Universidade de Lisboa

Faculdade de Ciências

Departamento de Biologia Vegetal

Splicing mutations:

making sense with antisense therapy

Sílvia Pires Lourenço

Dissertação

Mestrado em Biologia Molecular e Genética

2012

Universidade de Lisboa

Faculdade de Ciências

Departamento de Biologia Vegetal

Splicing mutations:
making sense with antisense therapy

Sílvia Pires Lourenço

Dissertação orientada por:

Orientadora externa: Prof.ª Doutora Isabel Antolin Rivera, iMed, FFUL.

Orientador interno: Prof. Doutor Francisco Dionísio, DBV, FCUL.

Mestrado em Biologia Molecular e Genética

2012

Index

Agradecimentos

É com grande orgulho que apresento esta tese em nome individual. Contudo, o resultado final deve-se a um trabalho conjunto e ao apoio de diversas pessoas às quais não posso deixar de agradecer.

À Professora Doutora Isabel Tavares de Almeida pela oportunidade que me foi dada de ingressar no seu grupo de investigação de Metabolismos e Genética, do iMed.UL, e que possibilitou a realização desta dissertação.

Um agradecimento muito especial à Professora Doutora Isabel Rivera. Pela confiança que depositou em mim desde o início, pela dedicação, ajuda, entusiasmo, disponibilidade, e discussões que tanto me ensinaram. Mas principalmente, por toda a amizade, apoio, carinho e compreensão que sempre demonstrou. Um MUITO obrigada por tudo.

Ao Professor Doutor Francisco Dionísio que se mostrou disponível para o que precisei.

À Ana I. Coelho que desde o dia um me ajudou em tudo e se mostrou disponível para esclarecer as minhas inúmeras dúvidas e sem a qual este trabalho não teria sido possível.

À Matilde, não só pela ajuda laboratorial que sempre facultou, mas especialmente pelas idas ao ginásio, pela amizade que criámos e que foi muito importante durante este ano.

À Doutora Belén Pérez pela ajuda e esclarecimentos prestados essenciais nesta tese.

À Sandra, por todos os e-mails trocados, fundamentais em muitas das minhas dúvidas.

Aos membros do Met&Gen, que não vou nomear não vá me esquecer de algum, que tanto me ajudaram. Por me integrarem no grupo e permitirem o bom ambiente que nele se vive. Pelos almoços, cafés, lanches e jantares que tornaram este ano muito mais especial.

A todos os meus amigos, por toda a força que sempre me deram, quer tenha sido nos nossos telefonemas, mensagens, cafés, jantares, saídas e cinemas, quer simplesmente pela ideia de saber que estavam ali para o que precisasse.

A vocês, Cori, Joana e Diana, por me aturarem e por serem mais que colegas de casa.

E por fim, porque não há agradecimentos suficientes, à minha família. O maior pilar que tenho na vida e que sei nunca me faltará. Aos meus pais, João e Etelvina, a quem devo tudo e com quem sei que poderei sempre contar. Pelo amor incondicional que me têm, e por possibilitarem a realização de um sonho. Por todos os telefonemas, especialmente naqueles momentos de incerteza em que nada há a dizer e resta ouvir a respiração do outro lado da linha. Às minhas irmãs, Ana, Cristiane e Andreia, que são essenciais na minha vida, que me apoiaram nesta jornada e me ensinaram a não desistir. Aos meus cunhados. E aos meus sobrinhos, Gonçalo, Cláudia, Lourenço, Lucas, Vasco e Catarina, que sempre me arrancam um sorriso, enchem a vida de alegria e tanto gostam de falar com a “titia Sílvia” ao telefone.

A todos um MUITO obrigada,

Sílvia Pires Lourenço

Glossary

3'ss	3' splice site
5'ss	5' splice site
A	adenosine
ATM	ataxia telangiectasia mutated
bp	base pair
BRCA1	breast cancer type 1 susceptibility protein
C	cytidine
cDNA	complementar DNA
CFTR	cystic fibrosis transmembrane conductance regulator
DMD	Duchenne muscular dystrophy
DMEM	Dulbecco's modified Eagle medium
DNA	deoxyribonucleic acid
E. coli	*Escherichia coli*
ESE	exonic splicing enhancer
ESS	exonic splicing silencer
G	guanosine
Gal1P	galactose 1-phosphate
GALE	UDP-galactose epimerase
GALK	galactokinase
GALM	galactose mutarotase
GALT	galactose 1-phosphate uridylyltransferase
HBB	hemogoblin beta
hnRNP	heterogeneous nuclear ribonucleoprotein
IEM	inborn error of metabolism
ISE	intronic splicing enhancer
ISS	intronic splicing silencer
kb	kilo base pairs
kDa	kilo Dalton
LNA	locked nucleic acid

LB	Luria-Bertani
min	minute(s)
mRNA	messenger RNA
mut	mutant
NBCS	new born calf serum
NMD	nonsense mediated decay
PCR	polymerase chain reaction
pre-mRNA	premature messenger RNA
PTC	premature termination codon
RBC	red blood cell
RNA	ribonucleic acid
RT-PCR	reverse transcription polymerase chain reaction
sec	second(s)
SMaRT	spliceosome-mediated RNA *trans*-splicing
snRNP	small nuclear ribonucleoprotein
SR	serine/arginine-rich proteins
T	thymidine
U	uridine
UDP	uridine diphosphate
UMP	uridine monophosphate
wt	wild-type

Abstract

Galactose-1-phosphate uridylyltransferase (GALT) transfers a UMP group from UDP-glucose to Gal1P in the second step of the Leloir pathway for galactose metabolism. Pathogenic mutations in the *GALT* gene cause deficient or absent activity of the enzyme and result in Classical Galactosemia.

Mutational analysis of 27 Portuguese patients confirmed Q188R as the prevalent molecular defect (≈60%), and revealed the intronic variation c.820+13a>g (IVS8+13a>g) as the second most frequent mutation, accounting for 12.5% of the mutant alleles.

In silico analysis revealed that the presence of the c.820+13a>g variation activates a cryptic splicing donor site and seems to create a strong ESE motif for the binding of the SRSF1 protein.

A minigene-based approach was used to analyze the effects of this presumptive pre-mRNA splicing mutation. The pSPL3 vector containing either the genomic wild-type or mutant fragments was transfected into COS-7 and Hek293 cell lines. 24h after transfection, RNA was purified and amplified by RT-PCR. Direct sequencing analysis of the reaction products clearly showed that c.820+13a>g favors the next GT dinucleotide (c.820+14_15) to be used as a new 5' splice donor site, leading to the inclusion of the first 13 nucleotides of intron 8 in the coding sequence, thus inducing a frameshift which generates a premature stop codon 17 amino acids downstream (p.D274fsX291).

Additionally, *in vivo* studies demonstrated that, contrarily to the control individuals, a homozygous patient for c.820+13a>g mutation only presented alternative transcripts with the inclusion of the first 13 nucleotides of intron 8.

Accordingly, c.820+13a>g may be considered indeed a disease-causing mutation, being the first intronic variation in the *GALT* gene whose molecular mechanism was elucidated.

In order to revert this alternative splicing, an antisense oligonucleotide approach was attempted, using a LNA (locked nucleic acid) that was cotransfected with the minigenes in COS-7 and Hek293 cells; however, this preliminary experience was not effective, so new alternatives will be developed.

Key-words: *GALT*, c.820+13a>g, alternative splicing, minigene assays, LNA.

Resumo

A galactose-1-fosfato uridililtransferase (GALT) é a enzima responsável pelo segundo passo da via de Leloir, a principal via catabólica da galactose no organismo humano. A GALT transfere um grupo UMP da UDP-glucose para a galactose-1-fosfato (Gal1P), libertando glucose-1-fosfato (Glc1P) e formando UDP-galactose. Uma ausente ou deficiente actividade de GALT é causadora de Galactosémia Clássica (GC). A GC é uma doença metabólica rara que apresenta prevalência diversa de acordo com as diferentes populações, afecta cerca de 1 em cada 47 mil nados vivos nas populações caucasianas europeias e está associada a mutações no gene *GALT*.

O gene *GALT* situa-se na região 13 do braço curto do cromossoma 9, cobre cerca de 4 kb de DNA genómico e organiza-se em 11 exões. Este gene codifica uma proteína de 379 aminoácidos com uma massa molecular de 43 kDa. A enzima activa é um homodímero com uma massa molecular total de 88 kDa. Até à data, foram já descritas mais de 200 variações no gene *GALT*. Embora a maioria sejam mutações pontuais do tipo *missense*, também foram detectadas mutações *nonsense*, alterações silenciosas, assim como deleções e mutações de *splicing*. Algumas destas variações são comuns, mas a maioria são raras e desconhece-se a sua importância funcional. As mais citadas são Q188R, K285N, S135L e N314D, sendo as duas primeiras as mais frequentes na população europeia.

Estudos mutacionais em 27 doentes Portugueses confirmaram que a mutação Q188R é a prevalente (≈60%), e revelaram que a variação intrónica c.820+13a>g (IVS8+13a>g) é a segunda mais frequente, estando presente em cerca de 12,5% dos alelos mutados.

Nas situações em que a conversão da Gal1P em Glc1P está comprometida, ocorre uma acumulação excessiva de galactose e de Gal1P, causando toxicidade no organismo. Actualmente, o único tratamento disponível consiste na restrição dietética de galactose, o qual diminui drasticamente a mortalidade neonatal por galactosémia. A longo prazo, apesar de salvar muitas vidas, esta terapia revela-se contudo insuficiente. Portanto, o desenvolvimento de novas terapias torna-se essencial.

Os objectivos do presente estudo foram a elucidação do mecanismo patogénico subjacente à mutação c.820+13a>g, a confirmação do seu efeito *in vivo* e, por fim, a tentativa da sua correção através de terapia com oligonucleótidos *antisense*.

Primeiramente, procedeu-se a um estudo *in silico* para prever a patogenicidade da mutação. O resultado revelou que a presença desta variação parece activar um sítio críptico (c.820+14_15) dador de *splicing*. Contudo, ao analisar a sequência mutada, este sítio críptico não foi reconhecido pelo programa NNSPLICE, mas apresentou uma pontuação de 0,70 no programa NetGene2. Por outro lado, aquando da análise da sequência mutada, este

5’GT críptico foi reconhecido pelos dois programas mencionados, apresentando uma pontuação de 0,95 e 0,93, respectivamente. Estas pontuações revelaram-se superiores às obtidas pelo sítio de *splicing* natural, o qual foi pontuado, na sequência mutada, com apenas 0,64 e 0,82 pelos programas NNSPLICE e NetGene2, respectivamente. Adicionalmente, na sequência mutada verificou-se o fortalecimento do motivo ESE (*exonic splicing enhancer*) para a ligação do factor de *splicing* SRSF1 (CCCAGGT, posições 9-15 do intrão 8). Com efeito, o ESE contendo o nucleótido mutado revelou uma pontuação de 3,906278 enquanto que a presença do nucleótido normal conferia apenas a pontuação de 1,986666, valor muito próximo do limiar do programa (1,956).

