UNIVERSIDADE DE LISBOA

FACULDADE DE FARMÁCIA

MOLECULAR BASES OF CLASSIC GALACTOSEMIA: SEARCHING FOR NEW THERAPEUTIC STRATEGIES

∞

BASES MOLECULARES DA GALACTOSÉMIA CLÁSSICA: PESQUISA DE NOVAS ESTRATÉGIAS TERAPÊUTICAS

Research Advisor: Professora Doutora Isabel Antolin Rivera

Co-advisor: Doutor João B. Vicente

Ana Isabel Cruz Coelho

Lisboa

2014

Dissertação apresentada à Faculdade de Farmácia da Universidade de Lisboa para obtenção do grau de Doutor em Farmácia (Biologia Celular e Molecular)

De acordo com o disposto no ponto 1 do artigo nº41 do Regulamento de Estudos Pós-Graduados da Universidade de Lisboa, deliberação nº93/2006, publicada em Diário da República – II série nº153 – 5 de Julho de 2003, a autora desta dissertação declara que participou na concepção e execução do trabalho experimental, interpretação dos resultados obtidos e redacção dos manuscritos.

The studies presented in this thesis were performed at Research Institute for Medicines and Pharmaceutical Sciences (*i*Med.UL), Faculty of Pharmacy, University of Lisbon - Portugal, under the scientific supervision of Professor Isabel Antolin Rivera and Doutor João B. Vicente. Also, scientific training was performed at Centro de Biologia Molecular Severo Ochoa (CBMSO), from Universidad Autónoma de Madrid, under the scientific supervision of Professor Belén Perez, and in the Department of Human Genetics, from Emory University Medical School, under the scientific supervision of Professor Judith Fridovich-Keil.

Ana Isabel Cruz Coelho is the recipient of a PhD felllowship (SFRH/BD/48259/2008) from Fundação para a Ciência e a Tecnologia (FCT), Lisbon, Portugal. This work was supported in part by PEst-OE/SAU/UI4013/2011 from Research Institute for Medicines and Pharmaceutical Sciences (*i*Med.UL), Faculty of Pharmacy, University of Lisbon.

Ciência.Inovação 2010 Programa Operacional Ciência e Inovação 2010
MINISTÉRIO DA CIÊNCIA, INOVAÇÃO E ENSINO SUPERIOR

UNIÃO EUROPEIA
Fundo Social Europeu

para os meus pais...

In life, nothing is to be feared.
Everything is to be understood.

Marie Skłodowska-Curie

ACKNOWLEDGMENTS

Depois de 5 anos no desenvolvimento deste trabalho que agora se concretiza nesta tese, há certos momentos que nunca esquecerei e, antes de concluir esta etapa tão importante da minha vida, não podia deixar de agradecer a todos aqueles que me acompanharam nesta viagem.

As minhas primeiras palavras de agradecimento dirigem-se à Professora Doutora Isabel Rivera, sem a qual não teria sido possível a concretização desta tese. Professora, muito obrigada por me ter introduzido no mundo da investigação, muito obrigada por todo o apoio dado ao longo destes 5 anos, e muito obrigada pela sua amizade. Muito mais do que uma orientadora, foi uma amiga e conselheira.

Gostaria também de agradecer à Professora Doutora Margarida Leite por me ter introduzido no mundo da galactosémia, e pela forma carinhosa como me acolheu.

À Professora Doutora Mª João Silva, à qual tenho de agradecer todo o apoio que me deu ao longo de todo o doutoramento.

Ao Professor Doutor João Vicente que me mostrou que a investigação pode ser um "rayo de sol", e que o lema "esforço e dedicação" pode superar muitos obstáculos. *Grazie ancora* pela música sugerida – apropriada para cada ocasião – e que preencheu os meus dias no laboratório.

Gostaria também de agradecer à Professora Doutora Isabel Tavares de Almeida, por me ter aceite no seu grupo de investigação e por ter reunido todas as condições que me permitiram desenvolver o trabalho que agora se concretiza nesta tese.

À Professora Doutora Fátima Ventura, pelos bons momentos passados no laboratório; à Professora Doutora Paula Leandro, por todos os ensinamentos que me transmitiu; à Professora Doutora Rita Castro, por ter sempre um sorriso; e à Professora Doutora Margarida Silva, por ter sempre uma palavra de encorajamento.

E gostaria também de agradecer à restante equipa Met&Gen que fez com que tudo fosse tão divertido.

À Ana Pinheiro e à Inês Faustino, com quem dei os meus primeiros passos no laboratório e cujo papel determinante na decisão de prosseguir com doutoramento nunca irei esquecer; à Sandra, por me ter acompanhado no projecto da PKU e que surpreendentemente acabaria por me acompanhar no projecto de doutoramento – muito obrigada por todo o apoio! Mesmo longe, foi como se tivesses estado sempre aqui! À Cátia, pela sua constante boa-disposição, com que enchia o laboratório de alegria, e pelo seu anel tão especial que um dia fez história! À Madalena, pelas boas gargalhadas que partilhámos e por ter sido a minha companheira da sequenciação – lembras-te das nossas tardes a preparar amostras para sequenciação? Como dizia a Ana P.: a dupla da sequenciação! À Angela, pelas suas palavras sábias proferidas naquele querido mês de Agosto! Ao Dr. M&M, aka Marco Moedas, pela sua boa-disposição e por conseguir sempre "arrancar-me uma gargalhada". Numa palavra: Mercadona! À Liliana, pelo poucos mas muito bons momentos

partilhados, muito obrigada pela tua amizade e constante encorajamento! À Cristina por ter sempre um sorriso e sempre uma palavra amiga! Muito obrigada por todo o apoio, muito obrigada pelos bons momentos partilhados, nomeadamente nas inúmeras boleias e operações stop com teus conterrâneos. Nunca desistas de concretizar os teus sonhos! És uma pessoa extremamente competente e nunca deixes que te digam o contrário. À Hana pelos bons momentos passados e pela constante disponibilidade para ajudar – děkuji! Ahoj e sbohem... Ao Rúben Esse, nunca pares de "breaking on through"! Ao Israel pelas longas conversas luso-italianas e pelos seus "indimenticabili" pitéus – grazie, bello! Às queridas "holandesas" – Paula, Mónica, Sara e Marisa – muito obrigada pelas deliciosas bolachas que alimentaram o cérebro em longas horas de trabalho no laboratório. À Liliana, a mais recente aquisição do grupo Met&Gen, pela imensa energia que transmite! Ao Rúben R., Ana Serrão e Dra. Elisa e D. Amélia, pelo excelente trabalho que fazem nas análises clínicas. À Conceição, por ter sempre disponível um ombro amigo, e por todo o importante apoio que me deu ao longo destes últimos anos.

Muito obrigada ainda ao Henrique e ao João Leandro que com toda a sua sabedoria e sensatez me transmitiram importantes conhecimentos, sem os quais eu não seria a pessoa que sou hoje. À Matilde quero agradecer todo o apoio e dedicação que prestou à galactosémia clássica, e dizer como me sinto honrada por ter participado na sua entrada no mundo da investigação. Ah, e claro agradecer pelos kits de western – indispensáveis para um bom western! À Sílvia, por todo o trabalho desenvolvido com a nossa querida "IVS8" e pela qual lhe estou muito grata. À Filipa, Francisca e Margarida, por serem as queridas meninas da MCAD. Muito obrigada ainda à Ana Calhelha e Isabel Serra por todo o apoio que me deram, pela frequente companhia ao almoço, e pelas inúmeras vezes em que "me salvaram"...

Por último gostaria de agradecer à minha família. Em primeiro lugar, uma palavra de agradecimento aos meus Pais que sempre acreditaram em mim; sempre me incentivaram a seguir os meus sonhos com força e determinação, e me mostraram que "vale a pena ser cientista".

Gostaria também de agradecer ao meu irmão, Bruno, e à minha cunhadinha, Mariana, por todo o apoio dado; e claro agradecer também ao Diogo, o sobrinho mais querido do mundo.

Aos meus padrinhos e afilhadinhos que também querem ser "cientistas como a Madrinha".

Às minhas grandes amigas, Carla, Joana, Susaninha e Serafim, por terem tido a paciência de me ouvir nos bons e maus momentos, e por serem muito mais do que amigas de infância. À Rita que sempre me admirou e me fez acreditar que eu conseguiria concretizar este sonho.

Ao Benny e à Inês, pelos jantares tardios e bons serões que passámos, depois das minhas longas e intermináveis culturas.

À Hélia e ao Vasco, pelos bons momentos compartidos, comigo e com a fiel companheira Fifi – momentos indeléveis!

Una parola di ringraziamento anche ai miei suoceri che, seppur lontani mi hanno dato sempre tanto supporto, sempre pronti ad alimentare il mio cervello con cinquemila chili di crema di ricotta – la crema più buona al mondo, c'è da dire! – e per incoraggiarmi a non desistire mai. E vorrei ringraziare anche il mio cognadinho che, seppur così differente in tanti aspetti, è uguale a me in tanti altri.

E por último gostaria de agradecer ao Andrea, porque esta tese é também para ti. A ti que sacrificaste tanto quanto eu ao longo destes 5 anos, nunca deixaste de acreditar em mim e sempre me incentivaste a "ir mais além". Obrigada!

TABLE OF CONTENTS

FIGURES

TABLES

ABBREVIATIONS

ΔΔG	free energy change
ϕ	hydrophobic amino acid
2'-OMe	2'-O-methyl
2'-MOE	2'-O-(2'-methoxy)ethyl
6His	hexa histidine
A	adenosine
ADP	adenosine diphosphate
ADA	adenosine deaminase
AMH	anti-Mullerian hormone
AR	aldose reductase
ATP	adenosine triphosphate
BBP	branch point bridging protein
BMD	bone mineral density
bp	base pairs
C	cytidine
CAS	childhood apraxia of speech
CD	circular dichroism
cDNA	complementary DNA
CGSC	coli genetic stock center
CH	congenital hypothyroidism
CUPSAT	Cologne University Protein Stability Analysis Tool
D	Duarte
D1	Los Angeles
D2	Duarte
DLS	dynamic light scattering
DNA	deoxyribonucleic acid
DSF	differential scanning fluorimetry
DTT	dithiothreitol
E. coli	*Escherichia coli*
EDTA	ethylenediaminetetraacetic acid
ER	endoplasmic reticulum
ESE	exonic splicing enhancer
ESI-MS/MS	electrospray ionization tandem mass spectrometry
ESS	exonic splicing silencer

FSH	follicle-stimulating hormone
G	guanosine
Gal-1-P	galactose-1-phosphate
GALE	UDP-galactose 4'-epimerase
GALK	galactokinase
GALM	galactose mutarotase
GALT	galactose-1-phosphate uridylyltransferase
GC/MS	gas chromatography-mass spectrometry
GHMP	galactokinase, homoserine kinase, mevalonate kinase and phosphomevalonate kinase
Glc-1-P	glucose-1-phosphate
GLUT	facilitative glucose transporter
HGMD	human gene mutation database
His	histidine
HIT	histidine triad
hnRNP	heterogeneous nuclear ribonucleoprotein
His	histidine
HPLC	high-performance liquid chromatography
HPO	hypothalamic-pituitary-ovarian
HRQoL	health-related quality of life
HRT	hormone replacement therapy
Hsp	heat-shock proteins
IMD	inherited metabolic disorder
IGF-I	insulin-like growth factor-I
IMAC	immobilized metal affinity chromatography
IMPase	inositol monophosphatase
ISE	intronic splicing enhancer
ISS	intronic splicing silencer
IQ	intelligence quotient
kan	kanamycin
kb	kilo base pairs
kDa	kilo Dalton
LA	Los Angeles
Leu	leucine
LH	luteinizing hormone
LNA	locked nucleic acid

min	minutes
mRNA	messenger RNA
MSD-NOS	motor speech disorder-not otherwise specified
MS	mass spectrometry
MS/MS	tandem mass spectrometry
n.a.	not available
NAD(H)	nicotinamide adenine dinucleotide
NADP(H)	nicotinamide adenine dinucleotide phosphate
NMR	nuclear magnetic resonance
non-snRNP	non-small nuclear ribonucleoprotein
OD_{600}	optical density
PAH	phenylalanine hydroxylase
PBS-T	phosphate buffered saline-tween 20
PCR	polymerase chain reaction
PKU	phenylketonuria
PMO	phosphoroamidate morpholino oligomer
PMSF	phenylmethanesulfonyl fluoride
PNA	peptide nucleic acid
PNSP	portuguese neonatal screening program
POI	primary ovarian insufficiency
PPi	pyrophosphate
PQC	protein quality control
Pro	proline
PS	phosphorothioate
PTC	premature termination codon
Py_n	polypyrimidine tract
RBC	red blood cells
RNA	ribonucleic acid
RNAi	RNA interference
RRM	RNA recognition motif
RT-PCR	reverse transcription polymerase chain reaction
SDH	sorbitol dehydrogenase
SDM	site-directed mutator
sec	second
Ser	serine
SGLT	sodium-glucose transporters

SIFT	sorting intolerant from tolerant
SmaRT	spliceosome-mediated RNA *trans*-splicing
snRNA	small nuclear RNA
snRNP	small nuclear ribonucleoproteins
SR	serine/arginine-rich proteins
SRE	splicing-regulatory elements
SVM	support vector machines
T	thymidine
Tyr	tyrosine
U	uridine
UMP	uridine monophosphate
UDP-Gal	uridine diphosphate galactose
UDP-Glc	uridine diphosphate glucose
UDP-GalNAc	uridine diphosphate-N-acetyl-galactosamine
UGGT	UDP-glucose:glycoprotein glucosyltransferase
UDP-GlcNAc	uridine diphosphate-N-acetyl-glucosamine
UGP	UTP-dependent glucose/galactose pyrophosphorylase
UTP	uridine-5'-triphosphate
UV	ultraviolet
Vc	column volume
WT	wild-type

SUMMARY

Classic galactosemia is a potentially lethal disease of the galactose metabolism caused by a severe deficiency of GALT, the second enzyme of the Leloir pathway. This disorder is associated with mutations in the *GALT* gene, and displays an autosomal recessive pattern of inheritance. Although apparently asymptomatic at birth, affected infants start developing escalating symptoms after 1 to 2 weeks of galactose ingestion due to milk feeding. Upon implementation of a galactose-restricted diet, infants show rapid clinical improvement and seem to be almost miraculously cured; however, the long-term outcome is often disappointing, as most patients go on to develop severe complications notwithstanding strict compliance with the dietary therapy.

Despite decades of intensive research, classic galactosemia pathophysiology is still largely unknown, resulting in a limited and poorly effective therapy. A comprehensive understanding of the molecular mechanisms underlying classic galactosemia potentially opens new therapeutic avenues, and prompted us to develop the present work.

The first part of this thesis (**chapter 1**) comprises a general introduction with a review of the literature concerning the molecular and biochemical basis of classic galactosemia, followed by an overview of the current therapeutic approaches in genetic disorders – with a special focus on inherited metabolic disorders – and finally the objectives of the present work.

Our studies initiated with the molecular characterization of all patients currently followed by the Portuguese metabolic centers (**chapter 2**). The rationale for this study was the previous description that *GALT* genotype represents a valuable prognostic tool for the outcome of galactosemic patients. Accordingly, after genotyping 42 Portuguese galactosemic patients, we searched for genotype-phenotype correlations by a 3-fold evaluation. Firstly, we employed *in silico* strategies to assess the structural-functional impact of previously uncharacterized mutations; secondly, we evaluated the biochemical phenotype at both the metabolite and enzymatic levels, represented by Gal-1-P values and GALT activity in erythrocytes; and finally, evaluated the resulting clinical outcome. Establishment of correlations between genotype and biochemical or clinical phenotypes, however, was poorly attained, reiterating the complexity of this disease and emphasizing the idea that other modifiers – possibly genetic, epigenetic and environmental – contribute to the pathophysiology of classic galactosemia.

The update of the Portuguese *GALT* mutational spectrum revealed the intronic mutation c.820+13A>G (IVS8+13A>G) as the second most prevalent variation, strongly suggesting its pathogenicity, which set the basis for the study described in **chapter 3**. We functionally characterized the c.820+13A>G variation by *ex vivo* and *in vivo* analyses, which were in full agreement with the previous *in silico* predictions. Indeed, we confirmed this variation is a disease-causing mutation, whose mechanism of action involves the activation of a cryptic donor

site, which, in turn, induces an aberrant splicing of the *GALT* pre-mRNA, thus causing a frameshift with inclusion of a premature stop codon. Structural-functional studies of the recombinant truncated GALT showed it was devoid of enzymatic activity and prone to aggregation. Finally, antisense oligonucleotides were designed to specifically recognize the mutation, and successfully restored the constitutive splicing.

Molecular studies in several inherited metabolic disorders have led to the realization that only a minor part of the mutations directly disrupt functional sites of ~~the~~ proteins. Accordingly, in **chapter 4**, we described the structural and functional characterization of the most prevalent mutations in the *GALT* gene – p.Q188R, p.S135L, p.K285N and p.N314D – and of other five clinically relevant mutations – p.R148Q, p.G175D, p.P185S, p.R231C and p.R231H. Interestingly, the analyzed mutations did not affect the global conformational stability of the GALT enzyme; rather, most mutations, notwithstanding their impact on the enzyme functionality, increased the propensity for aggregation, which at the cellular level reflects in a decrease of the enzyme's effective concentration. These results are in agreement with previous studies in classic galactosemia models, and suggest that GALT aggregation might be a major pathogenic mechanism underlying this disorder.

Previous studies have reported a yeast galactosemia model allowing the evaluation of human GALT mutations severity, by assaying the sensitivity of transformed yeast cultures to galactose addition to the medium. Thus, in **chapter 5**, we developed a prokaryotic model of galactose sensitivity to evaluate the ability of the above referred human GALT mutants in alleviating the galactose-induced toxicity. This model presents the inherent advantage of being assayed *in vivo*, thus providing valuable insights on mutations' impact on human GALT function. Additionally, arginine ability to ameliorate the galactose-induced toxicity was also evaluated for each human GALT mutant. The rationale for this approach was, not only the long-recognized anti-aggregation properties of arginine, but also its important therapeutic effect described in previous studies. In effect, arginine appears to exert a mutation-specific mode of action, alleviating the galactose toxicity in the p.Q188R, p.K285N, and p.G175D mutants, which suggests that might be of some benefit in classic galactosemia. Nonetheless, further studies are underway to ascertain arginine's potential therapeutic effect in this inherited disorder of the galactose metabolism.

C**hapter 6** presents a general discussion and major conclusions disclosed by this work, framing them in the current state of the art, proving also some perspectives for future studies.

Taken together, our results provide important insights on classic galactosemia, namely by shedding light on the underlying pathogenic molecular mechanisms, thus contributing for a better understanding of this enigmatic disorder. Finally, and importantly, these studies paved the way to the search, development and improvement of novel and alternative therapeutic strategies, so needed to overcome the overwhelming and burdensome long-term complications presented by

most classic galactosemic patients.

Keywords: classic galactosemia; GALT; genotype-phenotype correlation; antisense therapy; protein aggregation

RESUMO

Após a sua internalização celular, a galactose é rapidamente metabolizada a glucose-1-fosfato pela acção sequencial de três enzimas: galactocinase (GALK), galactose-1-fosfato uridililtransferase (GALT) e uridina difosfogalactose 4'-epimerase (GALE). Estas enzimas permitem, respectivamente, a fosforilação da galactose em galactose-1-fosfato (GALK), a conversão de galactose-1-fosfato e uridina difosfoglucose em glucose-1-fosfato e uridina difosfogalactose (GALT), e a interconversão de uridina difosfoglucose em uridina difosfogalactose (GALE). Estas enzimas constituem a via de Leloir e, apesar de o fígado ser o principal órgão envolvido no metabolismo da galactose, encontram-se na maioria das células e dos tecidos.

Uma deficiência enzimática em qualquer uma das três enzimas da via de Leloir resulta numa diminuição ou ausência de capacidade de metabolizar a galactose e, consequentemente, conduz à sua acumulação no sangue – hipergalactosémia.

A Galactosémia Clássica (OMIM #230400), a forma mais comum de hipergalactosémia primária, é uma doença genética de transmissão autossómica recessiva, que afecta 1 em cada 30.000 a 60.000 nados-vivos, apresentando uma prevalência variável entre populações, nomeadamente na Irlanda e Turquia onde atinge valores de 1/23.500 - 1/23.775 e no Japão onde a sua prevalência atinge o valor mais baixo (1/1.000.000). A ausência ou diminuição de actividade da GALT é causada por mutações no gene *GALT*, e encontram-se descritas mais de 250 variações, reflectindo a elevada heterogeneidade alélica desta doença metabólica. Para além disso, a maioria dos doentes são heterozigóticos compostos, um dos factores determinantes da ampla variabilidade fenotípica observada.

Aquando do nascimento, a criança aparenta ser assintomática. Os sintomas tornam-se evidentes após o início da ingestão de leite, e consistem inicialmente em dificuldades alimentares e de desenvolvimento, vómitos, diarreia, letargia e hipotonia, podendo evoluir para cataratas e septicémia, e eventualmente conduzir à morte. Após implementação de uma dieta restrita em galactose – o tratamento padrão – as crianças mostram notáveis melhorias em apenas 24 horas, e em apenas uma a duas semanas as disfunções hepática e renal desaparecem completamente. Esta resposta dramática à terapia dietética levou ao conceito da Galactosémia Clássica como uma doença relativamente benigna e fácil de tratar. No entanto, os doentes com Galactosémia Clássica desenvolvem complicações a longo prazo e a diversos níveis, como neurológico e psiconeurológico, crescimento e densidade óssea, e disfunção ovárica nas mulheres, os quais parecem ser independentes de um diagnóstico precoce e de uma adesão do doente à terapia.

Apesar de a primeira descrição datar de 1908, a Galactosémia Clássica continua a ser um enigma, quer ao nível do conhecimento aprofundado da sua fisiopatologia, quer ao nível do

desenvolvimento de alternativas terapêuticas que permitam, sobretudo, mitigar as complicações a longo prazo.

O presente trabalho pretende, assim, contribuir para a elucidação dos mecanismos moleculares subjacentes à patogénese das mutações *GALT* prevalentes na população galactosémica Portuguesa, assim como explorar novas estratégias terapêuticas direccionadas para cada tipo específico de mutação.

O **primeiro capítulo** da tese apresenta uma revisão geral da literatura sobre vários aspectos relacionados com a Galactosémia Clássica: as bases bioquímicas e moleculares, a fisiopatologia e as actuais abordagens terapêuticas. Em seguida, descrevem-se em pormenor os mecanismos moleculares subjacentes à maioria das mutações que originam Erros Hereditários do Metabolismo (onde se inclui a Galactosémia Clássica) e que causam perca de função, nomeadamente as mutações *missense* e as que afectam o *splicing*, mecanismos esses que condicionam as novas terapias desenhadas de acordo com o tipo de mutação. Por fim, são descritos os objectivos delineados para o presente trabalho.

O trabalho experimental teve início com a caracterização molecular de todos os doentes actualmente seguidos nos centros metabólicos nacionais de Lisboa, Porto e Coimbra, sendo o objectivo final averiguar se, como sugerido na literatura, o genótipo *GALT* pode ter um valor prognóstico sobre a evolução fenotípica dos doentes galactosémicos (**capítulo 2**). Os dados obtidos permitiram estabelecer o espectro mutacional da Galactosémia Clássica em Portugal e compará-lo com o de outras populações. O estudo de 42 doentes revelou a presença de 14 substituições nucleotídicas, sendo 10 *missense*, 2 *nonsense* e 2 de *splicing*. Identificaram-se 16 genótipos diferentes, mas metade dos doentes são homozigóticos para p.Q188R, a mutação prevalente não só em Portugal como a nível mundial. Surpreendentemente, a segunda mutação mais frequente é uma mutação de *splicing*, descrita até então como benigna.

Em seguida, e recorrendo a programas bioinformáticos adequados, procedeu-se à análise da potencial patogenicidade das mutações ainda não caracterizadas na literatura. Os resultados sugeriram que a maioria destas mutações *missense* afectará a estabilidade e a funcionalidade da proteína mutada, enquanto que a mutação de *splicing* mais frequente deverá induzir um mecanismo de *splicing* alternativo.

Por fim, este estudo revelou, na maioria dos casos, a ausência de uma correlação clara entre a gravidade das mutações prevista pela análise *in silico* e o fenótipo bioquímico dos doentes, determinado pelos níveis eritrocitários de galactose-1-fosfato, assim como com o fenótipo clínico e a evolução desfavorável manifestada pelos doentes. No entanto, tal resultado não surpreende dado a Galactosémia Clássica, apesar de ser uma doença monogénica, não originar fenótipos claros. Este facto prende-se com a falta de informação estrutural sobre os componentes enzimáticos da via de Leloir, colocando-se a hipótese de uma alteração na GALT poder afectar toda a via metabólica. Por outro lado, os metabolitos da galactose estão envolvidos

em diversas reacções fisiológicas, nomeadamente as de glicosilação que se reflectem aos mais variados níveis. Finalmente, a influência de outros genes modificadores, de alterações epigenéticas e de factores ambientais também não pode ser ignorada.

Os capítulos seguintes são dedicados ao estudo dos mecanismos moleculares subjacentes às mutações prevalentes na população Portuguesa e ao desenvolvimento de novas abordagens terapêuticas.

No **capítulo 3** descreve-se o estudo da mutação de *splicing* IVS8+13A>G, até então considerada como benigna, mas que revelou ser a segunda mais frequente na nossa população galactosémica. A caracterização funcional do mecanismo patogénico foi efectuada por transfeção de um minigene contendo a sequência mutada em duas linhas diferentes de células eucarióticas. A análise dos produtos de transcrição revelou que a mutação activa um sítio críptico de *splicing*, causando um *splicing* anómalo do pré-mRNA *GALT*, o qual induz um *frameshift* com inclusão de um codão de terminação prematuro (p.D274GfsX291).

Por outro lado, dado ter-se observado a presença do mensageiro mutado nas amostras biológicas dos doentes portadores desta mutação, colocou-se a hipótese de a proteína truncada ser produzida *in vivo*, o que levou à produção da proteína recombinante. Os estudos estruturais e funcionais subsequentes revelaram que esta proteína é propensa à formação de agregados e é destituída de actividade enzimática.

Finalmente, e utilizando oligonucleótidos *antisense* que hibridam especificamente com o local da mutação, impedindo assim a ligação do spliceossoma e forçando a sua ligação ao sítio canónico dador de *splicing*, foi possível corrigir o *splicing* alternativo induzido pela mutação e obter o mensageiro selvagem. Estas experiências constituíram a prova de conceito sobre a aplicabilidade da terapia *antisense* como alternativa estratégica para a claramente insuficiente dieta restrita em galactose.

O espectro mutacional da maioria dos erros hereditários do metabolismo é dominado por mutações do tipo *missense*. O **capítulo 4** aborda a caracterização estrutural e funcional das quatro variantes GALT prevalentes a nível mundial, assim como outras cinco variantes com relevância clínica, nomeadamente na população galactosémica Portuguesa. Diversas metodologias foram empregues para determinação da actividade enzimática e do perfil de inactivação térmica, bem como métodos biofísicos para avaliação das estruturas secundária, terciária e quaternária das proteínas mutadas. Os resultados revelaram que as mutações pontuais não afectam nenhuma das estruturas acima mencionadas, mas sim a propensão para uma agregação precoce destas variantes.

Este resultado constitui na realidade a principal novidade, dado que um estudo recente postulou que algumas mutações no gene *GALT* afectam o correcto *folding* das proteínas mutadas. Esta conclusão é extremamente relevante na medida em que, para além da diminuição da actividade uridililtransferásica, a acumulação de agregados proteicos interfere com a homeostase

celular. Efectivamente, diversos estudos relataram a presença, em doentes galactosémicos, do aumento de actividade dos sistemas ligados ao *stress* do retículo endoplasmático e da *unfolded protein response*, assim como de níveis elevados de *stress* oxidativo, fenómenos característico de uma proteotoxicidade.

Estes dados levam-nos então a colocar como potencial e nova hipótese terapêutica a utilização de moduladores da proteostase, os quais prolongam a semi-vida celular das variantes GALT, compensando assim parcialmente a diminuição de actividade enzimática e simultaneamente prevenindo a acumulação de agregados proteicos.

No **capítulo 5** desenvolvemos um modelo procariótico de sensibilidade à galactose que permitisse avaliar, não só o impacto causado pelas diversas mutações *missense*, mas também o efeito de compostos com potencial acção terapêutica. Utilizámos uma estirpe de *E. coli* com o gene *galT* endógeno deletado, de modo que toda a actividade GALT detectada é proveniente do mutante nela transformado. Os resultados, embora ainda preliminares, confirmaram a validade do modelo delineado e permitiram obter dados a dois níveis. Por um lado, replicámos *in vivo* os resultados previamente obtidos *in vitro* quanto à capacidade de aliviar a toxicidade induzida pela galactose, por comparação com a enzima selvagem. Por outro lado, e muito relevante, verificámos que o modelo é válido para testar moléculas com potencial terapêutico. Efectivamente, algumas destas enzimas mutadas revelaram-se sensíveis à arginina, um composto amplamente reconhecido como estabilizador de proteínas e anti-agregante, aliviando desse modo a toxicidade da galactose para a bactéria.

A presente tese termina com o **capítulo 6** no qual se apresenta uma discussão dos resultados obtidos, incluindo uma análise integrada de todo o trabalho e respectivas conclusões, bem como algumas perspectivas de trabalho futuro.

Em suma, este trabalho contribuiu para um melhor conhecimento da Galactosémia Clássica em Portugal e para a elucidação dos mecanismos patogénicos subjacentes às mutações prevalentes, bem como para a descoberta e desenvolvimento de novas alternativas terapêuticas, tão necessárias para minorar as graves sequelas que a maioria dos doentes galactosémicos apresenta a longo termo.

CHAPTER 1

GENERAL INTRODUCTION AND OBJECTIVES

1.1. Pathways of galactose metabolism

The main source of galactose is the disaccharide lactose, present in milk and dairy products (1, 2). After its ingestion, lactose is hydrolyzed in the intestinal lumen by the disaccharidase lactase into its constituent monosaccharides, glucose and galactose. Galactose is then transported across the brush border membrane of the enterocyte by the SGLT1 active transporter, which uses the electrochemical gradient provided by the basolateral Na^+/K^+ ATPase pump to transport galactose into enterocytes against its concentration gradient, or through facilitated diffusion, by the GLUT2 transporter (3, 4). Across the basolateral membrane of the enterocyte, galactose is then transported by GLUT2 out of the enterocyte toward the blood stream, and is delivered by the portal blood to the liver, the major site of galactose metabolism, where it is internalized by GLUT2 ($K_M \approx 92$ mM) (3-7).

1.1.1. The Leloir pathway

Galactose is often present as an anomeric mixture of α-D-galactose and β-D-galactose, which spontaneously interconvert in aqueous solution (1, 8-10). In the cytoplasm, however, the rate of the spontaneous interconversion is insufficient to provide the needs of metabolic pathways, since many enzymes of carbohydrate metabolism exhibit specificity for one anomer or the other (10). In fact, when released from the breakdown of lactose, galactose is in the beta conformation (1). Once inside the cells, however, β-D-galactose must be epimerized to α-D-galactose through the action of galactose mutarotase (GALM, EC 5.1.3.3), in order to provide the anomer specific of the first enzyme of the Leloir pathway (1, 8, 10-12).

Once in the alpha conformation, D-galactose is converted to glucose-1-phosphate (Glc-1-P) by the action of three consecutive enzymes: galactokinase (GALK), which converts α-D-galactose into galactose-1-phosphate (Gal-1-P); galactose-1-phosphate uridylyltransferase (GALT), which converts Gal-1-P and uridine diphosphate-glucose (UDP-Glc) into Glc-1-P and uridine diphosphate-galactose (UDP-Gal); and UDP-galactose 4'-epimerase (GALE), which is responsible for the interconversion of UDP-Glc and UDP-Gal (Figure 1.1) (1). UDP-Glc thus formed can then enter the reaction again in a cyclical fashion until all of the galactose coming into the pathway is converted to glucose via Glc-1-P (13).

These enzymes constitute the Leloir pathway, named after Luis Federico Leloir, one of the major contributors to the identification of this pathway, and ensure that most of the ingested galactose is rapidly converted into Glc-1-P (1, 11, 14, 15). The Glc-1-P produced by the Leloir pathway is converted by the enzyme phosphoglucomutase into glucose-6-phosphate, which can

be further metabolized via one of the following pathways: i) the glycolytic pathway; ii) the pentose phosphate pathway; or iii) the gluconeogenic pathway (14, 16, 17).

Although the liver is the major organ of galactose metabolism, enzymes of the Leloir pathway have been detected in many cell types and tissues, including red and white blood cells, fibroblasts, and amniocytes (1, 11).

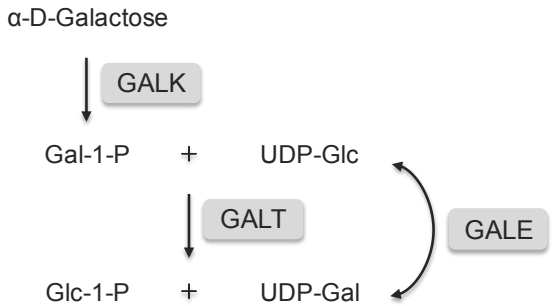


Figure 1.1. The Leloir pathway. Galactose is converted to glucose-1-phosphate (Glc-1-P) by the action of three consecutive enzymes: galactokinase (GALK), galactose-1-phosphate uridylyltransferase (GALT) and UDP-galactose 4'-epimerase (GALE), which constitute the Leloir pathway, the main pathway of galactose metabolism.

1.1.2. Hypergalactosemia

Hypergalactosemia refers to the presence of high levels of galactose in the blood ("emia" in Greek) and results from an impaired ability to metabolize galactose (1, 2). When an infant presents hypergalactosemia, the first question is whether this is primary hypergalactosemia or secondary hypergalactosemia (18).

There are several etiologies of secondary hypergalactosemia. They all center on liver dysfunction since the liver is the primary organ responsible for whole body galactose metabolism and disposal. In the presence of liver dysfunction, galactose ingestion leads to high levels of plasma galactose and erythrocyte Gal-1-P, even in the presence of normal erythrocyte GALT activity. Causes of liver dysfunction include: congenital infectious hepatitis; congenital hepatic arterio-venous malformations; patent ductus venosus; Fanconi–Bickel syndrome; and tyrosinemia type I, citrin deficiency (including citrullinemia type II), and other metabolic disorders producing hepatocellular disease (18).

Primary hypergalactosemia disorders result from a defect in the Leloir pathway, and include: GALK deficiency, GALT deficiency, and GALE deficiency (18). Although all forms of primary galactosemia are the result of impaired galactose metabolism, clinical presentation and severity vary widely among these patients (1).

1.1.2.1. GALT deficiency or type I galactosemia

Classic galactosemia or type I galactosemia (OMIM #230400) is caused by deficiency of galactose-1-phosphate uridylyltransferase (GALT; UDP-glucose:α-D-galactose-1-phosphate uridylyltransferase, EC 2.7.7.12), the second enzyme of the Leloir pathway, and is the most common cause of primary galactosemia (1, 19, 20). As the main theme of this thesis, it will be detailed further ahead.

1.1.2.2. GALK deficiency or type II galactosemia

In all living cells, α-D-galactose metabolism begins with the phosphorylation of galactose by galactokinase (GALK, EC 2.7.1.6), the first metabolic enzyme of the Leloir pathway (1, 21, 22). This enzyme phosphorylates the carbon-1 hydroxyl group of α-D-galactose at the expense of ATP (23). Despite GALK phosphorylating galactose, it does not belong to the same family of other sugar kinases, like hexokinase and glucokinase. GALK belongs to the family of small molecule kinases, the GHMP kinase superfamily, originally called GHMP due to its members constitution: galactokinase, homoserine kinase, mevalonate kinase and phosphomevalonate kinase (23-26). The members of this family share a highly conserved motif, which is thought to be involved in nucleotide binding (25, 26). In fact, the reaction mechanism is sequential, such that MgATP binding induces a conformational change in the enzyme, thereby creating a functional binding site for α-D-galactose; the reaction products, ADP and Gal-1-P, are released only after the binding of both substrates (1, 23, 24).

GALK deficiency (OMIM #230200) is an autosomal recessive disorder, associated with a markedly reduced erythrocyte GALK activity and an elevated blood galactose level (18). Deficiencies in this enzyme lead to type II galactosemia, of which the main symptom is early onset cataracts, and is less severe than the other primary hypergalactosemia disorders, GALT and GALE deficiencies (21, 24).

Provided GALK deficiency is detected early in life and patients are kept under a galactose-restricted diet throughout life, long-term complications should not develop (20, 24). This point is extremely important, as it provides compelling evidence that it is not the accumulation of galactose, but rather of Gal-1-P, or of some metabolic derivative of Gal-1-P, that leads to the complications, beyond cataracts, observed in treated patients with classic galactosemia (20). Indeed, the differences in clinical outcome between GALK and GALT deficiencies reflect the differences in tissue response to the characteristic changes in the levels of galactose metabolites originating from the respective enzyme deficiencies (8).

1.1.2.3. GALE deficiency or type III galactosemia

The third and final enzyme of the Leloir pathway is UDP-galactose 4'-epimerase (GALE, EC 5.1.3.2), which has a particularly important role, since it interconverts UDP-Gal and UDP-Glc, enabling endogenous synthesis of UDP-Gal in the absence of exogenous sources of galactose, and thereby rendering galactose a non-essential nutrient (1, 2, 20). In nearly all cells, GALE steady-state activity confers a ratio of UDP-Glc to UDP-Gal of about 3:1 (13). GALE also catalyzes the interconversion of a pair of larger substrates, uridine diphosphate-N-acetyl-galactosamine (UDP-GalNAc) and uridine diphosphate-N-acetyl-glucosamine (UDP-GlcNAc) (1, 20, 27, 28). GALE deficiency is the least understood type of galactosemia (27). However, it is relatively common in the African-American and Japanese populations with occurrence rates of 1/10,000 and 1/25,000 respectively (27).

GALE deficiency or type III galactosemia (OMIM #230350) is an autosomal recessive disorder, characterized by a partial to complete enzyme activity. GALE deficiency, initially thought to be a binary condition, is actually a continuum disorder, in which can be distinguished three forms of the disease: a benign peripheral form, where subjects only show impaired activity in their red blood cells (RBC) with consequent elevation of blood Gal-1-P levels; a severe generalized form, where impaired GALE activity and elevated Gal-1-P levels are detected in all tissues; and an intermediate form, where subjects demonstrate partially impaired GALE activity in nonperipheral cells and Gal-1-P levels in-between those of the mild and severe forms (1, 27, 29, 30).

One of the hallmarks of GALE deficiency is its variability: allelic heterogeneity, variability in the degree of enzyme defect, variability in which substrate pair is most affected by the enzyme defect (UDP-Gal/UDP-Glc or UDP-GalNAC/UDP-GlcNAc), and variability in most affected tissues. Variability in clinical outcome is therefore not surprising (1).

Patients with GALE deficiency accumulate intracellular Gal-1-P, and they may have hepatic and renal disease (31). Some present developmental or growth delay, cataracts or even seizures, emesis and hypoglycemia in response to lactose ingestion, but a cause and effect relationship between GALE deficiency and the clinical findings has not been established in each patient. The need for treatment remains unknown (18).

1.1.3. Accessory pathways of galactose metabolism

Even though the Leloir pathway is the predominant route for galactose disposal, three accessory pathways have been described (1, 11, 20). These include: i) reduction of galactose to galactitol by aldose reductase, ii) oxidation of galactose to galactonate, presumably by galactose dehydrogenase, and iii) conversion of galactose to UDP-Glc by the sequential activities of

GALK, UDP-glucose/galactose pyrophosphorylase (UGP) and GALE (Figure 1.2) (20). Under physiological conditions, these pathways are thought to collectively metabolize only trace quantities of galactose. In galactosemic patients, however, there is a raise in galactose intracellular levels, thereby approaching or exceeding the K_M parameters of the relevant enzymes, and causing these alternative pathways to assume a significant role (20).

The flow of galactose through these alternate pathways may limit the accumulation of potentially toxic intermediates, thereby mitigating negative outcomes, but, on the other hand, may also result in the accumulation of other end products that are themselves problematic (20).

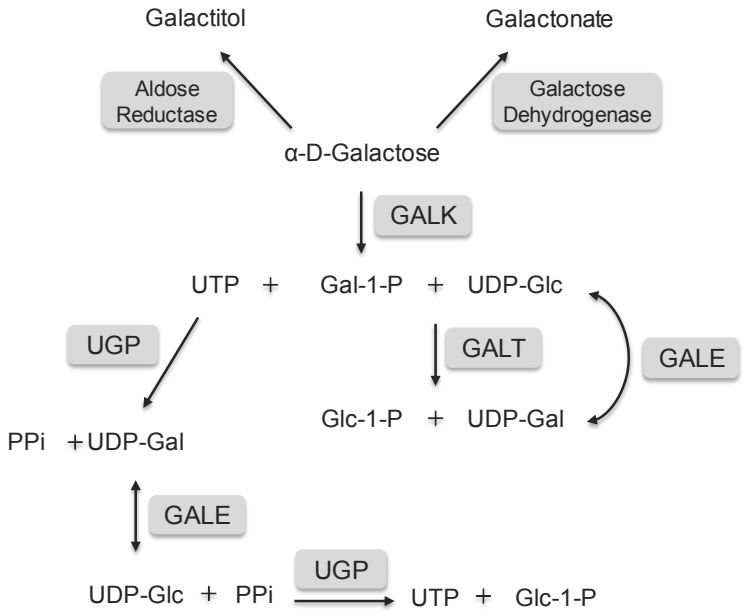


Figure 1.2. Acessory pathways of galactose metabolism. Galactose can be reduced to galactitol by aldose reductase, oxidized to galactonate, presumably by galactose dehydrogenase, or be converted into UDP-Glc, via the pyrophosphorylase pathway, by the sequential activities of GALK, UDP-glucose/galactose pyrophosphorylase (UGP) and GALE (PPi – pyrophosphate).

1.1.3.1. Galactose reduction

Galactose reduction is catalyzed by aldose reductase, the first enzyme of the polyol pathway (1, 32). This pathway consists of two enzymatic reactions involving aldose reductase (EC 1.1.1.21), which requires NADPH as a cofactor, and sorbitol dehydrogenase (EC 1.1.1.14), which uses NAD^+ as coenzyme (Figure 1.3) (32, 33).

Aldose reductase is a NADPH-dependent oxidoreductase, belonging to the aldo-keto reductase family (33, 34). Its transcription is highly regulated and appears to be tissue-specific (33). The proximal promoter of the aldose reductase gene, *AKR1B1*, presents several features of a regulated gene, namely a TATA box, a CCAAT box and an androgen-like response element,

whereas the region 1200 bp upstream the promoter presents three osmotic response elements (35, 36). In fact, aldose reductase is highly expressed the inner medulla of kidney, where it has an important role in the renal osmoregulation (33, 35, 37). Aldose reductase is also abundant in the sciatic nerve, lens, testis, heart, and cornea; and with lesser concentrations in liver, renal cortex, stomach, spleen, lung, small intestine, and colon (36). Several polymorphisms associated with increased susceptibility to diabetic complications have been described in the *AKR1B1* gene, as well as polymorphisms associated with inter-individual variability in aldose reductase levels in certain tissues (36).

Aldose reductase has broad specificity for monosaccharides, and can metabolize both glucose and galactose, yielding respectively sorbitol and galactitol (2, 32). Although both sugars are similarly processed, sorbitol can be converted to fructose via the sorbitol dehydrogenase, whereas galactitol is not a substrate for this enzyme and will, consequently, accumulate in cells and tissues (1, 11, 32). Galactitol accumulation produces a hyperosmotic effect, which results in a water uptake, via aquaporins 1 (AQ1) and 0 (AQ0), aiming to counter the osmotic gradient, leading to swelling of the cells, with consequent cell death and cataract formation (8, 24, 32).

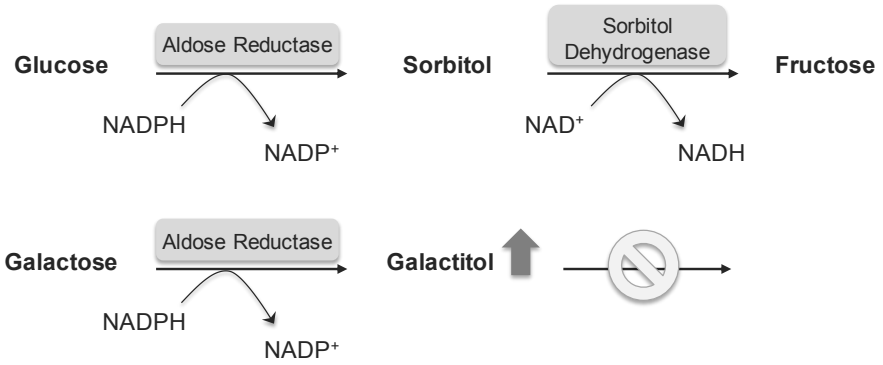


Figure 1.3. Galactose reduction into galactitol. The polyol pathway is comprised of the enzymes aldose reductase and sorbitol dehydrogenase. Aldose reductase has broad specificity for monosaccharides, and can metabolize both glucose and galactose, yielding respectively sorbitol and galactitol. Sorbitol is further metabolized by the second enzyme of the pathway, sorbitol dehydrogenase, whereas galactitol is not a substrate for this enzyme and will accumulate in cells and tissues.

1.1.3.2. Galactose oxidation

A further route of galactose disposal is its oxidation to galactonate. Galactonate was first identified in galactosemic patients by Bergen *et al.*, more than 40 years ago, and it is now well established that galactosemic patients produce galactonate (2, 38, 39). However, the route whereby galactonate is formed is still a controversial matter (1, 2).

The first description of the presence of galactose dehydrogenase in mammalian cells came from Cuatrecasas and Segal, who reported it in the liver of a variety of species and,

particularly, in the human liver (40). Indeed, galactose oxidation to galactonate is believed to reflect the activity of a soluble NAD^+-dependent galactose dehydrogenase (D-galactose:NAD^+oxidoreductase, EC 1.1.1.48) that converts galactose to galactonolactone, which then spontaneously or enzymatically converts to galactonate (1). However, the existence of this apparently highly active and soluble NAD^+-dependent galactose dehydrogenase has been questioned as a possible artifact, and the way by which galactonate is formed remains to be clarified (1, 2).

After galactose oxidation into galactonate, it can be excreted directly in this form or further metabolized via the reactions of the pentose phosphate pathway, where it is converted to β-keto-D-galactonate, which subsequently undergoes decarboxylation, liberating carbon-1 to form xylulose (Figure 1.4) (1, 13, 41).

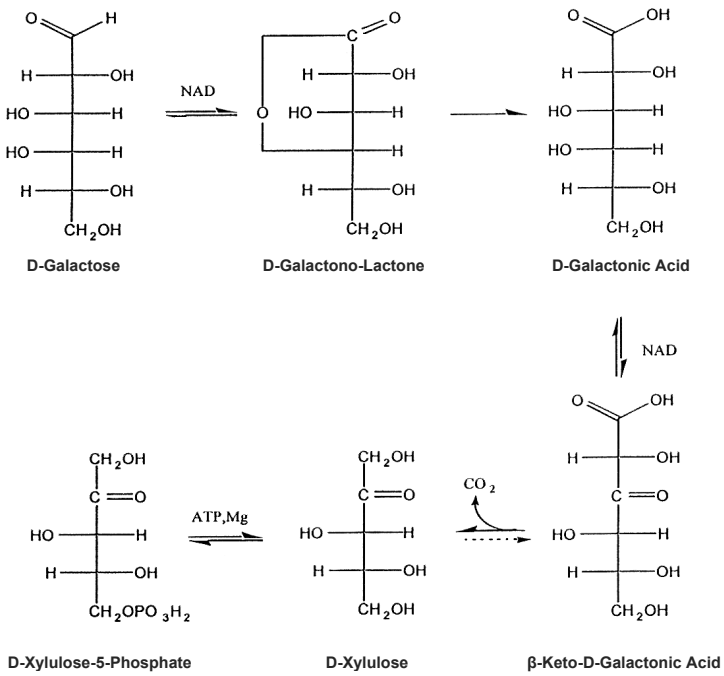


Figure 1.4. Galactose oxidation into galactonate. Galactose is first converted into galactonolactone, which then spontaneously or enzymatically is converted into galactonate. Galactonate can be excreted directly in this form or further metabolized via the reactions of the pentose phosphate pathway, where it is converted to β-keto-D-galactonate, which subsequently undergoes decarboxylation, liberating carbon-1 to form xylulose. Adapted from Wehrli *et al.*, 1997 (41).

In an attempt to quantitate galactose oxidation through this pathway, Segal and Cuatrecasas developed a study in which [1-^{14}C]galactose and [2-^{14}C]galactose were administered intravenously to both controls and patients, and its oxidation to $^{14}CO_2$ was quantified. They reported galactosemics oxidized greater amounts of administered [1-^{14}C]galactose to $^{14}CO_2$ than [2-^{14}C]galactose comparatively to healthy subjects, who showed no difference in the $^{14}CO_2$ yield from these specifically labeled galactoses, and put forward the hypothesis that the carbon-2

oxidation represented residual GALT activity, whereas carbon-1 represented direct oxidation of galactonate, which thereby plays an important role in the galactose metabolism of classic galactosemic patients (41, 42).

More recently, Berry *et al.* described a patient with a *GALT* deletion in homozygosity that could oxidize [1-^{13}C]galactose to $^{13}CO_2$ over a 24 hours period, thus providing unambiguous evidence for an alternate pathway of galactose utilization in the absence of any possible GALT mediation (42, 43).

1.1.3.3. Pyrophosphorylase pathway

The pyrophosphorylase pathway was first speculated by Isselbacher in 1957, as an explanation for the perceived ability to tolerate an increasing intake of galactose with age (44, 45). This pathway involves the phosphorylation of galactose to Gal-1-P, catalyzed by GALK, followed by a reversible UTP-dependent pyrophosphorylase reaction, in which the galactose moiety is incorporated into UDP-Gal (11, 45). UDP-Gal is then epimerized by GALE into UDP-Glc, from which a second pyrophosphorylase reaction generates Glc-1-P and UTP (Figure 1.2) (1, 44).

Neither the galactitol pathway nor the galactonate pathway involve any of the three Leloir enzymes (1, 20). This third alternative pathway, however, depends on the action of GALK and GALE. Indeed, the two reactions that are unique to this alternative pathway are the pyrophosphorylation reactions, both catalyzed by a single enzyme – UTP-dependent glucose/galactose pyrophosphorylase (UGP), also sometimes referred to as UTP-hexose-1-phosphate uridylyltransferase (EC 2.7.7.10) (1, 20, 44).

Essentially, conversion of Gal-1-P to UDP-Gal may be catalyzed either by GALT or by UGP (46). This observation led to the proposal that the enzyme UGP provides a GALT-independent backdoor route for galactose metabolism in classic galactosemia, catalyzing the conversion of Gal-1-P to UDP-Gal, albeit via a different mechanism and at a much lower rate than GALT (22, 47). Thus, UGP would play an important role in GALT deficiency, since it would reduce the intracellular levels of Gal-1-P and would replenish UDP-Gal (16).

However, in 1969, Gitzelmann presented evidence that the pyrophosphorylase pathway is probably more active in the reverse direction than that initially considered by Isselbacher, presenting scientific evidence of Gal-1-P production in GALT-deficient cells deprived of galactose, and suggested for the first time, an endogenous production of galactose (46, 48).

1.1.4. Endogenous synthesis of galactose

Gitzelmann was the first to propose the existence of a pathway for the endogenous synthesis of galactose and its derivatives (1, 48). Evidence for endogenous synthesis of galactose came from a case of a pregnant woman, to whom had been prescribed a galactose-restricted diet on the last six months of pregnancy. After birth, the child was put immediately on a soybean formula but, unexpectedly, presented gradually increasing levels of Gal-1-P in RBC (48).

Intrigued by this case, Gitzelmann used erythrocytes from galactosemic patients to show, *in vitro*, that Gal-1-P could be generated from UDP-Gal. Since UDP-Gal may be synthesized from UDP-Glc by GALE, he later introduced the concept of ‘‘self-intoxication’’, proposing that a pyrophosphorylase reaction with UDP-Gal and pyrophosphate as substrates is the biochemical basis for *de novo* synthesis of galactose (48, 49). According to this mechanism, the critical step is the epimerization of UDP-Glc to UDP-Gal, which produces galactose-containing substances necessary for normal growth and development independent of dietary galactose intake (50).

More recent studies have led to the proposal that the mechanism of endogenous galactose production involving the pyrophosphorylase pathway includes UDP-Glc epimerization by GALE to UDP-Gal, which is then cleaved by UGP to produce UTP and Gal-1-P, which finally is cleaved by inositol monophosphatase (IMPase) to yield free galactose (1, 16, 51, 52). Additionally, galactose may also derive from the turnover of glycoconjugates, in which lysosomal hydrolysis of glycoproteins and glycolipids releases free galactose (50, 53).

Several studies have provided quantitative evidence for whole body *de novo* galactose synthesis in healthy subjects and classic galactosemic patients (38, 49, 50, 54). In GALT-deficient patients, the calculated production rate of galactose was between 0.49 and 1.09 mg/kg per hour (1). Later studies aimed at the elucidation of the age dependence of endogenous galactose formation in galactosemia, and concluded the endogenous galactose burden is considerably higher in infants and children than in adults, gradually relieved with increasing age until adulthood, and that, therefore, the tolerance for dietary galactose increased with age in galactosemic patients (1, 38, 49).

A more recent study further demonstrated that the rate of endogenous galactose production is not influenced by short-term exogenous galactose intake, neither in patients with classic galactosemia nor in healthy individuals (55). As postulated by Gitzelmann and colleagues in 1975, endogenous production of galactose may be vital for biosynthesis of membrane constituents, but, at the same time, it may represent a process toxic to the galactosemic cell, since galactose, Gal-1-P, galactitol and galactonate accumulate, and may lead to a state of auto-intoxication, thereby contributing to long-term complications (50, 53, 54).

1.2. Classic galactosemia or GALT deficiency

Classic galactosemia is a potentially lethal disease of galactose metabolism, presenting an autosomal recessive pattern of inheritance (20, 56). It is secondary to a severe reduction or absence of GALT activity, caused by mutations in the *GALT* gene (57).

1.2.1. *GALT* gene

The *GALT locus* is located in the short arm of chromosome 9, in the region 9p13 (1, 58). The human *GALT* cDNA and gene have been cloned and characterized, which provided a means of studying the genetic basis of classic galactosemia (1, 59, 60).

The *GALT* gene is arranged into 11 exons, spanning ~4.0 kb of genomic DNA (NG_009029.1), and the *GALT* cDNA is 1407 bp in length (NM_000155.2). Some exons are highly conserved among species, notably exons 6 (which encodes the active site), 9 and 10, whereas others, such as exons 1, 2, 5, and 7, are poorly conserved (1).

Analysis of approximately 4 kb of the *GALT* promoter identified three GC-rich Sp1 sites, an imperfect non-palindromic AP-1 site (TCAGTCAG at -126 to -119), a CAAT box, two E-box motifs spaced 23 bp apart (CAGGTG at -146 to -141 bp and CACGTG at -117 to -112 bp); and no TATA box sequence (1, 61-63).

Despite the highly inducible nature of GALT expression in *Escherichia coli (E. coli*) and yeast, the mammalian gene appears to be expressed as a housekeeping gene (1). Indeed, there is no clear evidence of galactose induction and the 5' flanking region of *GALT* is very GC rich and lacks a TATA box, both features observed in housekeeping genes (1, 64). The classification of *GALT* as a housekeeping gene, however, does not imply lack of transcriptional regulation (63).

An *Alu* repeat motif with evidence of length variation has been identified in intron 10 (65). It consists of a polyadenine nucleotide repeat with three length variants, (A)17, (A)24 and (A)29 at frequencies of 47.5%, 50.0%, and 2.5%, respectively. These length variants have defined associations with particular *GALT* alleles. The p.Q188R and p.K285N mutant alleles were associated with the (A)17 and (A)24 *Alu* variants, respectively, while the (A)29 variant was found only in samples carrying the p.N314D allele (explained below) (1, 65).

At present, there are 266 different variations described at the *GALT locus* (available at: http://www.arup.utah.edu/database/GALT/GALT_display.php, last surveyed December 2013) (66). Out of the described variations, the great majority are missense mutations (62.0%), a common observation in genetic disorders. Additionally, 19 silent (7.1%), 17 nonsense (6.4%), 17 small deletions (6.4%), 6 insertions (2.3%), 3 indel (1.1%), 3 large gene deletions (1.1%), 2 no-stop changes (0.8%) and 1 frameshift (0.4%) mutations have also been reported. Regarding

non-coding regions, 15 splice site variations (5.6%) have been described. The remaining 18 variations are classified either as benign (16; 6.0%) or as uncertain (2; 0.8%) (66).

Even though a few of these mutations are common, and usually highly prevalent in particular populations or ethnic groups, most are rare (1). Of those mutations that have been characterized *in vitro*, it is clear that some are true nulls, whereas others are hypomorphs (1). Furthermore, classic galactosemic patients show phenotypic variability, which can be partially explained by the high allelic heterogeneity and the high prevalence of compound heterozygotes (63, 67-70).

1.2.2. Mutations at the *GALT locus*

GALT sequencing opened new avenues of research in classic galactosemia, allowing molecular studies of galactosemic patients and thereby providing new insights into this disorder. Indeed, mutational analysis of galactosemic patients has provided new understanding of the GALT protein, by pinpointing functionally and/or structurally important regions of the enzyme, and has uncovered a multiplicity of mutations and polymorphisms, which is noteworthy in light of the clinical heterogeneity of classic galactosemia (19, 60).

1.2.2.1. p.Q188R

The p.Q188R is the most frequent *GALT* mutation associated with classic galactosemia, accounting for 63%–64% of all galactosemic alleles in the Caucasian population, and was first reported in 1991 (1, 71, 72). It results from an A→G transition at nucleotide c.563 in exon 6 (CAG → CGG, conserved residue), leading to a substitution of a glutamine to an arginine at amino acid 188 of the GALT protein (1, 63, 66, 70, 73, 74).

This mutation is the most common mutation among people with European ancestry, particularly in Northwestern Europe (71, 72). The mutation shows a gradient increasing frequency through Europe in a North-Westernly direction, reaching its highest frequency in Ireland, where it accounts for 92%–94% of galactosemic alleles (71, 72, 75-77). In fact, p.Q188R was found to be the sole mutant allele among the Traveller population of the Republic of Ireland, where its incidence is estimated to be 1 in 480 (75). Interestingly, p.Q188R has never been identified in Asian patients or in Asian descendant patients (1, 70, 71, 78, 79).

This mutation is particularly remarkable due to its location, since it occurs two amino acids downstream from the active site His-Pro-His sequence (63). Glutamine located at 188 is essential for the imidazole-N interaction with the uridylyl intermediate, and the change in charge

imparted by p.Q188R has been proposed to alter protein conformation in the active site region (74, 80, 81).

When this mutation was initially expressed in monkey COS cells, it yielded approximately 10% of the wild-type GALT control activity. This incorrect observation was caused by endogenous GALT present in the monkey COS cells (73, 82). The same mutation expressed in bacterial and yeast systems, both devoid of endogenous GALT, presented essentially no GALT activity, and transformed lymphoblasts from homozygous-p.Q188R galactosemic patients presented less than 2% of control activity (69, 74, 82).

Patients homozygous for the p.Q188R mutation demonstrate essentially no residual GALT activity in RBC, and homozygosity for this allele represent a significant risk factor for developmental verbal dyspraxia and has been associated with a poor cognitive outcome (70, 83-85).

1.2.2.2. p.K285N

The second most common mutation of European origin is p.K285N (71, 86-89). Conversely, unlike p.Q188R, this mutation is particularly frequent in countries of Central and Eastern Europe, where it accounts for 26%–34% of galactosemic alleles (86-89). In fact, the Czech, Slovak, Polish and Austrian galactosemic populations present the highest frequencies of p.K285N mutation relatively to other European populations, which has led Kozák *et al.* to suggest this mutation could have a Slavic origin (86-89). This hypothesis was later confirmed by Lukac-Bajalo *et al.* (86). Notably, this mutation has never been identified in the Portuguese population (80). Analogously to p.Q188R, p.K285N has never been identified in Asian patients (70, 71, 78, 79).

The p.K285N mutation was first reported in 1992, and results from a G→T transition at nucleotide c.855 in exon 9 (AAG → AAT, non-conserved residue), leading to a substitution of a lysine to an asparagine at amino acid 285 of the GALT protein (1, 63, 66, 70, 74). This mutation is associated with essentially no GALT activity in RBC, and with a severe clinical phenotype, namely poor cognitive outcome and pronounced speech abnormalities (70, 89).

1.2.2.3. p.S135L

The p.S135L mutation is found almost entirely in individuals of African origin, accounting for approximately 48% of African American and 91% of South African *GALT* alleles (70, 90, 91). The p.S135L mutation was first reported in 1992 by Reichardt *et al.*, and results from a C→T transition at nucleotide c.404 in exon 5 (TCpG → TTG , non-conserved residue),

leading to a substitution of a serine to a leucine at amino acid 135 of the GALT protein (1, 63, 66, 70, 74, 92).

Notably, this mutation appears to exhibit some tissue specificity: homozygous patients have essentially no GALT activity in RBC, presenting approximately 5.5% of control activity in leukocytes, while liver and intestinal mucosa biopsy specimens express about 10% of normal GALT activity (91). Additionally, homozygous are capable of converting galactose to CO_2 at a rate comparable to control subjects (50, 91). In fact, this mutation is said to have a protective effect and p.S135L homozygosity is often associated with milder clinical outcomes (83, 91, 93). Nevertheless, patients carrying p.S135L should be diagnosed very early in life and diet should be implemented immediately (90, 91, 94).

1.2.3. GALT protein

GALT is an ubiquitous enzyme in living organisms and highly conserved through evolution (69, 74, 95). Comparison of the human (NP_000146.2) with the *E. coli* (NP_415279.1), *Saccharomyces cerevisiae* (NP_009574.1) and rat (NP_001013107.1) GALT amino acid sequences revealed overall identities of 51%, 40%, and 88%, respectively.

The three-dimensional structure of the human GALT has still not been experimentally solved. However, the *E. coli* GalT structure has been solved by X-ray crystallography, which revealed this enzyme is a dimer with two active sites, each one formed by amino acid residues from both subunits (12, 96-98). Furthermore, a direct comparison of the bacterial and human GALT sequences revealed these enzymes are superimposable, starting from the 20th amino acid of the human sequence, i.e., an extra 20 amino acids at the amino-terminus are present in the deduced human GALT sequence (63, 96, 99, 100).

GALT is a member of the transferase branch of the histidine triad (HIT) family of enzymes, so-called because of a hisϕhisϕhisϕϕ motif (ϕ referring to a hydrophobic amino acid) (101, 102). The active human enzyme is a homodimer (86.6 kDa), in which each monomer is constituted by a polypeptide of 379 amino acids (43.3 kDa). The active site sequence His-Pro-His is conserved among all known uridylyltransferases, and was firstly identified in *E. coli* GalT at residues 164 to 166, corresponding to residues 184 to 186 in the human sequence (63, 96, 99, 100).

GALT catalyzes the transfer of an uridyl group from UDP-Glc to Gal-1-P through a double displacement mechanism whereby a histidine residue is transiently modified (Figure 1.5) (12, 97, 103, 104). Detailed structural analyses of the dimeric protein from *E. coli* have revealed that each monomer displays a “half-barrel” motif and that His166 is the active site base (101, 105). In the first step of the ping-pong mechanism, a pair of electrons on the nitrogen ring of a

histidine residue (His166 in *E. coli* and His186 in humans) in the active site attacks the α-phosphate of UDP-Glc (12, 106). The result of this reaction is an UMP-enzyme adduct and Glc-1-P, which is released. In the second stage of the reaction, the UMP group is displaced from the enzyme by Gal-1-P, thereby regenerating the active site of the enzyme and producing the second product, UDP-Gal (12, 103, 106).

