

Universidade Federal do Rio Grande do Sul

CARACTERIZAÇÃO MOLECULAR DO GENE DA ARILSULFATASE A EM
PACIENTES BRASILEIROS COM LEUCODISTROFIA METACROMÁTICA E
ANÁLISE ESTRUTURAL DA ENZIMA

Madza Yasodara Farias Virgens

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

Orientadora: Profa. Dra. Maria Luiza Saraiva-Pereira

Porto Alegre, outubro de 2010

INSTITUIÇÕES E FONTES FINANCIADORAS

Este trabalho foi realizado no Laboratório de Identificação Genética do Centro de Pesquisa Experimental, no Laboratório de Genética Molecular do Serviço de Genética Médica, ambos no Hospital de Clínicas de Porto Alegre, e no Laboratório de Bioinformática Estrutural do Centro de Biotecnologia da Universidade Federal do Rio Grande do Sul, com o apoio financeiro do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Fundação de Apoio à Pesquisa do Rio Grande do Sul (FAPERGS) e Fundação de Incentivo à Pesquisa do Hospital de Clínicas de Porto Alegre (FIPE-HCPA).

AGRADECIMENTOS

À minha orientadora Profa. Dra. Maria Luiza Saraiva-Pereira,
ao Prof. Dr. Hugo Verli, líder do Laboratório de Bioinformática Estrutural,
aos amigos do Laboratório de Identificação Genética,
aos amigos do Laboratório de Bioinformática Estrutural,
ao PPGBM,
às Instituições que forneceram o apoio financeiro: CNPq, FAPERGS, UFRGS e HCPA
à minha família.

Dedico isso e tudo mais a Danilo Fraga

SUMÁRIO

LISTA DE ABREVIATURAS	iv
NOMENCLATURA DE MUTAÇÕES	v
RESUMO	vi
ABSTRACT	vii
1. INTRODUÇÃO	1
1.1. ASPECTOS CLÍNICOS E EPIDEMIOLÓGICOS DA MLD	1
1.2. ASPECTOS BIOQUÍMICOS DA MLD	3
1.3. ASPECTOS GENÉTICOS DA MLD	8
1.4. PSEUDODEFICIÊNCIA DE ARSA E OUTROS POLIMORFISMOS NÃO PATOGÊNICOS DO GENE DA ARSA	12
2. JUSTIFICATIVA	14
3. OBJETIVOS	15
3.1. OBJETIVO GERAL	15
3.2. OBJETIVOS ESPECÍFICOS	15
4. ARTIGO I	16
Genotypic Characterization of Brazilian Metachromatic Leukodystrophy patients	16
5. ARTIGO II	47
The Effects of Glycosylation and pH Conditions in Molecular Dynamics of Human Arylsulfatase A	47
6. DISCUSSÃO	82
6.1. CARACTERIZAÇÃO GENOTÍPICA	82
6.2. O PAPEL ESTRUTURAL DA GLICOSILAÇÃO DE ARSA	86
6.3. O PAPEL ESTRUTURAL DO pH PARA ARSA	88
7. CONCLUSÕES	90
8. REFERÊNCIAS BIBLIOGRÁFICAS	91

LISTA DE ABREVIATURAS

ARSA	arilsulfatase A
PD-ARSA	pseudodeficiência de ARSA
DLD	doença lisossômica de depósito
DM	dinâmica molecular
EIM	erros inatos do metabolismo
FGE	enzima geradora de formilglicina
FGly	formilglicina
HCPA	Hospital de Clínicas de Porto Alegre
HGMD	<i>Human Gene Mutation Database</i>
LEIM	laboratório de erros inatos do metabolismo
M6P	manose-6-fosfato
M6PR	receptor de manose-6-fosfato
MLD	Leucodistrofia metacromática
MSD	Deficiência de múltiplas sulfatases
RE	retículo endoplasmático
SAP-B	saposina B
Sp1	fator de transcrição específico da proteína 1
SUMF	fator modificador 1 de sulfatases

NOMENCLATURA DE MUTAÇÕES

Neste estudo, a nomenclatura adotada para as mutações no gene ARSA foi baseada nas recomendações da *Human Genome Variation Society* - HGVS (<http://www.hgvs.org>) (den Dunnen e Antonarakis, 2000).

RESUMO

A leucodistrofia metacromática (MLD) é um erro inato do metabolismo de herança autossômica recessiva, dividido em três subtipos clínicos de acordo com a idade de início dos sintomas. Na MLD ocorre deposição intralissossômica de glicolípídeos sulfatados principalmente no sistema nervoso central. A causa dessa doença é a deficiência de arilsulfatase A (ARSA), uma glicoproteína lisossômica que catalisa a hidrólise de glicolípídeos sulfatados constituintes da bainha de mielina. Até o momento cerca de 140 mutações relacionadas à manifestação de MLD foram descritas no gene da ARSA, das quais duas delas ocorrem com alta frequência na maioria das populações estudadas. Ademais, entre 7% e 12% das pessoas saudáveis têm deficiência *in vitro* de ARSA, condição denominada pseudodeficiência de ARSA (PD-ARSA). Não existem dados de caracterização genotípica de pacientes brasileiros com MLD. Da mesma forma, dados sobre a dinâmica molecular (DM) da ARSA também são escassos, em parte por dificuldades metodológicas implicadas no estudo desses aspectos em glicoproteínas. Nesse contexto, os objetivos desse trabalho foram identificar os alelos mutantes no gene da ARSA em pacientes brasileiros com MLD e avaliar os efeitos da glicosilação e do pH na DM da ARSA. A amostra foi composta por 27 pacientes com MLD. A mutação mais frequente (c.459+1G>A) e polimorfismos constituintes do alelo PD-ARSA foram detectados por PCR em tempo real e as demais mutações por sequenciamento direto do gene da ARSA. As DM realizadas foram de (i) ARSA não glicosilada em pH~7 e (ii) em pH~5, (iii) ARSA triplamente glicosilada e (iv) deficientemente glicosilada p.N350S, usando GROMACS. Nesse estudo a mutação c.459+1G>A foi a mais frequente (0,50), conforme esperado pela alta prevalência de pacientes com MLD infantil em nossa *coorte*. As mutações p.P426L (0,08) e c.103_110del8 (0,08) também apresentaram frequências relevantes quando comparadas às demais mutações. No conjunto, 11 mutações raras foram identificadas, incluindo 2 mutações novas: p.S44P e p.P284S. Além das mutações potencialmente deletérias, foram identificados 2 polimorfismos neutros frequentemente associados à mutação c.459+1G>A (p.W193C [0,54] e p.T391S [0,65]) e 2 polimorfismos constituintes do alelo PD-ARSA (p.N350S [0,15] e c.1524+95A>G [0,04]). As análises estruturais demonstraram um papel fundamental tanto da glicosilação, quanto do meio ácido na estabilidade da ARSA, o que é compatível com sua atividade lisossomal.

ABSTRACT

Metachromatic leukodystrophy (MLD) is an inborn error of metabolism inherited in an autosomal recessive manner. MLD is a neurodegenerative condition divided into three clinical types according to age at onset of symptoms. In MLD, intralysosomal deposition of sulfated glycolipids is observed mainly in the central nervous system. The biochemical defect associated to MLD is deficiency of arylsulfatase A (ARSA), a lysosomal glycoprotein that catalyzes degradation of sulfated glycolipids found in the myelin sheath. So far, around 140 MLD-associated mutations have been identified in ARSA gene. However, two of those remain at high frequency in the majority of studied populations. In addition to MLD, *in vitro* ARSA deficiency is observed in approximately 7-12% of healthy population, a condition named ARSA-pseudodeficiency (ARSA-PD). Up to date, there is no available data on genotypic characterization of Brazilian MLD patients. In the same way, data on ARSA molecular dynamics (MD) are also sparse, in part due by methodological difficulties related to studies of these aspects in glycoproteins. In this context, the aims of this work were to identify mutant alleles in the ARSA gene from MLD Brazilian patients, and evaluate the effects of pH and glycosylation in ARSA MD. Sample population was composed by 27 MLD patients. The most frequent MLD-associated ARSA mutation (c.459+1G>A) and the ARSA-PD polymorphisms were detected by real-time PCR, and the remaining mutations were detected by direct sequencing of ARSA gene. The performed MD were (i) non-glycosylated ARSA at pH~7 and (ii) at pH~5, (iii) fully-glycosylated ARSA at pH~5, and (iv) poorly-glycosylated ARSA-mutant p.N350S, using GROMACS. In this study the most frequent mutation was c.459+1G>A (0.50), as expected by the high prevalence of infantile-MLD cases. Mutations p.P426L (0.08) and c.103_110del8 (0.08) were also found at relevant frequencies when compared to other mutations. In total, 11 MLD-associated rare mutations were identified, including 2 novel putative disease-causing: p.S44P and p.P284S. Besides deleterious mutations, 2 polymorphisms frequently associated with c.459+1G>A (p.W193C [0.54] and p.T391S [0.65]) and 2 polymorphisms constituents of ARSA-PD allele (p.N350S [0.15] and c.1524+95A>G [0.04]) were observed. The structural analysis demonstrated the fundamental role related to glycosylation and the acid medium in the stability of ARSA that is compatible with its lysosomal activity.

1. INTRODUÇÃO

1.1. ASPECTOS CLÍNICOS E EPIDEMIOLÓGICOS DA MLD

A Leucodistrofia metacromática (MLD) é um erro inato do metabolismo (EIM), caracterizado pelo depósito intralisossômico de glicolipídeos sulfatados no sistema nervoso central e periférico e, em menor extensão, em órgãos viscerais, como rins e fígado. Estes glicolipídeos sulfatados acumulam-se como inclusões metacromáticas, que podem ser observadas principalmente nos lisossomos de macrófagos e promovem desmielinização progressiva, à qual estão relacionados eventos neurológicos clinicamente predominantes (Fluharty, 2006).

A primeira descrição de um caso de MLD foi publicada em 1910, pelo neurologista alemão Alois Alzheimer, que encontrou manchas metacromáticas no sistema nervoso de um paciente adulto com esclerose difusa (von Figura *et al*, 2001).

Em 1921, Whitte descreveu um paciente com acúmulo de material metacromático não somente no cérebro, mas também no fígado e rins. Pouco depois, em 1925, Scholz descreveu a ocorrência de três crianças aparentadas que apresentavam leucodistrofia progressiva, entretanto não foram observadas inclusões metacromáticas devido a inadequações no método de coloração utilizado na ocasião, em que o tratamento com álcool removia os lipídeos responsáveis pela metacromasia. Trinta anos depois, usando uma coloração diferente, Peiffer foi capaz de demonstrar metacromasia nas secções criopreservadas dos pacientes originais de Scholz por meio de um tratamento com ácido acético e violeta-cresil (von Figura *et al*, 2001).

De modo geral, o curso da MLD é marcado por paralisia, cegueira e retardo mental severo, entretanto diferenças podem ser observadas quanto à gravidade, velocidade de progressão e idade de início, que permitem a distinção de três subtipos clínicos:

MLD infantil – Na MLD infantil, os sintomas iniciam entre os 6 meses e 4 anos de idade e incluem fraqueza e hipotonia. A progressão do quadro clínico ocorre rapidamente, evoluindo com grave regressão mental, perda da fala, atrofia ótica e ataxia. Em seu estágio final, os pacientes perdem toda a capacidade de estabelecer contato com o ambiente; perdem a visão e são alimentados por meio de tubo gástrico. A morte ocorre geralmente cinco anos após o início dos sintomas clínicos (von Figura *et al*, 2001).

MLD juvenil – Na MLD juvenil, os sintomas iniciam entre os 4 e os 16 anos e caracterizam-se por disfunção motora e/ou deficiência mental com manifestações de psicose, dentre outros distúrbios do comportamento. A progressão da doença é semelhante à forma infantil, entretanto ocorre mais lentamente e é marcada por convulsões parciais em vez de generalizadas. A maioria dos pacientes morre entre 10 e 20 anos após o início dos sintomas clínicos (von Figura *et al*, 2001).

MLD adulta – Na MLD adulta, os sintomas iniciam após a maturidade sexual, não excedendo os 40 a 50 anos. Esse subtipo clínico apresenta predominantemente sintomas psiquiátricos, que são muitas vezes diagnosticados como esquizofrenia ou depressão psicótica. Os sinais iniciais envolvem declínio no rendimento escolar ou laboral por perda de habilidades intelectuais. O curso da doença é variável, com períodos de estabilidade e declínio alternados e, por esse motivo, pode variar de alguns poucos anos até décadas (von Figura *et al*, 2001).

Na maioria das populações estudadas, a MLD infantil é a forma mais comum, ocorrendo em 50% a 60% dos casos. As formas juvenil e adulta ocorrem mais raramente, compreendendo 20% a 30% e 15% a 20% dos casos, respectivamente (von Figura *et al*, 2001).

Tabela 1. Distribuição de prevalência de MLD em diferentes países europeus.

País	Número de pacientes			Total dos casos em cada país
	MLD infantil	MLD juvenil	MLD adulta	
Itália	32	-	-	32
Áustria	4	5	4	13
Bélgica	10	2	-	12
República Tcheca e Eslováquia	5	2	6	13
França	14	7	-	21
Alemanha	34	40	28	102
Inglaterra	9	7	2	18
Holanda	2	26	21	49
Grécia	4	2	1	7
Polônia	13	17	8	38
Portugal	13	1	3	17
Espanha	25	2	5	32
Suécia	12	-	1	13
Suíça	6	1	-	7
Ucrânia	3	6	1	10
Soma total dos casos em cada subtipo clínico	186	118	80	384
Total dos casos em cada subtipo clínico (%)	48	31	21	100

*Adaptado de Lugowska *et al*, 2005.

A incidência geral da MLD é de 1:40.000, entretanto esse número é ainda maior em algumas populações consanguíneas:

- 1:75 em judeus Habanitas israelenses (Zlotogora *et al*, 1995).
- 1:8.000 em árabes israelenses (Zlotogora *et al*, 1994).
- 1:10.000 em árabes cristãos israelenses (Zlotogora *et al*, 1994).
- 1:2.500 para a porção oeste da Nação Navajo nos E.U.A. (Holve *et al*, 2001).

1.2. ASPECTOS BIOQUÍMICOS DA MLD

Em 1959, a metacromasia presente nos tecidos de pacientes com MLD foi comprovada como sendo resultado da deposição de glicolípídeos sulfatados *in situ* (Austin *et al*, 1959). O acúmulo de sulfatídeos, principalmente no sistema nervoso e periférico, ocorre pelo prejuízo em sua hidrólise enzimática, que tem como primeira etapa a desulfatação dos resíduos de 3-O-sulfatogalactosil contidos nos glicolípídeos de membrana (Fluharty, 2006).

Os sulfatídeos compreendem cerca de 3,5% a 4,0% do total de lipídeos da mielina, onde estão ligados à proteína básica de mielina, o principal componente estrutural da bainha de mielina, que atua como isolante térmico e facilita a transmissão do impulso nervoso. Os sulfatídeos são formados a partir da reação de galactosilceramida com 5-fosfoadenosina-5-fosfosulfato, numa etapa catalisada por uma sulfotransferase. A síntese máxima ocorre durante o período de mielinização e continua mais lentamente no indivíduo adulto. Tanto neurônios quanto células oligodendrogliais são capazes de sintetizar sulfatídeos, entretanto a atividade de sulfotransferase é maior em células gliais (von Figura *et al*, 2001).

A desulfatação pela enzima lisossômica arilsulfatase A (ARSA) é a primeira etapa na degradação do 3-sulfatogalactosil contido nos glicolípídeos de membrana, numa reação combinada com um ativador protéico não enzimático, a saposina B (SAP-B). A deficiência de ARSA é o defeito enzimático relacionado à MLD, entretanto, eventualmente, pode ocorrer deficiência de SAP-B, quando a atividade de ARSA demonstra-se normal *in vitro*, o que resulta num quadro de MLD dado por deficiência de SAP-B (Grossi *et al*, 2008).

A ARSA é um membro prototípico da família das sulfatases (Gosh, 2007). As sulfatases compõem uma família altamente conservada evolutivamente de enzimas essenciais para a degradação e remodelamento de ésteres de sulfato e em todas elas o resíduo catalítico chave é gerado pela oxidação pós-traducional do grupo CH₂SH de uma

cisteína conservada em um grupo aldeídico, formando uma formilglicina (FGly), numa reação catalisada pela enzima geradora de FGly (FGE) (Dierks *et al*, 1997).

Em células de mamíferos, a distribuição correta de enzimas lisossômicas depende da presença de resíduos de manose-6-fosfato (M6P) em suas cadeias laterais de oligossacarídeos. Enzimas lisossômicas adquirem estes resíduos no aparelho de Golgi, onde são subsequentemente reconhecidas por uma fosfotransferase que transfere N-acetilglicosamina 1-fosfato aos resíduos de manose das cadeias laterais. Num segundo passo uma alfa-N-acetilglicosaminidase gera resíduos de manose-6-fosfato pela reação de resíduos de N-acetilglicosamina. O reconhecimento da manose-6-fosfato é essencial para a ligação aos receptores de manose-6-fosfato (M6PR) que medeiam à distribuição de enzimas recentemente sintetizadas para o lisossomo.

A ARSA é sintetizada a partir de um precursor polipeptídico de 507 aminoácidos (~52 KDa). Depois de sua produção no retículo endoplasmático (RE), um peptídeo sinal de 18 aminoácidos é clivado e removido, resultando em uma proteína madura de 489 aminoácidos (~51KDa) (Stein *et al*, 1989).

Ainda no RE, a ARSA é N-glicosilada; sofre conversão da sua cisteína crítica para formilglicina e forma dímeros. Os dímeros são transportados para o complexo de Golgi onde recebem marcadores de M6P e são reconhecidos por M6PR. Os complexos ARSA-M6PR são incorporados em endossomos onde se dissociam. Os M6PR retornam, então, às cisternas do complexo de Golgi e ARSA é entregue aos lisossomos. Dentro dos lisossomos os dímeros de ARSA oligomerizam-se em octâmeros de uma maneira dependente do pH (Sommerlade *et al*, 1994) [Fig1].

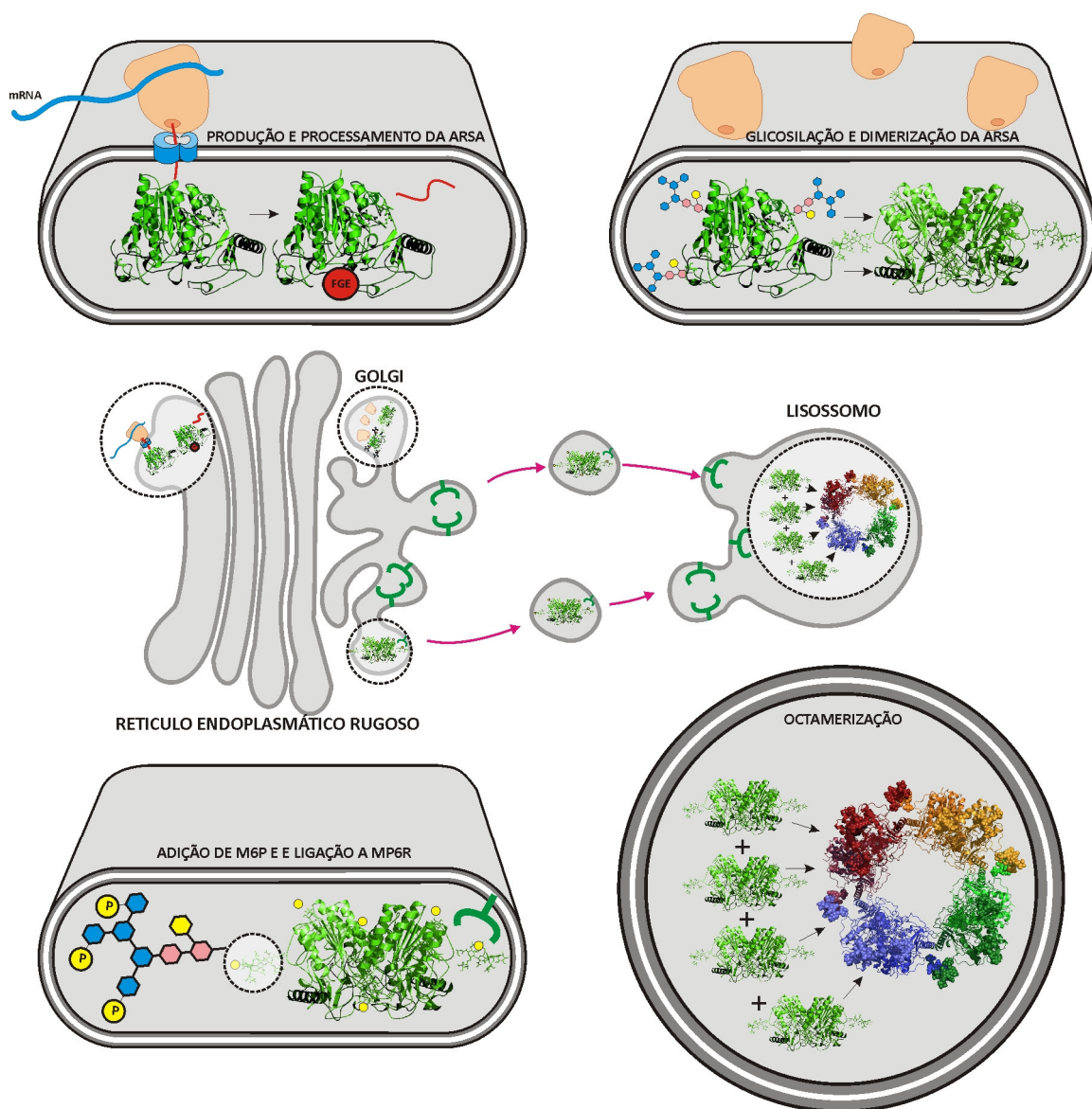


Figura 1. Rota metabólica da ARSA.

Essa dinâmica de associação/dissociação dependente do pH sugere a regulação a partir dos estados de protonação/desprotonação de Glu424, que é o único aminoácido de cadeia lateral ácida localizado na interface octamérica. Em pH~5, Glu424 torna-se protonado e estabiliza o octâmero devido a uma ligação de hidrogênio intermolecular com a Phe398 de um dímero adjacente. Já em pH~7, Glu424 é desprotonado e a ligação intermolecular com Phe398 não é mais possível. Dessa forma, a conformação preferida de Glu424 é agora aquela que permite uma ligação intramolecular com Gln460 (Vagedes *et al*, 2002).

ARSA é N-glicosilada *in vivo* em cada um dos seus três sítios de glicosilação (Asn158, Asn184 e Asn350) onde carrega oligossacarídeos ricos em manose com núcleos fucosilados. A fosforilação é restrita aos oligossacarídeos dos primeiro e terceiro sítios, uma vez que a presença de determinados oligossacarídeos fosforilados pode interferir na fosforilação de outros sítios (Laidler *et al*, 1997).

A estrutura secundária da ARSA é do tipo α/β , com 26% α -hélices e 3_{10} hélices, 16% de folhas β e 46% de voltas β ou 3_{10} . O núcleo enzimático é altamente conservado entre as sulfatases e consiste de duas folhas β pregueadas: a folha β menor é formada por quatro fitas β ($\beta 15$ a $\beta 18$) e a folha β maior é composta de dez fitas β ($\beta 8$ a $\beta 12$). A folha β maior é envolvida pelas hélices αA , αD , αE de um lado e αB , αC , αG e αH do outro lado. Na região C-terminal, um loop incluindo a longa hélice αI é formado por aproximadamente 80 aa e percorre todo o monômero (Lukatela *et al*, 1998)

A região de contato dimérico compreende a fita $\beta 13$ de um canto da folha β menor e quatro *loops*. A interface octamérica, por sua vez, é formada principalmente pela longa hélice αI , que é sustentada por dois segmentos da folha β menor: *loop* entre $\beta 15$ e $\beta 16$ e a fita $\beta 18$ (Lukatela *et al*, 1998).

Das cisteínas de ARSA, 12 formam pontes dissulfeto. Uma dessas pontes dissulfeto conecta as duas folhas β pregueadas, duas estabilizam um grampo formado entre as fitas $\beta 6$ e $\beta 7$ e outras três amarram a extremidade C-terminal de aproximadamente 20 aa, esse *cluster* de cisteínas é único de ARSA (Lukatela *et al*, 1998).

