vida	10 anos	20-30 anos	Normal, na maioria das vezes.
Principais Sintomas	Atrasos no desenvolvimento e estatura baixa (110 cm), características faciais grosseiras, insuficiência respiratória, doença cardíaca valvular, otite média recorrente, córnea opaca, disostose múltipla, e rigidez nas articulações, etc.	Opacificação da córnea, perda da audição, enrijecimento articular e problemas cardíacos, ausência de comprometimento cognitivo, podendo apresentar leves dificuldades na aprendizagem.	Rigidez articular, valvulopatia aórtica, opacidade da córnea entre outras. Curso clínico caracterizado por problemas ortopédicos, inteligência normal.

(Neufeld & Muenzer, 2001; Pastores *et al.*, 2007; Soliman *et al.*, 2007)

1.2 A Enzima Iduronidase, Biossíntese e Tráfego.

A enzima lisossomal α -L-iduronidase (IDUA) (EC: 3.2.1.76), é responsável pela clivagem de GAGs principalmente o heparan sulfato e o dermatan sulfato no

interior do lisossomo. Esta enzima é sintetizada no retículo endoplasmático rugoso como uma cadeia polipeptídica imatura contendo 653 aminoácidos. A região N-terminal (26 aminoácidos) que faz parte do peptídeo sinal é clivada dando origem à proteína madura. A sequência também apresenta seis sítios de glicosilação importantes para o endereçamento da proteína ao lisossomo (Schuchman & Desnick, 1988).

As enzimas que serão encaminhadas ao lisossomo são marcadas com um grupo de Manose-6-fosfato (M6P) no complexo de golgi, o qual é reconhecido pelos receptores transmembrana de M6P. Estas proteínas de membrana se ligam às hidrolases lisossomais e irão montar as vesículas de clatrina, que brotam da fase *trans-golgi*. Os conteúdos por sua vez, serão entregues aos endossomos, onde o baixo pH da organela causa a dissociação da enzima com o receptor M6P (M6PR), e finalmente chega ao lisossomo maduro (pH 5). Uma vez no lisossomo, estas hidrolases podem iniciar a digestão do material endossitado (**Figura 2**). A Via de receptores M6P é a maior rota de alvos lisossomais de hidrolases e é amplamente estudada devido ao seu envolvimento nas DLs (Coutinho *et al.*, 2012; Muenzer & Fisher, 2004; Muro, 2010).

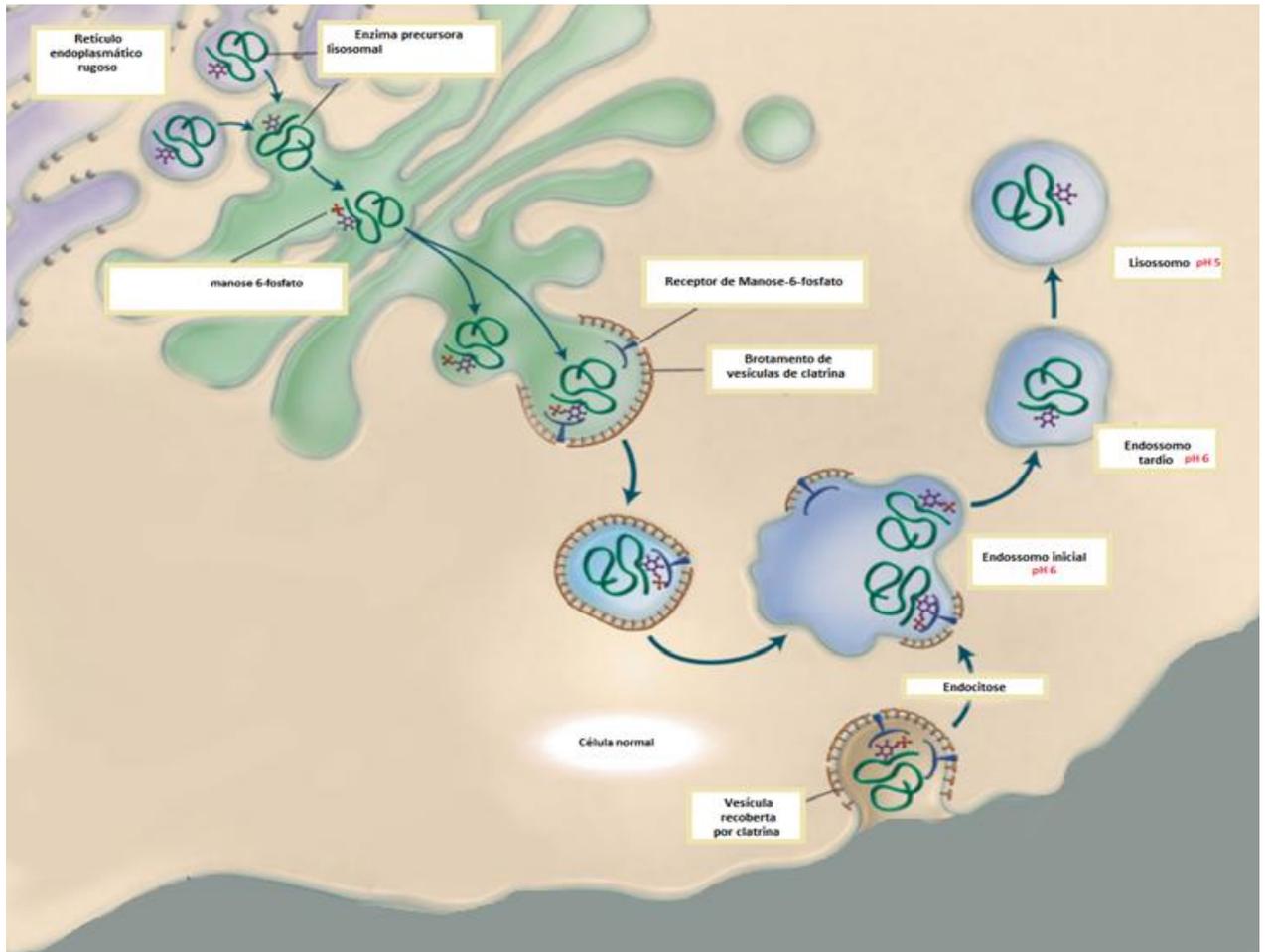


Figura 2: Transporte das enzimas lisossômicas em uma célula normal. O resíduo de M6P é adicionado à enzima lisossômica no Complexo de Golgi que por sua vez se liga aos receptores de M6P. O complexo enzima-receptor é entregue aos endossomos onde o Ph baixo promove a dissociação do mesmo. A enzima migra ao lisossomo maduro e o receptor M6P é reciclado ao Aparato de Golgi. Uma pequena quantidade de M6P é fosforilado e escapa da captura pelos receptores de M6P intracelulares sendo liberada para o exterior da célula. Esta enzima pode ser recapturada pela ligação ao receptor de M6P localizado na superfície da vesícula, recoberta por clatrina. Adaptado de (Muenzer & Fisher, 2004).

1.2.1 Caracterização da enzima

A enzima Iduronidase pertence à família das glicosil-hidrolases que tem como função a catálise e a hidrólise das ligações glicosídicas nos glicosídeos, oligossacarídeos e polissacarídeos. (Harvey *et al.*, 2000). As glicosil-hidrolases constituem-se em 58 famílias, agrupadas por similaridades de sequências entre os aminoácidos. Existem mais de 200 tipos de sequências distintas que agem em diferentes substratos, sendo proposto que evoluíram de um ancestral comum. São exemplos de enzimas que participam desta família a beta-glicosidase, beta-glicocerebrosidase, beta-galactosidase, beta-manosidase e a IDUA (Durand *et al.*,

1997).

1.3 O Impacto de Diferentes Mutações na IDUA em Relação ao Fenótipo da MPSI

Uma ampla variedade de manifestações clínicas de MPS I é causada por mutações distintas no gene da IDUA. Já foram descritas pelo menos 116 mutações diferentes até o momento de acordo com o banco de dados HGMD (*The Human Gene Mutation Database*). As mutações descritas incluem mutações sem sentido e *missense* (77), mutações nos sítios de *Splicing* (10), inserções e deleções (28) bem como rearranjos complexos (1). Um ponto importante a ser ressaltado é que 101 das mutações descritas no banco HGMD são responsáveis pelo fenótipo severo, 6 delas originam a forma moderada, 7 a forma leve e duas não foram caracterizadas.

As mutações sem sentido são responsáveis pelo fenótipo mais grave da doença, já que apresentam um *códon de terminação* prematuro gerando assim uma proteína truncada. Já as demais mutações podem apresentar um fenótipo variado, como acontece nos casos *missense*, onde a correlação entre fenótipo e genótipo é mais complicada. Dentro deste contexto, destacam-se algumas mutações pontuais que são responsáveis pelos fenótipos mais graves de MPS I, como: G51D, A75T, G208D\I, L218P, D349Y\N. Por sua vez as mutações T364M e R619G estão relacionadas com o fenótipo intermediário e as mutações R89Q, R89W e R383H originam a forma atenuada de MPS I (Hein *et al.*, 2003). As diferentes mutações podem interferir tanto na estabilidade da enzima devido a alterações na sua estrutura tridimensional e/ou interferir na sua atividade catalítica, devido a modificações no sítio ativo.

1.4 Diagnóstico

Quando existe suspeita clínica de uma MPS, solicita-se a dosagem dos GAGs na urina, uma vez que estes estão quantitativamente elevados em praticamente todos os casos de MPS. É importante salientar que a ocorrência de níveis normais não descarta o diagnóstico em paciente com quadro clínico sugestivo. A cromatografia ou eletroforese podem ser usadas para identificar qual

o tipo de GAG em excesso (dermatan-sulfato, heparan sulfato, queratan sulfato, por exemplo), auxiliando na escolha de quais enzimas deverão ser inicialmente testadas. O diagnóstico das MPS deve ser confirmado através de ensaio enzimático, documentando a atividade deficiente da enzima específica para cada tipo de MPS. A identificação do genótipo é de extrema importância para predição do fenótipo, uma vez que já foi mapeado o impacto de diferentes mutações na enzima. Além disso, as decisões das melhores terapias, o aconselhamento genético na família e o diagnóstico pré-natal dependem das mutações que o indivíduo carrega.

1.5 Tratamentos Para a Mucopolissacaridose Tipo I

Avanços nas áreas de medicina e genética contribuíram para o desenvolvimento de novos tratamentos para a MPS I, dentre eles destacam-se a terapia de reposição enzimática, o transplante de células tronco hematopoiéticas e, recentemente, a utilização de chaperonas farmacológicas.

1.5.1 Terapia de Reposição Enzimática

A terapia de reposição enzimática (TRE) é um tratamento que consiste na administração periódica, por via venosa, da enzima específica deficiente no paciente. A realização do primeiro tratamento eficaz com TRE em pacientes com a doença de Gaucher levou à busca de tratamentos semelhantes para outras doenças de depósito lisossômico. Subsequentemente este tratamento foi utilizado e aprovado para pacientes de MPS I, MPS VI e MPS II (Dietz, 2010).

A TRE para MPS I é realizada pela administração intravenosa exógena de laronidase (Aldurazyme®), uma proteína análoga à α -L-iduronidase humana produzida por engenharia genética em um sistema de expressão de células de ovário de hamster chinês (CHO). Esta terapia foi aprovada para o tratamento de pacientes nos Estados Unidos e na Europa em 2003 e no Brasil em 2005. Diversos estudos demonstraram a eficácia desta terapia em pacientes MPS I como, por exemplo, na redução de GAGs na urina, melhora dos problemas articulares, na função respiratória e na qualidade de vida (Braunlin *et al.*, 2006; Braunlin *et al.*, 2011; Sifuentes *et al.*, 2007). Trabalhos recentes enfatizam que a

TRE é muito mais eficaz quando o tratamento é iniciado precocemente, ou seja, antes do surgimento dos sintomas (D'Aco *et al.*, 2012; Tyłki-Szymanska *et al.*, 2010).

Um dos aspectos negativos desta terapia relaciona-se com a dificuldade que as enzimas intravenosas possuem de cruzar a barreira hematoencefálica, sendo ineficaz na melhora dos sintomas neurológicos apresentados por alguns pacientes com MPS I (Giugliani *et al.*, 2010a; Giugliani *et al.*, 2010b).

1.5.2 Transplante de Células Tronco Hematopoiéticas

O Transplante de células tronco hematopoiéticas tem sido utilizado como uma das opções de tratamento para pacientes com fenótipos graves de MPS I, principalmente quando envolve danos do sistema nervoso central. Este é o único dos tratamentos que possibilita uma correção metabólica em longo prazo e uma melhoria dos danos cognitivos. Isso é possível devido ao fato de que as células do doador (que produzem a enzima normal) migram para o cérebro e outros órgãos do paciente (Prasad & Kurtzberg, 2010). Embora este procedimento induza uma melhora significativa, existe ainda um grande risco associado ao índice de mortes pós-transplantes, desta forma este tratamento é indicado apenas em situações graves da doença como é o caso de MPS IH (doença de Hurler) quando detectado até os 24 meses de idade.

1.5.3 Terapia utilizando Chaperonas Farmacológicas

Uma nova abordagem terapêutica que vem sendo utilizada são as chaperonas farmacológicas que consistem de moléculas inibitórias que em uma concentração baixa apresentam a propriedade de se ligar preferencialmente no domínio catalítico da enzima mutada. Essa ligação induz uma reorganização conformacional da proteína evitando desta forma que a proteína alterada seja endereçada à via de degradação pelo controle de qualidade do retículo endoplasmático. Desta maneira enzimas com mutações que levam à troca de aminoácido muitas vezes apresentam uma atividade catalítica parcial que será dependente da capacidade da mesma ser encaminhada ao complexo de Golgi e finalmente ao lisossomo. Na tentativa de resolver essa questão, aumentando o

aporte de proteínas ao seu destino final, a indústria farmacêutica e os pesquisadores da área estão utilizando uma ampla variedade de compostos químicos que atuam como chaperonas moleculares em uma diversidade de doenças humanas e tem se mostrado eficazes no reestabelecimento da atividade de várias proteínas mutantes (Muro, 2010; Parenti, 2009).

Mayer avaliou o efeito do tratamento com geneticina e cloranfenicol em fibroblastos de pacientes com MPS I. O tratamento foi realizado em células de três pacientes com mutações sem sentido: um deles homocigoto para mutações sem sentido (p.W402X/p.W402X) e dois heterocigotos (p.Q70X/c.1739-1g<t), (p.R89W/pW402X). A geneticina não teve efeito sobre atividade enzimática dos genótipos avaliados, enquanto o cloranfenicol aumentou mais de 100 vezes a atividade enzimática da IDUA apenas nos heterocigotos (Mayer *et al.*, 2012).

O sequenciamento de cDNA mostrou que apenas o alelo sem mutações de códon de terminação prematuro estava sendo amplificado, sugerindo que os alelos com mutações sem sentido estão sendo degradados por *Nonsense-mediated mRNA Decay* e que por este motivo, o cloranfenicol deve ter outro mecanismo de ação além do SCRT (*stop codon readthrough*). A partir disso, os autores sugeriram que o cloranfenicol pode estar agindo como uma chaperona farmacológica, aumentando a estabilidade de IDUA e permitindo que a enzima chegue ao lisossomo (Mayer *et al.*, 2012).

1.6 Modelagem Molecular de Proteínas

Atualmente há um grande interesse nos estudos de estruturas de proteínas uma vez que detalhes atômicos de resíduos do sítio ativo das proteínas permitem uma maior compreensão do seu funcionamento possibilitando assim o desenvolvimento de medicamentos bem como a compreensão do impacto de mutações na sua estrutura. Experimentalmente a estrutura das proteínas é resolvida através de duas técnicas, a Cristalografia de Raios X e a Ressonância Magnética Nuclear (RMN). A cristalografia é a técnica mais utilizada, porém é uma técnica demorada, com alto custo e depende da obtenção de cristais da molécula de interesse. Neste sentido a modelagem molecular representa uma ferramenta importante para se desvendar a estrutura de proteínas que ainda não foram

cristalografadas (Hillisch *et al.*, 2004; Marti-Renom *et al.*, 2000).