Para a caracterização do mecanismo molecular subjacente a esta potencial mutação de *splicing*, recorreu-se à utilização da tecnologia dos minigenes. Para tal, os fragmentos genómicos de interesse, correspondentes à sequência génica normal e mutada do gene *GALT*, foram clonados no vector de splicing pSPL3 e designados pSPL3.wt e pSPL3.mut, respectivamente. De seguida, procedeu-se à sua transfeção em duas linhas celulares eucariotas, COS-7 e Hek293. 24 horas após transfeção, extraiu-se o RNA total celular que foi utilizado para amplificação por RT-PCR das sequências transcritas *in vitro*. A análise em gel de agarose do perfil transcricional das duas linhas celulares demonstrou a existência de *splicings* alternativos, quer nas células transfectadas com o pSPL3.wt quer com o pSPL3.mut. A sequenciação de alguns destes fragmentos revelou que, tal como previsto *in silico*, a mutação favorece a utilização do 5’GT críptico que se localiza nos nucleótidos imediatamente a seguir, causando um *splicing* alternativo. Em ambas as linhas celulares transfectadas com pSPL3.mut, mas não nas transfectadas com pSPL3.wt, ocorre a inclusão dos primeiros 13 nucleótidos do intrão 8 nos transcritos. Esta inclusão na sequência codificante leva à mudança do quadro de leitura e à criação de um codão stop prematuro 18 aminoácidos a jusante (D274fsX291).

Paralelamente, estudos *in vivo* desmonstraram que, contrariamente ao esperado, os transcritos alternativos contendo um codão stop prematuro não sofreram NMD (*nonsense mediated decay*). De facto, a análise dos transcritos de um doente homozigótico para a mutação c.820+13a>g revelou, tal como nos indivíduos controlo, a presença de dois transcritos. Nos indivíduos controlo, o fragmento maior com 378pb corresponde a uma reação de *splicing* em que são utilizados os sítios naturais de *splicing* e, como tal, inclui o fim do exão 7, exão 8, exão 9 e início do exão 10; o fragmento menor com 294pb corresponde a um evento de *splicing* alternativo que leva à não inclusão do exão 9 na sequência final do mRNA. A análise do doente revelou o mesmo padrão de transcrição, diferindo os fragmentos observados na presença adicional dos 13 primeiros nucleótidos do intrão 8. Tendo em conta os resultados obtidos, foi possível concluir que o *splicing*

alternativo do exão 9 não é causado pela mutação c.820+13a>g, uma vez que também ocorre nos indivíduos controlo.

Desta forma, confirmou-se que a mutação c.820+13a>g (IVS8+13a>g) é de facto causadora de doença, sendo a primeira variação intrónica no gene *GALT* cujo mecanismo molecular foi elucidado.

De seguida, e com o intuito de reverter o *splicing* alternativo causado pela mutação, procedeu-se a uma tentativa da sua correção utilizando oligonucleótidos *antisense*. Para tal, e tendo em conta a proximidade da mutação ao sítio de *splicing* natural, selecionou-se a utilização de um LNA (*locked nuclei acid*). Após cotransfeção dos minigenes e do IVS8-LNA nas linhas celulares COS-7 e Hek293, isolou-se o RNA total celular e procedeu-se à respectiva amplificação por RT-PCR.

Os padrões de *splicing* alternativo observados em ambas as linhas celulares foram semelhantes, quer em células não-tratadas, quer em células tratadas. Adicionalmente, a sequenciação de alguns dos fragmentos obtidos revelou a presença dos 13 primeiros nucleótidos do intrão 8 tanto nas células transfectadas com pSPL3.mut como nas células cotransfectadas com pSPL3.mut e IVS8-LNA. Assim, esta experiência preliminar revelou que o tratamento não foi eficaz na correção do *splicing* alternativo, realçando a necessidade de melhorar a presente metodologia ou de procurar outras alternativas terapêuticas.

Palavras-chave: *GALT*, c.820+13a>g, splicing alternativo, minigene, LNA.

Introduction

Health and disease depend on the delicate homeostasis of human organism. Gene expression is finely regulated to ensure that the correct complement of RNA and proteins is present in the right cell at the right time [1]. Any modification from transcription to protein function, may lead to phenotypic modifications, responsible for a disease state [2]. In addition to the exonic mutations that directly affect the protein, in the last two decades molecular genetics research has elucidated several mechanisms that alter gene function, such as epigenetic factors, distant regulatory elements and small interfering RNAs, as well as pre-mRNA splicing defects. Indeed, this last type of alteration seems to play a role in almost all known diseases with a genetic etiology [3].

Splicing involves precise removal of introns from pre-mRNA in a way exons are spliced together to form mature mRNAs with intact translation reading frames [2]. Exons are recognized through GT and AG intronic dinucleotides at the 5' (donor) and 3' (acceptor) exon-intron junctions, respectively, and also by the branch site (Figure 1) [2,4].

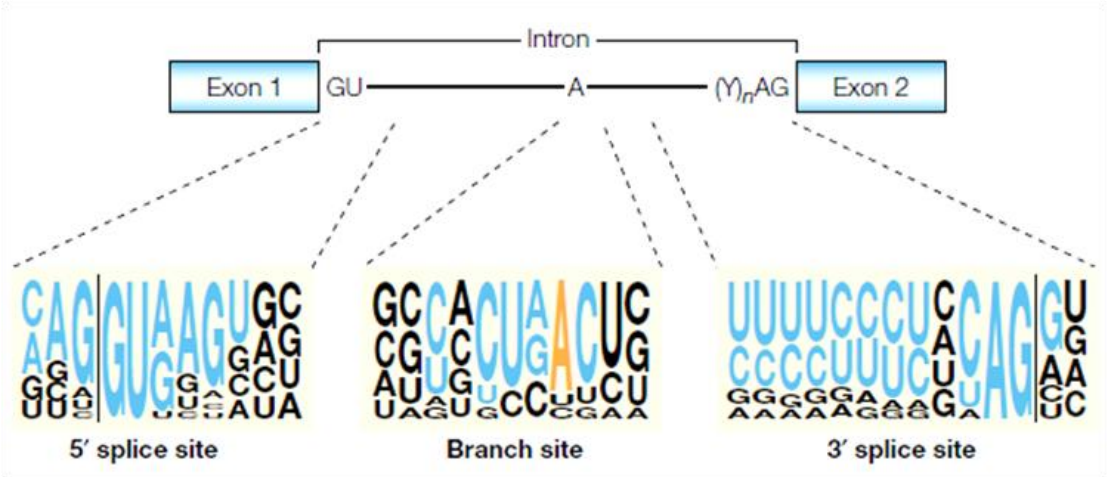

Figure 1 - Classical splicing signals.
Conserved motifs at or near the intron ends. The nearly invariant GU and AG dinucleotides at the intron ends, the polypyrimidine tract (Y)*n* preceding the 3′ AG, and the A residue that serves as a branchpoint are shown in a two-exon pre-mRNA. The sequence motifs that surround these conserved nucleotides are shown below. For each sequence motif, the size of a nucleotide at a given position is proportional to the frequency of that nucleotide at that position in an alignment of conserved sequences from 1,683 human introns. Nucleotides that are part of the classical consensus motifs are shown in blue, except for the branch-point A, which is shown in orange. The vertical lines indicate the exon–intron boundaries. [5].

Splicing is then carried out by the interaction of these sequences with the spliceosome and further cut-and-paste (transesterification) reactions [2,4]. The spliceosome is made up of five small nuclear ribonucleoproteins (snRNPs) – U1, U2, U4/U6 and U5 snRNPs - and a large number of non-snRNPs [2,4]. Each snRNP is composed of a single uridine-rich small nuclear RNA and multiple proteins [4]. Despite being necessary, these classical splicing signals are by no means sufficient to define exon-intron boundaries [5]. In fact, pseudo-exons

are very common in introns and although they match the consensus splice-site signals as efficiently as the natural splice sites do, or sometimes even better, they are normally not included in mature mRNAs [5].

To increase the overall fidelity of the splicing reaction, additional sequences are present in exons and introns. These *cis*-acting sequence elements can act by increasing or decreasing recognition and are, respectively, named exonic or intronic splicing enhancers (ESE, ISE) or silencers (ESS, ISS) [2,6].

ESEs stimulate the use of splice sites through the specific binding of regulatory proteins, such as serine/arginine-rich (SR) proteins, which are thought to have a role in the initial steps of spliceosome assembly. Sequences that act as exonic splicing silencers (ESSs) have also been described, but are less well characterized than ESEs. In some instances, ESSs have been shown to bind negative regulators belonging to the heterogeneous nuclear ribonucleoprotein (hnRNP) family [2, 5, 7]

ESEs and ESSs appear to play a relevant role in the definition of constitutive exons and, additionally, these sequences participate in the regulation of alternative splicing [7].

A very common event in mammalian cells, which makes the already formidable task of correctly identifying splice sites even more complex, is alternative splicing [5], corresponding to the joining of different 5' and 3' splice sites (Figure 2) [4]. This process is responsible for much of the complexity of the proteome because it allows individual genes to express multiple mRNAs that encode proteins with diverse and even antagonistic functions [4,5].

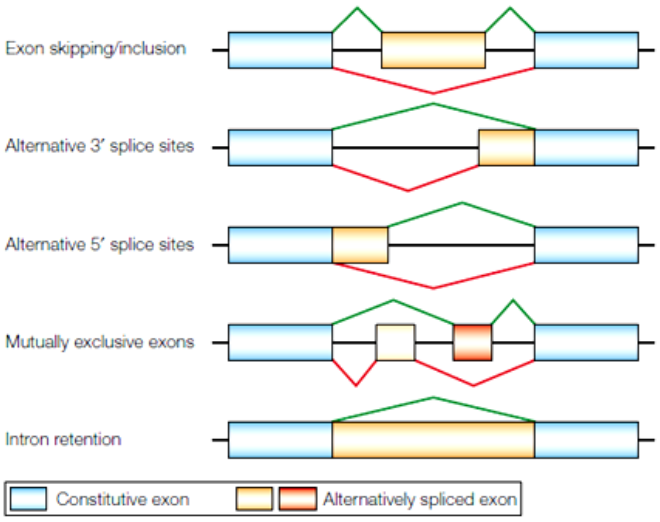

Figure 2 - Modes of alternative splicing.
Five common modes of alternative splicing. In each case, one alternative splicing path is indicated in green, the other path in red. In the last example, the alternative pathway corresponds to no splicing. In complex premRNAs, more than one of these modes of alternative splicing can apply to different regions of the transcript, and extra mRNA isoforms can be generated through the use of alternative promoters or polyadenylation sites. [5].