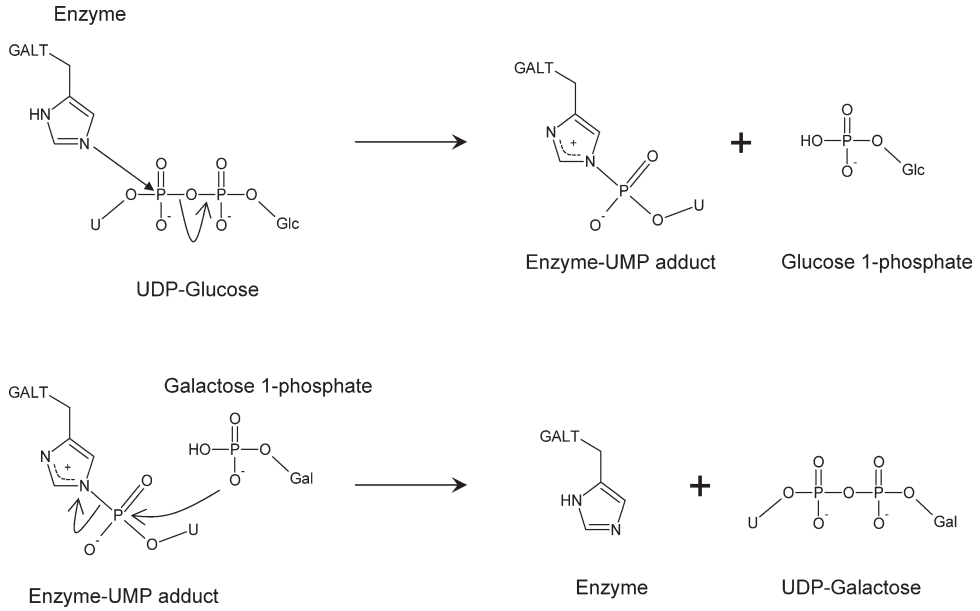


Figure 1.5. The catalytic mechanism of GALT. In the first step of the reaction, the Nε2 of the His186 attacks the α-phosphate of UDP-Glc, releasing Glc-1-P and forming the covalent uridylyl-enzyme intermediate; in the second half of the reaction, the intermediate reacts with Gal-1-P to produce UDP-Gal. Reproduced from McCorvie and Timson, 2011 (12).

Each monomer of the bacterial enzyme has been described as containing one zinc (II) ion and one iron (II) ion, in which the zinc ion is believed to be essential for activity, whereas the iron ion is thought to play a role in stabilizing the structure (22, 107, 108). The iron ion is coordinated by three histidine residues and a glutamate residue in a hydrophobic region close to the subunit interface, whereas the zinc ion is coordinated by two cysteine and two histidine residues and helps to orientate residues in, and close to, the active site of the enzyme. However, it has been noted that the human enzyme is different from *E. coli* GalT in that the zinc binding residues are not conserved and hence there is a considerable possibility that it may bind only one metal ion per monomer. In fact, the crystal structure predicted His115 to be the bacterial residue to serve as a ligand for zinc, which would correspond to Ser135 in the human GALT, raising thus the important question of whether human GALT is a metalloenzyme, and, if so, whether Ser135 could serve as a ligand to zinc in the human GALT (109, 110). In fact, considering serine is chemically unlikely to suffice as a ligand for zinc, it was suggested that the neighboring His132 would most likely serve as the corresponding zinc ligand in human GALT (109). To clarify these

questions, Wells *et al.* evaluated the zinc content in both wild-type and S135L human GALT, and demonstrated that both proteins contained stoichiometrically indistinguishable levels of zinc (109).

1.2.4. Duarte and Los Angeles variants

The first known variant of classic galactosemia was the Duarte variant, which was first described by Beutler *et al.* in 1965 and named after the city in which it was first defined (1, 62, 74).

The Duarte allele – or Duarte-2, D2 or simply D – is biochemically defined by an isoform distinguishable by gel electrophoresis and isoelectric focusing, and by a reduction of enzyme activity to approximately 50% of the wild-type allele levels (71, 111-113). Indeed, hemolysates of heterozygotes for Duarte allele present 75% of GALT activity, homozygotes present 50% activity, and compound heterozygotes for the Duarte allele and a classic galactosemia allele present 25% activity (74, 114).

In the early 1970's, Ng and colleagues reported several individuals that exhibited the same electrophoretic mobility profile than the Duarte variant, but presented mildly elevated rather than impaired activity, which led them to define a new variant – the Los Angeles variant, also known as Duarte-1, D1 or LA variant (62, 115).

After the molecular characterization of the *GALT* cDNA, Reichardt and Woo reported, for the first time, the p.N314D mutation, which results from an A→G transition at nucleotide c.940 in exon 10 (AAC → GAC, non-conserved residue), and leads to a substitution of an asparagine to aspartate at amino acid 314 of the GALT protein (19, 66, 70, 74, 116). This mutation was later shown to be associated with the Duarte and Los Angeles alleles, which posed an enigma, as p.N314D was associated in the Duarte variant with a decrease of enzymatic activity, and in the Los Angeles variant, with an increase of enzymatic activity (62, 63, 111).

The functional significance of the p.N314D mutation in both Duarte and Los Angeles alleles was a point of intense controversy and examination in the 1990's. p.N314D expression in a COS cell expression system showed 129% of the control activity, and so it was classified as a polymorphism. Later studies using a yeast expression system supported that this substitution produced an enzyme with activity and abundance levels indistinguishable from those of the wild-type protein. Additionally, this study demonstrated that the p.N314D mutation altered the protein's charge and therefore accounted, *per se*, to the aberrant migration on native starch and altered migration on isoelectric focusing electrophoresis (1, 19, 78, 117).

As the expression systems failed to detect impaired GALT activity with the p.N314D mutation, other undetected causes for altered enzyme activity were speculated, such as intron or

promoter variations (117). A study by Langley *et al.* identified the neutral polymorphism p.L218L (c.652C>T; CTA → TTA) in *cis* with p.N314D exclusively in the Los Angeles phenotype and, thus, suggested that the increased enzymatic activity in this variant was actually due to a "codon bias" increased translation, with subsequent increased abundance of protein, since the rarer leucine codon TTA could produce a more efficient translation than the CTA codon. Additionally, this study identified no point mutation other than p.N314D in fully sequenced Duarte alleles, which reinforced the later study by Lai *et al.* that showed that p.N314D caused a destabilization of the GALT protein and proposed this was the only molecular pathogenic mechanism in this variant, since the mRNA abundance was not affected in the Duarte alleles (118). However, later studies demonstrated the p.N314D allele does indeed give rise to the two variant forms depending on the presence of additional base changes: the Los Angeles allele carries p.N314D in *linkage disequilibrium* exclusively with the translationally silent mutation p.L218L, whereas the Duarte allele carries p.N314D in *linkage disequilibrium* with c.378-27G>C (IVS4-27G>C), c.507+62G>A (IVS5+62G>A), c.508-24G>A (IVS5-24G>A) and a 4-bp deletion in the 5' untranslated region 116 to 119 bases upstream from the first methionine codon (c.-119_-116delGTCA) (1, 66, 76, 87, 112, 113, 117, 119-121).

As previously mentioned, the *GALT* promoter presents an AP-1 site (TCAGTCAG from -126 to -119 bp) and two E-box motifs (at -146 to-141 bp – CAGGTG – and at -117 to -112 bp – CACGTG perfect). The -119 to -116 GTCA deletion removes the first two nucleotides of the -117 to -112 bp E-box motif (CA, located at -117 and -116 bp), which however does not change the E-box motif of CACGTG from -117 to -112 bp, since there are three repeats of GTCA in sequence, and the middle tetrad GTCA fills the deleted CA of the E-box motif (Figure 1.6). There is, however, a reduction in the spacing between the two E-box motifs that most likely alters *trans*–acting factors binding and thus impairs the positive regulation of *GALT* gene expression. Furthermore, the -119 to -116 GTCA deletion also affects the AP-1 site from -124 to -117 bp with the sequence TCAGTCAG, as it removes the terminal guanine of this octomeric domain, which can also reduce *GALT* gene expression (61, 122). In fact, a computer search for potential regulatory elements in the area of the deletion showed that it potentially eliminates the binding sites for the two activator proteins AP1Q2 and AP1Q4 (1, 61, 119, 122).

Functional analyses carried out by two independent groups – Trbušek *et al.* and Elsas *et al.* – strongly support the hypothesis that the GTCA deletion in the *GALT* promoter is indeed the major factor contributing to the diminished expression and activity of the Duarte variant (1, 61, 122).

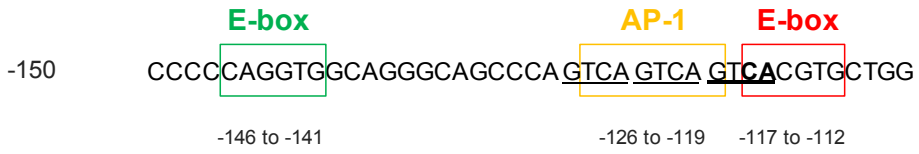


Figure 1.6. The 4-bp deletion in the 5′ untranslated region of the *GALT* gene. The *GALT* promoter presents an AP-1 site from -126 to -119 bp (TCAGTCAG) and two E-box motifs from -146 to -141 bp (CAGGTG) and from -117 to -112 bp (CACGTG). The -119 to -116 GTCA deletion removes the first two nucleotides (in bold) of the -117 to -112 bp E-box motif. The middle tetrad GTCA fills the deleted CA of the E-box motif; however, the spacing between the two E-box motifs is reduced, and the terminal G of the AP-1 site is removed. Nucleotide +1 is the A of the ATG-translation initiation codon. Adapted from Elsas *et al.*, 2002 (62).

At present, p.N314D is a common variant with an allele frequency of approximately 11% in European populations (*Centre d'Etude du Polymorphisme Humain*) and lower frequencies in other populations, for a pan-ethnic frequency of 8.3% (71, 112).

Interestingly, a study on the origin and distribution of the Duarte allele demonstrated that the p.D314 allele is, in fact, the ancestral allele, and that its distribution pattern suggests the p.D314N variant arose in Africa very early in human evolution. Furthermore, the persistence of the p.D314 allele in European and, to a lesser extent, Asian populations most likely reflects a founder effect acting upon descendants after the human migration out of Africa (112).

1.2.5. Abnormal accumulation of metabolites

Patients with classic galactosemia present a characteristic pattern of biochemical abnormalities: Gal-1-P, the substrate of the GALT enzyme, accumulates within cells, and surplus galactose is reduced to galactitol or oxidized to galactonate (123). Indeed, severe GALT deficiency results in the accumulation of toxic galactose metabolites and causes a decrease in UDP-Gal, a sugar required for the galactosylation of glycoproteins and glycolipids (77).

1.2.5.1. Galactose

Galactose is a hexose that differs from glucose only in the configuration of the hydroxyl group at the carbon-4 position (8). However, galactose is less stable, more susceptible to the formation of nonspecific glycoconjugates, and is incorporated into glycogen more efficiently than glucose, which would explain why glucose is the primary metabolic fuel for humans and most other species (1, 124, 125).

The human body is able to metabolize large amounts of galactose, as evidenced by rapid clearance of galactose from blood: 50% of radiolabeled galactose is found in glucose pools within 30 minutes after being administered intravenously (126). For nursing infants, galactose

conversion to glucose is crucial to maintain euglycemia, since 40% of calories derives from the hydrolysis of lactose to galactose and glucose (126). Indeed, the rapid conversion of galactose to glucose is of great biological importance, and explains the evolution and strong conservation of the Leloir pathway (1).

In opposition to healthy infants who rapidly metabolize galactose to glucose after a milk meal, newborn infants with classic galactosemia will accumulate galactose in the blood, as well as in other cells and tissues, and will present high galactose urinary concentrations, whereas blood glucose levels can fall to hypoglycemic levels (1, 8). Upon initiation of dietary galactose restriction, blood galactose levels fall quickly, but always remain elevated when compared with controls. In fact, in a study by Ning and Segal, the mean value of plasma galactose in 27 galactosemic patients was less than 2-fold higher than in controls; it was, however, about 4-fold higher than the levels in healthy infants on a galactose-restricted diet (1, 127).

When compared to other metabolites, blood galactose pool is very small. It may be that the galactose pool is maintained at its steady-state level because there is a high rate of conversion to RBC Gal-1-P and to alternate-pathway metabolites such as galactitol and galactonate in tissues (127).

1.2.5.2. Galactose-1-phosphate

Once galactose enters the cell, it is converted to Gal-1-P by GALK. In GALT deficiency, Gal-1-P cannot be further metabolized and accumulates in cells and tissues, namely RBC, liver, kidney, brain, tongue, adrenal, heart, and ocular lens (1, 11). In fact, Gal-1-P accumulation is a cardinal feature of the classic galactosemia phenotype, and it was its accumulation that first helped to pinpoint the position of the metabolic block in this disorder (1, 50, 128).

Schwarz *et al.* were the first to describe the accumulation of Gal-1-P in GALT-deficient patients' cells, and to conclude that the enzyme block was at the point of conversion of Gal-1-P to Glc-1-P (128). This observation was quickly confirmed by Kalckar *et al.*, who clearly demonstrated that the clinical picture was due to GALT deficiency and not to any deficiency of UDP-Glc (30). Furthermore, Komrower *et al.* used RBC as a representative of body tissue to evaluate Gal-1-P accumulation and, ever since, its measurement in RBC has become standard practice to assess the biochemical results upon initiation of a galactose-restricted diet in recently diagnosed newborns, as well as to monitor the adherence to diets during later life (30, 128, 129).

At the time of diagnosis, patients fed with galactose-containing milk present high Gal-1-P concentrations in RBC, which can be over 2.5-6.5 mM (1, 11). Following dietary restriction of galactose, Gal-1-P values in RBC fall dramatically, although they always remain elevated when compared to healthy controls (0.1-0.2 mM down to undetectable) (1, 11).

Gal-1-P levels in RBC at the time of diagnosis were initially thought to reflect the amount of Gal-1-P to which the infant had been exposed in the newborn period (130). However, it is now known that Gal-1-P also accumulates abnormally in patients prior to birth: two galactosemic fetuses at 20 weeks' gestation exhibited levels of Gal-1-P in their livers comparable with levels seen in newborns that died from the disorder (1).

1.2.5.3. Galactitol

Galactitol has been found in high levels in blood, tissues and urine of galactosemic patients, and thereby also constitutes a key element in the biochemical phenotype of affected individuals (1, 131-134). Galactitol accumulates in tissues with a high activity of aldose reductase, such as lenses and ovaries. Its effects are particularly notable in the lens, where galactitol causes osmotic swelling and cataract formation (135). Additionally, high levels of galactitol have also been found in the brain, where galactitol concentration is associated with diffuse white matter abnormalities, and in the liver of galactosemic patients (131, 133, 136, 137).

Galactitol builds up until dietary treatment is started, and is then cleared via urine, but it persists and does not entirely disappear most likely because of the generation of galactose by endogeneous metabolic reactions (132). Several studies have been conducted to understand galactitol relation to galactose and Gal-1-P levels, and its role in the pathogenesis of classic galactosemia.

In 1995, a study by Jakobs and colleagues demonstrated urinary galactose and galactitol excretion in controls is diet- and age-dependent, with the highest excretion at a younger age (132). In fact, untreated patients with null GALT activity excrete enormous amounts of galactitol in their urine that can reach 69,000 mmol/mol creatinine in the newborn toxic period. With dietary lactose restriction, however, urinary galactitol excretion decreases, ranging from 358-900 mmol/mol creatinine (n=3; <1y), while controls present a range of 3-81 mmol/mol creatinine (n=20; <1y) (132). Additionally, urinary galactitol excretion further declines with age, ranging from 120-890 mmol/mol creatinine in children with 1 to 6 years old (n=23) and ranging from 45-350 mmol/mol creatinine in children over 6 years (n=49) (132).

In a later study, Palmieri and colleagues reiterated the conclusions of Jakobs *et al.*, distinguishing three major age groups delineated according to urinary galactitol excretion (133). Additionally, Palmieri *et al.* found that classic galactosemic patients homozygous for the severe mutation p.Q188R, on a lactose-restricted diet, excrete 5 to 10-fold more urinary galactitol than controls, and over the age of 6 excrete on average about half the amount of younger patients (133). This study also reported, for the first time, a good agreement between RBC Gal-1-P and urinary galactitol excretion in a p.Q188R/p.Q188R patient, evaluated from the time the lactose-

restricted diet was instituted and over a 4-year period. The trends for the values of both compounds appeared similar over the 4-year period evaluation, with the high levels of both compounds from the newborn period decreasing over time, and reaching a relatively constant level by 6 months of age (133). Furthermore, Palmieri *et al.* also established a statistical correlation between RBC Gal-1-P and urinary galactitol excretion in 47 galactosemic patients (corresponding to 112 simultaneous measurements), with, however, a wide scatter, as pointed out by the authors (133). A previous study by Schweitzer *et al.* had already attempted to establish a correlation between these two parameters, but without success (133, 138).

In opposition to the decrease in urinary galactitol with age, Palmieri *et al.* further demonstrated that galactosemic patients do present increased galactitol levels in the plasma, which, however, remain in a narrow concentration range and do not vary with age or genotype (133). In contrast, healthy subjects present undetectable levels of plasma galactitol (133).

Similarly to Gal-1-P levels, galactitol has been found in the amniotic fluid of affected fetuses, which has been demonstrated not to be influenced by the strictness of the maternal dietary regime (30, 133).

1.2.5.4. Galactonate

It is still a matter of debate whether galactonate is rather disposed completely via the urine or partially degraded via the so-called galactonate pathway (16). Nevertheless, galactonate is generally regarded as terminal product of this alternative pathway, since it is excreted in significant amount in patients' urine (2).

In fact, analogously to galactitol, galactonate has been detected in urine and in several tissues of galactosemic patients (139). Unlike galactitol, however, galactonate is not detectable in the plasma of galactosemic patients, most likely because it is below the level of detection, and is not elevated in the RBC of all galactosemic patients, as several patients present RBC galactonate within the normal range (41, 127, 139).

Urinary galactonate excretion has not been thoroughly studied as galactitol urinary excretion (41). Nevertheless, it is well established that galactosemic patients present markedly increased urinary excretion of galactonate and, a study by Schadewaldt *et al.* reported a high statistically significant inverse correlation between urinary galactonate excretion with age and galactose endogenous synthesis (38, 42).

Therefore, urinary galactonate measurement offers a new parameter in addition to RBC Gal-1-P assay and urinary galactitol excretion to assess the metabolic status of galactosemic patients (41).

The levels of RBC Gal-1-P, galactitol and galactonate would be expected to reflect different aspects of galactose metabolism. Gal-1-P is a metabolite accumulating behind the block imposed by GALT deficiency while galactitol and galactonate are formed by two different enzymes with different metabolic fates. The measurement of all three RBC metabolites could further expand the clinical practice of RBC Gal-1-P determination as a means of following the efficacy of galactose-restricted diets in the treatment of galactosemics and monitoring dietary adherence (139). Furthermore, these pathways, and the genes and enzymes that comprise them, represent logical candidate modifiers of patient outcome in galactosemia, and represent potentially novel targets for small-molecule therapeutic intervention (20).

1.2.6. Epidemiology

The incidence of classic galactosemia in Western Europe has been estimated to be between 1/30,000-60/000, being relatively prevalent (1/23,775) in the Turkish population, because of a high consanguinity marriage rate, and relatively rare in Sweden (1/100,000) (1, 140, 141). Much higher frequencies have been observed in the Irish Travellers, in which the relative frequency among is 1/480, most probably due to the high degree of consanguinity and to a possible founder effect within the Traveller community (77).

The incidence in the United States, with most cases ascertained by newborn screening, is currently estimated at 1/50,000; while in Canada is reported to be in the range of 1/35,000 to 1/73,296 (142, 143).

In Japan, the frequency of the classic galactosemia is $1/20^{th}$ less than that in the United States (78).

1.2.7. Acute clinical presentation

Infants with classic galactosemia generally appear asymptomatic at birth. However, after one to two weeks of galactose ingestion through milk feeding, children start developing life-threatening symptoms that, if undiagnosed and untreated, may lead to death (144-146).

Initial symptoms include poor feeding with poor weight gain, vomiting and diarrhea, lethargy, and hypotonia (1). Failure to thrive is the most common presenting symptom and occurs in almost all patients, whereas vomiting or diarrhea may occur in 95% of patients (126). Progression of this acute neonatal toxicity syndrome may include the development of *E. coli* septicemia in the second week of life, hyperchloremic metabolic acidosis with aminoaciduria, vitreous hemorrhage, and *pseudotumor cerebri* causing a bulging fontanel (11, 123). Many infants present coagulopathy as a result from liver disease, with prolonged bleeding after venous

or arterial sampling or excessive bruising (1, 44, 126). Cataract may be detected on slit lamp examination (11). Among atypical presentations, purpura fulminans, hemophagocytic syndrome, hypothyroidism, hydrocele, and inguinal hernia have also been reported (140, 147, 148).

The results of initial investigations indicate liver disease – unconjugated or combined hyperbilirubinemia, raised liver transaminases, raised plasma amino acids, particularly phenylalanine, tyrosine, and methionine – and also renal tubular disease – metabolic acidosis, galactosuria, glycosuria, albuminuria, and aminoaciduria. In addition, hematologic abnormalities are common, disordered clotting results from liver disease, and hemolytic anemia may occur (1, 44, 149, 150). The early clinical presentation of galactosemia is by no means specific, and a high index of suspicion is necessary, particularly in countries where galactosemia is not included in a newborn screening program (1, 44).

1.2.8. Diagnosis

There are several approaches for classic galactosemia diagnosis, namely i) analysis of reducing substances in the urine, ii) quantitative analysis of galactose and galactose metabolite(s); iii) GALT activity analysis, the gold standard of diagnosis; and iv) *GALT* mutational analysis.

1.2.8.1. Reducing substances in the urine

Determination of the reducing substance in urine can be used as a first step screening test for classic galactosemia (11). The presence of reducing substances or galactose in the urine is neither sensitive nor specific and should not be used to either confirm or refute a diagnosis (1). Indeed, a positive result may be due to fructosuria, lactosuria resulting from intestinal lactase deficiency, or to other conditions that impair blood galactose clearance, such as severe liver disease (126). Moreover, glucose is also a reducing sugar, therefore its presence in the urine should be discriminated by a glucose dipstick test (11). On the other hand, it should not be used to reject the diagnosis, since galactosuria occurs only intermittently during periods of substantial food intake and completely resolves within 3–4 days of intravenous feeding (126). Indeed, no galactose will be present in the urine if the child is on intravenous fluids, as is often the case during a neonatal crisis (11).

1.2.8.2. Galactose and galactose metabolite(s)

A more specific approach is measurement of specific metabolites, such as galactose, Gal-1-P and/or galactitol, either in the blood and/or in urine (11).

Nursing infants accumulate galactose in blood, as well as in other cells and tissues, and present high galactose urinary concentrations. Upon implementation of a galactose-restricted diet, however, these levels fall quickly (1, 8, 11). Conversely, Gal-1-P and galactitol will always be detected, in spite of galactose elimination from the diet and in spite of RBC transfusions, and are, therefore, the most common metabolites measured in case of suspicion of classic galactosemia (1, 11).

Indeed, galactosemic patients present invariably high levels of Gal-1-P in RBC that decrease slowly over a period of weeks to months following implementation of a galactose-restricted diet (1). However, even though measurement of accumulated Gal-1-P in RBC represents an unquestionable diagnostic asset, benign variants can also present increased levels of this metabolite, particularly if the infant is consuming galactose, and caution should be taken before making a diagnosis of classic galactosemia (123).

Measurement of hemolysate Gal-1-P can be accomplished by a variety of methods: by indirect enzymatic assays, by isotope-dilution gas chromatography-mass spectrometry (GC/MS), or by tandem MS (MS/MS) (1, 129, 151, 152). Another approach is the Paigen test, which quantifies total galactose – galactose and Gal-1-P – by a microbiological assay. This test is, however, not suitable for automation and is sensitive to antibiotic treatment of newborns or their mothers (152, 153).

Nursing infants also excrete enormous amounts of galactitol in their urine and present high plasmatic levels of this metabolite, which are kept higher than in controls, even upon diet implementation. Galactitol can be measured in the urine by nuclear magnetic resonance (NMR) or by GC/MS, whereas in RBC more sensitive methods are required, and galactitol levels of galactosemic patients on diet can only be measured by GC/MS (41, 133, 139, 154, 155). However, despite its important informative value, many laboratories do not display a galactitol measurement method (11).

1.2.8.3. GALT activity

The gold standard for diagnosis is the measurement of GALT activity in RBC, which can be done by the semi-quantitative Beutler fluorescent spot test or by an actual quantitative assay (1, 11, 31, 123, 141, 156, 157).

The Beutler test uses the successive reactions from GALT, phosphoglucomutase (EC 5.4.2.2) and glucose-6-phosphate dehydrogenase (EC 1.1.1.49), all naturally present enzymes in

RBC, to visually evaluate the fluorescence produced by NADPH under UV-light. Indeed, if GALT is active, Glc-1-P will be produced and will be converted into Glc-6-P by phosphoglucomutase, which is then oxidized by glucose-6-phosphate dehydrogenase with the concomitant reduction of $NADP^+$ to NADPH, thereby producing fluorescence (1, 141, 158). This test does not demand a heavy workload and allows for immediate identification of patients with classic galactosemia, even in children that are not under a milk-based formula (141, 158).

Conversely, the Beutler test also presents notable disadvantages, particularly regarding false-positives results, which constitute a major concern for this screening test. Indeed, this assay yields a high rate of false-positives, as a result of several interfering factors, namely: samples exposure to high temperatures and humidity, which can deteriorate the GALT enzyme; collection of blood into EDTA tubes; benign deficiencies of any of the assay coupling enzymes; fluorescence quenching by hemoglobin; and finally from the fact that the result depends on visual evaluation (1, 141, 153, 158, 159). Additionally, this test is not suitable in children that have been subjected to blood transfusion within the last 3 to 4 months, since a false-negative might result (1, 149). Quantitative assays of both parents can be informative in these circumstances since they can determine potential carrier status (157).

Quantitative measurement of GALT activity, though more labor intensive, overcomes many of the weaknesses of the Beutler test, presenting the additional advantage of being able to distinguish variants with residual activity (157).

GALT activity quantification can be done indirectly, by an enzymatic assay, or directly by quantifying radioactively labeled Gal-1-P conversion to UDP-Gal or, alternatively, by quantifying unlabeled substrates and products by high-performance liquid chromatography (HPLC) (1, 11, 160, 161). Classic galactosemic patients usually present undetectable GALT activity or detected at less than 5% of controls levels (80, 162).

Many laboratories present a two-tier approach, in which a first analysis is performed and, in case of a positive result, the diagnosis suspicion is further confirmed by a second assay. The Beutler test is often used as a first approach and, in case of a positive result, a second test – usually involving metabolite determination – is done, thereby reducing the false-positive rate (1, 141, 152).

1.2.8.4. Mutational analysis

DNA analysis of the most common mutations is rapidly available in many laboratories, and has been used in some newborn screening programs to refine the screening process. However, as with other mutation-selective DNA-based assays, because of the large number of

distinct mutations, a positive result has predictive value, but a negative result does not exclude disease resulting from other mutations (123).

1.2.9. Newborn screening

The success of neonatal screening for several inherited metabolic disorders (IMD), particularly with phenylketonuria (PKU, OMIM #261600), in addition to the advent of rapid screening techniques, has resulted in an increased impetus to expand neonatal screening programs (143). Methods for mass screening of newborns for galactosemia have been available since 1964 and, ever since, newborn screening has been implemented in several countries, allowing for presymptomatic diagnosis and intervention (1, 163).

1.2.9.1. Galactosemia newborn screening – the eternal controversy

The efficacy of universal neonatal screening for galactosemia has been debated for years, but no consensus has been reached (143).

For many years, the principles originally articulated by Wilson and Jungner for screening, and later expanded by Pollitt *et al.*, have been accepted as the criteria for neonatal screening (164). In general, decisions to include a genetic disorder on the panel are based on frequency of the disorder, the feasibility and accuracy of the screening, the ability of early intervention to improve outcomes, the availability of treatment, and the cost of screening (165). Therefore, galactosemia has been excluded from some newborn screening programs based on three major arguments:

i) the disease can be diagnosed clinically,
ii) it presents a historically high rate of false positives,
iii) long-term complications still develop in spite of early treatment (141).

In 1997, two studies regarding newborn screening were published, and both concluded universal neonatal screening for galactosemia was not sustained (166, 167). Even though Pollitt *et al.* recognized that the availability of effective treatment was not an absolute prerequisite for a disease to be included in newborn screening programs, galactosemia specific screening was still not recommended. Nevertheless, the reviewers strongly advised all samples with increased phenylalanine to be screened for galactosemia as a secondary test (167).

A later study in favor of the early diagnosis by neonatal screening compared 139 children with metabolic diseases identified by neonatal screening (17 galactosemic) to 124 identified on the basis of clinical symptoms (9 galactosemic). Despite the similar rate of hospitalization (64%

versus 77%, respectively), 47% of the clinically identified children had mental retardation compared to only 14% of those identified by newborn screening, and parental stress tended to be greater in families of children identified clinically (57, 168). Indeed, early treatment implementation does not necessarily prevent long-term complications, but early diagnosis reduces the immediate impact of high doses of galactose, by cutting short the duration of exposure to the harmful metabolites and can, therefore, limit early morbidity and mortality from the disease (169).

1.2.9.2. Newborn screening worldwide

In contrast to newborn screening for PKU, which has become almost universal, galactosemia screening is included only in a minority of European newborn screening programs, namely Austria, Germany, Hungary, Ireland, Sweden, and Switzerland; while other four countries, specifically Turkey, Spain, Italy, and Belgium, present pilot programs (140, 141, 170).

Ireland presents a high prevalence of classic galactosemia, especially in the Traveller population, and newborn screening for classic galactosemia has been performed since 1972 as part of the *National Newborn Screening Programme* (77). Irish newborns undergo newborn screening within 72 to 120 hours after birth, while all high-risk Traveller neonates perform it at day 1 or 2 (77).

On the other hand, Sweden presents a relatively low frequency (1/100,000) and yet newborn screening for galactosemia has never been questioned. In fact, Sweden has a well-working strategy for the screening procedure, with screening methods well established, inexpensive, that do not demand a heavy workload, and, more importantly, with a low false-positive rate (141).

Other countries, such as France and The Netherlands do not offer neonatal screening and some, like Norway and Denmark, have stopped their programs (143, 170-172). In the United Kingdom, there is no screening program for classic galactosemia in England, Wales and Northern Ireland, but there is one in Scotland (143, 170).

In the United States of America, screening for galactosemia is currently offered in all states, and in Canada, in 1991, only 2 out of 10 provinces offered universal screening (1, 143).

1.2.9.3. Classic galactosemia newborn screening in Portugal

In 1979, the program for neonatal screening, so-called *Portuguese Neonatal Screening Program* (PNSP), was established in Portugal by the Ministry of Health with PKU screening, and further expanded in 1981 to congenital hypothyroidism (173, 174).

The later development of electrospray ionization tandem mass spectrometry (ESI-MS/MS) allowed, in 2004, expanding the screening to a total of 25 disorders, in which only potentially treatable diseases were authorized by the Ethics Committee to be included in the program (174). Despite galactosemia not being included in the PNSP, all samples with increased phenylalanine and/or tyrosine are screened for classic galactosemia, by a semi-quantitative GALT activity evaluation with the Beutler fluorescent spot test, followed by total galactose quantification by a colorimetric assay (Hugo Rocha, personal communication, November 8, 2013).

1.2.10. Therapy and long-term outcome

1.2.10.1. Initial management

In the first weeks of life, the most important part of managing patients with classic galactosemia is removing all galactose from the diet, immediately after starting the diagnostic investigations and without awaiting results, in order to prevent further life-threatening complications (175). Nevertheless, even after the removal of all galactose from the diet, the infant may be seriously ill and require additional therapy. Supportive care with intravenous fluids, phototherapy, antibiotics, and treatment of coagulopathy should be initiated as needed (1, 123).

Clotting abnormalities, derived from liver failure, may be managed by vitamin K and fresh-frozen plasma. Unconjugated hyperbilirubinemia is usually not severe enough to warrant phototherapy or exchange transfusion, but infants may be at increased risk of kernicterus, if albumin levels are particularly low secondary to liver dysfunction (1). In case of sepsis, gram-negative infection should be assumed, and an appropriate antibiotic should be administered intravenously (1).

Most children at diagnosis are able to tolerate enteral feedings, although these may have to be given initially by nasogastric tube because of poor sucking ability (1). If the infant is too sick to tolerate enteral feeding for more than 1 or 2 days, then parenteral feeding should be considered but prescribed cautiously if there is significant liver disease or thrombocytopenia (1).

After the initial management of the acute presentation, the infant is put on a galactose-free formula, and is checked regularly for pediatrician and nutrition support (1, 149, 176).

1.2.10.2. Therapy

Lifelong dietary restriction of galactose has been the basis of therapy of classic galactosemia since Mason and Turner, in 1935, described how removing galactose from the diet eliminated the acute toxicity syndrome (13).

As soon as there is a suspicion for galactosemia, breastfeeding or milk-based formula, which contain a significant amount of galactose, must be ceased immediately and replaced by a galactose-free formula (1, 146, 149, 176).

Commercial infant formulas appropriate for use in the management of classic galactosemia include elemental formulas and soy formulas (1, 126, 149). Infant soy formulas are most appropriate, unless there is significant liver disease, in which casein-based protein hydrolyzate formulas, that contain medium-chain triacylglycerol oil, should be used (1, 126, 149). In the absence of significant liver disease, however, casein hydrolyzates should be avoided, since they do contain small amounts of galactose (60 to 90 mg/100 ml). In that case, soy formulas are the formula of choice (1, 126, 149).

Most symptomatic infants with classic galactosemia show rapid clinical improvement following dietary restriction of galactose. Within 24 hours of removing galactose from the diet, the symptoms improve and, within 2 to 3 days, the infants seem to be almost miraculously cured (145). Features such as hepatomegaly, abnormal liver function, and renal tubular disease usually resolve fully within a week or two. Acceptable levels of Gal-1-P in RBC are usually achieved within 2-3 months of starting treatment, but may take up to 1 year (176).

1.2.10.3. Long-term complications

The dramatic response to dietary therapy for galactosemia led to the belief that this disorder was relatively benign and easy to treat (145). While the success of early treatment on the acute neonatal symptoms is unquestioned, the long-term outlook is not as good as expected (145). Indeed, even with early treatment commencement and optimal dietary compliance, patients often develop cognitive and neurologic complications. The complications may appear individually, as a complex, or not at all (56, 123).

Whether the long-term complications observed result from subtle developmental abnormalities initiated *in utero*, chronic exposure to endogenously produced galactose or other factors remains unknown.

1.2.10.4. Neurological and psychoneurological outcome

One of the most well-established long-term complications is cognitive disturbance, which affects approximately 50% of all treated patients (177). Indeed, after Komrower and Lee statement, in 1970, that "the intelligence of the early diagnosed and well treated is below average", several subsequent studies have supported cognitive impairment (138, 145, 178-181). Across most studies, intelligence quotient (IQ) has been found to be in the typical (85–115) to borderline-low (60–84) range, with considerable inter-individual variability (178, 181). There are, however, conflicting results regarding the IQ scores change with increasing age (178). In an international survey, Waggoner *et al.* reported a progressive decline in IQ scores over time, with patients who had had IQ scores evaluations at both 3-5 and 6-9 years presenting a mean decrease of 6.2 IQ points between the two ages (n = 42), and with patients who had had IQ scores evaluations scores at both 6-9 and 10-16 years presenting a mean decrease of 4.4 points (n = 46) (145). However, a later longitudinal study examining this hypothesis found no evidence of progressive decline in overall cognitive functioning with age (182). Regardless of the possibility of progressive neurodegeneration, studies concur that IQ scores do not appear to correlate with time of diagnosis, initiation of treatment, dietary compliance, exposure to Gal-1-P levels, or genotype (60, 123, 183).

Motor disturbances, such as coordination, gait, balance, fine motor tremors, and severe ataxia, are also quite frequent (11, 138, 145). Other neurologic findings, such as epilepsy and microcephaly, have also been reported, though more rare (11, 138).

Neuroradiologic studies of individuals with galactosemia have demonstrated white-matter abnormalities, although magnetic resonance imaging findings do not appear to correlate with severity of illness or IQ (178).

Moreover, several studies have estimated that 38–65% of children with galactosemia have neurodevelopmental speech disorders, of which more than 90% have been reported to have co-occurring language disorders, which include vocabulary and word retrieval deficits (145, 184). In fact, in Waggoner's retrospective study, problems with speech were reported in 56% of cases (> 3 years old), of which 92% were described as having delayed vocabulary and 90% had articulation problems (145).

Speech disorders in children with galactosemia have a neurologic origin and, in a study by Potter *et al.*, were divided into three subtypes of motor speech disorders: i) childhood apraxia of speech – CAS, a deficit in motor planning or programming; ii) dysarthria, a deficit in neuromuscular control; and iii) motor speech disorder-not otherwise specified – MSD-NOS, a cover term for speech, prosody, and voice behaviors that are consistent with a motor speech disorder, but not specific for CAS or dysarthria (177). Effectively, galactosemic children with speech disorders have a high co-occurrence of motor disorders, namely in strength and

coordination, which suggest a common underlying etiology (145, 177). Additionally, Potter *et al.* have also reported that galactosemic children with CAS or dysarthria presented poorer balance and manual dexterity than most of the children with MSD-NOS, and exhibited more speech errors and were less intelligible than the children with MSD-NOS. Accordingly, these authors concluded that speech, coordination, and strength disorders are most likely associated with diffuse cerebellar damage, since the cerebellum is involved with motor movements, refining speech and balance control (177). Indeed, histopathological brain examinations and multiple neuroimaging studies revealed cerebral and cerebellar atrophy, with demyelination of the white matter in galactosemic patients (77, 123, 138, 185, 186). The mechanism(s) for white matter lesions is unclear. It is possible that GALT deficiency leads to defective synthesis of glycoproteins and glycolipids critical for myelin synthesis, possibly as a result of defective transfer of galactose from UDP-Gal (11, 123, 185, 187-189).

Visual and perceptual impairments are also common in galactosemia and seem to be independent of IQ (138, 178, 186). Memory difficulties have also been reported, although these impairments do correlate with IQ (178, 186). Handwriting, reading, and mathematics difficulties are also common, and children with galactosemia are often academically retained (138, 178). In fact, school achievements are below the grades achieved by healthy siblings and a comprehensive multidisciplinary study on the adult galactosemic phenotype reported an average schooling of 1 to 2 years of college, with an occupational level of skilled manual laborer (138, 142, 179). Furthermore, as had been described by Bhat *et al.*, in a study concerning the social outcome of treated patients with IMD, many patients have no educational qualifications and are unemployed (190). Actually, the cognitive delay and the speech difficulties characteristic of galactosemia most likely interfere with the development of social skills and educational attainment (181, 191-193). Indeed, galactosemic patients are often described as shy and reserved, which may be related to word retrieval difficulties, and generally do not present with behavioral dysregulation – there are only a few reports of hyperactivity in galactosemic patients (178, 194).

Additionally, many patients are single and live with their parents, as only a minority is able to build up strong partnerships outside the core family (77, 142, 195). In fact, in a study by Hoffmann *et al.*, approximately 40% of the patients recognized that classic galactosemia did impair their company with other people (195).

In 2004, a study by Bosch *et al.*, aiming to evaluate patients' health-related quality of life (HRQoL) and specific concerns related to this disorder in patients and parents, reported galactosemia was seen as a burden by many patients (39%) (179). Actually, many felt different because of having galactosemia (34%), and believed their disease was not well understood by others (22%) (179). Furthermore, many parents believed the disorder influenced their contact with the child (73%) and that it was a burden to take care of a child with galactosemia (60%) – despite reporting treating and raising their child with galactosemia the same way as their healthy

children (69% and 77%, respectively) (179). This was the first study ever reporting the HRQoL of patients with classic galactosemia by using well-developed and validated instruments in different age groups, and the results were quite astonishing: galactosemia negatively influences the HRQoL, most strikingly on the cognitive domain, but also on the domain of communication and social function (179).

The underlying cause(s) behind the neurological and psychoneurological outcome in classic galactosemia is still an enigma. It most likely results from assaults from multiple fronts (123). In effect, all attempts to correlate the occurrence of neurological and psychoneurological outcome of classic galactosemia with differences in the course of the disease, such as age at diagnosis, severity of the acute clinical manifestations, age of galactose-restricted diet implementation, diet compliance, biochemical parameters or genotype, have failed (178, 183).

1.2.10.5. Growth and bone density

Several studies have reported growth delay and decreased bone mineral density in patients with classic galactosemia under dietary treatment.

The first report of decreased height in childhood and early adolescence in treated galactosemic patients was done by Waggoner *et al.*, who evaluated 350 patients aged from 2 weeks to 37 years, and found that, despite being severely delayed during childhood and early adolescence in many cases, growth often continued through the late teens so that final heights were normal (145).

In a more recent study, Panis *et al.* evaluated 40 children and adolescents (13 males and 27 females, aged 3–17 years) and reported decreased height (corrected for target height) in both genders and decreased height velocity in girls (196). Furthermore, predicted final height was lower than target height in most patients; however, as the authors emphasized, target height could be reached for the children who grow beyond the age of 18 (196).

In addition, treated galactosemic children, adolescents and adults are also at risk for a diminished bone mineral density (BMD) (197-200). Panis *et al.* reported a decreased BMD in the same group of galactosemic patients, in whom they had found no evidence of nutritional deficiencies. The authors did find, however, decreased levels of bone formation markers and bone resorption markers, namely decreased carboxylated osteocalcin (199). In another study with the same patients, Panis *et al.* evaluated, for the first time, body composition in galactosemic patients, and found decreased insulin-like growth factor-I (IGF-I), which, due to important role of hormones in height and body composition, the authors hypothesized could play a role in the decreased growth (201).

In another study, Panis *et al.* (2006) studied the effect of calcium, vitamins K_1 and D_3 supplementation on bone, since they had found a decreased carboxylated osteocalcin concentration, a marker for the vitamin K nutritional status (199, 202). Since vitamin K works best if given in combination with calcium and vitamin D, these authors developed a 2-year randomized clinical trial to investigate whether vitamin K could play a role in the pathophysiology of this complication. In fact, they found a statistically significant increase in osteocalcin carboxylation in both prepubertal and pubertal children receiving supplementation (202).

In another study, Gajewska *et al.*, aiming to assess bone formation and resorption processes in children and adolescents with galactosemia, reported similar mean values of bone turnover markers in both groups of patients, contrarily to healthy adolescents, who, in comparison to healthy children, presented significantly lower physiological values of the investigated markers (203). The higher concentration of the bone turnover markers of the galactosemic adolescents comparatively to age-matched controls suggested a higher rate of bone turnover with predominance of the bone formation process, which the authors hypothesized could be a reflection of an adaptive mechanism ensuring proper bone mass after disturbances of bone metabolism in early childhood (203).

A recent study evaluating 32 adult patients (17 males and 15 females) showed that bone density in galactosemic adults is low, thereby conferring a potential for increased fracture risk (200). These authors suggested early interventions, starting on childhood, in order to promote optimal bone accrual and thereby improve lifetime bone health for this high-risk group galactosemics represent (200).

The pathophysiological mechanism of the bone-related complications in classic galactosemia is not completely understood. Several mechanisms have been proposed, including nutritional deficiencies, alterations in the endocrine axis, and intrinsic factors related to bone metabolism (200). Because of their galactose-restricted diet, patients are at risk for nutritional deficiencies, particularly for calcium, since milk products are rich in casein-bound calcium – important for a good bioavailability of calcium – and are, therefore, considered the best source of dietary calcium (202). However, several studies have excluded major nutritional deficiencies, and concluded that other disease intrinsic factors must be involved (197, 200-202). In fact, Kaufman *et al.* found calcium intake to be correlated with BMD in adult galactosemia patients but not in prepubertal patients, implying that other than dietary factors must be involved (197, 202).

Nevertheless, inadequate calcium and vitamin D intake by galactosemic patients has been reported, as well as a decreased carboxylated osteocalcin concentration in the serum and, when considering strategies to improve bone mineralization and diminish morbidity, supplements like calcium, vitamins K_1 and D_3 may be taken into consideration (197, 200, 202). In fact, 90% of the 32 adult patients studied by Batey *et al.* reported they were not being followed by a nutritionist,

which suggests the frequency of visits with a nutritionist may decrease as the patients enter adulthood, highlighting the importance of providing appropriate nutritional counseling for maintenance of an ideal body weight and potentially as a strategy to optimize bone accrual during adolescence and to maintain bone mass during the adult years (200).

1.2.10.6. Gonadal function

The most common long-term complication reported for girls and women with classic galactosemia is primary ovarian insufficiency (POI), with an incidence between 80% to more than 90% (1, 125, 145, 204, 205). According to female patients and their parents, this complication represents the greatest psychological burden of classic galactosemia (179, 206, 207).

POI is a spectrum disorder of ovarian insufficiency, with clinical manifestations that may range from absent or delayed pubertal development to primary amenorrhea, secondary amenorrhea or oligomenorrhea and decreased fertility (1, 204, 206, 208). POI can be detected by measuring follicle-stimulating hormone (FSH), luteinizing hormone (LH), estradiol and anti-Mullerian hormone (AMH) in blood, but given the inactivity of the hypothalamic-pituitary-ovarian (HPO) axis in childhood, these levels may not be sensitive enough to identify ovarian dysfunction in a presymptomatic state (208). Elevated FSH level, an indirect measure of ovarian function, has been found elevated very early in life (4 months to 4 years) and also between early childhood and onset of puberty (5–12 years), while serum AMH level, an indirect marker for follicular function and ovarian reserve, is abnormally low in galactosemic girls relative to age-matched controls, even in girls younger than 2 years (125, 204, 205, 209).

All female patients should be followed by an endocrinologist during late childhood and puberty (1, 149). Hormone replacement therapy (HRT) with estrogen and progesterone is currently the recommended therapeutic approach; the age at which HRT should begin is, however, debatable (1, 149, 208).

The etiology underlying POI in classic galactosemia is not yet fully understood, and several mechanisms have been postulated, namely direct toxicity of metabolites – Gal-1-P and galactitol – and UDP-Gal deficiency (205, 206, 210). Gal-1-P may inhibit key enzymes that normally metabolize Glc-1-P, thereby impairing cellular metabolism and/or lead to increased cell stress, whereas galactitol causes osmotic stress and swelling of the ovarian cells, which would eventually lead to cell dysfunction (204). On the other hand, UDP-Gal deficiency can lead to apoptosis of oocytes and ovarian stromal cells or cause aberrant glycosylation of glycoproteins and galactolipids involved in ovarian function, especially since FSH contains galactose-rich oligosaccharide side chains (1, 204, 205, 211). FSH malfunction due to hypoglycosylation,

however, has been made less probable by the fact that FSH bioactivity was found normal in galactosemic girls (205, 209). However, a functionally altered FSH receptor due to hypoglycosylation remains a possibility that has not yet been studied (205). It is also conceivable that not one, but the different mechanisms acting in unison are responsible for this clinical picture (204, 205)

Moreover, determining the point in development at which POI begins has been difficult. It is still uncertain whether this ovarian insufficient is initiated in early development, perhaps even prenatally, or whether it represents a degenerative process compounded by ongoing galactose exposure (204, 205). Undoubtedly, fertility issues in female patients with classic galactosemia are complex and multifactorial (208). However, spontaneous pregnancies may occur, and a predicting factor for the possibility to conceive is spontaneous menarche (208).

Regarding the male gonadal function, there is a manifest paucity of data vis-à-vis to the female counterpart. Indeed, only few studies have examined the male reproductive system and, until recently, male fertility was believed not to be impaired by galactosemia (205, 212).

In 2006, Rubio-Gozalbo *et al.* described a high frequency of cryptorchidism in galactosemic individuals, presenting a prevalence of 25% (3 of the 12 males) in opposition to ≤1% in the healthy age-matched population (213). This study was, however, somewhat limited due to small sample size. A more recent investigation on the male reproductive system of 26 galactosemic patients demonstrated an increased prevalence of cryptorchidism (11.6%), with low semen volume, and lower testosterone, inhibin B and sperm concentration than in control subjects – although within the normal range on average – which might indicate mild defects in Sertoli and Leydig cell function (212).

With the exception of cryptorchidism, the clinical significance of these abnormalities should have little impact on fertility (212, 213). Again, there is a remarkable scarcity of fertility studies concerning galactosemic male patients, with only 3 fathered pregnancies in the literature (142, 212). The lack of documented paternity in men with galactosemia may be related to a publication bias and difficulty documenting paternity or to the known social interaction problems and delayed psychosocial development documented in these men (212).

In fact, in the study on the adult galactosemic phenotype, out of the 12 men involved, only 2 were fathers (142). This same study also reported depression and anxiety in 39% and 67%, respectively, of the patients, which may be a reflection of the impact of the long-term complications of classic galactosemia in the adult patients (142). Indeed, despite having a normal life expectancy, galactosemic patients do develop long-term complications that cannot yet be prevented, and greater support to psychosocial issues – depression, anxiety, and social relationships – as well as medical issues should be provided, in order to relieve the impact of this disorder in adulthood (142, 179, 190).

Classic galactosemia unexpected long-term outcome constitute one of the main enigmas of this metabolic disorder and set the basis for the development of new therapeutic approaches (further explained below).

1.2.10.7. *Classic galactosemia: dietary dilemmas* [175]

In the 1950's, different strategies to feed infants a diet without galactose were discussed in the literature. Nowadays, initiating the diet is much easier, as infant formulas with a very limited amount of galactose are widely available. Still, there are dilemmas in the treatment of galactosemia, and the dietary treatment varies widely around the world (175).

Both semi-elemental formulas and soy-based formulas contain very small amounts of galactose, and recent studies suggest that completely eliminating galactose from the diet by prescribing an elemental formula allows a significantly faster decrease of the high RBC Gal-1-P levels (175). Zlatunich and Packman have described the case of a classic galactosemic baby, diagnosed by newborn screening at the age of 6 days, and put on a soy-based formula containing 14 mg of bound galactose per liter (1.4 mg/100 mL). Galactose daily intake was estimated to be 13 mg. By the age of 5 months, erythrocyte Gal-1-P levels were still above 4 mg/dL, so an elemental formula was initiated. This formula resulted in a rapid decrease in Gal-1-P levels, which stayed at the low end of the treatment range (214). Two other patients, described by Ficicioglu *et al.*, with a similar clinical history, were also put on an elemental formula at age of 5 months, and they too presented statistically significant decreased levels of Gal-1-P in RBC (215).

Nevertheless, the possibility that patients' metabolite levels would eventually decrease on soy formula should not be ruled out (214). Soy protein-based formulas are safe and cost-effective, and are the first-line therapy in many countries, namely in The Netherlands and Australia (170, 176, 216). It remains to be elucidated whether a potential faster decrease by the use of elemental formulas**,** would positively affect long-term outcome. In fact, prospective studies need to be performed to clarify this issue, before routinely putting these infants on this expensive and unpalatable formula (175, 176).

Upon introduction of solids foods, some galactose is inevitably introduced into the diet (14, 176). Despite lactose continue being the primary source, there are other sources of galactose – either in the free or bound form – that are found naturally and widely distributed in foods (176). Though in much smaller quantities than in milk-containing foods, galactose is also present in cereals, offal meats, pulses, fruits and vegetables (11, 149, 176, 217).

A thorough study by Gross and Acosta quantified the free galactose content of 45 fruits and vegetables, which ranged from less than 0.1mg/100g in artichoke, mushroom, olive, and peanut to 35.4mg/100g in persimmon (217). In addition, galactose can also be found in its

complex form, namely galactans and galactolipids, which have the potential to be hydrolyzed in the gut by β-galactosidase or pancreatic lipase, respectively, and, therefore, to release galactose, rendering it available for absorption (217). The actual bioavailability of these complex galactose substances is, however, unknown (176, 218).

The significance of non-dairy foods to total intake of galactose in galactosemia outcome has long been questioned (176, 218-220). Several studies have shown that fruits and vegetables contribute with only small amounts of galactose compared to lactose rich foods (176). Berry *et al.* demonstrated that a lactose-free diet highly enriched in fruits and vegetables resulted in a mean galactose intake of 54 mg per day, and further demonstrated that the daily ingestion of 200 mg of fruit-derived galactose (50 mg of pure galactose in fruit juice 4 times per day) for 3 weeks had no impact on RBC Gal-1-P values and relatively little effect on urinary galactitol levels (220). In another study, Bosch *et al.* demonstrated that supplementation of galactose during a period of 6 weeks up to a maximum of 600 mg per day – corresponding to the amount of galactose in 7 kg of apples, 2.5 kg of tomatoes, or 12 kg of peas – did not result in any physical, ophthalmological, or biochemical abnormalities (219).

These data strongly suggest that the reduction of the daily dietary galactose intake by restricting fruit and vegetables is negligible and, in the short-term, appears to involve no significant clinical impact (219). Moreover, the realization that galactosemic patients endogenously synthesize galactose, and to an extent that far exceeds that from non-dairy foods, has minimized this concern (126, 149, 176, 219, 220). Therefore, some European countries – Denmark, Norway and United Kingdom – have liberalized the therapy, advising now a lactose-free diet (170, 219).

In Portugal, the diet shows slight variations according to the metabolic center. After management of the acute neonatal syndrome, the infant is kept on a soy formula. Upon introduction of solid foods, a galactose-restricted diet is implemented, with restriction of lactose-containing foods and of some fruits and vegetables of high galactose content, namely watermelon and beans. Some fully cured cheeses, such as Emmental and Gruyere, are allowed, as they have been described to contain undetectable amounts of lactose and galactose (221).

1.2.11. Current alternative therapeutic approaches in classic galactosemia

Inborn errors of metabolism comprise a large group of disorders, commonly caused by inactivity of specific enzymatic steps, leading to compromised flux of molecules in evolutionarily conserved metabolic pathways. Traditionally, IMD are categorized as those involved in i) synthesis or catabolism of complex molecules (e.g., congenital glycosylation disorders); ii) intermediate metabolism, (e.g., amino acidopathies, organic acidemias, urea cycle defects, and

disorders of carbohydrate metabolism including classic galactosemia); and iii) energy metabolism, (e.g., disorders of gluconeogenesis, lactic acidemias, fatty acid oxidation defects, and Krebs cycle and mitochondrial respiratory chain disorders) (222-224). The great majority of pathophysiological mechanisms of IMD are related to: i) accumulation of precursor(s); ii) alternative metabolites normally not encountered; or iii) absence (or deficiency) of metabolites past the enzymatic block. Any, or a combination of these possibilities, could be responsible for the phenotypes associated with an enzymatic blockage (222).

The pathophysiology of IMD varies greatly between disorders, and so will management and therapeutic approaches. The various therapeutic approaches of IMD can be categorized according to the level of management (Figure 1.7) (222).

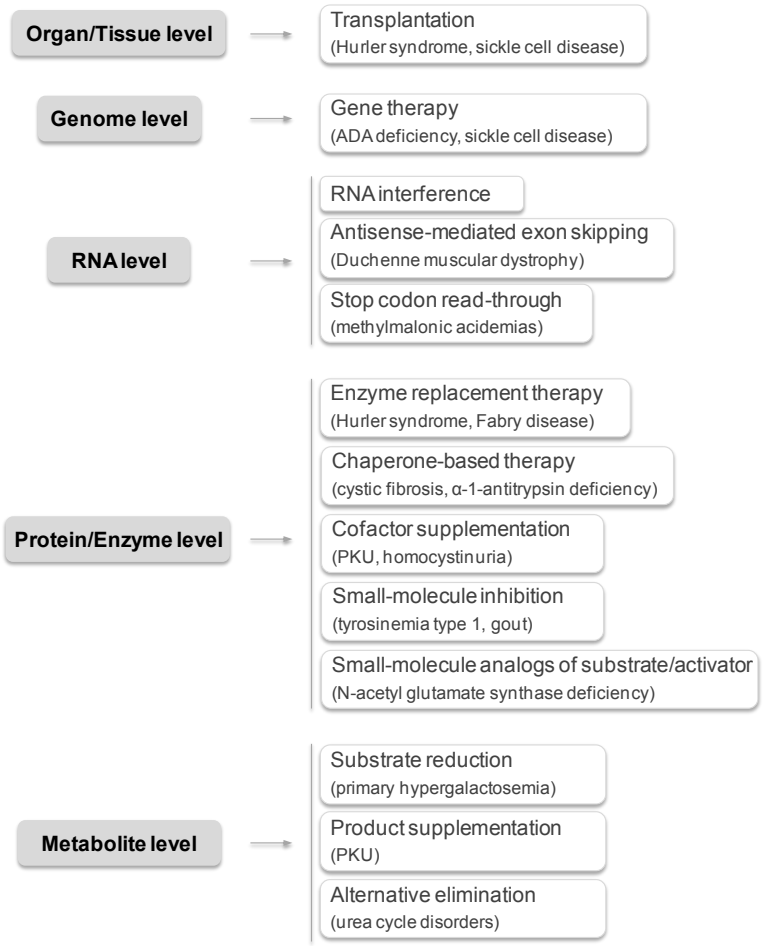


Figure 1.7. Overview of the current therapeutic approaches in IMD at several levels of intervention. (ADA – adenosine deaminase). Adapted from Tang *et al.*, 2011 (222).

In what regards classic galactosemia, its pathogenic mechanism(s), as well as its timing, are still inexplicable and puzzling (33, 222).

Several hypothesis have been put forward concerning timing of the cellular pathophysiology: i) the long-term complications of classic galactosemia are entirely due to

prenatal disease; ii) the complications are due to postnatal disease, owing to the exposure to relatively massive concentrations of lactose in the newborn period and/or the chronic exposure to exogenous and endogenously synthesized galactose; iii) the complications are both prenatal and postnatal in origin (13, 225).

Concerning the pathogenic mechanism(s), the hypothetical causative agents of cell and tissue dysfunction regard:

i) an abnormal *GALT* gene *per se*;
ii) an abnormal GALT protein;
iii) pleiotropic abnormalities, dependent on an elevation of cellular Gal-1-P;
iv) abnormal intracellular concentrations of galactitol;
v) a combination of the above (225).

Indeed, it is apparent that galactose and Gal-1-P accumulate in patients, with galactose being further metabolized through two alternative pathways to form galactitol and galactonate. Among all the metabolites formed, Gal-1-P and galactitol have received most attention; however, their potential toxicity targets and their exact pathogenic mechanism(s) are still largely unknown (222).

Despite its limited efficacy in preventing long-term complications, the current standard of care for classic galactosemia remains dietary restriction of galactose, a therapeutic approach at the substrate reduction level (1, 222).

An alternative therapeutic intervention that has been investigated is at the product supplementation level, with UDP-Gal. In light of the possible role of UDP-Gal deficiency in the pathogenesis of classic galactosemia put forward by Shin *et al.* in 1985, Ng *et al.* studied UDP-Gal levels in hepatocytes, erythrocytes and skin fibroblasts from galactosemic patients, thereby confirming its deficit in GALT-deficiency cells. Accordingly, the authors suggested uridine administration could be of therapeutic value by raising the intracellular concentrations of UDP-Gal (182, 226). Subsequently, Manis *et al.* began a 5-year study in which oral uridine was administered to two cohorts of galactosemic patients, whose results, however, did not support the hypothesis that oral uridine improved cognitive functioning in afflicted subjects (182). It is, however, unclear whether UDP-Gal would reach the cytosolic compartment in GALT-deficient cells; and thus there is still much controversy regarding this therapeutic intervention at the product supplementation level (222).

Another therapeutic strategy at the enzyme level involves inhibition of GALK, the first enzyme of the Leloir pathway. This approach was proposed more than 10 years ago by Bosch *et al.*, as a way of reducing the build-up of Gal-1-P, long known to play a key role in the pathophysiology of classic galactosemia, thereby converting this severe form of hypergalactosemia into the milder form of type II galactosemia (12, 227). Indeed, patients with galactokinase deficiency do not experience the long-term complications of classic galactosemia,

and a medically induced galactokinase deficiency could provide protection from the conversion of galactose to Gal-1-P. Therefore, Lai's research group has been searching for small-molecule inhibitors of galactokinase over the last years (25, 228). Another therapeutic option is the administration of an aldose reductase inhibitor to block galactitol synthesis, and, hopefully, perhaps both inhibitors could be given simultaneously (33, 229).

A more recent study regarding five GALT variants reported the pathogenic mechanism underlying impaired enzymatic activity is GALT misfolding and suggested this could be transversal to many GALT variants, and thus proposed chaperone-based therapy to classic galactosemia (110).

Despite decades of intensive research, the pathophysiology of classic galactosemia remains largely unknown, with limited and partly ineffective therapeutic options (222). Indeed, so many years have gone by since the first description of this disease and, aside from dietary lactose restriction in infancy, few therapeutic advances have been achieved (229).

Classic galactosemia does indeed constitute an enigma that, for now, continues to be unsolved.

1.3. Mutation-specific therapeutic approaches

Molecular studies in a number of genetic disorders have led to the realization that only a minor part of the mutations directly disrupt functional sites of the proteins. Rather, most sequence variations disturb either the folding of the newly synthesized variant proteins or the synthesis of the protein by affecting pre-mRNA processing (230-233).

Missense and small in-frame deletions and insertions may lead to aberrant proteins, depending on the residual function of the specific protein, and on the cellular and environmental conditions; whereas splicing variations and out-of-frame deletions and insertions result in total loss-of-function, often due to the introduction of a premature termination codon into the mRNA, typically resulting in its degradation by the mRNA surveillance mechanism (234, 235).

1.3.1. Pathogenic molecular mechanisms of genetic disorders – folding spoilers

The successful execution of cellular processes in the human body depends on the coordinated interactions of roughly 20,000 to 25,000 different proteins, which are responsible for most aspects of biological function (235, 236).

Approximately 20% to 30% of all proteins in the cell seem to be intrinsically devoid of any ordered three-dimensional structure and adopt folded conformations only after interaction with binding partners (237). However, the great majority of the proteins must be folded into their

native state, an ensemble of a few closely related three-dimensional structures, in order to achieve their function (237). How cells ensure the conformational integrity of their proteome constitutes one of the most important and relevant issues in biology, and provides a rational basis for tackling many genetic disorders in which protein (mis)folding plays a major role (237-239).

1.3.1.1. Protein folding – the challenge

Proper folding into the correct secondary structure and conformation is indispensable for the biological function of proteins and is achieved during, or immediately following protein synthesis (240). When the nascent polypeptide emerges from the ribosome, it must fold and assemble properly, in order for the protein to achieve its three-dimensional conformation (241). The native structure of a protein is mainly determined by its amino acid side chains, and is predominantly driven by the need to bury the hydrophobic residues within the protein core, hidden from the surrounding aqueous milieu (237, 241-243). Indeed, prior to folding, polypeptide hydrophobic domains are exposed to a complex and very crowded environment – estimated to reach protein concentrations as high as 300 mg/ml – that promotes inter- and intramolecular interactions, which could reduce the efficiency of the cellular protein folding (231, 235, 241). To prevent the undesired interactions and to provide a more favorable environment, folding is assisted and monitored by a class of proteins known as molecular chaperones. Molecular chaperones do not increase the rate of folding, but rather increase the probability of desirable interactions and prevent misguided interactions, thereby allowing proteins to fold efficiently and on a biologically relevant timescale (237, 241).

To picture protein folding, a topological landscape representing different energy levels has been suggested. In this model the process of folding is described as a constant strive towards minimal free energy with the native structure of the protein being the functional conformation with the lowest energy level, the global minimum of the landscape (Figure 1.8). Chaperones iteratively bind and release their substrates, each time raising the free energy and enabling evasion from wrong folding pathways. Eliminating protein by degradation can, in principle, take place from any location in the landscape, but occurs predominantly in areas with trapped misfolding or on pathways toward aggregated structures. Because the free energy of a protein is a function of its conformation defined by the interactions between the amino acid residues, even small alterations in the amino acid chain may change the surface of the landscape, leading to the possible formation of new local free-energy minima or, in the extreme case, formation of a new global free-energy minimum resulting in a different stable structure of the protein, which may be prone to aggregation. This implies that the native state of a protein is not always the global energy minimum. To refine the concept of energy folding landscapes to include the aggregation tendency, this model can be pictured with two deep valleys (Figure 1.8) (244).

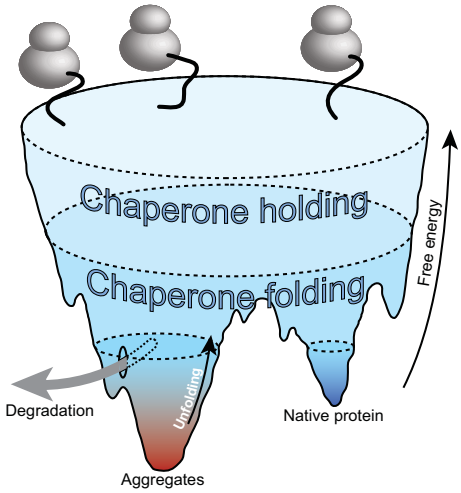


Figure 1.8. Free-energy folding landscape for chaperone-mediated protein folding. The landscape is drawn as a rugged funnel shape in a three-dimensional system with the free energy on the y axis and the conformational space or entropy as a two-dimensional projection on the x and z axes. Chaperones iteratively bind and release their substrates, each time raising the free energy and enabling escapes from wrong folding pathways, indicated by the "Unfolding" arrow. Eliminating protein by degradation can, in principle, take place from any location in the landscape, but occurs predominantly in areas with trapped misfolding or on pathways toward aggregated structures, as indicated in the aggregation funnel. Reproduced from Gregersen *et at.*, 2006 (244).

Given that proteins have an intrinsic tendency to misfold, cells must be able to refold misfolded proteins or to remove them by degradation. This sorting process is executed by the so-called protein quality control (PQC) system, whose main components are molecular chaperones – which supervise folding and protect folding intermediates from nonproductive interactions that could result in aggregation – and proteases – which assist in the elimination of aberrant folding intermediates and unstable proteins that would overload the chaperone systems and potentially damage the cell (Figure 1.9) (235).