1.6.1 Modelagem molecular de proteínas por homologia e o programa Modeller

O princípio da Modelagem Molecular por Homologia se baseia no fato de que ao longo da evolução as estruturas das proteínas se encontram mais conservadas do que a sua sequência (Branden & Tooze, 1991). A evolução biológica de proteínas obedece algumas regras como: (a) homologia entre sequência de aminoácidos implica em semelhança estrutural e funcional; (b) proteínas homólogas apresentam regiões internas conservadas (principalmente constituídas de elementos de estrutura secundária: α -hélices e folhas- β); (c) as alterações estruturais entre proteínas homólogas ocorrem nas regiões de alças (Branden & Tooze, 1991). Além disso, as proteínas agrupam-se em um número limitado de famílias tridimensionais possibilitando modelar as proteínas de interesse se houver um membro da família que já possui sua estrutura resolvida.

Um modelo construído por modelagem comparativa necessita que, pelo menos, uma estrutura 3D da família em questão tenha sido elucidada por técnicas experimentais. Outro ponto importante é a identidade entre as sequências (alvo e molde), esse valor deve ser acima de 25 % para que o modelo gerado possa ser confiável (Cavasotto & Phatak, 2009; D'Alfonso *et al.*, 2001; Marti-Renom *et al.*, 2000).

A Modelagem Molecular por homologia apresenta quatro etapas principais: a procura por sequências de proteínas homólogas, o alinhamento das sequências, a construção e otimização dos modelos e por fim a avaliação e validação das estruturas geradas (Cavasotto & Phatak, 2009; Marti-Renom *et al.*, 2000).

O MODELLER é um software amplamente utilizado para realizar a modelagem estrutural de proteínas por homologia e encontra-se disponível gratuitamente para instituições acadêmicas sem fins lucrativos (Eswar *et al.*, 2007; Sali & Blundell, 1993). O software realiza a modelagem molecular por homologia de proteínas utilizando como molde uma ou mais proteínas que já possuem a sua estrutura tridimensional resolvida (depositada no PDB). A partir do alinhamento entre a proteína alvo e a estrutura-molde é gerado um conjunto de restrições que são aplicadas a sequência a ser modelada. Estas restrições são realizadas a partir da

macromolécula de referência a partir da qual são assumidos os ângulos e distâncias. A violação das restrições, por incongruências entre o molde e a sequência alvo, é minimizada através de técnicas de otimização (Eswar *et al.*, 2007; Ginalski, 2006; Liu *et al.*, 2011). As restrições podem ocorrer nas distâncias, ângulos, ângulos diedros, pares de ângulos diedros e algumas outras características espaciais definidas por átomos ou pseudo-átomos (Marti-Renom *et al.*, 2000).

1.6.2 Modelagem Molecular utilizando o Servidor I-Tasser

Este software utiliza uma metodologia intermediária entre o método de modelagem por homologia e o *ab initio* chamada de *threading*. O método de *threading* é utilizado para sequências de proteínas que não possuem uma proteína homóloga no banco de dados de estrutura ou que apresenta identidade muito baixa. Esta metodologia opera através de um reconhecimento de padrões de enovelamento (*fold recognition*).

Este programa é composto por uma série de ferramentas capazes de gerar modelos precisos de proteínas através de uma sequência de vários passos: (Zhang, 2008).

- Alinhamento de múltiplos fragmentos (*multiple threading alignment*): é realizada uma busca em um banco de dados de estruturas de proteínas onde serão alinhados pequenos fragmentos com estrutura tridimensional conhecida com a nossa sequência de interesse. Nessa etapa também é considerado a semelhança da estrutura secundária predita para a proteína em estudo e a apresentada pelo fragmento molde naquela região (Roy *et al.*, 2007; Zhang, 2008)
- Agrupamento: Consiste na reunião dos fragmentos gerados no passo anterior bem como na modelagem *ab initio* das regiões não alinhadas. A organização dos fragmentos em uma estrutura tridimensional é feita através de simulações de Monte Carlo, do tipo troca entre réplicas, até que se encontre um modelo otimizado (Roy *et al.*, 2007; Zhang, 2008)
- Refinamento: É a última etapa onde são recuperados os cinco principais conformeros obtidos no passo anterior, para a realização de uma nova rodada de agrupamento por troca de réplicas com a finalidade principal de eliminar impedimentos estéricos. No final desta etapa são recuperados

cinco modelos onde serão adicionadas as cadeias laterais e submetido finalmente a um último passo de refinamento para otimização dos rotâmero (Roy *et al.*, 2007; Zhang, 2008)

1.6.3 Validação dos Modelos

A avaliação do modelo possibilita a identificação de possíveis erros no processo de modelagem constituindo-se em uma das etapas mais importantes do método de Modelagem molecular por Homologia (Cavasotto & Phatak, 2009). Nesse processo são avaliados alguns parâmetros como: a qualidade estereoquímica dos resíduos de aminoácidos, a qualidade e confiabilidade do empacotamento global (enovelamento da proteína) bem como a análise de possíveis erros em regiões específicas da estrutura. A validação fornecerá os níveis de qualidade e confiabilidade da estrutura modelada e será imprescindível para otimização do modelo (Cavasotto & Phatak, 2009). Caso o modelo não seja considerado satisfatório é necessário checar o alinhamento e refazer o processo de modelagem utilizando novas metodologias.

O Verify3D foi desenvolvido para verificar se a estrutura tridimensional do modelo final da proteína está correta (Eisenberg *et al.*, 1997; Luthy *et al.*, 1992). O objetivo, portanto, é fazer a distinção entre modelos “mal traçados” ou enovelados de forma errônea, daqueles modelos corretos, que não foram refinados adequadamente. Este programa avalia a compatibilidade entre a sequência de aminoácidos de uma proteína e o modelo da sua estrutura 3D, o que é realizado reduzindo-se o modelo da estrutura 3D em uma sequência de ambientes químicos (um ambiente para cada resíduo), utilizando-se em seguida a matriz 3D-1D desenvolvida por Bowie e colaboradores. (Bowie *et al.*, 1991). Esta matriz descreve a probabilidade de se encontrar cada um dos 20 aminoácidos em cada um dos ambientes químicos diferentes levando em consideração o tipo do aminoácido e a natureza do ambiente químico calculado no passo anterior. Por fim, é realizado o alinhamento da sequência primária com o seu perfil 3D. A qualidade deste alinhamento está relacionada com a medição da compatibilidade da sequência com a estrutura 3D descrita pelo seu perfil. Desta forma é possível detectar regiões que se encontram mal determinadas.

O PROCHECK foi desenvolvido por Laskowski e colaboradores e se

caracteriza por ser um dos programas mais utilizados para a validação estereoquímica de uma proteína. Através deste programa pode-se avaliar a distância de ligações, a quiralidade, os ângulos planos, a planaridade dos anéis de cadeias laterais bem como as suas conformações. O programa também analisa os ângulos de torção das cadeias principais (*phi* e *psi*) e laterais da molécula informando os maus contatos e a planaridade das ligações peptídicas (Laskowski *et al.*, 1993). Um dos arquivos de saída mais conhecido gerados pelo programa é o gráfico de Ramachandran que apresenta uma correlação entre os ângulos torcionais da cadeia principal (*phi* e *psi*). A análise da rotação desses ângulos levou à identificação de regiões permitidas e regiões não permitidas, onde ocorrem colisões entre os átomos (Laskowski *et al.*, 1993).

Outro programa que vem ganhando atenção é o MoldFOLD, este software realiza a predição global e local da qualidade da estrutura 3D dos modelos. Ele fornece um único score e um valor de P relacionado a qualidade do modelo avaliado, gerando predições de erro por resíduo.(McGuffin, 2008; McGuffin *et al.*, 2013)

1.7 Modelagem Molecular da IDUA

Rempel e colaboradores publicaram o primeiro modelo da enzima IDUA depositado no PDB com o código 1Y24 (Rempel *et al.*, 2005). Este modelo foi construído utilizando como molde a beta-D-xylosidase de *T. saccharolyticum* (1UHV, 1PX8) e o software online *Swiss model*. Esse programa realiza a modelagem por homologia de forma automatizada. Recentemente um segundo modelo desta enzima foi publicado por Chandar & Mahalingman utilizando a ferramenta automatizada *Schrodinger PRIME* (Chandar & Mahalingam, 2012).

1.8 Justificativa do Trabalho

O conhecimento da estrutura de uma proteína é de vital importância tanto para a predição de fármacos quanto para o entendimento do impacto de mutações na sua funcionalidade. Quando começamos este trabalho já havia dois modelos da IDUA publicados e a divulgação da estrutura obtida por cristalografia

de raios X ocorreu somente em abril de 2013.

Um dos objetivos iniciais deste trabalho era construir um modelo mais acurado da IDUA, e comparar com os já publicados. É importante ressaltar que os dois modelos publicados desta enzima não haviam sido avaliados através de dinâmica molecular. Esta avaliação é de suma importância para checar a estabilidade da proteína em solução aquosa bem como verificar a confiabilidade do modelo gerado.

Entretanto em abril de 2013, quando estávamos terminando a dissertação, a estrutura 3D da IDUA foi depositada no Protein Data Bank (PDB). A modelagem desta enzima era de um grau de dificuldade muito alto já que os moldes utilizados apresentavam uma identidade muito baixa. Desta forma, os modelos gerados da IDUA humana utilizando diferentes abordagens, quando ainda não havia o cristal, se tornaram um material precioso para ser utilizado na área de modelagem com *templates* de baixa identidade.

Além disso, pretendemos avaliar através de dinâmica molecular o efeito de duas mutações (R89W e R89Q) que alteram um resíduo chave para o mecanismo catalítico da enzima. De acordo com a literatura essas alterações possuem um impacto diferenciado uma vez que a mutação R89W altera apenas a atividade da enzima enquanto que a mutação R89Q tem um impacto tanto na sua atividade enzimática quanto na sua estabilidade estrutural. Embora ambas as alterações sejam responsáveis pelo fenótipo atenuado da doença, pacientes MPS I que apresentam a mutação R89W possuem um fenótipo clínico menos severo do que os portadores da mutação R89Q. Essa compreensão sobre o efeito das mutações na proteína visa contribuir para a escolha terapêutica a ser abordada. A escolha do tratamento com chaperonas farmacológicas é indicada em pacientes que apresentam mutações que induzam uma desestabilização da proteína (as quais irão compensar a estabilidade da proteína permitindo que esta seja entregue no lisossomo). Por outro lado, esta terapia não é indicada para pacientes que possuam mutações que afetam a atividade enzimática.

2 Objetivos

2.1 Objetivo Geral

O objetivo do presente trabalho é realizar a modelagem molecular da enzima α -L-Iduronidase humana utilizando diferentes metodologias. A partir dessas modelagens pretendemos realizar comparações com o cristal dessa proteína (4JXP - *publicado após a obtenção dos modelos*) para verificar qual método é o mais apropriado em modelagens comparativas com moldes de baixa identidade. O método que apresente o melhor resultado em relação ao cristal poderá ser aplicado na modelagem das outras proteínas (enzimas) da rota de degradação de glicosaminoglicanos. Adicionalmente, a partir do cristal da enzima IDUA pretendemos avaliar o impacto estrutural de 2 mutações descritas da mucopolissacaridose tipo I através de uma abordagem de dinâmica molecular.

2.2 Objetivos Específicos

- i. Construir 3 modelos moleculares da enzima iduronidase humana utilizando os métodos a seguir:
 - Modelagem automatizada utilizando o servidor I-TASSER,
 - Alinhamento por *fold recognition* (indicado para modelagens onde os moldes possuem baixa identidade com a sequência alvo) gerado pelo programa phyre2 e construção dos modelos através do programa MODELLER 9.9,
 - Alinhamento utilizado por Rempel e colaboradores (2005) e obtenção dos modelos através do programa MODELLER 9.8,
- ii. Avaliar o melhor modelo obtido a partir de cada uma das metodologias através de diferentes métodos de validação de modelos (Verify 3D, ModFOLD e Procheck) e por dinâmica molecular, em um período de 50 ns com o software GROMACS,
- iii. Comparar através dos procedimentos anteriormente citados o modelo já existente (1Y24) com os 3 modelos gerados a partir das diferentes metodologias especificadas acima.

- iv. Inferir qual o melhor método de modelagem a partir de moldes de baixa identidade com a sequência alvo para aplicar na construção de modelos das outras enzimas da via de degradação.
- v. Utilizar o cristal da IDUA humana como molde para gerar um modelo completo da enzima α -L-iduronidase selvagem e dos mutantes R89W e R89Q. A partir de simulações de dinâmica molecular de 50ns inferir o impacto das mutações R89W e R89Q na estrutura da IDUA e sua possível relação com o fenótipo da MPS I.

3 Metodologia

3.1 Predição da estrutura tridimensional da proteína α -L-Iduronidase

3.1.1 Modelo IDUA I-tasser

A sequência fasta da Iduronidase humana (UNIPROT: P35475) foi recuperada e os 26 primeiros aminoácidos que fazem parte do peptídeo sinal foram retirados. O Primeiro modelo da IDUA gerado foi obtido utilizando-se o software online I-tasser (Zhang, 2008). Após todos os passos de recuperação de segmentos moldes e da geração *ab initio* das regiões faltantes, como descrito na introdução, foram gerados 5 modelos ordenados pelos valores de C-score e TM-Score. O C-score é uma medida que estima a qualidade do modelo predito, variando de -5 a 2, de modo que quanto maior o valor de C-score maior o índice de confiança do modelo gerado. A acurácia do modelo é obtida através do TM-score, no qual valores entre 0,0 e 0,17 significam que a similaridade estrutural é ao acaso, enquanto valores entre 0,5 e 1,0 são indicativos de uma topologia correta do modelo em relação a proteína nativa.

Os modelos gerados pelo I-Tasser apresentaram uma região C-terminal que não foi modelada, desta forma os modelos gerados corresponderam à região entre os resíduos 27 e 635 da IDUA selvagem. O modelo melhor ranqueado pela modelagem com o I-tasser apresentou um C-score de -1,58 e um valor de TM-Score de $0,52 \pm 0,15$.

3.1.2 Modelo IDUA - Phyre 2.0/Modeller (IDUA - PM)

Um segundo modelo gerado a partir da sequência da IDUA madura (sem o peptídeo sinal) foi obtido utilizando-se como molde a estrutura da Beta-D-xylosidase de *T. saccharolyticum* (PDB: 1UHV). Esta estrutura foi obtida através da busca de sequências semelhantes através dos programas psiBLAST e BLASTp do NCBI, bem como do HHPRED, sendo este último disponível no site do Instituto Max Planck (<http://toolkit.tuebingen.mpg.de/hhpred>). O alinhamento entre a IDUA e o molde 1UHV foi gerado utilizando o programa Phyre 2 (Kelley & Sternberg, 2009). A região alinhada, compreendida entre os aminoácidos 27 e 542 da IDUA, apresentou 22% de identidade com o cristal 1UHV. Foram gerados 100 modelos utilizando o software Modeller 9v9 e o melhor modelo foi selecionado

avaliando-se os valores de DOPE Score e do Procheck.

3.1.3 Modelo IDUA - Rempel e cols /Modeller (IDUA - RM)

O terceiro modelo da IDUA foi gerado utilizando-se o software Modeller 9v9 (Marti-Renom *et al.*, 2000) e o molde descrito acima (PDB: 1UHV), mas baseando-se no alinhamento utilizado por Rempel e cols (2005) na confecção do modelo de IDUA previamente publicado (1Y24). Foram gerados 100 modelos da região entre os resíduos 36 e 504 da IDUA. O melhor modelo foi escolhido utilizando-se os valores de DOPE Score e do Procheck.

3.2 Validação dos modelos

Os modelos foram validados através de diferentes ferramentas como o Procheck, Verify 3D e ModFOLD versão 3 .

3.3 Simulação de Dinâmica Molecular

As simulações de Dinâmica Molecular foram realizadas com o pacote GROMACS v4.5.1 (Van Der Spoel *et al.*, 2005), na plataforma Linux (sistema operacional Ubuntu 12.04), de acordo com protocolo já estabelecido em nosso laboratório. O protocolo está dividido em uma série de passos, que podem ser agrupados em seis etapas principais: (1) preparação do sistema, (2) minimização de energia, (3) solvatação da proteína, (4) termalização, (5) produção e (6) análise dos dados, descritas a seguir:

1. Preparação do Sistema: A proteína estudada (modelos de IDUA) foi simulada em uma caixa cúbica com uma distância de 15 Å entre o soluto e a caixa, utilizando-se o campo de força GROMOS 53a6 e o modelo de água SPC. O sistema foi neutralizado com íons Sódio (Na⁺) e Cloro (Cl⁻) até a concentração final de 0,15 mol/L.
2. Minimização de Energia (EM): Foi composta por três etapas. Primeiramente o algoritmo *steepest-decent* foi aplicado para minimizar a energia do solvente (moléculas de água), enquanto a proteína era mantida rígida através da aplicação de um potencial harmônico (100 kJ⁻¹mol⁻¹nm⁻¹). Esta etapa foi seguida por uma nova aplicação do *steepest-decent*, agora sem restrições. Finalmente, o sistema foi submetido a uma etapa de

minimização de energia utilizando o algoritmo de gradiente conjugado (CG).