Outra característica estrutural funcionalmente importante da ARSA é a presença de uma ponte salina entre os resíduos Asp335 e Arg370. Asp335 está localizado no início de uma das hélices próxima à folha β centralmente localizada, enquanto Arg370 está localizada na folha β menor periférica. Apesar de não estarem estruturalmente envolvidas com o sítio catalítico, duas mutações que impedem a formação dessa ponte salina (p.D337V e p.R370) já foram descritas em pacientes com MLD grave e são responsáveis pela completa perda da atividade enzimática de ARSA *in vitro* (Schestag *et al*, 2002).

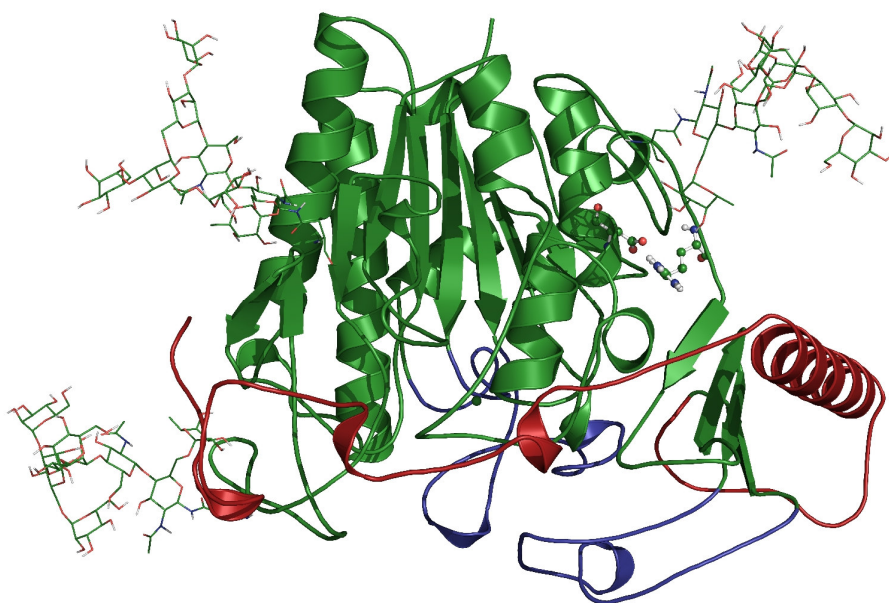


Figura 2. **Estrutura tridimensional do monômero de ARSA glicosilada em seus três sítios.** Em azul são mostrados os loops constituintes da interface dimérica de ARSA e em vermelho a hélice I, principal constituinte da interface octamérica, seguida de uma grande cauda C-terminal. Os resíduos Asp335 e Arg370, que participam da formação de uma ponte salina, estão representados como bolas e bastões coloridos por átomos e as cadeias laterais de oligossacarídeos ricos em manose com núcleo fucosilado (em Asn158, Asn184 e Asn350) são representadas como bastões coloridos por átomos.

A FGly essencial de ARSA está localizada em um bolsão positivamente carregado e atua como um ligante para um íon metálico Mg^{+2} que é octaedralmente coordenado. No mecanismo catalítico proposto para a clivagem do éster de sulfato, um dos oxigênios geminais do aldeído hidratado ataca o enxofre do éster de sulfato levando a uma transesterificação do grupo sulfato para o aldeído hidratado. Simultaneamente a porção álcool do substrato é liberada. Então, o sulfato é eliminado do intermediário enzima-sulfato por um rearranjo intramolecular produzido pelo segundo oxigênio geminal da FGly. Em ARSA um dos oxigênios do aldeído hidratado está próximo ao íon Mg^{+2} , que está complexado pelos resíduos Asp29, Asp30, Asp281 e Asn282, e o outro oxigênio está próximo ao anel imidazol de His125. Enquanto coordena o Mg^{+2} , o mesmo oxigênio ataca o grupamento sulfato no primeiro passo da clivagem dos sulfato. A His125 é candidata a desprotonar a segunda hidroxila do aldeído hidratado, o que está relacionado à liberação do sulfato e à regeneração do aldeído. Os resíduos positivamente carregados Lis123, Lis302, His229 e Ser150 fixam o grupo sulfato (Waldow *et al*, 1999).

O íon metálico Mg^{+2} presente no sítio ativo de ARSA acelera a catalise por cumprir pelo menos três diferentes funções: primeiramente, ele está envolvido na estabilização da orientação apropriada do gem-diol que é necessário para o ataque nucleofílico do grupo hidroxila ao substrato; secundariamente, ele serve na orientação apropriada do grupo sulfato e, finalmente, ele polariza a ligação $C\beta69-O\gamma2$, tornando $O\gamma2$ um nucleófilo melhor (Waldow *et al*, 1999)

Os resíduos do sítio ativo são extremamente conservados entre as sulfatases humanas, o que indica que sua estrutura, assim como o mecanismo catalítico, são também conservados (Gosh, 2007).

A ARSA é uma enzima constitutiva e pode ser purificada de uma variedade de tecidos, incluindo tecido hepático, placenta e urina. Várias formas moleculares de ARSA foram demonstradas em vários tecidos, o que pode ser atribuído à heterogeneidade de suas cadeias laterais de oligossacarídeos (von Figura *et al*, 2001). Além das variantes normais das cadeias laterais de oligossacarídeos existem associações de modificações anormais dessas cadeias como sialiação, fosforilação e sulfatação à manifestação de tumores (Nakamura *et al*, 1984).

Além da ARSA lisossômica, alguns autores têm sugerido também a participação de ARSA como um componente da superfície celular, onde ela existiria como uma proteína sem função enzimática, ou seja, continuaria sendo capaz de ligar sulfatídeos, mas não de hidrolisá-los, dado que o pH ótimo para atividade de ARSA é ácido, situado ao redor de 4,5. A importância do meio ácido para a ótima função de ARSA tem sido avaliada em termos de sua oligomerização, entretanto as possíveis diferenças estruturais entre a ARSA da superfície celular e a ARSA lisossômica que podem estar relacionadas às propriedades ionizantes de seus respectivos ambientes e com o desempenho da função enzimática ainda são desconhecidas (Mitsunaga-Nakatsubo *et al*, 2009)

1.3. ASPECTOS GENÉTICOS DA MLD

A maioria dos casos de MLD é causada por mutações alélicas no gene da ARSA, exceto nos casos de deficiência de SAP-B, uma condição não alélica relacionada à manifestação de sintomas comuns à MLD juvenil. Outra condição não alélica associada à manifestação de alguns sintomas comuns à MLD é denominada deficiência de múltiplas sulfatases (MSD), mas essa doença tem implicações clínicas gerais diferentes da MLD

(Barth *et al*, 1994). A MSD é causada pela deficiência combinada da atividade de uma série de sulfatases (ARSA, arilsulfatase B, arilsulfatase C, iduronato sulfatase, entre outras sulfatases envolvidas na degradação de mucopolissacarídeos), em decorrência de mutações no gene SUMF1, que codifica o fator modificador 1 de sulfatases, envolvido na modificação pós-traducional do resíduo de cisteína em FGly no sítio ativo destas sulfatases (Cosma *et al*, 2003).

O gene da ARSA foi mapeado em 22q13.3 e é dividido em oito exons, cobrindo aproximadamente de 3,2 Kb de DNA genômico. A região promotora está localizada cerca de 400 pb à montante do códon de iniciação e não possui caixas TATA ou CAAT, mas apresenta um alto conteúdo de GC, o que caracteriza um promotor de ação constitutiva (Kreysing *et al*, 1990).

Na região promotora do gene da ARSA, quatro potenciais sítios de iniciação da transcrição foram identificados e localizados entre 367 pb e 387 pb à montante do códon de iniciação da tradução. A posição em que as caixas GC estão dispostas (-10, -40, -50, -60 bases) é típica da interação com o fator Sp1. O gene da ARSA conta também com um sinal de poliadenilação que está distante 95 nucleotídeos à jusante do códon de terminação da tradução (Kreysing *et al*, 1990).

Cerca de 154 diferentes mutações já foram caracterizadas para o gene ARSA, sendo que 148 delas estão associadas à manifestação de MLD (HGMD). Essas mutações são, em sua maioria, substituições de ponto ou pequenas deleções e inserções, que podem alterar as propriedades bioquímicas da ARSA, seja causando retenção no RE, seja impedindo sua biossíntese ou dispersão intracelular, ou, ainda, promovendo sua instabilidade intralisossômica (Kappler *et al*, 1994).

Tabela 2. Classificação das mutações no gene da ARSA*.

Tipo de Mutação	Número de mutações
<i>Missense/nonsense</i>	115
<i>Splicing</i>	11
Regulatória	1
Pequenas deleções	15
Pequenas inserções	7
Pequenas deleções/inserções	2
Deleções grosseiras	1
Inserções grosseiras	1
Rearranjos complexos	1
Repetições variáveis	0
Total	154

*Adaptado de Human Gene Mutation Database (HGMD) em 09/2010

Três alelos diferentes relacionados à MLD ocorrem com alta frequência na maior parte das populações estudadas. O primeiro caracteriza-se uma mutação no sítio doador de *splicing*, na junção exon 2/intron 2, onde ocorre uma transição G>A, que destrói completamente a função enzimática, sendo identificado como c.459+1G>A (denominado antigamente alelo I) (Polten *et al*, 1991; Barth *et al*, 1993). Interessantemente, entre pacientes japoneses a mutação mais frequente é a p.G99D com uma prevalência de 45% (Kurosawa *et al*, 1998).

O segundo alelo mais frequente é caracterizado por uma transição C>T no exon 8, levando a uma substituição do aminoácido prolina por leucina na posição p.426, sendo identificado como p.P426L (denominado antigamente alelo A) (Polten *et al*, 1991; Barth *et al*, 1993). O aminoácido prolina na posição p.426, juntamente com Pro425 e Gln460, está envolvido na regulação da associação dímero/octâmero de ARSA, e é de fundamental importância para a atividade da ARSA nos lisossomos (Vagedes *et al*, 2002).

Na maioria das populações estudadas a prevalência descrita para a mutação c.459+1G>A varia entre 15% e 43% e para a mutação p.P426L entre 16% e 25% (Barth *et al*, 1993).

Tabela 3. Distribuição das mutações c.459+1G>A e p.P426L em diferentes países europeus.

País	Frequência alélica relativa (%)							
	MLD infantil		MLD juvenil		MLD adulta		Total em cada país	
	c.459+1G>A	p.P426L	c.459+1G>A	p.P426L	c.459+1G>A	p.P426L	c.459+1G>A	p.P426L
Itália	36	1,6	-	-	-	-	36	1,6
Áustria	37,5	2,5	20	50	12,5	62,5	37,5	42
Bélgica	40	-	25	25	-	-	37,5	4
República Tcheca e Eslováquia	70	-	25	25	17	17	38	11,5
França	-	-	-	-	-	-	14	14
Alemanha	46	-	16	38	9	57	24	30
Inglaterra	67	-	36	14	25	25	18	8
Holanda	-	25	15	37	-	33	8	35
Grécia	-	-	-	-	-	-	-	-
Polônia	26	-	34	15	16	50	76	17
Portugal	65	-	-	50	50	-	59	2,9
Espanha	24	-	25	-	20	40	23	6
Suécia	29	-	-	-	-	100	27	8
Suíça	33	-	50	-	-	-	33	-
Ucrânia	33	-	-	17	-	-	10	10
Total em cada subtipo clínico (%)	40	0,9	16	30	9	42,5	25	18,6

*Adaptado de Lugowska *et al*, 2005.

O terceiro alelo mais comum entre europeus é caracterizado por uma transversão T>G no exon 3, que resulta na substituição de um aminoácido isoleucina na posição p.179 por serina, sendo denominado p.I179S, e representa cerca de 12% do total de alelos mutantes que causam MLD nessa população. Esta mutação está associada a uma redução substancial, mas não à abolição da atividade enzimática, o que deriva da menor estabilidade da enzima mutante *in vitro* (Fluharty *et al*, 1991; Berger *et al*, 1997).

As demais mutações no gene da ARSA foram descritas com alta frequência entre pacientes de grupos étnicos isolados, um fato que pode ser explicado como resultante de um efeito do fundador ou, em alguns casos, de vantagem seletiva do heterozigoto (Heinisch *et al*, 1995). São exemplos, a alta frequência da mutação c.848-1G>A entre índios Navajo do sudoeste dos EUA e entre esquimós do Alaska (Draghia *et al*, 1997), a mutação p.T274M encontrada entre pacientes de origem libanesa da Austrália (Harvey *et al*, 1993) e entre cristão árabes da Galiléia (Heinisch *et al*, 1995) e a mutação p.L377P que foi encontrada com alta frequência entre judeus habanitas e yenemitas (Zlotogora *et al*, 1994).

Para muitos alelos que causam MLD, o efeito da mutação foi analisado em detalhes, o que permitiu a distinção em dois grupos específicos: o primeiro grupo compreende os alelos que resultam na supressão de toda a atividade enzimática de ARSA (alelos nulos) e o segundo grupo contém os alelos que resultam em alguma atividade enzimática residual de ARSA (alelos não nulos) (von Figura *et al*, 2001). Indivíduos homozigotos para alelos nulos (dos quais c.459+1G>A é um exemplo) comumente apresentam as formas mais severas de MLD, enquanto indivíduos homozigotos para alelos não nulos (p.P426L, por exemplo) comumente apresentam a forma adulta da doença (Polten *et al*, 1991; Barth *et al*, 1993).

A associação genótipo-fenótipo demonstra que a relação entre atividade enzimática residual e o genótipo pode ser determinante para o curso clínico da doença, entretanto o valor preditivo desta associação não é preciso, uma vez que em algumas populações estudadas não se observou correlação alguma entre atividade enzimática residual e gravidade ou idade de início da doença, ao mesmo tempo em que, por exemplo, foram descritos indivíduos heterozigotos para p.P426L apresentando a forma infantil de MLD, o que pode ser explicado pela ocorrência de uma segunda mutação que age abolindo completamente a atividade enzimática (Kappler *et al*, 1994).

Outros estudos mostram que a associação genótipo-fenótipo pode indicar os sintomas neurológicos e psiquiátricos predominantes durante o curso da MLD. Pacientes portando o alelo p.I179S demonstraram manifestação de sintomas psiquiátricos prevalentes no início da doença, ao passo que indivíduos portadores da mutação p.P426L apresentaram predominantemente sintomas neurológicos neste mesmo estágio da MLD (Tylki-Szymanska *et al*, 1996).

1.4. PSEUDODEFICIÊNCIA DE ARSA E OUTROS POLIMORFISMOS NÃO PATOGÊNICOS DO GENE DA ARSA

Eventualmente, a deficiência de ARSA é diagnosticada em indivíduos aparentemente saudáveis, uma condição conhecida como pseudodeficiência de ARSA (PD-ARSA) e sua prevalência geral é de cerca de 7% a 12% (Nelson *et al*, 1991). Na pseudodeficiência de ARSA, os indivíduos apresentam entre 5% e 15% de atividade enzimática normal de ARSA, o que é suficiente para a manutenção do metabolismo de sulfatídeos. A maioria dos indivíduos com esta pseudodeficiência são homocigotos para um determinado alelo composto de 2 polimorfismos combinados no gene da ARSA (alelo PD-ARSA) que juntos promovem a redução da atividade enzimática (Barth *et al*, 1993; Harvey *et al*, 1998). No entanto, indivíduos pseudodeficientes heterocigotos compostos para um alelo PD-ARSA e um alelo MLD também podem ocorrer (Gieselmann, 1991).

O alelo normal e o PD-ARSA diferem por duas transições A>G em desequilíbrio de ligação: A primeira delas leva a uma substituição de serina por asparagina na posição p.350 (p.N350S), o que resulta na perda de uma das cadeias laterais de oligossacarídeos da ARSA e é responsável pela diminuição de 2 KDa a 4 KDa em relação à ARSA produzida por indivíduos normais. A segunda transição altera o sinal de poliadenilação (c.1524+1A>G) do gene da ARSA, sendo responsável por uma redução na estabilidade do mRNA produzido e, conseqüentemente, por uma redução na síntese de ARSA ativa (von Figura *et al*, 2001).

A frequência de alelo PD-ARSA entre indivíduos brasileiros saudáveis é estimada em 7,9%, tendo o polimorfismo p.N350S prevalência de 20,7%, ao passo que o polimorfismo 1524+1A>G apresenta prevalência menor de 7,9% na nossa população (Pedron *et al*, 1999).

Historicamente, a redução da atividade de ARSA em indivíduos pseudodeficientes tem sido atribuída ao polimorfismo de alteração do sinal para poliadenilação, entretanto os recentes achados de indivíduos pseudodeficientes sem esse polimorfismo e portando apenas o polimorfismo de N-glicosilação têm apontado para um papel mais significativo do polimorfismo p.N350S no desenvolvimento dessa condição (Barth *et al*, 1994; Shen *et al*, 1993). Além disso, o polimorfismo p.N350S tem sido associado ao alcoolismo e algumas desordens psiquiátricas como esquizofrenia e transtorno bipolar, o que aponta ainda mais para sua natureza patogênica (Park *et al*, 1996). Entretanto, tanto a função biológica da tripla glicosilação de ARSA, quanto as alterações estruturais advindas da perda de glicosilação em p.N350S permanecem desconhecidas, em parte pela dificuldade metodológica associada ao estudo desses aspectos em glicoproteínas.

Além dos polimorfismos de PD-ARSA já foram descritos quatro outros polimorfismos codificantes não patogênicos no gene da ARSA. Entre eles, dois - p.W193C e p.T931S - apresentam desequilíbrio de ligação com a mutação c.459+1G>A, que, como visto anteriormente, é a mais frequente mutação relacionada à MLD na maioria da populações estudadas até o presente momento (Gort *et al*, 1999, Zlotogora *et al*, 1994).

Os outros dois polimorfismos codificantes não patogênicos no gene da ARSA são p.L76P e p.R496H, estando este último em desequilíbrio de ligação com o polimorfismo p.N350S, constituinte do alelo de PD-ARSA (Berger *et al*, 1996; Ricketts *et al*, 1998).

2. JUSTIFICATIVA

Desde a caracterização da estrutura do gene da ARSA, vários grupos têm analisado pacientes com MLD e caracterizado a associação entre essa doença e as várias mutações no gene da ARSA, o que tem ajudado na elucidação da natureza dos diversos domínios deste gene e do seu produto.

No Brasil, dados de caracterização epidemiológica e clínica foram recentemente publicados pelo nosso grupo (Artigálas *et al*, 2010). Entretanto, ainda não existem dados de caracterização genotípica de pacientes brasileiros com MLD. A caracterização molecular dos pacientes brasileiros com MLD é importante pelo desconhecimento atual da natureza da heterogeneidade genotípica desses pacientes e pode contribuir para a fixação de correlações fenótipo-genótipo em MLD.

Do mesmo modo, dados sobre aspectos estruturais básicos da ARSA são escassos. A compreensão da contribuição de fatores extrínsecos como o pH, fundamental para a atividade enzimática da ARSA e intrínsecos, como a glicosilação, para a manutenção de características estruturais funcionais da ARSA é de suma importância para a elucidação da patologia molecular da MLD e, conseqüentemente, para o melhor desenvolvimento de abordagens terapêuticas de sucesso.

3. OBJETIVOS

3.1. OBJETIVO GERAL

Identificar mutações no gene da ARSA em pacientes brasileiros com MLD e analisar aspectos da dinâmica molecular da ARSA humana em resposta a fatores como mudança de pH e glicosilação.

3.2. OBJETIVOS ESPECÍFICOS

- Identificar a mutação c.459+1G>A e os polimorfismos p.N350S e c.1524+1A>G no gene da ARSA em pacientes com baixa atividade *in vitro* de ARSA;
- Identificar alterações raras em toda a região codificante do gene da ARSA nos pacientes da amostra;
- Determinar a frequência das mutações encontradas na amostra em estudo e compará-las aos dados da literatura;
- Correlacionar o genótipo encontrado com o fenótipo do paciente, quando possível;
- Analisar a DM da ARSA monomérica não glicosilada em resposta a diferentes pH (em pH~5 e em pH~7);
- Analisar a DM da ARSA em resposta a diferentes estados de glicosilação (ARSA completamente glicosilada e ARSA deficientemente glicosilada N350S).

4. ARTIGO I

GENOTYPIC CHARACTERIZATION OF BRAZILIAN METACHROMATIC LEUKODYSTROPHY PATIENTS

(Artigo a ser submetido ao periódico *Molecular Genetics and Metabolism*)

GENOTYPIC CHARACTERIZATION OF BRAZILIAN METACHROMATIC LEUKODYSTROPHY PATIENTS

M.Y.F. Virgens^{1,2,3}, H. Bock^{1,3,4}, M. Siebert^{1,3}, M.G. Burin³, R. Giugliani^{1,2,3}, M.L. Saraiva-Pereira^{1,2,3,4}.

¹Laboratório de Identificação Genética, Centro de Pesquisas, Hospital de Clínicas de Porto Alegre, R. Ramiro Barcelos, 2350, Porto Alegre, 90035-903, RS, Brazil.

²Departamento de Genética, Universidade Federal do Rio Grande do Sul, Av. Bento Gonçalves, 9500, Porto Alegre, 91500-970, RS, Brazil.

³Serviço de Genética Médica – Hospital de Clínicas de Porto Alegre, R. Ramiro Barcelos, 2350, Porto Alegre, 90035-903, RS, Brazil.

⁴Departamento de Bioquímica, Universidade Federal do Rio Grande do Sul, Ramiro Barcelos, 2600, Porto Alegre, 90035-003, RS, Brazil.

***Corresponding author:**

Maria Luiza Saraiva-Pereira, PhD

Serviço de Genética Médica – Hospital de Clínicas de Porto Alegre

Ramiro Barcelos, 2350 – CP 90035-903

Porto Alegre, Rio Grande do Sul, Brazil

E-mail address: mlpereira@hcpa.ufrgs.br

Phone: +55 51 33598011

Fax: +55 51 33598010

ABSTRACT

Metachromatic leukodystrophy (MLD) is an inborn error of metabolism inherited as an autosomal recessive trait. MLD is a neurodegenerative condition subdivided in three types according to the age at onset of symptoms: infantile-MLD, juvenile-MLD and adult-MLD. The major biochemical defect associated with MLD is the deficiency of arylsulfatase A, a lysosomal glycoprotein that catalyzes sulfated glycolipids constituents of myelin sheet. So far, around 140 MLD-associated mutations have been identified in ARSA gene, among which 2 occur with high frequency in the majority of studied populations: c.459+1G>A and p.P426L. Besides, 7-12% of healthy people have *in vitro* deficiency of ARSA, a condition called ARSA pseudodeficiency (ARSA-PD). Despite the recent effort to a clinical and epidemiological characterization, there is no genotypic characterization of MLD Brazilian patients. In this context, the aim of this work was to identify mutant alleles in the ARSA gene from MLD Brazilian patients. The sample was constituted by 27 MLD patients. The most frequent MLD-associated ARSA mutation (c.459+1G>A) and the ARSA-PD polymorphisms were detected by real-time PCR, and remaining mutations were detected by direct sequencing of ARSA gene. In this study, the most frequent mutation was c.459+1G>A (0.50), as expected by the high prevalence of infantile-MLD cases. Interestingly, the second most frequent MLD-associated mutations in our cohort were p.P426L (0.08) and c.103_110del8 (0.08). In total, were identified 11 different MLD-associated uncommon mutations, including 2 novel putative disease-causing: p.S44P and p.P284S. In addition, 2 polymorphisms frequently associated with c.459+1G>A: p.W193C (0.54) and p.T391S (0.65) and the 2 polymorphisms constituents of ARSA-PD allele: p.N350S (0.15) and c.1524+95A>G (0.04) were identified.

Key-words: Metachromatic leukodystrophy, arylsulfatase A, pseudodeficiency of ARSA.

1. INTRODUCTION

Metachromatic leukodystrophy is an autosomal recessive disease with lysosomal involvement due to arylsulfatase A (ARSA) deficiency. This enzyme catalyzes the first step in the degradation of sulfated glycolipids, essential components of myelin sheath. The storage of these sulfated glycolipids can be mainly observed in the white matter of the central nervous system and in peripheral nerves, leading to progressive demyelination, and producing severe neurological consequences [1].

MLD is a rare disorder with an overall birth incidence of 1.4–1.8 per 100,000 newborns [2]. The disease is classified into three clinical types according to the age at onset of clinical manifestations: infantile-MLD, juvenile-MLD and adult-MLD. Infantile and juvenile MLD cases show more severe phenotype (psychomotor regression, ataxia, progressive spastic quadriplegia, dysphagia, blindness, deafness, seizures, diffuse muscular atrophy and weakness with areflexia) [2,3], while in adult MLD cases behavioral alterations are the most significant finding, leading to misdiagnoses of attention deficit disorder, schizophrenia or other psychosis [4].