Molecular chaperones are present in all types of cells and cellular compartments. Some chaperones interact with nascent chains as they emerge from the ribosome, whereas others are involved in guiding later stages of the folding process; and molecular chaperones often work in tandem to ensure that the various folding stages are all completed efficiently (245). Several classes of molecular chaperones can be distinguished, namely: the Hsp70 family, a paradigmatic group of molecular chaperones that control binding and release of folding intermediates; the chaperonins, which form a characteristic barrel structure whose interior cavity serves as a 'folding cage' where proteins can fold undisturbed; small heat-shock proteins (sHsps), which bind unfolded proteins, thus acting as 'holding chaperones' for folding intermediates; the lectin chaperones calnexin and calreticulin, chaperones of the endoplasmic reticulum (ER) that work in collaboration with glycosidases and the UDP-glucose:glycoprotein glucosyltransferase (UGGT), recognizing incompletely folded proteins with N-glucosyl chains carrying one glucose residue and not allowing them to move further ahead to the Golgi compartment; and the AAA-domain-containing chaperones, which represent a large superfamily involved in the disassembly of

proteins and unfolding of misfolded proteins to present them to the degradative system (231).

On the other hand, proteases selectively eliminate misfolded proteins, and thus require sorting or labeling of the polypeptide and unfolding before their processive peptidase activity executes peptide bond cleavage. The best-studied protease is the cytosolic proteasome, which is composed by a regulatory particle and a core proteolytic particle (231). Proteasome substrates are usually tagged by a system that recognizes proteins with a low folding capacity or structurally instable proteins and adds ubiquitin molecules to lysine residues in the polypeptide chain, after which the proteasome regulatory particle recognizes and binds the ubiquitinylated polypeptides, removes the ubiquitin tags, unfolds the chains and transports them into the proteolytic chamber of the core particle to be degraded (231). The mitochondria also contain a set of proteases with a similar architecture to the proteasome – proteolytic chamber and regulatory components – which recognize and unfold misfolded proteins, directing them to their proteolytic chamber (231). In contrast to the proteasome, the mitochondrial proteases substrates are not tagged with ubiquitin molecules; rather, they are recognized by regulatory domains or subunits of these proteases based on structural properties like folding status (231).

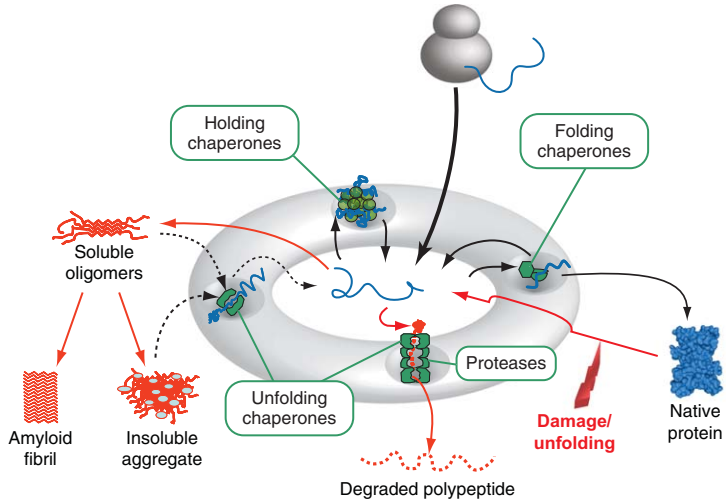


Figure 1.9. The PQC system. The PQC system manages the pool of unfolded and partially folded conformations (center). Molecular chaperones promote folding, protect folding intermediates and disaggregate aggregates or unfold misfolded proteins, whereas proteases eliminate misfolded proteins. Reproduced from Gregersen *et al.*, 2006 (244).

The PQC system regulates the interplay between molecular chaperones and intracellular proteases, being of prime importance in the regulation of protein levels and functions within the cell. Indeed, the PQC system constitutes the hub for the protein homeostasis network (or proteostasis), which, together with its other components, is in place to achieve proteome maintenance with alternative pathways to degradation (Figure 1.10) (246).

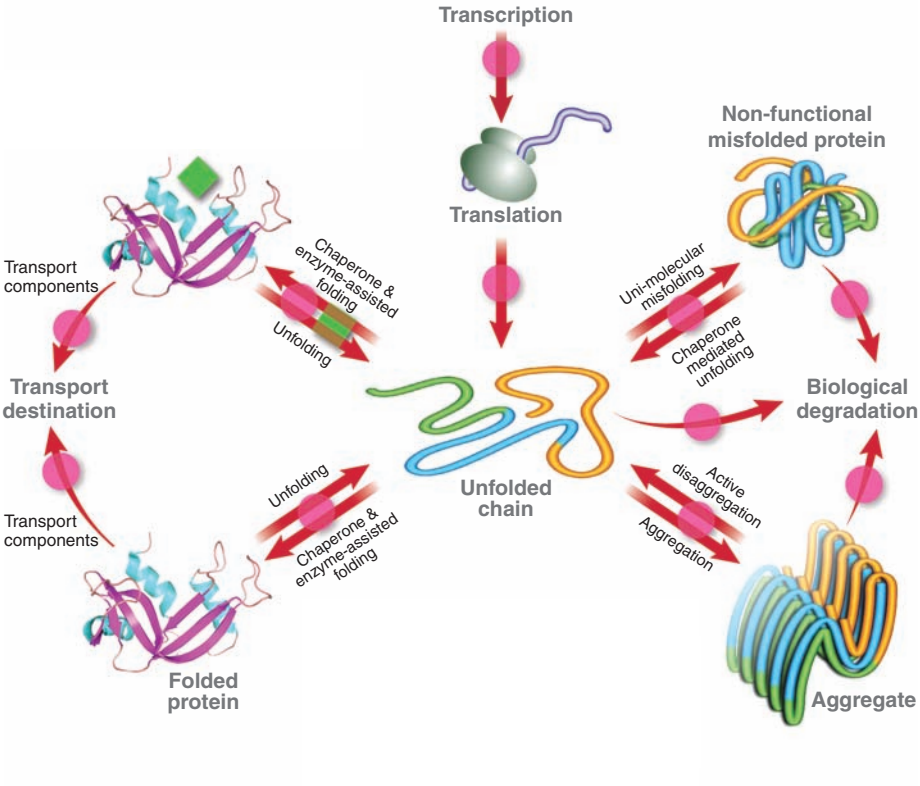


Figure 1.10. The several pathways of the protein homeostasis network. The PQC system constitutes the hub for the protein homeostasis network, managing the pool of unfolded and partially folded conformations, refolding misfolded proteins or removing them by degradation. Imbalances in proteostasis often lead to disease and, therefore, small-molecule-based therapy (magenta and green circles) that manipulate the proteostasis pathways can ameliorate both loss- and gain-of-function diseases (explained below). (ribonuclease A is shown; Protein Data Bank ID, 2BLP). Adapted from Balch *et al.*, 2008 (246).

However, despite the multiple regulatory systems and the intrinsic ability of evolved protein sequences to avoid pathogenic pathways, many genetic disorders are a consequence of protein misfolding, aggregation and degradation (242, 244, 247).

1.3.1.2. Protein misfolding underlies disease

As previously mentioned, in order for a protein to carry out its specific function, it must fold and adopt its three-dimensional conformation (240, 241, 248). A missense mutation affecting a protein structure, stability and/or binding affinity may cause significant functional perturbations (248, 249).

Protein stability is determined by many different factors, and the formation of hydrogen bonds is among the most important ones, as they are vital to many biological phenomena (248, 250). Indeed, hydrogen bonds contribute to protein structural integrity and participate in the function of many macromolecules, such as molecular recognition and catalytic reactions (248, 250). In fact, a mutation resulting in the removal or addition of a hydrogen donor or acceptor is

expected to have a significant impact on the structural integrity, particularly if the affected residue is important for the active site structural and/or functional integrity (248). Furthermore, by disrupting native hydrogen bonds, many missense mutations destabilize the local structure and thus expose the hydrophobic core of the protein to the water phase, which often triggers aggregation (248). The majority of missense mutations destabilize the corresponding protein, such as in PKU, in which phenylalanine hydroxylase (PAH) destabilization due to missense mutations has been considered the major determinant of disease (248, 251).

Despite being fundamental for a protein biological process, protein stability is, however, not sufficient to ensure functional integrity of a particular protein. In fact, proteins are not necessarily optimized to maximize their stability, and their functional regions may be energetically unfavorable. It is widely accepted that active site construction is thermodynamically unfavorable, and, accordingly, many active site mutations – particularly the removal of key catalytic residues – dramatically stabilize enzymes at the expense of activity (248, 252). Mutations of functionally important residues often result in more stable structures, especially when the substitutions occur in polar and charged residues, whereas mutations occurring at the protein surface display less effects on stability (248, 252).

Furthermore, proteins are dynamic molecules in constant motion that can adopt different conformations along the pathway of the corresponding biochemical reaction. The intrinsic flexibility that proteins exhibit provides the ability to sample alternative conformations, and is thus crucial for proteins biological processes (248, 253). Missense mutations can affect the structural flexibility of the entire protein or just a small region, and can either shift the equilibrium between different conformations or can affect the entire conformational dynamics of the molecule (248).

Several factors influence the impact of a mutation, namely the type of substituted amino acid and its location, structural environment of the affected site, and protein structural class and flexibility; all these factors should be considered when evaluating the effect of a mutation (248).

Many efforts have been made to develop *in silico* tools that can accurately predict the effect of a mutation on a particular protein (248). According to the approach used, three groups of bioinformatics tools can be distinguished:

i) programs that are based on different types of data relevant to protein stability and, in some cases, take into account experimental conditions such as temperature, salt concentration and pH values, which are important parameters for assessing the free-energy changes upon mutations at near physiological conditions;
ii) programs that are based on the evolutionary conservation data, and consider that changes occurring at conserved positions tend to be deleterious – despite not directly predicting the effect of a mutation on protein stability, these programs are typically used in conjunction with the first group to achieve consensus predictions;

iii) programs that are based on structural information and assume the ability of a protein to function properly depends on the fundamental physicochemical properties derived exclusively from structures (248).

Bioinformatics tools belonging to the first category of servers include programs that incorporate different physicochemical properties of amino acids and structural preferences of different sites and are trained on the experimental differences of folding free energy caused by mutations; I-Mutant2.0, which uses support vector machines (SVMs) and makes predictions based on either structure or sequence alone and predict actual values of the free energy change (ΔΔG); MuPro, which uses SVMs leveraging both sequence and structural information; MuStab, which predicts only the deleterious effects of mutations and ΔΔG values; PoPMuSiC-2.0, which uses a combination of statistical potential and neural networks to estimate the changes in stability, and exploits statistical potentials that take into account the coupling between four protein sequence and structure descriptors and the amino acid volume variation upon mutation (248, 254-258).

Bioinformatic tools belonging to the second category of servers include programs such as SIFT (Sorting Intolerant From Tolerant), which scores the normalized probabilities for all possible substitutions for a site and calculates the conditional probability that an amino acid is tolerated compared to the most frequent tolerated amino acid; PolyPhen (polymorphism phenotyping), which predicts damaging amino acid substitutions using sequence-based and structure-based features such as sequence conservation, structure and a position-specific independent count matrix derived from the multiple sequence alignment; PolyPhen-2, which uses 8 sequence-based predictive features, 3 structure-based predictive features and a Naive Bayes classifier to predict the functional significance of the mutation (248, 259-261).

Bioinformatic tools belonging to the third category include CUPSAT (Cologne University Protein Stability Analysis Tool), which uses structural environment-specific atom potentials and torsion angle potentials to predict ΔΔG; SDM (Site-Directed Mutator), which uses a statistical potential energy function; FoldX, which uses empirically derived energy functions that are quite accurate since they are based on experimental data (248, 262-265).

Due to the high frequency of missense mutations, a comprehensive understanding of their pathogenic mechanisms elucidates the physicochemical features of protein folding, sheds light on the molecular and biochemical basis of several disorders, and thus constitutes the starting point for the development of therapeutic strategies (239, 248).

1.3.1.3. Conformational disorders

The term ‘conformational disorders’ was originally coined to describe diseases

characterized by the appearance of alternatively folded forms that produced toxic aggregates (231, 266). However, the term has evolved and has been broadened to include all protein misfolding diseases (230, 231, 244).

Conformational disorders are typically caused by mutations that result either in a loss-of-function – in which the protein is prematurely eliminated by proteases, leading to a reduction in its level and/or activity – or in a gain-of-function – in which the protein is not eliminated, but rather accumulated, leading to an increase in its level and/or activity, along with the introduction of a novel pathological function often associated with activation of a pathway (Figure 1.11) (244, 267). While the second type of disorders is usually seen in dominantly inherited disorders, such as late-onset neurodegenerative diseases, the first one is typical of recessively inherited disorders, such as early-onset inborn errors of metabolism (230-233, 244, 267, 268).

The hallmarks of protein folding deficiencies are prolonged interaction with chaperones, decreased steady-state amounts of the affected protein and the occurrence of aggregates (231). On this basis, the therapeutic strategy for these disorders must be to stabilize, in the soluble state, the folding intermediates and target them for further folding instead of degradation and/or aggregation, notwithstanding their pathogenic mechanism (Figure 1.11) (235). Indeed, the conception framework of protein misfolding diseases comprising loss-of-function and gain-of-function pathogenic mechanisms is a practical one in the process of understanding the biochemistry and cell biology as well as in developing strategies for treating these diseases (244).

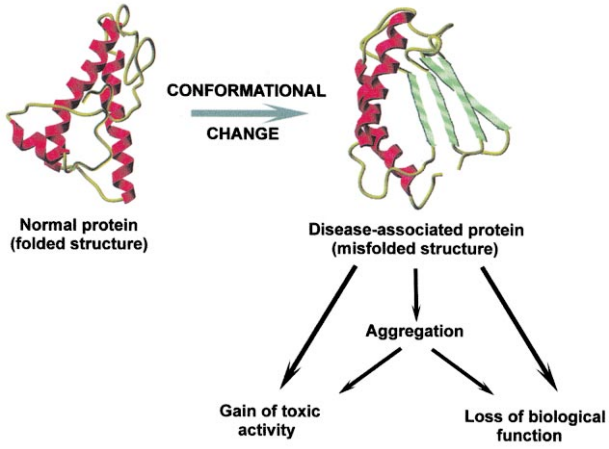


Figure 1.11. Protein misfolding in conformational disorders. Conformational disorders are typically caused by mutations that result either in loss-of-function, characterized by absence biological activity of the folded protein, or in gain-of-function, characterized by a gain of toxic activity by the misfolded protein. Reproduced from Soto, 2001 (268).

1.3.1.4. Therapeutic strategies in loss-of-function diseases

Many examples of conformational disorders in IMD have been described in which the main effect of the mutation is loss-of-function. The pathogenic effect of the mutant variant usually results from insufficiency of a metabolic or transport reaction and from the toxic effects of an accumulated substrate (231, 244).

Many IMD have been described as conformational disorders, as they major pathogenic mechanism is protein misfolding (231). PKU is considered a paradigm of misfolding metabolic diseases, since several reports substantiate that mutations in the *PAH* gene lead to PAH misfolding, increased degradation turnover and to a loss of enzymatic function; and, accordingly, it is believed that PKU presents a high potential for the development of therapeutic approaches using chaperone-based strategy (242, 251).

Over the last few years, molecular medicine has been exploring the usage of chemical and pharmacological chaperones – so-called small-molecules – in conformational disorders as a mean to improve the folding process, thereby ameliorating the clinical picture of these disorders (231, 242).

Chemical chaperones refer to small molecules that act unspecifically – do not bind directly to proteins – and instead act by increasing the fraction of the correctly folded variant (242). Chemical chaperones include osmolytes and protein stabilizers (e.g., glycerol, trehalose, gama-aminobutyric acid) (242, 269). In turn, the term pharmacological chaperone is used to refer small molecules that bind reversibly and weakly to a specific protein, thereby promoting its folding and function. This group is more heterogeneous as it comprises protein cofactors, ligands and competitive inhibitors, among others (242).

In the context of disease-related protein misfolding, the term "chemical chaperone" can be used to refer to the generality of these compounds. Actually, when a pharmacological chaperone – e.g., ligand or cofactor – binds to the protein and favors its native conformation, in practice, it is exerting a chemical chaperoning effect (242).

Chemical chaperones bind to and stabilize the folded functional form of a mutant protein, thereby shifting the folding equilibria away from degradation and aggregation (246). Indeed, chemical chaperones intervene in different cell compartments, at different stages of the protein folding process: i) by directing the folding pathways avoiding dead-end intermediate states; ii) by acting directly over the native conformation, stabilizing it and preventing aggregation; or iii) by promoting refolding of misfolded or destabilized variants (Figure 1.10) (242, 246). Chemical chaperoning increases the concentration and proper localization of the protein by coupling more of the folded state to the trafficking pathway, a pathway that strongly influences proteostasis (246).

Since the fate of the variant proteins is primarily guided by their physicochemical

properties, such as hydrophobicity and amino acid charges, a comprehensive characterization of mutational effects on protein stability and dynamics, and their role in enzyme loss-of-function in conformational diseases is crucial for the development of new therapeutic strategies (235, 270).

1.3.2. Pathogenic molecular mechanisms of genetic disorders – splicing spoilers

Splicing constitutes a critical process in the regulation of gene expression, and is a major contributor to proteomic diversity, as most genes express more than one mRNA by alternative splicing (234, 271-273). Actually, it is now estimated that more than 90% of the human genes with multiple exons have alternatively spliced mRNA isoforms, and that nearly 86% undergo alternative splicing to generate appreciable levels of two or more mRNA isoforms (274).

Pre-mRNA splicing must, therefore, be a tightly regulated process, so that it can occur with high efficiency and fidelity in order to maximize gene expression and to avoid the production of aberrant proteins (275). Indeed, any splicing errors will result in a disconnection between the coding gene and its encoded protein product, and can lead to a wide range of diseases (272, 273, 275, 276).

1.3.2.1. Pre-mRNA splicing reaction

Accurate pre-mRNA splicing requires exon–intron boundaries to be correctly recognized by the splicing machinery, for the introns to be correctly excised, and for the exons to be precisely joined so as to generate functional mature mRNA (271, 277). Paradoxically, the requirement for accurate splicing depends on several degenerate splicing signals weakly conserved intronic *cis*-elements: the 5' and 3' splice sites – almost invariant GT and AG dinucleotides, respectively – the branch site (YNYURAY; underlined A represents the branch point) located approximately 15-35 bases upstream the 3' splice site, and the polypyrimidine tract (Figure 1.12) (274, 277). Nevertheless, these elements are necessary but are by no means sufficient to define exon–intron boundaries (273). Additionally, there are auxiliary *cis*-elements known as splicing-regulatory elements (SREs) that recruit proteins and complexes that can either enhance or silence splicing and have been named descriptively: exonic and intronic splicing enhancers (ESEs and ISEs), and exonic and intronic splicing silencers (ESSs and ISSs) (273, 278).

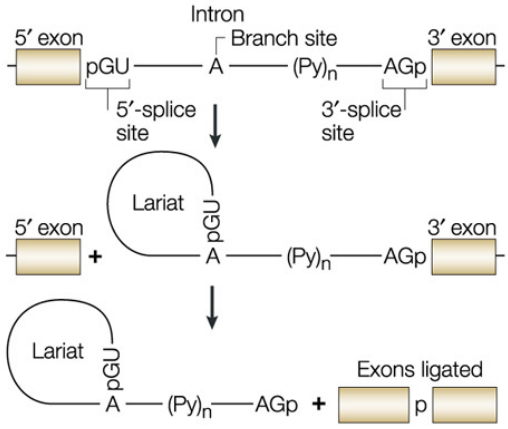


Figure 1.12. Splicing reaction and essential splicing signals. There are several weakly conserved intronic *cis*-elements that constitute essential splicing signals, namely the 5' splice site (GT or GU), the 3' splice site (AG), the branch site comprising the branch point (A), and the polypyrimidine tract (Py_n). Splicing takes places in two transesterification steps. In the first step, the 2'-hydroxyl group of the A nucleotide at the branch site attacks the phosphate at the donor splice site, which leads to cleavage of the 5′ exon from the intron and to the formation of a lariat intermediate. In the second step, the phosphate group (p) at the acceptor splice site and the 3′-hydroxyl group of the detached exon ligates the two exons, thus releasing the intron, still in the form of a lariat. Reproduced from Pagani and Baralle, 2004 (279).

The best characterized SREs are ESEs, which are required for definition and/or efficient splicing of the exons in which they reside, and, in fact, appear to be extremely frequent both in alternative and in constitutive exons (273, 278, 280). ESEs elements stimulate splicing and increase exon inclusion, by serving as binding sites for splicing factors, mainly serine/arginine-rich proteins – SR proteins. SR proteins are a family of non-small nuclear ribonucleoproteins (non-snRNPs) essential for splicing that act at multiple steps of spliceosome assembly and function in both constitutive and regulated splicing (Figure 1.13.A) (278, 281). The sequence characteristics that define SR proteins are the presence of an arginine/serine (RS) domain – composed of at least 50 amino acids with >40% of arginine and serine; and one or more RNA-binding domains of the RNA recognition motif (RRM) type (278).

On the other hand, ESSs suppress the removal of adjacent introns and/or lead to exon skipping, and their silencer activities are usually mediated through binding of members from the heterogeneous nuclear ribonucleoprotein (hnRNP) family (273, 281-285).

Unlike SR proteins, the mechanism whereby hnRNPs interfere with splicing is known only for a small number of cases. These include repressing assembly of components of the splicing machinery through multimerization along exons, blocking the recruitment of small nuclear ribonucleoprotein (snRNPs), or by looping out entire exons (Figure 1.13.B) (278).

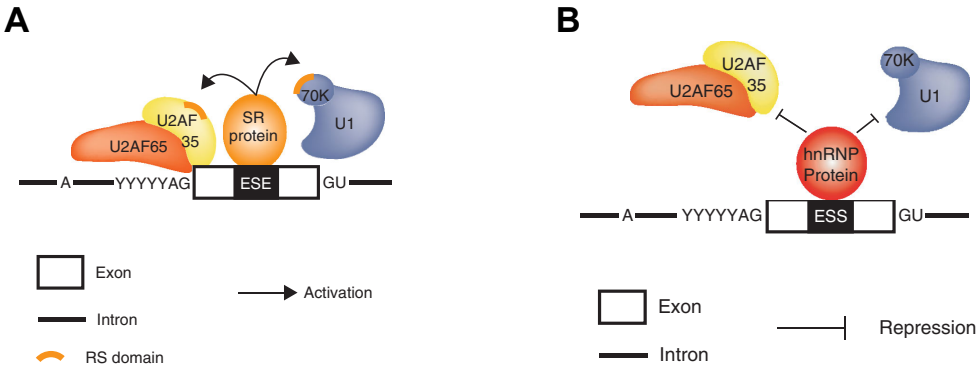


Figure 1.13. ESEs and ESSs elements. A. The ESEs elements stimulate splicing and increase exon inclusion, by serving as binding sites for mainly SR proteins. SR proteins interact with components of the general splicing machinery, such as U1 and U2 snRNPs (small nuclear ribonucleoproteins), which interact with the 5' splice site (GT or GU) and the 3' splice site (AG), respectively, thus promoting their recognition (further explained below). B. The ESSs elements repress splicing by serving as binding sites for mainly hnRNPs, whose mechanisms of action include repressing assembly of components of the splicing machinery and blocking the recruitment of snRNPs. Reproduced from Busch and Hertel, 2012 (278).

The splicing reaction is carried out by the spliceosome, which performs the two primary functions of splicing: recognition of the intron-exon boundaries and catalysis of the cut-and-paste reactions that remove introns and join exons (234).

Spliceosomes do not form a preassembled complex as most ribonucleoprotein (RNP) machines. Rather, their assembly onto pre-mRNA is an ordered process with several distinct intermediates (271, 286). The spliceosome is made up of snRNPs, comprising five small nuclear RNAs (snRNAs) – U1, U2, U4, U5, or U6 snRNA – and more than 300 proteins (234, 278, 287, 288). The U1 and U2 snRNPs form separate particles that bind, respectively, the 5' splice site and the 3' splice site, whereas the U4, U5 and U6 components join together in the U4/U6.U5 tri-snRNP (234, 287, 289). Spliceosome assembly initiates with the binding of the U1 snRNP to the 5' splice site and of the branch point bridging protein (BBP) to the branch site. In an ATP-dependent reaction, U2 snRNP subsequently displaces BBP and binds to the branch site, forming a complex that serves as a platform for the assembly of U4/U6.U5 tri-snRNP; the subsequent release of U1 and U4 results in the formation of the catalytically active complex, and the splicing reaction takes place (Figure 1.14) (287, 289).

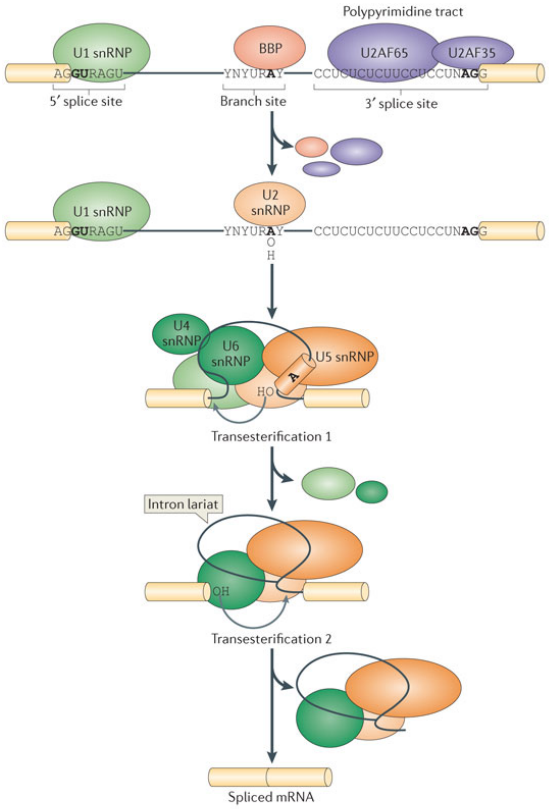


Figure 1.14. Spliceosome assembly. The spliceosome assembles onto pre-mRNA in an ordered process with several distinct intermediates. Spliceosome assembly initiates with the binding of the U1 snRNP to the 5' splice site and of the branch point bridging protein (BBP) to the branch site; U2 snRNP subsequently displaces BBP and binds to the branch site, forming a complex that serves as a platform for the assembly of the U4/U6.U5 tri-snRNP. The subsequent release of U1 and U4 results in the formation of the catalytically active complex, and the splicing reaction takes place. Reproduced from Kornblihtt *et al.*, 2013 (290).

The spliceosome plays an extremely important role in the splicing process, as it must identify bona fide exons, ignore pseudoexons (sequences that resemble an exon, both in size and in the presence of flanking splice-site sequences, but that should never be recognized as an exon by the splicing machinery), join contiguous exons without inadvertent skipping and appropriately regulate alternative splicing to meet the physiological requirements of cells and tissues (273, 291). These challenges are met by combinations of *cis*-regulatory elements and *trans*-acting factors that constitute what is now recognized as the splicing code (291-293).

Extensive efforts have been made to decipher the splicing code (291-293). Whereas the consensus splice sites were relatively easy to identify from alignments of exon–intron boundary sequences, the remainder of the code – particularly exon definition – remains to be fully elucidated (291). Nevertheless, at present, it is well-known that exon definition is accomplished by the accumulated recognition of multiple weak signals, as each exon has a specific set of identity elements – a unique pair of splice site signals and a unique group of SR proteins and hnRNPs binding sites – that ultimately define its overall recognition potential and the overall binding affinity for the spliceosome (234, 278).

Deciphering the splicing code is of paramount importance, as it would allow predicting the splicing pattern(s) in a developmental and cell-type-specific way uniquely from sequence inspection, and would, therefore, enable the reliable prediction of the effect(s) of a mutation (272, 291-293).

1.3.2.2. Splicing as a primary cause of disease

Until recently, the effect of a mutation on gene expression was based exclusively on its location and it was assumed to affect only the coding potential: point mutations in coding regions were usually scored as missense, nonsense or translationally silent mutations, whereas mutations in non-coding regions would have to affect the classical consensus splice-site signals to be considered splicing mutations (234, 273). Furthermore, knowing that the primary effect of an exonic mutation is a splicing defect, rather than the predicted protein-coding mutation, is crucial for understanding the detailed pathogenic mechanism of a disease, since it is important to distinguish if, for instance, a mutation results in loss of expression due to aberrant splicing and/or degradation, or instead results in expression of a protein containing a missense mutation (291). In fact, until recently, mutation detection was accomplished primarily by genomic DNA analysis, in which deep intronic sequences were not scanned, and transcript analysis was rarely performed (294).

In the recent years, however, there has been increasing evidence that both exonic and deep intronic mutations may affect splicing, which entails that many mutations might have been misclassified (273, 277, 294, 295). At present, splicing mutations represent 10% to 30% of the mutations reported to the Human Gene Mutation Database (HGMD® Professional 2013.1). These numbers are, however, probably underestimated, due to the misclassification of many mutations (296).

As explained above, splicing is a highly intricate process that shows extraordinary structural and compositional dynamics, both crucial features for splicing fidelity and accuracy. It is therefore not surprising that disruption in any of the events involved in the pre-mRNA processing may lead to disease (297).

Indeed, according to the mechanism whereby splicing defects result in a primary cause of disease, splicing mutations can be divided into two broad categories: i) mutations that disrupt a *cis*-acting element, in which a single gene is affected and in which the mutation can either lead to aberrant splicing or to aberrant regulation of splicing; ii) mutations that disrupt a *trans*-acting factor, in which multiple genes are affected, since the mutation will interfere with either a component of the splicing machinery or with the splicing regulatory complex (Figure 1.15) (234).

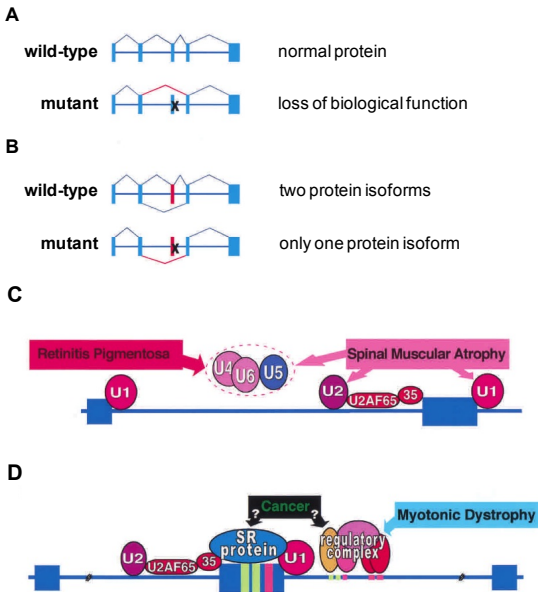


Figure 1.15. Molecular mechanisms whereby splicing defects lead to disease. Splicing mutations can be divided into two broad categories: *cis*- and *trans*-acting mutations. The first category includes *cis*-acting mutations, which alter the splicing of either a constitutive (A) or an alternative exon (B); this class constitutes the great majority of the splicing mutations. The second category includes *trans*-acting mutations, which can interfere with a component of the splicing machinery (C) or with the splicing regulatory complex (D). The distinction between *cis*- and *trans*- acting effects has important mechanistic implications, as effects in *cis* have a direct impact on the expression of only one gene, whereas effects in *trans* have the potential to affect the expression of multiple genes. Adapted from Faustino and Cooper, 2003 (234).

The majority of splicing mutations are single nucleotide substitutions that affect *cis*-acting elements. These mutations can result in exon skipping, retention of the mutated intron, usage of a cryptic 5' or 3' splice site, introduction of a new splice site within an exon or intron, or more rarely activation of a pre-existing cryptic splice site distal from the mutation. These mutations often cause a shift in the opening reading frame (ORF), which, in case of introduction of a premature termination codon (PTC) into the mRNA, usually triggers nonsense-mediated mRNA decay (NMD). The NMD is a post-transcriptional mechanism that, by eliminating abnormal transcripts that prematurely terminate translation, prevents the production of truncated proteins that could function in dominant-negative or other deleterious mechanisms (298). According to the established rule, transcripts with a PTC located more than 50-55 nucleotides upstream the next exon-exon junction are generally recognized as premature and elicit NMD (298, 299). This pathway, also known as mRNA surveillance mechanism, is an mRNA quality control system, particularly important in alternative splicing, as one-third of the resulting variant transcripts introduce PTCs and are degraded through NMD (291, 300).

Effectively, it is estimated that one-third of disease-causing mutations may induce aberrant splicing of pre-mRNA transcripts and a partially overlapping third to PTCs and NMD; and, in some diseases, the estimates even go up to 50% and >70% for aberrant splicing and NMD, respectively (300).

1.3.2.3. The deleterious effect of point mutations on pre-mRNA splicing in IMD

Several examples of *cis*-acting splicing mutations have been described for IMD, namely for organic acidemias, lysosomal disorders, congenital disorders of glycosylation, tetrahydrobiopterin deficiencies and fatty acid oxidation disorders (291, 294). Interestingly, in the fatty acid oxidation disorder resultant from deficiency in medium-chain acyl-CoA dehydrogenase (MCAD), one of the described missense mutations in the *MCAD* gene affects an ESE, which is inactivated, causing exon skipping, thus leading to disease (291). Additionally, a silent A→C polymorphism in the same exon located 11 nucleotides upstream from that ESE mutation modulates the severity of exon skipping, since the C variant inactivates an adjacent ESS, thereby lessening the skipping of the exon (291). This particular example reiterates the importance of exonic mutations in splicing defects, as it illustrates the complexity of the splicing code and the difficulties we currently have in deciphering it, and thus reinforce the importance of decoding it (291).

Indeed, in order to design a therapeutic strategy, it is crucial to know the molecular mechanism whereby a genetic variation affects splicing. Given the impossibility to do so directly from the genomic sequence, several methodologies were developed to aid in the elucidation of the role of a genetic variation in the pre-mRNA processing.

The most direct approach to determine how a genetic variation is affecting splicing is to perform a reverse transcription PCR (RT-PCR) analysis from the relevant tissue(s) of affected individuals (301). However, obtaining RNA from the disease-relevant tissues is frequently a major limitation; and, even if an RNA sample from the affected tissue is available, the detection of a particular splicing product is not straightforward, as aberrant transcripts are often degraded by NMD, thereby precluding their detection by normal RNA analysis (291). This problem of degradation can be solved by establishing stable cell lines from the patient lymphoblasts and treating them with NMD inhibitors, such as puromycin and cycloheximide, and then evaluate the transcription profile (294, 302, 303).

An alternative approach is to use minigene-based technologies (276, 301). Initially described 20 years ago, minigenes constitute a relatively fast approach to identify splicing spoilers and to study their underlying functional mechanisms (276). In this methodology, the relevant genomic segment is cloned into a plasmid between an upstream ubiquitous transcriptional promoter and a downstream gene segment necessary for mRNA 3' end formation. The plasmid is transfected into mammalian cells, and the transcription pattern is analyzed by RT-PCR with primers specifically designed to amplify processed transcripts derived from the minigene, to distinguish them from the cell endogenous transcripts (279, 280, 301). In addition to

their utility for analyzing the effects of mutations and allelic variants on splicing efficiency, minigenes have been a mainstay of investigations to identify the *cis*-acting elements and *trans*-acting factors that regulate tissue-specific alternative splicing (301).

A third method is the *in vitro* splicing assay. In this procedure, labelled pre-formed RNA molecules that are transcribed with bacterial polymerases are incubated in the presence of nuclear extracts and the resulting spliced products are resolved on polyacrylamide denaturing gel. With this assay, the intermediates of the splicing reactions, such as the lariat, can be evaluated and more molecular mechanistic studies can be performed. However, in contrast to the minigene procedure, the *in vitro* assay only allows relatively short sequences to be studied, does not reproduce nuclear organization and architecture and, most importantly, does not take into account the fact that transcription and splicing are intimately connected in the cell (274, 279).

A great boost for predicting the impact of nucleotide variations on splicing has been the development of *in silico* tools, which evaluate the many possible influences of a mutation on pre-mRNA splicing (304, 305). Bioinformatic tools include programs such as Splice-Site Prediction by Neural Network (NNSplice) and Cryptic Splice Finder (CSF), which predict splice sites; ESEfinder and Relative Enhancer and Silencer Classification by Unanimous Enrichment (RESCUE-ESE), which predict ESEs; PESX, which detects Putative ESEs and ESSs; Human Splicing Finder (HSF), which predicts splicing enhancer and silencer motifs by different sets of algorithms; *m*fold, which predicts all possible optimal and suboptimal secondary structures of a particular mRNA sequence, and *s*fold, which predicts only the best mRNA secondary structure; and SpliceAid 2, which, besides the 5' splice site, 3' splice site, polypyrimidine tract and branch point, also evaluates tissue-specific and cell-specific expression of SREs and *trans*-acting factors (304, 306-318). The main problem that remains for any algorithm is the degree of certainty of the predictions in the identification of splicing defects (302). Indeed, bioinformatic tools *per se* lack of reliability, and none of the aforementioned programs is 100% conclusive or predictive. Nevertheless, *in silico* approaches can and should be used in predicting the effect of a genetic variation on the splicing process, but as an integrated part of the experimental approach (276, 284, 302, 317, 319).

Unraveling the pathogenic mechanism of a mis-splicing mutation is, indeed, the starting point for further research and for the development of therapeutic interventions, in which directly targeting RNA is an increasingly compelling therapeutic strategy (302, 320).

1.3.2.4. Molecular approaches to splicing therapeutics

The screening of mutations in individuals at the genome and transcriptome scales allows the design of different therapeutic strategies for mutation-specific, personalized medicine based

on how the mutations take their toll (300).

Since the recognition of splicing as a primary cause of disease, several therapeutic strategies directed toward correcting or circumventing splicing abnormalities have arisen (234, 320). Approaches include overexpression of proteins that alter splicing of the affected exon, use of oligonucleotides to block use of aberrant splice sites and force use of beneficial splice sites, use of small molecules that indirectly modulate splicing (e.g., by altering phosphorylation of splicing factors or by stabilizing putative secondary structures, high-throughput screens to identify compounds that influence splicing efficiencies of target pre-mRNAs, and a *trans*-splicing approach to replace mutated exons with wild-type exons (Figure 1.16) (234).

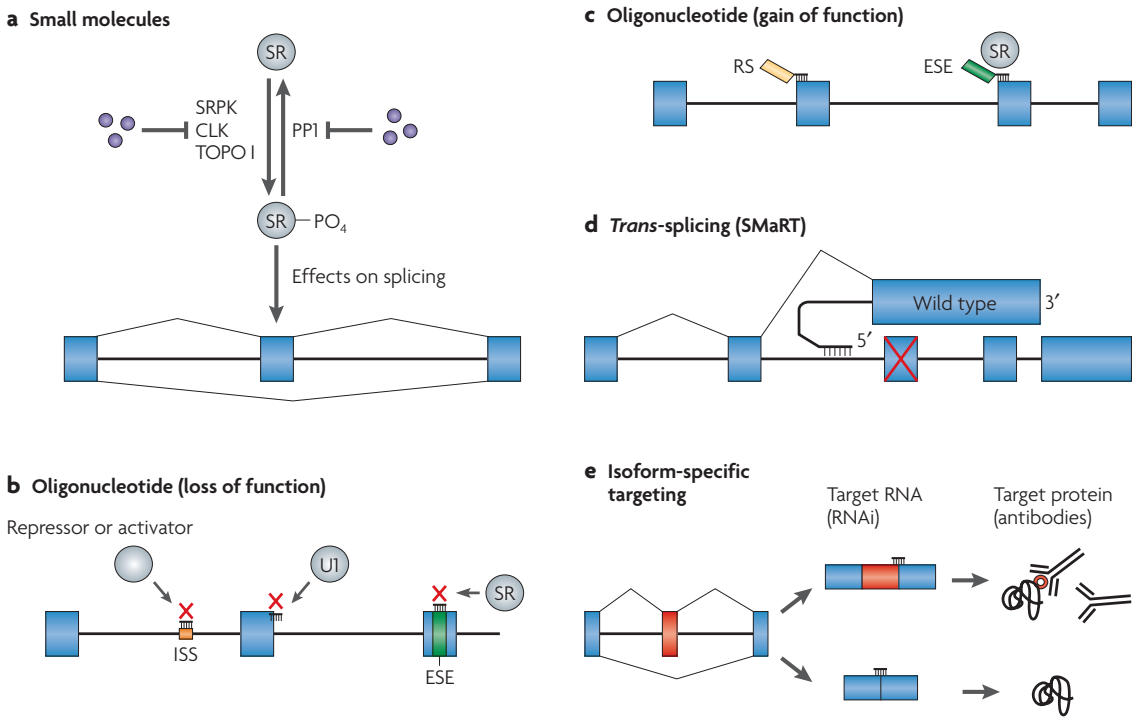


Figure 1.16. Molecular approaches to splicing therapeutics. Therapeutic approaches can either alter the splicing patterns of target genes (panels a–d) or target specific splice variants at the RNA or protein level to achieve a therapeutic affect (panel e). a) Small-molecule-based therapy, in which small-molecules modulate the splicing patterns by, e.g., altering the SR proteins phosphorylation state. b,c) Antisense therapy, in which antisense oligonucleotides are used to block the access of splicing factors to either the splice sites or to other components of the splicing code (panel b), or to enhance splicing of the targeted exon (panel c). d) *Trans*-splicing or SmaRT (spliceosome-mediated RNA *trans*-splicing) approach is used to reprogram the mRNA expressed from mutated genes, by tethering exogenous RNA into the endogenous mutated pre-mRNA. e) Isoform-specific targeting, in which alternative or aberrant splicing is targeted through the recognition of specific-mRNA or protein, by specific RNA interference- or antibodies-mediated degradation, respectively. As with all therapies, splicing-based approaches have issues with off-target effects, toxicity, efficiency and delivery. The potential applications are diverse; however, most applications are gene-specific and are therefore limited to specific diseases so they must be optimized for individual genes. Reproduced from Wang and Cooper, 2007 (291).

Currently, the most widely used therapy toward correcting aberrant splicing is pre-mRNA targeting anti-sense oligonucleotides (ASOs). ASOs are short synthetic, modified nucleic

acids that hybridize with the target mRNA and/or pre-mRNA through Watson-Crick base pairing and modulate its function (313, 321-323).

ASO binding to the mRNA and/or pre-mRNA result in decreased levels of translation of the target transcript by various mechanisms, depending on the chemistry properties of the ASO and on the location of hybridization (313, 322). These mechanisms can be broadly categorized as i) degradative, in which the ASO promotes degradation of the RNA either through endogenous enzymes, such as RNase H, or through cleavage mechanisms designed into the oligonucleotide; and as ii) non-degradative, in which the ASO binding to the mRNA or pre-mRNA modulates RNA function without promoting its degradation, rather ASO binding either sterically hinders translation of the mRNA or modulates pre-mRNA processing (320-322, 324). Actually, masking a regulatory sequence in the pre-mRNA is possible by designing an ASO complementary to the pre-mRNA area of interest, in a way that the *trans*-splicing regulatory factors are unable to bind and changes in splicing occur (325, 326). The application of oligonucleotide-based methods to redirect and modulate pre-mRNA splicing was first demonstrated for β-thalassemia and now shows promise for both spinal muscular atrophy and Duchenne muscular dystrophy (267, 327).

In order to be effective as therapeutics, ASOs must reach and selectively hybridize to the target RNA. This requires appropriate biodistribution of the ASO, efficient uptake, and a sufficient half-life within tissues (321).

Unmodified ASOs are rapidly degraded *in vivo* by the endogenous nucleases and their overall charge prevent them from penetrating the plasmatic membrane, which confers them poor potency (313, 321). Accordingly, in order to obtain efficacious activity, it would be necessary to administer high doses of unmodified ASOs, which would increase the possibility of pernicious off-target effects, both through hybridization to non-targeted RNAs and through interactions with proteins (321). Therefore, various chemical modifications have been developed, aiming to enhance nuclease resistance, prolong tissue half-life, increase affinity and potency, and reduce non-sequence-specific toxicity (Figure 1.17) (313).

The earliest and best characterized modified ASOs are the phosphorothioate (PS)-ASO, whose mechanism of action involves RNase H-mediated cleavage of target mRNA. The PS-backbone modification, in which one of the non-bridging oxygen atoms in the phosphodiester bond is replaced by a sulphur atom, confers higher resistance to the ASOs against nuclease degradation, thereby improving their bioavailability (Figure 1.17) (313, 321). Indeed, the antisense effect of the PS-ASOs can be observed for over 48 hours after a single application to tissue culture cells (328). However, PS modification may slightly reduce the affinity of the ASOs for the mRNA target due to the fact that the melting temperature of the heteroduplex decreases by approximately 0.5 ºC per nucleotide. In addition, PS-ASOs have also been reported to produce non-sequence-specific toxicities because of their affinity to bind to proteins, e.g., serum albumin (313, 321, 322, 329).

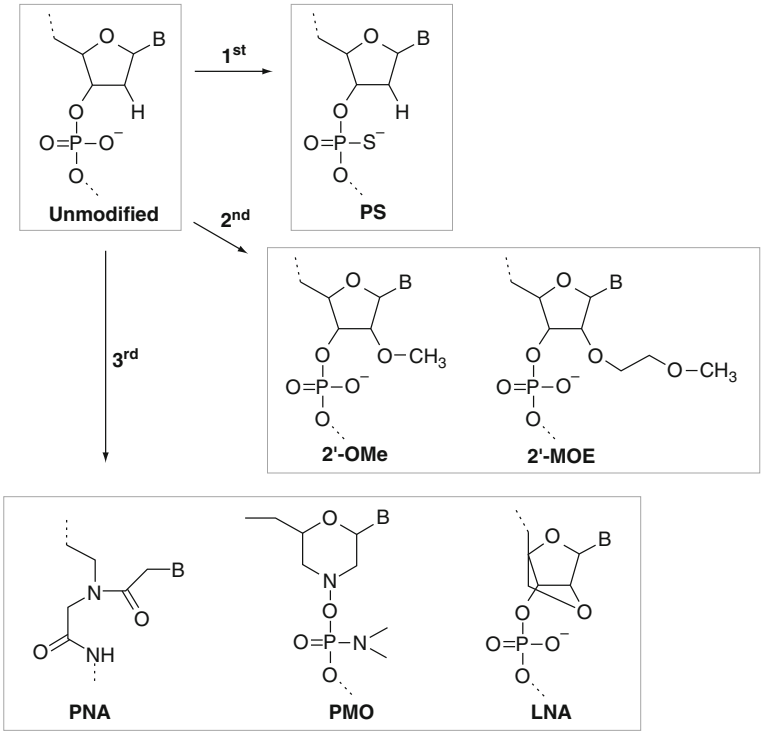


Figure 1.17. Three generations of ASOs. Modifications to the sugar and/or the backbone of the nucleotide provide ASOs with different biophysical, biochemical and biological properties: phosphorothioate (PS) – first generation; 2'-O-Methyl (2'-OMe) and 2'-O-(2'-methoxy)ethyl (2'-MOE) – second generation; peptide nucleic acid (PNA), phosphoroamidate morpholino oligomer (PMO) and locked nucleic acid (LNA) – third generation. Adapted from Kunze *et al.*, 2010 (322).

To further increase the binding affinity for target mRNA, second-generation ASOs with 2'-alkyl modifications of the ribose were developed, of which 2'-O-Methyl (2'-OMe) and 2'-O-(2'-methoxy)ethyl (2'-MOE) modifications of PS-ASOs are the two most widely studied (Figure 1.17) (313, 322). Chemical modification of the 2' position of the sugar moiety primarily increases binding affinity to target RNA, and so these ASOs show a slightly enhanced target affinity (321). Furthermore, 2'-OMe- or 2'-MOE-ASOs are less toxic than PS-ASOs, show an increased tissue half-life due to improved nuclease resistance, and present the potential for oral administration. Unexpectedly, 2'-OMe and 2'-MOE substitutions do not support RNase H-mediated cleavage of target mRNA, which implies they can only exert antisense activity via inhibition of translation, thereby dampening their efficacy (313, 322, 329). To circumvent this shortcoming, a chimeric ASO was developed in which the central "gap" region consists of approximately ten PS-modified 2'-deoxynucleotides and in which each flank consists of approximately five 2'-OMe or 2'-MOE PS-modified nucleotides. This chimeric 'gapmer' ASO allows RNase H to sit in the central gap to execute target-specific mRNA degradation, while the flanks resist to nuclease cleavage due to the 2'-alkyl modifications (313, 321, 322).

In the third generation of ASOs, further chemical modifications of the furanose ring were performed, conferring them enhanced target affinity and nuclease resistance. Of these ASOs,

peptide nucleic acid (PNA), phosphoroamidate morpholino oligomer (PMO) and locked nucleic acid (LNA) are the three most studied (Figure 1.17) (313).

PNA is a non-charged nucleotide analogue, in which the phosphodiester backbone is replaced with a flexible pseudopeptide polymer (N-(2-aminoethyl)glycine) and nucleobases are attached to the backbone via methylene carbonyl linkage. PNA-ASOs hybridize complementary RNA with high affinity and specificity and exert the antisense effect by forming a sequence-specific duplex with the target mRNA, causing steric hindrance of the translational machinery, thereby leading to protein knockdown (313).

PMO chemistry involves the replacement of the ribose sugar with a six-membered morpholino ring and the charged phosphodiester intersubunit linkage is replaced by an uncharged phosphorodiamidate linkage. PMO does not allow RNase H-mediated cleavage, and so its antisense effect is primarily mediated by steric interference either at the splicing level or at the translational level (330). Regarding splicing modulation, binding of a PMO-ASO to a 5' and/or 3' splice site in the pre-mRNA can cause the splicing machinery to effectively skip an entire exon, in order to induce removal of in-frame exons containing a missense, nonsense or frameshift mutation, thereby rescuing, at least partly, the protein function; or to skip a pseudoexon, by sterically blocking the access of the splicing machinery to the pseudoexonic regions, in order to force the usage of the natural splice sites, recovering normally spliced transcripts that encode the functional protein (294, 330-332). Regarding translational modulation, PMO-ASO binding to mRNA sterically hinders ribosomal assembly, thereby leading to translational arrest and thus to gene silencing (313).

Both PNA and PMO chemical modifications confer them excellent resistance to nucleases and proteases, and, since their backbones are uncharged, their ASOs are unlikely to form unwanted interactions with nucleic acid-binding proteins (329, 333-336).

LNA is a conformationally restricted nucleotide containing a methylene group linking the 2'-O and 4' position of the ribose. This chemistry serves to lock the furanose ring into the *N*-type sugar conformation (2'-O,4'-C-methylene-β-D-ribofuranosyl), thereby presenting a high hybridization affinity with RNA (337, 338).

LNA-ASOs present a high hybridization affinity towards target RNA and the resultant duplexes are not substrates for RNase H (313, 334, 337, 339). Notwithstanding, LNA is compatible with PS linkages and with other monomers, e.g., DNA and RNA, whose chimeric ASOs can elicit RNase H-mediated mRNA cleavage (313, 337). There are, however, contradictory reports regarding the properties the chimeric sequence must exhibit in order to elicit RNase H activity, namely in the length of the sequence and in the modification pattern of the ASO (337). Nevertheless, the overall conclusion is that LNA-ASOs can indeed be designed to present a degradative mechanism of action through RNase H, and with the advantage of improved binding and target accessibility conferred by the LNA monomers (337, 339).

Moreover, LNA presents various properties that make it a promising molecule in the field of nucleic acid research, namely good aqueous solubility, high biological stability, low toxicity and, more importantly, outstanding mismatch discrimination (339). LNA charged phosphate backbone allows ready delivery into cells using standard cationic transfection agents; and studies in a mouse model for splicing modulation activity demonstrated that LNA-ASOs are nontoxic and that the sequence-specific antisense effects of a single-dose regimen of a LNA-ASO persists for more than 3 weeks (336). Furthermore, oral delivery of the unformulated LNA-ASO induced minimal but clearly detectable splice modulation in a sequence-specific manner in targeted pre-mRNA (340). Moreover, in another study, LNA-ASOs were shown to be delivered to both liver and kidney of mice and non-human primates with remarkable efficacy following parenteral administration (341, 342).

The recent developments in oligonucleotides chemistry have provided ASOs with different biophysical, biochemical and biological properties based on various modifications to the sugar or the backbone of the nucleotides, which greatly enhanced the drug-like properties of ASOs, allowing rational design of diverse and powerful antisense therapeutic tools with extensive clinical applications (321, 343). Nevertheless, there are still some challenges for ASOs to overcome in order to constitute an effective antisense therapeutic strategy: i) the efficiency with which the treatment corrects the splicing defect; ii) the off-target effects, i.e., the side-effects on splicing of pre-mRNAs other than the target; and iii) the effectiveness and long-lasting delivery to the target cells (267, 313, 320, 323).

With the recent clinical success of several antisense-therapies, and establishment of proof-of-concept efficacy in several disease models, ASOs have established themselves as a promising and rapidly-developing therapeutic strategy covering a wide range of genetic disorders. With such dramatic improvements in antisense technology in a relatively short time frame, and with the current frenzied pace of antisense research, new and enhanced ASOs designs will likely be forthcoming and will facilitate their widespread application in the clinical realm (344).

1.3.3. *The future is now for rare genetic disorders* [(345)]

More than a decade after the first draft of the human genome sequence, the remarkable wealth of knowledge has profoundly affected our understanding of the human pathophysiology at a molecular level (345). Indeed, the growing knowledge of the human genome, epigenome, transcriptome, proteome and metabolome has led to development of many therapeutic strategies, in which not only clinical but especially molecular data are applied in the design of the appropriate drug for a particular person (346). In fact, treatment of genetic diseases is no longer

just disease-oriented, it is now stratified to fit the patient, the allele, and the process underlying the phenotype (347).

Undoubtedly, the post-genomic era has begun, and with it the promise of tailoring the practice of medicine to the individual (348).

1.4. Objectives

After more than 100 years since the first description of galactosemia, there are still many "clouds gathering over galactosemia". Indeed, despite early diagnosis and lifelong dietary restriction of galactose resolving the potentially lethal symptoms of classic galactosemia, the long-term outcome is often disappointing with most patients developing complications in the long run.

Advances in human genome research in the past decade have shifted the medicine's paradigm of disease-oriented care into the new paradigm of patient-focused care. Indeed, with the post-genomic era, has arisen the concept of personalized medicine, in which functional genomic analyses provide important insights into the molecular pathogenesis of genetic disorders, thus identifying new treatment targets that guide in the design and development of new therapeutic approaches.

Molecular studies in inherited metabolic disorders have led to the realization that only a minor part of the mutations directly disrupt functional sites of the proteins, as most genetic defects disturb the protein at its structural level, and/or disturb its synthesis by affecting the pre-mRNA processing, ultimately affecting the protein at its cellular level. Hence, mutation-specific structural and functional analyses are of paramount importance for the development of new therapeutic approaches, as a better understanding of their underlying molecular mechanisms of pathogenesis opens new avenues of research, paving the path for the development of new therapeutic strategies to tackling inborn errors of metabolism.

Accordingly, the aims of this study were:

- Genotyping all Portuguese galactosemic patients, and displaying the mutational spectrum of classic galactosemia in Portugal;
- Functional characterization of the highly prevalent intronic variation c.820+13A>G and its correction by antisense therapy;
- Structural-functional characterization of the most prevalent GALT missense mutants, and of other clinically important missense mutants;
- Development of a prokaryotic model of galactose sensitivity to evaluate the ability of human GALT variants in rescuing bacteria from galactose-induced toxicity;
- Development of an unicellular model and of biophysical methodologies to characterize human GALT mutants and ultimately to guide in the development of new therapeutic agents in classic galactosemia.

1.5. References

1.Fridovich-Keil JL, Walter JH. *Galactosemia*. In: Valle D, Beaudet AL, Vogelstein B, Kinzler KW, Antonarakis SE, Ballabio A, editors. The Online Metabolic and Molecular Bases of Inherited Disease: Mc-Graw Hill; 2008. p. 1-92.

2.Lai K, Klapa MI. *Alternative pathways of galactose assimilation: could inverse metabolic engineering provide an alternative to galactosemic patients?* Metab Eng. 2004;6(3):239-44.

3.Wood IS, Trayhurn P. *Glucose transporters (GLUT and SGLT): expanded families of sugar transport proteins*. Br J Nutr. 2003;89(1):3-9.

4.Leturque A, Brot-Laroche E, Le Gall M. *GLUT2 mutations, translocation, and receptor function in diet sugar managing*. Am J Physiol Endocrinol Metab. 2009;296(5):E985-92.

5.Steinhoff-Wagner J, Gors S, Junghans P, Bruckmaier RM, Kanitz E, Metges CC, Hammon HM. *Intestinal glucose absorption but not endogenous glucose production differs between colostrum- and formula-fed neonatal calves*. J Nutr. 2011;141(1):48-55.

6.Decombaz J, Jentjens R, Ith M, Scheurer E, Buehler T, Jeukendrup A, Boesch C. *Fructose and galactose enhance postexercise human liver glycogen synthesis*. Med Sci Sports Exerc. 2011;43(10):1964-71.

7.Schirmer WJ, Townsend MC, Schirmer JM, Hampton WW, Fry DE. *Galactose clearance as an estimate of effective hepatic blood flow: validation and limitations*. J Surg Res. 1986;41(5):543-56.

8.Lai K, Elsas LJ, Wierenga KJ. *Galactose toxicity in animals*. IUBMB Life. 2009;61(11):1063-74.

9.Lee SJ, Lewis DE, Adhya S. *Induction of the galactose enzymes in Escherichia coli is independent of the C-1-hydroxyl optical configuration of the inducer D-galactose*. J Bacteriol. 2008;190(24):7932-8.

10.Timson DJ, Reece RJ. *Identification and characterisation of human aldose 1-epimerase*. FEBS Letters. 2003;543(1-3):21-4.

11.Bosch AM. *Classical galactosaemia revisited*. J Inherit Metab Dis. 2006;29(4):516-25.

12.McCorvie TJ, Timson DJ. *The structural and molecular biology of type I galactosemia: Enzymology of galactose 1-phosphate uridylyltransferase*. IUBMB Life. 2011;63(9):694-700.

13.Segal S. *Galactosemia unsolved*. Eur J Pediatr. 1995;154(2):S97-S102.

14.Bosch AM. *Classical galactosemia revisited* [Ph.D. Dissertation]: Faculty of Medicine, University of Amsterdam; 2004.

15.Leloir LF. *Two decades of research on the biosynthesis of saccharides*. Science. 1971;172(990):1299-303.

16.Lai K, Elsas LJ. *Overexpression of human UDP-glucose pyrophosphorylase rescues galactose-1-phosphate uridyltransferase-deficient yeast*. Biochem Biophys Res Commun. 2000;271(2):392-400.

17.Novelli G, Reichardt JK. *Molecular basis of disorders of human galactose metabolism: past, present, and future*. Mol Genet Metab. 2000;71(1-2):62-5.

18.Berry GT. *Galactosemia: when is it a newborn screening emergency?* Mol Genet Metab. 2012;106(1):7-11.

19.Reichardt JK. *Molecular basis of galactosemia: Mutations and polymorphisms in the gene encoding human galactose-1-phosphate uridylyltransferase.* Proc Natl Acad Sci U S A. 1991.

20.Fridovich-Keil JL. *Galactosemia: the good, the bad, and the unknown*. J Cell Physiol. 2006;209(3):701-5.

21.Holden HM, Rayment I, Thoden JB. *Structure and function of enzymes of the Leloir pathway for galactose metabolism*. J Biol Chem. 2003;278(45):43885-8.

22.Frey PA. *The Leloir pathway: a mechanistic imperative for three enzymes to change the stereochemical configuration of a single carbon in galactose*. the FASEB. 1996;10:461-70.

23.Timson DJ, Reece RJ. *Functional analysis of disease-causing mutations in human galactokinase*. Eur J Biochem. 2003;270(8):1767-74.

24.Holden HM. *Galactokinase: structure, function and role in type II galactosemia*. Cell Mol Life Sci. 2004;61(19-20):2471-84.

25.Tang M, Wierenga K, Elsas LJ, Lai K. *Molecular and biochemical characterization of human galactokinase and its small molecule inhibitors*. Chem Biol Interact. 2010;188(3):376-85.

26.Zhou T, Daugherty M, Grishin NV, Osterman AL, Zhang H. *Structure and Mechanism of Homoserine Kinase: Prototype for the GHMP Kinase Superfamily.* Structure. 2000;8:1247-57.

27.McCorvie TJ, Timson DJ. *In silico prediction of the effects of mutations in the human UDP-galactose 4'-epimerase gene: towards a predictive framework for type III galactosemia*. Gene. 2013;524(2):95-104.

28.Sanders RD, Sefton JM, Moberg KH, Fridovich-Keil JL. *UDP-galactose 4' epimerase (GALE) is essential for development of Drosophila melanogaster*. Dis Model Mech. 2010;3(9-10):628-38.

29.Openo KK, Schulz JM, Vargas CA, Orton CS, Epstein MP, Schnur RE, Scaglia F, Berry GT, Gottesman GS, Ficicioglu C, Slonim AE, Schroer RJ, Yu C, Rangel VE, Keenan J, Lamance K, Fridovich-Keil JL. *Epimerase-Deficiency Galactosemia Is Not a Binary Condition*. Am J Hum Genet. 2006;78:89-102.

30.Komrower GM. *Galaetosaemia - Thirty Years on. The Experience of a Generation*. J Inher Metab Dis. 1982;5(2):96-104.

31.Bennett MJ. *Galactosemia diagnosis gets an upgrade*. Clin Chem. 2010;56(5):690-2.

32.Pintor J. *Sugars, the Crystalline Lens and the Development of Cataracts*. Biochemistry & Pharmacology: Open Access. 2012;01(04).

33.Berry GT. *The role of polyols in the pathophysiology of hypergalactosemia*. Eur J Pediatr. 1995;154(2):S53-S64.

34.Jez JM, Flynn G, Penning TM. *A New Nomenclature for the Aldo-Keto Reductase Superfamily*. Biochemical Pharmacology. 1997;54:639-47.

35.Ko BCB, Ruepp B, Bohren KM, Gabbay KH, Chung SSM. *Identification and characterization of multiple osmotic response sequences in the human aldose reductase gene*. J Biol Chem. 1997;272(26):16431-7.

36.Ramasamy R, Goldberg IJ. *Aldose reductase and cardiovascular diseases, creating human-like diabetic complications in an experimental model*. Circ Res. 2010;106(9):1449-58.

37.Yabe-Nishimura C. *Aldose reductase in glucose toxicity: a potential target for the prevention of diabetic complications*. Pharmacol Rev. 1998;50(1):21-34.

38.Schadewaldt P. *Age dependence of endogenous galactose formation in Q188R homozygous galactosemic patients*. Mol Genet Metab. 2004;81(1):31-44.

39.Ning C, Reynolds R, Chen J, Yager C, Berry GT, Leslie N, Segal S. *Galactose metabolism in mice with galactose-1-phosphate uridyltransferase deficiency: sucklings and 7-week-old animals fed a high-galactose diet*. Mol Genet Metab. 2001;72(4):306-15.

40.Cuatrecasas P, Segal S. *Mammalian Galactose Dehydrogenase: I. Identification and Purification in Rat Liver.* J Biol Chem. 1966;241(24):5904-9.

41.Wehrli SL, Berry GT, Palmieri M, Mazur A, Elsas II LJ, Segal S. *Urinary galactonate in patients with galactosemia: quantitation by nuclear magnetic resonance spectroscopy*. Pediatr Res. 1997;42(6):855-61.

42.Segal S, Wehrli S, Yager C, Reynolds R. *Pathways of galactose metabolism by galactosemics: evidence for galactose conversion to hepatic UDPglucose*. Mol Genet Metab. 2006;87(2):92-101.

43.Berry GT, Leslie N, Reynolds R, Yager CT, Segal S. *Evidence for alternate galactose oxidation in a patient with deletion of the galactose-1-phosphate uridyltransferase gene*. Mol Genet Metab. 2001;72(4):316-21.

44.Holton JB, Walter JH, Tyfield LA. *Galactosemia*. In: Scriver CR, Beaudet AL, Sly WS, Valle D, editors. The Metabolic and Molecular Bases of Inherited Disease. 8 ed. New York: McGraw-Hill; 2001. p. 1553–87.

45.Isselbacher KJ. *A mammalian uridinediphosphate galactose pyrophosphorylase*. J Biol Chem. 1958;232(1):429-44.

46.Holton JB. *Effects of galactosemia in utero*. Eur J Pediatr. 1995;154(1):S77-S81.

47.Schulz JM, Ross KL, Malmstrom K, Krieger M, Fridovich-Keil JL. *Mediators of galactose sensitivity in UDP-galactose 4'-epimerase-impaired mammalian cells*. J Biol Chem. 2005;280(14):13493-502.

48.Gitzelmann R. *Formation of galactose-1-phosphate from uridine diphosphate galactose in erythrocytes from patients with galactosemia*. Pediatr Res. 1969;3(4):279-86.

49.Berry GT, Moate PJ, Reynolds RA, Yager CT, Ning C, Boston RC, Segal S. *The rate of de novo galactose synthesis in patients with galactose-1-phosphate uridyltransferase deficiency*. Mol Genet Metab. 2004;81(1):22-30.