3. Solvatação do Complexo: Após as etapas de EM o sistema foi submetido a um procedimento que permite a formação das camadas de solvatação ao redor da molécula de IDUA. O sistema foi mantido aquecido (100 K) durante 50 ps induzindo agitação nas moléculas de água, enquanto a posição da molécula de IDUA foi mantida estável através da aplicação de um potencial harmônico com intensidade de $5.000 \text{ kJ}^{-1}\text{mol}^{-1}\text{nm}^{-1}$. A temperatura foi então reduzida a 20 K, o que foi seguido pela redução gradual da força de restrição sobre os átomos da IDUA, até o valor mínimo de $0,2 \text{ kJ}^{-1}\text{mol}^{-1}\text{nm}^{-1}$.
4. Termalização: Após remoção completa das restrições, o sistema passa a ser gradualmente aquecido até atingir a temperatura de 300 K. Ao todo são 7 passos de termalização, sendo que a diferença de temperatura entre as etapas foi sempre inferior a 50K.
5. Produção: Durante um período de 47,5 ns, a temperatura foi mantida estável em 300 K e o sistema foi mantido completamente flexível. Isso permite que a proteína se movimente no solvente de forma similar ao que ocorreria *in vivo*, bem como a ocorrência de alterações conformacionais, formação e rompimento de ligações não covalentes, etc. Para fins de análise, foram aproveitadas apenas as medidas realizadas durante o período de produção.
6. Análise dos dados: O pacote Gromacs possui vários programas que permitem realizar uma ampla série de análises sobre os dados gerados na etapa de produção. Dentre as análises que foram realizadas destacam-se o desvio médio da estrutura da IDUA ao longo da simulação (*g_rms*), o raio de giro da molécula (*g_gyrate*), a flutuação média de cada aminoácido ao longo do tempo (*g_rmsf*), entre outras.

3.3.1 Simulações envolvendo análise dos modelos gerados por homologia a partir do molde 1UHV e pela técnica de "threading"

Utilizando o protocolo descrito acima, uma simulação de 50ns foi realizada para os 3 modelos gerados pelas diferentes técnicas de modelagem (I-tasser, Modeller/Idua-PM e Modeller/Idua-RM), e para um modelo gerado com o

Modeller a partir do cristal de Idua recentemente publicado (4JXP), o qual não continha dois trechos resolvidos (55-61 e 103-106). Este último serviu como o padrão para inferência de estabilidade e de comportamento dos resíduos ao longo das simulações.

3.3.2 Simulações de modelos gerados a partir do cristal 4JXP para avaliação do impacto das mutações R89W e R89Q na estrutura da Idua

Dois modelos adicionais foram gerados a partir do cristal 4JXP para a geração de enzimas contendo as mutações R89W e R89Q, utilizando o software Modeller 9v9. Foi realizada uma dinâmica molecular de 50ns para cada um deles como descrito no item "3.4 Simulação de Dinâmica Molecular". Medidas comparativas com o cristal relativas a estabilidade e comportamento geral da proteína foram avaliadas, além de medidas relativas a resíduos importantes para a atividade normal da enzima.

4 Resultados e Discussão

4.1 Avaliação dos modelos e implicações para futuras modelagens das outras enzimas da via de degradação de GAGs

Em abril deste ano foi depositado no Protein Data Bank os cristais da enzima α -L-iduronidase humana sob os códigos 4JXP e 4JXO. Dentro deste contexto, os modelos construídos durante nosso projeto tornaram-se um material de extrema valia para a área da modelagem molecular. Neste período não havia ainda o cristal e o melhor molde possuía em torno de 20% de identidade (dependendo do alinhamento utilizado). A modelagem da IDUA era um grande desafio uma vez que apresentava um molde com identidade muito baixa e o conhecimento de sua estrutura seria de grande importância em estudos envolvendo a busca por novos fármacos, por exemplo. Considerando o exposto e de posse dos dados do cristal da IDUA, foi possível comparar os modelos gerados a partir de diferentes metodologias com a estrutura cristalográfica e escolher a melhor abordagem para a modelagem de enzimas que apresentam moldes com identidade muito baixa, como é o caso das outras enzimas da via de degradação dos glicosaminoglicanos envolvidas nas MPS II e III.

Uma primeira análise das informações das posições corretas dos ângulos dos diedros modelados e que fornece uma medida de qualidade dos mesmos é a análise dos ângulos *phi* e *psi*. Na (**Tabela 2**) são apresentados os resultados do mapa de Ramachandran para cada um dos modelos gerados, utilizando-se diferentes metodologias, e para o modelo previamente publicado 1Y24. Através desta análise os melhores modelos foram o IDUA-RM e o IDUA-PM apresentando respectivamente 87,3% e 85,0% dos aminoácidos em regiões mais favorecidas (valores similares ao do cristal) e 0,7% e 1,1% em regiões não permitidas. Esses dois modelos apresentaram valores melhores do que o modelo 1Y24 publicado por Rempel e colaboradores (2005) o qual apresentou 72,5% dos aminoácidos em regiões permitidas e 1,7% em regiões não permitidas. Em relação ao mapa de Ramachandran o pior modelo foi o gerado pelo programa I-tasser.

Tabela 2: Avaliação dos modelos de IDUA através do Gráfico de Ramachandran.

Regiões	Cristal 4JXP	Modelo 1Y24	IDUA- RM	IDUA- PM	IDUA I-tasser
Mais favorecidas	86,5%	72,5 %	87,3%	85,0%	76,7%
Adicionalmente permitidas	12,9%	23,4%	11,0%	12,3%	16,5%
Generosamente permitidas	0,6%	2,4%	1,0%	1,6%	4,1%
Não permitida	0,0%	1,7%	0,7%	1,1%	2,7%

Tabela 3: Avaliação dos modelos de IDUA nos programas ModFOLD e Verify 3D.

Modelos	ModFOLD			Verify 3D	
	Qualidade global	Valor de P	Confiança	% resíduos 3D-1D acima de 0,2	Resultado
1Y24	0,2081	2,663E-2	Média	56,29%	Reprovado
IDUA-RM	0,2160	2,288E-2	Média	56,08%	Reprovado
IDUA-PM	0,3734	1,898E-3	Alta	86,05%	Aprovado
IDUA-Itasser	0,2961	5,747E-3	Alta	80,79%	Aprovado

Além das análises das conformações permitidas da cadeia principal através do mapa de Ramachandran os modelos foram avaliados através dos programas ModFOLD e Verify3D (**Tabela 3**). A confiabilidade do enovelamento foi feita através do Verify3D que avaliou o ambiente químico para cada um dos resíduos traçando a compatibilidade da estrutura 3D modelada com a sua sequência. Os modelos 1Y24 e IDUA-RM apresentaram em torno de 56% dos aminoácidos com uma média 3D-1D acima de 0,2 e não foram aprovados no quesito enovelamento. Para considerar a qualidade do modelo satisfatória é necessário que pelo menos 80% dos resíduos possua um “*Score 3D-1D*” acima de 0,2. Além disso, esses dois modelos quando avaliados através do ModFOLD apresentaram um score de qualidade global menor do que o IDUA-PM e IDUA-Itasser e uma probabilidade de acerto **média** (no caso, a probabilidade do modelo estar incorreto é menor do que 1/20). Por outro lado os modelos IDUA-PM e IDUA-Itasser foram aprovados pelo verify3D e pelo ModFOLD apresentando um valor de confiança **alto**, ou seja, a probabilidade do modelo estar incorreto é

menor do que 1/100. Além disso, o modelo gerado com o alinhamento do Phyre (IDUA-PM) apresentou o maior valor de qualidade global 0,3734 um valor muito próximo de 0,4. É importante salientar que valores acima de 0,4 são um indicativo de confiabilidade da estrutura modelada, a qual pode ser considerada altamente similar com a estrutura nativa. Considerando estes três métodos de validação estrutural a IDUA-PM foi considerada como melhor modelo de Idua gerado neste trabalho.

Uma vez que a estrutura cristalográfica foi recentemente publicada, pode-se avaliar qual o modelo que mais se aproximou da estrutura nativa através da sobreposição dos modelos de IDUA com o cristal 4JXP (**Tabela 4 e Figura 3**). Na tabela 4 são apresentados os valores do desvio médio quadrático (RMSD) entre os modelos e o cristal, considerando tanto o cálculo em relação ao domínio "barril TIM" (domínio catalítico) quanto a proteína inteira. Nesta comparação o modelo que mais se aproximou da estrutura nativa também foi o construído utilizando o alinhamento por *fold recognition* através do programa Phyre (IDUA-PM), apresentando um desvio de 5,78Å considerando o domínio catalítico e 6,17Å para a proteína total. Embora esse valor de RMSD ainda seja alto, ele é duas vezes menor do que o obtido (domínio catalítico) pelo modelo publicado por Rempel e colaboradores (2005), sendo que para a proteína total o modelo 1Y24 apresentou um erro ainda maior (16,65 Å). O segundo melhor modelo foi o construído pelo programa I-tasser, apresentando um desvio de RMSD um pouco maior (6,35 Å) do que o da IDUA-PM.

Os modelos que foram construídos através da modelagem por homologia (IDUA-PM e IDUA-RM) utilizaram como molde a estrutura de beta-D-xylosidase de *T. saccharolyticum* (1UHV). Na **Figura 4** podemos visualizar a comparação estrutural entre a beta-D-xylosidase e a IDUA. Avaliando atentamente as duas estruturas observamos algumas diferenças que contribuíram para a geração de modelos com um RMSD alto em relação ao cristal. Embora exista uma grande semelhança no domínio catalítico formado pelo Barril ($\alpha\beta$)₈ é importante destacar a presença de dois grampos- β na IDUA que se expandem do barril, um dos quais não estava presente no molde e não foi corretamente formado em nenhuma das estruturas geradas (**Figura 3**). Observamos também a ausência do domínio da

fibronectina (não modelado) na estrutura de beta-D-xylosidase bem como pequenas alterações no domínio β -sanduíche (**Figura 4**). A IDUA é composta por 3 domínios, o domínio barril TIM, o β -sanduíche e um domínio C-terminal que se assemelha ao da fibronectina (**Figura 4**).

Dois pontos são cruciais na modelagem por homologia, o primeiro deles é o alinhamento e o segundo é a escolha do melhor modelo (Cavasotto & Phatak, 2009; Sali & Blundell, 1993). Quando a identidade da sequência entre o molde e o alvo é maior do que 90%, a precisão do modelo obtido a partir deste molde pode ser comparada com estruturas cristalográficas. Já quando a identidade varia de 50% a 90% o modelo pode apresentar erros locais de até 1,5 Å. No caso de proteínas onde a identidade é igual ou menor do que 25%, o alinhamento se constitui no principal gargalo para a modelagem por homologia, o qual pode induzir a erros cruciais nos modelos (Chothia & Lesk, 1986; Sippl, 1993). No caso da IDUA que apresentava um molde com identidade abaixo de 25%, o alinhamento foi determinante para que obtivéssemos modelos mais precisos do que o previamente publicado, como podemos visualizar na (**Figura 5**). O modelo gerado com o alinhamento do Phyre2 (IDUA-PM) foi o que mais se aproximou da estrutura nativa (em relação à análise do RMSD). Além disso, verificamos que o modelo já publicado da IDUA 1Y24 e o modelo IDUA-RM (o qual utilizou o mesmo alinhamento que Rempel e cols) apresentaram erros mais dramáticos, reiterando a importância de um alinhamento correto no processo de modelagem.

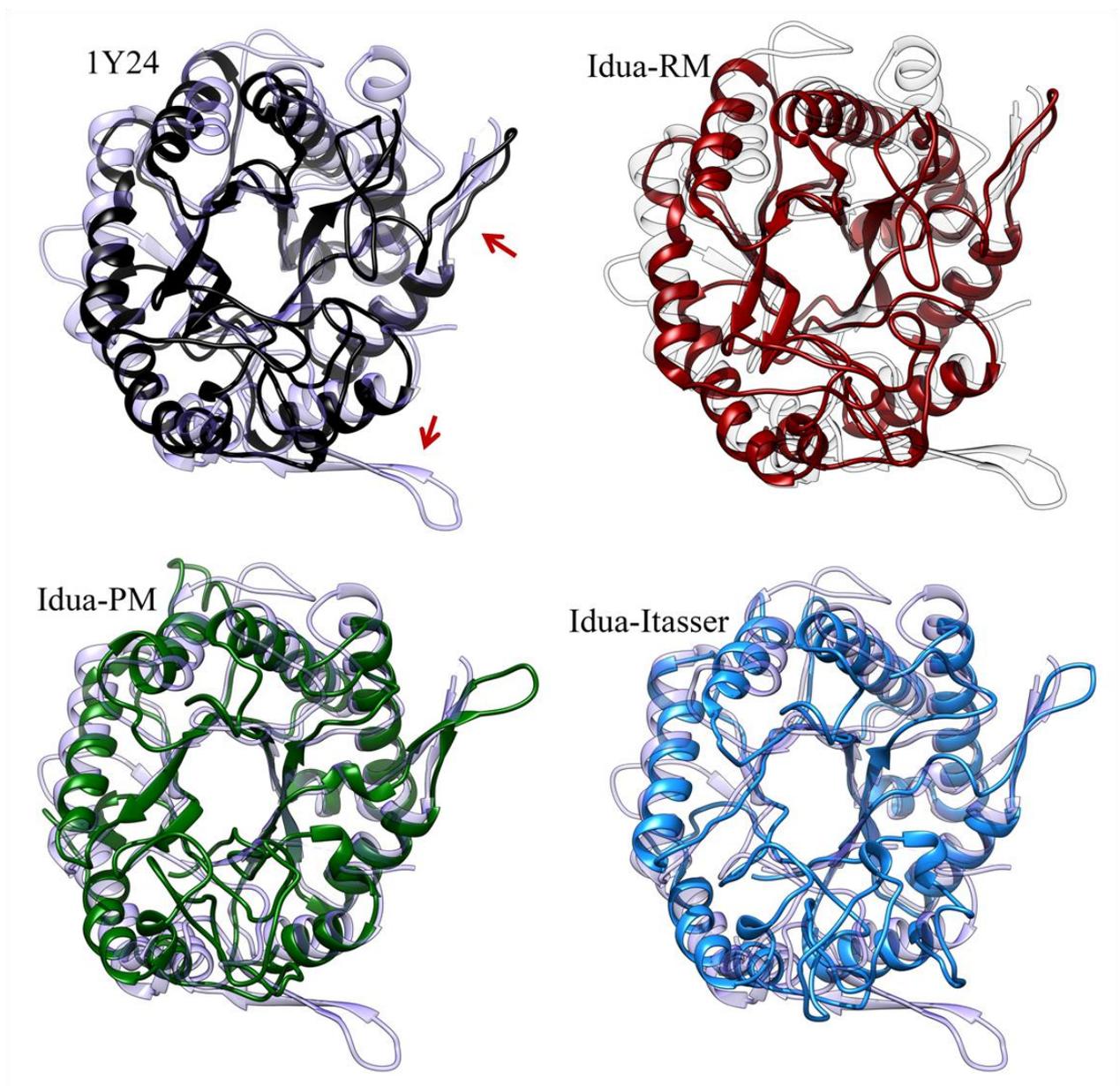


Figura 3: Domínio catalítico da α -L-iduronidase. Sobreposição dos modelos com a estrutura cristalográfica da IDUA (4JXP). O cristal está representado em cinza claro. A seta em vermelho destaca os dois grampos- β que protraem do barril.

Tabela 4: Comparação entre os modelos de IDUA e a estrutura cristalográfica 4JXP considerando o domínio catalítico e a proteína total.

	*RMSD Domínio Catalítico	*RMSD (proteína total)
Modelo 1Y24 x Cristal	12,24	16,65
IDUA- RM x Cristal	11,84	16,7
IDUA- PM x Cristal	5,78	6,17
IDUA I-tasser x Cristal	6,35	9,12

*RMSD realizado pelo Backbone.