Until recently, the effect of a mutation on gene expression was generally assumed according to its location. Exonic mutations were assumed to cause disease by affecting exclusively the coding potential: silent mutations were considered neutral; missense mutations were presumed to identify amino acid that are important for protein structure and/or function; and nonsense mutations were assumed to lead to expression of nonfunctional or deleterious truncated proteins [5]. These assumptions might be correct in some cases, but when they are not supported by characterization at the mRNA level, they could be misleading. In fact, many of the exonic mutations once thought to disrupt protein function actually disrupt splicing of the exon that contains the substitution [8]. Moreover, as research has progressed, it has become clear that almost any genomic variation, including translationally silent ones, even when found within intronic regions, should be considered as potential disease-causing mutation due to aberrant splicing [3]. Indeed, there is growing evidence that misclassification of mutations might commonly occur and that general extent of splicing mutations has been underestimated [5].

Most of the mutations that disrupt splicing are single nucleotide substitutions within the intronic or exonic segments of the classical splice sites and result in: i) complete exon skipping; ii) activation of a cryptic splice site; iii) intron retention; or iv) pseudo-exon inclusion. Also, mutations can disrupt splicing by introducing a new splice site within an exon or intron [4,9]. In most cases, use of unnatural splice sites or intron retention introduces premature termination codons (PTCs) into the mRNA, typically resulting in degradation by the nonsense-mediated decay (NMD) system [4]. Additionally, mutations may also affect the use of an alternative splice site leading to a shift of the ratio of natural protein isoforms [4]. Furthermore, some mutations may disrupt splicing by affecting a component of the splicing machinery, or by affecting a factor involved in splicing regulation [4].

The recognition of the importance of splicing defects as main pathogenic mechanism underlying numerous genetic disorders assumes crucial importance in clinical practice [3]. Indeed, it is extremely important to interpret the clinical impact of genetic variations that may cause splicing defects.

In order to investigate the pathogenicity of a splicing mutation, a minigene approach is usually used, relying on its first description in 1984. The minigenes contain a genomic segment from the gene of interest that includes the alternatively spliced region and flanking sequences, and represents a relatively fast approach for identifying splicing spoilers and ~~the~~ studying their underlying functional mechanisms [3,8].

Actually, it was soon realized that constitutive and regulated splicing reactions could be potential therapeutic targets and, accordingly, novel therapeutic strategies, directed toward correcting or circumventing splicing abnormalities, are now emerging [4,10]. Nowadays,

potential therapeutic approaches targeting splicing abnormalities include the use of small molecules, antisense oligonucleotides, bifunctional oligonucleotides, spliceosome-mediated RNA *trans*-splicing (SMaRT), isoform-specific RNA interference and modified U1 snRNPs [3,10].

One of the most successful approaches and a rapidly evolving field is the use of antisense oligonucleotides. These are thought to modulate splicing by steric hindrance of the recruitment of splicing factors to the targeted *cis*-elements, thus forcing the machinery to use the natural sites [3]. They can be used to inhibit the inclusion of unwanted exons and/or promote the production of a truncated but functional protein [3]. This type of therapeutic approach has already been tested in several genes involved in many common diseases, namely *CFTR* for cystic fibrosis, *HBB* for beta thalassemias, *BRCA1* and *ATM* for cancer and, in particular, *DMD* for Duchenne muscular dystrophy to which a human clinical phase III trial has already been launched [3].

Galactosemia results from an impaired ability to metabolize galactose and galactose-intolerant individuals were described in the medical literature as early as 1908 [11].

In the human organism, most of the ingested galactose is rapidly metabolized to glucose-1-phosphate by the action of three consecutive enzymes: galactokinase (GALK) responsible for turning D-galactose into galactose-1-phosphate (Gal1P); galactose-1-phosphate uridylyltransferase (GALT) which transfers a UMP group from UDP-glucose to Gal1P, thereby releasing glucose-1-phosphate and forming UDP-galactose; and UDP-galactose epimerase (GALE) that catalyzes the interconversion of UDP-glucose and UDP-galactose [9,12]. These three enzymes constitute the main pathway of galactose catabolism, the Leloir pathway (Figure 3). Besides this metabolic pathway, a number of alternative pathways have been described [11,12].

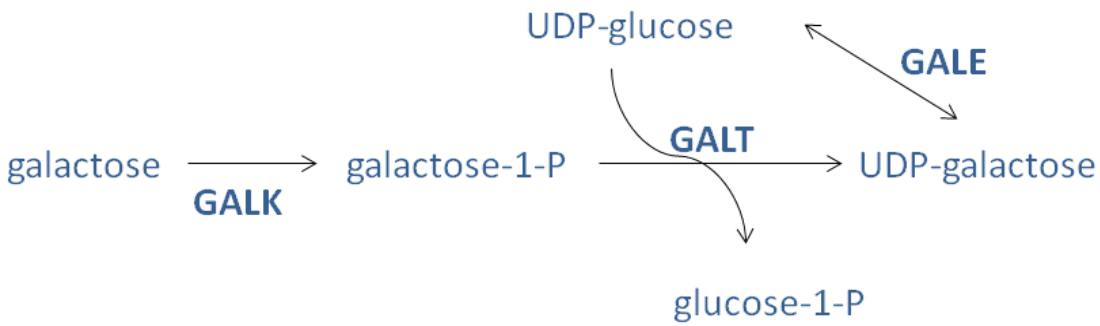

Figure 3 - The Leloir pathway of galactose metabolism.

Deficiency of any one of these enzymes in humans results in a form of galactosemia, being Classic Galactosemia the most common and clinically severe one. It is caused by

mutations in the *GALT* gene which profoundly impair the GALT enzyme, and affects about 1/47,000 live-births [11,13]. Acute symptoms generally appear in the first weeks of life and include poor feeding and weight loss, vomiting, diarrhea, lethargy, and hypotonia; liver dysfunction is present in most cases, but bleeding tendencies, cataracts, and septicemia may also occur [11].

The human *GALT* gene (Figure 4) is located in chromosome 9p13, is arranged into 11 exons spanning 4.37 kb of genomic sequence, and encodes a 379 amino acids protein with a molecular mass of 43 kDa. The active GALT enzyme is a homodimer with a total molecular mass of approximately 88 kDa [14].

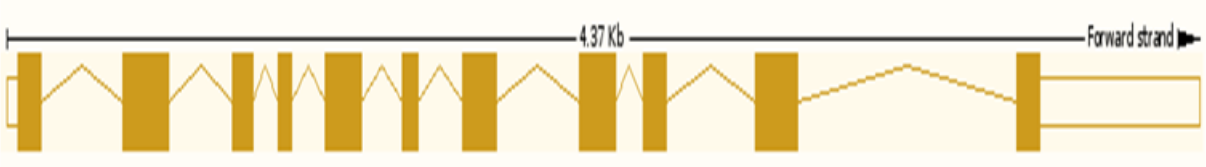

Figure 4 - The human GALT gene [15].

Like many other autosomal recessive metabolic disorders, Classical Galactosemia displays a great allelic heterogeneity with more than 200 variations described until now. Even though most of them are missense ones, silent, nonsense and noncoding changes have also been reported, as well as a large (~5kb) deletion, a number of small deletions and insertions and several splicing mutations (Figure 5) [11].

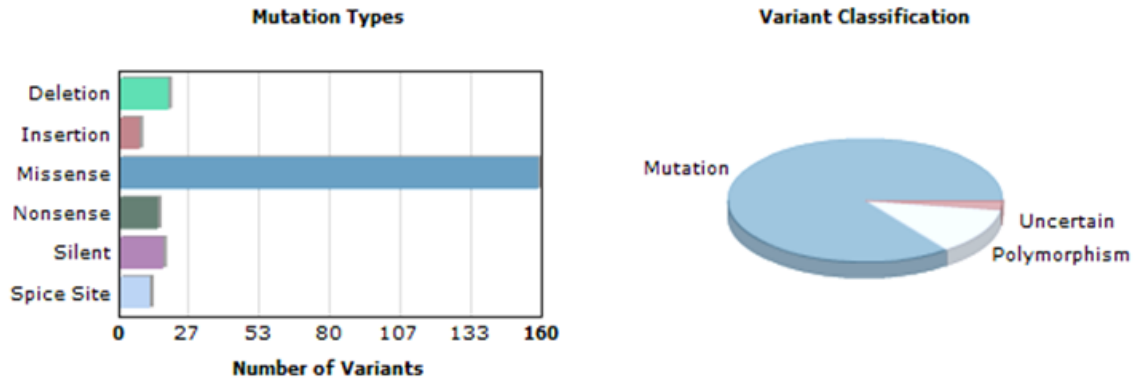

Figure 5 – Classical Galactosemia mutation and variant distributions [16].

The frequencies of specific *GALT* alleles appear to vary by population or racial group, likely reflecting founder effects or other genetic forces [11]. A few mutations are frequent, but most are rare. The most frequently cited are Q188R, K285N, S135L and N314D [17] being Q188R and K285N the two most prevalent mutations in European populations with frequencies depending on the studied population [18]. The S135L mutation is found almost

exclusively in the Afro-Americans being the most frequently reported mutation in this population [12].

Mutational analysis of 27 Portuguese patients confirmed the Q188R as the prevalent molecular defect (≈60%), and surprisingly revealed an intronic variation, c.820+13a>g (IVS8+13a>g), as the second most frequent variation, accounting for 12.5% of the mutant alleles. c.820+13a>g is believed to affect pre-mRNA splicing, and therefore to be a disease-causing mutation since it has never been identified in controls and was only found in Portuguese patients either in homozygous or compound heterozygous state. It is noteworthy that this change has also been found in a Spanish family from Galicia which is in the neighborhood of Portugal [19].

A diagnosis of GALT deficiency may arise from the investigation of acute illness or may be made presymptomatically following an abnormal newborn screening result [11], which utilizes a small amount of blood obtained from heel prick to assay GALT activity and quantify Gal1P and galactose concentrations in total red blood cells (RBC) [20]. When GALT activity is deficient, Gal1P and galactose accumulate and Gal1P competes with UTP-dependent glucose-1-P pyrophosphorylase to reduce UDP-glucose production [20].

Traditional treatment consists in a severe restriction of dietary galactose, which is life-saving, but clearly insufficient since long-term follow-up of patients has shown that they develop symptoms such as retarded mental development, verbal dyspraxia, motor abnormalities and hypergonadotrophic hypogonadism in women [12]. Endogenous production of galactose, which may be up to 1 gram per day in adult patients, has been suggested to be a major cause of these late complications [21]. Accordingly, new therapeutic approaches are urging.