The ARSA gene is located on *locus* 22q13, it covers 3.2 kb of genomic DNA and it is organized in 8 exons [5]. So far, around 140 MLD causing mutations have been identified in ARSA gene. Three of those mutations are frequent in European patients: (i) c.459+1A>G, a splice donor site mutation of the exon 2 / intron 2 border, which was reported as present in 15% to 43% of mutant alleles, (ii) p.P426L, a missense mutation that disturbs ARSA oligomerization, with an allelic prevalence of 16% to 25%, and (iii) p.I179S, with an allele prevalence of 12-13% [6]. Concerning non-Europeans patients, p.G199A is the most frequent alteration reported, being present in 45% of alleles with Japanese ancestry [7].

MLD-associated alleles can be placed into two groups, according to residual enzyme activity: (i) null-alleles, that do not result in functional enzyme, being represented by c.459+1A>G, and (ii) non-null or residual alleles, that show residual activity as those carrying p.P426L mutation [8]. Combination of alleles is responsible for different clinical manifestations that partially explain different forms of the disease, as previously reported extensively [9].

Low ARSA activity can also be detected in about 7-12% of healthy population where residual ARSA activity is reported to be about 10% of the normal range. This non-pathogenic condition is known as pseudodeficiency of ARSA (ARSA-PD, and is caused by the occurrence of ARSA-PD allele. This allele is characterized by two A>G transitions where one of them abolishes one N-glycosylation site in exon 6 (p.N350S), and the other leads to the loss of the polyadenylation signal (c.1524+96A>G) [10].

Following ARSA gene identification, several studies were performed in order to describe mutations and polymorphisms in this gene in different populations [6,8,11,12]. However, there is no report on MLD-associated mutations in South American MLD patients up to date. Our group has previously carried out a study to define frequency of ARSA-PD allele in normal individuals [13].

Therefore, the overall purpose of this study was to describe the genotypic spectrum of Brazilian MLD patients.

2. MATERIAL AND METHODS

2.1. Patients

Twenty seven unrelated MLD patients from different regions of Brazil were included in our sample population: 9 patients from South region, 14 patients from Southeast, 1 patient from Northeast, and 1 patient from Northern Brazil – just 2 patients could not be allocated. The gender distribution was as follows: 19 (70.3%) males and 8 (29.7%) females. The age at onset of clinical manifestations varied from 1 to 141 months. The inclusion criteria were low ARSA activity (<5 nmol/h/mg) and high levels of sulfatide in urine. This study was approved by the Research Ethics Committee of Hospital de Clínicas de Porto Alegre (HCPA).

2.2. Sample collection and DNA isolation

Blood samples (5 mL) from each patient were collected in EDTA. Genomic DNA was isolated from peripheral blood leukocytes, as described previously [14], and kept at -20°C until further analysis. DNA quantitation was carried out using a fluorescence-based assay (Quant-It™ ds DNA BR Assay Kit – Invitrogen) in the Qubit™ fluorometer (Invitrogen).

2.3. Screening of c.459+1G>A, p.N350S and c.1524+95A>G mutations in the ARSA gene

The mutation c.459+1G>A and the sequence alterations that characterize ARSA-PD allele (p.N350S and c.1524+95A>G) were screened by real time PCR qualitative assay using TaqMan® (Applied Biosystems) system that is based on fluorescent minor groove binding probes. Primers and probes were designed using PrimerExpress® software version 3.0 (Applied Biosystems) [Table 1]. PCR reaction was performed in a final volume of 8 µL containing 2 ng of genomic DNA, 0.3 µL of specific TaqMan® assay and 4 µL of 2x PCR

Genotyping Master Mix (Applied Biosystems). Amplification included an initial step at 50°C for 2 minutes (activation of AmpErase UNG function), AmpliTaq® Gold activation at 95°C for 10 minutes, followed by 40 cycles of denaturation at 95°C for 15 seconds and annealing extension at 60°C for 1 minute. Allelic discrimination step was performed at 60°C for 1 minute. PCR products were analyzed by the Sequence Detection System Software version 1.2.1 (Applied Biosystems) in the ABI Prism® 7500 Sequence Detector System (Applied Biosystems) through allelic discrimination plot.

2.4. PCR amplification

ARSA gene was divided into 4 different amplicons [Fig. 1]. Briefly, each PCR reaction was performed in a total volume of 25 µL, containing 200 ng of genomic DNA, 200 µM of each dNTP, 0.2 µM of each primer (forward and reverse), 20 mM of Tris (pH 8.4), 50 mM of KCl, 2.5 mM of MgCl₂ and 1.25 U of *Taq* DNA polymerase (Invitrogen). Amplification protocol was composed of initial denaturation at 94°C for 10 minutes, 30 cycles of denaturation at 94°C for 30 seconds, annealing at 57-63°C (according to each amplicon) for 30 seconds and extension at 72°C for 60 seconds, followed by final extension at 72°C for 10 minutes. Amplification was confirmed by running an aliquot of PCR products in a 1.5% (w/v) agarose gel and the produced amplicons were visualizing under UV light. Specific annealing temperature and primers sequences are available on Table 2.

2.5. Direct DNA sequencing

Prior to sequencing reaction, amplicons were purified using 2.5 U of Exonuclease I (USB) and 0.25 U of Shrimp Alkaline Phosphatase (USB). Direct DNA sequencing was performed using BigDye Terminator Cycle Sequencing kit version 3.1 (Applied Biosystems) following the manufacturer's instructions. Fragments were then separated by

electrophoresis in the ABI 3130*xl* genetic analyzer (Applied Biosystems) and sequences were analyzed by DNA Sequencing Analysis Software version 5.2 (Applied Biosystems). Sequence alterations were confirmed by analyzing an independent DNA sample using forward and reverse primers.

2.6. Computational comparative evaluation of novel mutations

Novel mutations detected in ARSA coding region were evaluated using PolyPhen-2 program server [15]. PolyPhen-2 predicts the possible impact of a non-synonymous mutation on the structure and function of a human protein using straightforward phylogenetic and structural comparative considerations (eight sequence-based and three structure-based features). The program classifies missense mutations as benign, possibly damaging or probably damaging, and a score 0-1.0 is attributed, with the most probably damaging mutations reaching 1.0.

3. RESULTS

3.1. Clinical characterization.

Twenty seven patients analyzed were grouped in MLD subtypes according to the age of onset of clinical manifestations: The majority of the patients were classified as presenting the infantile form of MLD (81% - 22 out of 27 patients), while a minor portion was classified as presenting the juvenile form (18% - 5 out of 27 patients). No adult MLD patient was included in our cohort.

3.2. Molecular characterization.

As 6 out of 27 patients (22%) were born from consanguineous marriages, our cohort was composed by 48 unrelated alleles. Distribution of sequence alterations, including polymorphisms and mutations, can be seen in Table 3. Most MLD-associated alleles were single changes, but we have also defined two complex alleles: [c.459+1G>A + p.D335V] and [c.103_110del8 + g.1190_1191insC][Table 3].

The most frequent MLD-associated mutation was c.459+1G>A with an allelic frequency of 0.50 (24 of 48 screened alleles). p.P426L and c.103_110del8 mutations were defined as second most frequent MLD-associated mutations with an allelic frequency of 0.08 each (4 of 48 mutant alleles) [Fig. 2]. The third most frequent MLD-associated mutation was g.1190_1191insC with an allelic frequency of 0.06 (3 of 48 mutant alleles)

The remaining alleles were rare and found only once with an allelic frequency of 0.02 each [Fig. 2]. In total, 11 different rare MLD-associated mutations were found in this cohort with an overall allelic frequency of 0.27 (13 of 48 mutant alleles), including 2 novel putative disease-causing mutations (p.S44P and p.P284S) [Fig. 3].

As expected, the most frequent genotype defined was homozygous patient for c.459+1G>A alleles (9 out 27 patients), followed by heterozygous patient for one

c.459+1G>A allele (7 out of 27 patients). The complete genotype distribution is shown in Table 3.

Besides this spectrum of MLD-associated mutations, in this study was evaluated the presence of non-pathogenic coding polymorphisms in ARSA gene, including the 2 polymorphisms constituents of the ARSA-PD allele [Fig. 2]. Five of six non-pathogenic polymorphisms already described in ARSA gene were found in this work. The p.W193C and p.T391S polymorphisms were more frequent, with allelic frequencies of 0.54 (26 of 48 screened alleles) and 0.65 (31 of 48 screened alleles), respectively. The non-pathogenic coding polymorphism p.R496H and the already described synonymous alteration c.453C>T (p.H151H) [16,17] presented isolated occurrences each one. The sixty non-pathogenic coding polymorphism in ARSA gene, p.L76P, has not been found in this cohort.

Concerning the polymorphisms ARSA-PD allele constituents, the p.N350S polymorphism was founded with allelic frequency of 0.15 (7 of 48 screened alleles) and the c.1524+95A>G polymorphism was founded with allelic frequency of 0.04 (2 of 48 screened alleles). The occurrences of the c.1524+95A>G polymorphism were always associated with p.N350S polymorphism, composing three ARSA-PD alleles allocate into one ARSA-PD homozygous and one ARSA-PD heterozygous patients.

4. DISCUSSION

The prevalence of infantile-MLD cases in our cohort (81%) was higher than that currently estimated for this MLD subtype among Europeans, located around 48% [18]. This finding indicates that other MLD subtypes are probably being under-diagnosed as suggested by a recent epidemiological study of MLD carried out by our group [19]. This fact can be related to the less aggressive symptoms presented by late onset forms of the disease, when compared to infantile form, and the clinical picture is in general atypical, with predominance of psychiatric symptoms and absence of motor alterations. In general, infantile-MLD is the more common MLD subtype found in the most of published studies.

In our cohort the c.459+1G>A allele alone was able found in 41% of infantile-MLD cases (9 of 22 patients), and account for approximately 56% of the MLD-associated alleles in this subtype (22 of 39 alleles). These findings are in agreement with the more recent overall prevalence estimated for this allele among infantile-MLD patients from European countries that was reported to be around 40% [18]. This strong association results from the large decrease in ARSA activity caused by the c.459+1G>A allele [20].

Among the 5 juvenile-MLD patients in this study, 2 were heterozygous for c.459+1G>A allele, covering 22% of MLD-associated alleles in this subtype (2 of 9 alleles), which is similar to the 16% estimated for European countries [18]. So, our dates corroborate the involvement of c.459+1G>A allele in juvenile MLD cases as a more uncommon event when compared with its occurrence in infantile MLD subtype.

The p.P426L allele was responsible for 60% of juvenile MLD cases (3 of 5 patients), representing 33% of the MLD-associated alleles in this subtype (3 of 9 alleles). The p.P426L allele is more implicated in the adult form of MLD, reaching an overall prevalence around 42% in this subtype in European countries, but it is also found in around 30% of the juvenile-MLD alleles found among Europeans MLD patients [18]. In general, it

is possible to affirm that there is a more conclusive involvement of p.P426L in later forms of MLD [20]. This more attenuated phenotypic effect of p.P426L is due to the fact that the mutation only denotes a decrease in stability of ARSA within the lysosomes, but both the trafficking to this organelles or its catalytic activity remain normal [21].

In addition to p.P426L, the rank of second most frequent MLD-associated mutation was also occupied by c.103_110del8. This mutation is located in exon 1, 16 bp after the signal peptide, and deletes three residues that are highly conserved among sulfatases, leading to an early termination codon 105 bp downstream [22]. The c.103_110del8 mutation was first described in one French infantile-MLD patient [22], but was also found in one infantile-MLD patient from Spain [23]. All of our patients that presented c.103_110del8 mutation also manifested the infantile form of MLD.

As above reported, the third most frequent MLD-associated mutation in our cohort was g.1190_1191insC instead of p.I179S, as in European countries. The g.1190_1191insC mutation occurs in exon 2 and leads to the appearance of a stop codon in exon 3. The g.1190_1191insC mutation was only described in one infantile-MLD patient from Portugal [24]. Our patients that carries g.1190_1191insC mutation are also an infantile-MLD cases.

The third most frequent MLD-associated mutation in European countries, p.I179S, is usually described as associated with the juvenile form of MLD [25]. Despite the p.I179S was not frequent in our patients, its only occurrence was also implicated in a juvenile-MLD case.

According to the previous genotype-phenotype correlations, homozygous patients for alleles which do not allow expression of any enzyme activity – null alleles, as c.459+1G>A –, usually suffer from the most severe form of disease, while homozygous patients for an allele that allows the expression of low amounts of enzyme activity – non-

null alleles, as p.P426L –, usually suffer of attenuated forms of MLD (mostly adult-MLD, but also juvenile-MLD) [8,10]. This fact is corroborated by our finding of one homozygous patient for the allele p.P426L that presented the juvenile form of MLD and presented the highest age at onset of manifestations in our cohort (around 12 years).

It has been described linkage disequilibrium between two coding polymorphisms of ARSA gene – p.W193C and p.T391S –, one non coding polymorphism located at the intron 7 – c.2213C>G –, and the most frequent MLD-associated mutation – c.459+1G>A [23, 26]. In the present work, it was possible to evaluate the association between p.W193C, p.T391S and c.459+1G>A. In our cohort, all 9 homozygous patients for the most frequent allele c.459+1G>A were also homozygous for each polymorphism p.W193C and p.T391S. In the same way, 7 heterozygous patients for one c.459+1G>A allele were equally heterozygous for one p.W193C allele, and the majority of these patients were also heterozygous for one p.T391S allele (4 of 7 patients). Among the 12 remaining patients, only one presented p.W193C in absence of c.459+1G>A, 6 patients presented p.T391S in absence of c.459+1G>A, and the remaining 5 patients presented none of these alterations.

In general, our results support the haplotype reported previously and reinforce the concept of a unique origin for all these mutations. However, further studies on the linkage phase are necessary for cases in which at least two of three these alterations are in heterozygosis. For cases in which at least two of three alterations are in homozygosis, the combination point to the linkage phase, despite the direct sequencing cannot detect it directly.

Concerning polymorphisms ARSA-PD allele constituents, the prevalence of ARSA-PD allele among the MLD alleles in our cohort (4.2% - 2 of 48 screened alleles) is similar of the range found in European countries (6%-30%) [27,28]. The occurrences of

c.1524+95A>G are frequently associated with p.N350S, while p.N350S is frequently found alone [10,29]. Similarly, the c.1524+95A>G was also associated with p.N350S in our results, but p.N350S was also found alone. These results are consistent with the hypothesis that the mutation at the N-glycosylation site first occurred in an ancestral allele before the emergence of modern *Homo sapiens* from Africa and, subsequently, the polyadenylation mutation occurred in this ancient allele with the N-glycosylation mutation [30].

The polymorphism p.R496H is described in linkage disequilibrium with the p.N350S allele, and in this study the single occurrence of this polymorphism was associated with p.N350S allele too. However, further studies about the linkage phase are also necessary [16].

Concerning the 11 less frequent mutation found in this study, the p.R84W mutation was previously described in juvenile-MLD patients from Canada, Spain, Italy, Poland and China [3,23,25,31,32]. This mutation is located in exon 2, which shows homology with other arylsulfatases of different origins and functions, being responsible for a residual ARSA activity of 6% [31], and is considered a non-null allele. In our study, the mutation p.R84W appears in a infantile-MLD patient together with another rare mutation p.R390W, also described in infantile-MLD patients from Canada, Poland, Italy and France [29,32,33,34].

The rare mutations g.2408DelC and p.G122C, as well as the others polymorphisms in ARSA gene, were found in two homozygous infantile-MLD patients from consanguineous marriages in this study, which suggests the identity by descent of these rare mutations. The mutation g.2408DelC was also described in one infantile-MLD patient from Portugal. This mutation is located in exon 6 and leads to the appearance of a stop

codon in exon 8, predicting the protein truncation of at about 84 amino acids before carboxi terminus [24]. The mutation p.G122C, in turn, was previous reported in one infantile-MLD patient from Poland. This mutation changes an uncharged neutral glycine with aliphatic chain to an uncharged neutral cysteine, which is potentially able to form an additional disulfide bond. Moreover, the residue Gly122 is a conserved amino acid between sulfatases, suggesting that it is essential for enzyme stability and/or activity [35]. In our study, the patient homozygous for the p.G122C allele also presented the infantile form of MLD.

In our study, one heterozygous compound patient for p.F247S and c.459+1G>A presented the infantile form of MLD. The mutation p.F247S had already been described in juvenile-MLD patients from Poland and Ukraine [34,36]. The Ukrainian patient presented the juvenile form of MLD, but he also carried the p.P426L allele, and his more attenuated phenotype may be explained by the presence of this non-null allele.

The mutation p.D335V, in turn, was first reported in one Arab infantile-MLD [37], and recently it was found in patients from Canada, Poland, Morocco and U.S. [32,38,39,40]. An apparent salt bridge is found between residues Asp335 and Arg370 and, despite this salt bridge does not contribute directly to the formation of the active site of ARSA, its conformational importance can be explained by the fact that, with the exception of iduronate sulfatase, an equivalent salt-bridge is conserved within the sulfatase family [38]. Besides, the p.D335V abolishes completely the ARSA activity, which explains its relation with the more severe cases of MLD, as illustrated in this study.

Another compound heterozygous patient for one allele p.R244H and one allele p.G325S presenting the infantile form of MLD was found. The mutation p.R244H was previously described in one infantile-MLD patient from Reunion Island (a French colony

in the African Indian coast). The codon p.244 is one of the most conserved among sulfatases, and it seems to be a hotspot for mutations [22]. The mutation p.G325S, in turn, was reported in a Canadian family in which two siblings had been diagnosed with juvenile-MLD [40]. The p.G325S mutation not only alters a conserved amino acid, but also changes the nucleotide at the -1 position relative to the canonical splice donor sequence of intron 5. Besides these two mutations, our patient also presented ARSA-PD allele. The coexistence of MLD and ARSA-PD polymorphisms may have serious consequences in the clinical picture of the disease [25] and may have contributed for the development of infantile-MLD by this patient.

The mutation c.848-1G>A was described with high frequency in Navajo Indian and Alaskan Eskimos with the infantile form of MLD [41,42]. This mutation eliminates the GT consensus sequence from the 5' end of intron 4, resulting in the skipping of exon 4 and in a frameshift that leads to the premature termination of ARSA synthesis. Interestingly, this mutation was found in one juvenile-MLD patient in our cohort, but this patient also presented the p.P426L mutation, which can explain his more attenuated phenotype.

The mutation p.D255H occurs in exon 4 of the ARSA gene and was first described in two Belgian patients with infantile-MLD. This mutation predicts the change of an aspartic acid to histidine at position p.255, which is a highly conserved spot among sulfatases, being characterized as a null allele. In this study, this mutation was also found in an infantile MLD patient [43].

Two putative disease-causing mutations were first reported in this work: p.S44P and p.P284S. The mutation p.S44P was found in one infantile-MLD patient that also presented one p.P426L allele. This mutation is the result of a transversion of T to C at position c.127, and leads to a serine to proline substitution at residue p.44 of ARSA. The other novel

mutation, p.P284S, was found in an infantile-MLD patient, that also presented one c.459+1G>A allele. This mutation results of a transversion of C to T at position c.851, and leads to a proline to serine substitution at residue p.284 of protein. Both of these mutations were predicted damaging by the Polyphen server.

This work constitutes the first systematic genotypic characterization of Brazilian MLD patients. Our findings lead to differences between the MLD allelic distribution in Brazilian patients and that one described for European patients concerning to most frequent MLD-associated mutations, represented by the mutations c.103_110del8 and g.1190_1191insC, instead of p.I179S, as in European patients. Besides, we have also described 2 novel putative disease-causing mutations (p.S44P and p.P284S), that add a contribution to the spectrum of know mutations in ARSA gene. Through this work, we hope to contribute for the better understanding of the genetic heterogeneity of MLD

ACKNOWLEDGEMENTS

We are grateful to patients and their families who kindly agreed to participate in this study. We also would like to thank Thais Santa Rita for her technical assistance. This work was supported by Brazilian Funding Agencies (CNPq, FAPERGS and FINEP-HCPA) and ONG Pela Vida). M.Y.V.F. and M.L.S.P. were supported by CNPq.

TABLES AND FIGURES

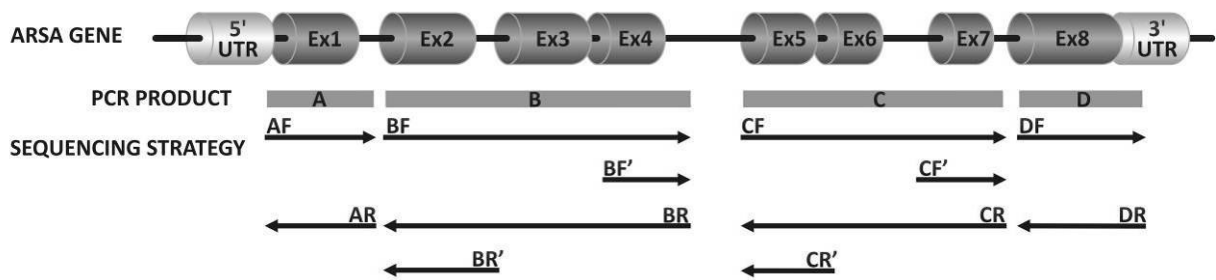


Figure 1. Scheme of PCR amplification and DNA sequencing of ARSA gene strategy.

	Mutation	cDNA	Allelic frequency
5' UTR	c.103_110del8	c.103_110del8	0.08
	p.S44P	c.127C>T	0.02
Ex1			
Ex2	p.R84W	c.250 C>T	0.02
	g.1190_1191insC	c.412 dup1	0.06
	p.G122C	c.364 G>T	0.02
	g.1231C>T	c.453 C>T	0.02
	c.459+1G>A	c.459+1G>A	0.50
Ex3	p.I179S	c.536 T>G	0.02
	p.W193C [†]	c.579 G>T [†]	0.54
Ex4	p.R244H	c.730 C>T	0.02
	p.F247S	c.740 T>C	0.02
	p.D255H	c.763 G>C	0.02
	c.849-1G>A	c.849-1G>A	0.02
Ex5	p.P284S	c.851C>T	0.02
	p.G325S	c.973 G>A	0.02
Ex6	p.D335V	c.1004A>T	0.02
	g.2409delC	c.1040delC	0.02
	p.N350S*	c.1049A>G*	0.15
Ex7	p.R390W	c.1168C>T	0.02
	p.T391S [†]	c.1172C>G [†]	0.65
Ex8	p.P426L	c.1277C>T	0.08
	p.R496H [†]	c.1487G>A [†]	0.02
3' UTR	c.1524+95A>G*	c.1524+95A>G*	0.04

Figure 2. Allelic frequencies and genetic distribution of identified mutations in ARSA gene. * PD polymorphisms, † Non-pathogenic polymorphisms, and bold letters represent novel mutations.

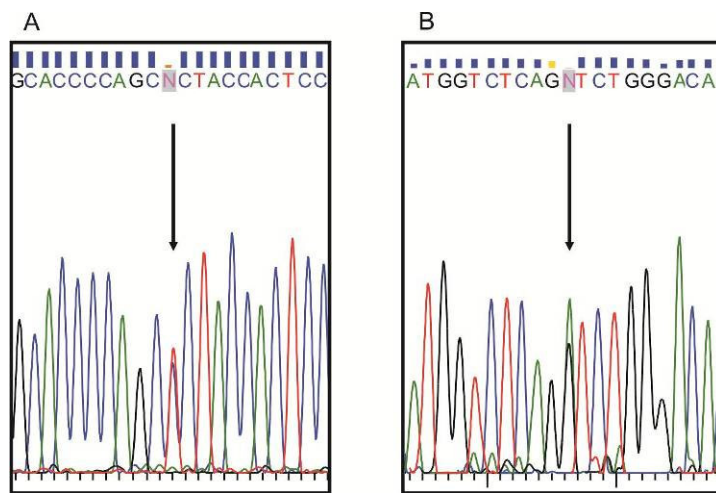


Figure 3. Novel putative disease-causing mutations in the ARSA gene. (A) The p.S44P mutation is caused by a transversion of T to C at position 127 of cDNA. **(B)** The p.P284S mutation is caused by a transversion of C to T at position 851 of cDNA.

Table 1. Primers and probes used to identify c.459+1G>A and ARSA-PD allele.