50.Berry GT, Nissim I, Gibson JB, Mazur AT, Lin Z, Elsas II LJ, Singh RH, Klein PD, Segal S. *Quantitative assessment of whole body galactose metabolism in galactosemic patients*. Eur J Pediatr. 1997;156(1):S43-S9.

51.Mehta DV, Kabir A, Bhat PJ. *Expression of human inositol monophosphatase suppresses galactose toxicity in Saccharomyces cerevisiae: possible implications in galactosemia*. Biochim Biophys Acta. 1999;1454:217-26.

52.Bhat PJ. *Galactose-1-phosphate is a regulator of inositol monophosphatase: a fact or a fiction?* Medical Hypotheses. 2003;60(1):123-8.

53.Ning C, Fenn PT, Blair IA, Berry GT, Segal S. *Apparent galactose appearance rate in human galactosemia based on plasma [^{13}C]galactose isotopic enrichment*. Mol Genet Metab. 2000;70(4):261–71.

54.Berry GT, Nissim I, Lin Z, Mazur AT, Gibson JB, Segal S. *Endogenous synthesis of galactose in normal men and patients with hereditary galactosaemia* The Lancet. 1995;346(8982):1073-74.

55.Huidekoper HH, Bosch AM, van der Crabben SN, Sauerwein HP, Ackermans MT, Wijburg FA. *Short-term exogenous galactose supplementation does not influence rate of appearance of galactose in patients with classical galactosemia*. Mol Genet Metab. 2005;84(3):265-72.

56.Elsas LJ. *Prenatal diagnosis of galactose-1-phosphate uridyltransferase (GALT)-deficient galactosemia*. Prenat Diagn. 2001;21(4):302-3.

57.Camelo JS, Jr., Fernandes MIM, Maciel LMZ, Scrideli CA, Santos JL, Camargo AS, Jr., Passador CS, Leite PC, Resende DR, Souza LO, Giugliani R, Jorge SM. *Galactosaemia in a Brazilian population: high incidence and cost-benefit analysis*. J Inherit Metab Dis. 2009;32 Suppl 1:S141-9.

58.Shi LY, Suslak L, Rosin I, Searle BM, Desposito F. *Gene dosage studies supporting localization of the structural gene for galactose-1-phosphate uridyl transferase (GALT) to band p13 of chromosome 9*. Am J Med Genet. 1984;19(3):539-43.

59.Ng WG, Xu YK, Kaufman FR, Donnel GN, Wolff J, Allen RJ, Koritala S, Reichardt JK. *Biochemical and molecular studies of 132 patients with galactosemia*. Hum Genet. 1994;94:359-63.

60.Reichardt JK. *Genetic basis of galactosemia*. Hum Mutat. 1992;1(3):190-6.

61.Elsas LJ, Lai K, Saunders CJ, Langley SD. *Functional analysis of the human galactose-1-phosphate uridyltransferase promoter in Duarte and LA variant galactosemia*. Mol Genet Metab. 2001;72(4):297-305.

62.Elsas II LJ, Webb AL, Langley SD. *Characterization of a carbohydrate response element regulating the gene for human galactose-1-phosphate uridyltransferase*. Mol Genet Metab. 2002;76(4):287-96.

63.Leslie ND, Immerrman EB, Flach JE, Florez M, Fridovich-Keil JL, Elsas II LJ. *The human galactose-1-phosphate uridyltransferase gene*. Genomics. 1992;14(2):474-80.

64.Leslie ND, Bai S. *Functional analysis of the mouse galactose-1-phosphate uridyl transferase (GALT) promoter*. Mol Genet Metab. 2001;72(1):31-8.

65.Flanagan JM, Tighe O, O' Neill C, Naughten E, Mayne PD, Croke DT. *Identification of sequence variation in the galactose-1-phosphate uridyl transferase gene by dHPLC*. Mol Genet Metab. 2004;81(2):133-6.

66.Calderon FR, Phansalkar AR, Crockett DK, Miller M, Mao R. *Mutation database for the galactose-1-phosphate uridyltransferase (GALT) gene*. Hum Mutat. 2007;28(10):939-43.

67.Wang BBT, Xu YK, Ng WG, Wong L-J. *Molecular and biochemical basis of galactosemia*. Mol Genet Metab. 1998;63(4):263-9.

68.Tyfield LA. *Galactosaemia and allelic variation at the galactose-1-phosphate uridyltransferase gene: a complex relationship between genotype and phenotype*. Eur J Pediatr. 2000;159 Suppl 3:S204-7.

69.Elsas II LJ, Lai K. *The molecular biology of galactosemia*. Genet Med. 1998;1(1):40-8.

70.Tyfield L, Reichardt JK, Fridovich-Keil JL, Croke DT, Elsas II LJ, Strobl W, Kozak L, Coskun T, Novelli G, Okano Y, Zekanowski C, Shin Y, Boleda MD. *Classical galactosemia and mutations at the galactose-1-phosphate uridyl transferase (GALT) gene*. Hum Mutat. 1999;13(6):417-30.

71.Suzuki M, West C, Beutler E. *Large-scale molecular screening for galactosemia alleles in a pan-ethnic population*. Human Genetics. 2001;109(2):210-5.

72.Flanagan JM, McMahon G, Brendan Chia SH, Fitzpatrick P, Tighe O, O'Neill C, Briones P, Gort L, Kozak L, Magee A, Naughten E, Radomyska B, Schwartz M, Shin JS, Strobl WM, Tyfield LA, Waterham HR, Russell H, Bertorelle G, Reichardt JK, Mayne PD, Croke DT. *The role of human demographic history in determining the distribution and frequency of transferase-deficient galactosaemia mutations*. Heredity (Edinb). 2010;104(2):148-54.

73.Reichardt JK, Packman S, Woo SLC. *Molecular characterization of two galactosemia mutations: correlation of mutations with highly conserved domains in galactose-1-phosphate uridyl transferase*. Am J Hum Genet. 1991;49(4):860-7.

74.Elsas II LJ, Langley S, Paulk EM, Hjelm LN, Dembure PP. *A molecular approach to galactosemia*. Eur J Pediatr. 1995;154(2):S21-S7.

75.Murphy M, McHugh B, Tighe O, Mayne P, O'Neill C, Naughten E, Croke DT. *Genetic basis of transferase-deficient galactosaemia in Ireland and the population history of the Irish Travellers*. Eur J Hum Genet. 1999;7(5):549-54.

76.Milankovics I, Schuler A, Kamory E, Csokay B, Fodor F, Somogyi C, Nemeth K, Fekete G. *Molecular and clinical analysis of patients with classic and Duarte galactosemia in western Hungary*. Wien Klin Wochenschr. 2010;122(3-4):95-102.

77.Coss KP, Doran PP, Owoeye C, Codd MB, Hamid N, Mayne PD, Crushell E, Knerr I, Monavari AA, Treacy EP. *Classical Galactosaemia in Ireland: incidence, complications and outcomes of treatment*. J Inherit Metab Dis. 2013;36(1):21-7.

78.Ashino J, Okano Y, Suyama I, Yamazaki T, Yoshino M, Furuyama JI, Lin HC, Reichardt JK, Isshiki G. *Molecular Characterization of Galactosemia (Type 1) Mutations in Japanese*. Hum Mutat. 1995;6:36-43.

79.Hirokawa H, Okano Y, Asada M, Fujimoto A, Suyama I, Isshiki G. *Molecular basis for phenotypic heterogeneity in galactosaemia: prediction of clinical phenotype from genotype in Japanese patients*. European Journal of Human Genetics. 1999;7:757-64.

80.Coelho AI, Ramos R, Gaspar A, Costa C, Oliveira A, Diogo L, Garcia P, Paiva S, Martins E, Teles EL, Rodrigues E, Cardoso MT, Ferreira E, Sequeira S, Leite M, Silva MJ, de Almeida IT, Vicente JB, Rivera I. *A frequent splicing mutation and novel missense mutations color the updated mutational spectrum of classic galactosemia in Portugal*. J Inherit Metab Dis. 2013;37(1):43-52.

81.Lai K, Willis AC, Elsas II LJ. *The biochemical role of glutamine 188 in human galactose-1-phosphate uridyltransferase*. J Biol Chem. 1999;274(10):6559-66.

82.Fridovich-Keil JL, Jinks-Robertson S. *A yeast expression system for human galactose-1-phosphate uridylyltransferase*. Proc Natl Acad Sci U S A. 1993;90(2):398-402.

83.Robertson A, Singh RH, Guerrero NV, Hundley M, Elsas II LJ. *Outcomes analysis of verbal dyspraxia in classic galactosemia*. Genet Med. 2000;2(2):142-8.

84.Sommer M, Gathof BS, Podskarbi T, Giuliani R, Kleinlein B, Shin YS. *Mutations in the galactose-l-phosphate uridyltransferase gene of two families with mild galactosaemia variants*. J Inher Metab Dis. 1995;18:567-76.

85.Shield JP, Wadsworth EJ, MacDonald A, Stephenson A, Tyfield L, Holton JB, Marlow N. *The relationship of genotype to cognitive outcome in galactosaemia*. Arch Dis Child. 2000;83(3):248-50.

86.Lukac-Bajalo J, Kuzelicki NK, Zitnik IP, Mencej S, Battelino T. *Higher frequency of the galactose-1-phosphate uridyl transferase gene K285N mutation in the Slovenian population*. Clin Biochem. 2007;40(5-6):414-5.

87.Kozák L, Francová H, Fajkusová L, Pijácková A, Štastná JMS, Peškovová K, Martincová O, Krijt J, Bzdúch V. *Mutation analysis of the GALT gene in Czech and Slovak galactosemia populations: identification of six novel mutations, including a stop codon mutation (X380R)*. Hum Mutat. 1999;15(2):1-8.

88.Greber-Platzer S, Guldberg P, Scheibenreiter S, Item C, Schuller E, Patel N, Strobl W. *Molecular Heterogeneity of Classical and Duarte Galactosemia: Mutation Analysis by Denaturing Gradient Gel Electrophoresis*. Hum Mutat. 1997;10:49-57.

89.Zekanowski C, Radomyska B, Bal J. *Molecular characterization of Polish patients with classical galactosaemia*. J Inher Metab Dis. 1999;22:679-82.

90.Manga N, Jenkins T, Jackson H, Whittaker DA, Lane AB. *The molecular basis of transferase galactosaemia in South African negroids*. J Inher Metab Dis. 1999;22:37-42.

91.Lai K, Langley SD, Singh RH, Dembure PP, Hjelm LN, Elsas II LJ. *A prevalent mutation for galactosemia among black Americans*. J Pediatr. 1996;128(1):89-95.

92.Reichardt JK, Levy H, Woo SLC. *Molecular characterization of two galactosemia mutations and one polymorphism: implications for structure-function analysis of human galactose-1-phosphate uridyltransferase*. Biochemistry. 1992;31(24):5430-3.

93.Lai K, Elsas LJ. *Structure-function analyses of a common mutation in blacks with transferase-deficiency galactosemia*. Mol Genet Metab. 2001;74(1-2):264-72.

94.Landt M, Ritter D, Lai K, Benke PJ, Elsas II LJ, Steiner RD. *Black children deficient in galactose 1-phosphate uridyltransferase: correlation of activity and immunoreactive protein in erythrocytes and leukocytes*. J Pediatr. 1997;130(6):972-80.

95.Kelley RI, Harris H, Mellman WJ. *Characterization of Normal and Abnormal Variants of Galactose-l-Phosphate Uridylyltransferase (EC 2.7.7.12) by Isoelectric Focusing*. Hum Genet. 1983;63:274-9.

96.Wedekind JE, Frey PA, Rayment I. *Three-dimensional structure of galactose-1-phosphate uridylyltransferase from Escherichia coli at 1.8 Å resolution*. Biochemistry. 1995;34(35):11049- 61.

97.Thoden JB, Ruzicka FJ, Frey PA, Rayment I, Holden HM. *Structural analysis of the H166G site-directed mutant of galactose-1-phosphate uridylyltransferase complexed with either UDP-glucose or UDP-galactose: detailed description of the nucleotide sugar binding site*. Biochemistry. 1997;36(6):1212-22.

98.Boutron A, Marabotti A, Facchiano A, Cheillan D, Zater M, Oliveira C, Costa C, Labrune P, Brivet M, French Galactosemia Working G. *Mutation spectrum in the French cohort of galactosemic patients and structural simulation of 27 novel missense variations*. Mol Genet Metab. 2012;107(3):438-47.

99.Marabotti A, Facchiano A. *Homology modeling studies on human galactose-1-phosphate uridylyltransferase and on its galactosemia-related mutant Q188R provide an explanation of molecular effects of the mutation on homo-and heterodimers*. J Med Chem. 2005;48(3): 773-9.

100.Field TL, Reznikoff WS, Frey PA. *Galactose-1-phosphate uridylyltransferase: identification of histidine-164 and histidine-166 as critical residues by site-directed mutagenesis*. Biochemistry. 1989;28(5):2094-9.

101.Brenner C. *Hint, Fhit, and GalT: function, structure, evolution, and mechanism of three branches of the histidine triad superfamily of nucleotide hydrolases and transferases*. Biochemistry. 2002;41(29):9003-14.

102.Leslie ND. *Insights into the pathogenesis of galactosemia*. Annu Rev Nutr. 2003;23:59-80.

103.Wong L-J, Frey PA. *Galactose-1-phosphate uridylyltransferase. Rate studies confirming a uridylyl-enzyme intermediate on the catalytic pathway*. Biochemistry. 1974;13(19):3889-94.

104.Wong L-J, Frey PA. *Galactose-1-phosphate uridylyltransferase Isolation of a uridylyl-enzyme intermediate*. J Biol Chem. 1974;249(7):2322-4.

105.Geeganage S, Frey PA. *Galactose-1-phosphate uridylyltransferase: kinetics of formation and reaction of uridylyl-enzyme intermediate in wild-type and specifically mutated uridylyltransferases*. Methods Enzymol. 2002;354:134-48.

106.Wedekind JE, Frey PA, Rayment I. *The structure of nucleotidylated histidine-166 of galactose-1-phosphate uridylyltransferase provides insight into phosphoryl group transfer*. Biochemistry. 1996;35(36):11560-9.

107.Geeganage S, Frey PA. *Significance of metal ions in galactose-1-phosphate uridylyltransferase: an essential structural zinc and a nonessential structural iron*. Biochemistry. 1999;38(40):13398-406.

108.Ruzicka FJ, Wedekind JE, Kim J, Rayment I, Frey PA. *Galactose-1-phosphate uridylyltransferase from Escherichia coli, a zinc and iron metalloenzyme*. Biochemistry. 1995;34(16):5610-7.

109.Wells L, Fridovich-Keil JL. *Biochemical characterization of the S135L allele of galactose-1-phosphate uridylyltransferase associated with galactosaemia*. J Inherit Metab Dis. 1997;20(5):633-42.

110.McCorvie TJ, Gleason TJ, Fridovich-Keil JL, Timson DJ. *Misfolding of galactose 1-phosphate uridylyltransferase can result in type I galactosemia*. Biochim Biophys Acta. 2013;1832(8):1279-93.

111.Elsas II LJ, Dembure PP, Langley S, Paulk EM, Hjelm LN, Fridovich-Keil JL. *A Common Mutation Associated with the Duarte Galactosemia Allele*. Am J Hum Genet. 1994;54:1030-6.

112.Carney AE, Sanders RD, Garza KR, McGaha LA, Bean LJ, Coffee BW, Thomas JW, Cutler DJ, Kurtkaya NL, Fridovich-Keil JL. *Origins, distribution and expression of the Duarte-2 (D2) allele of galactose-1-phosphate uridylyltransferase*. Hum Mol Genet. 2009;18(9):1624-32.

113.Langley SD, Lai K, Dembure PP, Hjelm LN, Elsas II LJ. *Molecular basis for Duarte and Los Angeles variant galactosemia*. Am J Hum Genet. 1997;60(2):366-72.

114.Elsas II LJ, Langley S, Steele E, Evinger J, Fridovich-Keil JL, Brown A, Singh R, Fernhoff P, Hjelm LN, Dembure PP. *Galactosemia: a strategy to identify new biochemical phenotypes and molecular genotypes*. Am J Hum Genet. 1995;56(3):630-9.

115.Ng WG, Bergren WR, Donnel GN. *A new variant of galactose-1-phosphate uridyltransferase in man: the Los Angeles variant*. Ann Hum Genet. 1973;37(1):1-8.

116.Reichardt JK, Berg P. *Conservation of short patches of amino acid sequence amongst proteins with a common function but evolutionarily distinct origins: implications for cloning genes and for structure-function analysis*. Nucleic Acids Research. 1988;16(18):9017-26.

117.Fridovich-Keil JL, Quimby BB, Wells L, Mazur LA, Elsevier JP. *Characterization of the N314D allele of human galactose-1-phosphate uridylyltransferase using a yeast expression system*. Biochem Mol Med. 1995;56(2):121-30.

118.Lai K, Langley SD, Dembure PP, Hjelm LN, Elsas II LJ. *The Duarte allele impairs biostability of galactose-1-phosphate uridyltransferase in human lymphoblasts*. Hum Mutat. 1998;11(1):28-38.

119.Kozak L, Francova H. *Presence of a deletion in the 5' upstream region of the GALT gene in Duarte (D2) alleles*. J Med Genet. 1999;36:576-8.

120.Gathof BS, Sommer M, Podskarbi T, Reichardt JK, Braun A, Gresser U, Shin YS. *Characterization of two stop codon mutations in the galactose-l-phosphate uridyltransferase gene of three male galactosemic patients with severe clinical manifestation*. Hum Genet. 1995;96:721-5.

121.Podskarbi T, Kohlmetz T, Gathof BS, Kleinlein B, Bieger WP, Gresser U, Shin YS. *Molecular characterization of Duarte-1 and Duarte-2 variants of galactose-l-phosphate uridyltransferase*. J Inherit Metab Dis. 1996;19:638-44.

122.Trbušek M, Francova H, Kozak L. *Galactosemia: deletion in the 5' upstream region of the GALT gene reduces promoter efficiency*. Human Genetics. 2001;109(1):117-20.

123.Ridel KR, Leslie ND, Gilbert DL. *An updated review of the long-term neurological effects of galactosemia*. Pediatr Neurol. 2005;33(3):153-61.

124.Burelle Y, Lamoureux M-C, Pèronnet F, Massicotte D, Lavoie C. *Comparison of exogenous glucose, fructose and galactose oxidation during exercise using C-labelling*. British Journal of Nutrition. 2007;96(01):56.

125.Berry GT. *Galactosemia and amenorrhea in the adolescent*. Ann N Y Acad Sci. 2008;1135:112-7.

126.Suchy FJ, Sokol RJ, Balistreri WF. *Inborn Errors of Carbohydrate Metabolism*. In: Suchy FJ, Sokol RJ, Balistreri WF, editors. Liver Disease in Children. 3 ed: Cambridge University Press; 2007. p. 595-625.

127.Ning C, Segal S. *Plasma galactose and galactitol concentration in patients with galactose-1-phosphate uridyltransferase deficiency galactosemia: determination by gas chromatography/mass spectrometry*. Metabolism. 2000;49(11):1460-6.

128.Schwartz V, Goldberg L, Komrower GM, Holzel A. *Some disturbances of erythrocyte metabolism in galactosaemia*. Biochem J. 1955;62(1):34-40.

129.Chen J, Yager C, Reynolds R, Palmieri M, Segal S. *Erythrocyte galactose 1-phosphate quantified by isotope-dilution gas chromatography-mass spectrometry*. Clin Chem. 2002;48(4):604-12.

130.Guerrero NV, Singh RH, Manatunga A, Berry GT, Steiner RD, Elsas Ii LJ. *Risk factors for premature ovarian failure in females with galactosemia*. The Journal of Pediatrics. 2000;137(6):833-41.

131.Berry GT, Hunter JV, Wang Z, Dreha S, Mazur A, Brooks DG, Ning C, Zimmerman RA, Segal S. *In vivo evidence of brain galactitol accumulation in an infant with galactosemia and encephalopathy*. J Pediatr. 2001;138(2):260-2.

132.Jakobs C, Schweitzer S, Dorland B. *Galactitol in galactosemia*. Eur J Pediatr. 1995;154(2):S50-S2.

133.Palmieri M, Mazur A, Berry G, Ning C, Wehrli S, Yager C, Reynolds R, Singh R, Muralidharan K, Langley S, Elsas II LJ, Segal S. *Urine and Plasma Galactitol in Patients With Galaetose-l-Phosphate Uridyltransferase Deficiency Galactosemia*. Metabolism. 1999;48(10):1294-302.

134.Hutchesson ACJ, Murdoch-Davis C, Green A, Preece MA, Allen J, Holton JB, Rylance G. *Biochemical monitoring of treatment for galactosaemia: Biological variability in metabolite concentrations*. J Inherit Metab Dis. 1999;22:139-48.

135.Karas N, Gobec L, Pfeifer V, Mlinar B, Battelino T, Lukac-Bajalo J. *Mutations in galactose-1-phosphate uridyltransferase gene in patients with idiopathic presenile cataract*. J Inherit Metab Dis. 2003;26:699-704.

136.Belman AL, Moshe SL, Zimmerman RD. *Computed Tomographic Demonstration of Cerebral Edema in a Child With Galactosemia*. Pediatrics. 1986;78(4):606-9.

137.Wang Z, Berry G, Dreha S, Segal S, Zimmerman RA. *In Vivo Proton Brain MRS of Galactosemia*.

138.Schweitzer S, Shin Y, Jakobs C, Brodehl J. *Long-term outcome in 134 patients with galactosaemia*. Eur J Pediatr. 1993;152(1):36-43.

139.Yager CT, Chen J, Reynolds R, Segal S. *Galactitol and galactonate in red blood cells of galactosemic patients*. Mol Genet Metab. 2003;80(3):283-9.

140.Karadag N, Zenciroglu A, Eminoglu FT, Dilli D, Karagol BS, Kundak A, Dursun A, Hakan N, Okumus N. *Literature review and outcome of classic galactosemia diagnosed in the neonatal period*. Clin Lab. 2013;59(10):1-8.

141.Ohlsson A, Guthenberg C, von Döbeln U. *Galactosemia screening with low false-positive recall rate: the Swedish experience*. J Inherit Metab Dis Reports-Case and Research Reports. 2012;2:113-7.

142.Waisbren SE, Potter NL, Gordon CM, Green RC, Greenstein P, Gubbels CS, Rubio-Gozalbo E, Schomer D, Welt C, Anastasoaie V, D'Anna K, Gentile J, Guo CY, Hecht L, Jackson R, Jansma BM, Li Y, Lip V, Miller DT, Murray M, Power L, Quinn N, Rohr F, Shen Y, Skinder-Meredith A, Timmers I, Tunick R, Wessel A, Wu BL, Levy H, Elsas II LJ, Berry GT. *The adult galactosemic phenotype*. J Inherit Metab Dis. 2012;35(2):279-86.

143.Shah V, Friedman S, Moore AM, Platt BA, Feigenbaum ASJ. *Selective screening for neonatal galactosemia: an alternative approach*. Acta Paediatr. 2001;90:948-9.

144.Panis B, Bakker JA, Sels JP, Spaapen LJ, van Loon LJ, Rubio-Gozalbo ME. *Untreated classical galactosemia patient with mild phenotype*. Mol Genet Metab. 2006;89(3):277-9.

145.Waggoner DD, Buist NRM, Donnel GN. *Long-term prognosis in galactosaemia: results of a survey of 350 cases*. J Inherit Metab Dis. 1990;13(6):802-18.

146.Berry GT, Walter JH. *Disorders of Galactose Metabolism*. In: Saudubray JM, van den Berghe G, Walter JH, editors. Inborn Metabolic Diseases: Diagnosis and Treatment. 5th ed: Springer; 2012.

147.Kundak AA, Zenciroglu A, Yarali N, Saygili Karagol B, Dursun A, Gokce S, Karadag N, Okumus N. *An unusual presentation of galactosemia: hemophagocytic lymphohistiocytosis*. Turk J Haematol. 2012;29(4):401-4.

148.Zenciroglu A, Ipek MS, Aydin M, Kara A, Okumus N, Kilic M. *Purpura fulminans in a newborn infant with galactosemia*. Eur J Pediatr. 2010;169(7):903-6.

149.Walter JH, Collins JE, Leonard JV. *Recommendations for the management of galactosaemia*. Arch Dis Child. 1999;80:93-6.

150.Henderson MJ, Shapiro L, McCowan C. *Galactosemia detection from phenylketonuria screening*. Clin Chem. 1988;34(1):188-9.

151.Schadewaldt P, Kamalanathan L, Hammen HW, Wendel U. *Stable-isotope dilution analysis of galactose metabolites in human erythrocytes*. Rapid Commun Mass Spectrom. 2003;17:2833-8.

152.Jensen UG, Brandt NJ, Christensen E, Skovby F, Norgaard-Pedersen B, Simonsen H. *Neonatal Screening for Galactosemia by Quantitative Analysis of Hexose Monophosphates Using Tandem Mass Spectrometry: A Retrospective Study*. Clin Chem. 2001;47(8):1364-72.

153.Beutler E. *Galactosemia: Screening and Diagnosis*. Clin Biochem. 1991;24:293-300.

154.Yager C, Wehrli S, Segal S. *Urinary galactitol and galactonate quantified by isotope-dilution gas chromatography-mass spectrometry*. Clin Chim Acta. 2006;366(1-2):216-24.

155.Chen J, Yager CT, Reynolds RA, Segal S. *Identification of galactitol and galactonate in red blood cells by gas chromatography/mass spectrometry*. Clinica Chimica Acta. 2002;322:37-41.

156.Fujimura Y, Kawamura M, Naruse H. *A New Method of Blood Galactose Estimation for Mass Screening of Galactosemia*. Tohoku J exp Med. 1981;133:371-8.

157.Broomfield AA, Brain C, Grunewald S. *Galactosaemia an update*. Paediatrics and Child Health. 2011;21(2):65-70.

158.Fujimoto A, Okano Y, Miyagi T, Isshiki G, Oura T. *Quantitative Beutler test for newborn mass screening of galactosemia using a fluorometric microplate reader*. Clin Chem. 2000;46(6):806-10.

159.Frazier DM, Clemons EH, Kirkman HN. *Minimizing False Positive Diagnoses in Newborn Screening for Galactosemia*. Biochemical Medicine and Metabolic Biology. 1992;48:199-211.

160.Cuthbert C, Klapper H, Elsas L. *Diagnosis of inherited disorders of galactose metabolism*. Curr Protoc Hum Genet. 2008;Chapter 17:Unit 17 5.

161.Lindhout M, Rubio-Gozalbo ME, Bakker JA, Bierau J. *Direct non-radioactive assay of galactose-1-phosphate:uridyltransferase activity using high performance liquid chromatography*. Clin Chim Acta. 2010;411(13-14):980-3.

162.Elsas II LJ. *Galactosemia*. In: Pagon R, MP A, TD B, Dolan CR, Fong C-T, Smith RJH, Stephens K, editors. Gene Reviews. Seattle (WA): University of Washington. 2010.

163.Schweitzer-Krantz S. *Early diagnosis of inherited metabolic disorders towards improving outcome: the controversial issue of galactosaemia*. Eur J Pediatr. 2003;162(1):S50-3.

164.Clarke JTR. *Newborn screening*. A clinical guide to inherited metabolic diseases. 3rd ed: Cambridge University Press; 2005. p. 228-40.

165.Waalen J, Beutler E. *Genetic screening for low-penetrance variants in protein-coding genes*. Annu Rev Genomics Hum Genet. 2009;10:431-50.

166.Walter JH. *Arguments for early screening: a clinician's perspective*. Eur J Pediatr. 2003;162(1):S2-4.

167.Pollitt RJ, Green A, McCabe CJ, Booth A, Cooper NJ, Leonard JV, Nicholl J, Nicholson P, Tunaley JR, Virdi NK. *Neonatal screening for inborn errors of metabolism: cost, yield and outcome*. Health Technol Assess Rep. 1997;1(7):1-202.

168.Waisbren SE, Read CY, Ampola M, Brewster TG, Demmer L, Greenstein R, Ingham CA, Korson M, Msall M, Pueschel S, Seashore MR, Shih VE, Levy H. *Newborn screening compared to clinical identification of biochemical genetic disorders*. J Inher Metab Dis. 2002;25:599-600.

169.Padilla CD. *Towards universal newborn screening in developing countries: obstacles and the way forward*. Ann Acad Med Singap. 2008;37(Suppl 3):6-9.

170.EGS. Galactosaemia in Europe - a comparison. EGS News: European Galactosaemia Society; 2003. p. 40.

171.Hansen TWR, Lie SO. *Galactosemia - to screen or not to screen?* Pediatrics. 1988;81(2):327-8.

172.Lund AM, Hougaard DM, Simonsen H, Andresen BS, Christensen M, Duno M, Skogstrand K, Olsen RK, Jensen UG, Cohen A, Larsen N, Saugmann-Jensen P, Gregersen N, Brandt NJ, Christensen E, Skovby F, Norgaard-Pedersen B. *Biochemical screening of 504,049 newborns in Denmark, the Faroe Islands and Greenland--experience and development of a routine program for expanded newborn screening*. Mol Genet Metab. 2012;107(3):281-93.

173.Vilarinho L, Rocha H, Marcão A, Sousa C, Fonseca H, Bogas M, Osório RV. *Diagnóstico precoce: resultados preliminares do rastreio metabólico alargado*. Acta Paediatr. 2006;37(5):186-91.

174.Vilarinho L, Rocha H, Sousa C, Marcão A, Fonseca H, Bogas M, Osório RV. Four years of expanded newborn screening in Portugal with tandem mass spectrometry. 2010.

175.Bosch AM. *Classic galactosemia: dietary dilemmas*. J Inherit Metab Dis. 2011;34(2):257-60.

176.Thompson S, Netting M. Dietary management of galactosaemia. Australasian Society for Inborn Errors of Metabolism: 2010.

177.Potter NL, Nievergelt Y, Shriberg LD. *Motor and speech disorders in classic galactosemia*. JIMD Rep. 2013;11:31-41.

178.Antshel KM, Epstein IO, Waisbren SE. *Cognitive strengths and weaknesses in children and adolescents homozygous for the galactosemia Q188R mutation: a descriptive study*. Neuropsychology. 2004;18(4):658-64.

179.Bosch AM, Grootenhuis MA, Bakker HD, Heijmans HSA, Wijburg FA, Last BF. *Living With Classical Galactosemia: Health-Related Quality of Life Consequences*. Pediatrics. 2004;113(5):e423-e8.

180.Lo W, Packman S, Nash S, Schmidt S, Ireland S, Diamond I, Ng W, Donnell G. *Curious Neurologic Sequelae in Galactosemia*. Pediatrics. 1984;73(3):309-12.

181.Doyle CM, Channon S, Orlowska D, Lee PJ. *The neuropsychological profile of galactosaemia*. J Inherit Metab Dis. 2010;33(5):603-9.

182.Manis FR, Cohn LB, McBride-Chang C, Wolff J, Kaufman FR. *A longitudinal study of cognitive functioning in patients with classical galactosaemia, including a cohort treated with oral uridine*. J Inher Metab Dis. 1997;20:549-55.

183.Hoffmann B, Wendel U, Schweitzer-Krantz S. *Cross-sectional analysis of speech and cognitive performance in 32 patients with classic galactosemia*. J Inherit Metab Dis. 2011;34(2):421-7.

184.Potter NL, Lazarus JA, Johnson JM, Steiner RD, Shriberg LD. *Correlates of language impairment in children with galactosaemia*. J Inherit Metab Dis. 2008;31(4):524-32.

185.Nelson MD, Wolff JA, Cross CA, Donnel GN, Kaufman FR. *Galactosemia: evaluation with MR imaging*. Radiology. 1992;184(1):255-61.

186.Kaufman FR, McBride-Chang C, Manis FR, Wolff J, Nelson L. *Cognitive functioning, neurologic status and brain imaging in classical galactosemia*. Eur J Pediatr. 1995;154(2):S2-S5.

187.Petry KG, Reichardt JK. *The fundamental importance of human galactose metabolism: lessons from genetics and biochemistry*. Trends Genet. 1998;14(3):98-102.

188.Hughes J, Ryan S, Lambert D, Geoghegan O, Clark A, Rogers Y, Hendroff U, Monavari A, Twomey E, Treacy EP. *Outcomes of siblings with classical galactosemia*. J Pediatr. 2009;154(5):721-6.

189.Coss KP, Byrne JC, Coman DJ, Adamczyk B, Abrahams JL, Saldova R, Brown AY, Walsh O, Hendroff U, Carolan C, Rudd PM, Treacy EP. *IgG N-glycans as potential biomarkers for determining galactose tolerance in Classical Galactosaemia*. Mol Genet Metab. 2012;105(2):212-20.

190.Bhat M, Haase C, Lee PJ. *Social outcome in treated individuals with inherited metabolic disorders: UK study*. J Inherit Metab Dis. 2005;28(6):825-30.

191.Jumbo-Lucioni PP, Garber K, Kiel J, Baric I, Berry GT, Bosch A, Burlina A, Chiesa A, Pico ML, Estrada SC, Henderson H, Leslie N, Longo N, Morris AA, Ramirez-Farias C, Schweitzer-Krantz S, Silao CL, Vela-Amieva M, Waisbren S, Fridovich-Keil JL. *Diversity of approaches to classic galactosemia around the world: a comparison of diagnosis, intervention, and outcomes*. J Inherit Metab Dis. 2012;35(6):1037-49.

192.Bosch AM, Maurice-Stam H, Wijburg FA, Grootenhuis MA. *Remarkable differences: the course of life of young adults with galactosaemia and PKU*. J Inherit Metab Dis. 2009;32(6):706-12.

193.Gubbels CS, Maurice-Stam H, Berry GT, Bosch AM, Waisbren S, Rubio-Gozalbo ME, Grootenhuis MA. *Psychosocial developmental milestones in men with classic galactosemia*. J Inherit Metab Dis. 2011;34(2):415-9.

194.Brazeal TJ, Farmer JE. *Natural course and treatment of neuropsychological deficits in a child with early-treated galactosemia*. Child Neuropsychology. 1999;5(3):197-209.

195.Hoffmann B, Dragano N, Schweitzer-Krantz S. *Living situation, occupation and health-related quality of life in adult patients with classic galactosemia*. J Inherit Metab Dis. 2012;35(6):1051-8.

196.Panis B, Gerver WJ, Rubio-Gozalbo ME. *Growth in treated classical galactosemia patients*. Eur J Pediatr. 2007;166(5):443-6.

197.Kaufman FR, Loro ML, Azen C, Wenz E, Gilsanz V. *Effect of hypogonadism and deficient calcium intake on bone density in patients with galactosemia*. The Journal of Pediatrics. 1993;123(3):365-70.

198.Rubio-Gozalbo ME. *Bone mineral density in patients with classic galactosaemia*. Archives of Disease in Childhood. 2002;87(1):57-60.

199.Panis B, Forget PP, van Kroonenburgh MJ, Vermeer C, Menheere PP, Nieman FH, Rubio-Gozalbo ME. *Bone metabolism in galactosemia*. Bone. 2004;35(4):982-7.

200.Batey LA, Welt CK, Rohr F, Wessel A, Anastasoaie V, Feldman HA, Guo CY, Rubio-Gozalbo E, Berry G, Gordon CM. *Skeletal health in adult patients with classic galactosemia*. Osteoporos Int. 2013;24(2):501-9.

201.Panis B, Forget PP, Nieman FH, van Kroonenburgh MJ, Rubio-Gozalbo ME. *Body composition in children with galactosaemia*. J Inherit Metab Dis. 2005;28(6):931-7.

202.Panis B, Vermeer C, van Kroonenburgh MJ, Nieman FH, Menheere PP, Spaapen LJ, Rubio-Gozalbo ME. *Effect of calcium, vitamins K1 and D3 on bone in galactosemia*. Bone. 2006;39(5):1123-9.

203.Gajewska J, Ambroszkiewicz J, Radomyska B, Chelchowska M, Oltarzewski M, Laskowska-Klita T, Milanowski A. *Serum markers of bone turnover in children and adolescents with classic galactosemia*. Adv Med Sci. 2008;53(2):214-20.

204.Fridovich-Keil JL, Gubbels CS, Spencer JB, Sanders RD, Land JA, Rubio-Gozalbo E. *Ovarian function in girls and women with GALT-deficiency galactosemia*. J Inherit Metab Dis. 2011;34(2):357-66.

205.Rubio-Gozalbo ME, Gubbels CS, Bakker JA, Menheere PP, Wodzig WK, Land JA. *Gonadal function in male and female patients with classic galactosemia*. Hum Reprod Update. 2010;16(2):177-88.

206.Gubbels CS, Land JA, Evers JL, Bierau J, Menheere PP, Robben SG, Rubio-Gozalbo ME. *Primary ovarian insufficiency in classic galactosemia: role of FSH dysfunction and timing of the lesion*. J Inherit Metab Dis. 2013;36(1):29-34.

207.Gubbels CS, Thomas CM, Wodzig WK, Olthaar AJ, Jaeken J, Sweep FC, Rubio-Gozalbo ME. *FSH isoform pattern in classic galactosemia*. J Inherit Metab Dis. 2011;34(2):387-90.

208.Knerr I, Coss KP, Doran PP, Hughes J, Wareham N, Burling K, Treacy EP. *Leptin levels in children and adults with classic galactosaemia*. JIMD Rep. 2013;9:125-31.

209.Sanders RD, Spencer JB, Epstein MP, Pollak SV, Vardhana PA, Lustbader JW, Fridovich-Keil JL. *Biomarkers of ovarian function in girls and women with classic galactosemia*. Fertil Steril. 2009;92(1):344-51.

210.Yamaguchi S, Mizushima Y, Fukushi M, Shimizu Y, Kikuchi Y, Takasugi N. *Microassay system for newborn screening for phenylketonuria, maple syrup urine disease, homocystinuria, histidinemia and galactosemia with use of a fluorometric microplate reader*. Screening. 1992;1(1):49-62.

211.Lai KW, Cheng LY, Cheung AL, O WS. *Inhibitor of apoptosis proteins and ovarian dysfunction in galactosemic rats*. Cell Tissue Res. 2003;311(3):417-25.

212.Gubbels CS, Welt CK, Dumoulin JC, Robben SG, Gordon CM, Dunselman GA, Rubio-Gozalbo ME, Berry GT. *The male reproductive system in classic galactosemia: cryptorchidism and low semen volume*. J Inherit Metab Dis. 2013;36(5):779-86.

213.Rubio-Gozalbo ME, Panis B, Zimmermann LJ, Spaapen LJ, Menheere PP. *The endocrine system in treated patients with classical galactosemia*. Mol Genet Metab. 2006;89(4):316-22.

214.Zlatunich CO, Packman S. *Galactosaemia: Early treatment with an elemental formula*. J Inherit Metab Dis. 2005;28:163-8.

215.Ficicioglu C, Hussa C, Yager C, Segal S. *Effect of galactose free formula on galactose-1-phosphate in two infants with classical galactosemia*. Eur J Pediatr. 2008;167(5):595-6.

216.Bhatia J, Greer F, American Academy of Pediatrics Committee on N. *Use of soy protein-based formulas in infant feeding*. Pediatrics. 2008;121(5):1062-8.

217.Gross KC, Acosta PB. *Fruits and vegetables are a source of galactose: implications in planning the diets of patients with galactosaemia*. J Inherit Metab Dis. 1991;14(2):253-8.

218.Acosta PB, Gross KC. *Hidden sources of galactose in the environment*. Eur J Pediatr. 1995;154(2):S87-S92.

219.Bosch AM, Bakker HD, Wenniger-Prick LJ, Wanders RJ, Wijburg FA. *High tolerance for oral galactose in classical galactosaemia: dietary implications*. Arch Dis Child. 2004;89(11):1034-6.

220.Berry GT, Palmieri M, Gross KC, Acosta PB, Henstenburg JA, Mazur A, Reynolds R, Segal S. *The effect of dietary fruits and vegetables on urinary galactitol excretion in galactose-1-phosphate uridyltransferase deficiency*. J Inherit Metab Dis. 1993;16(1):91-100.

221.Portnoi PA, MacDonald A. *Determination of the lactose and galactose content of cheese for use in the galactosaemia diet*. J Hum Nutr Diet. 2009;22(5):400-8.

222.Tang M, Odejinmi SI, Vankayalapati H, Wierenga KJ, Lai K. *Innovative therapy for classic galactosemia - tale of two HTS*. Mol Genet Metab. 2012;105(1):44-55.

223.Saudubray JM, Nassogne MC, de Lonlay P, Touati G. *Clinical approach to inherited metabolic disorders in neonates: an overview*. Semin Neonatol. 2002;7(1):3-15.

224.Saudubray JM, Sedel F, Walter JH. *Clinical approach to treatable inborn metabolic diseases: an introduction*. J Inherit Metab Dis. 2006;29(2-3):261-74.

225.Berry GT, Elsas LJ. *Introduction to the Maastricht workshop: lessons from the past and new directions in galactosemia*. J Inherit Metab Dis. 2011;34(2):249-55.

226.Ng DT, Xu YK, Kaufman FR, Donnel GN. *Deficit of Uridine Diphosphate Galactose in Galactosaemia*. J Inherit Metab Dis. 1989;12:257-66.

227.Bosch A, Bakker HD, Van Gennip AH, Van Kempen JV, Wanders RJ, Wijburg FA. *Clinical features of galactokinase deficiency: A review of the literature*. J Inherit Metab Dis. 2002;25:629-34.

228.Wierenga KJ, Lai K, Buchwald P, Tang M. *High-throughput screening for human galactokinase inhibitors*. J Biomol Screen. 2008;13(5):415-23.

229.Berry GT. *Is prenatal myo-inositol deficiency a mechanism of CNS injury in galactosemia?* J Inherit Metab Dis. 2011;34(2):345-55.

230.Gregersen N, Bross P, Jorgensen MM, Corydon TJ, Andersen BS. *Defective folding and rapid degradation of mutant proteins is a common disease mechanism in genetic disorders*. J Inherit Metab Dis. 2000;23:441-7.

231.Bross P, Andresen BS, Corydon TJ, Gregersen N. *Protein misfolding and degradation in genetic diseases (updated version)*. Encyclopedia of the Human Genome Nature Publishing Group. 2011.

232.Waters PJ. *Degradation of Mutant Proteins, Underlying "Loss of Function" Phenotypes, Plays a Major Role in Genetic Disease*. Curr issues Mol Biol. 2001;3(3):57-65.

233.Gregersen N, Bross P, Jorgensen MM. *Protein folding and misfolding: the role of cellular protein quality control systems in inherited disorders*. In: Valle D, Beaudet AL, Vogelstein B, Kinzler KW,

Antonarakis SE, Ballabio A, editors. The Online Metabolic and Molecular Bases of Inherited Disease (OMMBID). General Themes: Mc-Graw Hill; 2005.

234.Faustino NA, Cooper TA. *Pre-mRNA splicing and human disease*. Genes Dev. 2003;17(4):419-37.

235.Gregersen N. *Protein misfolding disorders: pathogenesis and intervention*. J Inherit Metab Dis. 2006;29(2-3):456-70.

236.Kim YE, Hipp MS, Bracher A, Hayer-Hartl M, Hartl FU. *Molecular chaperone functions in protein folding and proteostasis*. Annu Rev Biochem. 2013;82:323-55.

237.Hartl FU, Bracher A, Hayer-Hartl M. *Molecular chaperones in protein folding and proteostasis*. Nature. 2011;475(7356):324-32.

238.Roth DM, Balch WE. *Modeling general proteostasis: proteome balance in health and disease*. Curr Opin Cell Biol. 2011;23(2):126-34.

239.Stefani M. *Protein misfolding and aggregation: new examples in medicine and biology of the dark side of the protein world*. Biochim Biophys Acta. 2004;1739(1):5-25.

240.Ursini F, Davies KJA, Maiorino M, Parasassi T, Sevanian A. *Atherosclerosis: another protein misfolding disease?* Trends Mol Med. 2002;8(8):370-4.

241.Iram SH. *Protein misfolding in health and disease*. Queen's Health Sciences Journal. 2013;275:8.

242.Leandro P, Gomes CM. *Protein misfolding in conformational disorders: rescue of folding defects and chemical chaperoning*. Mini-Rev Med Chem. 2008;8(9):901-11.

243.Ellis RJ, Pinheiro TJ. *Danger — misfolding proteins*. Nature. 2002;416:483-4.

244.Gregersen N, Bross P, Vang S, Christensen JH. *Protein misfolding and human disease*. Annu Rev Genomics Hum Genet. 2006;7:103-24.

245.Dobson CM. *Protein folding and misfolding*. Nature. 2003;426:884-90.

246.Balch WE, Morimoto RI, Dillin A, Kelly JW. *Adapting Proteostasis for Disease Intervention*. Science. 2008;319:916-9.

247.Luheshi LM, Dobson CM. *Bridging the gap: from protein misfolding to protein misfolding diseases*. FEBS Lett. 2009;583(16):2581-6.

248.Stefl S, Nishi H, Petukh M, Panchenko AR, Alexov E. *Molecular mechanisms of disease-causing missense mutations*. J Mol Biol. 2013;425(21):3919-36.

249.Muntau AC, Gersting SW. *Phenylketonuria as a model for protein misfolding diseases and for the development of next generation orphan drugs for patients with inborn errors of metabolism*. J Inherit Metab Dis. 2010;33(6):649-58.

250.Derewenda ZS, Lee L, Derewenda U. *The occurrence of C - H · · · O hydrogen bonds in proteins*. J Mol Biol. 1995;252(2):248-62.

251.Pey AL, Stricher F, Serrano L, Martinez A. *Predicted effects of missense mutations on native-state stability account for phenotypic outcome in phenylketonuria, a paradigm of misfolding diseases*. Am J Hum Genet. 2007;81(5):1006-24.

252.Tokuriki N, Stricher F, Serrano L, Tawfik DS. *How protein stability and new functions trade off*. PLoS Comput Biol. 2008;4(2):e1000002.

253.Schlessinger A, Rost B. *Protein flexibility and rigidity predicted from sequence*. Proteins. 2005;61(1):115-26.

254.Capriotti E, Fariselli P, Casadio R. *I-Mutant2.0: predicting stability changes upon mutation from the protein sequence or structure*. Nucleic Acids Res. 2005;33(Web Server issue):W306-10.

255.Cheng J, Randall A, Baldi P. *Prediction of protein stability changes for single-site mutations using support vector machines*. Proteins. 2006;62(4):1125-32.

256.Teng S, Srivastava AK, Wang L. *Sequence feature-based prediction of protein stability changes upon amino acid substitutions*. BMC Genomics. 2010;11 Suppl 2:S5.

257.Dehouck Y, Kwasigroch JM, Gilis D, Rooman M. *PoPMuSiC 2.1: a web server for the estimation of protein stability changes upon mutation and sequence optimality*. BMC Bioinformatics. 2011;12:151.

258.Gonnelli G, Rooman M, Dehouck Y. *Structure-based mutant stability predictions on proteins of unknown structure*. J Biotechnol. 2012;161(3):287-93.

259.Ng PC, Henikoff S. *SIFT: predicting amino acid changes that affect protein function*. Nucleic Acids Res. 2003;31(13):3812-4.

260.Ramensky V, Bork P, Sunyaev S. *Human non-synonymous SNPs: server and survey*. Nucleic Acids Res. 2002;30(17):3894-900.

261.Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, Sunyaev SR. *A method and server for predicting damaging missense mutations*. Nat Methods. 2010;7(4):248-9.

262.Parthiban V, Gromiha MM, Schomburg D. *CUPSAT: prediction of protein stability upon point mutations*. Nucleic Acids Res. 2006;34(Web Server issue):W239-42.

263.Worth CL, Bickerton GRJ, Schreyer A, Forman JR, Cheng TMK, Lee S, Gong S, Burke DF, Blundell TL. *A structural bioinformatics approach to the analysis of nonsynonymous single nucleotide polymorphisms (nsSNPs) and their relation to disease*. J Bioinform Comput Biol. 2007;5(6):1297-318.

264.Worth CL, Preissner R, Blundell TL. *SDM--a server for predicting effects of mutations on protein stability and malfunction*. Nucleic Acids Res. 2011;39(Web Server issue):W215-22.

265.Schymkowitz J, Borg J, Stricher F, Nys R, Rousseau F, Serrano L. *The FoldX web server: an online force field*. Nucleic Acids Res. 2005;33(Web Server issue):W382-8.

266.Carrell RW, Lomas DA. *Conformational disease*. The Lancet. 1997;350(9071):134-8.

267.Beaulieu CL, Samuels ME, Ekins S, McMaster CR, Edwards AM, Krainer AR, Hicks GG, Frey BJ, Boycott KM, Mackenzie AE. *A generalizable pre-clinical research approach for orphan disease therapy*. Orphanet J Rare Dis. 2012;7(39):1-11.

268.Soto C. *Protein misfolding and disease; protein refolding and therapy*. FEBS Letters. 2001;498(2):204-7.

269.Chaudhuri TK, Paul S. *Protein-misfolding diseases and chaperone-based therapeutic approaches*. FEBS J. 2006;273(7):1331-49.

270.Pey AL. *The interplay between protein stability and dynamics in conformational diseases: The case of hPGK1 deficiency*. Biochim Biophys Acta. 2013;1834(12):2502-11.

271.Hastings ML, Krainer AR. *Pre-mRNA splicing in the new millennium*. Curr Opin Cell Biol. 2001;13(3):302-9.

272.Wang Z, Burge CB. *Splicing regulation: from a parts list of regulatory elements to an integrated splicing code*. RNA. 2008;14(5):802-13.

273.Cartegni L, Chew SL, Krainer AR. *Listening to silence and understanding nonsense: exonic mutations that affect splicing*. Nat Rev Genet. 2002;3(4):285-98.

274.Schor IE, Gomez Acuna LI, Kornblihtt AR. *Coupling Between Transcription and Alternative Splicing*. In: Wu JW, editor. RNA and Cancer. 158: Springer Berlin Heidelberg; 2013. p. 1-24.

275.Havens MA, Duelli DM, Hastings ML. *Targeting RNA splicing for disease therapy*. Wiley Interdiscip Rev RNA. 2013;4(3):247-66.

276.Buratti E, Baralle M, Baralle FE. *Defective splicing, disease and therapy: searching for master checkpoints in exon definition*. Nucleic Acids Res. 2006;34(12):3494-510.

277.Krawczak M, Thomas NS, Hundrieser B, Mort M, Wittig M, Hampe J, Cooper DN. *Single base-pair substitutions in exon-intron junctions of human genes: nature, distribution, and consequences for mRNA splicing*. Hum Mutat. 2007;28(2):150-8.

278.Busch A, Hertel KJ. *Evolution of SR protein and hnRNP splicing regulatory factors*. Wiley Interdiscip Rev RNA. 2012;3(1):1-12.

279.Pagani F, Baralle FE. *Genomic variants in exons and introns: identifying the splicing spoilers*. Genetics. 2004;5:389-96.

280.Cooper TA. *Use of minigene systems to dissect alternative splicing elements*. Methods. 2005;37(4):331-40.

281.Zhu J, Mayeda A, Krainer AR. *Exon identity established through differential antagonism between exonic splicing silencer-bound hnRNP A1 and enhancer-bound SR proteins*. Mol Cell. 2001;8(6):1351-61.

282.Homolova K, Zavadakova P, Doktor TK, Schroeder LD, Kozich V, Andresen BS. *The deep intronic c.903+469T>C mutation in the MTRR gene creates an SF2/ASF binding exonic splicing enhancer, which leads to pseudoexon activation and causes the cblE type of homocystinuria*. Hum Mutat. 2010;31(4):437-44.

283.Black DL. *Mechanisms of alternative pre-messenger RNA splicing*. Annu Rev Biochem. 2003;72:291-336.

284.Buratti E, Baralle FE. *Influence of RNA secondary structure on the pre-mRNA splicing process*. Mol Cell Biol. 2004;24(24):10505-14.

285.Roscigno RF, Weiner M, Garcia-Blanco MA. *A mutational analysis of the polypyrimidine tract of introns*. J Biol Chem. 1993;268(15):11222-9.

286.Fica SM, Small EC, Mefford M, Staley JP. *Mechanistic insights into mammalian pre-mRNA splicing*. In: Wu JY, editor. Posttranscriptional gene regulation: RNA processing in eukaryotes: Wiley-Blackwell; 2013. p. 133-61.

287.Hoskins AA, Moore MJ. *The spliceosome: a flexible, reversible macromolecular machine*. Trends Biochem Sci. 2012;37(5):179-88.

288.McManus CJ, Graveley BR. *RNA structure and the mechanisms of alternative splicing*. Curr Opin Genet Dev. 2011;21(4):373-9.

289.Bonnal S, Valcarcel J. *RNAtomy of the Spliceosome's heart*. EMBO J. 2013;32(21):2785-7.

290.Kornblihtt AR, Schor IE, Allo M, Dujardin G, Petrillo E, Munoz MJ. *Alternative splicing: a pivotal step between eukaryotic transcription and translation*. Nat Rev Mol Cell Biol. 2013;14(3):153-65.

291.Wang GS, Cooper TA. *Splicing in disease: disruption of the splicing code and the decoding machinery*. Nat Rev Genet. 2007;8(10):749-61.

292.Zhang C, Li WH, Krainer AR, Zhang MQ. *RNA landscape of evolution for optimal exon and intron discrimination*. Proc Natl Acad Sci U S A. 2008;105(15):5797-802.

293.Barash Y, Calarco JA, Gao W, Pan Q, Wang X, Shai O, Blencowe BJ, Frey BJ. *Deciphering the splicing code*. Nature. 2010;465(7294):53-9.

294.Perez B, Rodriguez-Pascau L, Vilageliu L, Grinberg D, Ugarte M, Desviat LR. *Present and future of antisense therapy for splicing modulation in inherited metabolic disease*. J Inherit Metab Dis. 2010;33(4):397-403.

295.Baralle D, Baralle M. *Splicing in action: assessing disease causing sequence changes*. J Med Genet. 2005;42(10):737-48.

296.Krawczak M, Ball EV, Fenton I, Stenson PD, Abeysinghe S, Thomas N, Cooper DN. *Human gene mutation database—a biomedical information and research resource*. Hum Mutat. 2000;15(1):45-51.

297.Hegele A, Kamburov A, Grossmann A, Sourlis C, Wowro S, Weimann M, Will CL, Pena V, Luhrmann R, Stelzl U. *Dynamic protein-protein interaction wiring of the human spliceosome*. Mol Cell. 2012;45(4):567-80.

298.Maquat LE. *Nonsense-mediated mRNA decay: splicing, translation and mRNP dynamics*. Nat Rev Mol Cell Biol. 2004;5(2):89-99.

299.Lareau LF, Brooks AN, Soergel DAW, Meng Q, Brenner S. *The coupling of alternative splicing and nonsense-mediated mRNA decay." Advances in experimental medicine and biology 623 (2008): 190.* In: Blencowe BJ, Graveley BR, editors. Alternative splicing in the postgenomic era. Advances in Experimental Medicine and Biology. New York: Springer; 2007. p. 190-211.

300.Xie J. *Making use of aberrant and nonsense: aberrant splicing and nonsense-mediated decay as targets for personalized medicine*. Int J Genomic Med. 2013;1(105):2332-0672.

301.Singh G, Cooper T. *Minigene reporter for identification and analysis of cis elements and trans factors affecting pre-mRNA splicing*. BioTechniques. 2006;41(2):177-81.

302.Baralle D, Lucassen A, Buratti E. *Missed threads*. EMBO reports. 2009;10(8):810-6.

303.Houdayer C, Caux-Moncoutier V, Krieger S, Barrois M, Bonnet F, Bourdon V, Bronner M, Buisson M, Coulet F, Gaildrat P, Lefol C, Leone M, Mazoyer S, Muller D, Remenieras A, Revillion F, Rouleau E, Sokolowska J, Vert JP, Lidereau R, Soubrier F, Sobol H, Sevenet N, Bressac-de Paillerets B, Hardouin A, Tosi M, Sinilnikova OM, Stoppa-Lyonnet D. *Guidelines for splicing analysis in molecular diagnosis derived from a set of 327 combined in silico/in vitro studies on BRCA1 and BRCA2 variants*. Hum Mutat. 2012;33(8):1228-38.

304.Houdayer C, Dehainault C, Mattler C, Michaux D, Caux-Moncoutier V, Pages-Berhouet S, d'Enghien CD, Lauge A, Castera L, Gauthier-Villars M, Stoppa-Lyonnet D. *Evaluation of in silico splice tools for decision-making in molecular diagnosis*. Hum Mutat. 2008;29(7):975-82.

305.Nielsen H. *RNA: methods and protocols*. Nielsen H, editor. Copenhagen Humana Press; 2011.

306.Reese MG, Eeckman FH, Kulp D, Haussler D. *Improved splice site detection in Genie*. Journal of computational biology. 1997;4(3):311-23.

307.Kapustin Y, Chan E, Sarkar R, Wong F, Vorechovsky I, Winston RM, Tatusova T, Dibb NJ. *Cryptic splice sites and split genes*. Nucleic Acids Res. 2011;39(14):5837-44.

308.Cartegni L. *ESEfinder: a web resource to identify exonic splicing enhancers*. Nucleic Acids Research. 2003;31(13):3568-71.

309.Fairbrother WG, Yeh RF, Sharp PA, Burge CB. *Predictive identification of exonic splicing enhancers in human genes*. Science. 2002;297(5583):1007-13.

310.Zhang XH, Chasin LA. *Computational definition of sequence motifs governing constitutive exon splicing*. Genes Dev. 2004;18(11):1241-50.

311.Geary SR. *Antisense oligonucleotide pharmacokinetics and metabolism*. 2009:381-91.

312.Desmet FO, Hamroun D, Lalande M, Collod-Beroud G, Claustres M, Beroud C. *Human Splicing Finder: an online bioinformatics tool to predict splicing signals*. Nucleic Acids Res. 2009;37(9):e67.

313.Chan JH, Lim S, Wong WS. *Antisense oligonucleotides: from design to therapeutic application*. Clinical and experimental pharmacology and physiology. 2006;33(5-6):533-40.

314.Zuker M. *Mfold web server for nucleic acid folding and hybridization prediction*. Nucleic Acids Research. 2003;31(13):3406-15.

315.Piva F, Giulietti M, Nocchi L, Principato G. *SpliceAid: a database of experimental RNA target motifs bound by splicing proteins in humans*. Bioinformatics. 2009;25(9):1211-3.

316.Piva F, Giulietti M, Burini AB, Principato G. *SpliceAid 2: a database of human splicing factors expression data and RNA target motifs*. Hum Mutat. 2012;33(1):81-5.

317.Aartsma-Rus A, van Vliet L, Hirschi M, Janson AA, Heemskerk H, de Winter CL, de Kimpe S, van Deutekom JC, t Hoen PA, van Ommen GJ. *Guidelines for antisense oligonucleotide design and insight into splice-modulating mechanisms*. Mol Ther. 2009;17(3):548-53.

318.Chen C, Kobayashi R, Helfman DM. *Binding of hnRNP H to an exonic splicing silencer is involved in the regulation of alternative splicing of the rat beta-tropomyosin gene*. Genes Dev. 1999;13:593-606.

319.Raponi M, Baralle D. *Alternative splicing: good and bad effects of translationally silent substitutions*. FEBS J. 2010;277(4):836-40.

320.Bennett CF, Swayze EE. *RNA targeting therapeutics: molecular mechanisms of antisense oligonucleotides as a therapeutic platform*. Annu Rev Pharmacol Toxicol. 2010;50:259-93.

321.Southwell AL, Skotte NH, Bennett CF, Hayden MR. *Antisense oligonucleotide therapeutics for inherited neurodegenerative diseases*. Trends Mol Med. 2012;18(11):634-43.

322.Kunze D, Kraemer K, Fuessel S. *Antisense oligonucleotides: insights from preclinical studies and clinical trials*. In: Erdmann VA, Barciszewski J, editors. RNA Technologies and Their Applications: Springer Berlin Heidelberg; 2010. p. 285-303.

323.O'Connor TP, Crystal RG. *Genetic medicines: treatment strategies for hereditary disorders*. Nat Rev Genet. 2006;7(4):261-76.

324.Dias N, Stein CA. *Antisense oligonucleotides: basic concepts and mechanisms*. Mol Cancer Ther. 2002;1(5):347-55.

325.Hammond SM, Wood MJ. *Genetic therapies for RNA mis-splicing diseases*. Trends Genet. 2011;27(5):196-205.

326.Garcia-Blanco MA, Baraniak AP, Lasda EL. *Alternative splicing in disease and therapy*. Nat Biotechnol. 2004;22(5):535-46.

327.Dominski Z, Kole R. *Restoration of correct splicing in thalassemic pre-mRNA by antisense oligonucleotides*. Proc Natl Acad Sci U S A. 1993;90:8673-7.

328.IDT. Designing antisense oligonucleotides. In: Technologies ID, editor. 2010.

329.Kurreck J. *Antisense technologies. Improvement through novel chemical modifications*. European Journal of Biochemistry. 2003;270(8):1628-44.

330.Ekker SC, Larson JD. *Morphant technology in model developmental systems*. Genesis. 2001;30(3):89- 93.

331.Yokota T, Lu QL, Partridge T, Kobayashi M, Nakamura A, Takeda S, Hoffman E. *Efficacy of systemic morpholino exon-skipping in Duchenne dystrophy dogs*. Ann Neurol. 2009;65(6):667-76.

332.Brasil S, Viecelli HM, Meili D, Rassi A, Desviat LR, Perez B, Ugarte M, Thony B. *Pseudoexon exclusion by antisense therapy in 6-pyruvoyl-tetrahydropterin synthase deficiency*. Hum Mutat. 2011;32(9):1019-27.

333.Braasch DA, Corey DR. *Novel Antisense and Peptide Nucleic Acid Strategies for Controlling Gene Expression*. Biochemistry. 2002;41(14):4503-10.

334.Mercatante DR, Sazani P, Kole R. *Modification of alternative splicing by antisense oligonucleotides as a potential chemotherapy for cancer and other diseases*. Curr Cancer Drug Targets. 2001;1(3):211-30.

335.Sazani P, Kole R. *Therapeutic potential of antisense oligonucleotides as modulators of alternative splicing*. Journal of Clinical Investigation. 2003;112(4):481-6.

336.Kaur H, Babu R, Maiti S. *Perspectives on chemistry and therapeutic applications of locked nucleic acid (LNA)*. Chem Rev. 2007;107(11):4672-97.

337.Vester B, Wengel J. *LNA (Locked Nucleic Acid): High-Affinity Targeting of Complementary RNA and DNA*. Biochemistry. 2004;43(42):13233-41.

338.Petersen M, Wengel J. *LNA: a versatile tool for therapeutics and genomics*. Trends in Biotechnology. 2003;21(2):74-81.

339.Jepsen JS, Sorensen M, Wengel J. *Locked nucleic acid: a potent nucleic acid analog in therapeutics and biotechnology*. Oligonucleotides. 2004;14(2):130-46.

340.Roberts J, Palma E, Sazani P, Orum H, Cho M, Kole R. *Efficient and persistent splice switching by systemically delivered LNA oligonucleotides in mice*. Mol Ther. 2006;14(4):471-5.

341.Straarup EM, Fisker N, Hedtjarn M, Lindholm MW, Rosenbohm C, Aarup V, Hansen HF, Orum H, Hansen JB, Koch T. *Short locked nucleic acid antisense oligonucleotides potently reduce apolipoprotein B mRNA and serum cholesterol in mice and non-human primates*. Nucleic Acids Res. 2010;38(20):7100-11.

342.Moschos SA, Frick M, Taylor B, Turnpenny P, Graves H, Spink KG, Brady K, Lamb D, Collins D, Rockel TD, Weber M, Lazari O, Perez-Tosar L, Fancy SA, Lapthorn C, Green MX, Evans S, Selby M, Jones G, Jones L, Kearney S, Mechiche H, Gikunju D, Subramanian R, Uhlmann E, Jurk M, Vollmer J, Ciaramella G, Yeadon M. *Uptake, efficacy, and systemic distribution of naked, inhaled short interfering RNA (siRNA) and locked nucleic acid (LNA) antisense*. Mol Ther. 2011;19(12):2163-8.

343.Aartsma-Rus A, Kaman WE, Bremmer-Bout M, Janson AAM, Den Dunnen JT, Van Ommen GB, Van Deutekom JCT. *Comparative analysis of antisense oligonucleotide analogs for targeted DMD exon 46 skipping in muscle cells*. Gene Ther. 2004;11(18):1391-8.

344.Lee J, Yokota T. *Antisense therapy in neurology*. J Pers Med. 2013;3(3):144-76.

345.Mackenzie A, Boycott KM. *The future is now for rare genetic diseases*. CMAJ. 2012;184(14):1603.

346.Rosenblum D, Peer D. *Omics-based nanomedicine: The future of personalized oncology*. Cancer Lett. 2013.

347.Treacy EP. *Treatment of Genetic Disease*. In: Valle D, Beaudet AL, Vogelstein B, Kinzler KW, Antonarakis SE, Ballabio A, editors. The Online Metabolic and Molecular Bases of Inherited Disease: Mc-Graw Hill; 2008. p. 1-13.

348.Ginsburg GS, McCarthy JJ. *Personalized medicine: revolutionizing drug discovery and patient care*. Trends in Biotechnology. 2001;19(12):491-6.