The objectives of the present project are to develop the adequate methodology allowing the elucidation of the pathogenic mechanism underlying the c.820+13a>g mutation, which potentially affects the natural splicing reaction, to confirm its effect in patients carrying the mutations and, finally, to eventually correct it.

We hope that this study will expand the knowledge on the molecular mechanisms underlying Classical Galactosemia pathogenesis and provide another step into the development of an alternative and efficient therapeutic approach.

Materials and methods

In Silico Analyses of Splice Site Strength and ESE and ESS Motifs

The strength of authentic and cryptic splice sites was assessed by web-based tools: NetGene2 (http://www.cbs.dtu.dk/services/NetGene2/) [20,21] for human sequences, and NNSPLICE 0.9 (http://www.fruitfly.org/seq_tools/splice.html) [22] with default settings for human or other organism, no reverse strand included and a minimum score for both 5' and 3' splice sites of 0.4. Both programs predict the presence of donor or acceptor sites by attributing a score to the possible splice sites. The scanned wild-type and mutant *GALT* gene sequences were 399bp long and included the last portion of intron 7, whole exon 8, intron 8, exon 9, and the first portion of intron 9 (Mutant Sequence in Supplementary data). Scores were then analysed and compared.

Predictive programs RESCUE-ESE (http://genes.mit.edu/burgelab/rescue-ese/) [23] and ESEfinder 3.0 (http://rulai.cshl.edu/tools/ESE/) [24,25] were used for ESE motifs, and FASS-ESS Web server (http://genes.mit.edu/fas-ess/) [26], using both set of hexamers (FAS-hex2 and FAS-hex3), for ESS 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. The same 399bp wild-type and mutant *GALT* gene fragments were analysed with default settings for human sequences and default thresholds, and the scores compared.

Cloning of the human *GALT* gene and minigene construction

Genomic fragments encompassing the region of the *GALT* gene containing the c.820+13a>g variation were isolated from a control individual and from a Portuguese patient heterozygous for this mutation. This fragment, including the last portion of intron 7, whole exon 8, intron, exon 9, and the first portion of intron 9, was PCR amplified using the *Taq* DNA Polymerase (Invitrogen Corporation, Carlsbad, CA, USA) and primers 8+9F and 8+9R (Table 1). The cycling conditions were: 5 min at 94ºC, 1 min at 58ºC, 2 min at 72ºC, 30 cycles of 40 sec at 94ºC, 40 sec at 58ºC, 90 sec at 72ºC, followed by a final extension during 5 min at 72ºC.

The fragments were cloned into pCR®-2.1TOPO vector (Invitrogen), according to manufacturer's instructions. The recombinant plasmids were further transformed into One Shot® TOP10 Chemically Competent *E. coli* cells (Invitrogen), according to manufacturer's

instructions and the positive clones were cultured overnight (37ºC, 150 rpm) in 10mL of LB medium containing kanamycin (50µg/ml). Plasmids were isolated using Ron's Plasmid Mini Kit (Bioron Gmbh, Ludwigshafen, Germany) and their concentration and purity was determined in a NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies, Wilmington, DE, USA). To confirm the integrity of the desired fragments, direct sequence analysis was then performed using the ABI Prism BigDye Terminator Cycle Sequencing Ready Reaction Kits, in an ABI PRISM 310 Genetic Analyzer (Applied Biosystems, Foster City, CA, USA).

Recombinant plasmids were then digested with *Eco*RI (NZYTech, Lisbon, Portugal), the restriction products were separated by 2% agarose gel electrophoresis and further purified using the QIAquick Gel Extraction Kit (Qiagen, Hilden, Germany). The pSPL3 exon-trapping vector (kindly supplied by Bélen Pérez, Madrid, Spain) was digested with *Eco*RI and the linearization was confirmed by electrophoretic analysis. Then it was purified using the QIAquick Gel Extraction Kit (Qiagen) and dephosphorylated with Alkaline Phosphatase (NZYTech).

The ligation reaction of the desired fragments with the vector was performed using Speedy Ligase (NZYTech) and the recombinant plasmids were transformed into One Shot® TOP10 Chemically Competent *E. coli* cells. A few clones were cultured overnight (37ºC, 150 rpm) in 10mL of LB medium containing ampicilin (50µg/ml). After isolation, concentration and purity were determined as previously described. Direct sequence analysis was performed in order to confirm the presence of the desired fragments, wild-type and mutant sequences, into the pSPL3 vector.

In vitro splicing analysis

Minigene constructs containing wild-type and mutant sequences, pSPL3.wt and pSPL3.mut respectively, as well as the empty vector, pSPL3, were transfected into appropriate eukaryotic cell lines, namely COS-7 and Hek293. Cells were cultured in 6-well plates with DMEM medium containing 10% NBCS, 1% penicillin/streptomycin, and 2mM glutamine (all products from Invitrogen) and further transfected in duplicates using 1µg of plasmid DNA and GeneJuice™ (Novagen, Merck KGaA, Darmstadt, Germany) according to manufacturer's instructions. After 24 hours, total RNA was extracted using the Trizol reagent (Invitrogen) and the first-strand cDNA was synthesized using the NZY First-Strand cDNA Synthesis Kit (NZYTech). PCR amplification was carried out using the *Taq* DNA Polymerase (Invitrogen) and primers SD6 and SA2 (Table 1), which hybridize with the vector sequence and thus prevent the amplification of the endogenous *GALT* transcripts. The cycling conditions were the following: 30 min at 50ºC, 15 min at 95ºC, 40 cycles of 30 sec at 95ºC,

30 sec at 55ºC, 1 min at 72ºC, followed by 10 min at 72ºC. PCR products were separated by 2% agarose gel electrophoresis, extracted and purified using the Isolate PCR and Gel Kit (Bioline, UK) and directly sequenced to characterize the effect of the mutation.

In vivo splicing analysis

Total cellular RNA was isolated from lymphocytes of a control individual and a patient homozygous for the IVS8+13a>g mutation using the Trizol reagent (Invitrogen). First-strand cDNA was synthesized using the NZY First-Strand cDNA Synthesis Kit (NZYTech) and then PCR amplified using the *Taq* DNA Polymerase (Invitrogen) and CF and CR primers (Table 1) with the following cycling conditions: 5 min at 94ºC, 1 min at 60ºC, 2 min at 72ºC, 30 cycles of 40 sec at 94ºC, 40 sec at 60ºC, 90 sec at 72ºC, followed by 7 min at 72ºC. Amplification products, encompassing the mutation region, were visualized by agarose gel electrophoresis. After separation by 2% NuSieve GTG agarose gel electrophoresis, PCR products were purified as previously described, and further analysed by direct sequencing.

Correction of alternative splicing

Cell culture and minigene transfection followed the same protocol previously referred. Five hours afterward, antisense oligonucleotide IVS8-LNA (5'CCAGGATCCTACCTG3') (Exiqon, Vedbaek, Denmark) was transfected using the INTERFERin reagent (Polyplus transfection, Illkirch, France), according to the manufacturer's instructions, and to final concentrations of 0, 15, 30 and 75nM. 24 hours later, conventional RT-PCR for splicing analysis was performed as described above.

Table 1 - Nucleotide sequences of primers used in this study.

Primer	Sequence
8+9F	5'CACCTTGATGACTTCCTATCCA3'
8+9R	5'GAAATGGTGTTGGGGCTAAA3'
SD6	5'TCTGAGTCACCTGGACAACC3'
SA2	5'GCTCACAAATACCACTGAGAT3'
CF	5'GTGAGGAGCGATCTCA3'
CR	5'TCATTACTACCCTCCGCTCC3'

Results

The intronic mutation c.820+13a>g (IVS8+13a>g) promotes the activation of a cryptic 5' splicing donor site.

In order to predict if the c.820+13a>g mutation directly affects the pre-mRNA splicing of the *GALT* gene or if alternatively is just a marker linked to another causative mutation, and to understand the possible mechanism underlying it, wild-type and mutant sequences of the previously referred 399bp fragment (for details see Materials and Methods) were scanned by two different programs for splice site prediction and the relevant scores were compared (Table 2).

Table 2 - *In silico* analyses for splice site prediction in wild-type and mutant *GALT* fragments.

	Position in the *GALT* fragment		**NNSPLICE**		**NetGene2**	
			wt	**mut**	**wt**	**mut**
5'GT	170_171	Intron 8	0.64	0.64	0.88	0.82
	183_184	Intron 8	-	0.95	0.70	0.93
	768_769	Intron 9	0.86	0.86	0.00	0.00
3'AG	35_36	Intron 7	0.98	0.98	0.00	0.00
	181_182	Intron 8	-	0.51	-	-
	271_272	Intron 8	0.45	0.45	0.55	0.55

Although both programs calculate scores according to different algorithms, analyses gave similar results. Authentic 5'GT of intron 8 presented the same or similar scores in both wild-type and mutant sequences in each program. Even thought scores were higher in NetGene2 (0.88 for wild-type and 0.82 for mutant sequence), in NNSPLICE the score was 0.64 in both sequences. Being so, these results are indicative that this is not a weak splicing donor site.

Both programs also predicted a new GT donor site (c.820+14_15) in the presence of the mutation. In the NNSPLICE program it was only recognized in the mutant sequence, in which it presented a score of 0.95, whereas in the NetGene2 program it was scored 0.70 and 0.93 in the wild-type and mutant sequences, respectively. This cryptic splice site is located immediately after the mutation (c.820+14_15) and, in the presence of the mutation, both programs conferred it a higher strength comparatively to the natural 5' splice site. These results suggested that *in vivo* c.820+14_15 splice site probably 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 5'GT from intron 9 showed the same scores (0.86) in wild-type and mutant sequences in the NNSPLICE program; however, it scored 0.00 in the NetGene2 program.

When analyzing the authentic 3' splice acceptor sites, we noticed that the one in intron 7 showed high scores (0.98) in the NNSPLICE program, while in the NetGene2 program it was scored 0.00. The acceptor site in intron 8 showed very low scores (0.45 and 0.55) in both programs, revealing its intrinsic weakness. The NNSPLICE program revealed that the mutation seams to create a new splice acceptor site in intron 8, immediately upstream the activated cryptic donor site, which however showed no significant strength (0.51).

The intronic mutation c.820+13a>g (IVS8+13a>g) enhances the binding score of the SR protein SRSF1.

Splicing mechanism is also dependent on the presence of splicing enhancers and silencers and, for that reason, we scanned the wild-type and mutant sequences for the presence of ESEs and ESSs using different web-based programs (for details see Materials and Methods).