Mutation	Primer sequence (5' > 3')	Probe sequence (5' > 3')
c.459+1G>A	Forward CGATTTCTAGGCATCCCGTACTC	Normal CACGACCAGGTAGGAA-VIC
	Reverse GGGAGGGTGGCTGAGG	Mutant CACGACCAGATAGGAA-FAM
p.N350S	Forward GGACCTGCTGCCTACCC	Normal CTGCCCAATGTCAC-VIC
	Reverse GCTGAGGTCAAAGCCATCCAA	Mutant CTGCCCAGTGTACAC-FAM
c.1524+95A>G	Forward GCTCCTCACTGGGAGTTGTG	Normal CCTGATAACGTAATAACAC-VIC
	Reverse GAATTGTCACAGCTGCAAGTCT	Mutant CTGATAACGTAGTAACAC-FAM

Table 2. Primers used for PCR amplification and DNA sequencing of ARSA gene*.

PCR fragment	Primer sequence (5' > 3')				Lenght (bp)
	Primer	Forward	Primer	Reverse	
A	AF	CTGCGGTATCGGAAAGAG	AR	CTGCAATCCATTGGGAGG	358
B	BF	ACATTCTGTCTGTCTCAGGG	BR	AAGCAGCTGAACTGCAAGGC	940
	BF'	CTCCAGTCTTTGCCCCCA	BR'	CGGGACAGTCAAGGGGCT	
C	CF	AGAGAGCTGGTCTTGGTAGGC	CR	AATTCTGTGCACAGGGCAAGG	756
	CF'	GGTGGGCAAGAAGCGGT	CR'	GTTCCGCATCAAGGTTGG	
D	DF	TCCCCGGAGCCCTAGATC	DR	GGCACCAGCACACAGCAT	554

*Bold letters represent primers also used for ARSA division in four amplicons.

Table 3. Patients' molecular characterization.

Patient	MLD-alleles		p.W193C	p.T391S	ARSA-PD allele	
	Allele 1	Allele 2			p.N3590S	c.1524+95A>G
1*	c.459+1G>A	c.459+1G>A	+/+	+/+	-/-	-/-
2	c.459+1G>A	c.459+1G>A	+/+	+/+	-/-	-/-
3*	g.1190_1191insC	g.1190_1191insC	+/+	+/+	-/-	-/-
4	p.S44P	p.P426L	-/-	-/-	-/+	-/-
5	c.459+1G>A	c.459+1G>A	+/+	+/+	-/-	-/-
6	c.459+1G>A	c.459+1G>A	+/+	+/+	-/-	-/-
7	p.R84W	p.R390W	-/-	-/-	-/-	-/-
8	c.103_110del8	c.459+1G>A	-/+	-/+	-/-	-/-
9	c.459+1G>A	?	-/+	-/+	-/-	-/-
10	c.459+1G>A	p.P426L	-/+	+/+	-/-	-/-
11	c.459+1G>A	p.P284S	-/+	+/+	-/-	-/-
12	c.459+1G>A	c.459+1G>A	+/+	+/+	-/-	-/-
13	c.103_110del8	c.459+1G>A	-/+	+/+	-/-	-/-
14	c.459+1G>A	c.459+1G>A	-/-	-/-	-/-	-/-
15	c.103_110del8 + g.1190_1191insC	g.1190_1191insC	+/+	+/+	-/-	-/-
16*	g.2408DelC	g.2408DelC	-/-	+/+	-/-	-/-
17*	p.P426L	p.P426L	-/-	-/-	-/-	-/-
18	c.103_110del8	p.I179S	-/-	-/-	-/-	-/-
19	c.848-1G>A	p.P426L	-/+	-/+	-/-	-/-
20	c.459+1G>A	c.459+1G>A	+/+	+/+	-/-	-/-
21	c.459+1G>A	c.459+1G>A	+/+	+/+	+/+	-/-
22	c.459+1G>A	c.459+1G>A	+/+	+/+	-/-	-/-
23*	p.G122C	p.G122C	-/-	+/+	+/+	+/+
24	c.459+1G>A	p.F247S	-/+	-/+	-/-	-/-
25*	c.459+1G>A + p.D335V	p.D335V	-/+	-/-	-/-	-/-
26	p.R244H	p.G325S	-/-	-/+	+/+	-/+
27	p.D255H	?	-/-	-/+	-/+	-/-

* denotes patients from consanguineous marriages; ? indicates non-identified alleles; -/- indicates homozygous patient for normal allele; +/- indicates heterozygous patient; +/+ indicates homozygous patient for mutant allele.

REFERENCES

- [1] K. Von Figura, V. Gieselmann, J. Jaeken, Metachromatic Leukodystrophy, in: C.R. Scriver, A.L. Beaudet, W.S. Sly, Valle, D. B. Childs, K.E. Kinzler, B. Vogelstein (Eds), *The Metabolic and Molecular Bases of Inherited Disease*, 8nd edition, McGraw-Hill, New York (2001), 5833-5843.
- [2] R. Pinto, C. Caseiro, M. Lemos, L. Lopes, A. Fontes, H. Ribeiro, E. Pinto, E. Silva, S. Rocha, A. Marcao, I. Ribeiro, L. Lacerda, G. Ribeiro, O. Amaral, M. C. Miranda, Prevalence of lysosomal storage diseases in Portugal, *Eur. J. Hum. Genet.* 12 (2004), 87-92.
- [3] A. Biffi, G. Lucchini, A. Rovelli, M. Sessa, Metachromatic leukodystrophy: an overview of current and prospective treatments, *Bone Marrow Transplant.* 42 Suppl 2 (2008), S2-6.
- [4] Y. Fukutani, Y. Noriki, K. Sasaki, K. Isaki, M. Kuriyama, K. Kurosawa, H. Ida, Adult-type metachromatic leukodystrophy with a compound heterozygote mutation showing character change and dementia, *Psychiatry Clin. Neurosci.* 53 (1999), 425-428.
- [5] J. Kreysing, K. von Figura, V. Gieselmann, Structure of the arylsulfatase A gene, *Eur. J. Biochem.* 191 (1990), 627-631.
- [6] J. Berger, B. Loschl, H. Bernheimer, A. Lugowska, A. Tylki-Szymanska, V. Gieselmann, B. Molzer, Occurrence, distribution, and phenotype of arylsulfatase A mutations in patients with metachromatic leukodystrophy, *Am. J. Med. Genet.* 69 (1997), 335-340.

- [7] K. Kurosawa, H. Ida, Y. Eto, Prevalence of arylsulphatase A mutations in 11 Japanese patients with metachromatic leukodystrophy: identification of two novel mutations, *J. Inherit. Metab. Dis.* 21 (1998), 781-782.
- [8] A. Biffi, M. Cesani, F. Fumagalli, U. Del Carro, C. Baldoli, S. Canale, S. Gerevini, S. Amadio, M. Falautano, A. Rovelli, G. Comi, M. G. Roncarolo, M. Sessa, Metachromatic leukodystrophy - mutation analysis provides further evidence of genotype-phenotype correlation, *Clin. Genet.* 74 (2008), 349-357.
- [9] A. Polten, A. L. Fluharty, C. B. Fluharty, J. Kappler, K. von Figura, V. Gieselmann, Molecular basis of different forms of metachromatic leukodystrophy, *N. Engl. J. Med.* 324 (1991), 18-22.
- [10] V. Gieselmann, A. Polten, J. Kreysing, K. von Figura, Arylsulfatase A pseudodeficiency: loss of a polyadenylation signal and N-glycosylation site, *Proc. Natl. Acad. Sci. USA.* 86 (1989), 9436-9440.
- [11] M. L. Barth, A. Fenson, A. Harris, Prevalence of common mutations in the Arylsulphatase A gene in Metachromatic Leukodystrophy patients diagnosed in Britain, *Hum. Genet.* 30 (1993), 73-77.
- [12] M. L. Barth, A. Fensom, A. Harris, Missense mutations in the arylsulphatase A genes of Metachromatic Leukodystrophy patients, *Hum. Molec. Genet.* 2 (1993) 2117-2121.
- [13] C. G. Pedron, P. A. Gaspar, R. Giugliani, M. L. Pereira, Arylsulfatase A pseudodeficiency in healthy Brazilian individuals, *Bras. J. Med. Biol. Res.* 199 (1999), 941-945.

- [14] S. A. Miller, D. D. Dykes, H. F. Polesky, A simple salting out procedure for extracting DNA from human nucleated cells, *Nucleic. Acids. Res.* 16 (1988), 1215.
- [15] V. Ramensky, P. Bork, S. Sunyaev, Human non-synonymous SNPs: server and survey, *Nucleic. Acids. Res.* 30 (2002), 3894-3900.
- [16] M. H. Ricketts, R. D. Poretz, P. Manowitz, The R496H mutation of arylsulfatase A does not cause metachromatic leukodystrophy, *Hum. Mutat.* 12 (1998), 238-239.
- [17] C. Perusi, M. Gomez-Lira, M. Mottes, P. F. Pignatti, N. Rizzuto, A. Salviati, A novel mutation which represents the fifth non-pathogenic polymorphism in the coding sequence of the arylsulfatase A gene, *Mol. Cell Probes.* 11 (1997), 449-451.
- [18] A. Lugowska, O. Amaral, J. Berger, L. Berna, N. U. Bosshard, A. Chabas, A. Fensom, V. Gieselmann, N. G. Gorovenko, W. Lissens, J. E. Mansson, A. Marcao, H. Michelakakis, H. Bernheimer, N. V. Ol'khovych, S. Regis, R. Sinke, A. Tytki-Szymanska, B. Czartoryska, Mutations c.459+1G>A and p.P426L in the ARSA gene: prevalence in metachromatic leukodystrophy patients from European countries, *Mol. Genet. Metab.* 86 (2005), 353-359.
- [19] O. Artigas, V. L. Lagranha, M. L. Saraiva-Pereira, M. G. Burin, C. M. Lourenco, H. van der Linden, M. L. Santos Jr., S. Rosemberg, C. E. Steiner, F. Kok, C. F. de Souza, L. B. Jardim, R. Giugliani, I. V. Schwartz, Clinical and biochemical study of 29 Brazilian patients with metachromatic leukodystrophy, *J. Inherit. Metab. Dis.* (2010).
- [20] V. Gieselmann, A. Polten, J. Kreysing, J. Kappler, A. Fluharty, K. von Figura, Molecular genetics of metachromatic leukodystrophy, *Dev. Neurosci.* 13 (1991), 222-227.
- [21] R. von Bulow, B. Schmidt, T. Dierks, N. Schwabauer, K. Schilling, E. Weber, I. Uson, K. von Figura, Defective oligomerization of arylsulfatase a as a cause of its

instability in lysosomes and metachromatic leukodystrophy, *J. Biol. Chem.* 277 (2002), 9455-9461.

[22] R. Draghia, F. Letourneur, C. Drugan, J. Manicom, C. Blanchot, A. Kahn, L. Poenaru, C. Caillaud, Metachromatic leukodystrophy: identification of the first deletion in exon 1 and of nine novel point mutations in the arylsulfatase A gene, *Hum. Mutat.* 9 (1997), 234-242.

[23] L. Gort, M. J. Coll, A. Chabas, Identification of 12 novel mutations and two new polymorphisms in the arylsulfatase A gene: haplotype and genotype-phenotype correlation studies in Spanish metachromatic leukodystrophy patients, *Hum. Mutat.* 14 (1999), 240-248.

[24] A. Marcao, O. Amaral, E. Pinto, R. Pinto, M. C. Sa Miranda, Metachromatic leucodystrophy in Portugal-finding of four new molecular lesions: C300F, P425T, g.1190-1191insC and g.2408delC, *Hum. Mutat.* 13 (1999), 337-338.

[25] A. Lugowska, J. Berger, A. Tylki-Szymanska, B. Loschl, B. Molzer, M. Zobel, B. Czartoryska, Molecular and phenotypic characteristics of metachromatic leukodystrophy patients from Poland, *Clin. Genet.* 68 (2005), 48-54.

[26] J. Zlotogora, Y. Furman-Shaharabani, A. Harris, M. L. Barth, K. von Figura, V. Gieselmann, A single origin for the most frequent mutation causing late infantile metachromatic leucodystrophy, *J. Med. Genet.* 31(1994), 672-674.

[27] A. Lugowska, B. Czartoryska, A. Tylki-Szymanska, M. Bisko, J. G. Zimowski, J. Berger, B. Molzer, Prevalence of arylsulfatase A pseudodeficiency allele in metachromatic leukodystrophy patients from Poland, *Eur. Neurol.* 44 (2000), 104-107.

- [28] S. Regis, M. Filocamo, M. Stroppiano, F. Corsolini, R. Gatti, Molecular analysis of the arylsulphatase A gene in late infantile metachromatic leucodystrophy patients and healthy subjects from Italy, *J. Med. Genet.* 33 (1996), 251-252.
- [29] S. Regis, F. Corsolini, V. Ricci, M. Di Duca, M. Filocamo, An unusual arylsulfatase A pseudodeficiency allele carrying a splice site mutation in a metachromatic leukodystrophy patient, *Eur. J. Hum. Genet.* 12 (2004), 150-154.
- [30] R. Ott, J. S. Waye, P. L. Chang, Evolutionary origins of two tightly linked mutations in arylsulfatase-A pseudodeficiency, *Hum. Genet.* 101 (1997), 135-140.
- [31] J. Wang, W. Zhang, H. Pan, X. Bao, Y. Wu, X. Wu, Y. Jiang, ARSA gene mutations in five Chinese metachromatic leukodystrophy patients, *Pediatr. Neurol.* 36 (2007), 397-401.
- [32] M. B. Coulter-Mackie, L. Gagnier, Spectrum of mutations in the arylsulfatase A gene in a Canadian DNA collection including two novel frameshift mutations, a new missense mutation (C488R) and an MLD mutation (R84Q) in cis with a pseudodeficiency allele, *Mol. Genet. Metab.* 79 (2003), 91-98.
- [33] S. Grossi, S. Regis, C. Rosano, F. Corsolini, G. Uziel, M. Sessa, M. Di Rocco, G. Parenti, F. Deodato, V. Leuzzi, R. Biancheri, M. Filocamo, Molecular analysis of ARSA and PSAP genes in twenty-one Italian patients with metachromatic leukodystrophy: identification and functional characterization of 11 novel ARSA alleles, *Hum. Mutat.* 29 (2008), E220-230.
- [34] A. Lugowska, R. Ploski, P. Wlodarski, A. Tytki-Szymanska, Molecular bases of metachromatic leukodystrophy in Polish patients, *J. Hum. Genet.* 55 (2010), 394-396.

- [35] A. Lugowska, P. Wlodarski, R. Ploski, H. Mierzewska, M. Dudzinska, A. Matheisel, H. Swietochowska, A. Tyłki-Szymanska, Molecular and clinical consequences of novel mutations in the arylsulfatase A gene, *Clin. Genet.* 75 (2009), 57-64.
- [36] N. V. Olkhovich, N. Takamura, N. A. Pichkur, N. G. Gorovenko, K. Aoyagi, S. Yamashita, Novel mutations in arylsulfatase A gene in three Ukrainian families with metachromatic leukodystrophy, *Mol. Genet. Metab.* 80 (2003), 360-363.
- [37] B. Hess, S. Kafert, U. Heinisch, D. A. Wenger, J. Zlotogora, V. Gieselmann, Characterization of two arylsulfatase A missense mutations D335V and T274M causing late infantile metachromatic leukodystrophy, *Hum. Mutat.* 7 (1996), 311-317.
- [38] F. Schestag, A. Yaghootfam, M. Habetha, P. Poepfel, F. Dietz, R. A. Klein, J. Zlotogora, V. Gieselmann, The functional consequences of mis-sense mutations affecting an intra-molecular salt bridge in arylsulphatase A, *Biochem. J* 367 (2002), 499-504.
- [39] Q. Yong, E. Shapira, R. J. Desnick, Metachromatic Leukodystrophy: Subtype Genotype/Phenotype Correlations and Identification of Novel Missense Mutations (P148L and P191T) Causing the Juvenile-Onset Disease, *Mol. Genet. Metab.* 67 (1999), 206–212.
- [40] B. Eng, L. N. Nakamura, N. O'Reilly, N. Schokman, M. M. Nowaczyk, W. Krivit, J. S. Wayne, Identification of nine novel arylsulfatase a (ARSA) gene mutations in patients with metachromatic leukodystrophy (MLD), *Hum. Mutat.* 22 (2003), 418-419.
- [41] N. M. Pastor-Soler, M. A. Rafi, J. D. Hoffman, D. Hu, D. A. Wenger, Metachromatic leukodystrophy in the Navajo Indian population: a splice site mutation in intron 4 of the arylsulfatase A gene, *Hum. Mutat.* 4 (1994), 199-207.

- [42] N. M. Pastor-Soler, E. M. Schertz, M. A. Rafi, G. de Gala, D. A. Wenger, Metachromatic leukodystrophy among southern Alaskan Eskimos: molecular and genetic studies, *J. Inherit. Metab. Dis.* 18 (1995), 326-332.
- [43] W. Lissens, R. Vervoort, N. Van Regemorter, P. Van Bogaert, M. Freund, C. Verellen-Dumoulin, S. Seneca, I. Liebaers, A D255H substitution in the arylsulphatase A gene of two unrelated Belgian patients with late-infantile metachromatic leukodystrophy, *J. Inherit. Metab. Dis.* 19 (1996), 782-786.

5. ARTIGO II

THE EFFECTS OF GLYCOSYLATION AND PH CONDITIONS IN MOLECULAR DYNAMICS OF HUMAN ARYLSULFATASE A

(Artigo a ser submetido ao periódico *Biochemistry*)

THE EFFECTS OF GLYCOSYLATION AND PH CONDITIONS IN THE DYNAMICS OF HUMAN ARYLSULFATASE A[†]

Key-words: Metachromatic leukodystrophy, Molecular dynamics, High mannose-type oligosaccharides, ARSA pseudodeficiency.

Virgens, M.Y.F.^{1,2}, Pol-Fachin, L.³, Verli, H.*^{3,4}, Saraiva-Pereira, M.L.^{1,2,3,4,5}

¹Laboratório de Identificação Genética, Centro de Pesquisas, Hospital de Clínicas de Porto Alegre, R. Ramiro Barcelos, 2350, Porto Alegre, 90035-903, RS, Brazil.

²Departamento de Genética, Universidade Federal do Rio Grande do Sul, Av. Bento Gonçalves, 9500, Porto Alegre, 91500-970, RS, Brazil.

³Centro de Biotecnologia, Universidade Federal do Rio Grande do Sul, Av. Bento Gonçalves 9500, Porto Alegre, CP 15005, 91500-970, RS, Brazil.

⁴Faculdade de Farmácia, Universidade Federal do Rio Grande do Sul, Av. Ipiranga 2752, Porto Alegre, 90610-000, RS, Brazil.

⁵Departamento de Bioquímica, Universidade Federal do Rio Grande do Sul, R. Ramiro Barcelos, 2600, Porto Alegre, 90035-003, RS, Brazil.

***Corresponding author:**

Hugo Verli PhD

Av. Bento Gonçalves 9500,

Porto Alegre, CP 15005, 91500-970, RS, Brazil

E-mail address: hverli@cbiot.ufrgs.br

Phone: +55 51 3308 7770

Fax: +55 51 3308 7309

[†] This work was supported by Brazilian Funding Agencies (CNPq, FAPERGS and FINE-HCPA).

ABSTRACT

Arylsulfatase A is a lysosomal sulfatase that catalyzes the hydrolysis of cerebroside sulfate. Usually, ARSA deficiency in humans results in Metachromatic Leukodystrophy, whereas a minor condition, called ARSA pseudodeficiency, occurs in healthy individuals. ARSA suffers post-translational oxidation of a conserved cysteine to an aldehyde, yielding formylglycine, which acts as a ligand of an octahedrally coordinated Mg^{+2} . ARSA is N-glycosylated with high mannose-type oligosaccharides at three sites. The absence of glycosylation in the p.N350S ARSA-mutant has a pronounced effect decreasing ARSA activity, and it is associated with ARSA pseudodeficiency condition. In this context, the current work intends to investigate ARSA dynamics in response to pH and glycosylation, employing molecular dynamics (MD) simulations, in order to obtain structural insights in to ARSA deficiencies. Accordingly, four systems were studied: (i) non-glycosylated ARSA at pH~7; (ii) non-glycosylated ARSA at pH~5; (iii) fully-glycosylated ARSA at pH~5; (iv) poorly-glycosylated p.N350S-mutant at pH~5. At pH~5, ARSA presented higher conformational stability when compared to pH~7, supporting the formation of an intramolecular salt-bridge between Asp335 and Arg370, compatible with ARSA lysosomal activity. The fully-glycosylated ARSA presented an even higher overall rigidity, notably at residues close to the glycosylation sites. The absence of glycosylation at the p.N350S-mutant destabilized its entire adjacent region, including the salt-bridge D335-R370 and the major core helix. Finally, the obtained data are discussed in the search for structural evidences able to contribute to the understanding of molecular etiology of ARSA deficiencies.

1. INTRODUCTION

Sulfatases are an evolutionary highly conserved family of proteins, essential for degradation and remodeling of sulfate esters (1,2,3). The physiological importance of these enzymes in humans is illustrated by seven distinct severe lysosomal storage disorders, each one caused by the deficiency of one specific sulfatase (4,5). The Arylsulfatase A (ARSA) deficiency causes Metachromatic Leukodystrophy (MLD), an autosomal recessive inborn error of metabolism, which results in a lysosomal accumulation of cerebroside sulfate, the major physiological substrate of ARSA, mainly in the white matter of the central nervous system and peripheral nerves, with severe neurological consequences (5). In addition, about 0.2-0.5% of the healthy population presents ARSA activity levels similar to those of MLD patients. This non-pathogenic condition, known as ARSA pseudodeficiency (ARSA-PD), is caused by the homozygosity for ARSA-PD allele. The normal and ARSA-PD differ by two A>G transitions, being one related to the substitution of a glycosylated Asn350 to serine (p.N350S), abolishing the glycosylation at this site, and the other concerned with the alteration of the polyadenylation signal (c.1524+95A>G) (6).

ARSA is synthesized as a 507 aa (~62KDa) polypeptide that, after translocation into endoplasmic reticulum (ER), is cleaved at its N-terminal region, yielding a 489 aa (~51KDa) mature enzyme, presenting a mixed α/β type secondary structure content. The ARSA core is conserved between sulfatases and consists of two β -pleated sheets: the minor β -sheet, formed by four antiparallel β -strands (β 15-18), and the major one, composed of ten β -strands (β 1,2,4,5,8-13). The major β -pleated sheet is sandwiched by seven helices (α A-G). At the C-terminus, a loop including the long helix α I is formed by nearly 80 aa and stretches over the almost whole ARSA monomer. As well, a cluster of six cysteine residues knots the C-terminal 20 aa by forming disulfides bonds, which is unique to ARSA

(3,7) [Fig. 1B].

Still in the ER, ARSA is N-glycosylated at three sites (Asn158, Asn184 and Asn350) with core fucosylated high mannose-type oligosaccharides (8) [Fig. 1C]. As well, formylglycine (FGly), a key catalytic residue in its active site, is generated by the FGly-generating enzyme (FGE) through oxidation of the CH₂SH group of a conserved cysteine to an aldehyde. Subsequently, dimers are formed and transported to the Golgi apparatus, where they receive mannose-6-phosphate (M6P) recognition markers at the Asn158 and Asn350 linked oligosaccharides, and bind to M6P receptors (M6PR). The ARSA-M6PR complexes are transported to endosomes, where the complexes dissociate. The receptors return to Golgi and ARSA is delivered to lysosomes, into which the enzyme octamerize in a pH-dependent manner [Fig. 1A], and exert its physiological activity (9,10).

Besides the ARSA lysosomal pathway, the controversial recent finding of ARSA involvement as an extracellular component has been discussed. Immunohistochemical data suggest that ARSA exists on the cell-surface as well as in lysosomes. However, while the cell-surface ARSA might be able to bind sulfates in the extracellular matrix, it would not hydrolyze them in such conditions, as its optimal pH is 4.5 (11). The ARSA pH dependence has been explained only in terms of its octameric association, but the pH importance for ARSA optimal activity or specific possible structural differences between lysosomal and cell-surface ARSA need further explanation.

Several evolutionary roles have been described to be performed by carbohydrates in glycoproteins, as stabilizing large regions of the backbone structure, being involved with protein folding and intracellular targeting (12,13,14,15). The most known function of ARSA high mannose-type oligosaccharides is the ARSA delivery to lysosomes, a common pathway with many lysosomal hydrolases (16,17). In addition, implications concerning

ARSA glycosylation in its conformational stability have been suggested by *in vitro* studies, pointing to reduced stability and activity in poor-glycosylated ARSA p.N350S-mutant (18,19). Moreover, ARSA carbohydrate component modifications, as sialylation, phosphorylation or sulfation, are often associated with the appearance of tumors (20,21).