Quando avaliamos o domínio catalítico, responsável pela clivagem do ácido hialurônico, verificamos que foram utilizados trechos menores da sequência de aminoácidos para formação deste domínio para os modelos 1Y24 e IDUA-RM (regiões não compreendidas pelo quadro vermelho na **Figura 5**). Já os modelos IDUA-PM e IDUA-I-tasser apresentaram um começo e um término do barril TIM na mesma região de aminoácidos do cristal (regiões externas ao quadro azul). Outro ponto a ser salientado na **Figura 5** se refere ao alinhamento das estruturas secundárias. No modelo IDUA-PM e IDUA-I-tasser verificamos uma boa concordância no alinhamento das estruturas secundárias, com poucas exceções, como é o caso da região entre as hélices $\alpha 7$ e $\alpha 8$. Já no modelo 1Y24 e na IDUA-RM verificamos uma maior discordância na sobreposição das estruturas secundárias, a qual pode ser evidenciada na região compreendida entre as hélices $\alpha 5$ e $\alpha 8$.

Outro ponto que deve ser salientado é o número de hélices e folhas betas presentes no domínio catalítico dos modelos em relação ao da estrutura nativa (8 alfa-hélices e 8 folhas-beta no barril TIM). Podemos verificar que no modelo IDUA-PM não foram formadas as folhas $\beta 1$ e $\beta 8$ entretanto as 8 hélices estão presentes. Além disso, um dos dois grampos- β (representado pela letra C) salientes ao domínio Barril TIM foi corretamente formado. No modelo I-tasser as folhas $\beta 1$, $\beta 2$ e $\beta 8$ estavam ausentes e apresentou a formação parcial de apenas um dos dois grampos- β . Nos modelos 1Y24 e IDUA-RM não foram formadas a folha $\beta 1$ e os dois grampos- β . Já as folhas $\beta 6$ e $\beta 8$ foram formadas porém a sua localização não estava correta. Em relação às hélices do domínio catalítico desses dois modelos houve problemas tanto em relação ao tamanho quanto a

localização das hélices $\alpha 3$, $\alpha 6$ e $\alpha 8$.

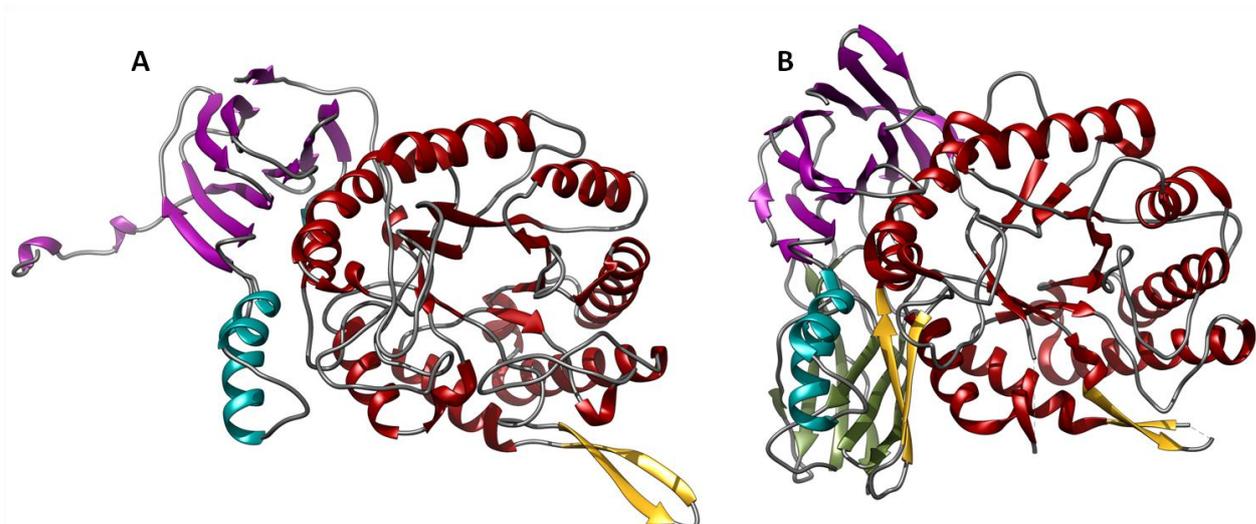
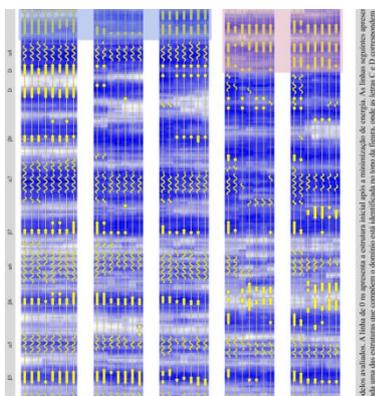


Figura 4: Comparação entre as estruturas cristalográficas de representantes da família 39 de Glicosil hidrolases. (A) beta-D-xylosidase de *T. saccharolyticum* (1UHV) e (B) α -L-iduronidase humana (4JXP). Em vermelho o domínio Barril ($\alpha\beta$)₈, em lilás o domínio β -sanduíche e em verde o domínio que se assemelha ao da fibronectina. Em amarelo são destacados os grampos- β que protraem do Barril.

Outra forma de avaliarmos os modelos gerados é através da dinâmica molecular, a qual utiliza os princípios de mecânica clássica para observar o comportamento dos modelos em solução aquosa e verificar alguns parâmetros como estabilidade e alteração de estrutura secundária no decorrer do tempo, por exemplo. A estrutura da IDUA depositada no PDB apresentava duas regiões (55-61 e 103-106) que não estavam resolvidas. Com intuito de obter uma estrutura completa para a realização da dinâmica molecular, realizamos a modelagem por homologia da IDUA utilizando como molde o próprio cristal da IDUA humana (4JXP), o qual foi chamado de Modelo-4JXP. Este modelo, além do modelo 1Y24 publicado por Rempel e cols (2005) e dos modelos construídos neste trabalho (IDUA-RM, IDUA-PM e IDUA-Itasser) quando ainda não havia o cristal da IDUA, foram submetidos à simulação de dinâmica molecular por um período de 50 ns.

Figura 5: Avaliação ao longo de 50ns de simulação da estrutura secundária do cristal (4JXP) e dos modelos de IDUA gerados.



Adaptado de [1]. A título de comparação a estrutura secundária do cristal (4JXP) e dos modelos de IDUA gerados ao longo de 50ns de simulação. Os dados foram gerados a partir de simulações de dinâmica molecular realizadas no software GROMACS 5.0.7.

Na **Figura 6** são apresentados os resultados do desvio médio quadrático (RMSD) do domínio catalítico ao longo de 50 ns de simulação. Essa medida permite verificar o quanto a proteína se deslocou ao longo do tempo em relação às suas coordenadas iniciais. O modelo-4JXP atingiu a estabilidade por volta dos 16 ns em 0,4 nm, esse comportamento é o esperado para uma proteína de 600 aminoácidos. Três simulações, dos modelos 1Y24, IDUA-Itasser e IDUA-RM apresentaram comportamentos semelhantes estabilizando em torno de 25 ns entre 0,55 nm – 0,60 nm. Já o modelo IDUA-PM ficou mais instável ao longo da simulação tendendo a uma estabilidade em 35ns com um RMSD em torno de 0,7 nm.

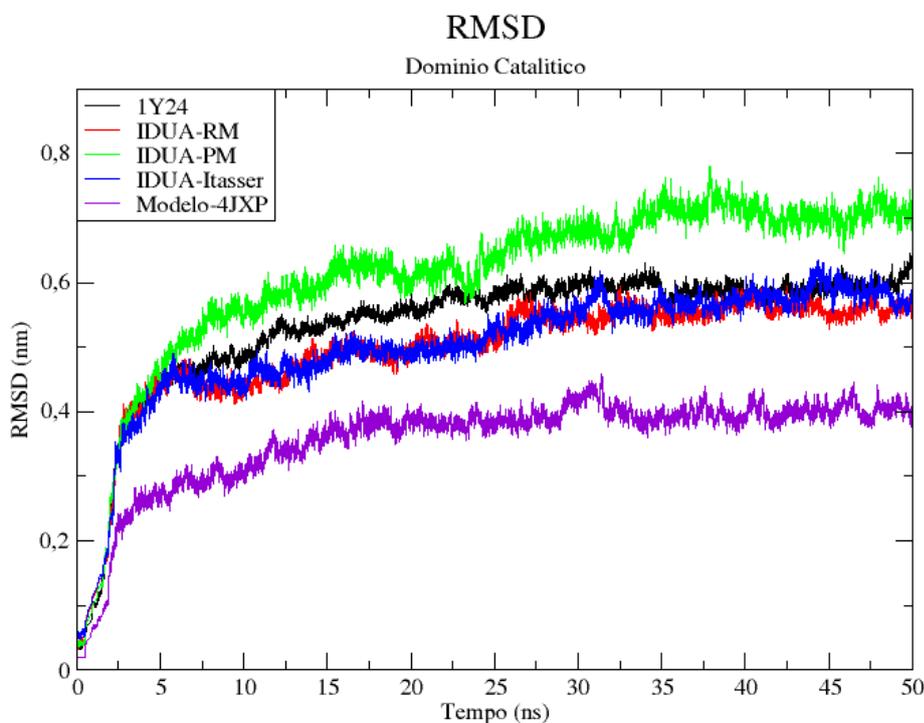


Figura 6: Comparação entre o desvio médio quadrático (RMSD) dos modelos de IDUA gerados através de diferentes metodologias e o Modelo-4JXP, ao longo de 50 ns.

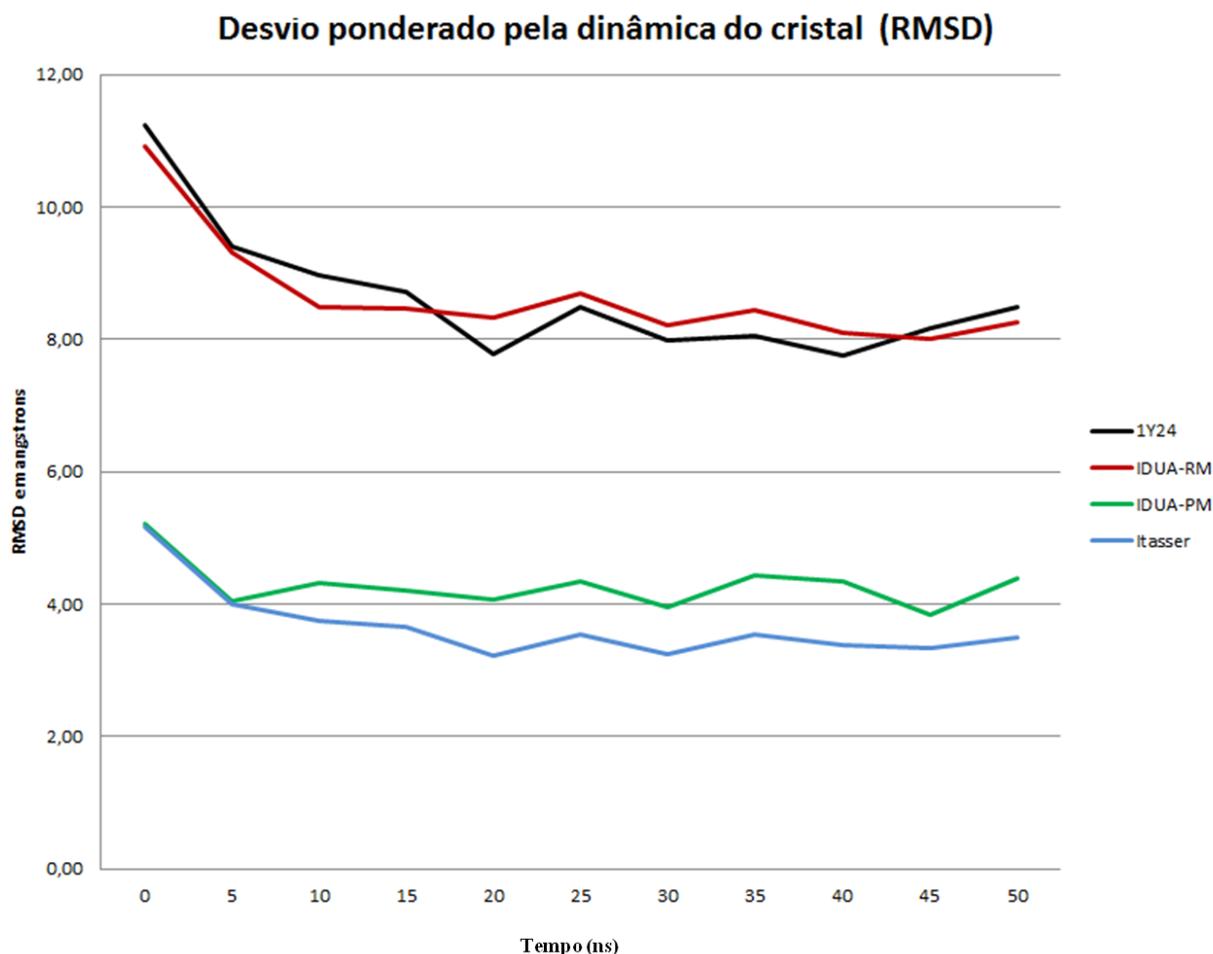


Figura 7: Desvio médio quadrático (RMSD) do domínio catalítico da IDUA (excluindo as regiões de alça) ponderado pela dinâmica do modelo-4JXP.

Outra abordagem realizada foi a comparação do RMSD dos modelos com o modelo-4JXP considerando somente as hélices e folhas betas do domínio catalítico (**Figura 7**). Para isso o RMSD dos modelos foi calculado a cada 5ns em relação a estrutura do modelo-4JXP (minimizado) durante a simulação. Na **Figura 7** verificamos que ao longo da simulação todos os modelos apresentaram uma diminuição do RMSD em relação ao modelo-4JXP alcançando um valor mínimo a partir dos 20ns. Nesta análise os dois modelos que mais se aproximaram do cristal foram IDUA I-tasser e IDUA-PM com um desvio médio ao longo da dinâmica de 3,66 Å e 4,28 Å, respectivamente.

É importante observar que as figuras 6 e 7 se referem a medidas

completamente distintas, o que explica a grande diferença nos valores de RMSD apresentados nas duas figuras. O primeiro gráfico (**Figura 6**) se refere ao desvio do domínio catalítico (incluindo alças) de cada um dos modelos com relação a sua própria conformação no tempo zero da simulação. Este gráfico foi obtido de forma contínua, incluído todas as conformações observadas ao longo da simulação. O valor do desvio em cada ponto deste gráfico é uma média do desvio de todos os resíduos da região analisada, considerando-se apenas o alinhamento estrutural. O objetivo neste caso era observar de maneira preliminar o quanto a estrutura inicial de cada modelo estava se mantendo estável ao longo da simulação.

Por outro lado, a **Figura 7** se refere ao desvio de cada um dos modelos com relação a uma mesma estrutura de referência (Modelo-4JXP). Além disso, nesta segunda análise foram desconsideradas as regiões de alça, por serem regiões desordenadas que normalmente apresentam grande variabilidade em solução. A análise foi realizada com conformações obtidas a cada 5 ns de simulação e o desvio em cada um destes pontos é uma média do desvio de todos os resíduos das estruturas secundárias da região analisada, com relação as respectivas regiões na estrutura referência. Observe que os pontos de início e término de cada estrutura secundária do domínio catalítico foram determinados com base na sequência da estrutura referência, ou seja, neste caso não bastava apenas ter uma estrutura secundária "equivalente", era preciso ter uma estrutura secundária equivalente e formada pelos resíduos corretos (em relação ao Modelo-4JXP). Por último, foi descontado dos valores em cada ponto o desvio observado pelo próprio Modelo-4JXP nos tempos equivalentes de simulação. O objetivo neste caso era observar de maneira mais refinada a precisão de cada modelo na determinação das estruturas secundárias do domínio catalítico, tanto em relação ao número de estruturas secundárias do barril TIM, quanto em relação a correta localização de cada estrutura dentro da sequência da proteína. Além disso, ponderando pelo desvio observado para o Modelo-4JXP, foi possível observar o grau de convergência entre os modelos e a estrutura referência, ao longo do período de simulação.