Concerning the presence of ESE sequences, the ESEfinder 3.0 program scored the same or very similar values for three of the SR proteins analysed (SRSF2, SRSF5, SRSF6) in both wild-type and mutant sequences. However, when analyzing the results for the SRSF1 protein, we could notice that its binding score increased from being 1.986666 in the wild-type sequence, which is close to the lower threshold (1.956), to having a much higher score of 3.906278 in the mutant sequence (CCCA<u>G</u>GT, positions 9-15 of intron 8, mutation position underlined). Nevertheless, the RESCUE-ESE program identified in the same region a binding sequence (CA<u>A</u>GTA, positions 11-16 of intron 8) for the same SRSF1 protein in the wild-type sequence, but not in the mutant one.

As candidate ESSs, the FASS-ESS program predicts a binding sequence corresponding to the same positions in the wild-type (<u>A</u>GTAGG, positions 13-18 of intron 8) and mutant (<u>G</u>GTAGG, positions 13-18 of intron 8) sequences. In addition, another hexamer is predicted when analyzing the mutant sequence (A<u>G</u>GTAG, positions 12-17 of intron 8), overlapping the previous one.

Altogether, these data suggested that c.820+13a>g mutation favors the next cryptic GT dinucleotide (c.820+14_15) to be used as a new splicing donor site, and simultaneously create in that region (CCCA<u>G</u>GT, positions 9-15 of intron 8) a stronger ESE motif, which potentially enhances the binding of the splicing factor SRSF1.

The intronic mutation c.820+13a>g (IVS8+13a>g) is sufficient to cause aberrant splicing of the *GALT* transcript.

Having into account the *in silico* results, and in order to investigate if the c.820+13a>g mutation directly causes activation of the intronic cryptic GT, we cloned the previously referred 399bp fragment containing the wild-type and mutant regions of the *GALT* gene into the pSPL3 exon-trapping vector. Minigene constructs, pSPL3.wt and pSPL3.mut, differ exclusively in the 13th nucleotide of the intron 8, as shown in Figure 6. Eukaryotic COS-7 and Hek293 cell lines were transiently transfected with each construct, splicing products were analysed by RT-PCR in an agarose gel electrophoresis (Figures 7 and 8), and further analysed by direct sequencing. Both cell lines revealed the presence of alternative splicing events with either pSPL3.wt or pSPL3.mut.

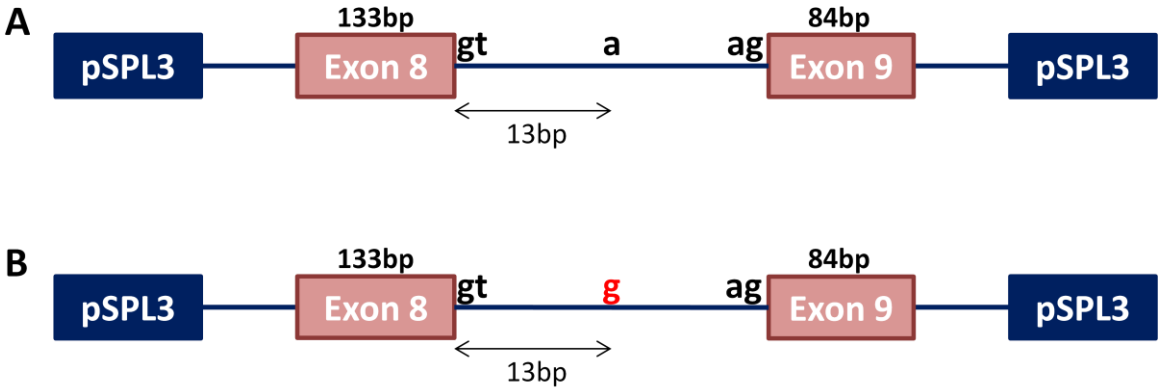

Figure 6 – Minigene constructions.
A 399bp fragment of the *GALT* gene, containing either the wild-type (A) or the mutant (B) sequence, was cloned in the *Eco*RI site of the pSPL3 vector.

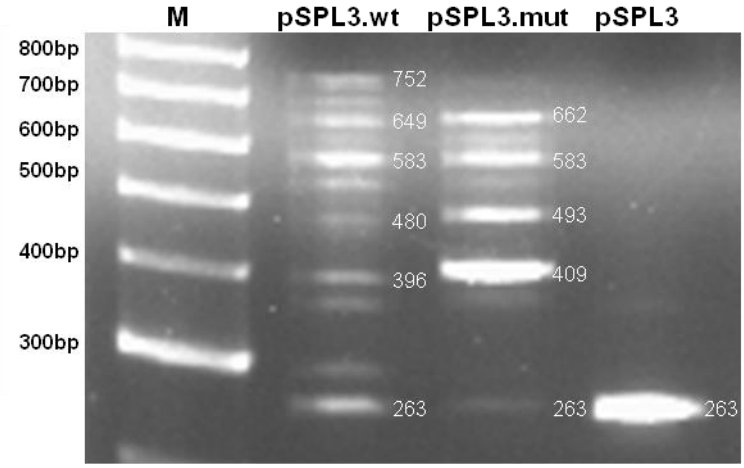

Figure 7 - Transcriptional profile of COS-7 cell line after transfection.
pSPL3.wt, (wild-type minigene), pSPL3.mut (mutant minigene), pSPL3 (empty vector).
M - Molecular marker Hypperladder II (Bioline).

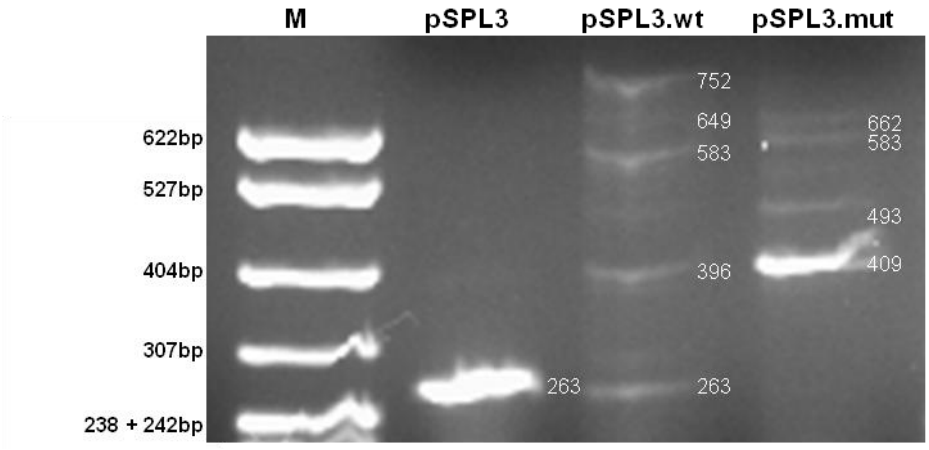

Figure 8 - Transcriptional profile of Hek293 cell line after transfection.
pSPL3.wt, (wild-type minigene), pSPL3.mut (mutant minigene), pSPL3 (empty vector).
M – Molecular marker pBR322x*MspI* (New England Biolabs).

A schematic interpretation of the most prominent transcripts is shown in Figures 9, 10 and 11 and, as we can see, the different fragments correspond to the usage of different splicing donor and acceptor sites.

COS-7 cells transfected with pSPL3.wt, pSPL3.mut and empty pSPL3 all presented a **263bp** fragment, corresponding to vector sequence only; on the other hand, in Hek293 cells, this same fragment was present when transfected with pSPL3.wt and empty pSPL3, but was not detected when transfected with pSPL3.mut. As depicted in Figure 9, this fragment arises from a single splicing event, in which the pSPL3 natural acceptor and donor splice sites are used. This fragment was expected, since the insertion of the fragments in pSPL3 had no evident effect in the vector splice sites strengths. *In silico* analysis with NNSPLICE program revealed a score of 0.94 for 5'GT with and without the cloned fragments, and 3'AG scored 0.64 in the empty vector and 0.67 in both constructs. Moreover, this analysis revealed also a cryptic 5'GT donor site in the pSPL3 sequence which scores 0.81 when fragments are present and 0.94 when only the empty pSPL3 is analysed.

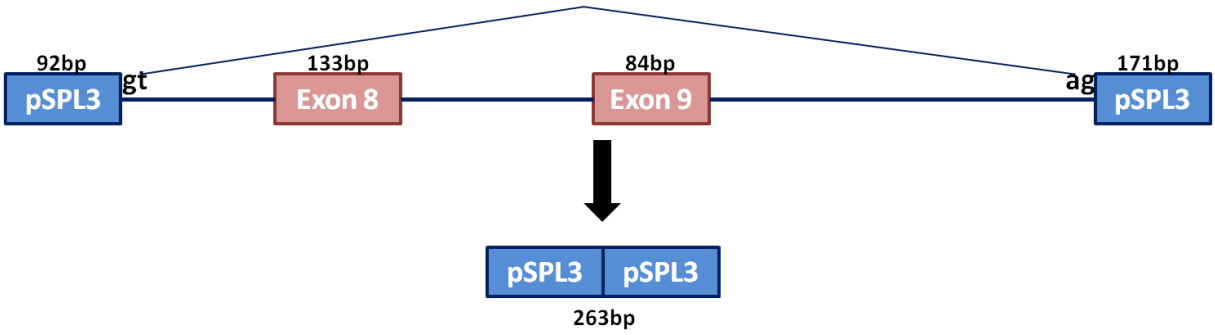

Figure 9 - Schematic interpretation of the 263bp minigene fragment.

The transcriptional profile of COS-7 cells transfected with the **pSPL3.wt** is rather complex, presenting several fragments: a stronger one with 583bp, followed by 649, 396, 752 and 480bp fragments. The **396bp** fragment included pSPL3 and exon 8 sequences (Figure 10.A).