CHAPTER 2

A frequent splicing mutation and novel missense mutations color the updated mutational spectrum of classic galactosemia in Portugal

Ana I. Coelho[1], Ruben Ramos[1], Ana Gaspar[3], Cláudia Costa[3], Anabela Oliveira[4], Luísa Diogo[5], Paula Garcia[5], Sandra Paiva[5], Esmeralda Martins[6], Elisa Leão Teles[7], Esmeralda Rodrigues[7], M. Teresa Cardoso[7], Elena Ferreira[8], Sílvia Sequeira[9], Margarida Leite[1,2], Maria João Silva[1,2], Isabel Tavares de Almeida[1,2], João B. Vicente[1,2], Isabel Rivera[1,2]

1 Metabolism & Genetics Group, Research Institute for Medicines and Pharmaceutical Sciences (*i*Med.UL), Faculty of Pharmacy, University of Lisbon, Portugal
2 Department of Biochemistry and Human Biology, Faculty of Pharmacy, University of Lisbon, Portugal
3 Department of Pediatrics, Hospital Santa Maria, Lisbon, Portugal
4 Department of Medicine, Hospital Santa Maria, Lisbon, Portugal
5 Metabolic Clinics, Pediatric Hospital, CHUC, Coimbra, Portugal
6 Department of Pediatrics, Hospital Santo António, Porto, Portugal
7 Metabolic Diseases Unit, Integrated Pediatric Hospital, Hospital São João, Porto, Portugal
8 Hospital Centre, Funchal, Madeira, Portugal
9 Department of Pediatrics, Hospital D. Estefânia, Lisbon, Portugal

This chapter is an expanded version of the paper published in
Journal of Inherited Metabolic Disease (2014) 37(1):43-52

2.1. Abstract

Classic galactosemia is an autosomal recessive disorder caused by deficient activity of galactose-1-phosphate uridylyltransferase (GALT). Patients develop symptoms in the neonatal period, which can be ameliorated by dietary restriction of galactose. Nevertheless, many patients develop long-term complications, with a broad range of clinical symptoms with a poorly understood pathophysiology. The high allelic heterogeneity that characterizes this disorder is thought to play a determinant role in biochemical and clinical phenotypes.

Herein, we aimed to characterize the mutational spectrum of GALT deficiency in Portugal and to assess potential genotype-phenotype correlations.

Direct sequencing of the *GALT* gene and *in silico* analyses were employed to evaluate the impact of uncharacterized mutations upon GALT functionality. Molecular characterization of 42 galactosemic Portuguese patients revealed a mutational spectrum comprising 14 nucleotide substitutions: 10 missense, 2 nonsense and 2 putative splicing mutations. Sixteen different genotypic combinations were detected, half of the patients being p.Q188R homozygotes. Notably, the second most frequent variation is a splicing mutation. *In silico* predictions complemented by a close-up on the mutations' locations in the protein structure suggest that the uncharacterized missense mutations have cumulative point effects on protein stability, oligomeric state, or substrate binding. One splicing mutation is predicted to cause an alternative splicing event. This study reinforces the difficulty in establishing a genotype-phenotype correlation in classic galactosemia, a monogenic disease whose pathogenesis and clinical features suggest it is actually a complex disease. Indeed, classic galactosemia pathophysiological mechanisms result in a broad range of clinical outcomes, such as ovarian insufficiency and neurological dysfunction, and emphasize the need to expand the knowledge on this "cloudy" disorder.

2.2. Introduction

Classic galactosemia or type I galactosemia (OMIM #230400) is an inherited disorder of the galactose metabolism, caused by deficient activity of galactose-1-phosphate uridylyltransferase (GALT; EC 2.7.7.12), the second enzyme of the Leloir pathway. This condition is inherited in an autosomal recessive pattern and occurs with a pan-ethnic frequency of approximately 1:30,000 to 1:60,000 live births (1, 2).

Generally asymptomatic at birth, infants born with classic galactosemia become ill within days-to-weeks after initiating breast milk or lactose-containing formula ingestion (1, 3). Initial symptoms include poor feeding, vomiting and diarrhea with poor weight gain progressing with liver failure, renal tubular disease, *Escherichia coli* sepsis, coma, and death (1). Life-long dietary

restriction of galactose, the current gold standard therapy, partially relieves or prevents these acute and potentially lethal symptoms. However, many patients go on to develop severe long-term complications (4).

The human *GALT* gene (GenBank Accession NG_009029.1) is located in chromosome 9p13 and spans about 4 kb of DNA arranged in 11 exons (5, 6). The 1295 bp-long open reading frame is translated into a 379 amino acid polypeptide with an estimated molecular mass of 44 kDa, the active enzyme being an 87 kDa homodimer (1).

GALT belongs to the histidine triad family of transferases, displaying a double-displacement or ping-pong mechanism (7). In the first half-reaction, UDP-glucose (UDP-Glc) enters the active site and binds covalently to a histidine in the active site (His186), and glucose-1-phosphate (Glc-1-P) is released. In the second half-reaction, galactose-1-phosphate (Gal-1-P) enters the active site, and the enzyme-bound UMP is transferred to Gal-1-P, forming and releasing UDP-galactose (UDP-Gal) (1).

Classic galactosemia is characterized by high allelic heterogeneity (8), and to date 264 different variations have been described at the *GALT locus* (9, 10). Although a few of these mutations are relatively common, most are rare (1). Approximately 60% of the described variations are missense mutations, a common observation in genetic disorders (11). Additionally, 19 silent (7.2%), 17 nonsense (6.4%), 17 small deletions (6.4%), 6 insertions (2.3%), 3 indel (1.1%), 2 large gene deletions (0.8%), 2 no-stop changes (0.8%) and 1 frameshift (0.4%) mutations have also been reported. Regarding non-coding regions, 14 splice site variations (5.3%) have been described. The remaining 18 variations are classified either as benign (16; 6.1%) or as uncertain (2; 0.8%) (9).

The most frequently observed mutations are p.Q188R, p.K285N, p.S135L and p.L195P (1, 12). The predominant mutation in galactosemic patients of European descent is p.Q188R, which accounts for about 64% of all *GALT* mutations (1, 12). The mutation reaches its highest frequency in Ireland, where it occurs at a relative allele frequency of 94% (13), and shows a South-Eastern declining gradient throughout Europe (14, 15). Homozygous patients bearing the p.Q188R mutation have essentially no residual GALT activity in red blood cells (RBC) and have been frequently associated with a poor outcome (15, 16). The p.K285N is the second most common European mutation, especially in the countries of Central and Eastern Europe where it can account for up to 34% of classic galactosemia cases (17). In homozygous patients, this mutation is associated with complete absence of GALT activity in RBC, representing a severe clinical phenotype (15, 16). The p.S135L mutation is found almost exclusively in individuals of African origin, accounting for approximately 91% of African *GALT* mutant alleles (18, 19). Notably, this mutation appears to exhibit some tissue specificity: homozygous patients have essentially no GALT activity in RBC, but present some low residual activity in leukocytes. Homozygosity for this mutation is often associated with mild clinical outcomes (1, 10, 20).

Besides these pathogenic mutations, several variants that do not result in galactosemia have also been reported and are named after the place of their discovery (21). The most common variant, Duarte or D2, has p.N314D mutation associated with three intronic substitutions and a 4-nucleotide promoter deletion. The promoter deletion appears to decrease promoter function in appropriate constructs (22) and may thus account for the lower activity of the Duarte variant. The Los Angeles or D1 variant has the p.N314D and the c.652C→T (p.L218L) mutations, with neither the intronic substitutions nor the promoter deletion found in the Duarte variant, presenting a GALT activity slightly higher than the wild-type one.

An essential question regarding classic galactosemia concerns the identification of predictive factors to distinguish patients who will thrive in the long run from those who will experience complications. Therefore, *GALT* genotyping may be a valuable prognostic tool providing information about the biochemical and clinical outcome (10), contributing to a more personalized and effective therapy (23). Considering the wide mutational spectrum observed throughout the world, as well as the broad range of functional consequences of the mutations, there is no doubt that molecular analysis contributes to a better understanding of Portuguese galactosemic population prognosis.

Most recent studies on the molecular characterization of classic galactosemia in Portugal proposed p.Q188R to be the only frequent mutation in Portuguese patients, as the other variations were found in no more than two alleles (24). The aim of this study is to characterize the mutational spectrum of the *GALT locus* in a group of 42 Portuguese galactosemic patients, updating the current mutational spectrum. Moreover, an *in silico* strategy is herein employed to evaluate the impact of uncharacterized mutations upon GALT functionality. Finally, comparing the modeling data with phenotypic parameters presented by the patients, we assess potential genotype-phenotype correlations.

2.3. Methods

2.3.1. Patients

Forty-two galactosemic Portuguese patients, encompassing four pairs of siblings, were investigated. The nineteen females (45%) and twenty-three males (55%) integrated in the study originated from all regions in Portugal, including Madeira and Azores Islands, thus being considered representative of the whole population. Diagnosis of patients from South/Center regions and Madeira Island was confirmed by absent or reduced (<5% of control mean) GALT activity in RBC. A positive Beutler test was used to confirm diagnosis of patients from North region and Azores Islands. A galactose-restricted diet was started upon disease suspicion. A

whole blood sample was collected from every patient in order to prepare RBC lysates for evaluation of GALT activity and to isolate nucleic acids for mutational analysis.

This study was approved by the local Ethics Committees and informed consents were obtained from the patients or their parents, who were also enrolled in the study, whenever possible.

2.3.2. GALT activity

Patients' GALT activity was determined by HPLC, according to a previously described method (25). Briefly, GALT activity was measured in RBC lysates using 40 mM Tris–HCl pH 7.5, 40 μM DTT, 6 mM Gal-1-P, 125 mM glycine and 1.5 mM UDP-Glc. Reaction substrate and product, respectively UDP-Glc and UDP-Gal, were separated using a reversed-phase Supelcosil (LC-18-S column, 150 mm × 4.5 mm, particle size 5 μm; Sigma-Aldrich) with UV detection at 262 nm (Waters 2996 Photodiode Array detector). The column temperature was maintained at 22°C. GALT activity was expressed as nmol of UDP-Gal formed *per* μmol of hemoglobin per hour.

Otherwise, the Beutler enzyme spot test was employed, in which GALT enzyme activity is monitored with the aid of phosphoglucomutase and glucose-6-phosphate dehydrogenase and visualization of the fluorescence of reduced $NADP^+$ under ultraviolet light (26, 27). Potential false positives were discarded by determination of galactose or Gal-1-P levels.

2.3.3. Galactose-1-phosphate levels

Gal-1-P levels were determined in RBC lysates by an in-house adaptation of a fluorometric assay (28), in a two-step reaction, using 500 mM Tris pH 8.6, 0.45 mM Mg^{2+}, 0.34 mM NAD^+. In the first reaction, free galactose is quantified by the addition of galactose dehydrogenase, which catalyses the oxidation of galactose and the reduction of NAD^+ to NADH, monitored by fluorescence spectroscopy. In the second reaction, Gal-1-P is quantified by the addition of alkaline phosphatase, thereby producing galactose, which is, once again, estimated by the fluorescence emitted by NADH. Gal-1-P is calculated by the difference between the second reaction, which measures total galactose (Gal-1-P plus free galactose), and the first reaction, which measures free galactose. Results were expressed as mg of Gal-1-P *per* 100 mL of RBC lysate.

2.3.4. Genotype Analysis

Genomic DNA was isolated from peripheral blood samples using the Citogene® DNA Blood Kit (Citomed, Lisbon, Portugal) according to the manufacturer's instructions. Total RNA was prepared from peripheral lymphocytes using the RNeasy® MiniKit (Qiagen, Hilden, Germany), and contamination by genomic DNA was eliminated by treatment with deoxyribonuclease I. One μg was reverse transcribed using the Reverse Transcription System A3500 kit (Promega, Madison, WI, USA) with random primers following the manufacturer´s instructions.

All the individual exons and respective intron boundaries, as well as the full-length cDNA, were PCR amplified and scanned for mutations by direct sequence analysis, using the ABI PRISM BigDye Terminator Cycle Sequencing Ready Reaction Kit, and ABI PRISM 3100 Genetic Analyzer (Applied Biosystems, Foster City, CA, USA). Additionally, in those cases revealing no alteration, the full introns were completely sequenced after PCR amplification. Table 2.1 lists the primers used for amplification of the *GALT* gene and cDNA.

2.3.5. *In silico* analysis of missense variants

GALT protein sequences from several organisms were retrieved from NCBI BLAST using as template the human GALT sequence (SWISS-PROT accession number: P07902.3). Selected sequences were aligned with ClustalX interface for Windows (29) and inspected with Genedoc. To predict the impact of selected missense mutations on the GALT stability and function, a structural model of wild-type human GALT was generated with the SWISS-PROT server (30, 31), using as template the structure of *Escherichia coli* GalT (1GUP) (32). Using the same approach, structural models of the p.F171C, p.G175D, p.P185S, p.S192G, p.R259W, p.P295T, and R333G mutants were also generated. The wild-type human GALT PDB file was modified to include only chain A. With the obtained model PDB, different servers were used to increase the robustness of the predictions of the mutations impact: PolyPhen-2 (33), SDM (34), PopMusic v2.1 (35, 36) and CUPSAT (37, 38). As controls of disease-causing mutations, the prevalent variants p.S135L, p.Q188R and p.K285N were analyzed using the same servers.

Table 2.1. Sequence of oligonucleotides used for the amplification of *GALT* gene and cDNA.

Primer ID	Sequence	Location in sequence	Fragment length
	Genomic (NG_009029.1)		
Exon 1F	TGG ACG GAG AAA GTG AAA GG	5002 – 5021	221 bp
Exon 1R	AGG GGA CGA AAG CTT CCT AA	5205 – 5224	
Exon 2F	GGA GAG AGG GAG CTA GAG AGC	5399 – 5419	300 bp
Exon 2R	CTG TGG GTG GAG GAC AGT TC	5679 – 5698	
Exon 3+4F	GTA TGG GGC AGT GAG TGC TT	5814 – 5833	300 bp
Exon 3+4R	TCC AAC TCT GGT TTG AAA GTT GT	6091 – 6113	
Exon 5F	GTA GCA CAG CCA AGC CCT AC	6142 – 6161	244 bp
Exon 5R	CCC AGA ACC AAA GCT TCA TC	6368 – 6387	
Exon 6F	TCA GGA GGG AGT TGA CTT GG	6418 – 6437	210 bp
Exon 6R	TTT CCT CTG TCC CAT CCA TT	6608 – 6627	
Exon 7F	CCT GCC TGT TCT TCT CTG CT	1662 – 1681	231 bp
Exon 7R	CCT TTA CAC ACC TCT CTC ATG G	6871 – 6892	
Exon 8+9F	CAC CTT GAT GAC TTC CTA TCC A	7089 – 7110	399 bp
Exon 8+9R	GAA ATG GTG TTG GGG CTA AA	7468 – 7487	
Exon 10F	CTG GTT GGG TTT GGG AGT AG	7691 – 7710	299 bp
Exon 10R	TCA GGT GCC TGC ACA TAC TG	7470 – 7499	
Exon 11F	CAG AGC AAG ACC TCG TCT CA	8627 – 8646	300 bp
Exon 11R	TTT TAT TAA TGC TAT ATC TGC CCA AA	8901 – 8926	
DelProm F	GCT CTG GGT GAA GTA GGA TCA	4856 – 4876	299 bp
DelProm R	GTT ACC GTT TGC CCG GAA G	5136 - 5154	
IVS1F	AGT GGA ACC GAT CCT CAG C	5077 – 5095	435 bp
IVS1R	GGT GAG CTG ACA CCA GCA	5494 – 5511	
IVS2+3F = 2F	GGA GAG AGG GAG CTA GAG AGC	5399 – 5419	715 bp
IVS2+3R = 3+4R	TCC AAC TCT GGT TTG AAA GTT GT	6091 – 6113	
IVS4F = 3+4F	GTA TGG GGC AGT GAG TGC TT	5814 – 5833	574 bp
IVS4R = 5R	CCC AGA ACC AAA GCT TCA TC	6368 – 6387	
IVS5F	GGC TGT TGT TGA TGC ATG G	6261 – 6279	347 bp
IVS5R	AAT CCT GTT CCC ATG TCC AC	6588 – 6607	
IVS6F = 6F	TCA GGA GGG AGT TGA CTT GG	6418 – 6437	475 bp
IVS6R = 7R	CCT TTA CAC ACC TCT CTC ATG G	6871 – 6892	
IVS7+8F = 7F	CCT GCC TGT TCT TCT CTG CT	6662 – 6681	826 bp
IVS7+8R = 8+9R	GAA ATG GTG TTG GGG CTA AA	7468 – 7487	
IVS9F = 9+9F	CAC CTT GAT GAC TTC CTA TCC A	7089 – 7110	901 bp
IVS9R = 10R	TCA GGT GCC TGC ACA TAC TG	7470 – 7499	
IVS10.1F	ATG GTT GGC TAC GAA ATG CT	7874 – 7893	492 bp
IVS10.1R	GGC CGA TGG TAT GGA GTT	8348 – 8365	
IVS10.2F	TCC AGC AAA ATG TCT CCT GA	8320 – 8339	487 bp
IVS10.2R	GAT GGT TGC TGT CTC CCT GT	8787 – 8806	

Primer ID	Sequence	Location in sequence	Fragment length
	mRNA (NM_000155.2)		
Fragment AF	TGG ACG GAG AAA GTG AAA GG	2 - 21	366 bp
Fragment AR	TGG GAA GTC GTT GTC AAA CA	348 - 367	
Fragment BF	CCC TCT CAA CCC TCT GTG TC	274 - 293	500 bp
Fragment BR	TGG TTA GGA CCA GAC GTT CC	754 - 773	
Fragment CF	GTG AGG AGC GAT CTC AGC A	669 - 687	378 bp
Fragment CR	GGA GCG GAG GGT AGT AAT GA	1027 - 1046	
Fragment DF	GAC GTC CTT TCC CTA CTC CA	940 – 959	343 bp
Fragment DR	TTC AAG GCC CTT TCT GCT TA	1263 – 1282	

2.3.6. *In silico* analysis of potential splice variants

The intronic mutations identified in the Portuguese galactosemic population were analyzed by bioinformatic tools, using different algorithms, which predict their influence upon splicing signals and splicing regulatory sequences in the context of the sequence variation. The strength of authentic and cryptic splice sites was assessed by NetGene2 (39, 40) and NNSPLICE 0.9 (41). Predictive programs RESCUE-ESE (42) and ESEfinder 3.0 (43, 44) were used for detection of exonic splicing enhancer (ESE) motifs. RESCUE-ESE identifies hexamers whose sequences are possible binding sites of several known RNA binding factors, and predicts whether a mutation disrupts any of these elements. ESEfinder 3.0 identifies ESE responsive to the human SR proteins: SRSF1, SRSF2, SRSF5 and SRSF6, formerly known as SF2/ASF, SC35, SRp40 and SRp55, respectively, and predicts whether a mutation disrupts or enhances the binding of any of these proteins. Default settings for human sequences and default thresholds were used for all analyses.

2.4. Results and Discussion

2.4.1. Mutational spectrum and genotypes found in the Portuguese GALT deficient patients

This study aimed to update characterization of the mutational spectrum of GALT deficiency in Portugal (24, 45), as well as to assess genotype-phenotype correlations. The molecular characterization of 42 galactosemic Portuguese patients was achieved, corresponding to 76 independent mutant alleles. The mutational spectrum encompassed 14 different mutations, which are distributed along the *GALT* gene (Table 2.2). All mutations are nucleotide

substitutions, corresponding to 10 missense (~71%), 2 nonsense (~14.5%), and 2 putative splicing (~14.5%) mutations. Moreover, 5 of the observed mutations were so far only identified in Portuguese galactosemic patients (p.F171C, p.G175D, p.P185S, p.S192G and p.P295T), the majority being located in exon 6, the most conserved exon in the *GALT* gene.

Table 2.2. Characterization of *GALT* mutations identified in 42 Portuguese galactosemic patients, corresponding to 76 independent mutant alleles.

Sequence variation	GALT mutation	Exon/Intron	Alleles (n)	Allele frequency (%)	*In vitro* GALT activity
c.238C>T	p.R80X	E2	1	1.3	n.a.
c.328+33G>A	IVS3+33G>A	I3	1	1.3	n.a.
c.404C>T	p.S135L	E5	3	4.0	5% (46, 47)
c.443G>A	p.R148Q	E5	2	2.6	n.a.
c.512T>G	p.F171C	E6	1	1.3	n.a.
c.524G>A	p.G175D	E6	3	4.0	n.a.
c.553C>T	p.P185S	E6	2	2.6	14% (48)
c.563A>G	p.Q188R	E6	51	67.1	0%
c.574A>G	p.S192G	E7	1	1.3	n.a.
c.610C>T	p.R204X	E7	1	1.3	0% (49)
c.775C>T	p.R259W	E8	2	2.6	<0.2 (50)
c.820+13A>G	IVS8+13A>G	I8	6	8.0	n.a.
c.883C>A	p.P295T	E9	1	1.3	n.a.
c.997C>G	p.R333G	E10	1	1.3	n.a.
Total			**76**	**100**	

n.a. – not available

Among the 42 patients, genotype analysis revealed the presence of 16 different genotypic combinations. Moreover, 26 patients were homozygotes (62%), corresponding to 46 independent mutant alleles, while the remaining patients were compound heterozygotes. Table 2.3 shows the genotypes found in the Portuguese galactosemic population, half of the patients being homozygous for the p.Q188R mutation.

The p.Q188R mutation, accounting for 67.1% of all mutant alleles, was once more confirmed as the most prevalent mutation in Portuguese galactosemic population (24, 45), as reported in all European descendant patients. Twenty-one patients have this mutation in homozigosity, and 14 patients in compound heterozygosity.

Surprisingly, the intronic variation c.820+13A>G (IVS8+13A>G), described as benign in the ARUP database (http://www.arup.utah.edu/database/GALT/GALT_display.php) (9), was found in 6 independent mutant alleles, thus being the second most frequent mutation among the Portuguese galactosemic patients (8.0%). This mutation was identified in a homozygous patient and in five compound heterozygotes, two of them being siblings. Four patients carry the p.Q188R

in the other allele, while one revealed the p.P295T mutation in his second allele. Interestingly, all these patients originate from central regions of Portugal.

The third most frequent mutations were p.S135L and p.G175D, each accounting for 4% of all mutant alleles. Concerning the p.S135L mutation, it was identified in 3 independent mutant alleles, and always in heterozygosity. Considering the strong African immigration Portugal experienced in the late XX century, it is noteworthy that only three patients carry the p.S135L mutation. Two of them are heterozygotes for p.Q188R variant, while the other patient carries the p.F171C (c.512T>G; TTT>TGT) mutation in the other allele. It is interesting to note that another mutation in the same nucleotide (c.512T>C, TTT>TCT, p.F171S) was described as the second most frequent mutation in African-Americans (15, 51). Both parents of the latter patient are from the Democratic Republic of São Tomé and Príncipe, in line with the African origin of the p.S135L and p.F171S mutations.

Up to now, the p.G175D mutation was exclusively detected in Portugal. A previous study from 2006 reported a patient carrying this mutation in only one allele (24, 45). Herein we report the identification of two additional independent mutant alleles identified in homozygous status in a pair of siblings from Madeira Island.

The p.R148Q, p.P185S and p.R259W mutations were detected with a frequency of 2.6%. Notably, the p.R148Q mutation, previously described as a common mutation in the Iberian Peninsula, was only found in two mutant alleles, whereas seven alleles have been identified in the Spanish galactosemic population (24, 45). On the other hand, up to now, p.P185S has been found exclusively in the Portuguese galactosemic population, in a homozygous patient. Similarly, the p.R259W was detected in homozygosity in one patient.

The remaining mutations – p.R80X, c.328+33G>A (IVS3+33G>A), p.F171C, p.S192G, p.R204X, p.P295T, p.R333G – occurred to a lesser extent, at a frequency of only 1.3%, corresponding to one allele of each mutation.

Two additional individuals whose phenotypes were suggestive of classic galactosemia have been initially enrolled in this study. Their molecular characterization revealed the presence, in heterozygous status, of the p.N314D mutation, along with the associated promoter deletion and intronic variations. Nevertheless, their GALT activity showed values compatible with the Duarte variant carrier status, thus being excluded from this study (52).

Moreover, we screened for the p.N314D variation among all the previously referred galactosemic patients and no mutant allele was identified. Accordingly, all the investigated patients carried only two mutant *GALT* alleles in *trans*.

Table 2.3. Genotypic and phenotypic data of 42 Portuguese galactosemic patients.

Patient	Birth Year	Gender	Allele 1	Allele 2	GALT activity (nmol UDP-gal/µmol Hb/h)	Gal-1-P (µmol/L) Diagnosis/Treatment Mean	Poor outcomes
1	2000	F	p.Q188R	p.Q188R	0.7	956 / 99.0	C, AMF, SI
2	1989	M	p.Q188R	p.Q188R	0.09	n.d. / 107	AMF, ID
3	1993	M	p.Q188R	p.Q188R	0.9	483 / 130	n.a.
4	1992	F	p.Q188R	p.Q188R	0	n.a. /82.7	SI, OD, WMA
5	2000	F	p.Q188R	p.Q188R	1.47	800 / 96.6	C, ID, SI
6	1989	F	p.Q188R	p.Q188R	0	603 / 82.8	C, OD
7	1984	M	p.Q188R	p.Q188R	0.1	292 / 61.7	C
8	1987	F	p.Q188R	p.Q188R	0	139 / 79.8	LD, SI
9*	1988	F	p.Q188R	p.Q188R	0	74 / 95.8	C
10*	2005	M	p.Q188R	p.Q188R	0.7	264/ 105.7	SI
11	2002	M	p.Q188R	p.Q188R	1.5	2772 / 108	C, ID, SI
12	2000	M	p.Q188R	p.Q188R	1.08	262 / 89.1	Absent
13	2009	F	p.Q188R	p.Q188R	n.d.	112 / 108	n.a.
14	1991	M	p.Q188R	p.Q188R	n.d.	n.d.	ID
15*	1991	M	p.Q188R	p.Q188R	n.d.	n.d.	ID
16*	2004	F	p.Q188R	p.Q188R	n.d.	n.d.	LD
17	1998	F	p.Q188R	p.Q188R	n.d.	n.d.	ID, OD
18	2006	F	p.Q188R	p.Q188R	0	536 / 119	No follow-up
19	2009	F	p.Q188R	p.Q188R	n.d.	797 / 89.1	n.a.
20	2000	M	p.Q188R	p.Q188R	n.d.	n.d.	LD, WMA
21	2004	F	p.Q188R	p.Q188R	n.d.	n.d.	LD
22	1993	M	p.Q188R	p.R80X	n.d.	n.d.	ID

Patient	Birth Year	Gender	Allele 1	Allele 2	GALT activity (nmol UDP-gal/µmol Hb/h)	Gal-1-P (µmol/L) Diagnosis/Treatment Mean	Poor outcomes
23	2004	M	p.Q188R	IVS3+33G>A	0.9	22 / 22.9	LD, SI, WMA
24	2001	M	p.Q188R	p.S135L	1.05	255 / 66.1	C, AMF, SI
25	2008	M	p.Q188R	p.S135L	n.d.	n.d.	SI
26	1989	F	p.Q188R	p.R148Q	0	n.a. / 79.8	ID, OD
27	1996	M	p.Q188R	p.R148Q	0.6	466 / 95.2	LD, SI
28	1987	M	p.Q188R	p.G175D	0.6	235 / 76.7	C, ID, SI
29	1979	M	p.Q188R	p.S192G	0	82 / 73	Absent
30	2006	M	p.Q188R	p.R204X	0	1320 / 102	LD
31*	2002	F	p.Q188R	IVS8+13A>G	n.d.	352 / 62.2	ID, SI
32*	2010	M	p.Q188R	IVS8+13A>G	4.4	307 / 78.5	Absent
33	2000	F	p.Q188R	IVS8+13A>G	0.1	132 / 46.2	SI
34	1989	F	p.Q188R	IVS8+13A>G	0	n.a. / 65.9	ID, OD
35	2007	F	p.Q188R	p.R333G	0.17	1644 / 53.8	Absent
36	2010	M	p.S135L	p.F171C	n.d.	140 / 71.0	SI
37*	1988	M	p.G175D	p.G175D	0	127 / 75.1	Absent
38*	1994	F	p.G175D	p.G175D	1.2	495 / 95.1	WMA
39	1996	M	p.P185S	p.P185S	4.65	n.a. / 50.6	ID, SI
40	1991	F	p.R259W	p.R259W	n.d.	n.d.	C, LD, OD
41	1991	M	IVS8+13A>G	IVS8+13A>G	0	125 / 37.3	LD
42	1988	M	IVS8+13A>G	p.P295T	0	125 / 52.6	C, LD

* Siblings

C – cataracts; AMF – anomalies of motor function; ID – intellectual disability; LD – learning disabilities; OD – ovarian dysfunction; SI – speech impairment; WMA – white matter anomalies. n.a. – not available; n.d. – not determined

It is noteworthy that two frequent mutations (p.K285N and p.L195P) observed in European populations, including the Spanish one, remain undetectable among the Portuguese patients. Interestingly, the Spanish *GALT* spectrum encompasses 21 different mutations, 7 more than the Portuguese one, but there are only 5 mutations in common with both populations. Nevertheless, these data are in accordance with what is observed for other genetic disorders, including phenylketonuria, which present remarkably different mutational spectra between both countries (53-55).

2.4.2. Prediction of mutations' impact upon GALT stability and functionality

Most of the mutations identified in the Portuguese galactosemic population affect strictly conserved residues in GALT sequences from organisms of all life domains, suggesting that the substituted positions are structurally and functionally relevant. Only a few missense mutations (e.g., p.Q188R, p.S135L) were studied by expressing the recombinant mutant proteins either in eukaryotic or prokaryotic systems (e.g. (56, 57)). While the effect upon the enzyme structural and functional properties of the newly described mutations is still under investigation and out of the scope of the present paper, we employed bioinformatic tools to predict their structural and functional impact (Table 2.4). To this end, we generated a structural model of human GALT using as template the *E. coli* GalT structure (PDB code 1GUP (32)), which was strengthened by the high sequence identity and similarity between both primary sequences (51% and 64%, respectively). To increase the robustness of the mutations predictions, we employed four different servers, and always tested the most prevalent disease-causing missense mutations identified in classic galactosemia patients, namely p.Q188R, p.S135L, and p.K285N.

Although there is a reasonable degree of consensus between prediction servers (Table 2.4), there are relevant point discrepancies, which may be due to these servers analyzing the overall structural perturbations and overlooking essential details such as dimer assembly, metal binding, active site and other functionally relevant residues. A striking example is p.Q188R, which is not considered disease-causing or destabilizing by every server; however, Q188 is a key site for reaction intermediates stabilization and the high impact of the p.Q188R mutation cannot be discriminated by the servers (58). These limitations and discrepancies weaken the conclusions drawn from the prediction servers, which prompted us to generate structural models of the GALT mutants herein studied, to visually inspect local effects in the mutations in terms of proximity to the active site, bound reaction intermediates, dimer interface, and appearance or loss of H-bonds and salt bridges (results summarized in Table 2.5).

The p.F171C mutant typifies the prediction servers' limitations: whereas two programs suggest a stabilizing effect, the other two classify it as a destabilizing mutation (Table 2.4). A

closer look to the mutation location (Supplementary Figure S2.2) shows that F171 is located near the active site and is part of a hydrophobic patch that includes W190 from the same monomer and Y339 from the opposing monomer. F171 has been shown to be tightly associated with Q188, the residue which stabilizes the bound intermediate, by establishing H bonds between both main chain carbonyl and amine groups. Indeed, substitutions in F171 by Ser, Leu and Tyr were shown to alter the position of Q188 and disturb the dimer interface and hexose binding (51). In the F171C model (Supplementary Figure S2.2), the H-bonds between the main chain C- and N-groups of C171 and Q188 are slightly shorter than in wild-type GALT, which could account for some distortion in the Q188 side-chain position, thus affecting hexose binding. Moreover, the C171 may disturb the monomer-monomer interaction by weakening the hydrophobic patch with W190 and Y339.

The substitution of a glycine by an aspartate residue in the p.G175D mutant is consensually destabilizing according to the servers' predictions (Table 2.4). G175 is strictly conserved (Supplementary Figure S2.1) and inserted in a coil region near the active site and bound substrate, as observed in the generated model (Supplementary Figure S2.2). The only striking structural perturbation resulting from the substitution of glycine by aspartate in this mutant is the moderately strong H-bond possibly made between the aspartate side-chain carboxylate and the main chain amine group of nearby M177 (D175 side-chain C=O and M177 main chain NH at a distance of 2.73 Å; Table 2.5). This additional H-bond may create an extra rigidity in this coil region that is marked by strong sequence conservation and is located near the dimer interface. In summary, this mutation is located not far from the active site and bound hexose, and may also introduce structural perturbations in a coil region close to the monomer-monomer interface (Table 2.5).

One of the most striking mutations is the substitution of proline by serine in the p.P185S mutant. This mutation is located at the core of the H-P-H active site and, thus, likely to have a strong impact on the GALT activity of the mutated enzyme. This mutation has already been characterized from a functional point of view along with other substitutions in the same position by alanine, glycine, glutamine and glutamate, by expressing the human GALT (wild-type and mutants) in a yeast GALT-null system (48). In this study, p.P185S was produced in lower abundance than wild-type, and displayed impaired conformational stability and ~7-fold decreased activity as compared to wild-type. Indeed, the prediction servers concur in assigning a destabilizing and disease-causing effect to this mutation (Table 2.4). A closer inspection of the possible structural impact of this mutation (Supplementary Figure S2.2) shows that the substituting serine may not only disturb the active site, but also introduce some rigidity in the H-(P)S-H tripeptide, since the S185 side-chain hydroxyl moiety is likely to form a strong H-bond with the main-chain C=O from valine 137 (S185 -OH at 2.48 Å from V137 main-chain C=O;

Table 2.5). Overall, this mutation is predicted to have a strong impact on GALT activity, while causing some local structural perturbations within each monomer.

The substitution of serine by glycine in p.S192G occurs in a position with poor sequence conservation, being replaced by a cysteine in the fruit fly GALT, a leucine in yeast GALT and asparagine in bacterial GalT proteins (Supplementary Figure S2.1). This position is, however, located near the active site, a region of strong sequence conservation. As observed in Table 2.4, the servers' predictions for this mutation are contradictory, since it is either considered to be benign/neutral, or destabilizing. It must, however, be taken into account that prediction servers focus solely on local contact and global impact concerning the structural model of the monomeric GALT. Indeed, a closer inspection of the mutated site shows that the substituted serine is likely to yield a weak H-bond between its side-chain -OH moiety and the -OH moiety of tyrosine 339 from the opposing monomer, Y339(B) in Supplementary Figure S2.2 (Ser192 side-chain -OH at 3.45 Å from side-chain -OH of Y339(B)). It is thus likely that, despite the fact that S192 is not close to the active site or bound substrate, the substitution to a glycine will prevent the formation of a H-bond with the neighbouring Y339(B) and thus disturb the dimer interface (Table 2.5). Notably, this tyrosine residue – Y339(B) – is strictly conserved, and is the same residue responsible for dimer interface perturbation caused by the p.F171C mutation.

The p.R259W mutation affects a position with no sequence conservation, although located amidst a highly conserved motif. Prediction servers diverged in ranking this mutation (Table 2.4). A closer inspection shows the mutated R259 is located at the GALT surface, with its side chain guanidinium moiety protruding into the solvent (Supplementary Figure S2.2). This residue is located far from the active site or dimer interface. The most striking perturbation resulting from substituting tryptophan side chain concerns the movement of an adjacent glutamate residue (E271) side chain. The tryptophan bulky indole moiety induces a 3.24 Å displacement of the E271 side chain carboxylate which enables the formation of two moderately strong H-bonds (2.98 Å and 3.01 Å) with the side chain -OH and main chain NH group of T268. These extra H-bonds will likely cause rigidity between the helix where E271 is located and the coil region of T268. This mutation highlights how even subtle structural perturbations in the GALT surface may propagate into functionally and structurally relevant sites and affect the overall GALT activity.

The substitution of proline by threonine in the p.P295T mutant occurs in a strictly conserved sequence motif (Supplementary Figure S2.1) in a coil region. In line with the structural perturbations arising from proline substitutions, the prediction servers align in scoring this as a probably destabilizing and disease-causing mutation (Table 2.4). Proline substitutions by threonine have been shown to induce disease-causing structural and functional perturbations in proteins related to other pathologies (59, 60). A closer look into the localization of the p.P295T mutation (Supplementary Figure S2.2) shows that the T295 side chain -OH moiety is within H-

bonding distance with the M177 side chain sulfur atom (3.70 Å; sulfur H-bonds may display longer distances (61, 62)). As mentioned for the p.G175D mutation, this M177 is located in a coil region near the dimer interface, and the rigidity derived from an extra H-bond may disturb the monomer-monomer interaction (Table 2.5).

The p.R333G mutation occurs in a strictly conserved position amidst a motif of high sequence conservation (Supplementary Figure S2.1). This substitution of arginine by glycine is ranked by the four prediction servers as a destabilizing mutation (Table 2.4). R333 is located in a coil region distant from the H-P-H active site or the bound substrate (Supplementary Figure S2.2). A closer inspection of this position reveals that the R333 side chain guanidinium moiety is located at the monomer surface close to the dimer interface, although no interaction with the other monomer can be envisaged. However, the R333 guanidinium moiety appears to be involved in a H-bond network with the main chain carbonyl groups of M177 and K334. The substitution of R333 by a glycine leads to the loss of these three H-bonds and is, thus, likely to disturb this whole region located at the monomer-monomer interface (Table 2.5).

In summary, the predicted structural and functional effects of the described mutations reveal that apparently innocuous substitutions may cause a number of point disturbances in the active site, substrate binding moieties, putative metal binding sites, and monomer-monomer interface. Altogether, these subtle perturbations may account for impaired conformational stability and/or increased proneness to aggregate, prompting the misfolded mutant GALT for degradation by the cell protein quality control systems, resulting in lower levels of a less active GALT protein, and ultimately leading to low GALT activity in patients. The inspection of local effects of the mutations, together with the knowledge on the wild-type structure and function, provide a more valuable tool than prediction servers which overlook relevant information about the analyzed proteins. Moreover, these predictions should be carefully considered, since they rely upon the observation of structural models based on the crystallographic structure of the *E. coli* enzyme.

2.4.3. GALT variations putatively affecting splice events

Bioinformatic programs were used to evaluate a possible missplicing effect of the two intronic variations identified in the Portuguese patients, c.820+13A>G (IVS8+13A>G) and c.328+33G>A (IVS3+33G>A). It is noteworthy that none of these nucleotide transitions was identified in 100 control alleles (24, 45), which strongly suggests they are indeed disease-causing mutations.

Table 2.4. Structural and functional effects of GALT mutations predicted by bioinformatics tools. A structural model of human GALT was generated by the SWISS-MODEL server using the *E. coli* GalT crystallographic structure as template (PDB code 1GUP), and uploaded to the prediction servers to analyze the effect of the selected mutations.

Mutation	***Prediction Server***									
	Polyphen-2		***SDM***			***PopMusic***		***CUPSAT***		
	Overall stability	Score	Overall stability	Effect	ΔΔG° (kcal/mol)	Overall stability	ΔΔG° (kcal/mol)	Overall stability	Torsion	ΔΔG° (kcal/mol)
F171C	probably damaging	1.00	stabilizing	non-disease associated	1.97	destabilizing	1.63	stabilizing	favourable	0.58
G175D	probably damaging	1.00	destabilizing	non-disease associated	-1.53	destabilizing	1.73	destabilizing	unfavourable	-1.97
P185S	probably damaging	1.00	highly destabilizing	causes protein malfunction and disease	-2.12	destabilizing	1.60	destabilizing	unfavourable	-3.87
S192G	benign	0.045	neutral	non-disease associated	-0.24	destabilizing	1.26	destabilizing	unfavourable	-3.87
R259W	probably damaging	1.00	stabilizing	non-disease associated	1.93	destabilizing	0.19	destabilizing	unfavourable	-12.46
P295T	probably damaging	1.00	highly destabilizing	causes protein malfunction and disease	-2.23	destabilizing	1.07	destabilizing	unfavourable	-2.38
R333G	possibly damaging	0.821	destabilizing	non-disease associated	-1.27	destabilizing	2.19	destabilizing	unfavourable	-2.17
S135L	probably damaging	0.94	highly destabilizing	causes protein malfunction and disease	2.25	destabilizing	0.89	destabilizing	unfavourable	-5.04
Q188R	probably damaging	1.00	neutral	non-disease associated	-0.22	destabilizing	0.55	stabilizing	favourable	1.93
K285N	probably damaging	0.94	slightly stabilizing	non-disease associated	0.83	destabilizing	1.45	destabilizing	unfavourable	-1.79

Table 2.5. Structural functional effects of GALT mutations predicted by inspection of structural models. Structural models of human GALT (wild-type and selected mutants) were generated by the SWISS-MODEL server using the *E. coli* GalT crystallographic structure as template (PDB code 1GUP). Conservation was determined by aligning selected sequences in Clustal X (Supplementary Figure S2.1). H-bond distances were calculated in PyMOL (63).

		Location / Effect of mutation				
Mutation	**Conservation[1]**	*Dimer Interface*	*Salt bridges*	*H-bonds*	*Near active site*	*Near bound substrate*
F171C	100%	part of hydrophobic patch with Y339 from the other monomer and W190	×	shorter H-bonds with Q188	✓	✓
G175D	100%	close to dimer interface	×	Asp side chain C=O with Met177 main chain NH	✓	✓
P185S	100%	×	×	Ser side chain -OH with Val137 main chain C=O	✓	✓
S192G	poor (mostly N)	×	×	Ser side chain -OH with Y339(B) side chain -OH	×	×
R259W	not conserved	×	×	Glu271 side chain C=O and C-OH with Thr268 main chain NH side chain -OH, respectively	×	×
P295T	100%	close to dimer interface	×	Thr side chain -OH with Met177 side chain S	×	×
R333G	100%	close to dimer interface	×	Arg side chain guanidinium moiety with Met177 C=O and Lys334 C=O	×	×

In order to predict if the c.820+13A>G mutation directly affects the pre-mRNA splicing of the *GALT* gene or if it is just a marker linked to another causative mutation, and to understand the possible mechanism underlying it, wild-type and mutant gene sequences comprising exon 8, intron 8 and exon 9 (320 bp) were scanned by two different programs for splice site prediction and the relevant scores were compared.

Although both programs calculate scores according to different algorithms, the analyses gave similar results. The authentic 5' splice donor site of intron 8 presents the same or similar scores in both wild-type and mutant sequences in each program (0.64 and 0.86 in NNSplice and NetGene2, respectively) and these results are indicative that this is not a weak splicing donor site. Both programs also predict a new GT donor site (c.820+14_15) in the presence of the mutation. In the NNSplice program, this cryptic donor site is only recognized in the mutant sequence, in which it scores 0.95, whereas in the NetGene2 program it scores 0.68 and 0.99 in the wild-type and mutant sequences, respectively. This cryptic splice site is located immediately downstream of the variant nucleotide site and, in the presence of the mutation, both programs assign it a higher strength comparatively to the canonical 5' splice site. These results suggest that, *in vivo*, the c.820+14_15 splice site prevails over the canonical 5' splice site, being used by the spliceosome machinery and leading to the inclusion of the first 13 nucleotides of intron 8 in the coding sequence.

The authentic 3' splice acceptor site of intron 8 shows very low scores in both programs, 0.45 in NNSplice and 0.04 in NetGene 2, revealing its intrinsic weakness. Moreover, both programs reveal that the mutation seems to create a new splice acceptor site in intron 8, immediately upstream the activated cryptic donor site (c.820+12_13) which, despite being also extremely weak, shows higher scores than the canonical one (0.51 and 0.09 in NNSplice and NetGene2, respectively).

The evaluation of possible modifications on ESE patterns caused by this nucleotide variation, was then assessed by two different programs which displayed contradictory results. The ESEfinder 3.0 program scores the same or very similar values for three of the SR proteins analyzed (SRSF2, SRSF5, SRSF6) in both wild-type and mutant sequences. However, when analyzing the results for the SRSF1 protein, we observe that its binding score increases from 1.986666 in the wild-type sequence, which is close to the threshold value (1.956), to a much higher score of 3.906278 in the mutant sequence (CCCAGGT, positions 9–15 of intron 8). On the other hand, the RESCUE-ESE program identifies in the same region a binding sequence (CAAGTA, positions 11–16 of intron 8) in the wild-type sequence, but not in the mutant one.

Concerning the c.328+33G>A, wild-type and mutant gene sequences encompassing exon 3, intron 3 and exon 4 (213 bp) were also scanned by the previously referred predictive servers. NNSplice and NetGene2 programs attribute similar scores, in the presence and absence of this

variation, to the natural donor (0.90 and 0.96) and acceptor (0.95 and 0.73) splice sites, respectively, and recognized no cryptic splice sites in either sequence.

The ESEfinder software shows this intronic change presents the same binding scores to all motifs covered by this software, with the exception of those for SRSF1. In effect, the wild-type sequence presents a motif (AGGAGGG, positions 28-34 of intron 3) scoring 2.05955 which is however no longer recognized in the presence of the mutation (AGGAGAG). On the contrary, the RESCUE-ESE program only identifies a new ESE motif for this same SR protein in the presence of the mutation (AGGAGA, positions 28-33 of intron 3).

Altogether, these data strongly suggest that the c.820+13A>G variation is most probably a splicing mutation, whereas the effect of the c.328+33G>A variation remained elusive. Unfortunately, aberrant splicing resulting from these mutations could not be investigated since no material other than genomic DNA was available from these patients or their parents. Efforts are underway to study these mutations in greater detail.

2.4.4. Assessing potential genotype-phenotype correlations in classic galactosemia

In order to evaluate the functional significance of *GALT* mutations, we searched for correlations between genotypes and several phenotypic manifestations of classic galactosemia, namely cataracts, anomalies in motor function, intellectual and learning disabilities, ovarian dysfunction, speech impairment and white matter anomalies.

Genotype-phenotype correlations are difficult to establish since most patients are compound heterozygotes and, for that reason, the resulting phenotype depends on the interactions between products of two different alleles. Moreover, genotype-phenotype correlations usually rely on the knowledge of the residual enzymatic activity of each mutation, determined by heterologous expression of recombinant proteins. Concerning *GALT* mutations, few have already been characterized (47, 49, 64-66) and, accordingly, herein we employed an *in silico* approach to predict the effect of the mutations on GALT functional properties.

The p.Q188R mutation is by far the most common in European populations, and one of the best characterized mutations (10). Homozygous patients for this mutation have undetectable RBC activity, which is consistent with the *in vitro* functional assays that show a complete loss of enzymatic activity in the homodimeric state. Indeed, half of the Portuguese GALT deficient patients are homozygous for the p.Q188R mutation and, as expected, their RBC Gal-1-P values are always in the upper levels of the advocated therapeutic range, and present a severe clinical phenotype with several negative outcomes (Table 2.3). On the other hand, the 14 heterozygous patients for p.Q188R present biochemical and clinical phenotypes strongly dependent on the second mutation. Severe or null mutations, such as p.R80X, p.R148Q, p.S192G and p.R204X,

confirm the severe phenotype; whereas milder mutations, such as p.S135L, alleviate the phenotype and strongly reduce the negative outcomes.

Indeed, some missense mutations appear to partially retain enzymatic activity, being associated with a milder biochemical phenotype, which is evidenced by *in vitro* functional analysis (10). Heterologous expression of p.S135L using a yeast expression system demonstrated ~5% of wild-type activity (47) and all the three patients carrying this mutation present a mild phenotype. Additionally, the p.P185S mutation presents 14% of residual activity when expressed in a yeast system (48); however, all prediction servers estimate a strong impact on GALT activity, due to the local structural perturbations within each monomer that this mutation causes. Actually, the homozygous patient for p.P185S seems to confirm the former data, since he presents some residual activity in RBC and his RBC Gal-1-P values tend to be close to the lower limit of the therapeutic range; and, in addition, this patient presents only slight speech and learning disabilities, which, overall, confers a mild phenotype.

The p.R259W mutation was also identified in a homozygous patient; however, in opposition to p.P185S, biochemical and clinical phenotypes are in agreement with the majority of prediction servers, which propose p.R259W as a destabilizing mutation whose structural perturbations affect GALT activity, thereby causing a severe phenotype.

The p.F171C mutation was identified exclusively in a Portuguese patient, carrying the p.S135L mutation in the other allele. In fact, the rarity of this mutation and the fact that it is in *trans* with a mild mutation make it difficult to estimate if the mild phenotype presented by this patient is due to p.S135L or if p.F171C itself alleviates the phenotype. Moreover, this mutation has never been expressed *in vitro*, and prediction servers estimate an ambiguous effect for this mutation (Table 2.4). Closer inspection of its location, however, suggests it may affect hexose binding and/or disturb the monomer-monomer interaction; additionally, this patient presents a similar phenotype to both patients who present the severe mutation, p.Q188R, *in trans* with the mild mutation p.S135L, which reinforces the idea that p.F171C mutation could be a severe mutation.

The predicted destabilizing effect of p.G175D is contrasted by the variability of the observed phenotypes. Although both homozygous siblings for p.G175D have a normal school career, the sister presents white matter anomalies, while the brother displays no negative outcomes. On the other hand, the patient who carries this mutation in heterozygous state with the p.Q188R mutation presents a severe phenotype, thereby confirming our prediction that the phenotype of heterozygous patients for p.Q188R strongly depends on the second mutation. Indeed, a potential severe mutation such as p.G175D is not capable of alleviating the phenotype and, as a result, emphasizes the severe phenotype conferred by the p.Q188R mutation.

A genotype-phenotype correlation was also attempted on the patients carrying the potential splicing mutations. The effect of the c.328+33G>A variation remained elusive.

Nevertheless, the patient harboring it in heterozygosity with the severe p.Q188R shows a mild clinical phenotype with no negative outcomes, and the control blood Gal-1-P values are always only slightly above the normal range. These data suggest that this intronic mutation should be a mild mutation.

Concerning the c.820+13A>G transition, all bioinformatic tools predict a severe splicing mutation, that induces a frameshift leading to a premature stop codon. Accordingly, it is anticipated that the resulting mRNA should be directed to the nonsense mediated decay system, with no detected enzymatic activity in biological samples. However, despite displaying null GALT activity in RBC, the patient who carries this mutation in homozygous state presents a mild clinical phenotype. No negative outcomes other than moderate learning disabilities were observed, which nevertheless have not hampered the patient in completing a secondary education level. On the other hand, patients carrying this mutation in heterozygosity with the p.Q188R mutation also display mild phenotypes but some negative outcomes, such as speech impediments and learning disabilities. Finally, the patient who carries the c.820+13A>G mutation in heterozygous state with the p.P295T mutation also reveals a mild phenotype. Altogether, these results suggest that c.820+13A>G transition, despite its potential severe disease effect, may alleviate the severe negative outcomes caused by highly deleterious mutations, such as p.Q188R. The mild phenotype afforded by this mutation could be due to tissue availability of specific spliceosomal factors, which may prevent the full expression of the alternative splicing effect.

2.5. Conclusion

The present work expands the knowledge on the mutational spectrum of classic galactosemia in Portugal, and also predicts the impact on the structural and functional effects of mutations identified in Portuguese patients that were not yet characterized. The establishment of potential correlations between genotype and phenotype is not trivial due to several factors. Firstly, *in silico* analysis servers and softwares are rather limited since they evaluate overall structural perturbations, overlooking relevant functional information, such as oligomeric arrangements, cofactor and ligand binding. These limitations were tentatively overcome by closely inspecting the local effects of the mutations and attempting to envisage structural-functional impairments. Secondly, despite being a monogenic disorder, the phenotype of classic galactosemia is not straightforward. There is a paucity of structural information on the Leloir pathway enzymes. It has been hypothesized they could form supra-molecular complexes, which would imply that a change in the GALT protein could affect the entire metabolic pathway (67). Moreover, galactose metabolites are implicated in several physiological pathways, including the intricate glycosylation reactions, which are reflected on many distinct levels. Finally, the

influence of other genetic modifiers, epigenetic alterations, and environmental factors must be considered (68).

2.6. Acknowledgments

The authors wish to acknowledge the patients and families enrolled in this study. This work was supported by SPDM Grant to Isabel Rivera, Fundação para a Ciência e Tecnologia SFRH/BD/48259/2008 PhD Grant to Ana Isabel Coelho, and PEst-OE/SAU/UI4013/2011.

2.7. References

1.Fridovich-Keil J, Walter JH. *Galactosemia*. In: Valle D, Beaudet AL, Vogelstein B, Kinzler KW, Antonarakis SE, Ballabio A, editors. The Online Metabolic and Molecular Bases of Inherited Diseases (OMMBID). Carbohydrates: McGraw Hill; 2008. p. 1-108.

2.Reichardt JK. *Molecular basis of galactosemia: Mutations and polymorphisms in the gene encoding human galactose-1-phosphate uridylyltransferase.* Proc Natl Acad Sci U S A. 1991.

3.Mumma JO, Chhay JS, Ross KL, Eaton JS, Newell-Litwa KA, Fridovich-Keil JL. *Distinct roles of galactose-1P in galactose-mediated growth arrest of yeast deficient in galactose-1P uridylyltransferase (GALT) and UDP-galactose 4'-epimerase (GALE)*. Mol Genet Metab. 2008;93(2):160-71.

4.Fridovich-Keil JL. *Galactosemia: the good, the bad, and the unknown*. J Cell Physiol. 2006;209(3):701-5.

5.Leslie ND, Immerman EB, Flach JE, Florez M, Fridovich-Keil JL, Elsas LJ. *The human galactose-1-phosphate uridyltransferase gene*. Genomics. 1992;14(2):474-80.

6.Reichardt JK, Berg P. *Cloning and characterization of a cDNA encoding human galactose-1-phosphate uridyl transferase*. Mol Biol Med. 1988;5(2):107-22.

7.McCorvie TJ, Timson DJ. *The structural and molecular biology of type I galactosemia: Enzymology of galactose 1-phosphate uridylyltransferase*. IUBMB Life. 2011;63(9):694-700.

8.Marabotti A, Facchiano AM. *Homology modeling studies on human galactose-1-phosphate uridylyltransferase and on its galactosemia-related mutant Q188R provide an explanation of molecular effects of the mutation on homo- and heterodimers*. J Med Chem. 2005;48(3):773-9.

9.Calderon FR, Phansalkar AR, Crockett DK, Miller M, Mao R. *Mutation database for the galactose-1-phosphate uridyltransferase (GALT) gene*. Hum Mutat. 2007;28(10):939-43.

10.Tyfield LA. *Galactosaemia and allelic variation at the galactose-1-phosphate uridyltransferase gene: a complex relationship between genotype and phenotype*. Eur J Pediatr. 2000;159 Suppl 3:S204-7.

11.Gregersen N, Bross P, Jorgensen MM. *Protein folding and misfolding: the role of cellular protein quality control systems in inherited disorders*. In: Valle D, Beaudet AL, Vogelstein B, Kinzler KW, Antonarakis SE, Ballabio A, editors. The Online Metabolic and Molecular Bases of Inherited Disease (OMMBID). General Themes: Mc-Graw Hill; 2005.

12.Flanagan JM, McMahon G, Brendan Chia SH, Fitzpatrick P, Tighe O, O'Neill C, Briones P, Gort L, Kozak L, Magee A, Naughten E, Radomyska B, Schwartz M, Shin JS, Strobl WM, Tyfield LA, Waterham HR, Russell H, Bertorelle G, Reichardt JK, Mayne PD, Croke DT. *The role of human demographic history in determining the distribution and frequency of transferase-deficient galactosaemia mutations*. Heredity (Edinb). 2010;104(2):148-54.

13.Murphy M, McHugh B, Tighe O, Mayne P, O'Neill C, Naughten E, Croke DT. *Genetic basis of transferase-deficient galactosaemia in Ireland and the population history of the Irish Travellers*. Eur J Hum Genet. 1999;7(5):549-54.

14.Padilla CD, Lam ST. *Issues on universal screening for galactosemia*. Ann Acad Med Singapore. 2008;37(12 Suppl):39-3.

15.Tyfield L, Reichardt J, Fridovich-Keil J, Croke DT, Elsas LJ, 2nd, Strobl W, Kozak L, Coskun T, Novelli G, Okano Y, Zekanowski C, Shin Y, Boleda MD. *Classical galactosemia and mutations at the galactose-1-phosphate uridyl transferase (GALT) gene*. Hum Mutat. 1999;13(6):417-30.

16.Sommer M, Gathof BS, Podskarbi T, Giugliani R, Kleinlein B, Shin YS. *Mutations in the galactose-1-phosphate uridyltransferase gene of two families with mild galactosaemia variants*. J Inherit Metab Dis. 1995;18(5):567-76.

17.Lukac-Bajalo J, Kuzelicki NK, Zitnik IP, Mencej S, Battelino T. *Higher frequency of the galactose-1-phosphate uridyl transferase gene K285N mutation in the Slovenian population*. Clin Biochem. 2007;40(5-6):414-5.

18.Crushell E, Chukwu J, Mayne P, Blatny J, Treacy EP. *Negative screening tests in classical galactosaemia caused by S135L homozygosity*. J Inherit Metab Dis. 2009;32(3):412-5.

19.Henderson H, Leisegang F, Brown R, Eley B. *The clinical and molecular spectrum of galactosemia in patients from the Cape Town region of South Africa*. BMC Pediatr. 2002;2:7.

20.Lai K, Langley SD, Singh RH, Dembure PP, Hjelm LN, Elsas LJ, 2nd. *A prevalent mutation for galactosemia among black Americans*. J Pediatr. 1996;128(1):89-95.

21.Suzuki M, West C, Beutler E. *Large-scale molecular screening for galactosemia alleles in a pan-ethnic population*. Hum Genet. 2001;109(2):210-5.

22.Elsas LJ. *Prenatal diagnosis of galactose-1-phosphate uridyltransferase (GALT)-deficient galactosemia*. Prenat Diagn. 2001;21(4):302-3.

23.Woodcock J. *The prospects for "personalized medicine" in drug development and drug therapy*. Clin Pharmacol Ther. 2007;81(2):164-9.

24.Gort L, Quintana E, Moliner S, Gonzalez-Quereda L, Lopez-Hernandez T, Briones P. *An update on the molecular analysis of classical galactosaemia patients diagnosed in Spain and Portugal: 7 new mutations in 17 new families*. Med Clin (Barc). 2009;132(18):709-11.

25.Lindhout M, Rubio-Gozalbo ME, Bakker JA, Bierau J. *Direct non-radioactive assay of galactose-1-phosphate:uridyltransferase activity using high performance liquid chromatography*. Clin Chim Acta. 2010;411(13-14):980-3.

26.Ohlsson A, Guthenberg C, von Döbeln U. *Galactosemia screening with low false-positive recall rate: the Swedish experience*. J Inherit Metab Dis Reports-Case and Research Reports. 2012;2:113-7.

27.Fujimoto A, Okano Y, Miyagi T, Isshiki G, Oura T. *Quantitative Beutler test for newborn mass screening of galactosemia using a fluorometric microplate reader*. Clin Chem. 2000;46(6):806-10.

28.Gitzelmann R. *Estimation of galactose-I-phosphate in erythrocytes: a rapid and simple enzymatic method*. Clin Chim Acta. 1969;26(2):313-6.

29.Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG. *The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools*. Nucleic Acids Res. 1997;25(24):4876-82.

30.Arnold K, Bordoli L, Kopp J, Schwede T. *The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling*. Bioinformatics. 2006;22(2):195-201.

31.Kiefer F, Arnold K, Kunzli M, Bordoli L, Schwede T. *The SWISS-MODEL Repository and associated resources*. Nucleic Acids Res. 2009;37(no. suppl 1):D387-92.

32.Thoden JB, Ruzicka FJ, Frey PA, Rayment I, Holden HM. *Structural analysis of the H166G site-directed mutant of galactose-1-phosphate uridylyltransferase complexed with either UDP-glucose or UDP-galactose: detailed description of the nucleotide sugar binding site*. Biochemistry. 1997;36(6):1212-22.

33.Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, Sunyaev SR. *A method and server for predicting damaging missense mutations*. Nat Methods. 2010;7(4):248-9.

34.Worth CL, Bickerton GR, Schreyer A, Forman JR, Cheng TM, Lee S, Gong S, Burke DF, Blundell TL. *A structural bioinformatics approach to the analysis of nonsynonymous single nucleotide polymorphisms (nsSNPs) and their relation to disease*. J Bioinform Comput Biol. 2007;5(6):1297-318.

35.Dehouck Y, Kwasigroch JM, Gilis D, Rooman M. *PoPMuSiC 2.1: a web server for the estimation of protein stability changes upon mutation and sequence optimality*. BMC Bioinformatics. 2011;12:151.

36.Gonnelli G, Rooman M, Dehouck Y. *Structure-based mutant stability predictions on proteins of unknown structure*. J Biotechnol. 2012;161(3):287-93.

37.Parthiban V, Gromiha MM, Abhinandan M, Schomburg D. *Computational modeling of protein mutant stability: analysis and optimization of statistical potentials and structural features reveal insights into prediction model development*. BMC Struct Biol. 2007;7:54.

38.Parthiban V, Gromiha MM, Schomburg D. *CUPSAT: prediction of protein stability upon point mutations*. Nucleic Acids Res. 2006;34(Web Server issue):W239-42.

39.Brunak S, Engelbrecht J, Knudsen S. *Prediction of human mRNA donor and acceptor sites from the DNA sequence*. J Mol Biol. 1991;220(1):49-65.

40.Hebsgaard SM, Korning PG, Tolstrup N, Engelbrecht J, Rouze P, Brunak S. *Splice site prediction in Arabidopsis thaliana pre-mRNA by combining local and global sequence information*. Nucleic Acids Res. 1996;24(17):3439-52.

41.Reese MG, Eeckman FH, Kulp D, Haussler D. *Improved splice site detection in Genie*. J Comput Biol. 1997;4(3):311-23.

42.Fairbrother WG, Yeh RF, Sharp PA, Burge CB. *Predictive identification of exonic splicing enhancers in human genes*. Science. 2002;297(5583):1007-13.

43.Cartegni L, Wang J, Zhu Z, Zhang MQ, Krainer AR. *ESEfinder: A web resource to identify exonic splicing enhancers*. Nucleic Acids Res. 2003;31(13):3568-71.

44.Smith PJ, Zhang C, Wang J, Chew SL, Zhang MQ, Krainer AR. *An increased specificity score matrix for the prediction of SF2/ASF-specific exonic splicing enhancers*. Hum Mol Genet. 2006;15(16):2490-508.

45.Gort L, Boleda MD, Tyfield L, Vilarinho L, Rivera I, Cardoso ML, Santos-Leite M, Giros M, Briones P. *Mutational spectrum of classical galactosaemia in Spain and Portugal*. J Inherit Metab Dis. 2006;29(6):739-42.

46.Fridovich-Keil JL, Langley SD, Mazur LA, Lennon JC, Dembure PP, Elsas II LJ. *Identification and functional analysis of three distinct mutations in the human galactose-1-phosphate uridyltransferase gene associated with galactosemia in a single family*. Am J Hum Genet. 1995;56(3):640-6.

47.Wells L, Fridovich-Keil JL. *Biochemical characterization of the S135L allele of galactose-1-phosphate uridylyltransferase associated with galactosaemia*. J Inherit Metab Dis. 1997;20(5):633-42.

48.Quimby BB, Wells L, Wilkinson KD, Fridovich-Keil JL. *Functional requirements of the active site position 185 in the human enzyme galactose-1-phosphate uridylyltransferase*. J Biol Chem. 1996;271(43):26835-42.

49.Chhay JS, Openo KK, Eaton JS, Gentile M, Fridovich-Keil JL. *A yeast model reveals biochemical severity associated with each of three variant alleles of galactose-1P uridylyltransferase segregating in a single family*. J Inherit Metab Dis. 2008;31(1):97-107.

50.Riehman K, Crews C, Fridovich-Keil JL. *Relationship between genotype, activity, and galactose sensitivity in yeast expressing patient alleles of human galactose-1-phosphate uridylyltransferase*. J Biol Chem. 2001;276(14):10634-40.

51.Crews C, Wilkinson KD, Wells L, Perkins C, Fridovich-Keil JL. *Functional consequence of substitutions at residue 171 in human galactose-1-phosphate uridylyltransferase*. J Biol Chem. 2000;275(30):22847-53.

52.Schweitzer-Krantz S. *Early diagnosis of inherited metabolic disorders towards improving outcome: the controversial issue of galactosaemia*. Eur J Pediatr. 2003;162(1):S50-3.

53.Arnaiz-Villena A, Martinez-Laso J, Gomez-Casado E, Diaz-Campos N, Santos P, Martinho A, Breda-Coimbra H. *Relatedness among Basques, Portuguese, Spaniards, and Algerians studied by HLA allelic frequencies and haplotypes*. Immunogenetics. 1997;47(1):37-43.

54.Perez B, Desviat LR, Ugarte M. *Analysis of the phenylalanine hydroxylase gene in the Spanish population: mutation profile and association with intragenic polymorphic markers*. Am J Hum Genet. 1997;60(1):95-102.