Assim sendo, apesar de a **Figura 6** indicar uma aparente maior

estabilidade dos modelos 1Y24 e IDUA-RM com relação ao modelo IDUA-PM, a **Figura 7** nos permite observar que esta estabilidade reflete uma configuração alternativa (incorreta) das estruturas secundárias do domínio catalítico, de modo que os modelos 1Y24 e IDUA-RM apresentam um desvio muito maior que o modelo IDUA-PM quando comparadas com a estrutura referência. Apesar de apresentar uma modelagem mais precisa das estruturas do domínio catalítico, a maior instabilidade observada para o modelo IDUA-PM na **Figura 6** pode ser resultado da ausência dos outros domínios da enzima. O modelo IDUA-Itasser apresentou bons resultados em ambas as análises, sendo que na análise apresentada na **Figura 7** apresentou comportamento ainda melhor que o de IDUA-PM.

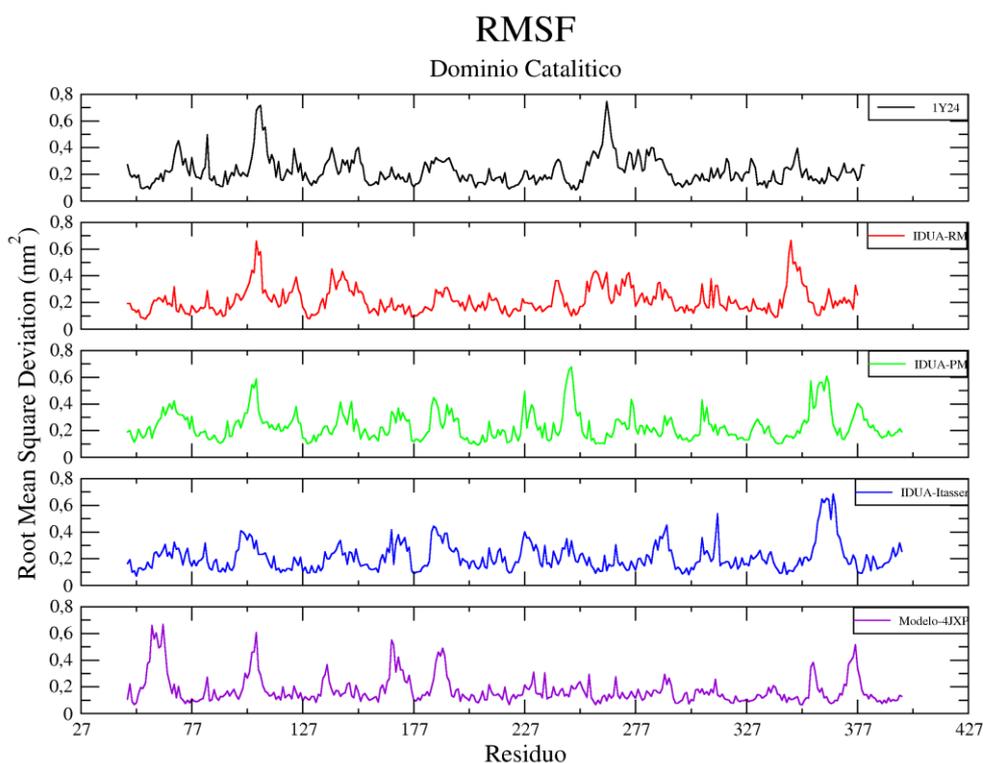


Figura 8: RMSF do domínio catalítico dos modelos de IDUA ao longo de 50 ns de simulação

A flutuação da raiz média quadrática (*root mean square fluctuations*, RMSF) é uma medida que permite avaliar as regiões de maior flexibilidade na proteína durante a simulação. A flexibilidade estrutural de uma proteína é essencial para o desempenho da sua função biológica (Gerstein *et al.*, 1994; Huber & Bennett Jr, 1983). Podemos observar que o modelo-4JXP apresentou uma baixa flutuação na maior parte dos seus resíduos que constituem o domínio catalítico (Barril $\alpha_8\beta_8$) apresentando seis picos de maior flexibilidade (**Figura 8**). Esses pontos de maior flutuação referem-se a regiões de alças (regiões 60, 187 e 350) e duas regiões de grampos- β (regiões 105 e 373) que protendem do Barril. Destes, somente o pico de número 3 (região 165) corresponde a uma região de alfa-hélice, mais precisamente o final da hélice α_3 que foi a região que apresentou desestruturação ao longo da dinâmica como pode ser visto na **Figura 5**. Em relação aos modelos, o gerado pelo I-Tasser foi o que apresentou um comportamento mais similar ao da estrutura nativa e o 1Y24 o que mais divergiu.

Avaliando-se a estrutura secundária ao longo da simulação pode-se observar a tendência de formação ou perda da estrutura secundária de uma proteína. Na **Figura 5** podemos observar o resultado dessa análise para o modelo-4JXP gerado a partir do cristal da IDUA humana e para os três modelos gerados (IDUA-PM, IDUA-RM e I-Tasser) utilizando como molde a estrutura da beta-D-xylosidase de *T. saccharolyticum*. Em relação ao modelo-4JXP observamos um comportamento mais estável em relação a manutenção das estruturas secundárias da região do domínio catalítico. Todas as hélices que compõem o domínio catalítico permaneceram estáveis com exceção da hélice α_3 que apresentou uma pequena desestruturação a partir de 20ns de simulação. Ao longo da dinâmica observamos oscilações no tamanho de algumas folhas beta do domínio catalítico. Também observamos o surgimento de pequenas folhas que não se mantiveram ao longo da simulação.

Em relação aos modelos podemos observar uma variação maior das estruturas secundárias ao longo da simulação. Somente os modelos IDUA-PM e IDUA-I-tasser apresentaram um comportamento similar ao do cristal e com estruturação secundária mais estável em relação aos modelos 1Y24 e IDUA-RM, principalmente em relação ao domínio catalítico.

O modelo IDUA-PM apresentou alterações parciais nas hélices $\alpha 2$, $\alpha 5$, $\alpha 6$, $\alpha 7$ e a desnaturação da $\beta 7$. Entretanto, verificamos também que o grampo- β (identificado pela letra C na **Figura 5**) corretamente predito neste modelo se mostrou estável ao longo da simulação. No modelo IDUA-Itasser observamos alterações semelhantes às apresentadas em IDUA-PM. É interessante salientar que houve a formação da oitava folha beta que não estava presente no início da simulação.

O modelo 1Y24 bem como o IDUA-RM apresentaram alterações mais críticas em relação à formação e deformação de estruturas secundárias. As hélices $\alpha 2$, $\alpha 5$, $\alpha 6$, $\alpha 7$ e $\alpha 8$ sofreram desestruturação completa ou pronunciada em 50ns. Diversas alterações nas folhas β também foram observadas pra esses dois modelos. A provável causa deste problema fica evidente na **Figura 5** na qual pode ser observado o erro no alinhamento dos modelos 1y24 e IDUA-RM em relação ao cristal. A folha $\beta 8$ e a hélice $\alpha 8$, as quais delimitam o final do domínio catalítico, estavam localizadas em uma região anterior a região correta. Isto foi evidenciado durante a dinâmica molecular a qual demonstrou um comportamento altamente instável para esta região.

Um dos grandes desafios na área de modelagem molecular é a etapa da seleção do melhor modelo. Para essa finalidade foram desenvolvidas ao longo dos anos várias ferramentas com capacidade de avaliar aspectos estereoquímicos das estruturas como, por exemplo: WHAT-CHECK (Hooft *et al.*, 1996), PROCHECK (Laskowski *et al.*, 1993) e MolProbity (Lovell *et al.*, 2003). Embora essas ferramentas sejam muito utilizadas e apresentem uma alta eficácia na avaliação de erros geométricos dos modelos, elas não fornecem um escore final, impossibilitando o ranqueamento dos diferentes modelos gerados. Programas como o ProSA (Sippl, 1993), DFIRE (Zhou & Zhou, 2002) e o VERIFY3D (Eisenberg *et al.*, 1997) apresentam como resultado um único valor relacionado a qualidade geral do modelo. Uma nova geração de programas de avaliação da qualidade de modelos baseados em aprendizado de máquina tem revolucionado esse campo, uma vez que possibilitaram uma predição mais precisa da qualidade global dos modelos. O ProQ (Wallner & Elofsson, 2003) QMEAN (Benkert *et al.*, 2008) e ModFOLD (McGuffin, 2008) fazem parte dessa

nova geração de programas, os quais utilizam uma análise combinada levando em consideração características estruturais e/ou de potencial de energia. Nesse trabalho, os modelos foram avaliados utilizando diferentes ferramentas que contemplam as diferentes técnicas acima citadas, como o Procheck (avaliação estequiométrica), Verify 3D (qualidade do enovelamento) e ModFOLD. O servidor ModFOLD4 providencia uma estimativa global e uma qualidade local (por resíduo) do modelo 3D. Além disso, ele disponibiliza uma análise visual da qualidade da estrutura terciária predita. Os melhores modelos ranqueados através do Verify3D e do ModFOLD foram o IDUA-PM e o IDUA-Itasser e esses dois modelos foram os que apresentaram menor RMSD comparado com o cristal da IDUA humana (4JXP) (**Figura 3 e Tabela 4**). Além disso, a avaliação de estrutura secundária entre os modelos de IDUA (**Figura 5**) apontam para erros na etapa do alinhamento na geração dos modelos 1Y24 (publicado por Rempel e cols 2005) e IDUA-RM (realizado nesse trabalho utilizando o mesmo alinhamento da 1Y24). De fato uma avaliação mais minuciosa evidencia que esses dois modelos apresentaram em torno de 16 aminoácidos a menos na região do domínio catalítico, a região mais importante para a atividade enzimática da enzima. É no barril que ocorre a catálise dos resíduos finais do ácido idurônico de GAGs não-reduzidos, heparan sulfato e dermatan sulfato (Zhao, 1997). Por sua vez, a região do domínio catalítico dos modelos IDUA-PM e IDUA-Itasser ficaram bem modeladas com a predição das α -hélices e folhas β do barril bem próximas a estrutura cristalográfica (4JXP) (**Figura 5**).

Considerando que a modelagem era de um alto grau de dificuldade devido à ausência de moldes com identidade elevadas optamos por realizar a avaliação dos modelos através de simulações de dinâmica molecular. A dinâmica molecular é uma técnica utilizada para avaliar os sistemas moleculares em detalhes atômicos, baseada em princípios da mecânica estatística e da física clássica (Daggett, 2002; Hansson *et al.*, 2002; Karplus & McCammon, 2002; McCammon *et al.*, 1977; Warshel, 2002). Na área de modelagem a dinâmica pode ser utilizada de duas formas, tanto para avaliar modelos como para refiná-los. Simulações longas utilizando campos de forças precisos deveriam em princípio ser úteis no refinamento de modelos com baixa resolução. Há vários trabalhos

que utilizaram a dinâmica para refinar modelos (Kwon *et al.* 2011; Platania *et al.*, 2012; Zhu *et al.*, 2008). Entretanto um trabalho recente onde foram simuladas 24 proteínas escolhidas através da categoria de refinamento de estruturas do CASP (*Critical Assessment of protein Structure Prediction*) demonstrou que na grande maioria dos casos os modelos se afastaram da estrutura nativa (Raval *et al.* 2012). Na verdade, o sucesso desta abordagem depende muito da acurácia do campo de força, sendo este um fator limitante.

No caso da avaliação de modelos, como na abordagem deste trabalho, a dinâmica tem se mostrado muito eficiente (Della-Longa & Arcovito; Kulleperuma *et al.*, 2013) tanto para extrair medidas mais simples como o RMSD e raio de giro, como para avaliações mais detalhadas como acompanhamento das estruturas secundárias ao longo da simulação.

Quando consideramos a dinâmica molecular observamos que os modelos 1Y24 e IDUA-RM apresentaram desestruturação quase que total das hélices $\alpha 6$, $\alpha 7$ e $\alpha 8$, apresentando uma alta instabilidade destas estruturas durante a simulação. Os modelos IDUA-PM e IDUA-Itasser ficaram mais estáveis na simulação de dinâmica molecular considerando a estrutura secundária e também foram os melhores ranqueados de acordo com o programa ModFOLD. Embora estes dois modelos tenham apresentado um comportamento melhor em relação à manutenção das estruturas secundárias, foram verificados alguns problemas como a desestruturação parcial da hélice $\alpha 7$. Essa desestruturação pode ser devido à ausência do domínio da fibronectina, que corresponde a região final da IDUA, o qual não foi modelado. Como comentado anteriormente, a IDUA apresenta três domínios: (i) o domínio Barril $(\alpha\beta)_8$, (ii) o β -sanduíche e o (iii) domínio da fibronectina (**Figura 4**). A hélice $\alpha 7$ fica na região inferior do barril, próxima ao domínio do β -sanduíche, enquanto a hélice $\alpha 1$ fica próxima ao domínio da fibronectina. Ambos domínios parecem ser importantes para a manutenção da arquitetura do barril.

4.2 Avaliação das mutações R89W e R89Q

A publicação do cristal da IDUA permitiu-nos dar um enfoque adicional ao trabalho, o qual foi avaliar o impacto de duas mutações R89W e R89Q através da

dinâmica molecular. O modelo da IDUA selvagem completo, bem como o dos mutantes, foi construído utilizando como molde a estrutura da IDUA humana (4JXP). A qualidade estereoquímica dos modelos foi avaliada através do mapa de Ramachandran **Tabela 3**. Os modelos apresentaram valores estereoquímicos superiores ao cristal com mais de 90% dos aminoácidos nas regiões mais favoráveis.

Tabela 3: Avaliação do Gráfico de Ramachandran para os modelos de IDUA selvagem e mutantes

Regiões	Cristal 4JXP	Modelo-4JXP	IDUA-R89W	IDUA-R89Q
Mais favorecidas	86,5%	91,5 %	91,3%	91,2%
Adicionalmente permitidas	12,9%	8,1%	8,3%	8,3%
Generosamente permitidas	0,6%	0,4%	0,4%	0,6%
Não permitida	0,0%	0,0%	0,0%	0,0%

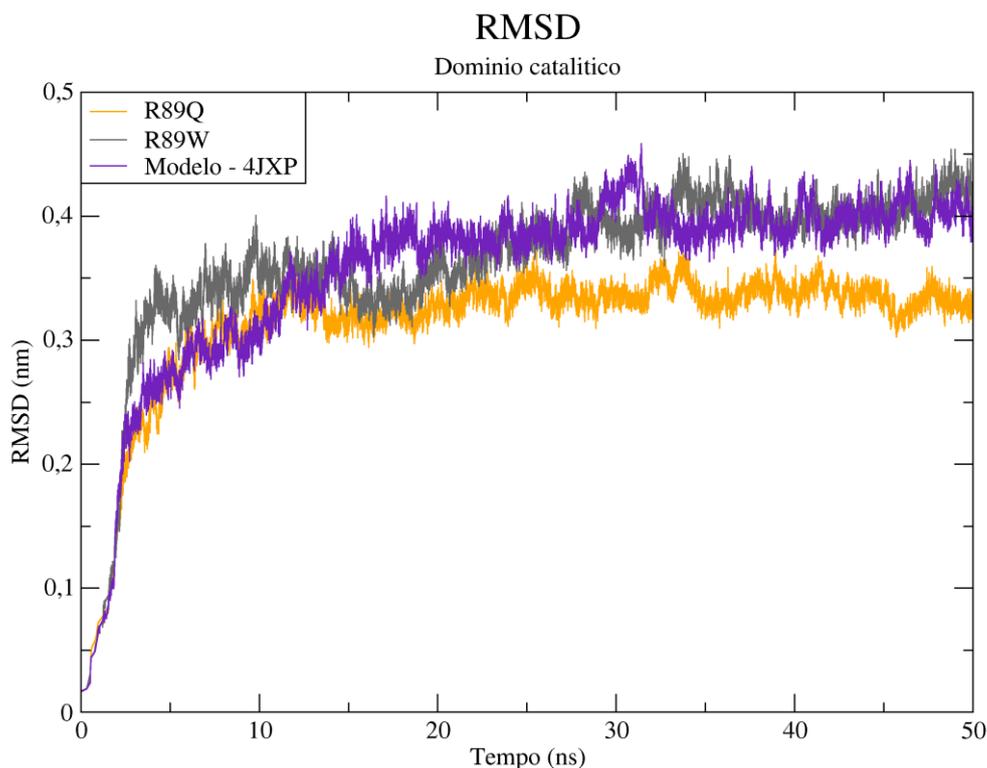


Figura 9: Desvio médio quadrático do domínio catalítico da IDUA selvagem e dos mutantes ao longo de 50 ns de simulação.