Despite the above mentioned *in vitro* demonstrations, the molecular basis related to the glycosylation effects over the ARSA activity and conformation remains unclear, in part because of the difficulties associated with the experimental determination of glycoprotein three-dimensional structure, mainly due to the high flexibility of the carbohydrate moieties. In view of that, X-ray derived structures usually cannot preserve their complete structure (22,23,24). In this scenario, molecular dynamics (MD) technique emerges as a promising tool to describe the conformational behavior of glycoconjugates in environments resembling the biological medium. Accordingly, MD simulations have already been demonstrated to accurately describe the conformational behavior of carbohydrates, being so able to explain time-variable aspects of the relationship between glycan and protein moieties (25,26,27,28,29,30).

In this context, the current work intends to investigate ARSA dynamics in response to changes in pH and glycosylation. Accordingly, ARSA MD simulations were performed in different conditions, and the retrieved data were compared with previous *in vitro* studies (including direct-site mutagenesis, pulse-chase and X-ray). The obtained data are expected to further contribute to the knowledge of pH and glycosylation significances to ARSA optimal activity and anomalous behavior.

2. EXPERIMENTAL PROCEDURES

2.1. Nomenclature and Software

The nomenclature recommendations and symbols are used as proposed by IUPAC (31). The relative orientation of a pair of contiguous carbohydrate residues is described, for different types of linkages, by two or three torsional angles at the glycosidic linkage.

For a (1→X) linkage, where “X” is 3, 4 or 6 for the (1→3), (1→4) or (1→6) linkages, respectively, the ϕ and ψ are defined as show below:

$$\phi = \text{O5-C1-OX'-CX'} \quad (1)$$

$$\psi = \text{C1-OX'-CX'-C(X-1)'} \quad (2)$$

For a (1→6) linkage, the ω is defined as show below:

$$\omega = \text{O6'-C6'-C5'-C4'} \quad (3)$$

The manipulation of structures was performed with MOLDEN (32), VMD (33), SPDBV (34) and PYMOL (35). The secondary structure content analyses were performed with DSSP (36) and PROCHECK (37), and all the MD calculations and remaining analysis were performed using GROMACS 3.3.3 simulation suite (38), and GROMOS96 43a1 force field (39), with all saccharide and FGly69 residue topologies generated using the PRODRG server (40).

2.2. Topology Construction

Human ARSA in its non-glycosylated monomeric form was retrieved from Protein Data Bank (PDB) under code 1AUK (3). The gap between Gly443 and Ala448 and the missing C-termini Asp504 to Ala507 were constructed based on its protein sequence retrieved from National Center for Biotechnology Information (NCBI) under code P15289 (7), using Swiss-PDB Viewer program. As well, the FGly69 residue was constructed using MOLDEN in its aldehydic non-hydrated form, being further submitted to the PRODRG

server, from which its topology was retrieved. The glycosylated ARSA structures were filled with core fucosylated high mannose-type oligosaccharides, as described by previous mass spectrometry studies (8), using glycosciences modeling tools (41). These obtained models for glycosylated ARSA had their glycosidic linkage geometries adjusted to the main conformations for each linkage, based on their relative abundance in the isolated disaccharides in water, as previously described (30). The glycan parameters (that is, for their composing monosaccharides) were obtained from the PRODRG server and added by HF/6-31**-derived Löwdin charges, as reported (25,27,29). Improper dihedrals were included to preserve the hexopyranose conformations in accordance with their expected forms in aqueous solution: 4C_1 for D-GlcpNAc, 1C_4 for L-Fucp, 4C_1 for D-Manp, and 4C_1 for D-Galp. Additionally, N-glycosidic linkage between the carbohydrate and amino acid residues was treated as previously described (29). This MD protocol has been proved to be accurate in conformational analysis of carbohydrates, being able to explain dihedral angle distributions for glycosidic linkages found in RMN experiments (25,26,27,28,29,30). Finally, the parameters utilized to describe the Mg^{+2} ion were the standard parameters of GROMOS96 43a1 force field.

2.3. MD Simulations

Four different MD simulations were performed: (i) non-glycosylated ARSA at pH~7; (ii) non-glycosylated ARSA at pH~5; (iii) fully-glycosylated ARSA at pH~5; and (iv) poorly-glycosylated ARSA p.N350S-mutant at pH~5. Each structure was solvated in a rectangular box using periodic boundary conditions and SPC water model (42). The protonation states were automatically adjusted according to the standard pKa values to the amino acids side chains and additionally verified using PDB2PQR (43). Counter ions (Na+ and Cl-) were added to neutralize the systems. The employed MD protocol was based in

previous studies, as described (25,27,29). The Lincs method (44) was applied to constrain covalent bond lengths, allowing an integration step of 2 fs after an initial energy minimization using Steepest Descent algorithm. Electrostatic interactions were calculated with Particle Mesh Ewald Method (45). Temperature and pressure were kept constant by coupling glycoproteins, ions and solvent to external temperature and pressure baths with coupling constants of $\tau = 0.1$ and 0.5 ps, respectively (46). The dielectric constant was treated as $\epsilon = 1$, and the reference temperature was adjusted to 310K. The systems were slowly heated from 50K to 310K, in steps of 5ps, each one increasing the reference temperature by 50K, and extended to 50 ns.

3. RESULTS AND DISCUSSION

3.1. Structural Effects of pH and Triple Glycosylation

In order to search for structural evidences for the importance of acid pH to the optimal ARSA lysosomal activity we performed MD simulations of monomeric non-glycosylated ARSA in two different conditions: (i) at pH~7, condition close to plasmatic medium, and (ii) at pH~5, condition close to lysosomal medium.

At pH~5, ARSA presented lower values of overall RMSD when compared to ARSA at pH~7. While ARSA at pH~7 overall RMSD gradually increased, the ARSA at pH~5 RMSD remained stable during most of simulation time, except by the sudden increase at the end of simulation time [Fig. 2B]. This sudden RMSD increase was associated to alterations in its helix secondary structure content, especially the helix D folding and helix E and I unfolding [Fig. 3B].

While the global flexibility of ARSA at pH~5 was reduced, specific regions presented an increase in flexibility [Fig. 3B]. The β -twisted enzyme hydrophobic core proved to be very rigid in ARSA at pH~5 [region X in Fig. 3B]. This profile is compatible with optimal ARSA lysosomal activity. Nevertheless, regions close to glycosylation sites became more flexible. This changing in the patten of flexibility between the protein core structures and regions near to glycosylation sites lead us to hypothesize about the likely stabilizing role of glycans, as reported for others glycoproteins (12,13,15).

In this context, a MD simulation of monomeric, fully-glycosylated, ARSA at pH~5 was performed in order to obtain insights into ARSA glycosylation structural role. The fully-glycosylated ARSA presented the lowest overall RMSD and global flexibility when compared to non-glycosylated ARSA at pH~5 or non-glycosylated ARSA at pH~7 [Fig.

2B]. As expected, the regions near to glycosylation sites also became less flexible, and its entire secondary structure content was very constant during all simulation [Fig. 3C].

The glycans at Asn158 and Asn184 composed an arrangement with the protein C-termini, a region in which lies the ARSA-unique cluster of six cysteine residues that knot the ARSA C-terminal through three disulfide bonds [Fig. 4A and C]. As noted for other proteins (47), the glycans arrangement may work as a steric protective factor against proteolytic attacks in this region. Likewise, the glycosylation at Asn350 interacts with the salt-bridge D335-R370 loop region and may be important not only to its stabilization but also to the steric protection against proteolysis of this entire critical region [Fig. 2A and C].

An increase in ARSA positive electrostatic potential surface occurs from pH~7 to pH~5 [Supplementary Fig.4], which may be resultant from protonated histidine residues (48). The three major positive ARSA charged sites at both pH~7 and pH~5 are: (i) the octameric interface, (ii) the loop components of dimeric interface, and (iii) the C-termini tail [Supplementary Fig. 3]. This heterogenic pattern of ARSA electrostatic potential is more evident at pH~7 and more diffuse at pH~5. The majority of ARSA ionizable residues are located at the enzyme surface, especially on the hat-shaped monomer base, where ARSA association interfaces are located.

3.2. ARSA-p.N350S Disturbing Effects

Historically, the reduction in ARSA activity caused by ARSA-PD has been attributed to the polyadenylation signal (18,19). However, the hypothesis that p.N350S mutation had no effect on ARSA enzymatic activity has been questioned by the identification of MLD heterozygous patients who carried the p.N350S allele without the polyadenylation defect (49,50,51). Taking into account these findings and the high structural stability of fully-glycosylated ARSA state observed in this work, we performed a MD simulation of monomeric p.N350S ARSA-mutant at pH~5 in order to contribute for

the enlightenment of this issue.

While ARSA p.N350S-mutant overall RMSD presented a behavior similar to the non-glycosylated ARSA at pH~7 [Fig. 2B], its global flexibility increases when compared with the fully-glycosylated ARSA. Specifically, the absence of the glycosylation at Asn350 destabilized its adjacent region, including the salt-bridge D335-R370 [Fig. 2C]. At the end of the simulation time, the helix I is moved away from protein core [Fig. 2D], as occurred with the non-glycosylated ARSA at pH~7, and the C-termini tail movement lead to modifications in the arrangement between the glycans at Asn158, Asn184 and the C-termini [Fig. 4A and D]. The glycan at Asn184 departs from the C-termini and start interacting with neighbor helices D and E (the longest helix in the system) which were, respectively, destabilized and unfolded [Fig. 4D]. In view of that, the loss of ~57% in ARSA-p.N350S *in vitro* activity may be related to changes in ARSA-p.N350S entire secondary content, as could be seen through its behavior during this simulation.

Missense mutations may be able to disturb ARSA structure and cause partial or complete loss of activity (5). This hypothesis is corroborated by two mutations (p.D335V and p.R370W), that disturb the functionally important salt-bridge D335-R370 (52,53,54). The instability promoted by the ARSA salt-bridge loss may be illustrated by pulse-chase experiments, in which p.D335V and p.R370W mutants, in contrast to the wild-type ARSA, were degraded within 16h (53). These mutations have been described in patients with severe MLD and cause a complete loss of enzyme activity (54).

In this work, the role of the salt-bridge D335-R370 in ARSA conformational stability was observed in both non-glycosylated ARSA at pH~5 system and in fully-glycosylated ARSA at pH~5 [Fig. 2C]. It has been speculated that the formation of salt-bridges in the interior of a protein could be less favorable and even disfavors the protein

folding (55,56). However, empirical studies of direct-site mutagenesis had demonstrated that the removal of ion pair members usually destabilize the native protein structure by 3-5 Kcal.mol⁻¹, so the maintenance of Coulomb interactions in the protein interior may be related to an increase in conformational stability and specific folding of some proteins (57).

The salt-bridge D335-R370 in ARSA is located in the interior of the enzyme, and the residues involved are only partially accessible to solvent, and, with exception of iduronate sulfatase, different but equivalent salt-bridges are conserved within the sulfatase family (2). The salt-bridge D335-R370 importance for ARSA correct folding has been recently described, and the tridimensional disposition of ARSA amino acid residues suggests that this salt-bridge may work as a hook, helping to sustain the oligomerization domain close to the enzyme core, thus favoring a proper hat-shaped monomer. This hypothesis may be also evidenced by the previously described data, in which the retention of ARSA-D335V in the ER occurs as resulted of the ARSA inappropriate fold (53).

3.3. Glycans Analysis

The N-glycosidic linkage conformation has been extensively studied by both NMR and X-ray methods (13,24,58,59). These studies show the asparagine linkage to be relatively rigid and planar, with a tendency to extend the first glycan residue away from the peptide backbone and into the solvent. Moreover, previous X-ray data indicates that the φ_{N-GL} dihedral angle presents higher flexibility compared to ψ_{N-GL} dihedral angle, with 80° and 40° amplitude, respectively (58). This large amplitude of the φ_{N-GL} also suggests the co-existence of multiple conformer populations in solution (29,30).

In agreement with these findings, all ARSA glycans at fully-glycosylated ARSA at pH~5 system presented similar conformational patterns around the GlcNAc-(1→N)-Asn during the performed simulation, in a behavior mostly independent on the surrounding

protein scaffold [Supplementary Fig. 1]. However, the φ_{N-GL} in ARSA-p.N350S system presented an irregular behavior [Fig. 4B]. The distance between the glycan at Asn184 and the ARSA C-termini progressively increase in ARSA-p.N350S system [Fig. 4D]. As a result of this loss of interactions between the glycan at Asn184 and ARSA C-termini, changes are observed to occur in the glycan orientation, and then the Asn158 side chain undergoes torsion. This φ_{N-GL} irregular behavior in the ARSA-p.N350S may be result of this torsion. In the same way, the irregular behavior of the φ angle from β -D-GlcpNAc-(1 \rightarrow 4)- β -D-GlcpNAc at the ARSA-p.N350S follows the glycan at Asn184 reorientation [Supplementary Fig. 2].

Additionally, in order to search for possible conformational influences of the ARSA proteic core on glycans conformation beyond its linkage to asparagine residues, each dihedral angle from glycosidic linkages was analyzed and compared to the conformational patten observed for the proper disaccharide units in solution [Supplementary data – Fig. 1 and 2]. Accordingly, disaccharides linkages in different carbohydrate trees from ARSA did not present major new conformer populations depending on the site of attachment to ARSA, but distinct preferences over the conformations observed in solution for isolated disaccharides units, as observed in previous MD studies (29,30). Such behavior denotes a minor influence of the protein scaffold on glycan conformational preference.

3.4. Active-site Considerations

The ARSA active site residues and architecture are highly conserved among eukaryotic and prokaryotic sulfatases, indicating a common catalytic mechanism shared by members of the family, and even more the active center regions in ARSA and Arylsulfatase B (ARSB) show remarkable structural homology (1,2). These sulfatases contain divalent metal cations that are coordinated by three aspartate side chains, an

asparagine residue and the key residue FGly. In the case of ARSA, the active site is a positively charged pocket, where the residues Asp29, Asp30, Asp281 and Asn282 and FGly69 are responsible for the Mg^{+2} octahedral coordination. The positively charged residues Lys123, Lys302, His229 and Ser150 are proposed to allocate the sulfate group (60).

In all systems simulated in this work the Mg^{+2} remains correctly positioned in the active site pocket, except in the ARSA-p.N350S system. Accordingly, the overall distances between oxygens (Asp29-O ϑ , Asp30-O δ , Asp281-O δ and O ϑ , Asn282-O ϑ , and FGly-O γ) and Mg^{+2} were measured [Table 1]. The smaller distances between the oxygens donors and the Mg^{+2} were noted for fully-glycosylated ARSA, and the FGly-O γ only reaches an optimal distance for coordination in this system. On the other hand, in the ARSA-p.N350S system the metal escape as a result of the previous global destabilization of this system [Fig. 5]. The metal escape leads a loss of electrostatic potential at the catalytic pocket, which became strongly negative (Supplementary Fig. 4). This behavior agrees to the already suggested hypothesis about the large importance of Mg^{+2} for ARSA binding sulfate by giving rise to a more favorable positive microenvironment (48).

4. CONCLUSIONS

In the current work we were able to explain how ARSA dynamics can be affected by environmental changes, as pH, and intrinsic components, as different patterns of glycosylation. Several general and specific hypotheses, early formulated by *in vitro* findings, could be confirmed by the time-variable approach presented here:

- At cytoplasmatic pH, non-glycosylated ARSA may have its catalytic function abolished by structural changes associated with flexibility increase and loss of important interactions, as the salt-bridge D335-R370, which may disturb the proper folding of hat-shaped monomer.
- At lysosomal pH, non-glycosylated ARSA is more stable than at cytoplasmatic pH. However, this transient conformational stability only becomes more permanent in a triple glycosylation state, allowing the maintenance of secondary structure and the specific disposition of glycans. Glycosylation also presents a role as a steric protective factor against proteolytic attacks.
- The absence of ARSA third glycan results in global structural changes and in the most drastic changes in secondary structure, which can be associated with the loss of 57% in ARSA-p.N350S *in vitro* activity, noted in the ARSA pseudodeficiency.
- While the ARSA protein moiety suffers important modifications derived of the interactions with glycan, as becoming more rigid, the glycan moieties do not present new conformer populations. This fact denotes a minor influence of the protein scaffold on glycan conformation.

With these findings, we hope to contribute to the enlightening of the molecular etiology of ARSA deficiencies as well to the better comprehension of the relationship between glycan and protein moieties in glycoproteins.

ACKNOWLEDGEMENTS

We are grateful to Guilherme Geisel for technical help. This work was supported by the following Brazilian Funding Agencies: Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), MCT, Fundação de Amparo à Pesquisa do Rio Grande do Sul (FAPERGS), and Fundo de Incentivo à Pesquisa e Eventos do Hospital de Clínicas (FIPE-HCPA).

TABLES AND FIGURES

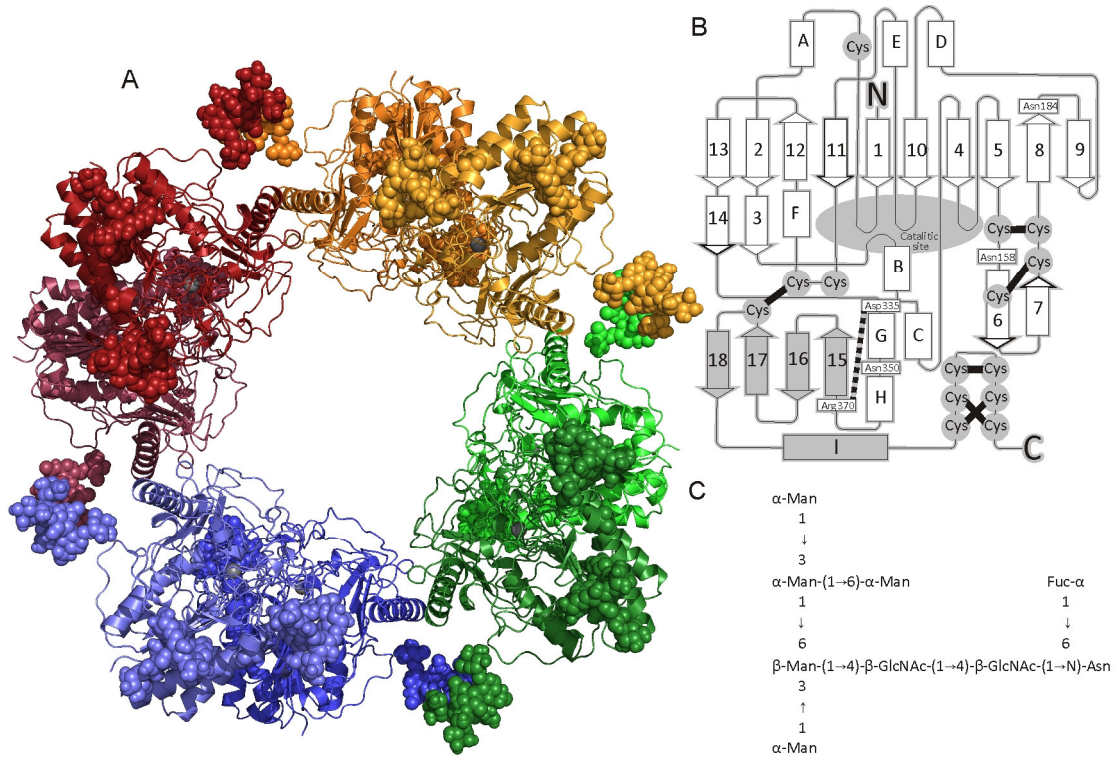


Figure 1. **(A) Model for fully-glycosylated ARSA octameric association.** The dimeric interface is composed by loops, and the main structure in the octameric interface is composed by the helix I. The ARSA glycans are presented in the oligomeric structure. **(B) ARSA topology.** α - helices are shown as columns and β -strands as arrows, numbered from N- to C-termini. Cysteine residues are circled, with disulfide bridges presented as thick black lines. The active site region is represented by the shaded ellipse. The glycosylation sites and amino acid residues forming the salt-bridge D335-R370 are boxed, and the salt-bridge residues are shown within dashed lines. The oligomerization domain is marked as grey. Adapted from Lukatela *et al*, 1998 (3). **(C) Scheme of ARSA core fucosylated high mannose-type oligosaccharides.**

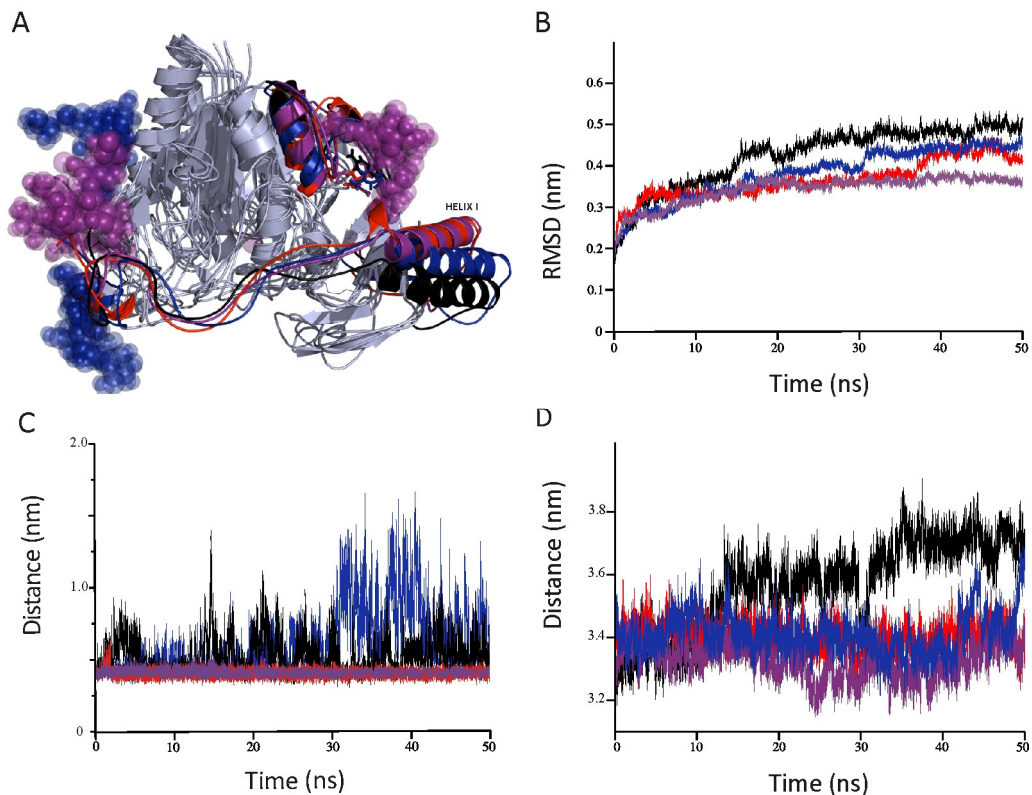


Figure 2. **ARSA RMSD analysis.** (A) Superimposition of the final frames of each system. (B) All-atom RMSD for non-glycosylated ARSA at pH~7 (black), non-glycosylated ARSA at pH~5 (red), fully-glycosylated ARSA at pH~5 (purple), and poorly-glycosylated p.N350S-mutant at pH~5 (blue). (C) Distance between Asp335-C γ and Arg370-N centers of mass for non-glycosylated ARSA at pH~7 (black), non-glycosylated ARSA at pH~5 (red), fully-glycosylated ARSA at pH~5 (purple), and poorly-glycosylated p.N350S-mutant at pH~5 (blue). (D) Distance between ARSA core domain and Helix I mass centers as a function of the time for non-glycosylated ARSA at pH~7 (black), non-glycosylated ARSA at pH~5 (red), fully-glycosylated ARSA at pH~5 (purple), and poorly-glycosylated p.N350S-mutant at pH~5 (blue).