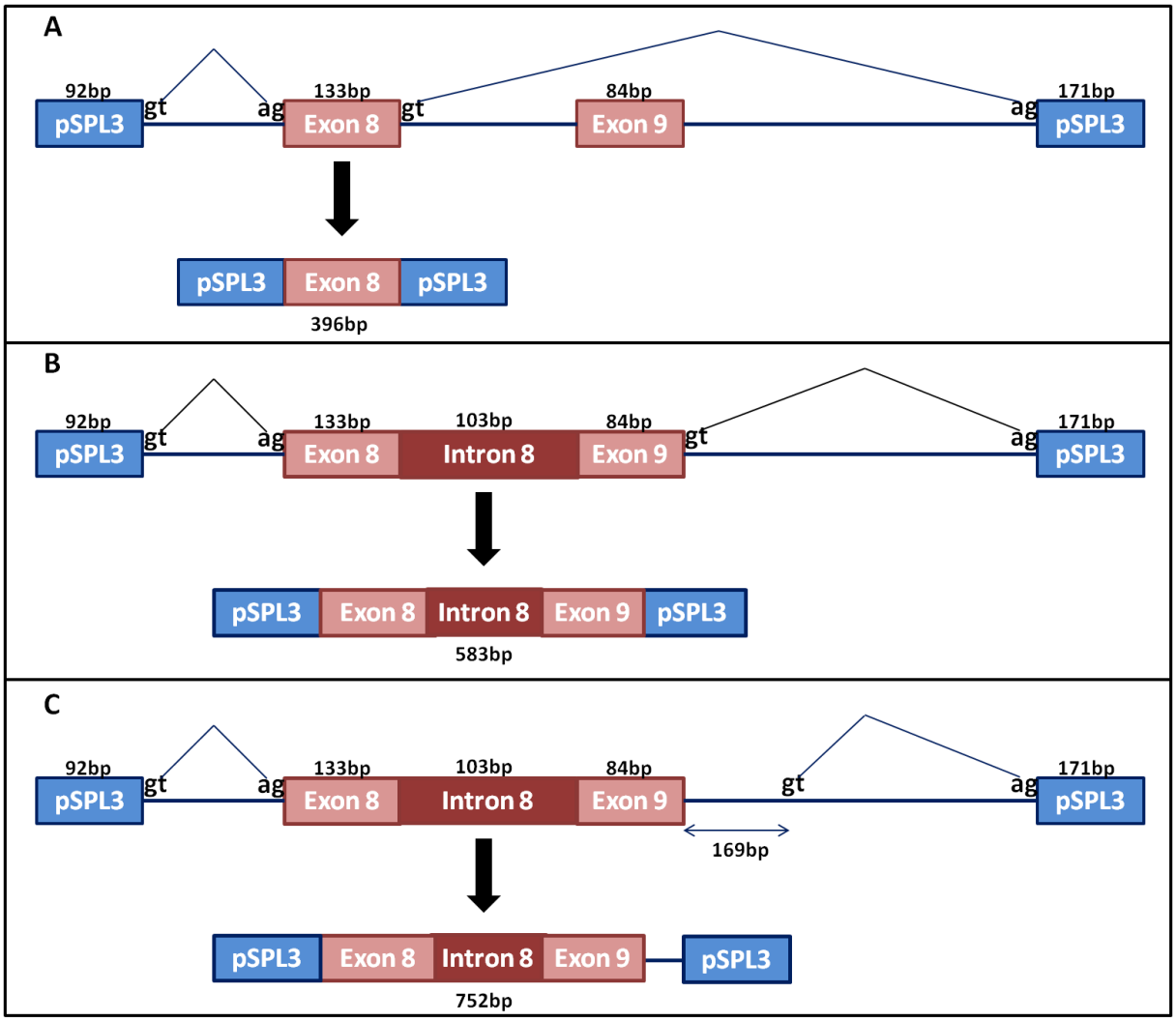

Figure 10 - Schematic interpretation of the wild-type minigene transcriptional fragments.

Indeed, this splicing event used 5'GT of pSPL3 with 3'AG of intron 7, and 5'GT of intron 8 with 3'AG of pSPL3 (Figure 10.A). The **583bp** fragment included both exon 8, intron 8 and exon 9 (Figure 10.B), in which splicing event used 5'GT of pSPL3 with 3'AG of intron 7, and 5'GT of intron 9 with 3'AG of pSPL3. In this case, occurred retention of intron 8. Analogously to the previous fragment, the **752bp** fragment, included exon 8, intron 8 and exon 9; however, splicing event differs from the former, since 5'GT of intron 9 was not used, and instead cryptic 5'GT of pSPL3 was used preferentially, leading to the inclusion of an additional 169bp fragment (Figure 10.C). Two additional bands, putatively corresponding to

fragments with **480bp** and **649bp**, were not sequenced but most probably correspond to well-defined splicing events. In the first case, to the usage of all the canonical splice sites leading to a fragment containing the vector sequences plus exons 8 and 9. In the second case, the usage of all the canonical splice sites, except the 5'GT of intron 9 that was replaced by the cryptic GT of pSPL3 vector, led to a fragment containing vector sequences plus exons 8 and 9 and the additional 169 nucleotides of the vector sequence (Figure 7).

The transcription process in COS-7 cells transfected with **pSPL3.mut** presented a simpler profile with only four bands of similar intensities (Figure 7).

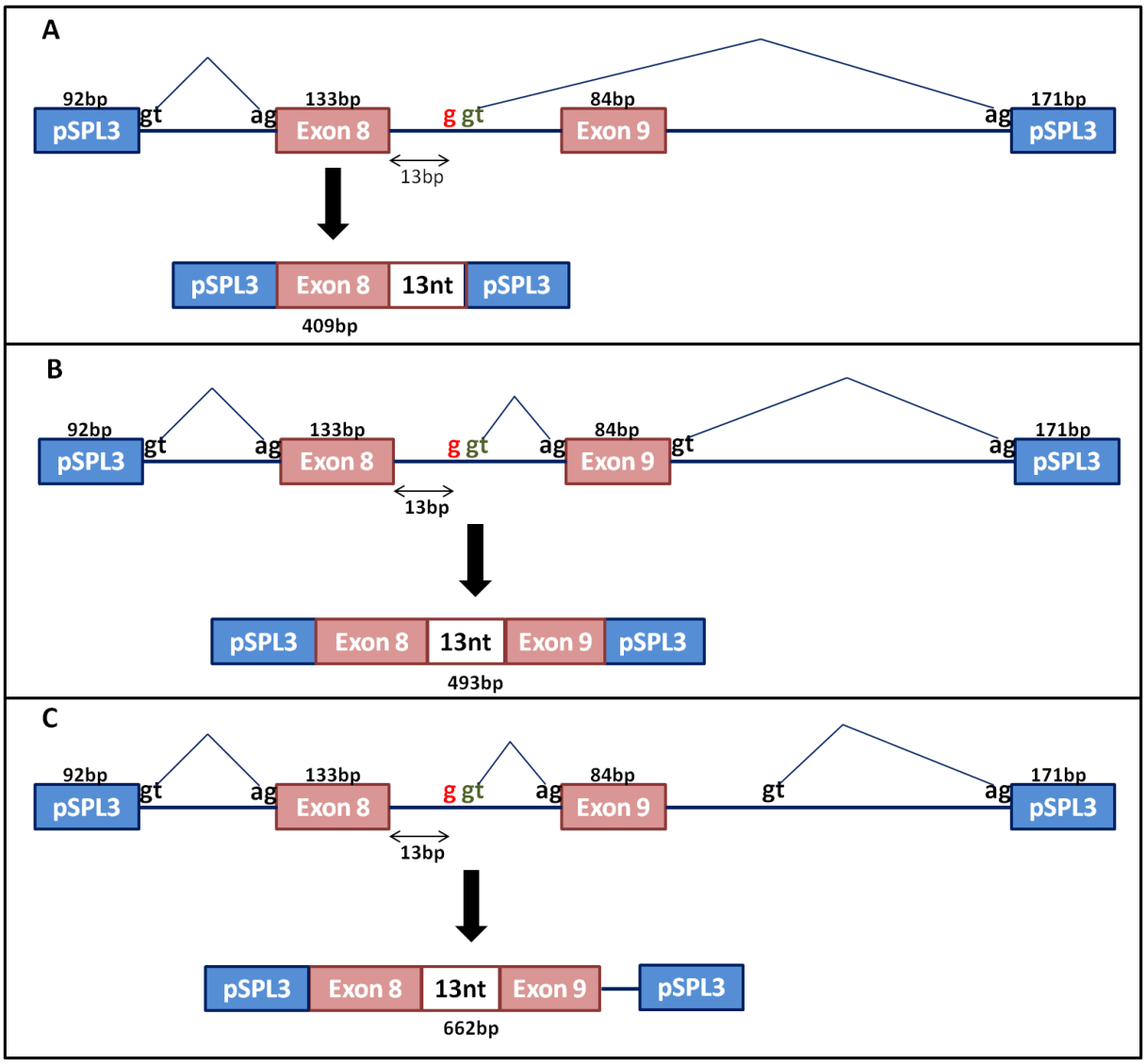

Figure 11 - Schematic interpretation of the mutant minigene transcriptional fragments.

The smallest and more abounding fragment, with **409bp**, corresponded to a splicing event using the 5'GT of pSPL3 with 3'AG of intron 7, and the cryptic 5'GT (c.820+14_15) of

intron 8 with 3'AG of pSPL3 (Figure 11.A). The **493bp** fragment includes both exons 8 and 9 as well as the first 13 nucleotides of intron 8 (Figure 11.B). The **662bp** fragment results from the utilization of the previous splice sites, except the canonical 5'GT of intron 9 which was replaced by the cryptic one in the vector sequence, inducing the inclusion of the additional 169bp (Figure 11.C). Finally, a fourth band putatively with **583bp** was visualized, but not sequenced, and it most probably corresponds to a splice site selection causing the inclusion in the transcript of exon 8, whole intron 8 and exon 9.

The splicing of wild-type and mutant minigenes in Hek293 cells revealed a similar but cleaner transcriptional profile (Figure 8). The wild-type minigene, besides the 263bp fragment containing only vector sequences, only revealed three bands (396, 583 and 752bp) similar to the ones detected in COS-7 cells; the 480 and 649bp fragments were present at a residual level. On the other hand, the transcription profile of the mutant minigene, when compared to the COS-7 cells one, revealed an extremely significant band with 409bp and three additional faint bands with 493, 583 and 662bp.

Results obtained with both constructs in two different eukaryotic cell lines (one from monkey and the other from human) showed that c.820+13a>g mutation *per se* activates a strong cryptic 5' splicing donor site, which leads to the inclusion of the first 13 nucleotides of intron 8 in the resulting mature mRNA, being responsible for inducing aberrant splicing of the *GALT* transcript.

In order to confirm *in vivo* the pathogenic mechanism of c.820+13a>g mutation, total RNA was isolated from a control individual and from a patient homozygous for the mutation. RT-PCR of the transcript region of interest was performed and the amplified cDNA products visualized by agarose gel electrophoresis (Figure 12).

The results revealed that both control and patient presented two fragments. Direct sequence analysis of the larger control band, 378bp long, 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, as expected (Figures 12). However, the smaller 294bp fragment revealed to correspond to an alternative splicing event leading to the skipping of exon 9. Surprisingly, the alternative transcript was more abundant than the wild-type one. When analyzing the patient's fragments, we confirmed that they only differ from control individuals' ones by the additional 13bp corresponding to the first nucleotides of intron 8 (Figure 12 and 13). Nevertheless, the alternative transcript corresponding to exon 9 skipping was much less abundant.