O comportamento das proteínas em solução foi avaliado através de simulações de 50ns, onde foram avaliados o RMSD, raio de giro e RMSF. Na **Figura 9** é apresentada a comparação do desvio médio quadrático do domínio catalítico da IDUA selvagem e das mutantes R89W e R89Q. O mutante R89Q foi o que apresentou um comportamento mais estável ao longo de 50 ns de simulação. A proteína selvagem estabilizou em torno de 15 ns de simulação. O mutante R89W apresentou um comportamento instável até os 30 ns de simulação após esse período parece estabilizar num desvio de 0,4 nm, semelhante ao da proteína selvagem.

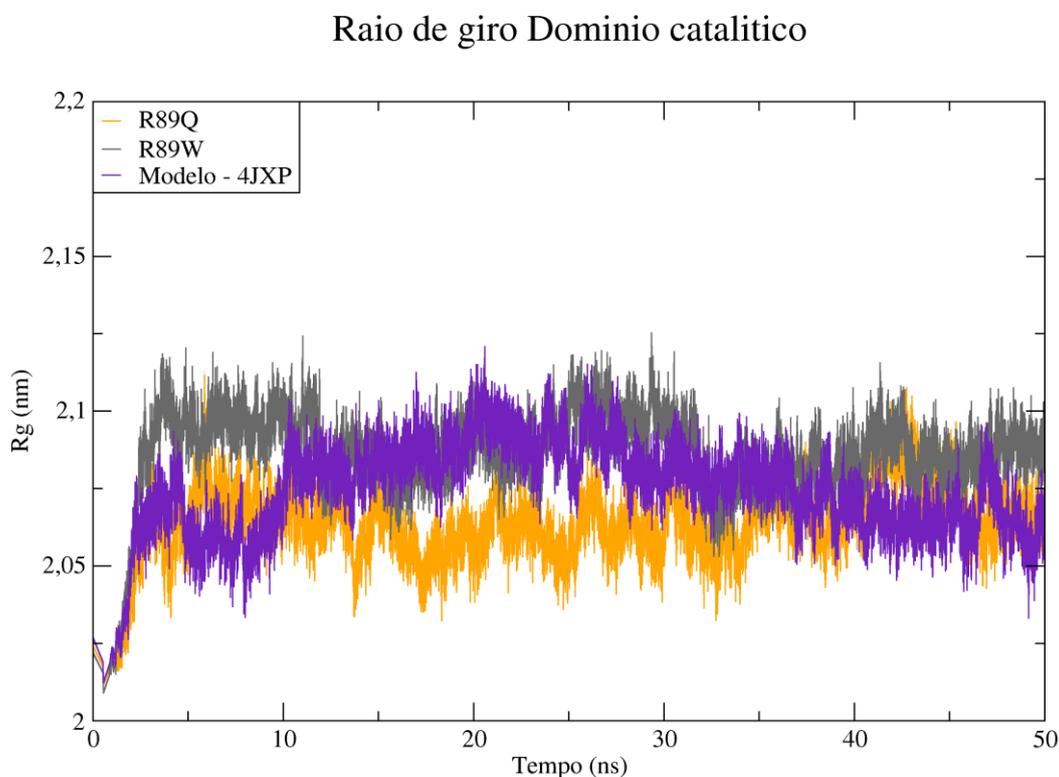


Figura 10: Raio de giro do domínio catalítico da IDUA selvagem e mutante ao longo de 50 ns.

Outro parâmetro avaliado refere-se aos movimentos de expansão e contração do domínio catalítico, realizado através da medida do raio de giro ao longo de 50 ns de simulação (**Figura 10**). Neste gráfico podemos observar que a

proteína selvagem apresenta movimentos de expansão e contração do seu barril, movimentos que ocorrem devido à ausência do substrato no sítio ativo tornando-o mais maleável. Em relação ao mutante R89Q verificamos que ele se mantém mais estável ao longo da dinâmica, dando um aspecto mais rígido ao barril. Já o mutante R89W parece apresentar três momentos de expansão. O que chama a atenção é que mesmo que esses movimentos de contração e expansão do barril sejam menores, o tamanho do barril se manteve maior no conjunto da simulação, quando comparado com a proteína selvagem e a mutante R89Q.

RMSF Dominio catalítico 0 - 50 ns

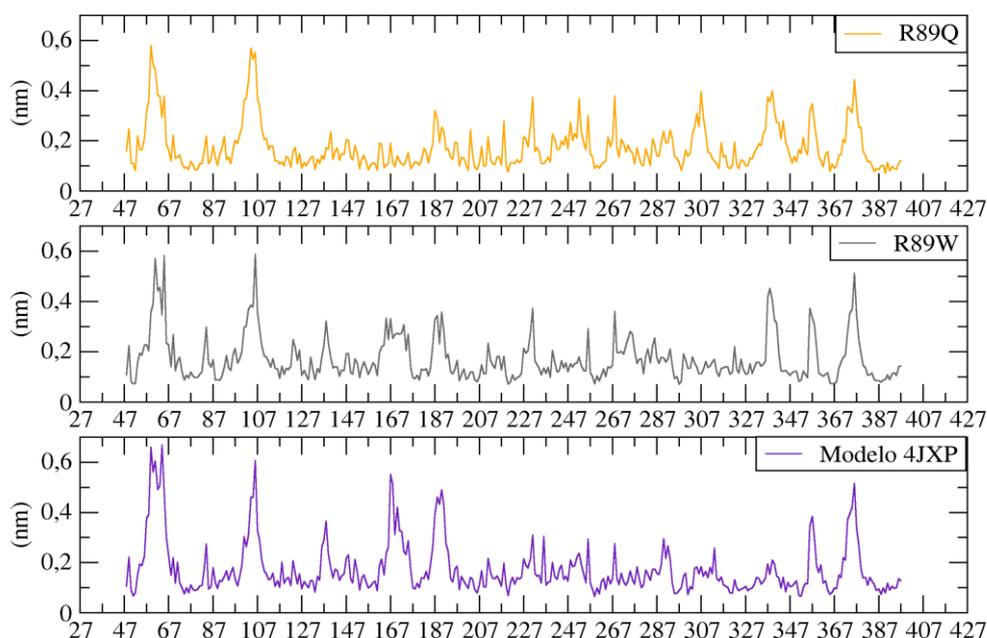


Figura 11: Flutuação quadrática média dos resíduos do domínio catalítico da IDUA.

Uma vez avaliado o comportamento do Barril em relação aos seus movimentos de contração e expansão passamos agora para uma avaliação mais detalhada da flutuação dos resíduos que fazem parte desse domínio. Verificamos que a proteína selvagem apresenta alguns picos de elevada mobilidade que se encontram em regiões de alça **Figura 11**. Há somente dois picos que não correspondem a regiões de alça, o da região 167 e o da região 287, os quais

correspondem ao final das hélices $\alpha 3$ e $\alpha 6$, respectivamente. O pico da região 140 foi mais pronunciado na proteína selvagem e ausente no mutante R89Q. Na região 137 (alça entre a $\beta 3$ e $\alpha 3$) verificamos uma mobilidade maior na selvagem e no mutante R89W em relação a R89Q. Outra diferença encontra-se na região 187 (alça entre $\beta 4$ e $\alpha 4$) onde se localiza o glutamato catalítico que apresentou menor flutuação no mutante R89Q. Por sua vez a região 305 (alça entre $\beta 7$ e $\alpha 7$), onde se localiza o glutamato nucleofílico, apresentou maior flutuação no mutante R89Q do que na selvagem e no mutante R89W. Outra alteração observada se refere a região 337 (alça entre a $\alpha 7$ e $\beta 8$), que apresenta uma maior flutuação nas mutantes do que na selvagem.

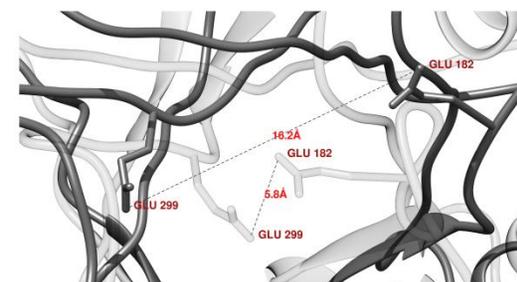
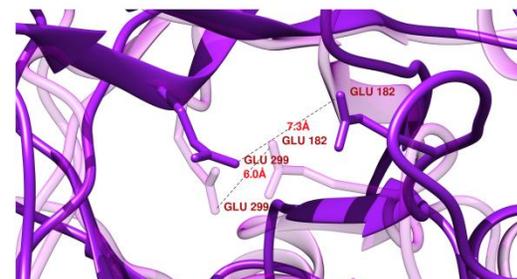
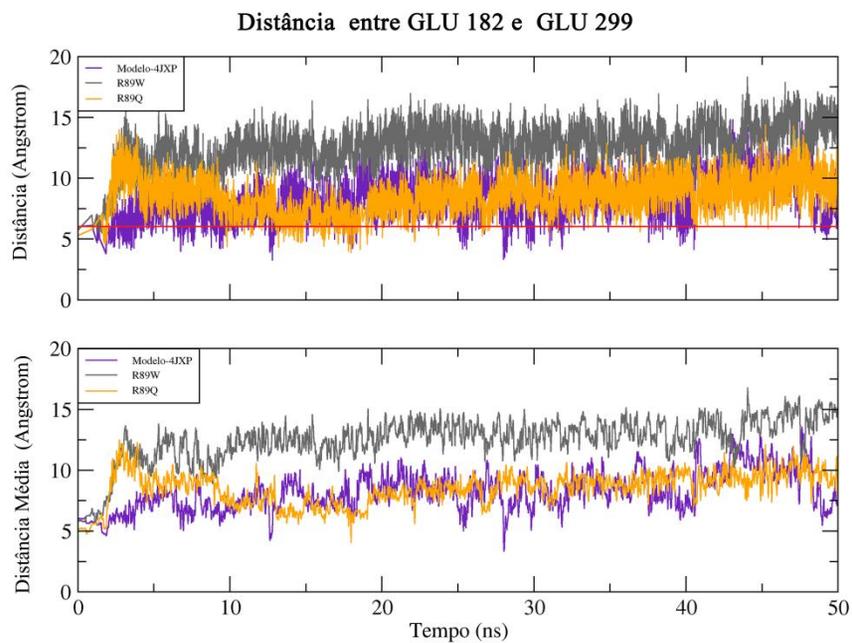


Figura 12: Variação da distância entre os resíduos catalíticos ao longo de 50 ns de simulação. (A) Distância absoluta, linha vermelha representa a região abaixo de 5,5 Å. (B) Distância média. (C) sobreposição da estrutura inicial após minimização de energia (tonalidade clara) com a estrutura no final da simulação (50ns).

Uma vez avaliado o domínio do Barril ($\alpha\beta$)₈ passamos para a avaliação dos resíduos responsáveis pela reação catalítica da IDUA, o glutamato 182 (E182) e o glutamato 299 (E299). O E182 tem a função na catálise e o E299 atua como nucleófilo. Na **Figura 12** são apresentadas as medidas da distância entre os dois glutamatos realizadas nas estruturas antes e após a simulação de 50 ns. A distância inicial observada entre os dois glutamatos após minimização de energia foi de 6 Å para a IDUA selvagem e 5,3 e 5,7 Å para os mutantes R89Q e R89W, respectivamente. Essa distância foi monitorada ao longo da simulação de 50 ns de dinâmica molecular e os resultados podem ser visualizados na Figura 12. Na proteína selvagem as distâncias entre o E182 e E299 permaneceram em média de 7,5 Å. O mutante R89Q apresentou um aumento grande nesta distância nos 10 ns iniciais da simulação chegando a alcançar uma distância máxima de 12,5 Å, após esse período houve uma estabilização em torno de 7,5 Å, semelhante a selvagem. Por outro lado o mutante R89W apresentou um comportamento bem diferenciado atingindo uma distância média entre os dois resíduos de 12,5 Å ao longo da simulação. A análise do raio de giro mostrou que embora o mutante R89W não apresente movimentos grandes de expansão e contração, o seu raio de giro do domínio catalítico se manteve maior do que a selvagem e a mutante R89Q durante a maior parte do período simulado. Este comportamento global do barril, por sua vez, pode estar afetando a distância entre os dois glutamatos. Análises de bioinformática e estudos experimentais identificaram dois resíduos no sítio ativo da IDUA, o E182 atuando como catalista ácido/base e o E299 como sendo o nucleófilo (Brooks *et al.*, 2001; Durand *et al.*, 2000; Henrissat & Davies, 1997). A família 39 das Glicosil-hidrolases além de compartilhar o mesmo domínio estrutural Barril Tim compartilham a mesma função catalítica, hidrolisando a ligação glicosídica através de um mecanismo de catalise ácida envolvendo dois passos. Estas hidrolases apresentam dois resíduos de glutamatos em direções opostas com uma distância aproximada de 5,5 Å (Henrissat & Davies, 1997). A proteína β -D-xylosidase de *T. saccharolyticum* foi a primeira proteína cristalografada desta família, e foi demonstrado que a média da

distância entre esses dois resíduos é de 5,05 Å, semelhante ao que foi observado nesse trabalho para a IDUA selvagem (Yang *et al.*, 2004). É importante ressaltar que embora houve um comportamento semelhante em relação a distância média ao longo da simulação entre a selvagem e a R89Q, foi a proteína selvagem que apresentou uma maior frequência da distância ideal de 5,5 Å entre os resíduos catalíticos ao longo da dinâmica (Figura 12 delimitada pela região abaixo da linha vermelha).

Sabe-se da importância das ligações de hidrogênio e de outras interações moleculares para a manutenção da estrutura da proteína. Considerando o exposto, outra abordagem utilizada foi avaliar o número de ligações de hidrogênio na proteína selvagem e nas mutantes. A estrutura inicial do modelo-4JXP apresentou 532 ligações de hidrogênio, esse número de ligações foi menor para os mutantes R89W e R89Q, os quais apresentaram 527 e 515 respectivamente.

Na **Figura 13** são apresentadas as quantidades de ligações de hidrogênio realizadas pelo resíduo 89 na proteína selvagem e nos mutantes na região do domínio catalítico da IDUA. Observamos que no início da simulação a proteína selvagem apresenta um maior número de interações quando comparado aos dois mutantes. Entretanto, no decorrer da dinâmica esse número diminuiu um pouco, mantendo uma média de aproximadamente 2,5 ligações. Na simulação do mutante R89Q são observadas grandes flutuações nos primeiros 25 ns, após esse período o número médio de interações fica em torno de 3. Por sua vez, a troca de uma arginina por um triptofano (R89W) diminuiu o número de interações do resíduo 89, o qual oscila entre uma e duas ligações de hidrogênio ao longo dos 50 ns de simulação.

Além disso, foram computados os resíduos que fazem pontes de hidrogênio com a arginina 89, através da recuperação de uma estrutura em cada um dos intervalos; 0 (estrutura submetida a minimização de energia), 10, 20, 30, 40 e 50 ns além da estrutura inicial. Na **Tabela 5** pode-se observar o número de ligações de hidrogênio no modelo inicial (antes da minimização de energia), sendo que a proteína selvagem apresentou 9 ligações enquanto as mutantes R89W e R89Q

apresentaram 2 e 3 ligações respectivamente. Pode-se verificar que o número de ligações varia ao longo da simulação bem como os resíduos com os quais o aminoácido 89 faz contato. A troca de uma arginina (R) por um triptofano (W) no mutante R89W causa uma diminuição no número de ligações de hidrogênio tanto na proteína inicial como durante a simulação, uma vez que foram observadas um total de duas a três ligações nos pontos amostrados. Já a substituição de uma arginina na posição 89 por uma glutamina (Q) altera o número de pontes na estrutura inicial (3 ligações), entretanto ao decorrer da simulação o número de ligações de hidrogênio aumenta, se aproximando ao observado para a selvagem. Outro ponto a ser salientado refere-se aos resíduos que formam as ligações com o resíduo R89. Chamo a atenção para os resíduos F46, T50, E178, N181 e E299, observados na estrutura inicial da IDUA selvagem e que se mantém em grande parte das avaliações durante a simulação. Embora uma interação com o glutamato 299 (E299) tenha sido observada somente na estrutura inicial, esta merece destaque uma vez que corresponde ao resíduo nucleófilo. Ao longo da simulação, esta interação parece ser substituída por uma ligação com o resíduo anterior, D298. O resíduo R89 - localizado no final da segunda folha beta do barril - estabelece contato entre a primeira (F48, T50) e quarta (E178) folha β . Além disso, também realiza contato com a asparagina 181 (N181), localizada em uma região de alça ao lado do glutamato catalítico (E182), e com o resíduo E299 (nucleófilo), localizado na alça posterior a folha β 7. Esses contatos são alterados ao longo da simulação. Uma vez delineada a rede de interações do resíduo R89 que estabelece contatos com as estruturas secundárias do barril, partimos para a visualização da rede dos mutantes. A rede para o mutante R89Q também apresenta interações ao longo da simulação com resíduos da quarta (E178) e sétima (D298) folhas betas como, observado para a proteína selvagem. O resíduo W89 estabelece contato com a primeira folha β (F48 e T50) e com a terceira (E131) folha β , perdendo as interações com a quarta e sétima folha β . Interações com as folhas β 4 e β 7 parecem ser importantes, uma vez que os resíduos E182 e E299 se localizam próximos a estas estruturas (regiões de alças adjacentes).