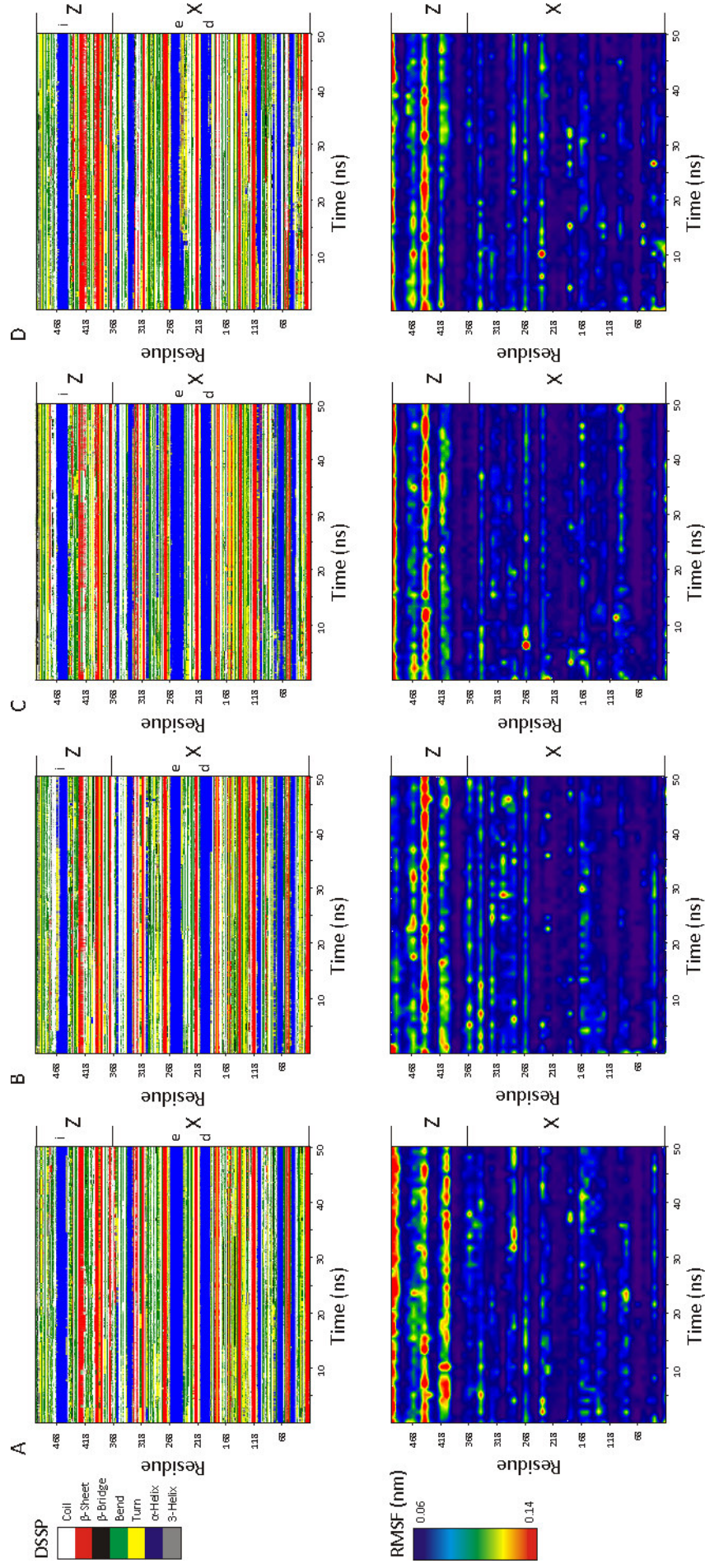


Figure 3. **Global flexibility and secondary modifications.** DSSP and RMSF analysis per residue as a function of time for ARSA at pH~7 (A), non-glycosylated ARSA at pH~5 (B), fully-glycosylated p.N350S-mutant at pH~5 (C), and poorly-glycosylated ARSA at pH~5 (D). The oligomeric and catalytic domains are represented by Z and X regions, respectively, and helices D, E and I are highlighted in DSSP plots by small letters.

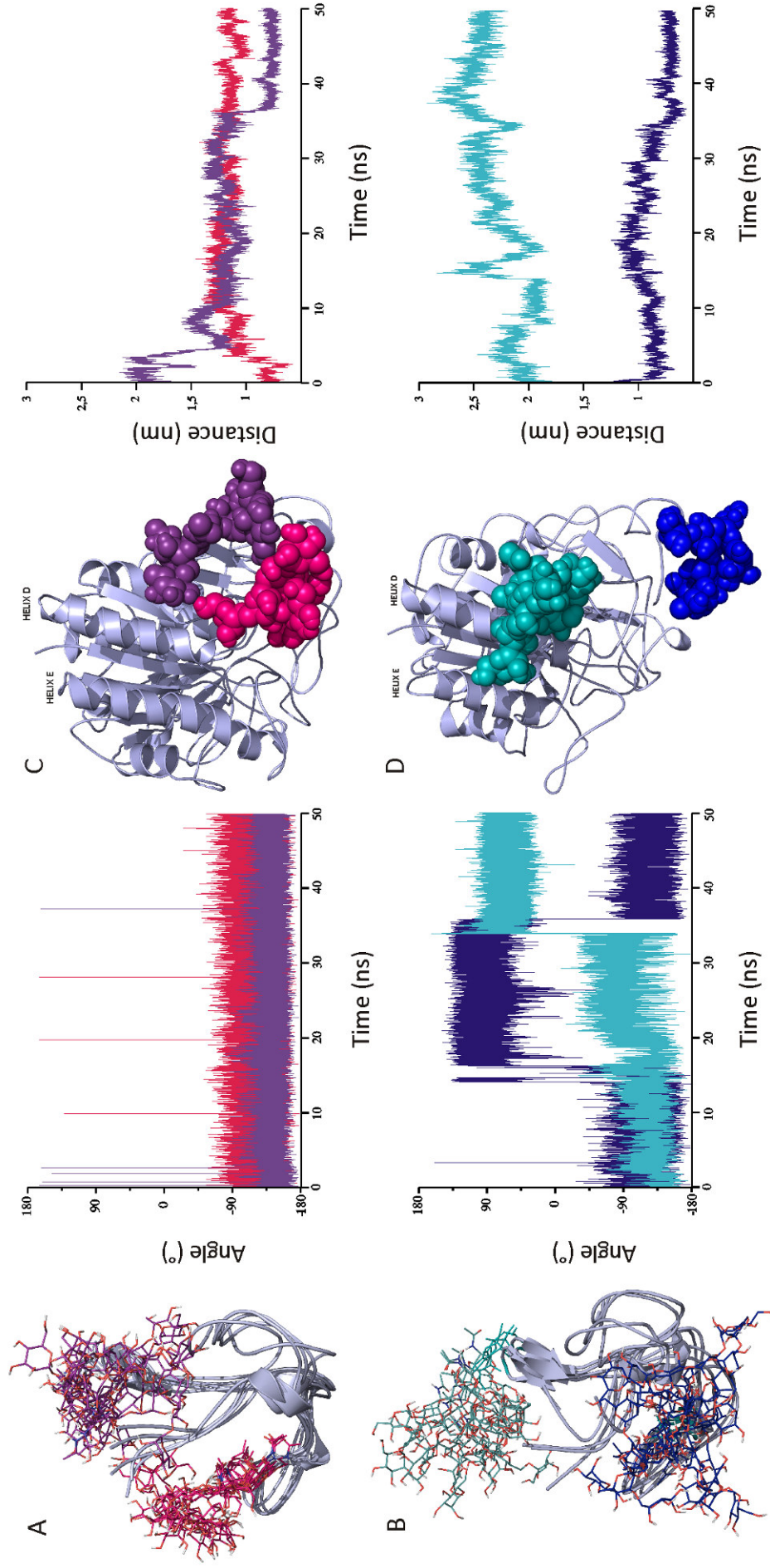


Figure 4. **N-linkage conformational analysis and glycan reorientation.** φ_{N-GL} measurements as a function of time for glycans at Asn158 (pink) and at Asn184 (purple) at fully-glycosylated ARSA system (4A), and for the glycans at Asn158 (light blue) and at Asn184 (dark blue) at ARSA-p.N350S system (4B). The glycans reorientation is presented in more details through of a frame superimposition of each 10ns of simulation evolution. Compound formation between the glycan at Asn158 (pink), the glycan at Asn184 (purple), and ARSA C-termini in fully glycosylated ARSA-pH~5 system (4C). Detachment between the glycan at Asn158 (light blue), the glycan at Asn184 (dark blue), and ARSA C-termini in ARSA-p.N350S system (4D). The distances between each glycan and the 18aa C-termini mass centers as a function of time is also presented.

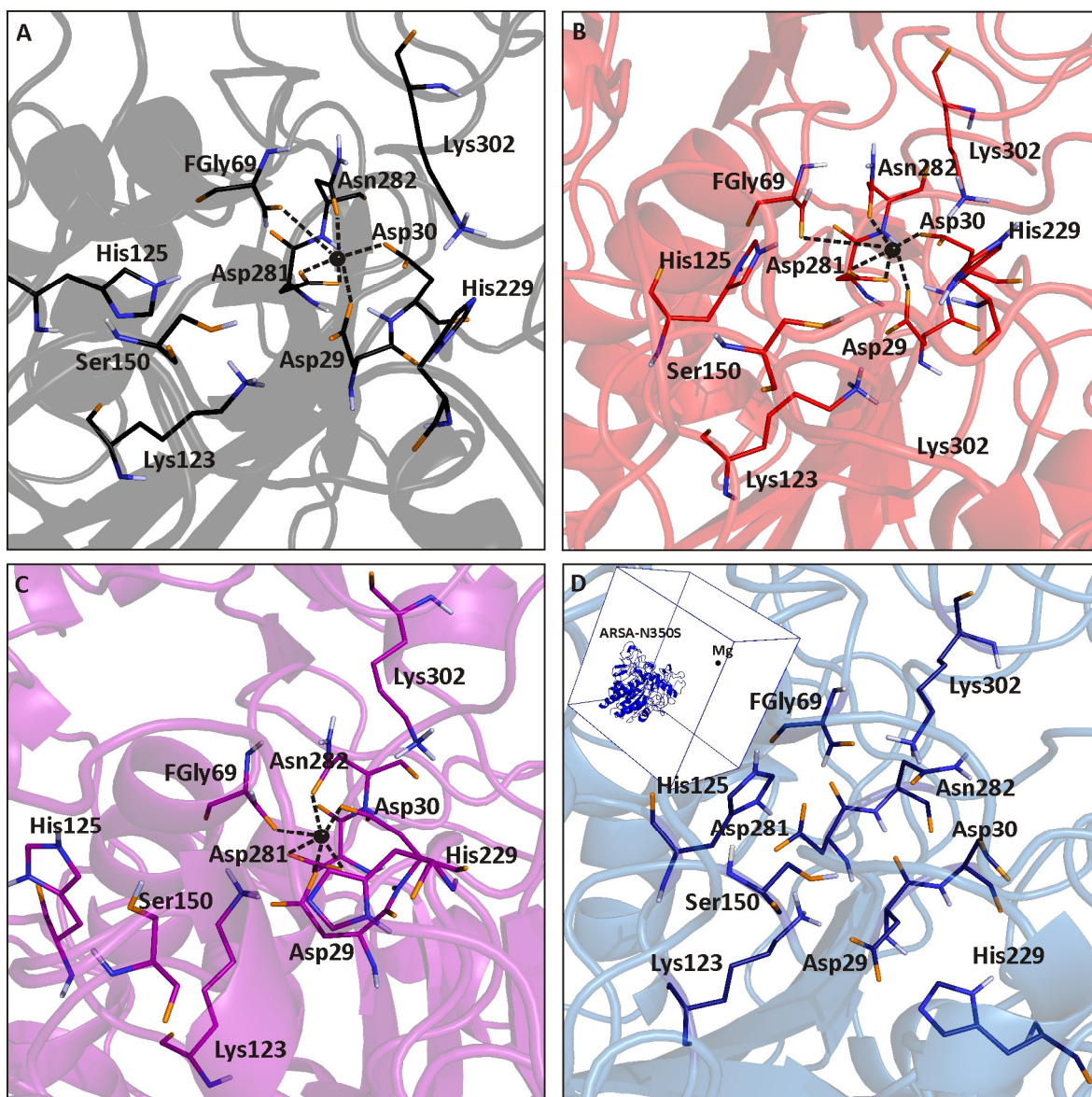
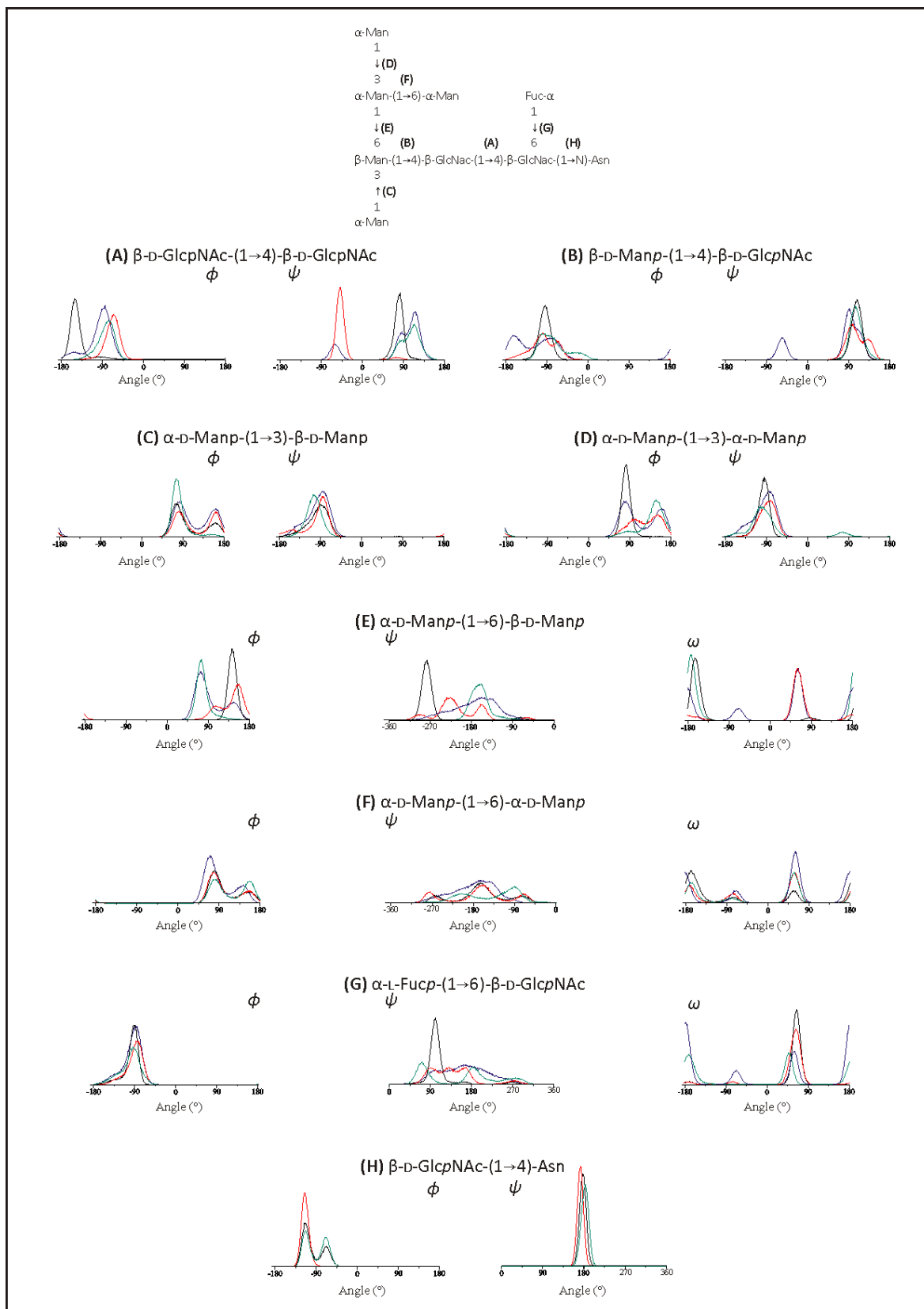


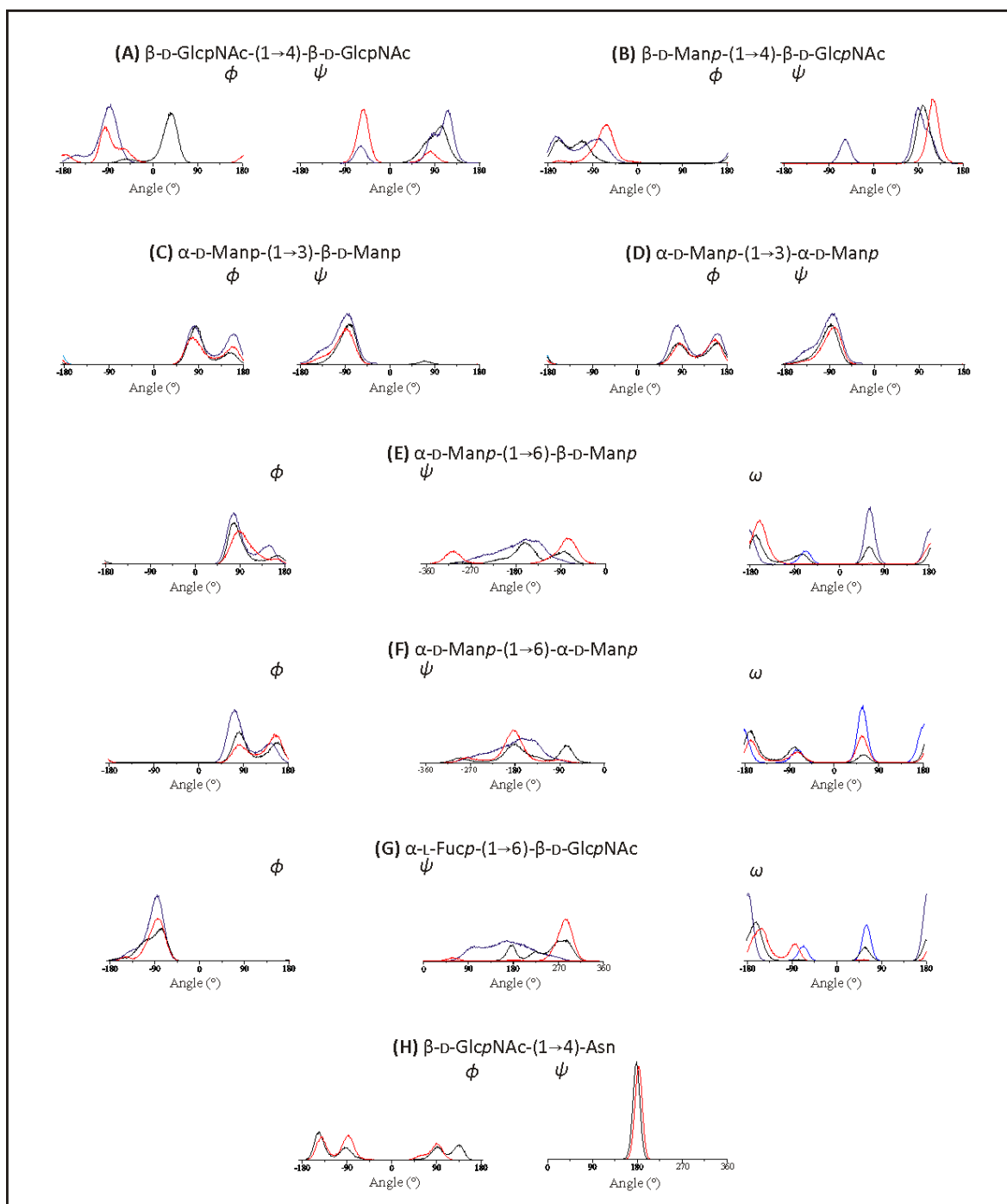
Figure 5. **Catalytic site analysis.** Amino acids forming the catalytic site of ARSA at non-glycosylated ARSA at pH~7 (A), non-glycosylated ARSA at pH~5 (B), fully-glycosylated ARSA at pH~5 (C), and poorly-glycosylated p.N350S-mutant at pH~5 (D) systems. Mg^{+2} octahedral coordination by Asp29, Asp30, Asp281, Asn282, and FGly69 is represented by dashed lines. At ARSA-p.N350S system is also presented the simulation box in order to illustrate the Mg^{+2} escape.

Table 1. Octahedral Mg⁺² coordination.

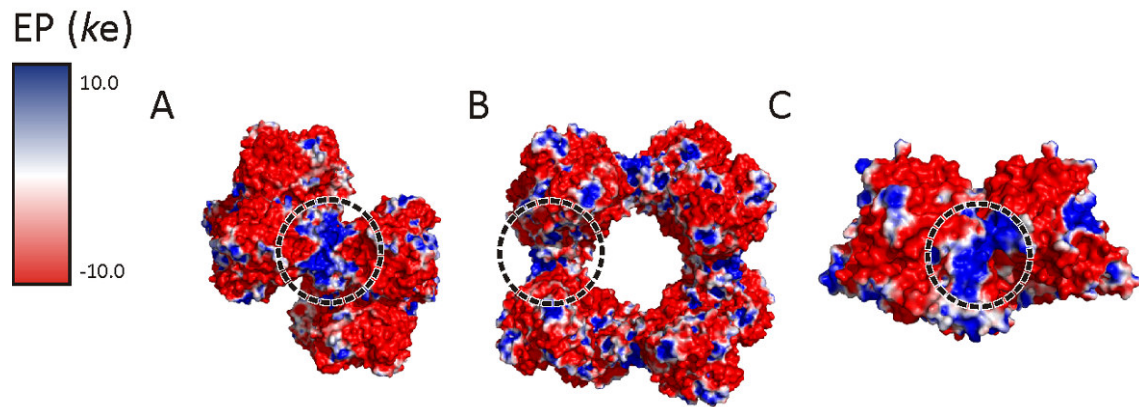
Oxygen donors		Distance from Mg ⁺² (Å)				
		1AUK	ARSA MD simulations			
			pH~7	pH~5	fully-glycosylated	p.N350S
Asp29	O β	4.30	2.07±0.01	2.60±0.09	2.79±0.09	28.54±12.94
Asp30	O δ	2.23	2.11±0.01	2.07±0.01	2.08±0.01	25.52±16.31
Asp281	O δ	2.66	2.08±0.01	2.07±0.01	2.06±0.01	29.30±14.31
	O β	2.19	2.09±0.01	2.11±0.01	2.06±0.01	29.37±14.12
Asn282	O β	2.59	2.19±0.02	2.18±0.02	2.02±0.01	26.58±16.20
FGly69	O γ	3.02	4.21±0.05	4.10±0.06	2.01±0.01	28.06±16.02



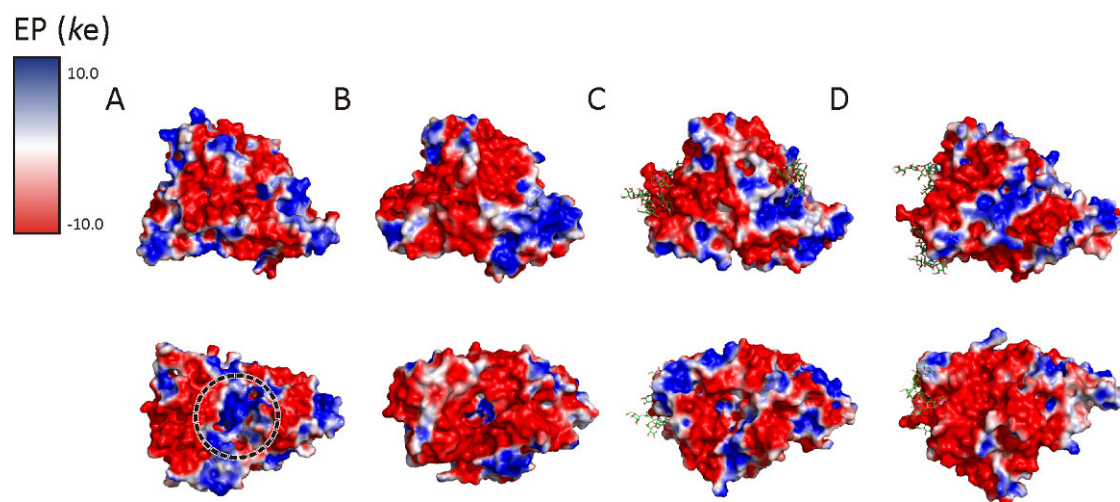
Supplementary Figure 1. **Distribution of the dihedral angles from fully-glycosylated ARSA, as obtained from MD simulations.** The ARSA glycans at Asn158, (black), Asn184 (red) and Asn350 (green) are shown together with the conformational patterns associated with respective disaccharides isolated in water (blue). The A-H letters indicate the relation of a distribution plots to its location to its location on the respective disaccharide unit from the complete glycan.