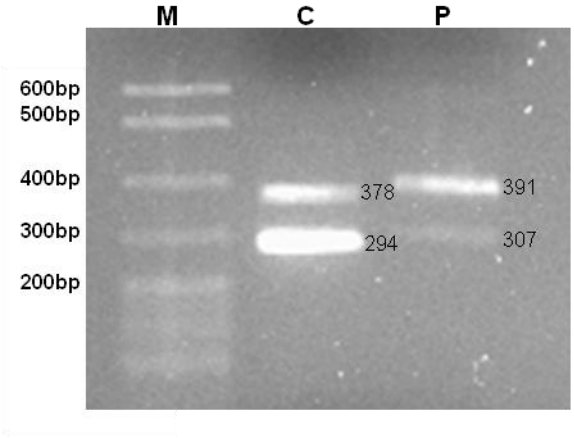

Figure 12 - Agarose gel electrophoresis of the *GALT* cDNA fragments encompassing the desired sequence.
C: control, P: patient, M: molecular weight marker Hypperladder II (Bioline),

These results revealed that the alternative splicing which leads to the skipping of exon 9 is not caused by the c.820+13a>g mutation since it also occurs in homozygous wild-type individuals. Besides, we could confirm the *in vitro* results, namely that the c.820+13a>g (IVS8+13a>g) mutation activates *in vivo* a cryptic 5' splicing donor site in intron 8, leading to a splicing defect, and that it is in fact a disease-causing mutation.

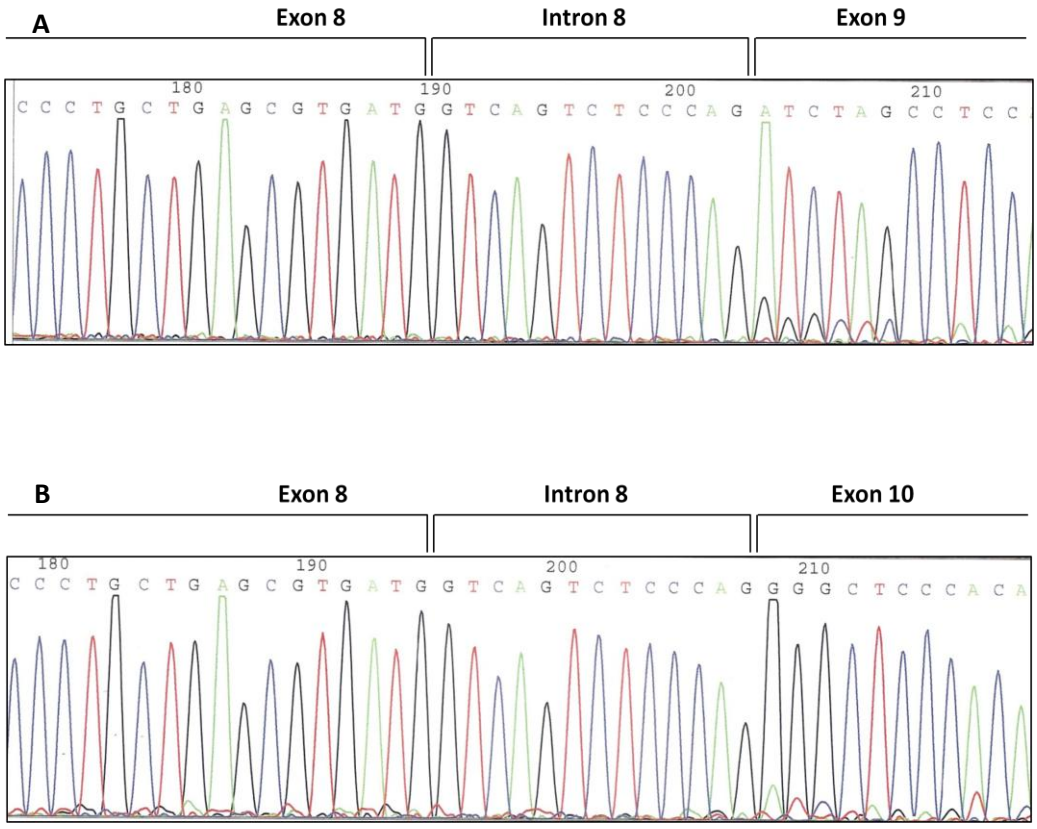

Figure 13 - Partial sequences of the patient's RT-PCR bands corresponding to canonical 3'ss splicing (A) and to alternative splicing involving skipping of exon 9 (B). In both cases the first 13 nucleotides of intron 8 are present.

The LNA approach was not effective in the correction of the splicing defect caused by c.820+13a>g (IVS8+13a>g) mutation.

The correction of the splicing defect was attempted using a locked nucleic acid (LNA) approach. After transfection of COS-7 and Hek293 cell lines with the wild-type (control) and mutant minigenes, and subsequent transfection with the IVS8-LNA, total RNA was isolated and RT-PCR analysis was performed.

The alternative splicing patterns observed in both cell lines were similar when comparing the untreated *versus* treated cells (Figure 12). Sequence analysis of some of the fragments revealed the presence of the first 13 nucleotides of intron 8, not only in the cells transfected with the mutant minigene, but also in the cells cotransfected with the mutant minigene and the IVS8-LNA. The cells transfected with the wild-type minigene were used as negative control and we could confirm that the treatment with LNA had no effect on their splicing process.

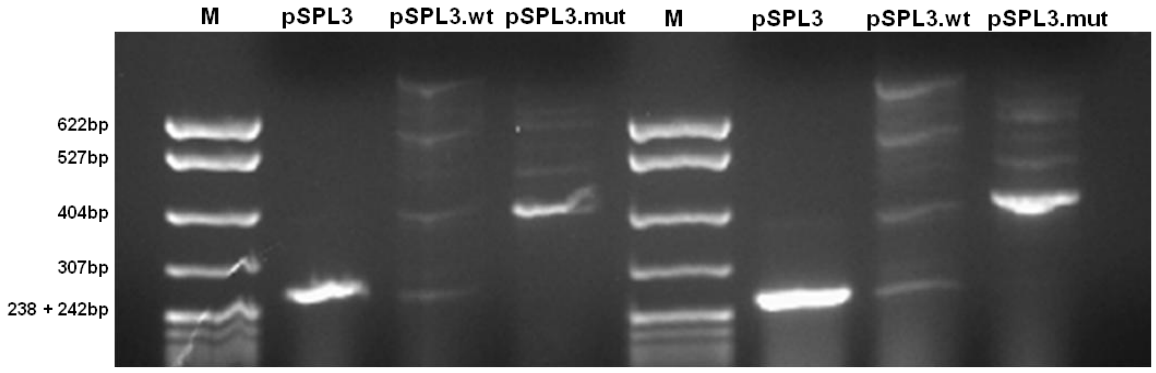

Figure 14 – Transcriptional profile of Hek293 untreated (left) *versus* treated (right) with IVS8-LNA cells. M – Molecular marker pBR322x*MspI* (New England Biolabs).

Discussion

Most population studies revealed that Classical Galactosemia displays a wide mutational spectrum, essentially composed by missense mutations (≈ 70%) [11,14,21]. A significant number of them have already been characterized as disease-causing mutations by heterologous expression of the recombinant proteins, showing that they affect the enzyme structure and/or function [30]. Nevertheless, the number of *GALT* splicing variations is also significant but none of them has already been functionally characterized [16]. On the other hand, the study of the Portuguese galactosemic population evidenced the c.820+13a>g (IVS8+13a>g) mutation as the second most prevalent, accounting for 12.5% of all mutant alleles [19,31,32]. This mutation is classified as benign in the *GALT* mutation database [16]; however, it has been exclusively observed in patients and never in control individuals, which strongly suggests its responsibility on the galactosemic phenotype. Accordingly, these data were the rationale for the present study.

The first approach was to evaluate the mutation pathogenicity by *in silico* analyses. Using web-based programs that predict splice site strength and search for ESE and ESS motifs, we could determine that the c.820+13a>g mutation activates a cryptic splice donor site (Table 2), immediately downstream of the nucleotide variation (c.820+14_15), which also creates a strong ESE motif for the binding of the SRSF1 protein. These first results strongly suggested c.820+13a>g as a pathogenic mutation.

In order to elucidate the underlying pathogenic mechanism, we used an *in vitro* minigene approach. After cloning the *GALT* sequence of interest in an exon-trapping vector (Figure 6), we evaluated the splicing mechanism occurring in two types of eukaryotic cells, Hek293 and COS-7. The transcriptional profile was always complex, revealing the presence of alternative transcripts (Figures 7 and 8). This fact was expected once the conditions are dissimilar from the ones in the human cell, where the entire *GALT* pre-mRNA is spliced. However, the most important result was the observation that the minigene construct carrying the *GALT* mutant sequence, encompassing the c.820+13a>g mutation, originated transcripts containing the first 13 nucleotides of intron 8 (Figures 7, 8, 9, 11).

This result showed that the presence of the c.820+13a>g mutation activates the downstream cryptic 5' splice donor site, whose strength completely overlaps those of the canonic one. The activation of cryptic splice donor sites is a pathogenic mechanism already described in several diseases [33,34,35]. Additionally, this mutation creates a strong ESE motif (CCCAGGT) for the binding of the SRSF1 protein. It is known the involvement of this and other SR proteins in the recruitment of U1 snRNP and its importance for the early

spliceosome assembly [36]. Therefore, we hypothesize that c.820+13a>g mutation induces an alternative SRSF1 binding, which may cause a shift in the spliceossomal assembly, preventing splicing in the natural 5'GT and leading to the c.820+14_15 usage instead.

In parallel to the *in vitro* analyses, we have evaluated the *in vivo* process. Accordingly, we selected a GALT-deficient patient, homozygous for this mutation, from whose lymphocytes we prepared total RNA. His transcriptional profile revealed two bands: the larger one contained the additional first 13 nucleotides of intron 8, as expected (Figures 12 and 13.A). The smallest band translated a splicing event leading to the skipping of exon 9 (Figures 12 and 13.B) and has never been reported before. Nevertheless, this exon skipping was surely induced by the weakness of the 3'AG acceptor site in intron 8, already detected by the predictive programs used in the *in silico* analysis.

It is interesting to note that both transcripts displayed by the patient, containing the additional 13 nucleotides, cause a frameshift in the open reading frame and induce a premature stop codon. Usually, these transcripts are directed to the nonsense mediated decay (NMD) system and are not detected [37,38,39]. Presently, after cloning the mutant cDNA on an appropriate expression vector, we are performing the heterologous expression of the mutant protein in E.coli cells to evaluate the structure and function of the protein potentially produced.