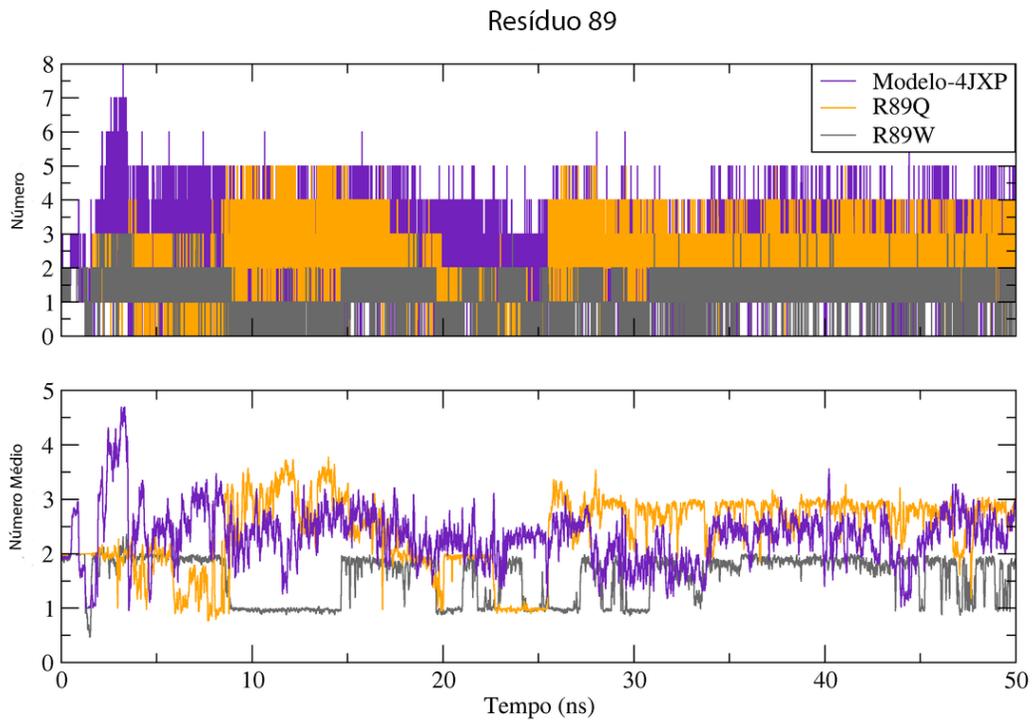


Figura 13: Impacto das mutações R89W e R89Q em relação ao número de ligações de hidrogênio realizadas pelo resíduo 89. É possível observar uma redução no número de ligações da mutante R89W, quando comparada com a proteína selvagem. Gráfico superior apresenta o número absoluto de ligações de hidrogênio ao longo do tempo, enquanto o gráfico inferior apresenta uma média dos pontos do gráfico superior.

Tabela 5: Ligações de hidrogênio formadas pelo resíduo 89 em diferentes pontos da simulação. Arginina na proteína selvagem, W e Q nos mutantes.

Modelo-4jxp (IDUA selvagem)			R89W		R89Q	
Estruturas	Resíduos	Total LH ^a	Resíduos	Total LH ^a	Resíduos	Total LH ^a
Inicial	F46, T50, E178, N181, H262, N297, E299	9	F46, T50	2	F46, T50, D298	3
0 ^b	T50, E178, N181	4	F46, T50, E131	3	F46, T50, Q87, E131, D298	5
10	F46, T50, E178, D298	5	F46, T50	2	F46, T50, E131, E178,	4
20	F46, T50, G129	3	F46, T50	2	F46, T50	2
30	F46, T50, Q87, N181	4	F46, T50	2	F46, T50, E131, E178,	4
40	F46, T50, E178, N181	4	F46, T50, E131	3	S49, F46, T50, E131, E178, N181	6
50	F46, T50, H91, G129, E178,	6	F46, T50, E131	3	S49, F46, T50, Q87, E131	5

^aLH – Ligações de hidrogênio

^b0 – Estrutura após minimização de energia

Outra abordagem realizada foi a verificação das estruturas secundárias no início e no final das simulações. Ao avaliar a sobreposição das estruturas nos extremos da simulação verificamos que os dois mutantes apresentaram uma desestruturação da hélice α_6 , mostrando que ambas mutações afetam a estrutura da proteína, mas com efeito mais acentuado no mutante R89W (**Figura 14**).

A MPS tipo I é caracterizada pelo acúmulo progressivo de glicosaminoglicanos não digeridos devido a inativação, ausência ou baixa atividade da enzima α -L-iduronidase. Esta proteína é responsável pela clivagem de resíduos terminais do ácido-idurônico do heparan sulfato e dermatan sulfato (Hopwood & Morris, 1990). Até o momento, aproximadamente 119 mutações diferentes no gene da IDUA foram identificadas e depositadas no banco de dados HGMD. Dentre estas,

duas mutações R89W e R89Q são responsáveis por um fenótipo clínico atenuado da doença (Hein *et al.*, 2003). Yang e colaboradores (2004) através da cristalização da enzima β -xylosidase, identificaram os resíduos catalíticos no sítio ativo desta enzima e demonstraram a importância da arginina 89. O sítio catalítico é conservado dentro da família 39 das glicosidases, onde a IDUA se encontra, apresentando dois glutamatos catalíticos (E182 e E299). Segundo os autores, a arginina 89 interage com o nucleófilo catalítico (E299), assegurando que ele permaneça desprotonado. Avaliando a estrutura cristalizada da IDUA depositada no PDB com o código 4JXP observamos que o resíduo R89 apresenta ligação de hidrogênio com o resíduo E299. Da mesma forma, o modelo da estrutura selvagem construído neste trabalho (Modelo-4JXP) também apresentou interação com este resíduo (E299), assegurando a qualidade do modelo utilizado neste trabalho. Cada aminoácido apresenta características específicas em relação ao tamanho, carga e hidrofobicidade. A proteína selvagem apresenta uma arginina como ativador catalítico, o qual é carregado positivamente. Entretanto, essa carga é perdida quando ocorre a troca por uma glutamina (Q), que é neutra. Além da diferença na carga há também alteração no tamanho, uma vez que o resíduo mutante (Q) é menor do que o selvagem (R). O impacto no tamanho influenciou tanto a quantidade de ligações de hidrogênio realizada por este resíduo quanto os resíduos que são contatados. Na **Figura 13** observamos que a R89 apresenta um grande número de ligações de hidrogênio nos 8 ns iniciais da simulação, alcançando um valor máximo de 8 ligações. Este valor não foi alcançado pelas mutantes Q89 e W89, que atingiram valores máximos de 5 e 3 ligações de hidrogênio, respectivamente. Em relação aos resíduos que são contatados, verificamos que o Q89 estabelece contatos novos que não foram observados na selvagem, como é o caso da interação com o E131 (**Tabela 5**). Além disso, a interação com a asparagina 181 (N181) só foi observada uma vez em 40 ns, enquanto que na R89 foi observada em todas as conformações amostradas (exceção de 20 ns).

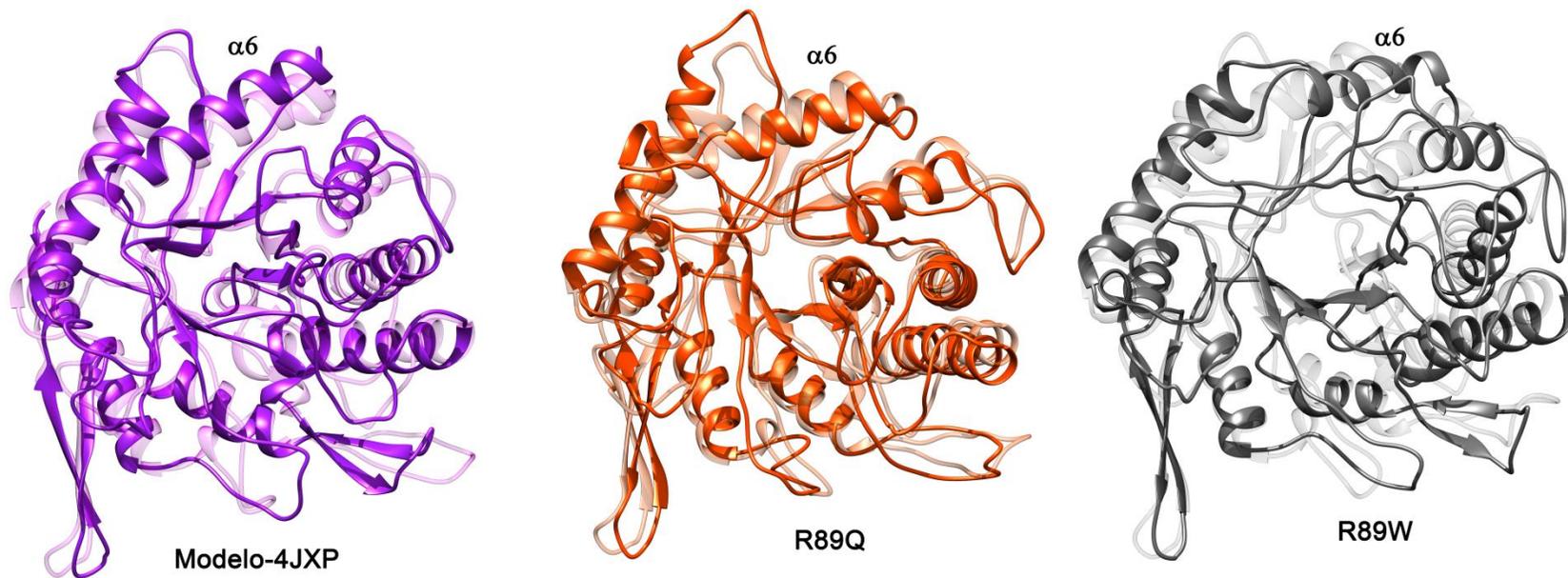


Figura 14: Comparação do domínio catalítico da IDUA selvagem e dos mutantes no início e ao final de 50 ns de simulação. Tonalidades mais claras representam a estrutura inicial, tonalidades mais escuras a estrutura final aos 50 ns. A hélice α_6 é indicada.

Com relação a análise de contatos realizados pelo resíduo W89, verificou-se que estes são estabelecidos com a fenilalanina 46 (F46), treonina 50 (T50) e glutamato 131 (E131). Estas interações contatam as folhas β 1 e β 2 do barril hidrofóbico, sendo perdido o contato com as folhas β 4 e β 7 próximo a localização dos glutamatos catalíticos (E182, E199). Estas alterações podem ser explicadas pelo fato do triptofano ser um aminoácido neutro enquanto que a arginina é carregada positivamente. Outro ponto importante é que o triptofano é maior e mais hidrofóbico do que a arginina. O encaixe de um aminoácido maior no barril interfere nas interações estéricas e no número de ligações de hidrogênio, podendo interferir no aumento do diâmetro do próprio barril. Este efeito pode ser observado no gráfico do raio de giro (**Figura 10**), onde o mutante R89W apresentou um valor maior do que a selvagem e a mutante R89Q. Outro ponto a ser salientado é o fato do mutante R89W apresentar uma distância muito maior entre os dois glutamatos catalíticos, chegando a valores de 15 Å. Todas estas alterações apontam para uma maior alteração no sítio catalítico do mutante R89W, quando comparado ao mutante R89Q. De fato, Hein e colaboradores (2003) verificaram que substituições da arginina 89 ocasionam alterações na atividade catalítica, sem ocasionar a perda completa da atividade. Através de ensaios *in vitro*, os autores determinaram que a atividade catalítica específica da IDUA R89W foi 10 vezes menor do que o controle, enquanto que a atividade específica da mutante R89Q foi apenas duas vezes menor do que o controle (Hein *et al.*, 2003). Outro ponto importante levantado pelos autores é o fato de que pacientes MPS I portadores do alelo R89W apresentavam níveis normais da proteína em extratos de fibroblastos, enquanto que os portadores da mutação R89Q apresentavam níveis reduzidos da enzima. Os pesquisadores sugeriram que a alteração de uma arginina por uma glutamina poderia estar provocando alterações estruturais na enzima, dificultando o seu transporte ao lisossomo. Quando foi avaliada a atividade total, o impacto da mutação R89Q (redução de 20 vezes) foi maior do que a R89W (redução de 10 vezes), sendo que esta diferença foi justificada

pela menor quantidade da enzima R89Q presente. Esse resultado explica o fato dos pacientes portadores do alelo R89W apresentarem um fenótipo mais brando em relação aos portadores do R89Q.

Neste trabalho observamos que a troca de uma arginina por uma glutamina na região 89 apresentou um impacto no sítio catalítico menor do que a substituição por um triptofano (W), estes resultados estão de acordo com os obtidos por Hein et al., 2003 que demonstraram através de testes *in vitro* uma atividade catalítica específica menor do R89W quando comparado ao R89Q.

Uma questão a ser discutida refere-se ao fato da mutação R89Q apresentar um impacto maior no trajeto ao lisossomo do que propriamente a sua atividade catalítica. De fato a IDUA humana apresenta seis sítios de glicosilação (Asparaginas 110, 190, 336, 372, 415 e 451) importantes para o endereçamento e transporte para o lisossomo. No complexo de golgi, enzimas específicas convertem as N-glicanas do tipo *high-mannose* em glicanas complexas, um processo que envolve uma série de passos e liberação na acessibilidade das cadeias das glicanas para o processamento pelo complexo golgi (Kermode 1996, Lerouge et AL., 1998). Dentre os seis sítios de glicosilação destaca-se a importância da glicana ligada a asparagina 372 (N372), a qual faz parte de um sistema de qualidade impedindo o transporte de proteínas que não estejam completamente dobradas. Essa etapa é realizada através da fosforilação de um oligossacarídeo, de forma que a ausência de fosforilação indica que este sítio não está acessível a fosfotransferase que catalisa o primeiro passo na síntese do sinal. Outros dois sítios de glicosilação importantes são o N336 e N451 que são marcados com a manose-6-fosfato para serem transportados via receptor manose-6-fosfato até o lisossomo (Zhao, 1997). Considerando o exposto, a mutação R89Q deve afetar estruturalmente algum ponto chave da proteína que altera o processo de glicosilação ou processamento das glicanas, entretanto, esse processo não foi avaliado neste trabalho.

5 Conclusões

A realização da modelagem molecular de proteínas que possuem moldes com baixa identidade deve ser realizada com cautela e, pelos nossos resultados, utilizar mais de uma técnica de construção de modelos. No nosso caso, as técnicas que apresentaram os melhores desempenhos foi a de modelagem mista (utilizando a ferramenta I-tasser) bem como a modelagem por homologia (através da ferramenta Modeller 9v9) utilizando um alinhamento baseado em reconhecimento de padrões de enovelamento. A escolha do melhor modelo também é outro ponto crítico para a modelagem e deve contemplar validadores com diferentes características como os apresentados e utilizados neste trabalho. Além disso, destacamos a importância da inclusão da simulação de dinâmica molecular como passo final na avaliação dos modelos levando em consideração a manutenção das estruturas secundárias.

A sobreposição do domínio catalítico dos dois melhores modelos gerados neste trabalho (IDUA-PM e I-tasser) com o cristal da IDUA (4JXP) apresentou um RMSD de 5,78 e 6,35 Å. Esses valores são duas vezes menores do que o do modelo previamente publicado 1Y24 (12,24 Å). Isto demonstra o ganho de desempenho de nossas abordagens, o que justifica a utilização deste protocolo para a modelagem das outras enzimas da via de degradação de GAGs.