55.Rivera I, Leandro P, Lichter-Konecki U, Tavares de Almeida I, Lechner MC. *Population genetics of hyperphenylalaninaemia resulting from phenylalanine hydroxylase deficiency in Portugal*. J Med Genet. 1998;35(4):301-4.

56.Fridovich-Keil JL, Jinks-Robertson S. *A yeast expression system for human galactose-1-phosphate uridylyltransferase*. Proc Natl Acad Sci U S A. 1993;90(2):398-402.

57.Lai K, Elsas LJ. *Structure-function analyses of a common mutation in blacks with transferase-deficiency galactosemia*. Mol Genet Metab. 2001;74(1-2):264-72.

58.Geeganage S, Frey PA. *Transient kinetics of formation and reaction of the uridylyl-enzyme form of galactose-1-P uridylyltransferase and its Q168R-variant: insight into the molecular basis of galactosemia*. Biochemistry. 1998;37(41):14500-7.

59.Saul B, Kuner T, Sobetzko D, Brune W, Hanefeld F, Meinck HM, Becker CM. *Novel GLRA1 missense mutation (P250T) in dominant hyperekplexia defines an intracellular determinant of glycine receptor channel gating*. J Neurosci. 1999;19(3):869-77.

60.Yang G, Xiong C, Li S, Wang Y, Zhao J. *A recurrent mutation in CRYGD is associated with autosomal dominant congenital coralliform cataract in two unrelated Chinese families*. Mol Vis. 2011;17:1085-9.

61.Gregoret LM, Rader SD, Fletterick RJ, Cohen FE. *Hydrogen bonds involving sulfur atoms in proteins*. Proteins. 1991;9(2):99-107.

62.Elsas II LJ, Webb AL, Langley SD. *Characterization of a carbohydrate response element regulating the gene for human galactose-1-phosphate uridyltransferase*. Mol Genet Metab. 2002;76(4):287-96.

63.DeLano WL. *PyMOL: an open-source molecular graphics tool*. 2002.

64.Shield JP, Wadsworth EJ, MacDonald A, Stephenson A, Tyfield L, Holton JB, Marlow N. *The relationship of genotype to cognitive outcome in galactosaemia*. Arch Dis Child. 2000;83(3):248-50.

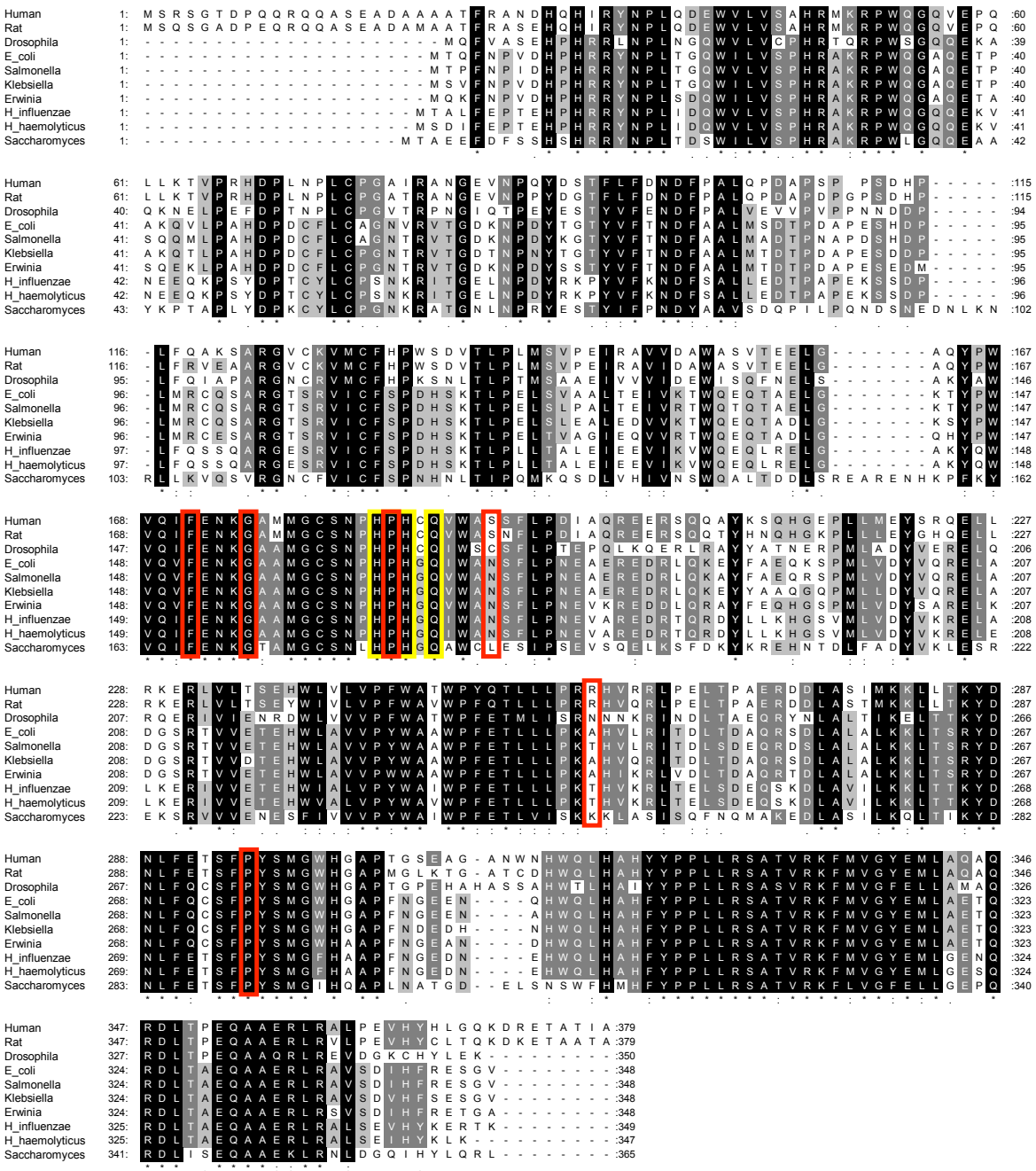
65.Riehman K, Crews C, Fridovich-Keil JL. *Relationship between genotype, activity, and galactose sensitivity in yeast expressing patient alleles of human galactose-1-phosphate uridylyltransferase*. Journal of Biological Chemistry. 2001;276(14):10634-40.

66.Fridovich-Keil JL, Langley SD, Mazur LA, Lennon JC, Dembure PP, Elsas JL, 2nd. *Identification and functional analysis of three distinct mutations in the human galactose-1-phosphate uridyltransferase gene associated with galactosemia in a single family*. Am J Hum Genet. 1995;56(3):640-6.

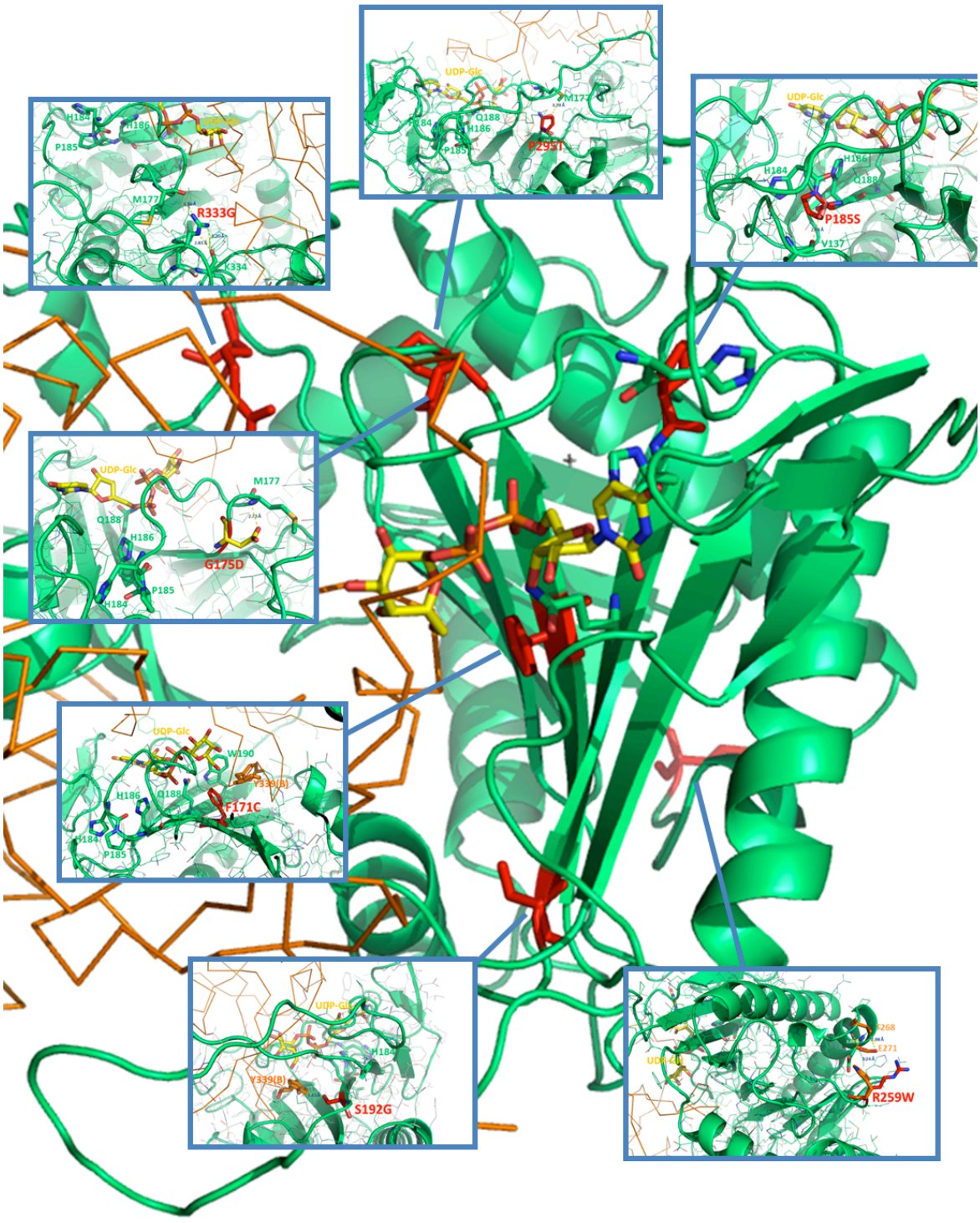
67.McCorvie TJ, Timson DJ. *Structural and molecular biology of type I galactosemia: disease-associated mutations*. IUBMB Life. 2011;63(11):949-54.

68.Waisbren SE, Potter NL, Gordon CM, Green RC, Greenstein P, Gubbels CS, Rubio-Gozalbo E, Schomer D, Welt C, Anastasoaie V, D'Anna K, Gentile J, Guo CY, Hecht L, Jackson R, Jansma BM, Li Y, Lip V, Miller DT, Murray M, Power L, Quinn N, Rohr F, Shen Y, Skinder-Meredith A, Timmers I, Tunick R, Wessel A, Wu BL, Levy H, Elsas II LJ, Berry GT. *The adult galactosemic phenotype*. J Inherit Metab Dis. 2012;35(2):279-86.

2.8. Supplementary Material



Supplementary Figure 2.1. Sequence alignment of GALT proteins. Protein sequences retrieved from NCBI BLAST and aligned with Clustal X for Windows. GALT protein sequence accession numbers: Human (SwissProt P079023); Rat, *Rattus norvegicus* (SwissProt Q03249.3); Drosophila, *Drosophila melanogaster* (SwissProt Q9VMA2.2); E_coli, *Escherichia coli* str. K12 substr. MG1655 (SwissProt P09148); Salmonella, *Salmonella enterica* subsp. *enterica* (NCBI ZP_03217106.1); Klebsiella, *Klebsiella pneumoniae* subsp. *pneumoniae* NTUH-K2044 (NCBI YP_006637266.1); Erwinia, *Erwinia billingiae* Eb661 (NCBI YP_003740707.1); H_influenziae, *Haemophilus influenziae* Rd KW20 (SwissProt P31764.2); H_haemolyticus; *Haemophilus haemolyticus* M19107 (GenBank EGT77758.1); Saccharomyces, *Saccharomyces cerevisiae* (GenBank AAA34627.1). Yellow boxes highlight the active site residues His_{184}-Pro_{185}-His_{186} and the substrate stabilizing residue Q188; red boxes highlight the mutations herein analyzed in detail.



Supplementary Figure 2.2. Structural model of human GALT with highlighted mutations. Structural model of human GALT (green cartoon) and a monomer of *E. coli* GalT (orange ribbon, PDB code 1GUP, B chain). Red sticks, GALT mutations herein studied. Green sticks, His_{184}-Pro_{185}-His_{186} active site and Q188 residue (human GALT numbering), which stabilizes bound UDP-glucose (yellow sticks). Inset panels highlight structural effects of studied mutations. Figure generated with PyMOL.

CHAPTER 3

Antisense therapy for classic galactosemia: functional correction of a splicing mutation in the *GALT* gene

Ana I. Coelho[1,2], Sílvia Lourenço[1,2], Matilde Trabuco[1,2], Maria João Silva[1,2], Anabela Oliveira[3], Ana Gaspar[4], Luísa Diogo[5], Isabel Tavares de Almeida[1,2], João B. Vicente[1,2], Isabel Rivera[1,2]

[1] Metabolism & Genetics Group, Research Institute for Medicines and Pharmaceutical Sciences (*i*Med.UL), Faculty of Pharmacy, University of Lisbon, Portugal

[2] Department of Biochemistry and Human Biology, Faculty of Pharmacy, University of Lisbon, Portugal

[3] Department of Medicine, Hospital Santa Maria, Lisbon, Portugal

[4] Department of Pediatrics, Hospital Santa Maria, Lisbon, Portugal

[5] Metabolic Clinics, Pediatric Hospital, CHUC, Coimbra, Portugal

European Journal Human Genetics
(submitted)

3.1. Abstract

In recent years, antisense therapy has emerged as an increasingly important therapeutic approach to tackle several genetic disorders, particularly inherited metabolic disorders. Intronic mutations activating cryptic splice sites are particularly amenable to antisense therapy, since the canonical splice sites remain intact, thus retaining the potential for restoring constitutive splicing. Mutational analysis of Portuguese galactosemic patients revealed the intronic variation c.IVS8+13A>G as the second most prevalent mutation, strongly suggesting its pathogenicity. The aim of this study was to functionally characterize this intronic variation, to elucidate its pathogenic molecular mechanism(s) and, ultimately, to correct it by antisense therapy. Minigene splicing assays in two distinct cell lines and patients transcript analyses showed that the mutation activates a cryptic donor splice site, inducing an aberrant splicing of the *GALT* pre-mRNA, which in turn leads to a frameshift with inclusion of a premature stop codon (p.D274GfsX291). Functional-structural studies of the recombinant wild-type and truncated GALT showed that the latter is devoid of enzymatic activity and prone to aggregation. Finally, two LNA oligonucleotides, designed to specifically recognize the mutation, successfully restored the constitutive splicing, thus establishing a proof-of-concept for the application of antisense therapy as an alternative strategy for the clearly insufficient dietary treatment in classic galactosemia.

3.2. Introduction

During the last years, splicing mutations have emerged as an important pathogenic mechanism, underlying 10-30% of genetic diseases (HGMD® Professional 2013.1) (1). Splicing accuracy depends on the recognition of exon–intron junctions, which are defined by intronic *cis*-elements: 5′ splice site, 3′ splice site, branch site and poly-pyrimidine tract (2-4). Even though these are essential elements, splicing also requires more discrete elements, entitled splicing regulatory elements (SRE). Indeed, several *cis*-regulatory sequences direct the splicing machinery to use the correct splice sites, which include exonic/intronic splicing enhancers (ESE/ISE) and silencer elements (ESS/ISS). Exonic splicing enhancers (ESE) and intronic splicing enhancers (ISE) stimulate splicing and serve as binding sites mainly for serine/arginine-rich proteins (SR proteins). Exonic splicing silencers (ESS) and intronic splicing silencers (ISS) repress splicing, and often function by binding of proteins from the heterogenous nuclear ribonucleoprotein (hnRNP) family (2, 4-6). Although most reported splicing mutations directly abolish an authentic splice site or create a novel one, an increasing number of disease-associated mutations that alter splicing enhancers or silencers have been reported (2, 7-9). Indeed, each nucleotide modification should be considered a potential candidate for splicing alterations since

not only intronic but also nonsense, missense and silent modifications may impact splicing (10). Accordingly, constitutive and regulated splicing reactions are considered potential therapeutic targets and novel strategies for their correction are evolving. Among these novel approaches, antisense oligonucleotides must be highlighted, since many are in phase II/III of clinical trials, namely AVI4658 (Eteplirsen®), a morpholino oligonucleotide for Duchenne muscular dystrophy (11). Furthermore, one of the hallmarks of antisense oligonucleotides is their exquisite specificity, being capable of distinguishing a single nucleotide mismatch (12).

Classic galactosemia (OMIM #230400), which affects about 1/47,000 live-births, results from deficient activity of galactose-1-phosphate urydylyltransferase (GALT, EC 2.7.7.12). This is a key enzyme in galactose metabolism, constituting an essential route for nursing infants, as lactose represents their primary carbohydrate source (13, 14). The present gold standard of care of galactosemia patients is a lifelong dietary galactose restriction which, however, has proven to be inadequate to prevent the long-term sequelae (13-17). The disorder is caused by mutations in the *GALT* gene, which profoundly impair GALT enzymatic activity (14, 18, 19). The *GALT* gene is located in chromosome 9p13, is arranged into 11 exons spanning about 4.0 kb of genomic sequence, and encodes a 379 amino acid protein, which is assembled as a ~87 kDa homodimer (19-22).

As many other autosomal recessive metabolic disorders, classic galactosemia displays great allelic heterogeneity with more than 260 variations described (available at: http://www.arup.utah.edu/database/GALT/GALT_display.php, last surveyed November 2013) (23). Although most are missense mutations, other variations have been reported, namely silent, nonsense and noncoding changes (18, 24, 25). Moreover, several intronic mutations have been identified in the *GALT* gene, approximately half of which are known to affect splicing (23). Mutational analysis of 42 Portuguese patients confirmed p.Q188R as the prevalent molecular defect (67.1%), and surprisingly revealed an intronic variation, IVS8+13A>G (c.820+13A>G), as the second most frequent mutation, accounting for 8.0% of the mutant alleles (26). Although IVS8+13A>G transition is currently classified as benign, complete *GALT* sequence of a galactosemic patient homozygous for this variation revealed no other alteration in his genomic DNA (23). Furthermore, this variation has never been identified in controls and was found in Portuguese patients either in homozygous or compound heterozygous state (26).

The aim of this study is to confirm that the IVS8+13A>G transition is indeed a disease-causing mutation, to elucidate its pathogenic mechanism, by *in vitro*, *ex vivo* and *in vivo* approaches, to functionally characterize the resulting protein and, finally, to use antisense oligonucleotides, namely locked nucleic acids (LNA), to modulate the deleterious effect of the mutation.

3.3. Materials and Methods

3.3.1. Patients

This study was approved by the local Ethics Committees and written informed consents were obtained from patients and family members enrolled in this study. The IVS8+13A>G variation was identified in six galactosemic patients: one homozygous for this variation and five compound heterozygotes, four bearing the p.Q188R mutation, and one the p.P295T mutation (26).

The clinical presentation was common to all patients, and included poor feeding, vomiting, weight loss, jaundice and hepatomegaly. Only the younger brother of one of the patients was diagnosed pre-symptomatically and put on a galactose-restricted diet immediately after birth. At present, all patients are under dietary treatment, with calcium and vitamin D supplementation.

3.3.2. *In vivo* splicing analysis

Total RNA was isolated from lymphocytes of homozygous and compound heterozygous patients, and of control wild-type *GALT* homozygous individuals using the TRIZOL reagent (Invitrogen Corporation, Carlsbad, CA, USA). RNA was reverse transcribed using the NZY First-Strand cDNA Synthesis Kit (NZYTech, Lisbon, Portugal) and PCR analysis was performed using the CF and CR primers (Table 3.1) with the following cycle conditions: 5 min at 94 °C, 1 min at 60 °C, 2 min at 72 °C, 30 cycles of 40 sec at 94 °C, 40 sec at 60 °C, 90 sec at 72 °C, followed by 7 min at 72 °C. After separation by 2 % NuSieve GTG agarose gel electrophoresis, PCR products were purified (Isolate PCR and Gel Kit, Bioline, London, UK), and further analyzed by direct sequencing analysis using the ABI Prism BigDye Terminator Cycle Sequencing Ready Reaction Kits, in an ABI PRISM 310 Genetic Analyzer (Applied Biosystems, Foster City, CA, USA).

3.3.3. Minigenes construction

Genomic fragments encompassing the region of the *GALT* gene containing the IVS8+13A>G variation were PCR-amplified from one heterozygous patient carrying this mutation, using the primers 7F and 9R (Table 3.1). The fragment includes the last portion of intron 7, exon 8, intron 8, exon 9, and the initial portion of intron 9, and is 399 bp long. Both wild-type and mutant fragments were cloned into the pCR®2.1TOPO vector (Invitrogen Corporation, Carlsbad, CA, USA), and positively selected by blue/white screening. Fragments

were then subcloned into the exon-trapping vector pSPL3, kindly provided by Prof. Bélen Pérez (Centro de Biologia Molecular Severo Ochoa, Universidad Autónoma de Madrid, Madrid, Spain). Clones bearing the wild-type and mutant inserts were analyzed by direct sequencing for confirmation of fragment insertion and evaluation of their correct orientation.

Table 3.1. Sequence of DNA oligonucleotides used in this study.

Oligonucleotides	Sequence (5' → 3')
CF	GTG AGG AGC GAT CTC AGC A
CR	GGA GCG GAG GGT AGT AAT GA
7F	CAC CTT GAT GAC TTC CTA TCC
9R	GAA ATG GTG TTG GGG CTA AA
SD6	TCT GAG TCA CCT GGA CAA CC
SA2	GCT CAC AAA TAC CAC TGA GAT
6His-F [a]	CCA GCG GAT CCC CCT CAA AAA TGC ATC ATC ACC ATC ACC ACA TGT CGC GCA GTG GAA CCG ATC
6His-R [a]	GAT CGG TTC CAC TGC GCG ACA TGT GGT GAT GGT GAT GAT GCA TTT TTG AGG GGG ATC CGC TGG
IVS8-Mut-F [b]	CCC CTG CTG AGC GTG ATG GTC AGT CTC CCA GAT CTA GCC TCC ATC ATG
IVS8-Mut-R [b]	CAT GAT GGA GGC TAG ATC TGG GAG ACT GAC CAT CAC GCT CAG CAG GGG

[a] underlined residues encode the hexa-histidyl tag
[b] underlined residues represent the first 13 nucleotides of intron 8

3.3.4. *Ex vivo* splicing analysis

Empty pSPL3 vector and minigene constructs containing wild-type (pSPL3.wt) and mutant sequences (pSPL3.mut) were transfected into eukaryotic cell lines (HeLa and COS-7) using Lipofectamine 2000 (Invitrogen Corporation, Carlsbad, CA, USA), according to the manufacturer's instructions. Twenty-four hours after transfection, total RNA was extracted using the TRIZOL reagent (Invitrogen Corporation, Carlsbad, CA, USA), reverse transcribed using the NZY First-Strand cDNA Synthesis Kit (NZYTech) and PCR reaction was performed using the pSPL3-specific primers SD6 and SA2 (Table 3.1) with the following cycle conditions: 40 sec at 95 °C, 40 sec at 55 °C, 90 sec at 72 °C and a final extension for 10 min at 72 °C. PCR products were separated by 2 % Nusieve GTG agarose gel electrophoresis, purified and analyzed by direct sequencing analysis.

3.3.5. Correction of alternative splicing with LNA oligonucleotides

Two locked nucleic acid (LNA) oligonucleotides were designed and synthesized and purified by Exiqon (Vedbaek, Denmark). Antisense oligonucleotide IVS8-LNA1 is a

15-mer (5'-CCAGGATCCTACCTG-3'), whereas IVS8-LNA2 is a 16-mer (5'-GATCCTACCTGGGAGA/3Phos); the mutant nucleotide is underlined.

Cell cultures and minigene transfection followed the previously referred protocol for *ex vivo* splicing analysis. Five hours after minigene transfection, the antisense oligonucleotides IVS8-LNA1, IVS8-LNA2 or saline solution (as control) were transfected using Lipofectamine 2000, as previously mentioned. We used two different concentration ranges, namely low concentrations (5-150 nM) and high concentrations (0.5-1.0 μM). Conventional RT-PCR for splicing analysis was performed after 24 hours, using the pSPL3-specific primers SD6 and SA2 (Table 3.1), as described above. PCR products were analyzed by agarose gel electrophoresis and further sequenced.

3.3.6. Production of recombinant wild-type and mutant GALT

GALT cDNA, a generous gift of Prof. Judith Fridovich-Keil (Department of Human Genetics, Emory University, USA) was cloned into the pET24b(+) plasmid. The T7 tag was deleted using the NdeI and EcoRI enzymes; subsequently, *GALT* cDNA was cloned using the enzymes HindIII and SalI. By site-directed mutagenesis (QuikChange® II XL Mutagenesis kit, Stratagene) using the primers 6His-F and 6His-R (Table 3.1), a hexa-histidyl tag was introduced at the N-terminus. All introduced changes were confirmed by direct sequencing in both forward and reverse orientations, using the universal primers T7 promoter (#69348-3) and T7 terminator (#69337-3). The resultant recombinant vector was named pET24.6His.GALT.

The IVS8+13A>G mutation was generated by site directed mutagenesis (NZY mutagenesis kit, NZYTech) using the IVS8-Mut-F and IVS8-Mut-R primers (Table 3.1), and confirmed by direct sequencing in both forward and reverse orientations. The resultant recombinant vector was named pET24.6His.GALT.p.IVS8. The vectors bearing the wild-type or mutant *GALT* were transformed into *E. coli* BL21(DE3) Rosetta cells, which were grown in M9 minimal medium (27), containing kanamycin (25 μg/mL) and chloramphenicol (34 μg/mL). At an $O.D._{600}$ of 0.3, isopropyl-D-thiogalactoside (IPTG) was added to a final concentration of 400 μM to induce GALT expression, and, simultaneously, medium was supplemented with 100 μM ferrous ammonium sulfate and 100 μM zinc sulfate. Culture was incubated for an additional 4 hours, at 21 °C, 120 rpm, at which point, cells were collected by centrifugation at 8,000 xg for 5 min, at 4 °C.

The bacterial pellets were resuspended in buffer A (50 mM Tris pH 7.5, 300 mM KCl, and 10 % glycerol) containing 1 mM phenylmethanesulfonyl fluoride (PMSF) and 1 mg/mL lysozyme. After a 30-min incubation at 4 °C, cells were disrupted by sonication (three 60 sec pulses, 50% efficiency, with 30 sec gaps in between for cooling). The fusion proteins expressed

in *E. coli* were purified by immobilized metal affinity chromatography (IMAC) using a HisTrap FF column (Amersham, GE Healthcare). After cell disruption and centrifugation at 13,000 xg, for 5 min, at 4 °C, imidazole was added at a final concentration of 20 mM to the supernatant cell lysates, which was loaded into a HisTrap FF column pre-equilibrated with five column volumes (Vc) of buffer A with 20 mM imidazole. The column was then washed with ten Vc of buffer A containing 20 mM imidazole, and with five Vc of buffer A with 50 mM imidazole. Pure proteins were eluted with 2.5 mL of buffer A containing 500 mM imidazole, which was subsequently eliminated with a PD-10 desalting column (Amersham, GE Healthcare). Protein solutions were concentrated by ultrafiltration using Vivaspin 15R (Sartorius Stedim, Germany) 30 kDa-cutoff filters, at 3,000 xg and 4 °C in a swinging-bucket centrifuge. Immunoblotting and GALT enzymatic activity determination were performed on the same day as purification; the remaining samples were aliquoted, flash-frozen in liquid nitrogen and stored at -80 °C.

3.3.7. Immunoblotting

The purified proteins were run on a 12.5 % SDS-PAGE and transferred into nitrocellulose membrane (Hybond ECL, Amersham, GE Healthcare). After a 2-hour blocking with 5 % milk in PBS-T buffer, the membrane was incubated for 1 hour at room temperature with the primary mouse monoclonal anti-GALT antibody (sc-365577, Santa Cruz Biotechnology, Santa Cruz, CA, USA) at a 1:1000 dilution. After three washes with PBS-T buffer, the membrane was incubated for 1 hour at room temperature with the secondary peroxidase-conjugated Affinipure goat anti-mouse IgG (H + L) antibody (Jackson ImmunoResearch Laboratories, West Grove, PA, USA) at a 1:1000 dilution. The ECL Prime Western Blotting Detection Reagent was used for protein detection (Amersham, GE Healthcare).

3.3.8. GALT enzymatic activity

GALT enzymatic activity was measured as previously described (28). All assays were carried out at 37 °C, in a reaction mixture containing 40 mM Tris–HCl, pH 7.5, 2 mM Gal-1-P, 0.5 mM UDP-Glc, 40 µM dithiothreitol (DTT) and 125 mM glycine. UDP-Glc and UDP-Gal were separated by HPLC and analyzed by UV detection at 262 nm (26, 28).

3.3.9. Far-UV circular dichroism

Far-UV circular dichroism (far-UV CD) spectra and thermal denaturation profiles were recorded in a Jasco J-710 spectropolarimeter, coupled to a Jasco PTC-348WI Peltier temperature

controller and a Haake G/D8 water bath. GALT protein samples were at 0.15 mg/mL (wild-type) or 0.25 mg/mL (p.IVS8), each spectrum being the result of six accumulations at a 50 nm/min scan rate. Thermal denaturation profiles were monitored at 222 nm in a 0.1 cm light path cuvette, in the 20-90 °C temperature range (1 °C /min slope; data pitch: 1°C; delay time: 0 sec). Temperature scan curves were fit according to a two-state model.

3.3.10. Differential scanning fluorimetry

Differential scanning fluorimetry (DSF) was performed in a C1000 Touch thermal cycler equipped with a CFX96 optical reaction module (Bio-Rad, Hercules, CA, USA).

All fluorescence measurements were performed at a final protein concentration of 0.1 mg/mL in buffer A, and SYPRO orange (Invitrogen Corporation, Carlsbad, CA, USA) at a 5x working concentration, in a 50 µL total volume. The PCR plate was sealed with Optical-Quality Sealing Tape (Bio-Rad) and centrifuged at 400 xg for 1 min. DSF assays were carried out with a 10-min incubation step at 20 °C followed by ramping the temperature from 20 °C to 90 °C at 1 °C /min, with a 1 sec hold time every 0.2 °C and fluorescence acquisition using the HEX channel. Data were processed using CFX Manager software V3.0 (Bio-Rad). Temperature scan curves were fitted to a biphasic sigmoidal function and the T_m values were obtained from the midpoint of the first and second transitions.

3.3.11. Dynamic light scattering

Dynamic light scattering (DLS) analysis was performed on a ZetaSizer Nano-S (Malvern Instrument, UK) particle size analyzer, coupled to a precision Peltier temperature control unit; a He-Ne laser was used as the light source (633 nm). Prior to analysis, samples were centrifuged at 20,000 xg for 30 min at 4 °C, diluted in buffer A to a final concentration of 0.15 mg/mL, and filtered with a 0.22 µm membrane in order to remove larger aggregates. Data were processed using Zetasizer Nano DTS software v7.01 (Malvern Instrument). Temperature was ramped from 20 °C to 70 °C at 0.5 °C/min, with the particle size average, distribution and total scattering intensity being collected. The melting temperature (T_{agg}), at which both size and intensity start to increase significantly, was determined by fitting the obtained data to a plateau followed by one phase association equation. The kinetics of thermal aggregation was monitored at 37 °C and 42 °C for 60 min.

3.4. Results

3.4.1. *Ex vivo* analysis revealed the intronic mutation IVS8+13A>G is sufficient to cause aberrant splicing of the GALT transcript

Since RNA samples from the patients were initially not available, the mutation characterization started by minigene analysis. To investigate if the intronic mutation IVS8+13A>G was sufficient to cause aberrant splicing, we cloned the previously referred 399 bp fragment containing the wild-type or the mutant regions of the *GALT* gene into the pSPL3 exon-trapping vector. Minigene constructs, pSPL3.wt and pSPL3.mut, differed exclusively in the 13th nucleotide of intron 8 (Figure 3.1A). Eukaryotic HeLa and COS-7 cell lines were transiently transfected with each construct and with empty vector, and the resulting splicing products were analyzed by direct sequencing. Transfection with the empty vector pSPL3 showed a single 263 bp fragment resulting from the vector canonical splicing sites. Concerning the wild-type and mutant constructs, both cell lines revealed a different splicing pattern: whereas the COS-7 cell line exhibited a complex transcript profile, the HeLa cell line displayed a much simpler splicing profile.

Transfection of HeLa cells with the wild-type minigene pSPL3.wt showed a single 396 bp fragment, resulting from the inclusion of exon 8 and exclusion of exon 9; and transfection with the mutant minigene pSPL3.mut originated a single 409 bp fragment, also resulting from the inclusion of exon 8 and exclusion of exon 9, including however the first 13 nucleotides of intron 8 (Figure 3.1B).

Transfection of COS-7 cells with the wild-type and mutant minigenes revealed several fragments, corresponding to the usage of all available splicing sites, including a cryptic donor site in the vector sequence (data not shown). Nevertheless, comparing with the wild-type minigene, transcript profiles of the mutant minigene construct were clearly different, since all transcripts presented the additional first 13 nucleotides of intron 8 and the canonical donor splice site was never used in any of the splicing events.

Results obtained in two different eukaryotic cell lines showed that the IVS8+13A>G mutation *per se* activates a cryptic donor splice site (c.820+14_15), which leads to the inclusion of the first 13 nucleotides of intron 8 in the resulting mature mRNA, being thus responsible for inducing an aberrant splicing of the *GALT* transcript.

3.4.2. *In vivo* analysis confirmed that IVS8+13A>G is a disease-causing mutation

Classic galactosemic patients carrying the IVS8+13A>G mutation were selected for transcript analysis. Lymphocytes were isolated from two patients, one homozygous and one

heterozygous, as well as from controls, and total RNA was extracted. RT-PCR analysis of the *GALT* cDNA revealed the presence of two fragments when analyzing either control individuals or the homozygous patient, and the presence of four fragments when analyzing the heterozygous patient.

Direct sequencing analysis of the expected fragment – 378 bp – of the control individuals revealed a splicing reaction using the canonical splice sites and thus including the end of exon 7, entire exons 8 and 9 and the beginning of exon 10 (Figure 3.2). However, the smallest fragment – 294 bp – resulted from an alternative splicing event, with skipping of exon 9, most likely derived from the weakness of the 3' acceptor site in intron 8. When analyzing the homozygous patient' fragments (391 and 307 bp), we confirmed they only differ from those of control individuals by the additional 13 bp corresponding to the first nucleotides of intron 8, and resulting from transcription of the allele carrying the IVS8+13A>G transition (Figure 3.2).

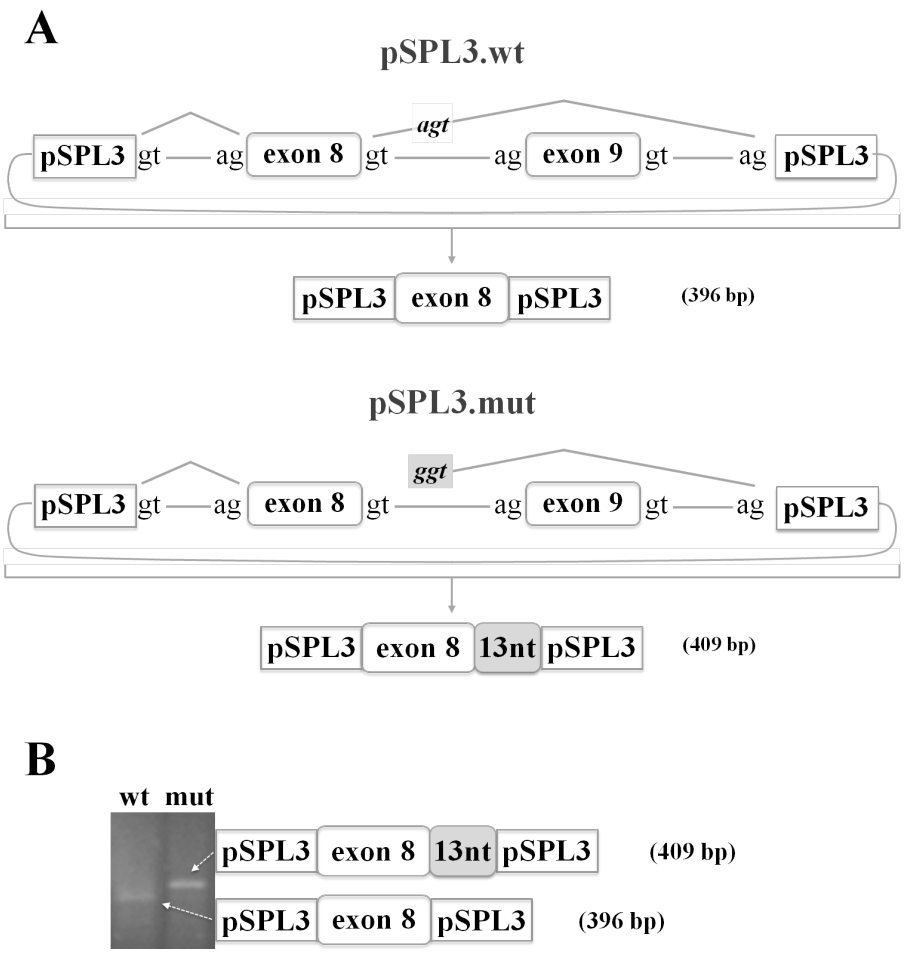


Figure 3.1. Minigene constructs and transcript analysis of transfected HeLa cells. Panel A, schematic representation depicting the minigene constructs. The PCR fragment comprising the last portion of intron 7, exon 8, intron 8, exon 9 and the first portion of intron 9 was cloned in the pSPL3 vector: pSPL3.wt minigene presents the genomic wild-type fragment; pSPL3.mut minigene presents the genomic mutant fragment (IVS8+13A>G mutation underlined); Panel B, RT-PCR pattern and sequence analysis of the corresponding fragment: pSPL3.wt fragment (396 bp) resulted from the inclusion of exon 8 and skipping of exon 9; pSPL3.mut fragment (409 bp) resulted from the inclusion of exon 8 and the first 13 nucleotides of intron 8, and skipping of exon 9.

Concerning the analysis of the heterozygous patient, we observed the presence of the four previously mentioned fragments corresponding to both alleles, either not carrying or carrying the splicing mutation (data not shown). Moreover, we determined that IVS8+13A>G mutation is not responsible for the alternative splicing causing exon 9 skipping.

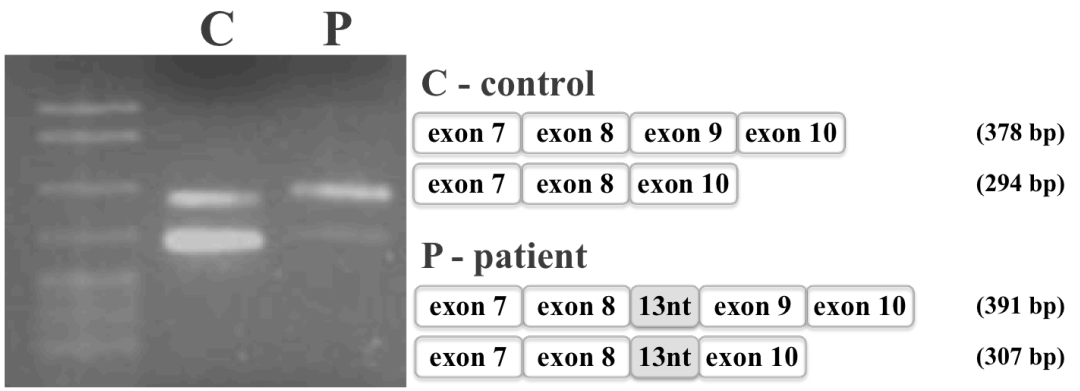


Figure 3.2. Transcript analysis of control and patient's lymphocytes confirm different splicing patterns. Total RNA was isolated from lymphocytes of control and a homozygous patient with the IVS8+13A>G mutation. RT-PCR analysis of *GALT* mRNA revealed the presence of two fragments: control showed the expected fragment (378 bp) and a smaller fragment (294 bp) that resulted from an alternative splicing event with skipping of exon 9; homozygous patient also showed the two fragments, although both presented the first 13 nucleotides of intron 8 (391 bp and 307 bp).

3.4.2. The use of antisense therapy allowed the *ex vivo* reversion of the alternative splicing caused by IVS8+13A>G mutation

To investigate the possibility of restoring correct splicing by antisense therapy, two LNA oligonucleotides were designed. Both oligonucleotides were targeted to the cryptic splice site, sterically blocking the access of the spliceosome machinery to this site, and thus forcing the usage of the canonical donor splice site in the vicinity. HeLa and COS-7 cell lines were transfected with both minigene constructs and LNAs at two different concentration ranges. Transcript analysis of the wild-type minigene was not affected by any of the antisense oligonucleotides, at both concentration ranges (data not shown). Transcript analysis of the mutant minigene revealed that both antisense oligonucleotides at low doses were not able to hamper the binding of the spliceosomal machinery to the cryptic splice site (data not shown). When the LNAs were raised to the higher concentration range, the aberrant splicing was completely abolished and the natural splicing mechanism was recovered (Figure 3.3). This process suggests there is an LNA concentration threshold (0.5 μM), after which the splicing correction appears to be dose-independent.

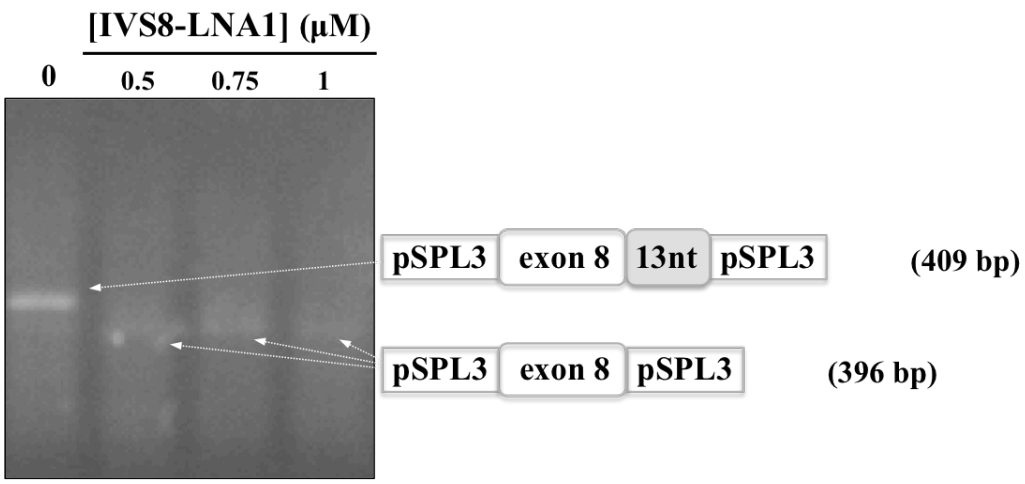


Figure 3.3. Antisense nucleotides correct the splicing pattern of the mutant minigene. Transcript analysis of HeLa cells expressing the wild-type and mutant minigene constructs and further treated with antisense oligonucleotides. RT-PCR pattern and sequence analysis of the corresponding fragment in untreated (0 μM) and treated (0.5, 0.75 and 1 μM IVS8-LNA1). IVS8-LNA2 showed identical transcript profile, also subsequently confirmed by direct sequencing. HeLa cells confirmed that both LNA oligonucleotides fully corrected the aberrant splicing (herein represented IVS8-LNA1).

3.4.2. The truncated mutant GALT is stable but inactive and prone to aggregation

The IVS8+13A>G mutation, despite inducing a frameshift in the open reading frame and introducing a premature stop codon 17 amino acids downstream (p.D274GfsX291 or p.D274Gfs*17), does not seem to elicit the nonsense mediated decay (NMD) system, as reverse transcript analysis detected the aberrant mature transcript in all patients. There is thus a considerable possibility that the mutant protein will be translated *in vivo*.

To further understand the molecular mechanism of disease, the recombinant wild-type and truncated mutant GALT were produced in a prokaryotic expression system and analyzed from a functional-structural viewpoint.

Wild-type and mutant proteins were purified by affinity chromatography. Immunoblotting analysis showed two main bands with molecular masses of ~45 kDa for the wild-type and ~34 kDa for the truncated GALT (Supplementary Figure S3.1A). Besides the main band, both proteins display bands with lower molecular mass and much lower intensity.

The impact of the mutation on the protein stability and proneness for aggregation was probed by different biophysical techniques. By far-UV CD spectroscopy, we observed nearly identical spectra for the wild-type and truncated p.IVS8 GALT, particularly two minima at 208 and 222 nm (Figure 3.4A). Monitoring CD ellipticity at 222 nm as a function of increasing temperature and fitting data to a two-state model, secondary structure thermal denaturation profiles yielded T_m values of 53.0±1.3 °C and 54.6±2.1 °C, respectively for wild-type and p.IVS8 GALT. Differential scanning fluorimetry assays were employed to obtain thermal denaturation

profiles focused on the proteins' tertiary structure elements. Two melting transitions were observed (Figure 3.4B), with similar T_m values for both proteins: T_{m1} 43.7±0.6 °C and T_{m2} 52.5±1.1 °C for wild-type, and T_{m1} 43.4±0.1 °C and T_{m2} 51.8±0.1 °C for p.IVS8 GALT. The propensity to aggregate in solution of both wild-type and mutant GALT was analyzed by DLS. Both proteins displayed quite similar T_{agg} (Figure 3.4C; 41.3±0.1 °C for wild-type and 41.1±0.1 °C for p.IVS8) and overlapping aggregation kinetics profiles at 37 °C (26.6±0.0 °C for wild-type and 25.6±2.7 °C for p.IVS8; data not shown). However, the aggregation kinetics at 42 °C showed a faster aggregation of p.IVS8 GALT comparatively to wild-type (Figure 3.4D), with a $t_{1/2}$ of 5.0±0.1 min for mutant GALT and 7.0±0.2 min for wild-type.

Recombinant wild-type GALT displayed a specific enzymatic activity of 10.2 μmol.mg^{-1}.h^{-1}, whereas p.IVS8 GALT had null activity, under standard conditions.

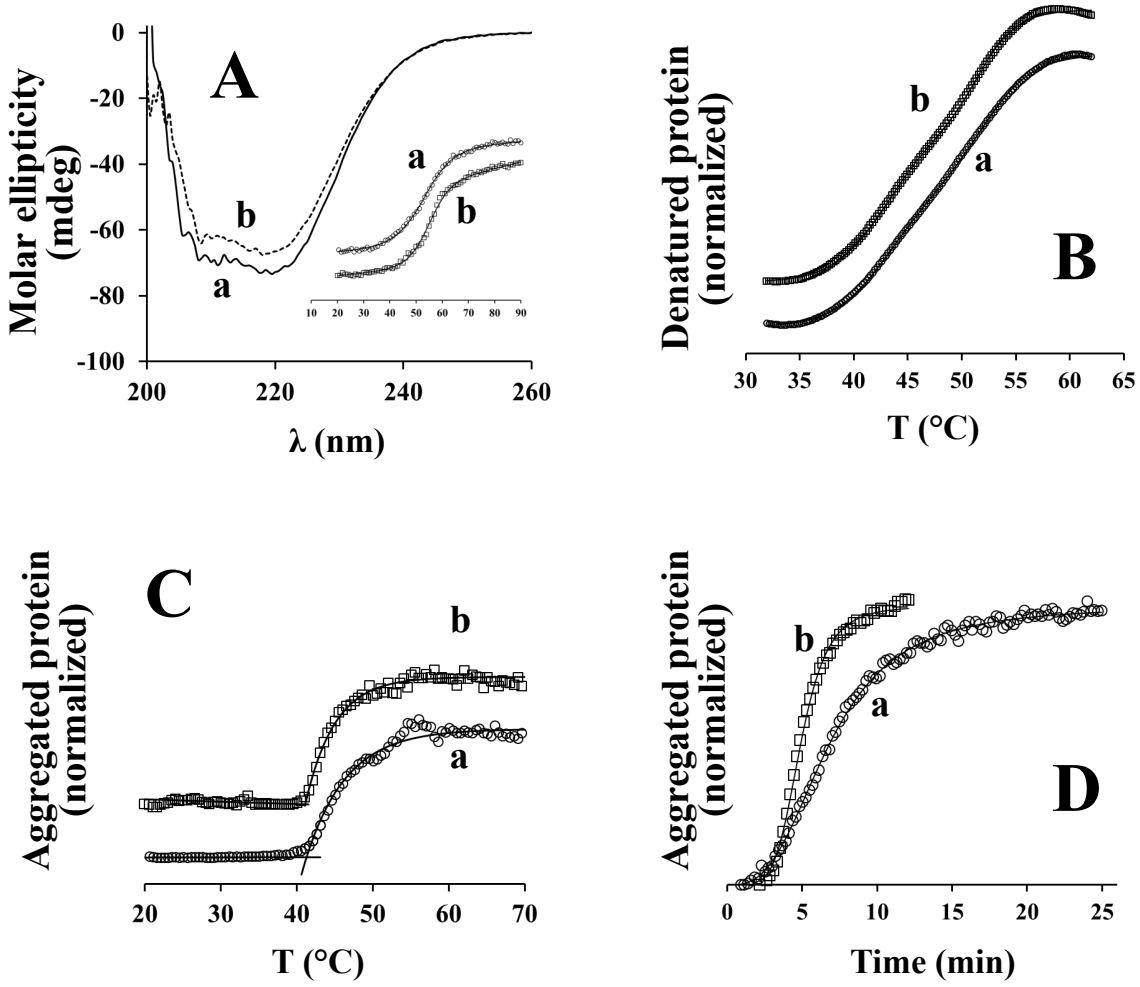


Figure 3.4. Mutant GALT is more prone to aggregation. Biophysical methodologies employed to analyze the structural impact of the mutation IVS8+13A>G on the recombinant GALT protein. Panel A, far-UV circular dichroism spectra and thermal denaturation profiles (Inset) for wild-type (a) and p.IVS8 (b); Panel B, thermal denaturation profiles obtained by DSF for wild-type (a) and p.IVS8 (b); Panel C, thermal denaturation analyzed by DLS for wild-type (a) and p.IVS8 (b); Panel D, aggregation kinetics analyzed by DLS at 42 °C for wild-type (a) and p.IVS8 (b).

3.5. Discussion

We report the first functional characterization of an intronic mutation in the *GALT* gene, and the first description of antisense oligonucleotides as therapeutic agents in classic galactosemia. Previous *in silico* analysis of the target sequence revealed the presence of a cryptic donor splice site (c.820+14_15) which, in the presence of the mutation IVS8+13A>G, undergoes activation, leading to the exonization of the first 13 nucleotides of intron 8 (26). Since RNA samples were initially not available from patients, the splicing effect predicted by bioinformatic programs was first confirmed by minigene studies. Minigene-based technologies are widely used to functionally characterize mis-splicing mutations and to evaluate the efficacy of antisense oligonucleotides in modulating splicing in a number of genetic disorders, namely in inherited metabolic disorders (29, 30).

A minigene carrying the genomic region of *GALT* spanning from exon 8 to exon 9 and respective intronic boundaries was generated, and splicing was evaluated in two eukaryotic cell lines, HeLa and COS-7. The transcript profiles were different between cell lines: HeLa cells displayed a very simple profile with one single fragment, while COS-7 cells displayed a more complex profile. Nevertheless, a key observation common to both cell lines is that the presence of the IVS8+13A>G mutation always induced the usage of the cryptic donor site instead of the canonical one.

The weakness of the acceptor splice site in intron 8, also predicted by the previous *in silico* analysis (26), was also fully confirmed by the minigene functional assay, as transcript profiling analyses of the wild-type minigene also showed exon 9 skipping, thereby confirming this phenomenon is not related with the mutation herein under study.

Pre-mRNA splicing is a process highly dependent on the availability of tissue-specific factors. Accordingly, the different transcript profiles of HeLa and COS-7 cell lines most likely reflect these differences. Indeed, recognition of splicing targets and splicing efficiency are modulated by auxiliary proteins that recognize and bind to *cis*-acting elements in the pre-mRNA, thereby either stimulating or inhibiting the recruitment of other splicing factors for the spliceosome assembly (31, 32). Moreover, the binding of these auxiliary proteins to either exonic or intronic sequences must be determinant for modulating the usage of weak splice sites. Each eukaryotic cell line presents a particular pool of splicing factors, and therefore it is expected that both cell lines present different splicing profiles. Specifically, the HeLa cell line, despite its wide usage in splicing assays, most probably lacks specific factor(s) that are crucial for exon 9 inclusion in the mature *GALT* mRNA.

Further genetic analysis of homozygous and heterozygous patients carrying the IVS8+13A>G mutation fully confirmed the results obtained by the minigene analyses. Indeed, a fragment containing the insertion of the first 13 nucleotides of intron 8 was always detected by

RT-PCR analysis, in contrast to controls, who presented a fragment resulting from the constitutive splicing event. Surprisingly, because never reported before, a second fragment was identified, in both patients and controls, corresponding to the skipping of exon 9, thus evidencing *in vivo* the minigene results (26). The occurrence of exon 9 skipping in homozygous wild-type individuals confirms this alternative splicing event is not caused by the IVS8+13A>G mutation. Indeed, transcripts resulting from alternative splicing have been previously described as normally rare transcripts that are part of the "background" noise of the splicing process (33). Accordingly, this alternative splicing of the *GALT* transcript should correspond to a physiological event, at least in human lymphocytes, which we herein report for the first time.

Altogether, *ex vivo* and *in vivo* analyses demonstrated that the IVS8+13A>G variation is actually a disease-causing mutation. Our results reiterate that the *ex vivo* approach using a minigene splicing assay constitutes a reliable tool to unravel the molecular mechanism(s) whereby a mutation affects splicing (34). Nonetheless, despite being a valuable tool *per se*, whenever possible, the minigene approach should be complemented by the *in vivo* analysis of patients' cells, as in the present study.

Once proven the pathogenic effect of the IVS8+13A>G mutation, we investigated the possibility of restoring correct splicing by antisense therapy with two LNA oligonucleotides. Both oligonucleotides were targeted to the cryptic splice site, so that the access of this splicing site would be sterically blocked to the spliceosome machinery, which would, thus, be forced to use the canonical donor splice site in the vicinity.

The need for two oligonucleotides stemmed from the fact that this mutation activates a cryptic donor splice site extremely close to the natural one and, accordingly, its steric blocking by antisense oligonucleotides could affect the correct splicing, by preventing the recognition and/or binding of the canonical GT by the spliceosome machinery. For that reason, we have chosen to design two oligonucleotides, IVS8-LNA1 and IVS8-LNA2, whose distance from the canonical GT is 10 and 5 nucleotides, respectively. Furthermore, we have chosen to design LNA oligonucleotides for their high mismatch discrimination ability (35). LNA is an RNA analogue that is conformationally locked in an N-type (C3'-*endo*) conformation (36-38), which confers it an extremely high affinity to complementary RNA and greatly improves mismatch discrimination. In addition, LNA oligonucleotides display high stability and low toxicity in biologic systems (37).

In the first experiments, we tested LNA concentrations in the reported ranges, i.e., below 100 nM. However, as the inhibition of the alternative splicing process was never achieved, higher concentrations were attempted (0.5-1 μM), which successfully restored the normal splicing profile. After the lowest LNA concentration at which the normal splicing profile was obtained (0.5 μM), no improvement in efficiency was observed at increasing doses. These results suggest that the effective LNA concentration may be potentially sequence-dependent and are in

accordance with previous pharmacological studies which provided evidence for LNA oligonucleotides to exhibit 'threshold affinity' – described as the concentration that the LNA must achieve in order to exhibit high potency, and beyond which further increase in affinity will not necessarily correlate with additional potency (39).

The transcript displaying the additional 13 nucleotides of intron 8 induces a frameshift in the open reading frame and introduces a premature termination codon (PTC) 17 residues downstream: p.D274GfsX291 or p.D274Gfs*17. PTC-containing transcripts usually trigger the nonsense-mediated decay (NMD) system, an mRNA surveillance mechanism by which the cell controls the quality of the mRNA (2, 40). Abnormal transcripts that prematurely terminate translation are eliminated by the NMD pathway, thereby preventing the production of truncated proteins that could have deleterious gain-of-function or dominant-negative effects (40, 41).

The NMD pathway is elicited upon recognition of a PTC located more than 50–55 nucleotides upstream the following exon junction (40). The *GALT* IVS8+13A>G mutation leads to the inclusion of a PTC 44 nucleotides upstream of the exon 9 - exon 10 junction, which should prevent the NMD from recognizing and degrading this transcript, thus allowing for a truncated protein to be formed. Indeed, we could detect the presence of the mutant transcript in all patients carrying this mutation.

The splicing mutation herein reported results in an mRNA encoding a truncated polypeptide with 290 amino acid residues, 89 shorter than wild-type GALT (Supplementary Figure S3.1B). Moreover, the 17 residues of p.IVS8 in the C-terminus differ from the wild-type sequence due to the frameshift caused by the aberrant splicing event. Structural models of both wild-type human GALT and p.IVS8 (Supplementary Figure S3.1C) were obtained based on the structure of bacterial GALT (PDB code: 1GUP) (42). Notably, the missing residues in p.IVS8 include i) a helix that protrudes into the other monomer to assemble the functional dimeric form, and ii) three of the four ligands (the three histidines) that bind the mononuclear non-heme iron in the native bacterial enzyme. Native GALT is proposed to function as a homodimer, and the non-catalytic iron center has been assigned a structural role (21, 43). These observations thus led us to anticipate considerable functional and structural perturbations in p.IVS8 as compared to the wild-type enzyme. To further understand the molecular mechanism dictating the functional and structural impairment regarding p.IVS8, recombinant wild-type and mutant GALT were expressed and purified from bacterial cells. A set of biophysical methodologies were employed to analyze the structural impact of the mutation on p.IVS8 GALT. We observed that the wild-type and mutant GALT displayed nearly identical far-UV CD spectra (Figure 3.4A), consistent with a high content in α-helical structural elements, balanced with β-sheets and coil regions (Supplementary Figure S1C). Moreover, the thermal denaturation profiles obtained for both proteins yielded very similar T_m values, ruling out perturbations at the secondary structure level. DSF assays allowed analyzing the mutation's effect on the tertiary structure and, although no

clear structural domains are detected in GALT, two transitions could be observed for both wild-type and p.IVS8 GALT (Figure 3.4B). Very similar T_m values were obtained between the two variants for both transitions, which suggest the absence of major structural perturbations at the tertiary structure level. By DLS, we analyzed the propensity of wild-type and p.IVS8 GALT to aggregate in solution. Both proteins exhibited nearly identical thermal aggregation profiles and T_{agg} values (Figure 3.4C), and overlapping aggregation kinetics at 37 °C. However, p.IVS8 was shown to aggregate significantly faster at 42 °C, as compared to the wild-type GALT (Figure 3.4D), which indicates a higher propensity for aggregation. Since soluble aggregates are recognized and marked for degradation by the cellular protein quality control machinery, it is likely that in a cellular context p.IVS8 has a significantly shorter lifetime. Whereas recombinant wild-type GALT displayed a specific enzymatic activity in the range of reported values, p.IVS8 had null activity under the same conditions. This indicates that not only p.IVS8 is likely to be less stable in a cellular context, but it also appears to be completely inactive. Therefore, despite the possibility that the stable mutant mRNA is translated into p.IVS8 GALT in the patients' cells, it should be completely devoid of any functional ability. These observations clearly enforce the potential of antisense therapy to correct splicing mutations and to restore GALT function in a physiological context.

In conclusion, this study constitutes the first functional characterization of an intronic mutation in classic galactosemia by two distinct yet complementary approaches: minigene assays in two cells lines and patients' genetic analysis. Additionally, functional-structural analyses of the resultant protein revealed its lack of activity and high propensity to aggregation, despite no major structural perturbations having been identified. Furthermore, two LNA oligonucleotides, specifically designed to recognize the IVS8+13A>G mutation, have successfully restored the splicing profile, and thus establish a proof-of-concept for the application of antisense therapy for mis-splicing mutations in classic galactosemia.

3.6. Acknowledgments

The authors wish to acknowledge the patients enrolled in this study, and Prof. Bélen Pérez and Sandra Brasil (Centro de Biologia Molecular Severo Ochoa, Universidad Autónoma de Madrid, Madrid, Spain) for their assistance with the minigene assays. This work was supported by Fundação para a Ciência e Tecnologia SFRH/BD/48259/2008 PhD Grant to Ana I. Coelho, FEBS Short-term Fellowship to João B. Vicente, SPDM Grant to Isabel Rivera, and PEst-OE/SAU/UI4013/2011.

3.7. References

1.Krawczak M, Ball EV, Fenton I, Stenson PD, Abeysinghe S, Thomas N, Cooper DN. *Human gene mutation database—a biomedical information and research resource*. Hum Mutat. 2000;15(1):45-51.

2.Cartegni L, Chew SL, Krainer AR. *Listening to silence and understanding nonsense: exonic mutations that affect splicing*. Nat Rev Genet. 2002;3(4):285-98.

3.Sironi M, Menozzi G, Riva L, Cagliani R, Comi GP, Bresolin N, Giorda R, Pozzoli U. *Silencer elements as possible inhibitors of pseudoexon splicing*. Nucleic Acids Res. 2004;32(5):1783-91.

4.Black DL. *Mechanisms of alternative pre-messenger RNA splicing*. Annu Rev Biochem. 2003;72:291-336.

5.Buratti E, Baralle FE. *Influence of RNA secondary structure on the pre-mRNA splicing process*. Mol Cell Biol. 2004;24(24):10505-14.

6.Roscigno RF, Weiner M, Garcia-Blanco MA. *A mutational analysis of the polypyrimidine tract of introns*. J Biol Chem. 1993;268(15):11222-9.

7.Wang GS, Cooper TA. *Splicing in disease: disruption of the splicing code and the decoding machinery*. Nat Rev Genet. 2007;8(10):749-61.

8.Buratti E, Baralle M, Baralle FE. *Defective splicing, disease and therapy: searching for master checkpoints in exon definition*. Nucleic Acids Res. 2006;34(12):3494-510.

9.Homolova K, Zavadakova P, Doktor TK, Schroeder LD, Kozich V, Andresen BS. *The deep intronic c.903+469T>C mutation in the MTRR gene creates an SF2/ASF binding exonic splicing enhancer, which leads to pseudoexon activation and causes the cblE type of homocystinuria*. Hum Mutat. 2010;31(4):437-44.

10.Houdayer C, Caux-Moncoutier V, Krieger S, Barrois M, Bonnet F, Bourdon V, Bronner M, Buisson M, Coulet F, Gaildrat P, Lefol C, Leone M, Mazoyer S, Muller D, Remenieras A, Revillion F, Rouleau E, Sokolowska J, Vert JP, Lidereau R, Soubrier F, Sobol H, Sevenet N, Bressac-de Paillerets B, Hardouin A, Tosi M, Sinilnikova OM, Stoppa-Lyonnet D. *Guidelines for splicing analysis in molecular diagnosis derived from a set of 327 combined in silico/in vitro studies on BRCA1 and BRCA2 variants*. Hum Mutat. 2012;33(8):1228-38.

11.Malik R, Roy I. *Making sense of therapeutics using antisense technology*. Expert Opin Drug Discov. 2011;6(5):507-26.

12.Bennett CF, Swayze EE. *RNA targeting therapeutics: molecular mechanisms of antisense oligonucleotides as a therapeutic platform*. Annu Rev Pharmacol Toxicol. 2010;50:259-93.

13.Leslie ND. *Insights into the pathogenesis of galactosemia*. Annu Rev Nutr. 2003;23:59-80.

14.Fridovich-Keil JL, Walter JH. *Galactosemia*. In: Valle D, Beaudet AL, Vogelstein B, Kinzler KW, Antonarakis SE, Ballabio A, editors. The Online Metabolic and Molecular Bases of Inherited Disease: Mc-Graw Hill; 2008. p. 1-92.

15.Fridovich-Keil JL. *Galactosemia: the good, the bad, and the unknown*. J Cell Physiol. 2006;209(3):701-5.

16.Bosch AM. *Classical galactosaemia revisited*. J Inherit Metab Dis. 2006;29(4):516-25.

17.Waisbren SE, Potter NL, Gordon CM, Green RC, Greenstein P, Gubbels CS, Rubio-Gozalbo E, Schomer D, Welt C, Anastasoaie V, D'Anna K, Gentile J, Guo CY, Hecht L, Jackson R, Jansma BM, Li Y, Lip V, Miller DT, Murray M, Power L, Quinn N, Rohr F, Shen Y, Skinder-Meredith A, Timmers I, Tunick R, Wessel A, Wu BL, Levy H, Elsas II LJ, Berry GT. *The adult galactosemic phenotype*. J Inherit Metab Dis. 2012;35(2):279-86.