As previously referred in the Introduction, dietary restriction of galactose, the standard of care for Classical Galactosemia treatment, is clearly insufficient and new alternative therapeutics are urging. Accordingly, we tried to correct this splicing defect using antisense oligonucleotides (AO), the most used approach in these cases [40,41,42]. The rationale for its usage relies on the AO hybridization to the mutant region, hindering the assembly of the spliceosome. Once the mutation is too close to the canonical 5'GT donor site, the use of morpholinos seemed not adequate due to their large volume. The selected alternative was the use of locked nucleic acids (LNA), ribose nucleotides which RNA moieties are modified with an extra bridge connecting the 2' oxygen and the 4' carbon. However, the cotransfection of the eukaryotic cells with the mutant construct and the IVS8-LNA produced no splicing correction (Figure 14). The first approach to solve this problem will be the design of new LNA sequences, trying to find the effective binding site of the AO in order to block the access of the first spliceosomal protein, the U1 snRNP to the cryptic splice site and to force its binding to the natural one.

Conclusion

The use of web-based programs for *in silico* prediction of the pathogenic mechanisms underlying nucleotide substitutions, as well as the minigene approach to *in vitro* evaluate the results of the splicing events, revealed to be excellent tools for studying nucleotide variations potentially affecting the splicing process. Finally, the study of a homozygous patient allowed us to validate the *in silico* and *in vitro* results and definitively classify c.820+13a>g (IVS8+13a>g) as a disease-causing mutation.

References

[1] Licatalosi DD, Darnell RB. RNA processing and its regulation: global insights into biologicalnetworks. Nat Rev Genet 2010;11:75-87.

[2] Nissim-Rafinia M, Kerem B. Splicing regulation as a potential genetic modifier. Trends Genet 2002;18:123-127.

[3] Baralle D, Lucassen A, Buratti E. Missed threads. EMBO Rep 2009;10:810-816.

[4] Faustino NA, Cooper TA. Pre-mRNA splicing and human disease. Genes Dev 2003;17:419-437.

[5] Cartegni L, Chew SL, Krainer AR. Listening to silence and understanding nonsense: exonic mutations that affect splicing. Nat Rev Genet 2002;3:285-298.

[6] Baralle D, Baralle M. Splicing in action: assessing disease causing sequence changes. J Med Genet 2005; 42:737–748.

[7] Sironi M, Menozzi G, Riva L. Silencer elements as possible inhibitors of pseudoexon splicing. Nucleic Acids Res. 2004;32(5):1783-1791.

[8] Cooper TA. Use of minigene systems to dissect alternative splicing elements. Methods 2005;37:331-340.

[9] Buratti E, Baralle M, Baralle FE. Defective splicing, disease and therapy: searching for master checkpoints in exon definition. Nucleic Acids Res 2006; 34: 3494-3510.

[10] Feero WG, Guttmacher AE. New therapeutic approaches to mendelian disorders. N Engl J Med 2010;363:852-863.

[11] Fridovich-keil JL, Walter JH. (2008). PART 7: CARBOHYDRATES Chapter 72: Galactosemia. 1-92. *The Online Metabolic & Molecular Bases of Inherited Disease*.

[12] Bosch AM. Classical galactosemia revised. J Inher Metab Dis 2006; 29:516-525.

[13] Fridovich-Keil JL. Galactosemia: the good, the bad, the unknown. J Cell Phsysiol 2006; 209: 701-705.

[14] Calderon FRO, Phansalkar AR, Crockett DK, Miller M, Mao R. Mutation database for the galactose-1-phosphate uridyltransferase (GALT) gene. Hum Mutat 2007; 28:939-943.

[15] Ensemble: http://www.ensembl.org/Homo_sapiens/Transcript/Summary?db=core;g=ENSG00000213930;r=9:34638130-34651032;t=ENST00000378842.

[16] Fernanda R.O. Calderon, Amit R. Phansalkar, David K. Crockett, Martin Miller, Rong Mao. Mutation database for the galactose-1-phosphate uridyltransferase (GALT) gene. *Hum Mutat*. 2007; 28(10): 939-43. http://www.arup.utah.edu/database/GALT/GALT_welcome.php.

[17] Tyfield L, Reichardt J, Fridovich-Keil J, 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;13:417-430.

[18] Gort L, Quintana E, Moliner S, González-Quereda L, López-Hernández T, Briones P. An update on the molecular analysis of classical galactosemia patients diagnosed in Spain and Portugal: 7 new mutations in 17 new families. Med Clin 2009;132(18):709-711.

[19] Gort L, Boleda MD, Tyfield L, Vilarinho L, Rivera I, Cardoso ML, Santos-Leite M, Girós M, Briones P. Mutational spectrum of classical galactosemia in Spain and Portugal. J Inherit Metab Dis 2006;29:739-742.

[20] Elsas LJ II. Galactosemia. 2000 Feb 4 [Updated 2010 Oct 26]. In: Pagon RA, Bird TD, Dolan CR, et al., editors. GeneReviews™ [Internet]. Seattle (WA): University of Washington, Seattle; 1993. Available from: http://www.ncbi.nlm.nih.gov/books/NBK1518/

[21] Bosch AM, IJlst L, Oostheim W, Mulders J, Bakker HD, Wijburg FA, Wanders RJA, Waterham HR. Identification of novel mutations in classical galactosemia. Hum Mutat 2005;25:502-508.

[22] Hebsgaard SM, Korning PG, Tolstrup N, Engelbrecht J, Rouze P, Brunak S. Splice site prediction in Arabidopsis thaliana DNA by combining local and global sequence information. Nucleic Acids Res 1996;24:3439-52.

[23] Brunak S, Engelbrecht J, and Knudsen S. Prediction of Human mRNA Donor and Acceptor Sites from the DNA Sequence. J Mol Biol 1991;220:49-65.

[24] Reese MG, Eeckman FH, Kulp D, Haussler D. Improved Splice Site Detection in Genie. J Comp Biol 1997;4(3):311-23.

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

[26] 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):2490-2508.

[27] Cartegni L, Wang J, Zhu Z, Zhang M Q, Krainer AR. ESEfinder: a web resource to identify exonic splicing enhancers. Nucleic Acid Res 2003;31(13):3568-3571.

[28] Wang Z, Rolish ME, Yeo G, Tung V, Mawson M, Burge CB. Systematic identification and analysis of exonic splicing silencers. Cell 2004;119:831-845.

[29] 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 Acid Res 2009;37(9):e67.

[30] Christacos NC, Fridovich-Keil JL. Impact of patient mutations on heterodimer formation and function in human galactose-1-P uridylyltransferase. Mol Genet Metab 2002;76(4):319–326.

[31] Lourenço SP, Coelho AI, Silva MJ, Tavares de Almeida I, Vicente JB, Rivera I. Correction of a splicing mutation in the GALT gene using antisense therapy. FEBS International Workshop on "New Developments in RNA Biology: State of the art and future perspectives", September 1th – 4th, 2012, Tavira, Portugal (abstract).

[32] Lourenço SP, Coelho AI, Silva MJ, Tavares de Almeida I, Vicente JB, Rivera I. Classical Galactosemia: Characterization of the first intronic variation and its correction by antissense therapy. SSIEM, J Inherit Metab Dis 2012;35(Suppl 1):S1-S182.

[33] Jorge-Finnigan A, Aguado C, Sánchez-Alcudia R, Abia D, Richard E, Merinero B, Gámez A, Banerjee R, Desviat LR, Ugarte M, Belen Pérez, B. Functional and Structural Analysis of Five Mutations Identified in Methylmalonic Aciduria *cblB* Type. Hum Mutat 2010;31(9):1033–1042.

[34] Stump MR, Gong Q, Zhou Z. Multiple splicing defects caused by hERG splice site mutation 2592_1G_A associated with long QT syndrome. Am J Physiol Heart Circ Physiol 2011;300: H312–H318.

[35] Attanasio C, Moerloose P, Antonarakis SE, Morris MA, Neerman-Arbez M. Activation of multiple cryptic donor splice sites by the common congenital afibrinogenemia mutation, *FGA* IVS411G3T. Blood 2001;97:1879-1881.

[36] Long JC, Caceres JF. The SR protein family of splicing factors: master regulators of gene expression. Biochem J 2009;417:15–27.

[37] Baker KE, Parker R. Nonsense-mediated mRNA decay: terminating erroneous gene expression. Curr Opin Cell Biol 2004;16:293–299.

[38] Lejeune F, Maquat LE. Mechanistic links between nonsense-mediated mRNA decay and pre-mRNA splicing in mammalian cells. Curr Opin Cell Biol 2005;17:309–315.

[39] McGlincy NJ, Smith CWJ. Alternative splicing resulting in nonsense-mediated mRNA decay: what is the meaning of nonsense?. Trends Biochem Sci 2008;33(8):385-393.

[40] Brasil S, Viecelli HM, Meili D, Rassi A, Desviat LR, Pérez B, Ugarte M,Thony B. Pseudoexon Exclusion by Antisense Therapy in 6-Pyruvoyl-Tetrahydropterin Synthase Deficiency. Hum Mutat 2011;32:1019–1027.

[41] Gupta N, Fisker N, Asselin M, Lindholm M, Rosenbohm C, Ørum H, Elmén J, Seidah NH, Straarup EM. A Locked Nucleic Acid Antisense oligonucleotide (LNA) Silences PCSK9 and Enhances LDLR Expression In Vitro and In Vivo. Plos One 2010,5(5):e10682.

[42] Sanake PS, Toompuu M, McClorey G, Bindoff, LA. Antisense oligonucleotide corrects splice abnormality in hereditary myopathy with lactic acidosis. Gene 2012;494:231–236.

Supplementary data

CACCTTGATGACTTCCTATCCATTCTGTCTTCCTAGGAACGTCTGGTCCTAACCAGTGAGCACTGGTTAGTACTG
GTCCCCTTCTGGGCAACATGGCCCTACCAGACACTGCTGCTGCCCCGTCGGCATGTGCGGCGGCTACCTGAGCTG
ACCCCTGCTGAGCGTGATGGTCAGTCTCCCAGGTAGGATCCTGGGGCTAGGCACTGGATGGAGGTTGCTCCCAGT
AGGGTCAGCATCTGGACCCCAGGCTGAGAGTCAGGCTCTGATTCCAGATCTAGCCTCCATCATGAAGAAGCTCTT
GACCAAGTATGACAACCTCTTTGAGACGTCCTTTCCCTACTCCATGGGCTGGCATGGTGAGGCTTTTCAAGTACC
TATATTTAGCCCCAACACCATTTC

Figure S1 – Sequence of the mutant 399bp fragment of *GALT* gene including the last portion of intron 7, whole exon 8, intron 8 and exon 9, and the first portion of intron 9. In red are the exons 8 and 9. The sequences in yellow and green correspond to primers 8+9F and 8+9R. In blue is the sequence complementar to IVS8-LNA. The c.820+13a>g mutation is in dark green.