Supplementary Figure 2. **Distribution of the dihedral angles from poorly-glycosylated ARSA. N350S as obtained from MD simulations.** The ARSA glycans at Asn158 (black) and Asn184 (red) are shown together with the conformational patterns associated with respective disaccharides isolated in water (blue). The A-H letters indicate the relation of a distribution plots to its location to its location on the respective disaccharide unit from the complete glycan.



Supplementary Figure 3. **Electrostatic potential distribution of association interfaces in the final frame simulation.** Lateral view of ARSA octamer of non-glycosylated ARSA at pH~5 (A). Frontal view of ARSA octamer of non-glycosylated ARSA at pH~5 (B). Lateral view of ARSA dimer of non-glycosylated ARSA at pH~5 (C). The circled positive charged regions are the octameric interface, and the dimeric interface, respectively.



Supplementary Figure 4. **Electrostatic potential distribution of final frames simulation.** Frontal and inferior view of ARSA monomer of non-glycosylated ARSA at pH~7 (A), non-glycosylated ARSA at pH~5 (B), fully-glycosylated ARSA at pH~5 (C), and ARSA-p.N350S at pH~5 (D). The circulated positive charged region is the catalytic pocket.

REFERENCES

- 1 Galperin, M. Y., and Jedrzejak, M. J. (2001) Conserved core structure and active site residues in alkaline phosphatase superfamily enzymes, *Proteins* 45, 318-324.
- 2 Ghosh, D. (2007) Human sulfatases: a structural perspective to catalysis, *Cell Mol Life Sci* 64, 2013-2022
- 3 Lukatela, G., Krauss, N., Theis, K., Selmer, T., Gieselmann, V., von Figura, K., and Saenger, W. (1998) Crystal structure of human arylsulfatase A: the aldehyde function and the metal ion at the active site suggest a novel mechanism for sulfate ester hydrolysis, *Biochemistry* 37, 3654-3664.
- 4 Neufeld, E.F., and Muenzer, J. The mucopolysaccharidoses. In: Scriver CR, Beaudet A.L., Sly W.S., Valle D., Childs B., Kinzler K.E., and Vogelstein B. (eds) *The Metabolic and Molecular Bases of Inherited Disease*, 8th edition. McGraw-Hill, New York, pp 5833-5843.
- 5 Von Figura K., Gieselmann V., and Jaeken J. (2001) Metachromatic Leukodystrophy. In: Scriver C.R., Beaudet A.L., Sly W.S., Valle D., Childs B., Kinzler K.E., and Vogelstein B. *The Metabolic and Molecular Bases of Inherited Disease*, 8th edition. McGraw-Hill, New York, 5833-5843.
- 6 Gieselmann, V., Polten, A., Kreysing, J., and von Figura, K. (1989) Arylsulfatase A pseudodeficiency: loss of a polyadenylation signal and N-glycosylation site, *Proc Natl Acad Sci U S A* 86, 9436-9440.
- 7 Stein, C., Gieselmann, V., Kreysing, J., Schmidt, B., Pohlmann, R., Waheed, A., Meyer, H. E., O'Brien, J. S., and von Figura, K. (1989) Cloning and expression of human arylsulfatase A, *J Biol Chem* 264, 1252-1259.

- 8 Laidler, P., and Litynska, A. (1997) Arylsulfatase A from human placenta possesses only high mannose-type glycans, *Int J Biochem Cell Biol* 29, 475-483.
- 9 Dierks, T., Dickmanns, A., Preusser-Kunze, A., Schmidt, B., Mariappan, M., von Figura, K., Ficner, R., and Rudolph, M. G. (2005) Molecular basis for multiple sulfatase deficiency and mechanism for formylglycine generation of the human formylglycine-generating enzyme, *Cell* 121, 541-552.
- 10 Vagedes, P., Saenger, W., and Knapp, E. W. (2002) Driving forces of protein association: the dimer-octamer equilibrium in arylsulfatase A, *Biophys J* 83, 3066-3078.
- 11 Mitsunaga-Nakatsubo K., Kusunoki, S., Kawakami, H., Akasaka, K., and Akimoto, Y. (2009) Cell-surface arylsulfatase A and B on sinusoidal endothelial cells, hepatocytes, and Kupffer cells in mammalian livers, *Med Mol Morphol* 42, 63–69.
- 12 Drickamer, K., and Taylor, M. E. (1998) Evolving views of protein glycosylation, *Trends Biochem Sci* 23, 321-324.
- 13 Wormald, M. R., and Dwek, R. A. (1999) Glycoproteins: glycan presentation and protein-fold stability, *Structure* 7, R155-160.
- 14 Imperiali, B. (1997) Protein glycosylation: the clash of the titans, *Acc Chem Res* 30, 452-459.
- 15 Imperiali, B., and O'Connor, S. E. (1999) Effect of N-linked glycosylation on glycopeptide and glycoprotein structure, *Curr Opin Chem Biol* 3, 643-649.
- 16 Sommerlade, H. J., Selmer, T., Ingendoh, A., Gieselmann, V., von Figura, K., Neifer, K., and Schmidt, B. (1994) Glycosylation and phosphorylation of arylsulfatase A, *J Biol Chem* 269, 20977-20981.

- 17 Gieselmann, V., Schmidt, B., and von Figura, K. (1992) In vitro mutagenesis of potential N-glycosylation sites of arylsulfatase A. Effects on glycosylation, phosphorylation and intracellular sorting, *J Biol Chem* 267, 13262-13266.
- 18 Harvey, J. S., Carey, W. F., and Morris, C. P. (1998) Importance of the glycosylation and polyadenylation variants in metachromatic leukodystrophy pseudodeficiency phenotype, *Hum Mol Genet* 7, 1215-1219.
- 19 Fluharty, A.L., Meek, W.E., and Kihara, H. (1983) Pseudo arylsulfatase A deficiency: evidence for a structurally altered enzyme, *Biochem Biophys Res Commun* 112, 191-197.
- 20 Waheed, A., Steckel, F., Hasilik, A., and von Figura, K. (1983) Two allelic forms of human arylsulfatase A with different numbers of asparagine-linked oligosaccharides, *Am J Hum Genet* 35, 228-233.
- 21 Nakamura, M., Gasa, S., and Makita, A. (1984) Arylsulfatase A from normal human lung and lung tumors showed different patterns of Microheterogeneity, *J Biochem* 96, 207-203.
- 22 Bohne, A., and von der Lieth, C. W. (2002) Glycosylation of proteins: a computer based method for the rapid exploration of conformational space of N-glycans, *Pac Symp Biocomput*, 285-296.
- 23 Petrescu, A. J., Wormald, M. R., and Dwek, R. A. (2006) Structural aspects of glycomes with a focus on N-glycosylation and glycoprotein folding, *Curr Opin Struct Biol* 16, 600-607.

- 24 Petrescu, A. J., Petrescu, S. M., Dwek, R. A., and Wormald, M. R. (1999) A statistical analysis of N- and O-glycan linkage conformations from crystallographic data, *Glycobiology* 9, 343-352.
- 25 Verli, H., and Guimaraes, J. A. (2004) Molecular dynamics simulation of a decasaccharide fragment of heparin in aqueous solution, *Carbohydr Res* 339, 281-290.
- 26 Verli, H., and Guimaraes, J. A. (2005) Insights into the induced fit mechanism in antithrombin-heparin interaction using molecular dynamics simulations, *J Mol Graph Model* 24, 203-212.
- 27 Becker, C. F., Guimaraes, J. A., and Verli, H. (2005) Molecular dynamics and atomic charge calculations in the study of heparin conformation in aqueous solution, *Carbohydr Res* 340, 1499-1507.
- 28 Pol-Fachin, L., and Verli, H. (2008) Depiction of the forces participating in the 2-O-sulfo-alpha-L-iduronic acid conformational preference in heparin sequences in aqueous solutions, *Carbohydr Res* 343, 1435-1445.
- 29 Pol-Fachin, L., Fernandes, C. L., and Verli, H. (2009) GROMOS96 43a1 performance on the characterization of glycoprotein conformational ensembles through molecular dynamics simulations, *Carbohydr Res* 344, 491-500.
- 30 Fernandes, C. L., Sachett, L. G., Pol-Fachin, L., and Verli, H. (2010) GROMOS96 43a1 performance in predicting oligosaccharide conformational ensembles within glycoproteins, *Carbohydr Res* 345, 663-671.
- 31 IUPAC-IUB (1983) Commission on Biochemical Nomenclature, *Pure Appl Chem* 55.

- 32 Schaftenaar, G., and Noordik, J. H. (2000) Molden: a pre- and post-processing program for molecular and electronic structures, *J Comput Aided Mol Des* 14, 123-134.
- 33 Humphrey, W., Dalke, A., and Schulten, K. (1996) VMD: visual molecular dynamics, *J Mol Graph* 14, 33-38, 27-38.
- 34 Guex, N., and Peitsch, M. C. (1997) SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling, *Electrophoresis* 18, 2714-2723.
- 35 DeLano, W.L. (2002) The PyMOL Molecular Graphics System, DeLano Scientific, San Carlos, CA, USA.
- 36 Kabsch, W., and Sander, C. (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features, *Biopolymers* 22, 2577-2637.
- 37 Laskowski, R. A., McArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) PROCHECK: a program to check the stereochemical quality of protein structures, *J M J Appl Crystallog.* 26, 283-291.
- 38 Van Der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A. E., and Berendsen, H. J. (2005) GROMACS: fast, flexible, and free, *J Comput Chem* 26, 1701-1718.
- 39 Van Gunsteren, W. F. B., Eising, A. A., Hünenberger, P. H., Krüger, P., Mark, A. E., Scott, W. R. P., and Tironi, I. G. (1996) *Biomolecular Simulation: The Gomos96 Manual and User Guide*, ETH Zürich, BIOMOS b.v., Zürich, Groningen
- 40 Schuttelkopf, A. W., and van Aalten, D. M. (2004) PRODRG: a tool for high-throughput crystallography of protein-ligand complexes, *Acta Crystallogr D Biol Crystallogr* 60, 1355-1363.

- 41 Lutteke, T., Bohne-Lang, A., Loss, A., Goetz, T., Frank, M., and von der Lieth, C. W. (2006) GLYCOSCIENCES.de: an Internet portal to support glycomics and glycobiology research, *Glycobiology* 16, 71R-81R.
- 42 Berendsen, H. J. C., Grigera, J. R., and Straatsma, T. P. (1987) The missing term in effective pair potentials, *J Phys Chem* 91, 6269-6271.
- 43 Dolinsky, T. J., Czodrowski, P., Li, H., Nielsen, J. E., Jensen, J. H., Klebe, G., and Baker, N. A. (2007) PDB2PQR: Expanding and upgrading automated preparation of biomolecular structures for molecular simulations. *Nucleic Acids Res* 35, W522-525.
- 44 Hess, B., Bekker, H., Berendsen, H. J. C., and Fraaije, J. G. E. M. (1997) LINCS: a linear constraint solver for molecular simulations, *J Comput Chem* 18, 1463-1472.
- 45 Darden, T., York, D., and Pedersen, L. (1993) Particle mesh Ewald an Nlog(n) method for Ewald sums in large systems, *J Chem Phys* 98, 10089-10092.
- 46 Berendsen, H. J. C., Postma, J. P. M., DiNicola, A., and Haak, J. R. (1984) Molecular dynamics with coupling to an external bath, *J Chem Phys* 81, 3684-3690.
- 47 Van Veen, H. A., Geerts, M. E., van Berkel, P. H., and Nuijens, J. H. (2004) The role of N-linked glycosylation in the protection of human and bovine lactoferrin against tryptic proteolysis, *Eur J Biochem* 271, 678-684.
- 48 Schenk, M., Koppisetty, C. A., Santos, D. C., Carmona, E., Bhatia, S., Nyholm, P. G., and Tanphaichitr, N. (2009) Interaction of arylsulfatase-A (ASA) with its natural sulfoglycolipid substrates: a computational and site-directed mutagenesis study, *Glycoconj J* 26, 1029-1045.

- 49 Francis, G. S., Bonni, A., Shen, N., Hechtman, P., Yamut, B., Carpenter, S., Karpati, G., and Chang, P. L. (1993) Metachromatic leukodystrophy: multiple nonfunctional and pseudodeficiency alleles in a pedigree: problems with diagnosis and counseling, *Ann Neurol* 34, 212-218.
- 50 Shen, N., Li, Z. G., Waye, J. S., Francis, G., and Chang, P. L. (1993) Complications in the genotypic molecular diagnosis of pseudo arylsulfatase A deficiency, *Am J Med Genet* 45, 631-637.
- 51 Barth, M. L., Fensom, A., and Harris, A. (1994) The arylsulphatase A gene and molecular genetics of metachromatic leucodystrophy, *J Med Genet* 31, 663-666.
- 52 Hess, B., Kafert, S., Heinisch, U., Wenger, D. A., Zlotogora, J., and Gieselmann, V. (1996) Characterization of two arylsulfatase A missense mutations D335V and T274M causing late infantile metachromatic leukodystrophy, *Hum Mutat* 7, 311-317.
- 53 Schestag, F., Yaghootfam, A., Habetha, M., Poeppel, P., Dietz, F., Klein, R. A., Zlotogora, J., and Gieselmann, V. (2002) The functional consequences of mis-sense mutations affecting an intra-molecular salt bridge in arylsulphatase A, *Biochem J* 367, 499-504.
- 54 Lugowska, A., Ploski, R., Wlodarski, P., and Tylki-Szymanska, A. Molecular bases of metachromatic leukodystrophy in Polish patients, *J Hum Genet* 55, 394-396.
- 55 Hendsch, Z. S., and Tidor, B. (1994) Do salt bridges stabilize proteins? A continuum electrostatic analysis, *Protein Sci* 3, 211-226
- 56 Honig, B., and Nicholls, A. (1995) Classical electrostatics in biology and chemistry, *Science* 268, 1144-1149.

- 57 Honig, B., and Yang, A. S. (1995) Free energy balance in protein folding, *Adv Protein Chem* 46, 27-58.
- 58 Imberty, A., and Perez, S. (1995) Stereochemistry of the N-glycosylation sites in glycoproteins, *Protein Eng* 8, 699-709.
- 59 Davis, J. T., Hirani, S., Bartlett, C., and Reid, B. R. (1994) ¹H NMR studies on an Asn-linked glycopeptides, GlcNAc-1 C2-N2 bond is rigid in H₂O, *J Biol Chem* 269, 3331-3338.
- 60 Waldow, A., Schmidt, B., Dierks, T., von Bulow, R., and von Figura, K. (1999) Amino acid residues forming the active site of arylsulfatase A. Role in catalytic activity and substrate binding, *J Biol Chem* 274, 12284-12288.

6. DISCUSSÃO

6.1. CARACTERIZAÇÃO GENOTÍPICA

As doenças de lisossômicas de depósito (DLD), dentre as quais está a MLD, formam um grupo heterogêneo de pelo menos 41 doenças metabólicas herdadas, em sua maioria de forma recessiva, caracterizadas pelo prejuízo de vias catabólicas intralisossômicas. Com exceção de estudos da prevalência de DLD específicas em algumas populações restritas, dados de prevalência conjunta das diversas DLD são escassos na literatura, destacando-se apenas os relatos da Holanda (Poorthuis *et al*, 1999), Austrália (Meiklee *et al*, 1999) e Portugal (Pinto *et al*, 2004). Em todos esses países a MLD está entre as mais frequentes DLD.

As DLD correspondem a 59,8% dos EIM diagnosticados no Laboratório de Referência para Erros Inatos do Metabolismo do Hospital de Clínicas de Porto Alegre (LEIM-HCPA), centro de referência para esse diagnóstico (Coelho *et al*, 1997).

O diagnóstico de MLD depende da detecção da deficiência da atividade de ARSA *in vitro*, entretanto, esse tipo de análise bioquímica pode levar a erros devido à considerável variabilidade de atividade enzimática entre indivíduos e à presença de alelos PD-ARSA, responsáveis por uma atividade enzimática residual de cerca de 5% a 15%. Nesses casos, a distinção entre MLD ou PD-ARSA requer a análise de excreção urinária de sulfatídeos. A excreção urinária de sulfatídeos é aumentada em pacientes com MLD, mas normal em indivíduos pseudodeficientes.

Ademais, a atividade normal de ARSA pode ser observada em alguns pacientes com sintomas comuns à MLD que tenham deficiência de SAP-B, nesses pacientes, entretanto, a excreção urinária de sulfatídeos também está aumentada (Biffi *et al*, 2008).

Nesse panorama, a pesquisa de mutações no gene da ARSA constitui uma ferramenta adicional importante para o diagnóstico preciso de MLD. Nesse trabalho, a eficácia da pesquisa de mutações para o diagnóstico de MLD foi comprovada pelo sucesso na identificação de 96% dos alelos relacionados à MLD (46 dos 48 alelos pesquisados). O fato de dois dos alelos não terem sido identificados pode ser explicado pela adoção nesse trabalho de uma cobertura restrita à região codificante do gene da ARSA, desse modo porções não codificantes que podem conter mutações igualmente deletérias, como introns e regiões regulatórias, não foram pesquisadas.

Além de sua importância para o diagnóstico, a caracterização genotípica de pacientes é importante para o estudo de associações genótipo-fenótipo. No caso da MLD, associações desse tipo são estudadas em termos da relação entre a atividade enzimática de ARSA e a gravidade dos fenótipos associados. Como visto anteriormente, a divisão dos alelos deletérios quanto à atividade enzimática permite distinção em dois subgrupos: alelos nulos, responsáveis por uma perda completa da atividade enzimática e alelos não-nulos, que permitem uma atividade enzimática residual. Esses tipos de alelos podem ser combinados em três categorias de genótipos que são associadas à gravidade dos sintomas e, conseqüentemente, aos subtipos clínicos de MLD. Indivíduos homocigotos para alelos nulos comumente apresentam a forma mais severa, representada pelo subtipo clínico infantil de MLD, enquanto indivíduos heterocigotos compostos para um alelo nulo e um alelo não nulo comumente apresentam uma forma mais atenuada de MLD, representada pelo subtipo juvenil, e indivíduos homocigotos para alelos não nulos frequentemente apresentam a forma mais branda de MLD, representada pelo subtipo adulto.

Essa associação genótipo-fenótipo é ainda mais estrita para os alelos mais freqüentes em que: indivíduos homocigotos para o alelo c.459+1A>G comumente apresentam MLD infantil, enquanto indivíduos heterocigotos compostos para c.459+1A>G e p.P426L demonstram manifestações clínicas comuns à MLD juvenil e indivíduos homocigoto o alelo p.P426L comumente manifestam de MLD adulta.

Nesse estudo, a alta freqüência da mutação c.459+1A>G está associada à prevalência de casos de MLD infantil, como previsto pela associação genótipo fenótipo estabelecida para MLD. Entretanto, também foi possível notar a existência de casos que fugiram à regra, como um indivíduo homocigoto para o alelo p.P426L que começou a manifestar os sintomas ao redor de 12 anos de idade, ou seja, que não apresentou a forma adulta da doença, mas sim a forma juvenil. Apesar disso, esse paciente foi o que apresentou uma idade de início mais avançada entre todos os demais pacientes.

Algumas outras mutações, além das supracitadas, também já haviam sido caracterizadas em termos de sua contribuição para a perda de atividade enzimática, o que permitiu categorizá-las em alelos nulos ou não-nulos. Nesse estudo, para 17 dos 27 pacientes foi possível distinguir os três tipos de genótipos completos quanto à combinação entre alelos nulos e não nulos e sua relação com a atividade enzimática e subtipos clínicos de MLD. Os demais pacientes apresentaram pelo menos uma mutação não caracterizada

em termos bioquímicos e, portanto, não puderam ter seus genótipos categorizados nesses termos.

Tabela 4. Classificação genotípica dos pacientes quanto à combinação de alelos nulos e não nulos.

Pacientes*	Idade de início (meses)	Atividade enzimática (nmol/h/mg)	Genótipo	Combinação de alelos
1	12	0,9	[c.459+1G>A] + [c.459+1G>A]	nulo/nulo
2	18	0,4	[c.459+1G>A] + [c.459+1G>A]	nulo/nulo
3	3	nd	[g.1190_1191insC] + [g.1190_1191insC]	nulo/nulo
5	3	nd	[c.459+1G>A] + [c.459+1G>A]	nulo/nulo
8	36	0,1	[c.103_110del8] + [c.459+1G>A]	nulo/nulo
10	60	0,5	[c.459+1G>A] + [p.P426L]	nulo/não nulo
12	24	1,0	[c.459+1G>A] + [c.459+1G>A]	nulo/nulo
13	38	0,4	[c.103_110del8] + [c.459+1G>A]	nulo/nulo
14	36	1,0	[c.459+1G>A] + [c.459+1G>A]	nulo/nulo
15	24	0,8	[c.103_110del8 + g.1190_1191insC] + [g.1190_1191insC]	nulo/nulo
16	5	0,8	[g.2409DelC] + [g.2409DelC]	nulo/nulo
17	141	1,1	[p.P426L] + [p.P426L]	não nulo/não nulo
18	72	0,7	[c.103_110del8] + [p.I179S]	nulo/não nulo
19	72	1,0	[c.848-1G>A] + [p.P426L]	nulo/não nulo
20	48	0,3	[c.459+1G>A] + [c.459+1G>A]	nulo/nulo
21	24	0,2	[c.459+1G>A] + [c.459+1G>A]	nulo/nulo
25	8	nd	[c.459+1G>A + p.D335V] (+) [p.D335V]	nulo/nulo

nd = atividade enzimática de ARSA não-detectada.

*A numeração utilizada para os pacientes é referente àquela utilizada no artigo “*Genotypic Characterization of Brazilian Metachromatic leukodystrophy Patients*”.

De modo geral, houve um ajuste parcial entre os parâmetros de idade de início, atividade enzimática e genótipos dos pacientes. Todos os 13 pacientes que apresentavam MLD infantil também eram homocigotos para alelos nulos e 61,5% dos mesmos (8 de 13 pacientes) apresentaram-se distribuídos entre a metade inferior dos valores de atividade de ARSA, sendo a variação total da atividade enzimática de 0 a 1,1 nmol/h/mg. Dentre os 4 pacientes que apresentavam MLD juvenil, 3 eram heterocigotos para um alelo nulo e um alelo não nulo, como o esperado pela associação genótipo fenótipo já estabelecida, mas, como já referido, um deles era homocigoto para alelos não nulos, o que estaria ligado à manifestação de MLD adulta. Com respeito à atividade enzimática dos pacientes com MLD juvenil, 75,0% (3 de 4 pacientes) apresentaram-se distribuídos na metade superior dos valores de atividade de ARSA, sendo o paciente homocigoto para alelos não nulos o que apresentou o maior dos valores listados (1,1 nmol/h/mg prot) [Fig. 3].

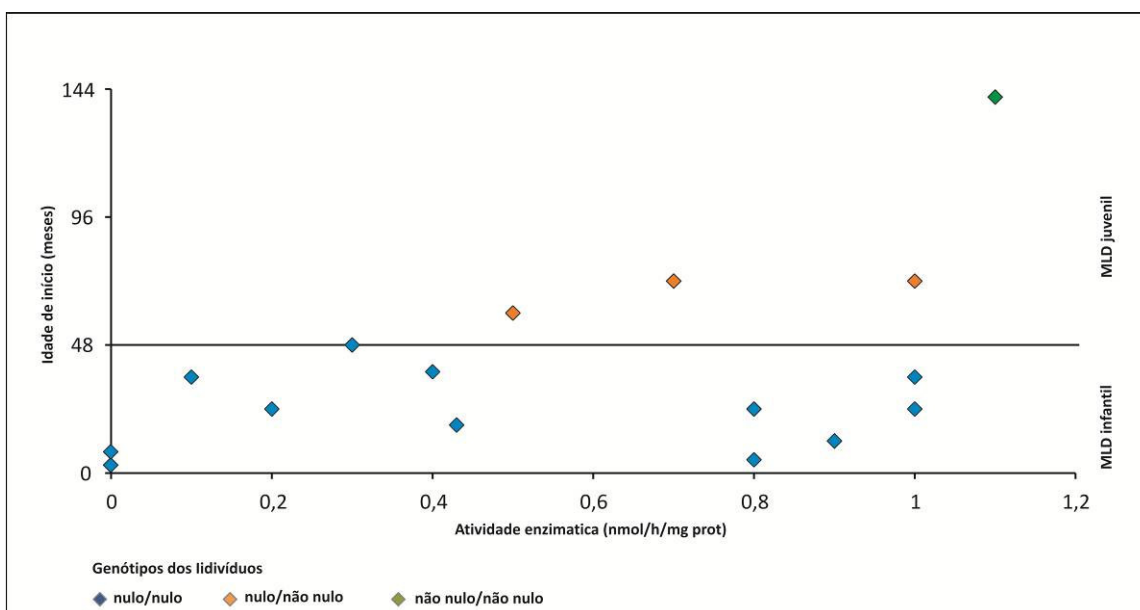


Figura 3. **Análise de associação genótipo-fenótipo.** Associação entre idade de início dos sintomas, atividade enzimática da ARSA e genótipos completos para os 17 pacientes em que foi possível estabelecer a combinação entre alelos nulos e não nulos. Os pacientes 3 e 4 (Tabela 1) aparecem sobrepostos.