18.Tyfield L, Reichardt JK, Fridovich-Keil JL, Croke DT, Elsas II LJ, Strobl W, Kozak L, Coskun T, Novelli G, Okano Y, Zekanowski C, Shin Y, Boleda MD. *Classical galactosemia and mutations at the galactose-1-phosphate uridyl transferase (GALT) gene*. Hum Mutat. 1999;13(6):417-30.

19.Leslie ND, Immerrman EB, Flach JE, Florez M, Fridovich-Keil JL, Elsas II LJ. *The human galactose-1-phosphate uridyltransferase gene*. Genomics. 1992;14(2):474-80.

20.Elsas LJ, Lai K, Saunders CJ, Langley SD. *Functional analysis of the human galactose-1-phosphate uridyltransferase promoter in Duarte and LA variant galactosemia*. Mol Genet Metab. 2001;72(4):297-305.

21.Williams VP. *Purification and Some Properties of Galactose 1-phosphate Uridylyltransferase from Human Red Cells*. Archives of Biochemistry and Biophysics. 1978;191(1):182-91.

22.Shi LY, Suslak L, Rosin I, Searle BM, Desposito F. *Gene dosage studies supporting localization of the structural gene for galactose-1-phosphate uridyl transferase (GALT) to band p13 of chromosome 9*. Am J Med Genet. 1984;19(3):539-43.

23.Calderon FR, Phansalkar AR, Crockett DK, Miller M, Mao R. *Mutation database for the galactose-1-phosphate uridyltransferase (GALT) gene*. Hum Mutat. 2007;28(10):939-43.

24.Reichardt JK. *Genetic basis of galactosemia*. Hum Mutat. 1992;1(3):190-6.

25.Reichardt JK. *Molecular basis of galactosemia: Mutations and polymorphisms in the gene encoding human galactose-1-phosphate uridylyltransferase.* Proc Natl Acad Sci U S A. 1991.

26.Coelho AI, Ramos R, Gaspar A, Costa C, Oliveira A, Diogo L, Garcia P, Paiva S, Martins E, Teles EL, Rodrigues E, Cardoso MT, Ferreira E, Sequeira S, Leite M, Silva MJ, de Almeida IT, Vicente JB, Rivera I. *A frequent splicing mutation and novel missense mutations color the updated mutational spectrum of classic galactosemia in Portugal*. J Inherit Metab Dis. 2013;37(1):43-52.

27.Maniatis T, Fritsch EF, Sambrook J. *Molecular cloning. A laboratory manual*. New York: Cold Spring Harbor; 1982.

28.Lindhout M, Rubio-Gozalbo ME, Bakker JA, Bierau J. *Direct non-radioactive assay of galactose-1-phosphate:uridyltransferase activity using high performance liquid chromatography*. Clin Chim Acta. 2010;411(13-14):980-3.

29.Perez B, Rodriguez-Pascau L, Vilageliu L, Grinberg D, Ugarte M, Desviat LR. *Present and future of antisense therapy for splicing modulation in inherited metabolic disease*. J Inherit Metab Dis. 2010;33(4):397-403.

30.Hammond SM, Wood MJ. *Genetic therapies for RNA mis-splicing diseases*. Trends Genet. 2011;27(5):196-205.

31.Schor IE, Gomez Acuna LI, Kornblihtt AR. *Coupling Between Transcription and Alternative Splicing*. In: Wu JW, editor. RNA and Cancer. 158: Springer Berlin Heidelberg; 2013. p. 1-24.

32.Faustino NA, Cooper TA. *Pre-mRNA splicing and human disease*. Genes Dev. 2003;17(4):419-37.

33.Rincon A, Aguado C, Desviat LR, Sanchez-Alcudia R, Ugarte M, Perez B. *Propionic and methylmalonic acidemia: antisense therapeutics for intronic variations causing aberrantly spliced messenger RNA*. Am J Hum Genet. 2007;81(6):1262-70.

34.Baralle D, Baralle M. *Splicing in action: assessing disease causing sequence changes*. J Med Genet. 2005;42(10):737-48.

35.You Y, Moreira BG, Behlke MA, Owczarzy R. *Design of LNA probes that improve mismatch discrimination*. Nucleic Acids Res. 2006;34(8):e60.

36.Koshkin AA, Singh SK, Nielsen P, Rajwanshi VK, Kumar R, Meldgaard M, Olsen CR, Wengel J. *LNA (Locked Nucleic Acids): Synthesis of the adenine, cytosine, guanine, 5-methylcytosine, thymine and uracil bicyclonucleoside monomers, oligomerisation, and unprecedented nucleic acid recognition*. Tetrahedron. 1998;54(14):3607-30.

37.Jepsen JS, Sorensen M, Wengel J. *Locked nucleic acid: a potent nucleic acid analog in therapeutics and biotechnology*. Oligonucleotides. 2004;14(2):130-46.

38.Lindow M, Kauppinen S. *Discovering the first microRNA-targeted drug*. J Cell Biol. 2012;199(3):407-12.

39.Straarup EM, Fisker N, Hedtjarn M, Lindholm MW, Rosenbohm C, Aarup V, Hansen HF, Orum H, Hansen JB, Koch T. *Short locked nucleic acid antisense oligonucleotides potently reduce apolipoprotein B mRNA and serum cholesterol in mice and non-human primates*. Nucleic Acids Res. 2010;38(20):7100-11.

40.Maquat LE. *Nonsense-mediated mRNA decay: splicing, translation and mRNP dynamics*. Nat Rev Mol Cell Biol. 2004;5(2):89-99.

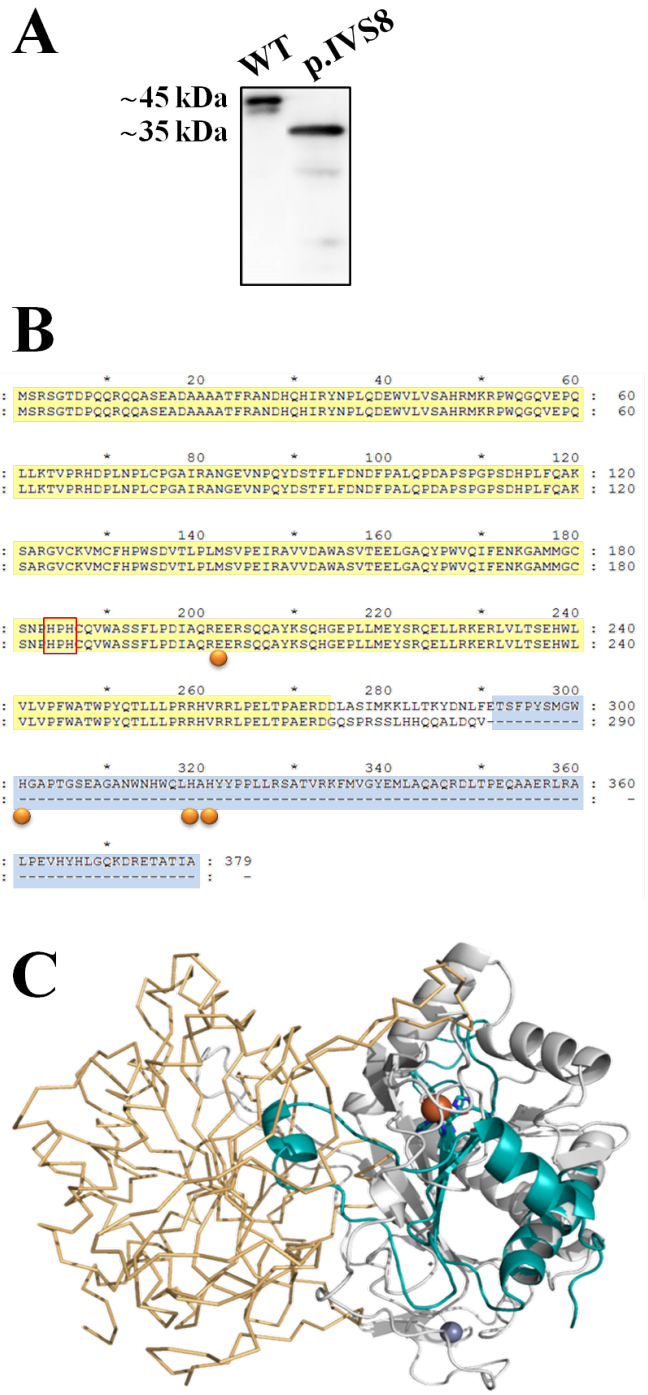
41.Wilkinson MF. *The cycle of nonsense*. Mol Cell. 2003;12(5):1059–66.

42.Thoden JB, Ruzicka FJ, Frey PA, Rayment I, Holden HM. *Structural analysis of the H166G site-directed mutant of galactose-1-phosphate uridylyltransferase complexed with either UDP-glucose or UDP-galactose: detailed description of the nucleotide sugar binding site*. Biochemistry. 1997;36(6):1212-22.

43.Geeganage S, Frey PA. *Significance of metal ions in galactose-1-phosphate uridylyltransferase: an essential structural zinc and a nonessential structural iron*. Biochemistry. 1999;38(40):13398-406.

44.DeLano WL. *PyMOL: an open-source molecular graphics tool*. 2002.

3.8. Supplementary Material



Supplementary Figure S3.1. Recombinant production and *in silico* analysis of wild-type and truncated mutant GALT. Panel A, immunoblotting analysis of purified recombinant wild-type (lane 1, WT) and mutant (lane 2, p.IVS8) GALT proteins; Panel B, sequence alignment between wild-type and p.IVS8 mutant obtained with Clustal X; red box highlights the His_{184}-Pro_{185}-His_{186} active site; orange spheres highlight the conserved mononuclear iron ligands; light blue box highlights the missing residues in the p.IVS8 mutant; Panel C, structural models of wild-type and p.IVS8 GALT obtained with Swiss-Modeller with the *Escherichia coli* GALT (PDB code 1GUP) as the structural template; grey cartoon depicts the overlapping structure of wild-type and p.IVS8, whereas the light blue cartoon depicts the extra residues in wild-type GALT; orange and purple spheres depict respectively the iron and zinc ions in the bacterial enzyme; orange ribbons depict the opposing monomer from the bacterial GALT PDB; figure elaborated with PyMOL (44).

CHAPTER 4

Functional and structural impact of the most prevalent missense mutations in classic galactosemia

Ana I. Coelho[1], Matilde Trabuco[1], Ruben Ramos[1], Maria João Silva[1,2], Isabel Tavares de Almeida[1,2], Paula Leandro,[1,2] Isabel Rivera[1,2], João B. Vicente[1,2]

[1] Metabolism & Genetics Group, Research Institute for Medicines and Pharmaceutical Sciences (*i*Med.UL), Faculty of Pharmacy, University of Lisbon, Portugal

[2] Department of Biochemistry and Human Biology, Faculty of Pharmacy, University of Lisbon, Portugal

Human Mutation
(submitted)

4.1 Abstract

Galactose-1-phosphate uridylyltransferase is a key enzyme in the galactose metabolism, particularly important in the neonatal period due to the ingestion of galactose-containing milk. GALT deficiency results in the life-threating genetic disorder classic galactosemia, whose pathophysiology, despite several decades of intensive research, is still not fully elucidated. Classic galactosemia is a loss-of-function disease, hypothesized to result from GALT misfolding. A thorough functional-structural characterization of GALT most prevalent variants is still lacking, hampering the development of a small-molecule based therapeutic approach. The aim of this study is to investigate the structural-functional effects of nine mutations in the *GALT* gene, four of which account for the vast majority of the mutations identified worldwide in galactosemic patients. Several methodologies were employed, from enzymatic activity and thermal inactivation to biophysical methods, to evaluate the mutations' impact on the GALT protein secondary, tertiary and quaternary structure. The results indicate no major effects on the secondary and tertiary structures, and reveal that the key structural impact concerns aggregation in solution. In particular, p.Q188R, the most frequent allele at a worldwide scale, accounting for ~60% of mutant alleles, displayed a much lower T_{agg} than WT GALT, and aggregated much faster both at 37 °C and upon thermal insult at 42 °C. Thus, GALT aggregation, rather than misfolding, appears to be the major pathogenic mechanism underlying the most prevalent missense mutations, and so we hypothesize that proteostasis modulators might potentially improve GALT function, opening new avenues for therapeutic research in classic galactosemia.

4.2 Introduction

Classic galactosemia (OMIM #230400) is an autosomal recessive disorder caused by mutations in the *GALT* gene, resulting in deficient activity of galactose-1-phosphate uridylyltransferase (GALT, EC 2.7.7.12), a key enzyme in galactose metabolism (1). GALT catalyzes the second step of the Leloir pathway, converting galactose-1-phosphate (Gal-1-P) and UDP-glucose (UDP-Glc) into glucose-1-phosphate and UDP-galactose (UDP-Gal) (1).

In classic galactosemia, acute symptoms generally appear soon after birth upon exposure to milk, and include: vomiting, diarrhea, excessive weight loss, lethargy, hypotonia, liver dysfunction, and, in the absence of intervention, can escalate to cataracts, *E. coli* sepsis, and eventually to neonatal death (1-4). These symptoms generally self-resolve once the patient is placed on a stringent life-long dietary restriction of galactose, which is the current standard of care (5). However, despite resolving the acute and potentially lethal symptoms, the dietetic treatment does not prevent the development of serious long-term complications, namely cognitive

and neurologic disabilities, and premature ovarian insufficiency in females (1, 6).

Thus far 266 variations have been described at the *GALT locus* (available at http://www.arup.utah.edu/database/GALT/GALT_display.php, last surveyed on December 2013), of which missense mutations constitute the majority (>60%), despite the high allelic heterogeneity (7). In particular, the p.Q188R mutation (c.563A>G; CAG→CGG) is by far the most frequent, accounting for ~63% of *GALT* mutant alleles (8, 9). Its incidence is particularly high in European descendant patients, reaching >90% of mutant alleles in Ireland; it has however never been reported in Asian descendant patients (1, 10, 11). Other frequent mutations include p.S135L, p.K285N and p.N314D. The p.S135L mutation (c.404 C>T; TCG → TTG) affects mostly African descendant patients, ranging from approximately half of mutant alleles in African Americans to ~90% in South African patients (8, 9). The second most frequent *GALT* mutant allele in European descendant patients is p.K285N (c.855G>T; AAG→AAT), with a higher incidence in Eastern Europe, reaching 34% in Poland (8, 9, 12, 13). The p.N314D mutation (c.940A>G; AAC→GAC) appears to be an evolutionary remnant, since the D314 is actually the ancestral allele that persists nowadays at a pan-ethnic frequency of nearly 10% (12, 14). Notwithstanding the several studies on the molecular basis of these mutations' pathogenicity (15-26), a characterization of these variants focusing on different structural features is still missing.

To date there is no solved three-dimensional structure of the human GALT. However, based on the availability of the *E. coli* GalT structure (27) and on the high sequence identity and similarity among the human and the prokaryotic GALT, a three-dimensional model of the human GALT was constructed which provided important insights into the structural and functional features of this protein (Figure 4.1). GALT is a member of the transferase branch of the histidine triad (HIT) family of enzymes; the catalytic site sequence His-Pro-His is conserved in nature, and was firstly identified in the *E. coli* enzyme at residues 164 to 166, corresponding to residues 184 to 186 in the human sequence (27-30). The reaction displays ping-pong kinetics and a double displacement mechanism, involving an uridylyl-enzyme, in which the nucleophilic histidine at residue 186 is transiently nucleotidylated (27, 31, 32). The active enzyme is an 86.6-kDa homodimer (Figure 4.1) composed of ~43.3-kDa monomers, with two active sites, each formed by residues contributed by both subunits (33). Whereas the *E. coli* GalT has two mononuclear metal-binding sites (one for zinc and the other for iron) with proposed structural roles, the human GALT lacks two of the zinc ligands, thus remaining to be established whether metal binding in the human protein is comparable to that of the bacterial GalT (34).

A recent study reported that five missense mutations in the *GALT* gene led to misfolding of the GALT protein, suggesting classic galactosemia as a conformational disorder (35). However, little is known about the conformational impact of the most prevalent mutations, which hampers the design of alternative therapies for this monogenic disorder based on the use of stabilizing low molecular weight compounds (36). Accordingly, the aim of this study is to further

investigate the structural-functional effects of the most prevalent mutations in the *GALT* gene, p.Q188R, p.S135L, p.K285N and p.N314D, and of five other clinically relevant mutations, p.R148Q, p.G175D, p.P185S, p.R231C and p.R231H.

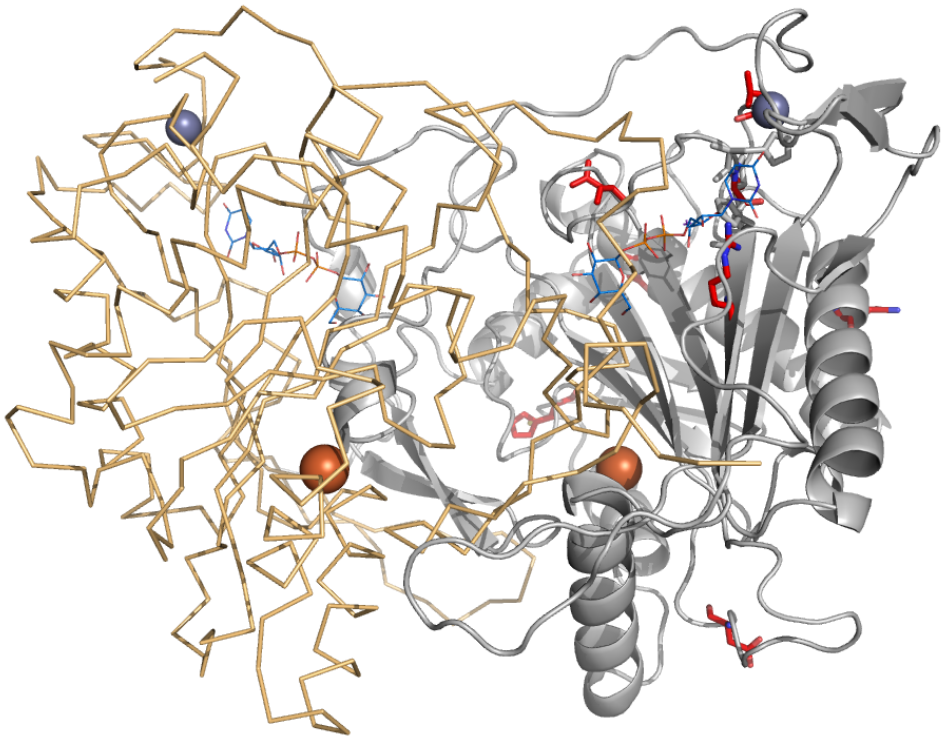


Figure 4.1 - Structural model of GALT dimer. Structural model of human GALT (grey cartoon representation) obtained using *Escherichia coli* GalT crystallographic structure as template (orange ribbon representation, PDB ID: 1GUP). Iron (orange sphere), zinc (purple sphere), and bound UDP-glucose (light blue lines) originate from 1GUP PDB. Mutations herein studied are represented in red sticks. Figure generated with PyMOL (37).

4.3 Materials and Methods

4.3.1 Production of recombinant human GALT variants

Recombinant human GALT was produced by cloning the human *GALT* cDNA sequence into pET24, to yield a N-terminally hexa-histidyl-tagged protein, as previously described (38). Site-directed mutagenesis (NZY mutagenesis kit, NZYTech, Lisbon, Portugal) was employed to introduce all the mutations herein under study using the primers listed in (Supplementary Table S4.1). Direct sequencing in both forward and reverse orientations was used to confirm the correct introduction of mutations and to exclude additional mutational events.

Vectors bearing the cDNA encoding the GALT variants were transformed into *E. coli* BL21 (DE3) Rosetta cells. For protein production, cells were grown in M9 minimal medium (39) supplemented with 100 μM ferrous ammonium sulfate and 100 μM zinc sulfate, at 37 °C. Protein expression was induced by addition of 400 μM IPTG once the OD_{600nm} reached 0.3, the cultures

were placed at 21 °C, and the cells were harvested after 4 hours. Bacterial cells were resuspended in buffer A (50 mM Tris pH 7.5, 300 mM KCl, and 10% glycerol) with 1 mg/mL lysozyme and 500 μM phenymethanesulfonyl fluoride, disrupted by sonication, and clarified by centrifugation (5 minutes at 8,000 xg).

The fusion proteins were purified by immobilized metal affinity chromatography (IMAC), by loading the cellular extracts into a 1-mL FF-Crude column (Amersham, GE Healthcare) and eluting the proteins with buffer A containing increasing concentrations of imidazole (pure GALT eluted at 500 mM imidazole). After purification, imidazole was eliminated with a desalting column pre-equilibrated and eluted with buffer A, and protein solutions were concentrated by ultrafiltration, aliquoted, flash-frozen in liquid nitrogen and stored at −80 °C. Protein purity was assessed by SDS-PAGE, and protein concentration was determined by the Bradford assay (40).

4.3.2 GALT activity assays and thermal inactivation profiles

GALT enzymatic activity was measured as previously described (41), and performed on the same day as purification. All assays were carried out for 30 min at 37 °C, in a reaction mixture containing 2.0 mM Gal-1-P, 0.5 mM UDP-Glc, 40 μM dithiothreitol (DTT) and 125 mM glycine, in 40 mM Tris–HCl, pH 7.5. UDP-Glc and UDP-Gal were separated by HPLC and analyzed by UV detection at 262 nm (41, 42). The enzyme activity was expressed in μmol UDP-Gal formed per hour per mg protein at 37 °C (μmol UDP-Gal.h^{-1}.mg^{-1}). Adequate controls lacking either substrate or the GALT protein were routinely performed.

Wild-type (WT) GALT kinetic parameters for UDP-Glc and Gal-1P were determined in the same conditions as in (42) with minor modifications namely the use of nine concentrations of UDP-Glc (0.02-1.5 mM; [Gal-1-P] = 2.0 mM), and nine concentrations of Gal-1-P (0.05- 6.0 mM; [UDP-Glc] = 0.5 mM). The steady-state kinetic constants were determined by nonlinear regression analysis using the GraphPad Prism 6 software (GraphPad, Software, Inc), the Michaelis-Menten equation for Gal-1-P and the modified Michaelis-Menten equation to account for substrate inhibition for UDP-Glc.

Thermal inactivation profiles were obtained by analyzing enzyme activity as a function of temperature in the 20-65 °C range. Aliquots of protein (WT and p.N314D: 4.28 μg/mL; remaining mutants: 21.4 μg/mL) were incubated at the different temperatures for 10 min, immediately chilled on ice for 10 min, and enzyme activity was determined by adding the reaction mixture described above and incubating it at 37 °C for 30 min. Enzymatic activity values plotted as a function of temperature yielded sigmoidal curves, from which the midpoints of thermal inactivation ($T_{½}$) were obtained from the inflexion point. Two assays were performed for

each temperature data point, and the WT GALT thermal inactivation profile was repeated in parallel with each tested mutant.

4.3.3 Far-UV circular dichroism spectropolarimetry

Far-UV circular dichroism (far-UV CD) spectra and thermal denaturation profiles were recorded in a Jasco J-710 spectropolarimeter, coupled to a Jasco PTC-348WI Peltier temperature controller and a Haake G/D8 water bath. All GALT protein samples were at 0.15 mg/mL, each spectrum being the result of six accumulations at a 50 nm/min scan rate, at 20 °C, in a 0.1 cm light path cuvette. Thermal denaturation profiles were obtained by monitoring molar ellipticity at 222 nm, in the 20-90 °C temperature range (1 °C/min slope; data pitch: 1 °C; delay time: 0 sec). Temperature scan curves were fitted to a two-state model.

4.3.4 Differential scanning fluorimetry

Differential scanning fluorimetry (DSF) experiments were performed in a C1000 Touch thermal cycler equipped with a CFX96 optical reaction module (Bio-Rad), by having the GALT variants at a 0.1 mg/mL (~2.5 μM in monomer) final concentration in buffer A, SYPRO orange (Invitrogen Corporation) at a 5x working concentration, in a 50 μL total volume. A 10-min incubation step at 20 °C preceded the temperature ramp from 20 to 90 °C at 1 °C/min, with a 1-sec hold time every 0.2 °C and fluorescence acquisition using the HEX channel (excitation maximum at 535 nm, emission maximum at 555 nm). Assays using 2.0 mM Gal-1-P, 0.5 mM UDP-Glc, 100 μM Fe^{2+} and 100 μM Zn^{2+} were also performed. Control assays in the absence of protein were routinely performed.

Data were processed using CFX Manager software V3.0 (Bio-Rad). Temperature scan curves were fitted to a biphasic sigmoidal function and the T_m values were obtained from the inflexion points of the first and second transitions. Variations in T_m values are considered significant when $|\Delta T_m| \geq 2$ °C (above the standard deviation).

4.3.5 Dynamic light scattering

Dynamic light scattering (DLS) data were acquired in a ZetaSizer Nano-S (Malvern Instrument, UK) particle size analyzer, coupled to a Peltier temperature control unit, using a He-Ne laser as the light source (633 nm). Prior to data collection, protein samples were centrifuged at 15,000 xg for 30 min at 4 °C, diluted in buffer A to a final concentration of 0.15 mg/mL, and filtered with a 0.22 μm membrane to remove large aggregates. Temperature was ramped from 20

°C to 70 °C at 0.5 °C/min, with the particle size average, distribution and total scattering intensity being collected. Data were processed using Zetasizer Nano DTS software v7.01 (Malvern Instrument). The aggregation temperature (T_{agg}), defined as the temperature at which both size and intensity start to increase significantly, was determined by fitting the obtained data to a plateau followed by one phase association equation.

The kinetics of thermal aggregation was monitored at 37 °C and 42 °C for 60 min. By plotting light scattering intensity as a function of time, sigmoidal curves were obtained and the $t_{1/2}$ was determined as the time elapsed to reach half saturation of aggregated protein in the sample.

4.3.6 *In silico* analysis

Structural models of WT and mutant human GALT, based on the *E. coli* GalT structure (PDB ID: 1GUP), were obtained from the SWISS-MODEL server (43, 44). Comparative analysis of the structural models and the corresponding electrostatic surface maps was done with the PyMOL software (37).

4.4 Results

4.4.1 Impaired catalytic ability of GALT mutants

The WT recombinant human GALT was isolated in its active state, displaying for Gal-1-P a V_{max} of 59.1 μmol UDP-Gal.h^{-1}.mg^{-1} and a K_M of 1.08 mM, and for UDP-Glc a V_{max} of 75.5 μmol UDP-Gal.h^{-1}.mg^{-1} and a K_M of 425 μM.

Aside from the p.N314D mutant, which displayed nearly identical enzymatic activity to the WT protein, all the studied GALT mutants presented markedly reduced (≤0.2% of WT for p.Q188R, p.S135L and p.G175D) or null enzymatic activity (Table 4.1). Thermal inactivation profiles were obtained for the GALT variants exhibiting measurable catalytic activity. All the analyzed mutants, namely p.Q188R, p.S135L, p.N314D and p.G175D, displayed lower $T_{½}$ than that of the WT GALT, with $\Delta T_{½}$ ranging from -8.1 °C to -19.9 °C (Table 4.1).

4.4.2 Limited impact of GALT mutations on the secondary and tertiary structure

Far-UV CD spectra of all GALT variants were very similar to that of WT, with two minima at 208 and 222 nm (Figure 4.2). Thermal denaturation curves, obtained by monitoring the molar ellipticity at 222 nm as a function of constantly increasing temperature, presented an

apparently single transition and were fitted according to a two-state model. Thermal denaturation of the GALT variants appeared to be irreversible, since the spectra collected at 20 °C after cooling the denatured samples had lost the spectral features assigned to the different secondary structure elements (data not shown). The thermal denaturation profiles of all variants yielded similar T_m values (Table 4.1), ranging from 52.6 to 56.7 °C. With the exception of p.R231H, all mutants displayed slightly higher T_m values than WT GALT, although all the $\Delta T_m \pm$ SD fall below the 2 °C threshold.

Differential scanning fluorimetry was employed to analyze the mutations' effects on tertiary structure elements. The fluorescence intensity measured in the first asymptote of the sigmoidal thermal denaturation profiles (flat over the 20-30 °C range) was normalized with respect to the WT values (Supplementary Figure S4.1). As observed, all the variants exhibited similar ground-state fluorescence, the sole exception being the p.Q188R mutant, which displayed values ~30% higher than those of the WT GALT.

The DSF thermal denaturation profiles for all variants exhibited two apparent transitions, each accounting for 40-60 % of the overall fluorescence increase (Supplementary Figure S4.2). The inflexion points of the two transitions, T_{m1} and T_{m2}, fell within a narrow range of temperatures, with T_{m1} ranging from 42.0 ± 0.4 to 44.3 ± 0.2 °C and T_{m2} ranging from 51.0 ± 0.4 to 53.9 ± 0.1 °C (Table 4.1). None of the GALT mutants exhibited T_m values ≥2 °C higher or lower than those determined for the WT GALT.

Binding of the GALT substrates Gal-1-P and UDP-Glc was tested by DSF assays. Neither substrate yielded significant changes in either T_m values (all ΔT_m <1 °C, Supplementary Table S4.2). Since GALT has two putative mononuclear metal binding sites, one for iron and another for zinc, DSF assays were carried out in the presence of either metal. The only GALT variant exhibiting a response to Fe^{2+} was p.P185S, its T_{m1} increasing by 2.5 ± 0.5 °C (Supplementary Table S4.2). The presence of Zn^{2+} had two levels of impact on the thermal denaturation profiles and their corresponding T_m values (Supplementary Table S4.2). Whereas the T_{m1} values for p.Q188R, p.N314D, and p.R148Q remained unvaried, WT and all other mutants exhibited a decrease in T_{m1} from -2.5 ± 0.3 °C to -5.5 ± 0.3 °C. The effect of Zn^{2+} had a deeper impact on the T_{m2} values, which decreased significantly (ΔT_{m2} between -2.9 ± 0.6 °C and -5.2 ± 1.5 °C) for the WT GALT and all the mutants except p.Q188R.

4.4.3 Disturbed aggregation of GALT mutants

The propensity of GALT variants to aggregate in solution was analyzed by DLS, evaluating the T_{agg} and also the aggregation kinetics at two different temperatures (37 °C and 42 °C). Scanning the particle size as a function of temperature, the estimated T_{agg} were essentially

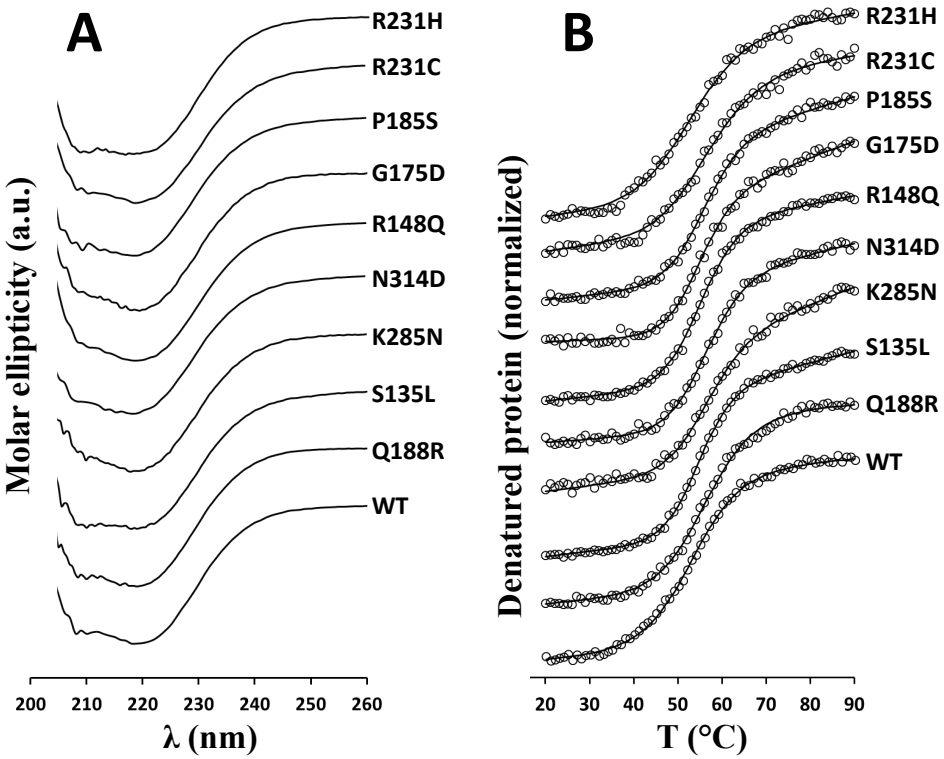


Figure 4.2 - No impact of the studied mutations on human GALT secondary structure. Effect of missense mutations on the secondary structure of GALT variants, probed by far-UV circular dichroism spectropolarimetry. Panel A: Far-UV CD spectra, collected for GALT variants, at 20 °C, at 0.15 mg/mL, in 50 mM Tris-HCl, 300 mM KCl, 10 % glycerol, pH 7.5. Spectra resulted from 6 accumulations at a 50 nm/min scan rate, in a 0.1 cm light path cuvette. Panel B: thermal denaturation profiles obtained by monitoring molar ellipticity at 222 nm in the 20-90 °C temperature range (1 °C/min slope; data pitch: 1 °C; delay time: 0 sec); temperature scan curves were normalized and fitted according to a two-state model (respective T_m values in Table 4.1).

identical for all GALT variants herein studied (ranging from 40.1 ± 1.0 °C to 41.4 ± 0.1 °C), except the p.Q188R mutant, which started to aggregate at a lower temperature (ΔT_{agg} of -3.9 °C, with respect to the WT GALT) (Figure 4.2.A and Table 4.1).

The aggregation kinetics was monitored by determining $t_{1/2}$ at 37 °C and 42 °C, representing respectively a physiological body temperature and a thermal insult. At 37 °C, whereas the p.K285N, p.R148Q, and p.R231H mutants displayed similar $t_{1/2}$ as the WT GALT (~27 min), the other mutants exhibited disturbed aggregation profiles, aggregating either faster (p.Q188R, p.G175D and p.P185S, approximate $\Delta t_{1/2}$ respectively -20, -12, and -16 min) or slower (p.S135L, p.N314D and p.R231C, approximate $\Delta t_{1/2}$ respectively +10, +15, and +22 min) than the WT GALT (Supplementary Figure S4.3 and Table 4.1). At 42 °C, the aggregation was sped-up, lowering the $t_{1/2}$ of all GALT variants, and the mutants either aggregate as fast as the WT GALT ($t_{1/2}$ ~7 min; p.S135L, p.K285N, p.N314D, and p.R231C) or faster ($t_{1/2}$ between 2.3 and 5 min; p.Q188R, p.R148Q, p.G175D, p.P185S, and p.R231H) (Figure 4.2.B and Table 4.1).

Table 4.1. Structural and functional parameters determined for recombinant wild-type and mutant GALT. Enzyme activity and thermal inactivation profiles determined by HPLC; secondary structure probed by far-UV circular dichroism; tertiary structure analyzed by differential scanning fluorimetry; aggregation propensity studied by dynamic light scattering.

	Enzyme activity	Thermal Inactivation	Circular Dichroism	Differential scanning fluorimetry		Dynamic light scattering		
				Thermal denaturation		Thermal aggregation	Aggregation kinetics	
							37°C	42°C
	(% WT)	$T_{½}$ (°C)	T_m (°C)	T_{m1} (°C)	T_{m2} (°C)	T_{agg} (°C)	$t_{½}$ (min)	$t_{½}$ (min)
Wild-type	100	55.5	53.0 ± 1.5	43.7 ± 0.7	52.4 ± 1.2	41.3 ± 0.1	26.6 ± 0.1	7.0 ± 0.3
Q188R	0.2	46.6	56.4 ± 0.9	42.0 ± 0.4	52.3 ± 0.3	37.4 ± 0.2	6.7 ± 0.1	2.3 ± 0.1
S135L	0.1	37.3	54.4 ± 1.0	44.2 ± 0.1	51.8 ± 0.2	41.3 ± 0.1	36.0 ± 1.4	6.9 ± 0.1
K285N	n.d.[a]	n.a.	55.9 ± 0.3	42.7 ± 0.7	51.4 ± 0.2	41.2 ± 0.1	27.9 ± 1.0	7.5 ± 0.1
N314D	101	47.4	56.1 ± 0.7	43.8 ± 0.1	53.9 ± 0.1	41.4 ± 0.1	41.8 ± 6.3	6.7 ± 0.1
R148Q	n.d.[a]	n.a.	55.0 ± 0.9	44.3 ± 0.2	53.3 ± 0.4	40.1 ± 1.0	28.4 ± 8.1	5.0 ± 0.4
G175D	0.2	35.6	54.4 ± 0.3	43.1 ± 0.1	51.3 ± 0.1	40.4 ± 0.5	15.0 ± 0.1	3.6 ± 0.1
P185S	n.d.[a]	n.a.	55.6 ± 0.1	42.9 ± 0.5	52.1 ± 0.4	41.0 ± 0.1	10.4 ± 0.1	3.3 ± 0.4
R231C	n.d.[a]	n.a.	56.7 ± 3.5	43.9 ± 0.2	52.1 ± 0.1	42.0 ± 0.1	48.3 ± 9.8	7.6 ± 0.1
R231H	n.d.[a]	n.a.	52.6 ± 3.8	42.4 ± 0.1	51.0 ± 0.4	41.0 ± 0.1	24.7 ± 0.6	4.9 ± 0.1

[a] below the detection limit of the assay.

4.5 Discussion

Like the majority of genetic disorders, the mutational spectrum of classic galactosemia is dominated by missense mutations. Since the current standard of care based on a galactose-restricted diet is mostly ineffective in preventing the long-term complications, a deeper knowledge on the molecular basis of *GALT* mutations pathogenicity will guide in the design of new therapeutic strategies. Herein we report the structural-functional characterization of nine clinically relevant GALT mutants, four of which are the most prevalent among *GALT* mutant alleles: p.Q188R, p.S135L, p.K285N and p.N314D.

The purified recombinant WT GALT displayed enzymatic an activity and kinetic parameters towards Gal-1-P and UDP-Glc compatible with reported values obtained by a direct UDP-Gal quantitation HPLC assay (41), rather than the often used two-enzyme coupled GALT activity assay (35, 45). From all the tested GALT variants, only four mutants displayed detectable activity (p.Q188R, p.S135L, p.N314D and p.G175D), even using a 5-fold higher protein concentration than WT GALT. Whereas p.N314D had the same specific activity as the WT enzyme, the other mutants displayed ≤0.2 % of residual enzyme activity, consistent with previous reports (15, 16, 19, 21, 23). Thermal inactivation profiles were obtained, showing that the WT GALT is strikingly stable ($T_{1/2}$ = 55.5 °C), whereas all the mutants with detectable enzymatic activity were more sensitive to thermal inactivation, revealing an impaired functional stability (Table 4.1).

Far-UV CD spectropolarimetry was used to probe the impact of the studied mutations on the secondary structure elements of GALT mutants. The WT and mutant GALT displayed overlapping CD spectral features (Figure 4.2), particularly minima at ~222 nm and ~208 nm, consistent with a combination of α-helical and β-sheet secondary structure content (six α-helices and thirteen β-sheets). These data show that the studied mutations have no significant effects on the secondary structure topology, replicating the previous study on the *E. coli* GalT and its p.Q168R variant, equivalent to human p.Q188R (46). Structural models obtained for each variant (Supplementary Figures S4.4) suggest that the side-chains of the substituting residues likely have limited or null effects on secondary structure elements. Indeed, thermal denaturation profiles (Figure 4.2) confirmed that all mutants displayed thermal stability at the secondary structure level very similar to WT GALT, with $|\Delta T_m|\pm$ S.D. < 2 °C (Table 4.1).

Impact of the mutations on the tertiary structure of GALT variants was evaluated by DSF, a method whereby a fluorescent dye binds to the proteins buried hydrophobic regions exposed upon thermal unfolding. The 'ground-state' fluorescence was evaluated for all GALT variants and a >30% increase was observed for p.Q188R (Supplementary Figure S4.1), whereas the remaining mutants were similar to the WT GALT. This indicates that p.Q188R shows a basal destabilization level, possibly caused by a more open conformation and a higher exposure of

hydrophobic residues. However, this ~30 % increase is modest, when compared with destabilizing mutations in other proteins, such as medium-chain acyl-coA dehydrogenase (47). DSF thermal denaturation curves showed two distinct transitions (Supplementary Figure S4.2), indicative of two protein regions unfolding almost as separate domains, contrarily to the previous report of a single transition in similar experiments, for a different set of GALT variants (35). Both T_m values determined for the WT GALT (43.7 ± 0.7 °C and 52.4 ± 1.2 °C) are lower than the previously reported single T_m (63 °C) (35). Different experimental details could partially explain these distinct observations, particularly a pH closer to physiologic in our assays (7.5 vs. pH 8.8 reported in (35)), besides the protein concentrations and the temperature slope (35). Concerning the mutants, the T_m values for each transition lie within two temperature ranges separated by ~9 °C, and displayed no significant differences relative to the WT GALT (Table 4.1), since all $|\Delta T_m| < 2$ °C. This observation appears to rule out any significant effect of the studied mutations on the protein thermal stability, particularly concerning their tertiary structure.

DSF assays were also employed to evaluate the effect of substrate binding (Gal-1-P and UDP-Glc) on the conformational stability of the GALT variants (Supplementary Table S4.2). Under the tested conditions, we observed no effect of either substrate on the T_m values for the WT GALT and the studied variants, whereas McCorvie *et al.* previously reported a stabilizing effect of both substrates towards the WT GALT and the p.D28Y and p.F194L mutants (35). Notably, the structures of the *E. coli* GalT in the native and nucleotylated states (PDB ID 1HXP and 1HXQ) are totally overlapping (27, 48), ruling out major structural conformational changes upon substrate binding.

Zinc and iron were shown to have a structural role in bacterial GalT (49). While the mononuclear iron-binding ligands are fully conserved in human GALT, the zinc ligands are only partially conserved, which raised the question whether this metal is indeed essential (19). Whereas iron had no effect on the WT GALT stability in the DSF assays, zinc had the puzzling effect of destabilizing the protein, lowering both T_m values by ~4 °C (Supplementary Table S4.2). Since the zinc binding pocket is not fully conserved in human GALT, our results raise the possibility that zinc can partly occupy the mononuclear iron-binding site, previously proposed to be unable to completely discriminate between iron and zinc (34, 50), as commonly observed in mononuclear and binuclear iron-binding proteins. Concerning the mutants, iron had no significant effects on any variant except for a slight increase in p.P185S T_{m1}, whereas zinc had major effects on most mutants (Supplementary Table S4.2). The p.Q188R mutant presented the most deviant behaviour, being totally insensitive to zinc. Whereas p.N314D and p.R148Q showed a decrease only in T_{m2} in the presence of zinc, all the other mutants behaved essentially as the WT GALT. The mutated serine in p.S135L structurally overlaps with the zinc-binding ligand H115 in *E. coli* GalT. The fact that this mutant displays the same zinc sensitivity as WT GALT enforces the idea that the observed zinc-induced destabilization might be related to non-

specific binding at the iron-binding site and not at the putative zinc-binding site. Although the effect of zinc cannot as yet be rationalized in functional terms, the different impact on the conformational stabilities may be suggestive of subtle structural differences between these variants.

Since aggregation in solution is a hallmark of protein misfolding, DLS was used to compare the proneness of the different GALT variants to aggregate. For each protein, we evaluated the T_{agg} and the aggregation kinetics at 37 °C and upon a thermal insult at 42 °C (Figure 4.3 and Table 4.1). In terms of thermal aggregation profiles, all but the p.Q188R mutant exhibited T_{agg} values close to 41 °C, nearly identical to the WT GALT. The p.Q188R mutant had a remarkably disturbed thermal aggregation profile, with a T_{agg} close to 37 °C, indicating a higher propensity to aggregate in solution as compared to WT GALT. Since parameters similar to the WT were obtained in the far-UV CD and DSF experiments, differences in conformational stability can be discarded. Nevertheless, and taking into consideration that this variant presented ~30 % higher ground-state fluorescence than the remaining GALT variants, we hypothesize that the higher exposure of hydrophobic residues in this mutant may be directing its increased tendency to aggregate. The aggregation kinetics at the two different temperatures also highlighted the disturbed aggregation behaviour of other variants. At 37 °C, the p.G175D and p.P185S mutants also aggregate significantly faster than the WT GALT, whereas the p.S135L, p.N314D and p.R231C mutants presented delayed aggregation kinetics, suggestive of subtle structural differences that are exclusively reflected on the aggregation propensity. Upon thermal insult at 42 °C, some of the effects observed at 37 °C were majored, others levelled out, and new effects were observed. Besides p.Q188R, p.G175D and p.P185S, also p.R148Q and p.R231H aggregate faster than WT GALT at this higher temperature, indicating that the latter two mutants are actually less resistant to aggregation under this thermal insult. All the mutants that presented slower aggregation at 37 °C were levelled to the WT aggregation kinetics at 42 °C. p.K285N was the only mutant showing aggregation parameters essentially identical to the WT GALT.

The results from the different methodologies herein employed indicate that the major structural impact of the studied mutations concerns the aggregation in solution, with no significant effects on the secondary and tertiary structures. To support our understanding of the molecular basis of each mutation's pathogenicity, we generated structural models of each mutant and attempted to extract subtle structural differences that may provide clues for the functional and structural impairment (Supplementary Figure S4.4). The p.Q188R has generally been regarded as a functional mutation, since the substituted glutamine establishes through its amide moiety two H-bonds towards UDP-Gal, as observed in the bacterial structure (46). In the bacterial mutant p.Q168R – equivalent to human p.Q188R – one of these H-bonds is absent. However, in the generated p.Q188R model (Supplementary Figure S4.4), it appears there is an actual gain in H-bonds in the mutant, since the guanidinium moiety is able to establish three

'new' H-bonds towards the intermediate phosphate and sugar moieties. The functional impairment of p.Q188R may therefore result from an over-stabilization of the substrates and/or products in place, thus blocking the enzyme active site for further reaction turnover. On top of the local H-bond network differences, we observed a significant change in the electrostatic surface surrounding this position (Supplementary Figure S4.4), consistent with the substitution of a globally neutral amide with the positively charged guanidinium of arginine. These global charge differences may also affect binding of reaction substrates and products, which are all mostly negatively charged. The structural model did not provide clues for the increased aggregation of p.Q188R, which constitutes the major novelty regarding this mutant. It remains also to be clarified how this substitution renders the mutated protein insensitive to zinc-induced destabilization. Taken together, these two observations demonstrate how a local subtle change can propagate into other regions of the protein with such dramatic global effects.

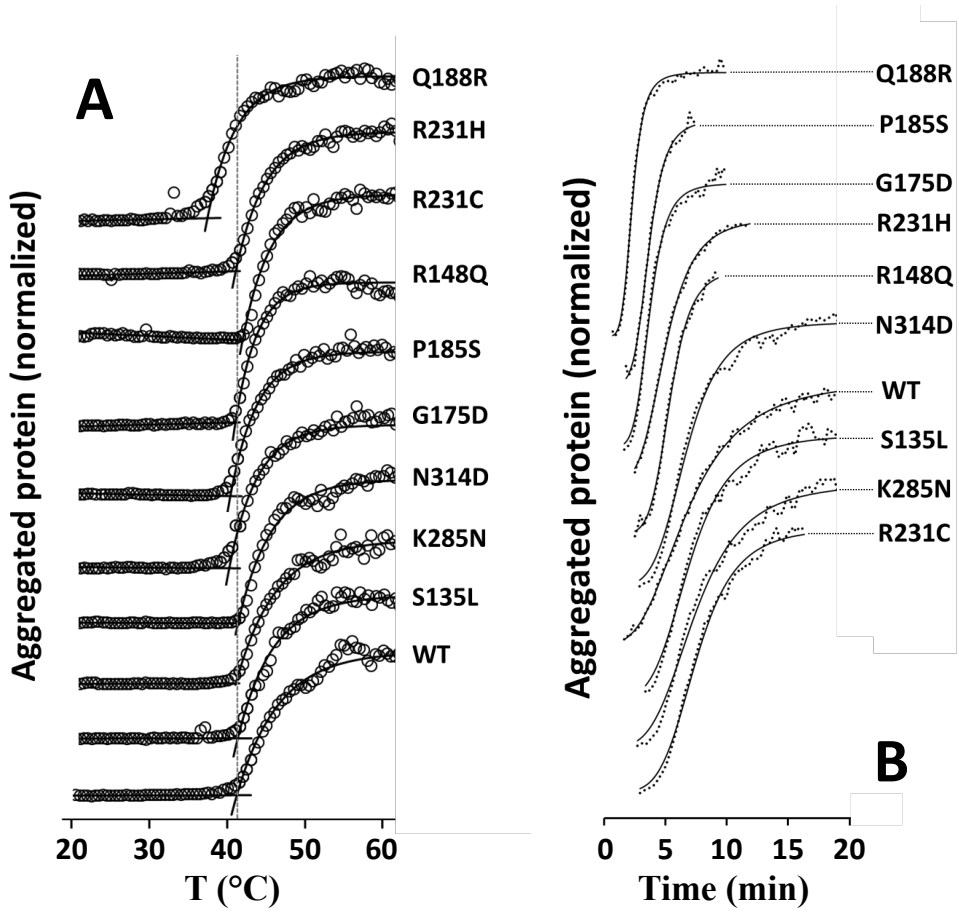


Figure 4.3 - Dynamic light scattering analysis of GALT variants reveals disturbed aggregation. Impact of missense mutations on the aggregation of GALT variants in solution, studied by dynamic light scattering (DLS). All proteins samples were diluted in 50 mM Tris-HCl, 300 mM KCl, 10 % glycerol, pH 7.5, to a final concentration of 0.15 mg/mL. *Panel A*, temperature-induced aggregation profiles, obtained by a linear temperature increase from 20 °C to 70 °C at 0.5 °C/min, collecting the particle size average, distribution and total scattering intensity. Scattering intensity data were normalized and fitted to a plateau followed by one phase association equation, the aggregation temperature (T_{agg}) being defined as the temperature at which the intensity starts to increase significantly. *Panel B*, kinetics of thermal aggregation monitored at 42 °C for 60 min. Light scattering intensity are plotted as a function of

time, sigmoidal curves were obtained and the $t_{1/2}$ was defined as the time elapsed to reach 50% of maximum of aggregation. Asymptotes were removed for clarity, due to the data noise in those regions of the profiles.

Most studied mutants that presented disturbed aggregation kinetics also displayed predicted effects on their surface electrostatic charges, either inverting the polarity of the local charges, or neutralizing them. The surrounding areas of p.G175D and p.N314D (Supplementary Figure S4.4) change from slightly positive to negative, though the first mutant residue is slightly buried but close to the surface while the latter is fully exposed to surface. More dramatic effects are predicted for the p.R148Q, p.R231C and p.R231H mutations (Supplementary Figure S4.4), where the highly positive surface charges from arginine guanidinium moieties are obliterated to neutrality. Altogether, the predicted significant impact of these mutations on the GALT surface charges may account for the observed disturbed aggregation profiles.

The p.S135L, p.K285N and p.P185S mutations (Supplementary Figure S4.4) remain as outliers, as the observed structural and functional impairment cannot be completely rationalized. The p.P185S mutation is buried and appears to have no effect on the external surface charge, although it aggregates almost as fast as the p.Q188R mutant. Notably, p.Q188R and p.P185S have a nearby location and affect catalytically-relevant residues. The p.S135L mutation, displaying no predicted dramatic local charges or H-bonds alterations, remains puzzling. Again, we can only envisage subtle structural changes protruding onto the surface and causing p.S135L to display a slower aggregation profile at 37 °C. The p.K285N is also an intriguing mutation, since it displays nearly identical aggregation properties as the WT GALT, despite being slightly surface exposed and the substitution resulting in a significant change in the surface charge (from positive to negative). Moreover, no net gain or loss of H-bonds or other types of electrostatic interactions can be envisaged from the structural model.

Several studies have aimed to unravel the molecular bases of the most prevalent mutations in classic galactosemia. A recent report suggests that misfolding is the underlying pathogenic mechanism of several *GALT* missense mutations, as commonly observed for genetic diseases (35, 51). Herein we aimed to shed light on the pathogenic mechanism(s) of the most frequent mutations in classic galactosemia that altogether account for the vast majority of mutant alleles at a global scale (p.Q188R, p.S135L, p.K285N and p.N314D), by analyzing their impact from both functional and structural perspectives. The most striking and novel observation is that the mutants display disturbed aggregation profiles, despite the absence of structural effects on their secondary and tertiary structures. This is particularly relevant for p.Q188R, the most prevalent mutation accounting for ~60 % of the mutant alleles. This observation is extremely important, since at the cellular level, the accumulation of aggregation-prone proteins interferes dramatically with the cellular homeostasis. Notably, studies on galactosemia models have revealed increased ER stress (52), unfolded protein response (53) and oxidative stress levels (54), hallmarks of perturbations in the cellular protein homeostasis. At present, there is an increasing

awareness that accumulation of damaged or abnormal proteins is the underlying pathogenic molecular mechanism of several diseases, and several studies in inherited metabolic disorders have unveiled that protein aggregation is actually a more common pathogenic mechanism that previously thought (55). Accordingly, the results from the structural analyses of the GALT mutants herein analyzed strongly suggest that GALT aggregation might be the major pathogenic mechanism, and so we hypothesize that proteostasis modulators might potentially improve GALT function in classic galactosemia models. Such proteostasis modulators may extend the life-time of the GALT variants in the cell, partially compensating impaired function with enzyme availability, and simultaneously preventing accumulation of protein aggregates, which appears to be a key feature in classic galactosemia pathogenesis.

4.6 Acknowledgments

This work was supported by Fundação para a Ciência e Tecnologia SFRH/BD/48259/2008 PhD Grant to Ana I. Coelho, FEBS Short-term Fellowship to João B. Vicente, SPDM Grant to Isabel Rivera, and PEst-OE/SAU/UI4013/2011. Paulo R. Lino (Metabolism & Genetics Group, *i*Med.UL, Faculty of Pharmacy, University of Lisbon) is acknowledged for his support with dynamic light scattering measurements.

4.7 References

1.Fridovich-Keil JL, Walter JH. *Galactosemia*. In: Valle D, Beaudet AL, Vogelstein B, Kinzler KW, Antonarakis SE, Ballabio A, editors. The Online Metabolic and Molecular Bases of Inherited Disease: Mc-Graw Hill; 2008. p. 1-92.

2.Bosch AM. *Classical galactosaemia revisited*. J Inherit Metab Dis. 2006;29(4):516-25.

3.Suchy FJ, Sokol RJ, Balistreri WF. *Inborn Errors of Carbohydrate Metabolism*. In: Suchy FJ, Sokol RJ, Balistreri WF, editors. Liver Disease in Children. 3 ed: Cambridge University Press; 2007. p. 595-625.

4.Holton JB, Walter JH, Tyfield LA. *Galactosemia*. In: Scriver CR, Beaudet AL, Sly WS, Valle D, editors. The Metabolic and Molecular Bases of Inherited Disease. 8 ed. New York: McGraw-Hill; 2001. p. 1553–87.

5.Fridovich-Keil JL. *Galactosemia: the good, the bad, and the unknown*. J Cell Physiol. 2006;209(3):701-5.

6.Waggoner DD, Buist NRM, Donnel GN. *Long-term prognosis in galactosaemia: results of a survey of 350 cases*. J Inherit Metab Dis. 1990;13(6):802-18.

7.Calderon FR, Phansalkar AR, Crockett DK, Miller M, Mao R. *Mutation database for the galactose-1-phosphate uridyltransferase (GALT) gene*. Hum Mutat. 2007;28(10):939-43.

8.Tyfield L, Reichardt JK, Fridovich-Keil JL, Croke DT, Elsas II LJ, Strobl W, Kozak L, Coskun T, Novelli G, Okano Y, Zekanowski C, Shin Y, Boleda MD. *Classical galactosemia and mutations at the galactose-1-phosphate uridyl transferase (GALT) gene*. Hum Mutat. 1999;13(6):417-30.

9.Elsas II LJ, Langley S, Paulk EM, Hjelm LN, Dembure PP. *A molecular approach to galactosemia*. Eur J Pediatr. 1995;154(2):S21-S7.

10.Coss KP, Doran PP, Owoeye C, Codd MB, Hamid N, Mayne PD, Crushell E, Knerr I, Monavari AA, Treacy EP. *Classical Galactosaemia in Ireland: incidence, complications and outcomes of treatment*. J Inherit Metab Dis. 2013;36(1):21-7.

11.Hirokawa H, Okano Y, Asada M, Fujimoto A, Suyama I, Isshiki G. *Molecular basis for phenotypic heterogeneity in galactosaemia: prediction of clinical phenotype from genotype in Japanese patients*. European Journal of Human Genetics. 1999;7:757-64.

12.Suzuki M, West C, Beutler E. *Large-scale molecular screening for galactosemia alleles in a pan-ethnic population*. Human Genetics. 2001;109(2):210-5.

13.Zekanowski C, Radomyska B, Bal J. *Molecular characterization of Polish patients with classical galactosaemia*. J Inher Metab Dis. 1999;22:679-82.

14.Carney AE, Sanders RD, Garza KR, McGaha LA, Bean LJ, Coffee BW, Thomas JW, Cutler DJ, Kurtkaya NL, Fridovich-Keil JL. *Origins, distribution and expression of the Duarte-2 (D2) allele of galactose-1-phosphate uridylyltransferase*. Hum Mol Genet. 2009;18(9):1624-32.

15.Fridovich-Keil JL, Jinks-Robertson S. *A yeast expression system for human galactose-l-phosphate uridylyltransferase*. Proc Natl Acad Sci U S A. 1993;90(2):398-402.

16.Lai K, Willis AC, Elsas II LJ. *The biochemical role of glutamine 188 in human galactose-1-phosphate uridyltransferase*. J Biol Chem. 1999;274(10):6559-66.

17.Reichardt JK, Packman S, Woo SLC. *Molecular characterization of two galactosemia mutations: correlation of mutations with highly conserved domains in galactose-1-phosphate uridyl transferase*. Am J Hum Genet. 1991;49(4):860-7.

18.Fridovich-Keil JL, Langley SD, Mazur LA, Lennon JC, Dembure PP, Elsas II LJ. *Identification and functional analysis of three distinct mutations in the human galactose-1-phosphate uridyltransferase gene associated with galactosemia in a single family*. Am J Hum Genet. 1995;56(3):640-6.

19.Wells L, Fridovich-Keil JL. *Biochemical characterization of the S135L allele of galactose-1-phosphate uridylyltransferase associated with galactosaemia*. J Inherit Metab Dis. 1997;20(5):633-42.

20.Lai K, Langley SD, Singh RH, Dembure PP, Hjelm LN, Elsas II LJ. *A prevalent mutation for galactosemia among black Americans*. J Pediatr. 1996;128(1):89-95.

21.Lai K, Elsas LJ. *Structure-function analyses of a common mutation in blacks with transferase-deficiency galactosemia*. Mol Genet Metab. 2001;74(1-2):264-72.

22.Chhay JS, Openo KK, Eaton JS, Gentile M, Fridovich-Keil JL. *A yeast model reveals biochemical severity associated with each of three variant alleles of galactose-1P uridylyltransferase segregating in a single family*. J Inherit Metab Dis. 2008;31(1):97-107.

23.Fridovich-Keil JL, Quimby BB, Wells L, Mazur LA, Elsevier JP. *Characterization of the N314D allele of human galactose-1-phosphate uridylyltransferase using a yeast expression system*. Biochem Mol Med. 1995;56(2):121-30.

24.Langley SD, Lai K, Dembure PP, Hjelm LN, Elsas II LJ. *Molecular basis for Duarte and Los Angeles variant galactosemia*. Am J Hum Genet. 1997;60(2):366-72.

25.Lai K, Langley SD, Dembure PP, Hjelm LN, Elsas II LJ. *The Duarte allele impairs biostability of galactose-1-phosphate uridyltransferase in human lymphoblasts*. Hum Mutat. 1998;11(1):28-38.

26.Riehman K, Crews C, Fridovich-Keil JL. *Relationship between genotype, activity, and galactose sensitivity in yeast expressing patient alleles of human galactose-1-phosphate uridylyltransferase*. J Biol Chem. 2001;276(14):10634-40.

27.Wedekind JE, Frey PA, Rayment I. *Three-dimensional structure of galactose-1-phosphate uridylyltransferase from Escherichia coli at 1.8 Å resolution*. Biochemistry. 1995;34(35):11049- 61.

28.Leslie ND. *Insights into the pathogenesis of galactosemia*. Annu Rev Nutr. 2003;23:59-80.

29.Brenner C. *Hint, Fhit, and GalT: function, structure, evolution, and mechanism of three branches of the histidine triad superfamily of nucleotide hydrolases and transferases*. Biochemistry. 2002;41(29):9003-14.

30.Marabotti A, Facchiano A. *Homology modeling studies on human galactose-1-phosphate uridylyltransferase and on its galactosemia-related mutant Q188R provide an explanation of molecular effects of the mutation on homo-and heterodimers*. J Med Chem. 2005;48(3): 773-9.

31.Field TL, Reznikoff WS, Frey PA. *Galactose-1-phosphate uridylyltransferase: identification of histidine-164 and histidine-166 as critical residues by site-directed mutagenesis*. Biochemistry. 1989;28(5):2094-9.

32.Wong L-J, Frey PA. *Galactose-1-phosphate uridylyltransferase. Rate studies confirming a uridylyl-enzyme intermediate on the catalytic pathway*. Biochemistry. 1974;13(19):3889-94.

33.Thoden JB, Ruzicka FJ, Frey PA, Rayment I, Holden HM. *Structural analysis of the H166G site-directed mutant of galactose-1-phosphate uridylyltransferase complexed with either UDP-glucose or UDP-galactose: detailed description of the nucleotide sugar binding site*. Biochemistry. 1997;36(6):1212-22.

34.Geeganage S, Frey PA. *Significance of metal ions in galactose-1-phosphate uridylyltransferase: an essential structural zinc and a nonessential structural iron*. Biochemistry. 1999;38(40):13398-406.

35.McCorvie TJ, Gleason TJ, Fridovich-Keil JL, Timson DJ. *Misfolding of galactose 1-phosphate uridylyltransferase can result in type I galactosemia*. Biochim Biophys Acta. 2013;1832(8):1279-93.

36.Leandro P, Gomes CM. *Protein misfolding in conformational disorders: rescue of folding defects and chemical chaperoning*. Mini-Rev Med Chem. 2008;8(9):901-11.

37.DeLano WL. *PyMOL: an open-source molecular graphics tool*. 2002.

38.Coelho AI, Lourenço S, Trabuco M, Silva MJ, Oliveira A, Gaspar A, Diogo L, Tavares de Almeida I, Vicente JB, Rivera I. *Antisense therapy for classic galactosemia: functional correction of a splicing mutation in the GALT gene*.submitted.

39.Maniatis T, Fritsch EF, Sambrook J. *Molecular cloning. A laboratory manual*. New York: Cold Spring Harbor; 1982.

40.Bradford MM. *A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding*. Anal Biochem.72(1):248-54.

41.Lindhout M, Rubio-Gozalbo ME, Bakker JA, Bierau J. *Direct non-radioactive assay of galactose-1-phosphate:uridyltransferase activity using high performance liquid chromatography*. Clin Chim Acta. 2010;411(13-14):980-3.

42.Coelho AI, Ramos R, Gaspar A, Costa C, Oliveira A, Diogo L, Garcia P, Paiva S, Martins E, Teles EL, Rodrigues E, Cardoso MT, Ferreira E, Sequeira S, Leite M, Silva MJ, de Almeida IT, Vicente JB, Rivera I. *A frequent splicing mutation and novel missense mutations color the updated mutational spectrum of classic galactosemia in Portugal*. J Inherit Metab Dis. 2013;37(1):43-52.

43.Arnold K, Bordoli L, Kopp J, Schwede T. *The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling*. Bioinformatics. 2006;22(2):195-201.

44.Kiefer F, Arnold K, Kunzli M, Bordoli L, Schwede T. *The SWISS-MODEL Repository and associated resources*. Nucleic Acids Res. 2009;37(no. suppl 1):D387-92.

45.Tang M, Facchiano A, Rachamadugu R, Calderon F, Mao R, Milanesi L, Marabotti A, Lai K. *Correlation assessment among clinical phenotypes, expression analysis and molecular modeling of 14 novel variations in the human galactose-1-phosphate uridylyltransferase gene*. Hum Mutat. 2012;33(7):1107-15.

46.Geeganage S, Frey PA. *Transient kinetics of formation and reaction of the uridylyl-enzyme form of galactose-1-P uridylyltransferase and its Q168R-variant: insight into the molecular basis of galactosemia*. Biochemistry. 1998;37(41):14500-7.

47.Maier EM, Gersting SW, Kemter KF, Jank JM, Reindl M, Messing DD, Truger MS, Sommerhoff CP, Muntau AC. *Protein misfolding is the molecular mechanism underlying MCADD identified in newborn screening*. Hum Mol Genet. 2009;18(9):1612-23.

48.Wedekind JE, Frey PA, Rayment I. *The structure of nucleotidylated histidine-166 of galactose-1-phosphate uridylyltransferase provides insight into phosphoryl group transfer*. Biochemistry. 1996;35(36):11560-9.

49.Ruzicka FJ, Wedekind JE, Kim J, Rayment I, Frey PA. *Galactose-1-phosphate uridylyltransferase from Escherichia coli, a zinc and iron metalloenzyme*. Biochemistry. 1995;34(16):5610-7.