Com relação aos dados de dinâmica molecular, a substituição de uma arginina (R) na posição 89 por um triptofano (W) apresentou um impacto maior no sítio ativo da IDUA do que a substituição por uma glutamina (Q). Esse impacto se refletiu tanto no número de ligações de hidrogênio realizadas pelo resíduo na posição 89 e nas distâncias dos dois glutamatos catalíticos (E182 e E299) quanto na desestruturação da hélice α_6 . Estes resultados corroboram os testes *in vitro* onde o mutante R89W apresentou uma atividade catalítica específica menor do que R89Q quando comparado à enzima selvagem (R89).

6 Perspectivas

A via de degradação de GAGs é composta por 11 enzimas, cinco delas (Iduronato sulfatase - MPS II, heparan-N-sulfatase - MPS IIIA, N-acetilglicosaminidase - MPSIIIB, acetil-CoA-alfa-glucosamina acetiltransferase - MPS IIIC, N-acetilglicosamina 6-sulfatase- MPSIIID) ainda não possuem cristal e os moldes disponíveis apresentam uma baixa identidade. Desta forma, pretendemos utilizar a metodologia deste trabalho para a modelagem da IDUA humana para modelar a estrutura tridimensional das outras enzimas da via envolvidas nas diferentes formas de MPS.

Neste trabalho escolhemos duas mutações que ocorrem no mesmo resíduo e apresentam impactos diferenciados na proteína. Até o momento conseguimos avaliar as mutações em relação ao sítio ativo inferindo relações com a sua atividade catalítica. Como perspectiva, pretendemos avaliar a simulação destas mutantes considerando tanto os sítios de glicosilação quanto os diferentes pHs encontrados pela proteína, uma vez que estes são pontos importantes tanto para o trajeto da proteína ao lisossomo quanto para a sua atividade enzimática. É importante salientar que a compreensão do efeito de cada mutação é importante tanto para a escolha do tratamento a ser administrado quanto para o desenvolvimento de novas terapias.

7 Referências

- Benkert P, Tosatto SC and Schomburg D (2008) QMEAN: A comprehensive scoring function for model quality assessment. *Proteins* 71:261-277.
- Bowie JU, Luthy R and Eisenberg D (1991) A method to identify protein sequences that fold into a known three-dimensional structure. *Science* 253:164-170.
- Branden C and Tooze J (1991) *Introduction to protein structure.*, New York.
- Braunlin EA, Berry JM and Whitley CB (2006) Cardiac findings after enzyme replacement therapy for mucopolysaccharidosis type I. *Am J Cardiol* 98:416-418.
- Braunlin EA, Harmatz PR, Scarpa M, Furlanetto B, Kampmann C, Loehr JP, Ponder KP, Roberts WC, Rosenfeld HM and Giugliani R (2011) Cardiac disease in patients with mucopolysaccharidosis: presentation, diagnosis and management. *J Inherit Metab Dis* 34:1183-1197.
- Brooks Da, Fabrega S, Hein LK, Parkinson EJ, Durand P, Yogalingam G, Matte U, Giugliani R, Dasvarma a, Eslahpazire J *et al.* (2001) Glycosidase active site mutations in human alpha-L-iduronidase. *Glycobiology* 11:741-750.
- Cavasotto CN and Phatak SS (2009) Homology modeling in drug discovery: current trends and applications. *Drug discovery today* 14:676-683.
- Chandar SS and Mahalingam K (2012) Mucopolysaccharidosis type I: Homology modeling and docking analysis of the lysosomal enzyme, human α -L-iduronidase. *African Journal of Pharmacy and Pharmacology* 6:2027.
- Chothia C and Lesk AM (1986) The relation between the divergence of sequence and structure in proteins. *EMBO J* 5:823-826.
- Coutinho MF, Prata MJ and Alves S (2012) Mannose-6-phosphate pathway: a review on its role in lysosomal function and dysfunction. *Molecular genetics and metabolism* 105:542-550.
- D'Aco K, Underhill L, Rangachari L, Arn P, Cox GF, Giugliani R, Okuyama T, Wijburg F and Kaplan P (2012) Diagnosis and treatment trends in mucopolysaccharidosis I: findings from the MPS I Registry. *Eur J Pediatr* 171:911-919.
- D'Alfonso G, Tramontano A and Lahm A (2001) Structural conservation in single-

- domain proteins: implications for homology modeling. *J Struct Biol* 134:246-256.
- Daggett V (2002) Molecular dynamics simulations of the protein unfolding/folding reaction. *Acc Chem Res* 35:422-429.
- Della-Longa S and Arcovito A (2013) Structural and functional insights on folate receptor alpha (FR α) by homology modeling, ligand docking and molecular dynamics. *J Mol Graph Model* 44C:197-207.
- Dietz HC (2010) New therapeutic approaches to mendelian disorders. *N Engl J Med* 363:852-863.
- Durand P, Fabrega S, Henrissat B, Mornon JP and Lehn P (2000) Structural features of normal and mutant human lysosomal glycoside hydrolases deduced from bioinformatics analysis. *Human molecular genetics* 9:967-977.
- Durand P, Lehn P, Callebaut I, Fabrega S, Henrissat B and Mornon JP (1997) Active-site motifs of lysosomal acid hydrolases: invariant features of clan GH-A glycosyl hydrolases deduced from hydrophobic cluster analysis. *Glycobiology* 7:277-284.
- Eisenberg D, Luthy R and Bowie JU (1997) VERIFY3D: assessment of protein models with three-dimensional profiles. *Methods Enzymol* 277:396-404.
- Eswar N, Webb B, Marti-Renom MA, Madhusudhan MS, Eramian D, Shen MY, Pieper U and Sali A (2007) Comparative protein structure modeling using MODELLER. *Curr Protoc Protein Sci* Chapter 2:Unit 2 9.
- Gerstein M, Lesk AM and Chothia C (1994) Structural mechanisms for domain movements in proteins. *Biochemistry* 33:6739-6749.
- Ginalski K (2006) Comparative modeling for protein structure prediction. *Curr Opin Struct Biol* 16:172-177.
- Giugliani R, Federhen A, Munoz Rojas MV, Vieira TA, Artigalás O, Pinto LL, Azevedo AC, Acosta AX, Bomfim C, Lourenço CM *et al.* (2010a) [Enzyme replacement therapy for mucopolysaccharidoses I, II and VI: recommendations from a group of Brazilian F experts]. *Rev Assoc Med Bras* 56:271-277.
- Giugliani R, Federhen A, Rojas MVM, Vieira T, Artigalás O, Pinto LL, Azevedo AC, Acosta A, Bomfim C, Lourenço CM *et al.* (2010b) Mucopolysaccharidosis I, II, and VI: Brief review and guidelines for treatment. *Genetics and molecular biology* 33:589-604.

- Hansson T, Oostenbrink C and van Gunsteren W (2002) Molecular dynamics simulations. *Curr Opin Struct Biol* 12:190-196.
- Hartung SD, Frandsen JL, Pan D, Koniar BL, Graupman P, Gunther R, Low WC, Whitley CB and McIvor RS (2004) Correction of metabolic, craniofacial, and neurologic abnormalities in MPS I mice treated at birth with adeno-associated virus vector transducing the human alpha-L-iduronidase gene. *Mol Ther* 9:866-875.
- Harvey AJ, Hrmova M, De Gori R, Varghese JN and Fincher GB (2000) Comparative modeling of the three-dimensional structures of family 3 glycoside hydrolases. *Proteins* 41:257-269.
- Hein LK, Hopwood JJ, Clements PR and Brooks Da (2003) The α -L-iduronidase mutations R89Q and R89W result in an attenuated mucopolysaccharidosis type I clinical presentation. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease* 1639:95-103.
- Henrissat B and Davies G (1997) Structural and sequence-based classification of glycoside hydrolases. *Curr Opin Struct Biol* 7:637-644.
- Hillisch A, Pineda LF and Hilgenfeld R (2004) Utility of homology models in the drug discovery process. *Drug Discov Today* 9:659-669.
- Hooft RW, Vriend G, Sander C and Abola EE (1996) Errors in protein structures. *Nature* 381:272.
- Hopwood JJ and Morris CP (1990) The mucopolysaccharidoses. Diagnosis, molecular genetics and treatment. *Mol Biol Med* 7:381-404.
- Huber R and Bennett Jr WS (1983) Functional significance of flexibility in proteins. *Biopolymers* 30:239–256.
- Karplus M and McCammon JA (2002) Molecular dynamics simulations of biomolecules. *Nat Struct Biol* 9:646-652.
- Kelley LA and Sternberg MJ (2009) Protein structure prediction on the Web: a case study using the Phyre server. *Nat Protoc* 4:363-371.
- Kulleperuma K, Smith SM, Morgan D, Musset B, Holyoake J, Chakrabarti N, Cherny VV, DeCoursey TE and Pomes R (2013) Construction and validation of a homology model of the human voltage-gated proton channel hHV1. *J Gen Physiol* 141:445-465.

- Kwon T, Harris AL, Rossi A and Bargiello TA (2011) Molecular dynamics simulations of the Cx26 hemichannel: evaluation of structural models with Brownian dynamics. *J Gen Physiol* 138:475-493.
- Laskowski RA, Moss DS and Thornton JM (1993) Main-chain bond lengths and bond angles in protein structures. *J Mol Biol* 231:1049-1067.
- Liu T, Tang GW and Capriotti E (2011) Comparative modeling: the state of the art and protein drug target structure prediction. *Comb Chem High Throughput Screen* 14:532-547.
- Lovell SC, Davis IW, Arendall WB, 3rd, de Bakker PI, Word JM, Prisant MG, Richardson JS and Richardson DC (2003) Structure validation by Calpha geometry: phi,psi and Cbeta deviation. *Proteins* 50:437-450.
- Luthy R, Bowie JU and Eisenberg D (1992) Assessment of protein models with three-dimensional profiles. *Nature* 356:83-85.
- Marti-Renom MA, Stuart AC, Fiser A, Sanchez R, Melo F and Sali A (2000) Comparative protein structure modeling of genes and genomes. *Annu Rev Biophys Biomol Struct* 29:291-325.
- Mayer FQ, 2011 Desenvolvimento de Novas Abordagens Mucopolissacaridose Tipo I. Tese de Doutorado.
- Mayer FQ, Artigalás OA, Lagranha VL, Baldo G, Schwartz IV, Matte U, Giugliani R (2012) Chloraphenicol enhances IDUA activity on fibroblasts from mucopolysaccharidosis I patients *Curr Pharm Biotechnol*. 2013 Feb;14(2):194-8.
- McCammon JA, Gelin BR and Karplus M (1977) Dynamics of folded proteins. *Nature* 267:585-590.
- McGuffin LJ (2008) The ModFOLD server for the quality assessment of protein structural models. *Bioinformatics (Oxford, England)* 24:586-587.
- McGuffin LJ, Buenavista MT and Roche DB (2013) The ModFOLD4 server for the quality assessment of 3D protein models. *Nucleic Acids Res* 41:W368-W372.
- Muenzer J and Fisher A (2004) Advances in the treatment of mucopolysaccharidosis type I. *The New England journal of medicine* 350:1932-1934.
- Muro S (2010) New biotechnological and nanomedicine strategies for treatment of

lysosomal storage disorders. Wiley interdisciplinary reviews. Nanomedicine and nanobiotechnology 2:189-204.

Neufeld EF and Muenzer J (2001) The mucopolysaccharidoses. In: Scriver CR, Beaudet AL, Sly WS, Valle D, Childs B, Kinzler KW and Vogelstein B (eds) The Metabolic and Molecular Bases of Inherited Disease. 8th edition edition. McGraw-Hill, Medical Publishing Division, pp 3421.

Parenti G (2009) Treating lysosomal storage diseases with pharmacological chaperones: from concept to clinics. EMBO Mol Med 1:268-279.

Pastores GM, Arn P, Beck M, Clarke JTR, Guffon N, Kaplan P, Muenzer J, Norato DYJ, Shapiro E, Thomas J *et al.* (2007) The MPS I registry: design, methodology, and early findings of a global disease registry for monitoring patients with Mucopolysaccharidosis Type I. Molecular genetics and metabolism 91:37-47.

Platania C, Salomone S, Leggio G, Drago F and Bucolo C (2012) Homology Modeling of Dopamine D2 and D3 Receptors: Molecular Dynamics Refinement and Docking Evaluation. Plos One 7:1-12.

Prasad VK and Kurtzberg J (2010) Transplant outcomes in mucopolysaccharidoses. Seminars in hematology 47:59-69.

Raval A, Piana S, Eastwood MP, Dror RO and Shaw DE (2012) Refinement of protein structure homology models via long, all-atom molecular dynamics simulations. Proteins 80:2071-2079.

Rempel BP, Clarke LA and Withers SG (2005) A homology model for human alpha-l-iduronidase: insights into human disease. Molecular genetics and metabolism 85:28-37.

Roy A, Kucukural A and Zhang Y (2007) I-TASSER: a unified platform for automated protein structure and function prediction. Nat Protoc 5:725-738.

Sali A and Blundell TL (1993) Comparative protein modelling by satisfaction of spatial restraints. J Mol Biol 234:779-815.

Sánchez MB and Fuentes FJR (2001) Mucopolisacaridosis. In: Sanjurjo P and Baldellou A (eds) Diagnóstico y tratamiento de las enfermedades metabólicas hereditarias. Ediciones Argon, Madrid.

Schuchman EH and Desnick RJ (1988) Mucopolysaccharidosis type I subtypes. Presence of immunologically cross-reactive material and in vitro enhancement

of the residual alpha-L-iduronidase activities. *J Clin Invest* 81:98-105.

Sifuentes M, Doroshov R, Hoft R, Mason G, Walot I, Diament M, Okazaki S, Huff K, Cox GF, Swiedler SJ *et al.* (2007) A follow-up study of MPS I patients treated with laronidase enzyme replacement therapy for 6 years. *Mol Genet Metab* 90:171-180.

Sippl MJ (1993) Recognition of errors in three-dimensional structures of proteins. *Proteins* 17:355-362.

Soliman OI, Timmermans RG, Nemes A, Vletter WB, Wilson JH, ten Cate FJ and Geleijnse ML (2007) Cardiac abnormalities in adults with the attenuated form of mucopolysaccharidosis type I. *J Inher Metab Dis* 30:750-757.

Tylki-Szymanska A, Rozdzyńska A, Jurecka A, Marucha J and Czartoryska B (2010) Anthropometric data of 14 patients with mucopolysaccharidosis I: retrospective analysis and efficacy of recombinant human alpha-L-iduronidase (laronidase). *Mol Genet Metab* 99:10-17.

Van Der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE and Berendsen HJ (2005) GROMACS: fast, flexible, and free. *J Comput Chem* 26:1701-1718.

Vijay S and Wraith JE (2005) Clinical presentation and follow-up of patients with the attenuated phenotype of mucopolysaccharidosis type I. *Acta Paediatr* 94:872-877.

Wallner B and Elofsson A (2003) Can correct protein models be identified? *Protein Sci* 12:1073-1086.

Warshel A (2002) Molecular dynamics simulations of biological reactions. *Acc Chem Res* 35:385-395.

Winchester B (2005) Lysosomal metabolism of glycoproteins. *Glycobiology* 15:1R-15R.

Yang JK, Yoon HJ, Ahn HJ, Lee BI, Pedelacq JD, Liong EC, Berendzen J, Laivenieks M, Vieille C, Zeikus GJ *et al.* (2004) Crystal structure of beta-D-xylosidase from *Thermoanaerobacterium saccharolyticum*, a family 39 glycoside hydrolase. *J Mol Biol* 335:155-165.

Zhang Y (2008) I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics* 9:40.

Zhao KW (1997) Carbohydrate Structures of Recombinant Human alpha -L-

Iduronidase Secreted by Chinese Hamster Ovary Cells. *Journal of Biological Chemistry* 272:22758-22765.

Zhou H and Zhou Y (2002) Distance-scaled, finite ideal-gas reference state improves structure-derived potentials of mean force for structure selection and stability prediction. *Protein Science* 11:2714–2726.

Zhu J, Fan H, Periole X, Honig B and Mark AE (2008) Refining homology models by combining replica-exchange molecular dynamics and statistical potentials. *Proteins* 72:1171-1188.