Dentre as mutações já caracterizadas em termos bioquímicos, está p.I179S, descrita como a terceira mutação mais frequente em pacientes europeus e comumente relacionada à MLD juvenil. Interessantemente, entre os alelos identificados nesse trabalho, essa mutação ocorre uma única vez, ao passo que as mutações g.1190_1191insC e c.103_110del8, duas mutações anteriormente descritas como raras em outras populações, alcançam frequências mais relevantes, ocupando os postos de terceira e segunda mutações mais frequentes, respectivamente. Os 5 pacientes que apresentaram pelo menos uma dessas mutações estão geograficamente separados e espalhados e em diversos estados do Brasil (Paraná, São Paulo, Amazonas e Rio Grande do Sul), de modo que não se tratam de integrantes de uma comunidade endogâmica e se pode descartar esse fato como causa para o compartilhamento de mutações raras.

Dentre os países onde as mutações g.1190_1191insC e c.103_110del8 foram previamente identificadas, estão os países ibéricos. Dos vários povos que colonizaram o território brasileiro desde 1500, os povos ibéricos contaram com maior número de imigrantes. Desse modo, é possível especular sobre o fluxo desses alelos mutantes durante

o período inicial de colonização do Brasil, entretanto um estudo haplotípico mais detalhado seria necessário para apoiar tal afirmação.

No total, foram identificadas 11 diferentes mutações raras com baixa frequência na amostra, dentre as quais 2 foram descritas primeiramente nesse trabalho e são propostas estarem relacionadas a manifestação de MLD: p.S44P e p.P824S. Ambas foram encontradas em pacientes com MLD infantil e preditas como provavelmente deletérias pelo programa de análise *in silico* PolyPhen. A mutação p.S44P foi encontrada em um alelo em um paciente também heterozigoto para a mutação p.P426L, como visto, comumente associada à manifestações mais brandas de MLD. É possível sugerir, então, que a manifestação de um fenótipo mais severo (idade de início da manifestação dos sintomas de 5 meses e atividade enzimática de ARSA de 0,39 nmol/h/mg) esteja mais ligada aos efeitos perturbadores da mutação p.S44P, que pode representar um alelo nulo.

6.2. O PAPEL ESTRUTURAL DA GLICOSILAÇÃO DE ARSA

Entre os vários papéis desempenhados por carboidratos, adquiridos evolutivamente, estão: (i) a função primordial como constituinte da superfície celular importante na manutenção de interações celulares, mais comum em procariotos, (ii) o mais recente envolvimento na marcação intracelular através de elaborações terminais que funcionam como marcadores de reconhecimento para receptores endógenos, uma aquisição de seres multicelulares, (iii) o papel estrutural intermediário no dobramento proteico, mais encontrado em eucariotos que em procariotos, (iv) a função colateral de fator protetor contra ataques proteolíticos, isolando porções da superfície proteica e (v) a função de fator estabilizante de grandes porções da cadeia central de proteínas, promovendo uma maior compactação da estrutura secundária protéica (Drickamer *et al*, 1998; Wormald *et al*, 1999; Imperiali *et al*, 1999).

Somente recentemente, o papel da glicosilação na manutenção estrutural de glicoproteínas tem sido verificado, entretanto as dificuldades metodológicas implicadas nesse tipo de estudo por meio das ferramentas usuais como cristalografia de raios-X e espectroscopia por ressonância magnética constituem um entrave a esse tipo de abordagem. Nesse cenário, a DM surge como uma ferramenta acessível e robusta para o estudo da importância da glicosilação na manutenção conformacional de glicoproteínas.

Embora o desenvolvimento e aprimoramento de campos de força para carboidratos estejam crescendo gradativamente nos últimos anos, a adequação desses parâmetros no estudo de glicoconjugados, tais como glicoproteínas, anda se encontra bastante incipiente. Isso pode ser evidenciado pelo pequeno número de estudos envolvendo DM de glicoproteínas (Pol-fachin *et al*, 2009), não ultrapassando uma dezena de trabalhos.

O polimorfismo p.N350S, constituinte do alelo de PD-ARSA, é responsável pela perda de uma das cadeias laterais de glicanos da ARSA, levando a uma perda de 57% da atividade *in vitro* de ARSA (Harvey *et al*, 1998). A associação de p.N350S com alguns transtornos psiquiátricos (como o alcoolismo) e a capacidade de causar sozinho pseudodeficiência de ARSA têm sido amplamente discutidas (Park *et al*, 1999).

Além dessas associações, existem achados que apontam para a influência do alelo de pseudodeficiência, do qual o polimorfismo p.N350S é integrante, na evolução clínica da MLD, levando a quadros mais severos (Lugowska *et al*, 2000). Entre os 27 pacientes analisados nesse estudo, 5 apresentaram pelo menos um dos polimorfismos constituintes do alelo de PD-ARSA, entre os quais, 2 pacientes apresentaram alelo PD-ARSA, sendo um paciente homozigoto e outro heterozigoto para o alelo de PD-ARSA.

Tendo em vista os aspectos já citados, esse trabalho propôs o estudo detalhado do papel estrutural possivelmente desempenhado pelos estados de glicosilação conhecidos da ARSA (a tripla glicosilação funcional e o estado deficientemente glicosilado representado pelo polimorfismo p.N350S) através de uma abordagem computacional, a fim de ajudar na construção de uma explicação sólida para o defeito catalítico da ARSA encontrado em indivíduos pseudodeficientes e na avaliação da possível relação entre essa alteração e MLD.

No mais, o conhecimento acerca da função conformacional do padrão normal de glicosilação da ARSA está na base do entendimento de como modificações anômalas dos glicanos, como fosforilação, sialiação e sulfatação, podem estar relacionadas ao aparecimento de cânceres, como previamente descrito e, não obstante, podem ajudar também na compreensão dos aspectos gerais da importância da glicosilação para a estabilidade de glicoproteínas.

Esse trabalho demonstrou, através da comparação entre as dinâmicas estruturais de ARSA não glicosilada e completamente glicosilada, que o estado de tripla glicosilação é responsável por um substancial ganho de estabilidade conformacional e manutenção de

estrutura secundária de ARSA, ao mesmo tempo em que o arranjo específico da glicanos, dispostas próximas a porções possivelmente expostas ao meio (a cauda C-terminal de ARSA e o loop onde está localizada uma ponte salina), indica o envolvimento das cadeias laterais de glicanos na proteção a ataques proteolíticos.

O papel estrutural da tripla glicosilação foi ainda verificado a partir da comparação entre ARSA completamente glicosilada e ARSA deficientemente glicosilada (p.N350S). A ausência de uma das três glicanos na ARSA mutante p.N350S promove não só ganho de flexibilidade geral do sistema como também a perda de estrutura secundária. Os efeitos perturbadores dessa ausência se dão em cascata, como resultado da reorientação geral da glicanos restantes que perdem suas ótimas interações com a cauda C-terminal de ARSA e passam a interagir com hélices formadoras do núcleo protéico. Esse novo padrão de interação, por sua vez, leva a perdas significativas de estrutura secundária que ecoam por todo o núcleo protéico e culminam com a desorientação de resíduos do sítio catalítico.

Desse modo a diminuição de 57% da atividade catalítica na ARSA mutante p.N350S pode estar ligada ao ganho de instabilidade conformacional resultante da perda de seu estado essencial de tripla glicosilação.

6.3. O PAPEL ESTRUTURAL DO pH PARA ARSA

Recentemente a certeza de que ARSA era uma enzima de distribuição apenas lisossômica foi abalada pela descoberta de ARSA como componente da superfície celular e da matriz extracelular de células endoteliais, hepatócitos e células de Kupffer do fígado de mamíferos. A ARSA funcionaria como componente estrutural tanto da superfície celular quanto da matriz extracelular, sendo capaz de ligar glicolípídeos sulfatados, mas não de hidrolisá-los, uma vez que o pH ótimo para catálise se situa ao redor de 4,5. Nesse esquema, ARSA teria um papel significativo para a morfologia das células onde ocorre e os efeitos desmielinizantes oriundos da deficiência de ARSA encontrados na MLD ocorreriam não só pelo depósito intralisossômico de sulfatídeo como também por mudanças na estrutura da matriz extracelular e pela perturbação da diferenciação celular (Mitsunaga *et al*, 2009).

Os efeitos do pH para a atividade de ARSA têm sido verificados em termos de sua da regulação da associação octamérica intralisossômica, mas as evidências supracitadas apontam para a necessidade de uma abordagem mais detalhada da importância de

mudanças de pH pra a dinâmica estrutural da ARSA, de modo a esclarecer quais as prováveis diferenças entre a ARSA lisossômica e a ARSA da superfície celular que se refletem em suas diferentes funções.

No presente estudo, avaliamos os efeitos do pH sobre a dinâmica estrutural de ARSA a partir da comparação entre ARSA monomérica não glicosilada em pH~5 (pH próximo ao lisossômico) e em pH~7 (pH próximo ao da matriz extracelular). Pôde ser verificado que a ARSA em pH~5 demonstrou maior estabilidade global que em pH~7, inclusive de uma importante ponte salina que ajuda na união entre os domínios de oligomerização e catalítico de ARSA e contribui para a manutenção da globularidade da enzima pelo seu correto dobramento, como já sugerido em estudos prévios. Desse modo a ARSA existente na superfície celular pode ser estruturalmente diferente da ARSA lisossômica no que diz respeito à estabilidade e globularidade.

7. CONCLUSÕES

- Duas novas mutações no gene da ARSA foram identificadas e propostas como estando relacionadas à manifestação de MLD: p.S44P e p.P824S.
- 96% do total de alelos pesquisados relacionados à MLD (46 dos 48 alelos pesquisados) foram completamente identificados.
- A mutação mais frequentemente encontrada entre os pacientes analisados nesse trabalho foi a c.459+1A>G.
- As mutações p.P426L e c.103_110del8 foram aquelas com a segunda mais elevada frequência, sendo que a c.103_110del8 foi descrita anteriormente como rara em outras populações.
- A tripla glicosilação é responsável por um substancial ganho de estabilidade conformacional e manutenção da estrutura secundária de ARSA.
- Os efeitos perturbadores resultantes da deficiência de glicosilação provocada pela mutação p.N350S podem levar à redução da atividade de ARSA por intermédio de alterações do padrão ótimo de interação das glicanos restantes e consequente instabilidade da ARSA.
- ARSA em pH~5 demonstrou maior estabilidade global que em pH~7 o que está diretamente ligado à sua função catalítica restrita à localização lisossomal.

8. REFERÊNCIAS BIBLIOGRÁFICAS

- Artigalás, O., Lagranha, V. L., Saraiva-Pereira, M. L., Burin, M. G., Lourenco, C. M., van der Linden, H., Santos Jr, M. L., Rosemberg, S., Steiner, C. E., Kok, F., de Souza, C. F., Jardim, L. B., Giugliani, R., Schwartz, I. V. (2010) Clinical and biochemical study of 29 Brazilian patients with metachromatic leukodystrophy, *J. Inherit. Metab. Dis.*
- Austin, J.H. (1959) Metachromatic sulfatides in cerebral white matter and kidney. *Proc Soc Exp Biol Med* 100, 361-364.
- Barth, M.L., Fenson, A., Harris, A. (1993) Missense mutations in the arylsulfatase A genes of metachromatic leukodystrophy patients. *Hum Mol Genet* 2, 2117-2121.
- Barth, M.L., Fenson, A., Harris A. (1994) The arylsulfatase A gene and molecular genetics of metachromatic leucodystrophy. *J Med Genet* 31, 663-666.
- Berger, J., Gmach, M., Fae, I., Molzer, B., Bernheimer, H. (1996) A new polymorphism of arylsulfatase A within the coding region, *Hum Genet* 98, 348-350.
- Berger, J., Loschl, B., Bernheimer, H., Lugowska, A., Tylki-Szymanska, A., Gieselmann, V., Molzer, B. (1997) Occurrence, distribution, and phenotype of arylsulfatase A mutations in patients with metachromatic leukodystrophy, *Am J Med Genet* 69, 335-340.
- Biffi, A., Lucchini, G., Rovelli, A., Sessa, M. (2008) Metachromatic leukodystrophy: an overview of current and prospective treatments, *Bone Marrow Transplant* 42 Suppl 2, S2-6.
- Coelho, J.C., Wajner, M., Burin, M.G., Vargas, C.R., Giuliani R. (1997) Selective screening of 10.000 high-risk Brazilian patients for the detection of inborn errors of metabolism. *Eur J Pediatr* 156:650-654.
- Cosma, M. P., Pepe, S., Annunziata, I., Newbold, R. F., Grompe, M., Parenti, G., Ballabio, A. (2003) The multiple sulfatase deficiency gene encodes an essential and limiting factor for the activity of sulfatases, *Cell* 113, 445-456.
- Den Dunnen, J.T., Antonarakis, S.E. (2000) Mutation nomenclature extensions and suggestions to describe complex mutations: a discussion, *Hum Mutat*; 15: 7-12.
- Dierks, T., Dickmanns, A., Preusser-Kunze, A., Schmidt, B., Mariappan, M., von Figura, K., Ficner, R., Rudolph, M. G. (2005) Molecular basis for multiple sulfatase deficiency and mechanism for formylglycine generation of the human formylglycine-generating enzyme, *Cell* 121, 541-552.
- Drickamer, K., Taylor, M. E. (1998) Evolving views of protein glycosylation, *Trends Biochem Sci* 23, 321-324.

- Draghia, R., Letourneur, F., Drugan, C., Manicom, J., Blanchot, C., Kahn, A., Poenaru, L., Caillaud, C. (1997) Metachromatic leukodystrophy: identification of the first deletion in exon 1 and of nine novel point mutations in the arylsulfatase A gene, *Hum Mutat* 9, 234-242.
- Fluharty, A.L. (2006) Arylsulfatase Deficiency, Disponível em: <http://www.ncbi.nlm.nih.gov/bookshelf/br.fcgi?book=gene&part=mld> >. Acessado em agosto, 2008.
- Gieselmann, V. (1991) An assay for the rapid detection of the arylsulfatase A pseudodeficiency allele facilitates diagnosis and genetic counseling for metachromatic leukodystrophy. *Hum Genet* 86, 251-255.
- Ghosh, D. (2007) Human sulfatases: a structural perspective to catalysis, *Cell Mol Life Sci* 64, 2013-2022.
- Gort, L., Coll, M. J., Chabas, A. (1999) Identification of 12 novel mutations and two new polymorphisms in the arylsulfatase A gene: haplotype and genotype-phenotype correlation studies in Spanish metachromatic leukodystrophy patients, *Hum Mutat* 14, 240-248.
- Grossi, S., Regis, S., Rosano, C., Corsolini, F., Uziel, G., Sessa, M., Di Rocco, M., Parenti, G., Deodato, F., Leuzzi, V., Biancheri, R., Filocamo, M. (2008) Molecular analysis of ARSA and PSAP genes in twenty-one Italian patients with metachromatic leukodystrophy: identification and functional characterization of 11 novel ARSA alleles, *Hum Mut* 9999, 1098-1004.
- Harvey, J. S., Carey, W. F., Morris, C. P. (1998) Importance of the glycosylation and polyadenylation variants in metachromatic leukodystrophy pseudodeficiency phenotype, *Hum Mol Genet* 7, 1215-1219.
- Harvey, J. S., Nelson, P. V., Carey, W. F., Robertson, E. F., Morris, C. P. (1993) An arylsulfatase A (ARSA) missense mutation (T274M) causing late-infantile metachromatic leukodystrophy, *Hum Mutat* 2, 261-267.
- Heinisch, U., Zlotogora, J., von Figura, K., Kafert, S., Gieselmann, V. (1995) Multiple mutations are responsible for the high frequency of metachromatic leukodystrophy in a small geographic area, *Am J Hum Genet* 56, 51-57.
- Holve, S., Hu, D., McCandless, S.E. (2001) Metachromatic leukodystrophy in the Navajo: fallout of the American-Indian wars of the nineteenth century, *Am J Med Genet* 101, 203-208.
- Human Gene Mutation Database (HGMD). Disponível em: <http://www.hgmd.cf.ac.uk/ac/index.php> >. Acessado em setembro, 2010.

- Imperiali, B., O'Connor, S. E. (1999) Effect of N-linked glycosylation on glycopeptide and glycoprotein structure, *Curr Opin Chem Biol* 3, 643-649.
- Kappler, J., Sommerlade, H., von Figura, K., Gieselmann, V. (1994) Complex arylsulfatase A alleles causing metachromatic leukodystrophy, *Hum. Mutat* 4, 119-127.
- Kreysing, J., von Figura, K., Gieselmann, V. (1990) Structure of the arylsulfatase A gene, *Eur J Biochem* 191, 627-631.
- Kurosawa, K., Ida, H., Eto, Y. (1998) Prevalence of arylsulphatase A mutations in 11 Japanese patients with metachromatic leukodystrophy: identification of two novel mutations, *J Inher Metab Dis* 21, 781-782.
- Laidler, P., Litynska, A. (1997) Arylsulfatase A from human placenta possesses only high mannose-type glycans, *Int J Biochem Cell Biol* 29, 475-483.
- Lugowska, A., Amaral, O., Berger, J., Berna, L., Bosshard, N.U., Chabas, A., Fensom, A., Gieselmann, V., Gorovenko, N.G., Lissens, W., Masson, J.E., Marcao, A., Michelakakis, H., Bernheimer, H., Ol'khovych, N.V., Regis, S., Sinke, R., Tylki-Szymanska, A., Czartoryska, B. (2005) Mutations c.459+1G>A and p.P426L in the ARSA gene: prevalence in metachromatic leukodystrophy patients from European countries, *Mol Genet Metab* 86, 353-359.
- Lugowska, A., Czartoryska, B., Tylki-Szymanska, A., Bisko, M., Zimowski, J. G., Berger, J., Molzer B. (2000) Prevalence of arylsulfatase A pseudodeficiency allele in metachromatic leukodystrophy patients from Poland, *Eur Neurol* 44, 104-107.
- Lukatela, G., Krauss, N., Theis, K., Selmer, T., Gieselmann, V., von Figura, K., Saenger, W. (1997) Crystal structure of human arylsulfatase A: the aldehyde function and the metal ion at the active site suggest a novel mechanism for sulfate ester hydrolysis, *Biochemistry* 37, 3654-3664.
- Meikle, P. J., Hopwood, J. J., Clague, A. E., Carey, W. F. (1999) Prevalence of lysosomal storage disorders, *JAMA* 281, 249-254.
- Mitsunaga-Nakatsubo, K., Kusunoki, S., Kawakami, H., Akasaka, K., Akimoto, Y. (2009) Cell-surface arylsulfatase A and B on sinusoidal endothelial cells, hepatocytes, and Kupffer cells in mammalian livers, *Med Mol Morphol* 42, 63-69.
- Nakamura, M., Gasa, S., Makita, A. (1984) Arylsulfatase A from normal human lung and lung tumors showed different patterns of Microheterogeneity, *J Biochem* 96, 207-203.
- Nelson, P.V., Carey, W.F. Morris, C.P. (1991) Population frequency of the arylsulphatase A pseudo-deficiency allele, *Hum Genet* 87, 87-88.

- Park, D. S., Poretz, R. D., Stein, S., Nora, R., Manowitz, P. (1996) Association of alcoholism with the N-glycosylation polymorphism of pseudodeficient human arylsulfatase A, *Alcohol Clin Exp Res* 20, 228-233.
- Pedron, C.G., Gaspar, P.A., Giugliani, R., Pereira, M.L. (1999) Arylsulfatase A pseudodeficiency in healthy Brazilian individuals, *Bras J Med Biol Res* 199, 941-945.
- Pol-Fachin, L., Fernandes, C. L., Verli, H. (2009) GROMOS96 43a1 performance on the characterization of glycoprotein conformational ensembles through molecular dynamics simulations, *Carbohydr Res* 344, 491-500.
- Pinto, R., Caseiro, C., Lemos, M., Lopes, L., Fontes, A., Ribeiro, H., Pinto, E., Silva, E., Rocha, S., Marcao, A., Ribeiro, I., Lacerda, L., Ribeiro, G., Amaral, O., Sa Miranda, M. C. (2004) Prevalence of lysosomal storage diseases in Portugal, *Eur J Hum Genet* 12, 87-92.
- Polten, A., Fluharty, A.L., Fluharty, C.B., Kappler, J., von Figura, K., Gieselmann, V. (1991) Molecular basis of different forms of metachromatic leukodystrophy, *N Engl J Med* 324,18-22.
- Poorthuis, B. J., Wevers, R. A., Kleijer, W. J., Groener, J. E., de Jong, J. G., van Weely, S., Niezen-Koning, K. E., van Diggelen, O. P. (1999) The frequency of lysosomal storage diseases in The Netherlands, *Hum Genet* 105, 151-156.
- Ricketts, M. H., Poretz, R. D., Manowitz, P. (1998) The R496H mutation of arylsulfatase A does not cause metachromatic leukodystrophy, *Hum Mutat* 12, 238-239.
- Shen, N., Li, Z. G., Wayne, J. S., Francis, G., Chang, P. L. (1993) Complications in the genotypic molecular diagnosis of pseudo arylsulfatase A deficiency, *Am J Med Genet* 45, 631-637.
- Sommerlade, H. J., Selmer, T., Ingendoh, A., Gieselmann, V., von Figura, K., Neifer, K., Schmidt, B. (1994) Glycosylation and phosphorylation of arylsulfatase A, *J Biol Chem* 269, 20977-20981.
- Stein, C., Gieselmann, V., Kreysing, J., Schmidt, B., Pohlmann, R., Waheed, A., Meyer, H. E., O'Brien, J. S., von Figura, K. (1989) Cloning and expression of human arylsulfatase A, *J Biol Chem* 264, 1252-1259.
- Tylki-Szymanska, A., Berger, J., Loschl, B., Lugowska, A., Molzer, B. (1996) Late juvenile metachromatic leukodystrophy (MLD) in three patients with a similar clinical course and identical mutation on one allele, *Clin Genet* 50, 287-292.
- Vagedes, P., Saenger, W., Knapp, E. W. (2002) Driving forces of protein association: the dimer-octamer equilibrium in arylsulfatase A, *Biophys J* 83, 3066-3078.

- Von Figura, K., Gieselmann, V., Jaeken, J., (2001) Metachromatic Leukodystrophy. In: Scriver, C.R., Beaudet, A.L., Sly, W.S., Valle, D., Childs, B., Kinzler, K.E., Vogelstein, B. (eds) *The Metabolic and Molecular Bases of Inherited Disease*. 8th edition. McGraw-Hill, New York, 5833-5843.
- Waheed, A., Hasilik, A., von Figura, K. (1982) Synthesis and processing of arylsulfatase A in human skin fibroblasts, *Hoppe Seyler Z Physiol Chem* 363, 425-430.
- Waldow, A., Schmidt, B., Dierks, T., von Bulow, R., von Figura, K. (1999) Amino acid residues forming the active site of arylsulfatase A. Role in catalytic activity and substrate binding, *J Biol Chem* 274, 12284-12288.
- Wormald, M. R., Dwek, R. A. (1999) Glycoproteins: glycan presentation and protein-fold stability, *Structure* 7, R155-160.
- Zlotogora, J., Bach, G., Bosenberg, C., Barak, Y., von Figura, K., Gieselmann, V. (1995) Molecular basis of late infantile metachromatic leukodystrophy in the Habbanite Jews, *Hum Mutat* 5, 137-143.
- Zlotogora, J., Furman-Shaharabani, Y., Harris, A., Barth, M. L., von Figura, K., and Gieselmann, V. (1994) A single origin for the most frequent mutation causing late infantile metachromatic leucodystrophy, *J Med Genet* 31, 672-674.
- Zlotogora, J., Gieselman, V., von Figura, K., Zeigler, M., Bach, G. (1994) Late infantile metachromatic leukodystrophy in Israel, *Biomed Pharmacother* 48, 347-350.