50.Holden HM, Rayment I, Thoden JB. *Structure and function of enzymes of the Leloir pathway for galactose metabolism*. J Biol Chem. 2003;278(45):43885-8.

51.Gregersen N, Bross P, Vang S, Christensen JH. *Protein misfolding and human disease*. Annu Rev Genomics Hum Genet. 2006;7:103-24.

52.Mulhern ML, Madson CJ, Danford A, Ikesugi K, Kador PF, Shinohara T. *The unfolded protein response in lens epithelial cells from galactosemic rat lenses*. Invest Ophthalmol Vis Sci. 2006;47(9):3951-9.

53.De-Souza EA, Pimentel FS, Machado CM, Martins LS, da-Silva WS, Montero-Lomeli M, Masuda CA. *The unfolded protein response has a protective role in yeast models of classic galactosemia*. Dis Model Mech. 2014;7(1):55-61.

54.Jumbo-Lucioni PP, Hopson ML, Hang D, Liang Y, Jones DP, Fridovich-Keil JL. *Oxidative stress contributes to outcome severity in a Drosophila melanogaster model of classic galactosemia*. Dis Model Mech. 2013;6(1):84-94.

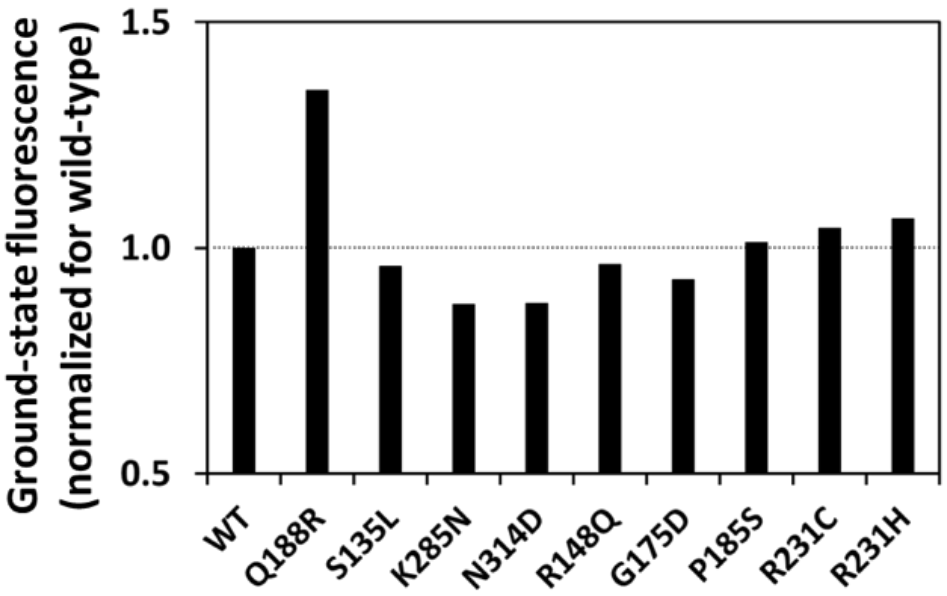
55.Pedersen CB, Bross P, Winter VS, Corydon TJ, Bolund L, Bartlett K, Vockley J, Gregersen N. *Misfolding, degradation, and aggregation of variant proteins. The molecular pathogenesis of short chain acyl-CoA dehydrogenase (SCAD) deficiency*. J Biol Chem. 2003;278(48):47449-58.

4.8 Supplementary material

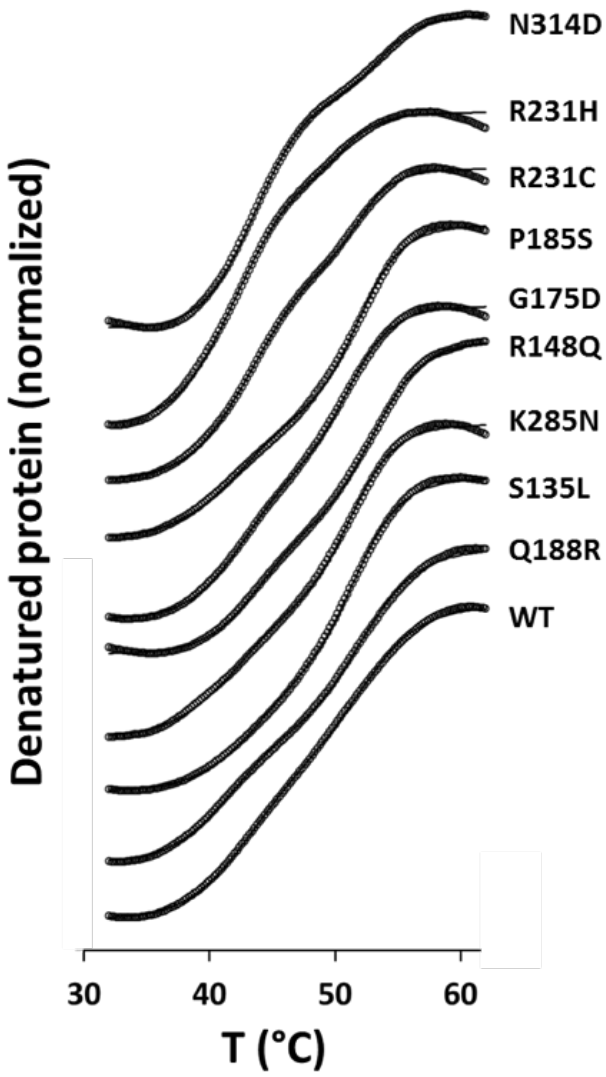
Supplementary Table S4.1. Oligonucleotides used for site-directed mutagenesis.

Primers [a]	**Sequence (5'→3')** [b]
Q188R-F	CCC CAC CCC CAC TGC CGG GTA TGG GCC AGC AG
Q188R-R	CTG CTG GCC CAT ACC CGG CAG TGG GGG TGG GG
S135L -F	GCT TCC ACC CCT GGT TGG ATG TAA CGC TGC
S135L-R	GCA GCG TTA CAT CCA ACC AGG GGT GGA AGC
K285N-F	GAA GAA GCT CTT GAC CAA TTA TGA CAA CCT CTT TGA G
K285N-R	CTC AAA GAG GTT GTC ATA ATT GGT CAA GAG CTT CTT C
N314D-F	GGC TGG GGC CAA CTG GGA CCA TTG GCA GCT GC
N314D-R	GCA GCT GCC AAT GGT CCC AGT TGG CCC CAG CC
R148Q-F	GGT CCC TGA GAT CCA GGC TGT TGT TGA TGC
R148Q-R	GCA TCA ACA ACA GCC TGG ATC TCA GGG ACC
G175D-F	GCA GAT CTT TGA AAA CAA AGA TGC CAT GAT GGG CTG TTC
G175D-R	GAA CAG CCC ATC ATG GCA TCT TTG TTT TCA AAG ATC TGC
P185S-F	GCT GTT CTA ACC CCC ACT CCC ACT GCC AGG
P185S-R	CCT GGC AGT GGG AGT GGG GGT TAG AAC AGC
R231C-F	GCT ACT CAG GAA GGA ATG TCT GGT CCT AAC CAG TG
R231C-R	CAC TGG TTA GGA CCA GAC ATT CCT TCC TGA GTA GC
R231H-F	GCT ACT CAG GAA GGA ACA TCT GGT CCT AAC CAG TG
R231H-R	CAC TGG TTA GGA CCA GAT GTT CCT TCC TGA GTA GC

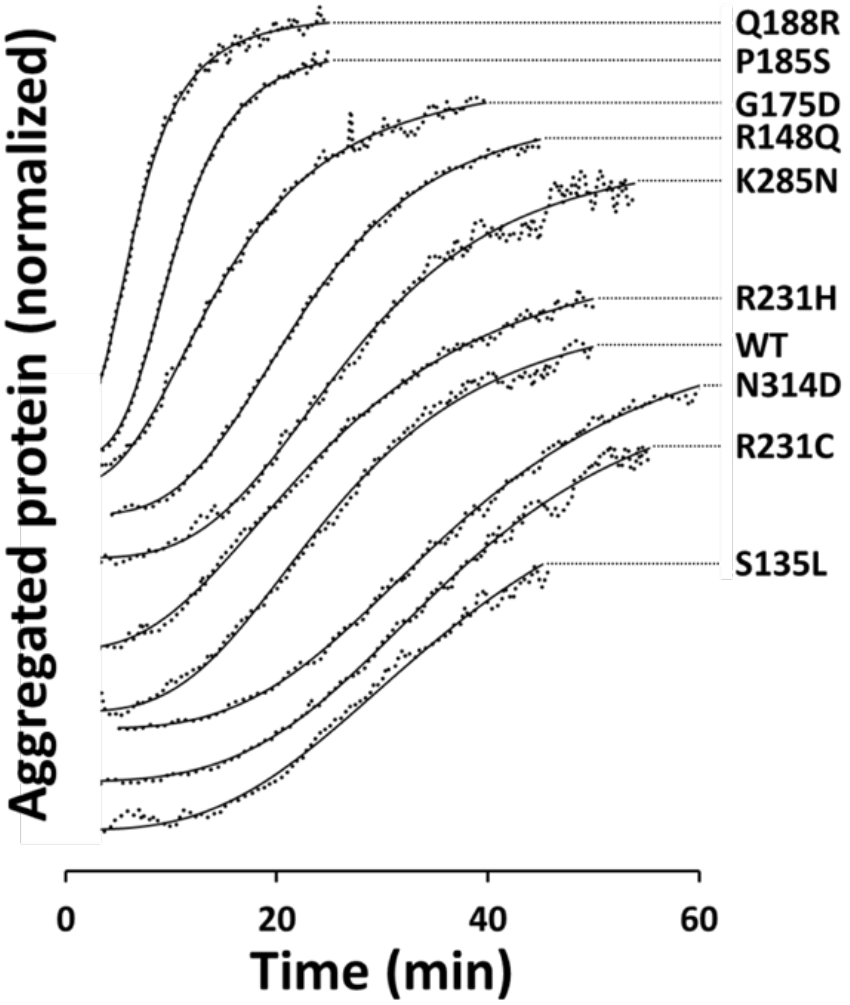
[a] F, forward; R, reverse. [b] mutagenesis sites are underlined



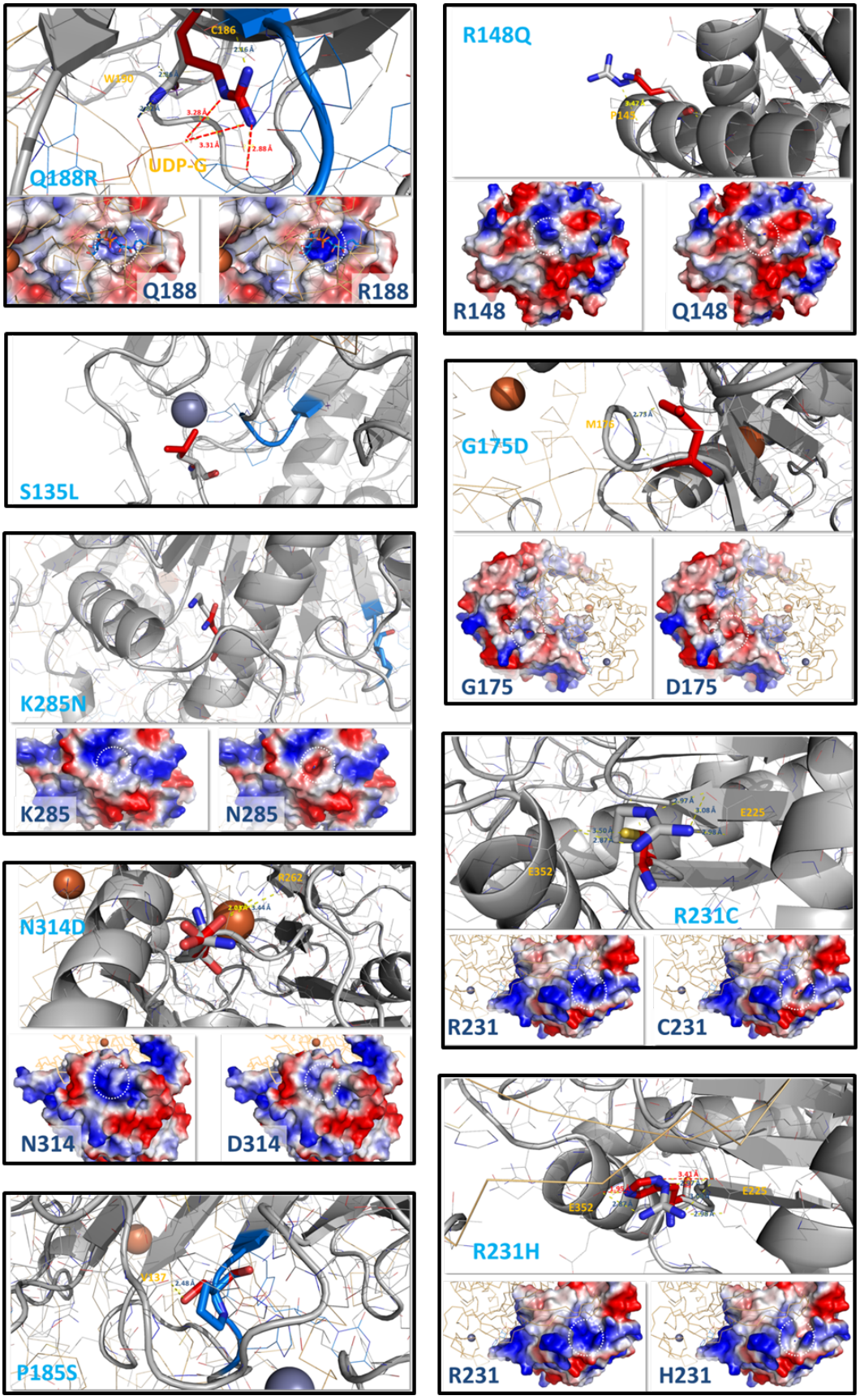
Supplementary Figure S4.1. 'Ground-state' extrinsic fluorescence of GALT variants in the presence of fluorescent dye targeting hydrophobic regions. Thermal denaturation profiles were obtained by differential scanning fluorimetry assays, probing the impact of mutations on the tertiary structure of GALT variants. 'Ground-state' extrinsic fluorescence was estimated as the averaged fluorescence intensities recorded in the first asymptote (20-30 °C) of the sigmoidal thermal denaturation profiles. Histogram represents 'ground-state' fluorescence of GALT variants normalized for the wild-type GALT values. Dashed line represents the wild-type GALT level.



Supplementary Figure S4.2 - The studied mutations have a limited impact on the tertiary structure of GALT variants. Thermal denaturation profiles were obtained by differential scanning fluorimetry assays, probing the impact of somatic mutations on the tertiary structure of GALT variants. The reaction mixture, totaling 50 µl, contained 0.1 mg/mL protein (~2.5 µM in monomer) in 50 mM Tris-HCl, 300 mM KCl, 10 % glycerol, pH 7.5, and 5x SYPRO Orange. After a 10-min incubation at 20 °C, temperature was linearly increased from 20 to 90 °C at 1 °C/min, with HEX channel fluorescence acquisition every 0.2 °C. Temperature scan curves were averaged, normalized and fitted to a biphasic dose-response. T_m values were estimated from the inflexion points of the first and second transitions (Table 4.1).



Supplementary Figure S4.3 - Thermal aggregation kinetics probed by dynamic light scattering. Kinetics of thermal aggregation monitored at 37 °C for 60 min. Light scattering intensity are plotted as a function of time, sigmoidal curves were obtained and the $t_{1/2}$ was defined as the time elapsed to reach 50% of maximum of aggregation. Asymptotes were removed for clarity, due to the data noise in those regions of the profiles.



Supplementary Figure 4.4. Structural impact of the studied mutations in human GALT. Structural model of human mutant GALT (grey cartoon) and opposing monomer from *Escherichia coli* GalT (orange ribbon, PDB ID: 1GUP). Substituting residues in red sticks. Surface electrostatics maps: red, negative charge; blue, positive charge; white, neutral. Figure generated with PyMOL.

CHAPTER V

Studying the potential role of arginine in rescuing GALT mutants

Ana I. Coelho[1], Matilde Trabuco[1], Maria João Silva[1,2], Isabel Tavares de Almeida[1,2], Paula Leandro[1,2], Isabel Rivera[1,2], João B. Vicente[1,2]

[1] Metabolism & Genetics Group, Research Institute for Medicines and Pharmaceutical Sciences (*i*Med.UL), Faculty of Pharmacy, University of Lisbon, Portugal

[2] Department of Biochemistry and Human Biology, Faculty of Pharmacy, University of Lisbon, Portugal

unpublished results

5.1. Abstract

Classic galactosemia is a rare metabolic disease resulting from deficient activity of galactose-1-phosphate uridylyltransferase (GALT), a key enzyme in the galactose metabolism. This disorder is characterized by a remarkable allelic heterogeneity in the *GALT* gene, in which the vast majority of the variations thus far described are missense mutations.

Functional-structural studies on several human GALT mutants have demonstrated that the most prevalent *GALT* mutations originate proteins with disturbed aggregation behaviour, which is actually an increasingly recognized pathogenic mechanism in several disorders.

Arginine has been described as a stabilizer of aggregation-prone proteins, with an evidence-based therapeutic effect in inherited metabolic disorder. Thus, we have developed a prokaryotic model of galactose sensitivity to evaluate human GALT function, particularly of clinically relevant mutants.

Thus far, no clear correlation between specific activity determined *in vitro* and the degree of toxicity could, however, be established. Nevertheless, it is possible that some hGALT mutants present some residual activity *in vivo* that is not detectable when assayed *in vitro* with purified protein. Moreover, arginine revealed to have a mutation-specific beneficial effect, specifically in the p.Q188R, p.K285N, and p.G175D GALT mutants, which led us to hypothesize that arginine might be of some benefit in classic galactosemia. Further studies are underway to ascertain the potential therapeutic effect of arginine. Nonetheless, these results lay a foundation for future studies using the prokaryotic model of galactose sensitivity.

5.2. Introduction

Classic galactosemia (OMIM #230400) is a rare metabolic dis resulting from deficient activity of galactose-1-phosphate uridylyltransferase (GALT, EC 2.7.7.12), the second enzyme of the Leloir pathway (1). This inherited metabolic disorder is a potentially lethal disorder that develops in the neonatal period, upon exposure to galactose in milk (1-3). The present gold standard of care is a lifelong dietary galactose restriction, which, notwithstanding its irrefutable life-saving role against the acute symptoms, fails to prevent cognitive, motor and fertility impairments (1, 4, 5).

Classic galactosemia is caused by mutations in the *GALT* gene, in which more than 260 variations have been described, the majority being missense mutations (>60%) (6). The *GALT* gene is arranged into 11 exons spanning ~4.0 kb of genomic sequence, and encodes a 379 amino acid protein, which is assembled as a ~87 kDa homodimer (7, 8). GALT is highly conserved in nature, particularly its active site sequence $His_{184}Pro_{185}His_{186}$, and each monomer has been

proposed to contain one zinc ion and one iron ion, based on studies on bacterial GalT (9-11).

Functional-structural studies on several human GALT mutants have demonstrated that the most prevalent *GALT* mutations do not seem to significantly affect the protein structure either at the secondary or tertiary levels; rather, they seem to affect the protein's aggregation behaviour in solution. In particular, p.Q188R, the most prevalent *GALT* mutation, originates a protein considerably more prone to aggregation than the wild-type GALT (12). In fact, protein aggregation is an increasingly recognized pathogenic mechanism in several disorders – including inherited metabolic disorders – and is believed to severely interfere with the cellular homeostasis (13, 14). Notably, studies on galactosemia models have revealed increased ER stress (15), unfolded protein response (16) and oxidative stress levels (17), hallmarks of perturbations in the cellular protein homeostasis.

The well-established effect of arginine as aggregation suppressant in globular proteins (18, 19), and the previous report of a pyruvate dehydrogenase complex deficient patient whose biochemical and clinical symptoms significantly improved due to arginine intake (20, 21), prompted us to develop the present study.

The aim of this study is to develop a prokaryotic model of galactose sensitivity of several human GALT mutants and to evaluate the effect of arginine on rescuing galactose-induced toxicity.

5.3. Materials and Methods

The *Escherichia coli* K-12 Δ*galT* strain (JW0741-1; *ΔgalT730::kan*) with a deletion of the endogeneous *galT* gene was purchased from the Coli Genetic Stock Center (CGSC) (22). The human *GALT* cDNA sequence was cloned into the ampicillin-resistant pTrcHis A plasmid, yielding an N-terminally hexa-histidyl-tagged protein. Site-directed mutagenesis (NZY mutagenesis kit, NZYTech, Lisbon, Portugal) was employed to introduce all the mutations herein under study using the primers listed in Table 1. Direct sequencing in both forward and reverse orientations was used to confirm the correct introduction of mutations and to exclude additional mutational events.

Vectors bearing the cDNA encoding the human GALT (hGALT) variants were transformed into the above mentioned bacterial strain. Additionally, the human *PAH* (phenylalanine hydroxylase) cDNA, cloned into the pTrcHis A vector, was assayed in parallel, as a negative control. All cultures were grown in M9 minimal medium (23) containing glycerol (1%) as sole carbon source. A starter culture was used to normalize cell concentration in all cultures at a starting optical density at 600 nm (OD_{600nm}) of 0.05. Cultures were grown at 37 °C and when an OD_{600nm} of 0.3 was reached, hGALT expression was induced by adding 250 μM

isopropyl-D-thiogalactoside (IPTG). Ferrous ammonium sulfate (100 μM) and zinc sulfate (100 μM) were also added to all cultures, and arginine (25 mM) was added to conditions III and IV (Table 2). One hour after inducing hGALT expression, 1% galactose was added to conditions II and IV (Table 2). Table 2 describes the four different culture conditions that were used. In order to follow culture growth, OD_{600nm} was measured hourly, starting at induction time and up to 9 hours. Assays were performed in at least three independent experiments.

Table 5.1. Oligonucleotides used for site-directed mutagenesis.

Primers [a]	**Sequence (5'→3')** [b]
Q188R-F	CCC CAC CCC CAC TGC CGG GTA TGG GCC AGC AG
Q188R-R	CTG CTG GCC CAT ACC CGG CAG TGG GGG TGG GG
S135L -F	GCT TCC ACC CCT GGT TGG ATG TAA CGC TGC
S135L-R	GCA GCG TTA CAT CCA ACC AGG GGT GGA AGC
K285N-F	GAA GAA GCT CTT GAC CAA TTA TGA CAA CCT CTT TGA G
K285N-R	CTC AAA GAG GTT GTC ATA ATT GGT CAA GAG CTT CTT C
N314D-F	GGC TGG GGC CAA CTG GGA CCA TTG GCA GCT GC
N314D-R	GCA GCT GCC AAT GGT CCC AGT TGG CCC CAG CC
R148Q-F	GGT CCC TGA GAT CCA GGC TGT TGT TGA TGC
R148Q-R	GCA TCA ACA ACA GCC TGG ATC TCA GGG ACC
G175D-F	GCA GAT CTT TGA AAA CAA AGA TGC CAT GAT GGG CTG TTC
G175D-R	GAA CAG CCC ATC ATG GCA TCT TTG TTT TCA AAG ATC TGC
P185S-F	GCT GTT CTA ACC CCC ACT CCC ACT GCC AGG
P185S-R	CCT GGC AGT GGG AGT GGG GGT TAG AAC AGC
R231C-F	GCT ACT CAG GAA GGA ATG TCT GGT CCT AAC CAG TG
R231C-R	CAC TGG TTA GGA CCA GAC ATT CCT TCC TGA GTA GC
R231H-F	GCT ACT CAG GAA GGA ACA TCT GGT CCT AAC CAG TG
R231H-R	CAC TGG TTA GGA CCA GAT GTT CCT TCC TGA GTA GC

[a] F, forward; R, reverse. [b] mutagenesis sites are underlined

Growth curves were obtained from plotting OD_{600} from cultures I and II (Table 2) as a function of time. To directly evaluate the galactose toxicity and the arginine effect, ratios – r and r_{arg} – were calculated according to equations 1 and 2, respectively, and were plotted as a function of time.

Eq. 1
$$r = \frac{OD_{600}\ \text{gal (II)}}{OD_{600}\ \text{gly (I)}}$$

Eq. 2
$$r_{arg} = \frac{OD_{600}\ \text{gal+arg (IV)}}{OD_{600}\ \text{gly+arg (III)}}$$

The growth of the non-transformed bacterial strain was also assessed in parallel to the bacterial strain expressing the wild-type hGALT, in a M9 minimal medium containing the following sugars as sole carbon source: i) glucose (1%); ii) glucose (1%) and galactose (1%); iii) glycerol (1%); or iv) galactose (1%).

Table 5.2. Culture conditions used in this study.

	Supplementation*	
	Galactose	**Arginine**
I	-	-
II	+	-
III	-	+
IV	+	+

* All cultures were grown in M9 minimal medium containing glycerol as carbon source supplemented with 100 μM ferrous sulfate and 100 μM zinc sulfate.

Recombinant production of hGALT and hPAH was confirmed by immunoblotting analysis. Briefly, 8 μg of total protein (bacterial soluble lysate) were applied on a 12.5% SDS-PAGE and blotted into a nitrocellulose membrane (Hybond ECL, Amersham, GE Healthcare Biosciences, Uppsala, Sweden), which, after a 2-hour blocking with 5 % milk in PBS-T buffer, was probed at room temperature with the primary mouse monoclonal anti-GALT antibody at a 1:1000 dilution (sc-365577, Santa Cruz Biotechnology, Santa Cruz, CA, USA) for hGALT variants detection, and with the primary mouse monoclonal anti-His antibody at a 1:1000 dilution (27-4710-01, GE Healthcare Biosciences) for hPAH detection. After three washes with PBS-T buffer, the membrane was incubated for 1 hour at room temperature with the secondary peroxidase-conjugated Affinipure goat anti-mouse IgG (H + L) antibody (Jackson ImmunoResearch Laboratories, West Grove, PA, USA), the ECL Prime Western Blotting Detection Reagent was used for protein detection (Amersham, GE Healthcare Biosciences).

5.4. Results and Discussion

5.4.1. Bacterial model of galactose sensitivity

Previous studies have reported a yeast galactosemia model allowing the evaluation of hGALT mutations severity, by assaying the sensitivity of transformed yeast cultures to galactose added to the medium. This model presents the inherent advantage of being assayed *in vivo*, and provides valuable insights on mutations impact on hGALT function (24, 25). Since yeast cultures are technically more demanding and time-consuming, we sought to develop a prokaryotic model to evaluate the ability of previously uncharacterized mutants hGALT in alleviating galactose toxicity.

The growth of the non-transformed Δ*galT* strain, in parallel with wild-type hGALT transformed strain, was assessed in the presence of glucose (i), glucose and galactose (ii), glycerol (iii), and galactose (iv) as sole carbon sources. As expected, in conditions (i) and (ii), the growth curves are indistinguishable, since glucose is the favored carbon source, and galactose represents no toxicity in the presence of this hexose; in the presence of glycerol (condition iii),

the growth is ~50% of that observed in the presence of glucose, confirming that this polyol is not as good as glucose as a carbon source; whereas in the presence of galactose (condition iv), the growth is practically null, suggesting a high toxicity of this sugar (data not shown). In turn, the bacterial strain expressing the wild-type hGALT presented growth curves identical to the non-transformed strain in conditions i, ii and iii, whereas in the presence of galactose (condition iv), the growth was significantly higher. These data strongly suggest the expression of hGALT is alleviating galactose toxicity in *E. coli* Δ*galT*.

5.4.2. Recombinant expression of hGALT variants in the bacterial model of galactose sensitivity

Since the bacterial *galT* gene in the used strain is deleted, the transformed pTrcHis A vectors carrying the human *GALT* cDNAs (wild-type and mutants) are the only source of uridylyltransferase activity. Bacterial cultures expressing each of the hGALT variants or hPAH, in four different medium conditions (Table 2), were followed by recording OD_{600} as a function of time; the resulting profiles were evaluated focusing on galactose toxicity (Figure 5.1 to 5.4, upper panels). Moreover, production of recombinant hGALT variants and hPAH was confirmed by immunoblotting analysis (data not shown).

The bacterial strain expressing wild-type hGALT was used as the positive control, whose growth should not be affected by the presence of galactose. Indeed, we could observe that the addition of 1% galactose had no significant influence on the bacterial growth, indicating that wild-type hGALT is able to alleviate galactose toxicity (Figure 5.1.A, upper panel).

Bacteria expressing hPAH were used as the negative control, with galactose severely arresting the culture growth. Interestingly, using glycerol as sole carbon source (culture condition I, Table 2), hPAH culture exhibits a much higher growth rate than hGALT (wild-type and mutants) in the same conditions (Figure 5.1.B, upper panel). This may arise from the fact that producing different recombinant proteins has different energy costs to the bacterial cells.

The p.N314D, p.S135L, p.R231C and p.R231H variants show essentially no growth arrest upon galactose addition, presenting a galactose growth curve essentially superimposable to that of glycerol (Figure 5.2, upper panel). The p.Q188R, p.K285N and p.G175D hGALT mutants show a growth profile intermediate between that of positive and negative controls, exhibiting a variable galactose toxicity (Figure 5.3, upper panel). On the other hand, the p.R148Q and p.P185S hGALT mutants show a galactose-sensitive profile similar to the negative control hPAH (Figure 5.4, upper panel). Despite the galactose sensitivity, p.P185S presents the highest absolute growth of all hGALT variants (maximum OD_{600} ~4) (Figure 5.4.B, upper panel). This

observation suggests that, not only the production of different proteins, but specifically the production of different variants poses different energy costs to the bacteria.

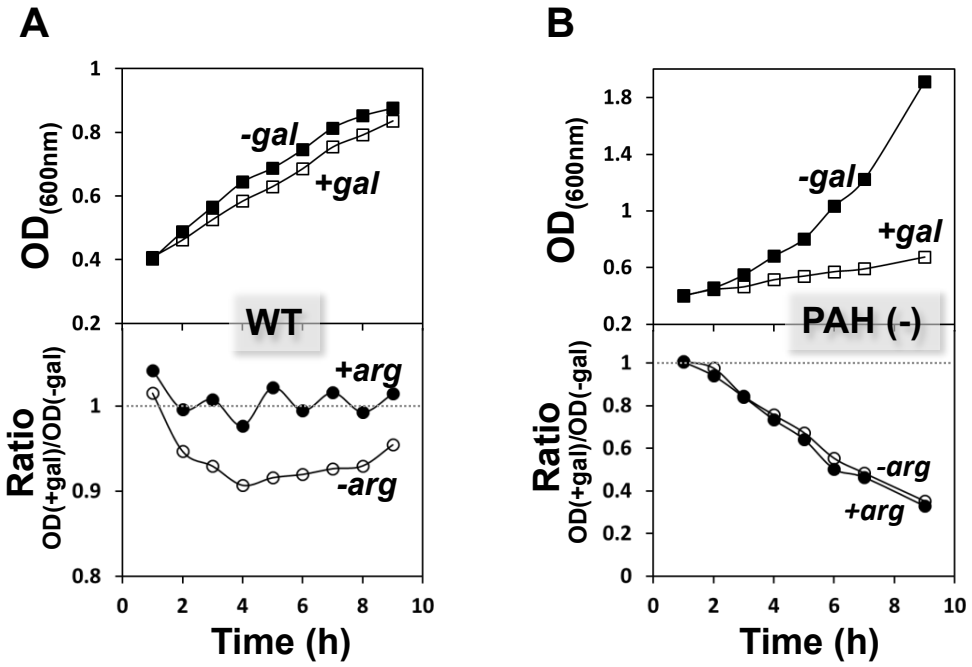


Figure 5.1. Growth and ratio curves of wild-type hGALT and hPAH. Growth curves (upper panel) of wild-type hGALT (A., positive control) show growth is not affected by the presence of galactose, whereas those of hPAH (B., negative control) show growth is severely arrested by the presence of galactose. The ratio curves (lower panel) of wild-type hGALT (A., positive control) show a slight degree of response to arginine, whereas those of hPAH (B., negative control) show no decrease in galactose toxicity due to the presence of arginine. (WT – wild-type)

No clear correlation between specific activity and degree of toxicity could be established. Most of the studied variants present specific activities either essentially null (p.K285N, p.R148Q, p.P185S, p.R231C and p.R231H) or ≤ 0.2% of wild-type activity (p.Q188R, p.S135L, p.G175D) (12). Indeed, taking also into account the severe galactose toxicity exhibited by hGALT mutants in the yeast model (25), many of these variants were expected to exhibit growth profiles similar to the negative control. Therefore, it is possible that some hGALT mutants present some residual activity *in vivo* that, however, when determined *in vitro* with the isolated proteins, is not detectable. Another hypothesis for the absence of a clear inverse correlation between hGALT activity *in vitro* and galactose sensitivity *in vivo* would be existence of an alternative bacterial pathway to the galactose metabolism. This hypothesis was, however, excluded by evaluating the bacterial growth of the non-transformed strain, which, as described above, exhibited an essential null growth in the presence of galactose.

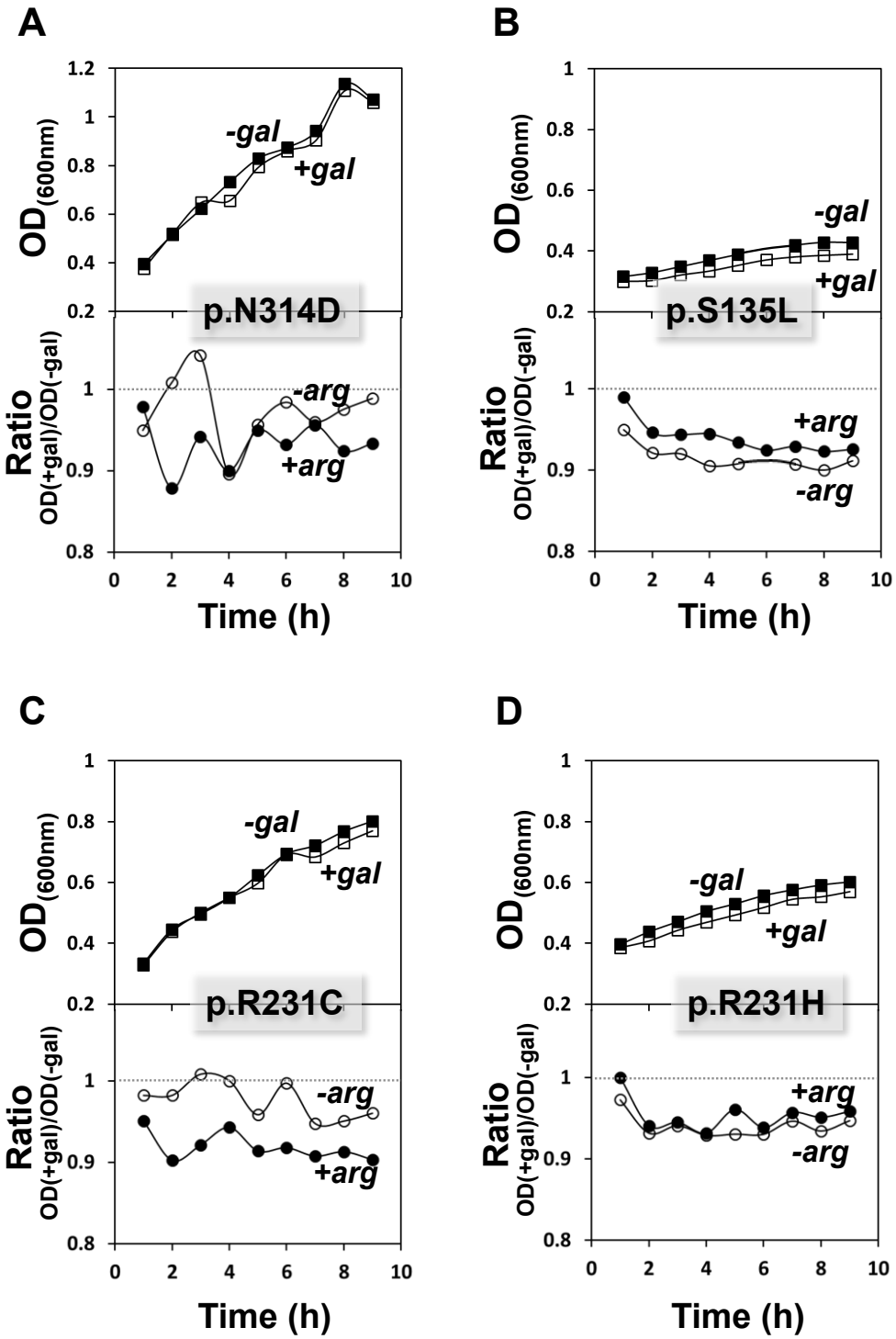


Figure 5.2. Growth and ratio curves of p.N314D, p.S135L, p.R231C and p.R231H hGALT mutants. Growth curves show essentially no growth arrest upon galactose addition (upper panel); and ratio curves suggest an overall insensitivity to arginine supplementation (lower panel).

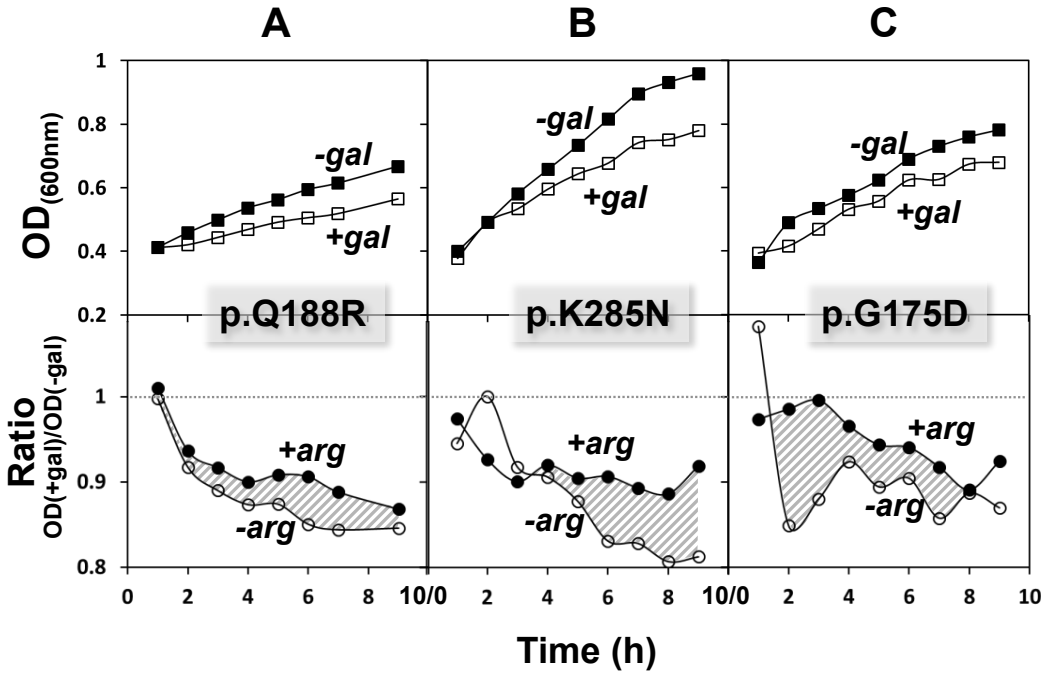


Figure 5.3. Growth and ratio curves of p.Q188R, p.K285N and p.G175D hGALT mutants. Growth curves show a growth profile intermediate between that of positive and negative controls (upper panel); and ratio curves (lower panel) suggest a partial rescue by arginine ($r_{arg} > r$, shaded area).

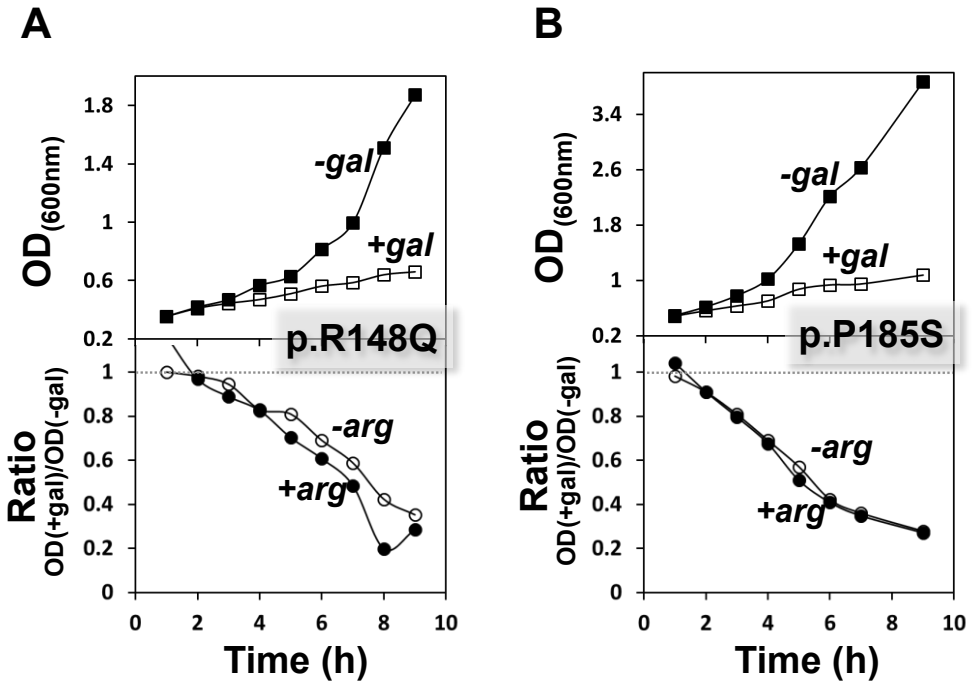


Figure 5.4. Growth and ratio curves of p.R148Q and p.P185S hGALT mutants. Growth curves suggest a galactose-sensitivity profile similar to that of the negative control hPAH (upper panel) and ratio curves show unresponsiveness to arginine supplementation (lower panel).

5.4.3. Arginine effect on the bacterial model of galactose sensitivity

Arginine has been described as therapeutically beneficial in a number of studies (20, 26). In fact, arginine is a long-recognized protein stabilizer that has been proposed to exert an anti-aggregation effect by increasing the activation energy of protein association/aggregation (18). Since the most prevalent hGALT variants revealed to be more prone to aggregation than the wild-type hGALT (12), arginine could potentially exert a protective stabilizing effect and constitute a novel therapeutic agent in classic galactosemia.

From the ratio curves, we can observe that the wild-type hGALT culture presents a slight degree of response to the medium supplementation with arginine (Figure 5.1.A, lower panel), which is not surprising since it has been described that arginine is also able to stabilize the proteins' native state (27). In contrast, the hPAH ratio curves are superimposable, indicating its insensitivity to arginine supplementation.

Interestingly, p.N314D, which is believed to be the ancestral allele (28), appears to be essentially insensitive to arginine (figure 5.2 A lower panel); as the same for the p.S135L and p.R231H cultures, which show essentially overlapping ratio curves (Figure 5.2, B. and D., lower panel). The p.R231C mutant, affecting the same residue as the latter variant, also presents almost superimposable ratio curves, but with the puzzling feature of a slightly increased toxicity in the presence of arginine. The p.Q188R, p.K285N, and p.G175D cultures are partially rescued by arginine, with $r_{arg} > r$ (Figure 5.3, lower panel). The rescue of p.Q188R is particular intriguing, since it is considered a severe mutation and has been described as particularly prone to aggregation (12). Immunoblotting analysis of the soluble lysate revealed that this mutant is actually highly expressed, which may account for the unexpected high tolerance (data not shown). In turn, p.R148Q and p.P185S ratio curves are very similar to the hPAH negative control, revealing their functional impairment and unresponsiveness to the potential stabilizing effect of arginine (Figure 5.4, lower panel).

The reproducibility of the ratio curves clearly indicate that the arginine mode of action is mutation-specific, suggesting it is indeed alleviating the galactose toxicity in the p.Q188R, p.K285N, and p.G175D mutants. These results here presented put forward the hypothesis that arginine might be of some benefit in classic galactosemia and, despite being preliminary, set the stage for further detailed studies.

5.5. References

1.Fridovich-Keil JL, Walter JH. *Galactosemia*. In: Valle D, Beaudet AL, Vogelstein B, Kinzler KW, Antonarakis SE, Ballabio A, editors. The Online Metabolic and Molecular Bases of Inherited Disease: Mc-Graw Hill; 2008. p. 1-92.

2.Berry GT, Walter JH. *Disorders of Galactose Metabolism*. In: Saudubray JM, van den Berghe G, Walter JH, editors. Inborn Metabolic Diseases: Diagnosis and Treatment. 5th ed: Springer; 2012.

3.Fridovich-Keil JL. *Galactosemia: the good, the bad, and the unknown*. J Cell Physiol. 2006;209(3):701-5.

4.Bosch AM. *Classical galactosaemia revisited*. J Inherit Metab Dis. 2006;29(4):516-25.

5.Waggoner DD, Buist NRM, Donnel GN. *Long-term prognosis in galactosaemia: results of a survey of 350 cases*. J Inherit Metab Dis. 1990;13(6):802-18.

6.Calderon FR, Phansalkar AR, Crockett DK, Miller M, Mao R. *Mutation database for the galactose-1-phosphate uridyltransferase (GALT) gene*. Hum Mutat. 2007;28(10):939-43.

7.Leslie ND, Immerrman EB, Flach JE, Florez M, Fridovich-Keil JL, Elsas II LJ. *The human galactose-1-phosphate uridyltransferase gene*. Genomics. 1992;14(2):474-80.

8.McCorvie TJ, Gleason TJ, Fridovich-Keil JL, Timson DJ. *Misfolding of galactose 1-phosphate uridylyltransferase can result in type I galactosemia*. Biochim Biophys Acta. 2013;1832(8):1279-93.

9.Geeganage S, Frey PA. *Significance of metal ions in galactose-1-phosphate uridylyltransferase: an essential structural zinc and a nonessential structural iron*. Biochemistry. 1999;38(40):13398-406.

10.Ruzicka FJ, Wedekind JE, Kim J, Rayment I, Frey PA. *Galactose-1-phosphate uridylyltransferase from Escherichia coli, a zinc and iron metalloenzyme*. Biochemistry. 1995;34(16):5610-7.

11.Wedekind JE, Frey PA, Rayment I. *Three-dimensional structure of galactose-1-phosphate uridylyltransferase from Escherichia coli at 1.8 Å resolution*. Biochemistry. 1995;34(35):11049- 61.

12.Coelho AI, Trabuco M, Ramos R, Silva MJ, Tavares de Almeida I, Leandro P, Rivera I, Vicente JB. *Functional and structural impact of the most prevalent missense mutations in classic galactosemia*.submitted.

13.Pedersen CB, Bross P, Winter VS, Corydon TJ, Bolund L, Bartlett K, Vockley J, Gregersen N. *Misfolding, degradation, and aggregation of variant proteins. The molecular pathogenesis of short chain acyl-CoA dehydrogenase (SCAD) deficiency*. J Biol Chem. 2003;278(48):47449-58.

14.Gregersen N, Bross P, Jorgensen MM. *Protein folding and misfolding: the role of cellular protein quality control systems in inherited disorders*. In: Valle D, Beaudet AL, Vogelstein B, Kinzler KW, Antonarakis SE, Ballabio A, editors. The Online Metabolic and Molecular Bases of Inherited Disease (OMMBID). General Themes: Mc-Graw Hill; 2005.

15.Mulhern ML, Madson CJ, Danford A, Ikesugi K, Kador PF, Shinohara T. *The unfolded protein response in lens epithelial cells from galactosemic rat lenses*. Invest Ophthalmol Vis Sci. 2006;47(9):3951-9.

16.De-Souza EA, Pimentel FS, Machado CM, Martins LS, da-Silva WS, Montero-Lomeli M, Masuda CA. *The unfolded protein response has a protective role in yeast models of classic galactosemia*. Dis Model Mech. 2014;7(1):55-61.

17.Jumbo-Lucioni PP, Hopson ML, Hang D, Liang Y, Jones DP, Fridovich-Keil JL. *Oxidative stress contributes to outcome severity in a Drosophila melanogaster model of classic galactosemia*. Dis Model Mech. 2013;6(1):84-94.

18.Baynes BM, Wang DIC, Trout BL. *Role of arginine in the stabilization of proteins against aggregation*. Biochemistry. 2005;44(12):4919-25.

19.Schneider CP, Trout BL. *Investigation of cosolute-protein preferential interaction coefficients: new insight into the mechanism by which arginine inhibits aggregation*. J Phys Chem B. 2009;113(7):2050-8.

20.Silva MJ, Pinheiro A, Eusebio F, Gaspar A, Tavares de Almeida I, Rivera I. *Pyruvate dehydrogenase deficiency: identification of a novel mutation in the PDHA1 gene which responds to amino acid supplementation*. Eur J Pediatr. 2009;168(1):17-22.

21.Florindo C. *Expressão heteróloga da forma selvagem e mutante da piruvato desidrogenase humana (hPDHE1): efeito de chaperones químicos* 2010.

22.Baba T, Ara T, Hasegawa M, Takai Y, Okumura Y, Baba M, Datsenko KA, Tomita M, Wanner BL, Mori H. *Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection*. Mol Syst Biol. 2006;2:2006 0008.

23.Maniatis T, Fritsch EF, Sambrook J. *Molecular cloning. A laboratory manual*. New York: Cold Spring Harbor; 1982.

24.Ross KL, Davis CN, Fridovich-Keil JL. *Differential roles of the Leloir pathway enzymes and metabolites in defining galactose sensitivity in yeast*. Mol Genet Metab. 2004;83(1-2):103-16.

25.Riehman K, Crews C, Fridovich-Keil JL. *Relationship between genotype, activity, and galactose sensitivity in yeast expressing patient alleles of human galactose-1-phosphate uridylyltransferase*. J Biol Chem. 2001;276(14):10634-40.

26.Berendse K, Ebberink MS, IJlst L, Wanders RJ, Waterham HR. *Arginine improves peroxisome functioning in cells from patients with a mild peroxisome biogenesis disorder*. Orphanet J Rare Dis. 2013;8(1):138-45.

27.Arakawa T, Tsumoto K. *The effects of arginine on refolding of aggregated proteins: not facilitate refolding, but suppress aggregation*. Biochem Biophys Res Commun. 2003;304(1):148-52.

28.Carney AE, Sanders RD, Garza KR, McGaha LA, Bean LJ, Coffee BW, Thomas JW, Cutler DJ, Kurtkaya NL, Fridovich-Keil JL. *Origins, distribution and expression of the Duarte-2 (D2) allele of galactose-1-phosphate uridylyltransferase*. Hum Mol Genet. 2009;18(9):1624-32.

CHAPTER VI

CONCLUDING REMARKS

The present work aimed to provide new insights on the molecular bases of classic galactosemia, and thus broaden the understanding on this enigmatic disorder. Our first goal focused on further elucidating the genetic bases of classic galactosemia, by characterizing the mutational spectrum in the Portuguese galactosemic population, since previous studies postulated that *GALT* genotyping provides added value in the establishment of prognosis for galactosemic patients (1, 2). This disorder is characterized by a wide clinical variability, and a long-pursued goal has been the identification of predictive factors to distinguish patients who will thrive in the long run from those who will experience complications (1). Accordingly, this work initiated with the molecular characterization of GALT deficient Portuguese patients, which revealed a marked allelic heterogeneity, as described for other populations (3, 4). Nevertheless, some mutations present a noteworthy prevalence, such as the missense p.Q188R mutation, which represents the vast majority of all mutant alleles (67%), followed by the splicing mutation, c.820+13A>G (8%). Establishment of genotype-phenotype correlations was subsequently attempted; however, no correlation was observed between *GALT* genotype and either biochemical or clinical phenotypes, even among siblings. In fact, recent studies involving siblings have shown that a too stringent diet may actually be harmful, reinforcing the idea that *GALT* genotyping is a determinant factor in classic galactosemia pathophysiology but only to a certain degree (5). Classic galactosemia has, in fact, been described as a complex disorder (1), characterized by a large biochemical and clinical variability (6), and whose pathophysiology, albeit being essentially unknown, is unquestionably also determined by other factors, such as genetic, epigenetic and environmental factors. Ultimately, the phenotype is beyond the control of the single gene itself (1).

In recent years, there has been an increasing awareness that deep intronic variations, and not just variations in the canonical splice sites, may affect splicing. Molecular characterization of Portuguese galactosemic patients revealed that the second most prevalent mutation is an intronic variation, until now classified as benign (7). Thus, we have characterized this mutation from a functional viewpoint by both *in vitro* and *in vivo* approaches. Our results confirmed that c.820+13A>G is a disease-causing mutation, whose underlying pathogenic mechanism involves the activation of a cryptic donor site located immediately downstream the variation. Intronic mutations activating cryptic splice sites are particularly amenable to antisense therapy, since the canonical splice sites remain intact, retaining the potential for restoring constitutive splicing (8). Hence, LNA oligonucleotides were designed to specifically recognize the mutation and have proved to successfully restore the constitutive splicing profile, thus establishing a proof-of-concept for the application of antisense therapy for mis-splicing mutations in classic galactosemia. It should be noted, however, that antisense therapy is a mutation-specific therapeutic approach and, thus, if proved to be applicable for galactosemic patients, it would be targeted to specific patients.

Missense mutations dominate the mutational spectrum of classic galactosemia, as for many other genetic disorders, and are undeniably the most common type of mutations in inherited metabolic disorders. Initially thought to affect exclusively the protein's functional residues, missense mutations are known to also affect the protein's conformational stability. In the *GALT* gene, patient-identified missense mutations are scattered throughout the whole coding region, which strongly suggests that GALT deficiency does not arise exclusively from mutations in functional sites. In reality, a number of *GALT* mutations have been described to originate misfolded proteins (9). In what regards the most prevalent missense mutations, several studies have focused on their functional characterization (10-14); however, a characterization of these variants focusing on different structural features was still lacking. Thus, we have performed a thorough structural-functional characterization of nine recombinant GALT variants, four of which result from the most prevalent mutations identified in *GALT* mutant alleles. Remarkably, no major structural effects were detected on their secondary and tertiary structures. Rather, most mutants displayed disturbed aggregation profiles. In particular, p.Q188R, the most prevalent *GALT* mutation, was shown to give rise to a protein considerably more prone to aggregation than the wild-type GALT. This observation, a novelty resulting from this work, is extremely important since, at the cellular level, the accumulation of aggregation-prone proteins interferes dramatically with the cellular homeostasis. Thus, classic galactosemia pathogenesis may be related, not merely with a deficient GALT activity, but also with its proteotoxicity. Notably, studies on galactosemia models have revealed increased ER stress, unfolded protein response and oxidative stress levels, hallmarks of perturbations in the cellular protein homeostasis (15-17). Furthermore, studies on inherited metabolic disorders had already disclosed that protein aggregation was actually a more common pathogenic mechanism than previously thought (18, 19), and classic galactosemia may now constitute another example of a metabolic disorder whereby protein aggregation is linked to disease. Identifying the possible layers of interaction between GALT mutants and the intricate protein quality control system will likely provide new insights on the molecular mechanisms involved in GALT mutants proteotoxicity, and possibly provide new clues on the phenotypic heterogeneity.

Our findings have important therapeutic implications, since one of the treatment strategies intensively explored in the last decade consists of the discovery and development of GALK inhibitors (20, 21). Nevertheless, if GALT mutants do indeed lead to proteotoxicity, this line of approach might be insufficient, since it would essentially control the Gal-1-P intracellular levels, and would, thus, fail to prevent GALT mutants aggregation toxicity.

We hypothesize that a reasonable strategy might be the use of proteostasis modulators, which would potentially extend the life-time of the GALT variants in the cell, partially compensating impaired function with enzyme availability, and simultaneously preventing accumulation of protein aggregates. Hopefully, this approach has the potential advantage of not

being mutation-specific and would thus have a therapeutic indication for a wider number of galactosemic patients.

We also sought to develop a prokaryotic model of galactose sensitivity, based on the previous description of a yeast model (22, 23). This bacterial model, lacking the endogenous *galT* gene, presents the inherent advantage of being assayed *in vivo*, which, in fact, might be crucial for some human GALT mutants, since their expression in our bacterial model was able to alleviate the galactose-induced toxicity, whereas *in vitro* exhibited essentially null specific activity. Furthermore, this unicellular model also presents the remarkable advantage of allowing the evaluation of potential therapeutic agents inferred from their ability to rescue galactose toxicity when expressing a specific human GALT mutant.

The previous report of a pyruvate dehydrogenase complex deficient patient whose biochemical and clinical symptoms significantly improved upon arginine intake (24, 25), and the long-recognized effect of arginine as anti-aggregation stabilizer in globular proteins (26, 27), set the stage for this particular study. Thus, we have expressed several human GALT in the *Escherichia coli* Δ*galT* strain, and evaluated the alleviation of galactose-toxicity. The p.R148Q and p.P185S curves suggested a severe functional impairment and a complete unresponsiveness to arginine, whereas the p.S135L, p.N314D, p.R231C and p.R231H present no significant improvement upon arginine supplementation, in line with the essentially absent galactose-induced toxicity. Hence, only the p.Q188R, p.K285N, and p.G175D variants were partially rescued and stand that arginine might be of some benefit for these GALT mutants.

These are, however, preliminary results, and further studies are underway to understand this mutation-specific mode of action of arginine. Nevertheless, this prokaryotic model shows high potential for the screening of possible therapeutic agents in classic galactosemia.

With the post-genomic era, has arisen the concept of personalized medicine, whereby the design of new therapeutic strategies is based on how the mutations take their toll (28). Accordingly, in this work, we have performed detailed molecular studies in classic galactosemia. We have described a particularly high prevalent intronic variation in the *GALT* gene and, accordingly, have developed a mutation-specific antisense therapy, thus establishing a proof-of-concept for this treatment approach in classic galactosemia. Moreover, this work provides a new piece in the complicated puzzle that classic galactosemia constitutes, by elucidating that the underlying molecular mechanism of several mutants – particularly of the most prevalent mutants – is mainly related to protein aggregation. Accordingly, we hypothesize that new therapeutic approaches, focusing on either preventing aggregation or promoting the degradation of these potentially toxic aggregates, may be beneficial in classic galactosemia. These insights constitute, indeed, a breakthrough in this enigmatic disorder, opening new avenues of research with the

long-term goal of developing new alternative therapies that can overcome the dramatic long-term complications experienced by most patients.

Despite the marked clinical heterogeneity and the complex pathophysiology, classic galactosemia is a single gene disorder, whose prospectives of treatment are, like for other monogenic disorders, closer than for non-mendelian disorders (29). Indeed, the future is now for rare genetic disorders (29). The future is now for classic galactosemia.

6.1 References

1.Tyfield L. *Galactosaemia and allelic variation at the galactose-1-phosphate uridyltransferase gene: a complex relationship between genotype and phenotype*. Eur J Pediatr. 2000;159(3):S204-S7.

2.Shield JPH. *The relationship of genotype to cognitive outcome in galactosaemia*. Arch Dis Child. 2000;83(3):248-50.

3.Tyfield L, Reichardt JK, Fridovich-Keil JL, Croke DT, Elsas II LJ, Strobl W, Kozak L, Coskun T, Novelli G, Okano Y, Zekanowski C, Shin Y, Boleda MD. *Classical galactosemia and mutations at the galactose-1-phosphate uridyl transferase (GALT) gene*. Hum Mutat. 1999;13(6):417-30.

4.Elsas II LJ, Lai K. *The molecular biology of galactosemia*. Genet Med. 1998;1(1):40-8.

5.Hughes J, Ryan S, Lambert D, Geoghegan O, Clark A, Rogers Y, Hendroff U, Monavari A, Twomey E, Treacy EP. *Outcomes of siblings with classical galactosemia*. J Pediatr. 2009;154(5):721-6.

6.Bosch AM. *Classical galactosaemia revisited*. J Inherit Metab Dis. 2006;29(4):516-25.

7.Calderon FR, Phansalkar AR, Crockett DK, Miller M, Mao R. *Mutation database for the galactose-1-phosphate uridyltransferase (GALT) gene*. Hum Mutat. 2007;28(10):939-43.

8.Rincon A, Aguado C, Desviat LR, Sanchez-Alcudia R, Ugarte M, Perez B. *Propionic and methylmalonic acidemia: antisense therapeutics for intronic variations causing aberrantly spliced messenger RNA*. Am J Hum Genet. 2007;81(6):1262-70.

9.McCorvie TJ, Gleason TJ, Fridovich-Keil JL, Timson DJ. *Misfolding of galactose 1-phosphate uridylyltransferase can result in type I galactosemia*. Biochim Biophys Acta. 2013;1832(8):1279-93.

10.Lai K, Willis AC, Elsas II LJ. *The biochemical role of glutamine 188 in human galactose-1-phosphate uridyltransferase*. J Biol Chem. 1999;274(10):6559-66.

11.Lai K, Elsas LJ. *Structure-function analyses of a common mutation in blacks with transferase-deficiency galactosemia*. Mol Genet Metab. 2001;74(1-2):264-72.

12.Fridovich-Keil JL, Jinks-Robertson S. *A yeast expression system for human galactose-l-phosphate uridylyltransferase*. Proc Natl Acad Sci U S A. 1993;90(2):398-402.

13.Wells L, Fridovich-Keil JL. *Biochemical characterization of the S135L allele of galactose-1-phosphate uridylyltransferase associated with galactosaemia*. J Inherit Metab Dis. 1997;20(5):633-42.

14.Fridovich-Keil JL, Quimby BB, Wells L, Mazur LA, Elsevier JP. *Characterization of the N314D allele of human galactose-1-phosphate uridylyltransferase using a yeast expression system*. Biochem Mol Med. 1995;56(2):121-30.

15.Mulhern ML, Madson CJ, Danford A, Ikesugi K, Kador PF, Shinohara T. *The unfolded protein response in lens epithelial cells from galactosemic rat lenses*. Invest Ophthalmol Vis Sci. 2006;47(9):3951-9.

16.De-Souza EA, Pimentel FS, Machado CM, Martins LS, da-Silva WS, Montero-Lomeli M, Masuda CA. *The unfolded protein response has a protective role in yeast models of classic galactosemia*. Dis Model Mech. 2014;7(1):55-61.

17.Jumbo-Lucioni PP, Hopson ML, Hang D, Liang Y, Jones DP, Fridovich-Keil JL. *Oxidative stress contributes to outcome severity in a Drosophila melanogaster model of classic galactosemia*. Dis Model Mech. 2013;6(1):84-94.

18.Gregersen N, Bross P, Jorgensen MM. *Protein folding and misfolding: the role of cellular protein quality control systems in inherited disorders*. In: Valle D, Beaudet AL, Vogelstein B, Kinzler KW, Antonarakis SE, Ballabio A, editors. The Online Metabolic and Molecular Bases of Inherited Disease (OMMBID). General Themes: Mc-Graw Hill; 2005.

19.Pedersen CB, Bross P, Winter VS, Corydon TJ, Bolund L, Bartlett K, Vockley J, Gregersen N. *Misfolding, degradation, and aggregation of variant proteins. The molecular pathogenesis of short chain acyl-CoA dehydrogenase (SCAD) deficiency*. J Biol Chem. 2003;278(48):47449-58.

20.Bosch A, Bakker HD, Van Gennip AH, Van Kempen JV, Wanders RJ, Wijburg FA. *Clinical features of galactokinase deficiency: A review of the literature*. J Inherit Metab Dis. 2002;25:629-34.

21.Tang M, Wierenga K, Elsas LJ, Lai K. *Molecular and biochemical characterization of human galactokinase and its small molecule inhibitors*. Chem Biol Interact. 2010;188(3):376-85.

22.Ross KL, Davis CN, Fridovich-Keil JL. *Differential roles of the Leloir pathway enzymes and metabolites in defining galactose sensitivity in yeast*. Mol Genet Metab. 2004;83(1-2):103-16.

23.Riehman K, Crews C, Fridovich-Keil JL. *Relationship between genotype, activity, and galactose sensitivity in yeast expressing patient alleles of human galactose-1-phosphate uridylyltransferase*. J Biol Chem. 2001;276(14):10634-40.

24.Silva MJ, Pinheiro A, Eusebio F, Gaspar A, Tavares de Almeida I, Rivera I. *Pyruvate dehydrogenase deficiency: identification of a novel mutation in the PDHA1 gene which responds to amino acid supplementation*. Eur J Pediatr. 2009;168(1):17-22.

25.Florindo C. *Expressão heteróloga da forma selvagem e mutante da piruvato desidrogenase humana (hPDHE1): efeito de chaperones químicos* 2010.

26.Baynes BM, Wang DIC, Trout BL. *Role of arginine in the stabilization of proteins against aggregation*. Biochemistry. 2005;44(12):4919-25.

27.Schneider CP, Trout BL. *Investigation of cosolute-protein preferential interaction coefficients: new insight into the mechanism by which arginine inhibits aggregation*. J Phys Chem B. 2009;113(7):2050-8.

28.Xie J. *Making use of aberrant and nonsense: aberrant splicing and nonsense-mediated decay as targets for personalized medicine*. Int J Genomic Med. 2013;1(105):2332-0672.

29.Mackenzie A, Boycott KM. *The future is now for rare genetic diseases*. CMAJ. 2012;184(14):1603.