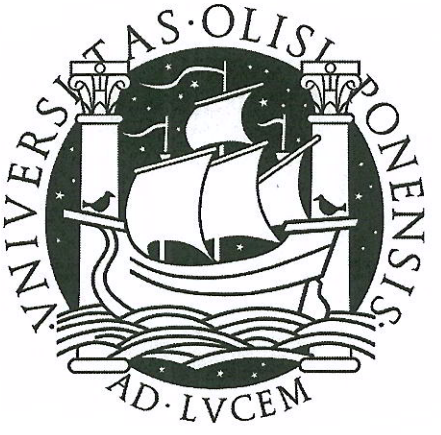
Universidade de Lisboa

Faculdade de Ciências

Departamento de Biologia Vegetal



Characterization of post-transcriptional control mechanisms regulating SMN2 gene expression

Ana Luísa Rodrigues Gomes

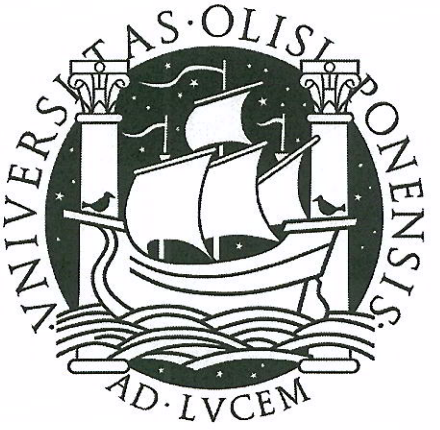
Mestrado em Biologia Molecular e Genética

2010

Universidade de Lisboa

Faculdade de Ciências

Departamento de Biologia Vegetal



Characterization of post-transcriptional control mechanisms regulating SMN2 gene expression

Dissertação orientada por:

Professora Doutora Margarida Gama Carvalho

(Instituto de Biologia Molecular, Faculdade de Medicina da Universidade de Lisboa)

Professora Doutora Margarida Telhada

(Departamento de Química e Bioquímica, Faculdade de Ciências da Universidade de Lisboa)

Ana Luísa Rodrigues Gomes

Mestrado em Biologia Molecular e Genética

2010

Agradecimentos

À Professora Margarida Gama Carvalho por me ter acolhido, pela sua orientação, exigência e pela sempre disponibilidade e confiança. Muito obrigado por ter tornado possível este trabalho e por todas as oportunidades e partilha de entusiasmo.

À Professora Margarida Telhada que sempre demonstrou uma elevada disponibilidade e ajuda.

Ao Departamento de Biologia Vegetal que me tem acompanhado neste últimos 5 anos, em especial aos professores que contribuíram para a minha formação.

Ao Laboratório de Bioinformática e Genética Evolutiva do Center for Biodiversity, Functional & Integrative Genomics pelo acolhimento e por ter proporcionado as condições necessárias à realização deste trabalho.

Ao pessoal do laboratório: Ana Isabel, Francisco, Inna, Luisa, Mariana, Marta, Professor Carlos, Simão, agradeço-vos a boa recepção, apoio e companheirismo e por terem feito do lab um bom local de trabalho. Obrigada pelas risadas, conversas e ajudas com as dúvidas existenciais!

À Professora Maria Prazeres um muito obrigado por ter sido a grande "despertadora" do meu bichinho pela ciência, porque foram as suas aulas, a sua paixão pela ciência que me incentivaram a estar aqui hoje.

À "Tia" Carlota um enorme muito obrigado por ser um grande ponto de referência na minha vida em todos os sentidos..."pequenina nascia e as "Tias" me entregaram, com elas eu cresci e muito me amaram"...toda a vivência no Patinho Feio acompanha-me até aos dias de hoje.

Ás minhas queridas biólogas compinchas: Marina, Sara, Sofia...adoro as nossas longas refeições e conversas como cerejas...agradeço o vosso carinho e amizade!

Aos meus queridos e "enormes" amigos: Andreia minha querida amiga...agradeço-te todo o teu carinho, todos os nossos cafezinhos e conversas! Célia agradeço-te por tudo, minha Mana mais velha...e eu já sei...devo-te umas feriazitas num spa! Edi és para mim um exemplo de força, coragem e luta...adoro-te e admiro-te muito! Minha Inês adoro que a nossa amizade tenha crescido...gosto imenso de ti e agradeço-te todo o apoio e força que tens dado...em tudo...na vida científica, mas principalmente no resto! Ao meu querido João obrigada por tudo...acho que sabes o quanto sempre me ajudaste e adoro ser a tua "crominha"...és um óptimo amigo! Sóninhas...a nossa longa amizade diz o quanto eu te estimo e muito obrigado por me acompanhares há tanto tempo...adoro ver-nos crescer e já sabes... seremos bem velhinhas e a fazer tricô!

Orgulho-me tanto de pertencer à vossa vida e de vos ter presentes na minha! A vossa amizade...vocês sabem o quanto significam! Muito obrigada por aturarem as minhas tolices, as minhas "cromices"...obrigada por esta partilha...principalmente neste último ano.

À minha Família, a todos...graúdos e caçulas, agradeço o carinho, preocupação e orgulho que partilhamos!

Aos meus Pais...acho que nunca conseguirei exprimir como deve ser o quanto significam...sou quem sou por vossa causa, mas acreditem que o orgulho que sentimos é mútuo. Vocês são os pilares da minha vida e admiro a vossa força, coragem e carinho na construção da "nossa família". É a vossa confiança, o vosso amor que me guia todos os dias, que me faz acreditar que sou capaz...muito obrigada por tudo e por nada. Muito obrigada por todo o apoio e amor incondicional principalmente neste último ano.

A vós dedico este trabalho.

Table of contents

Abstract

Research has demonstrated that a significant part of gene expression control is taking place at the post-transcriptional level, determining the stability and translation efficiency of mRNA molecules. These regulatory switches are under the control of RNA-binding proteins and microRNAs that associate predominantly with the 5' and 3' UTRs.

The prospect for development of effective therapies for Spinal Muscular Atrophy is enhanced by the existence of an almost identical copy of the SMN1 gene-mutated in SMA. The SMN2 gene can produce the functional protein at low levels; however variations in its copy number and thus, protein expression output have been shown to modulate disease severity in affected individuals.

Although it is clear that the control of mRNA translation and stability can influence dramatically the protein output of a given gene, and in spite of the advantages that increased SMN2 expression may provide to patients, no systematic study has thus far been performed on the post-transcriptional control mechanisms acting on the SMN2 mRNA. Therefore, the identification of these mechanisms is of great interest, as it may provide the basis for new therapies for SMA.

To identify and characterize the SMN2 post-transcriptional mechanisms regulating the SMN2 mRNA we established a Luciferase reporter system for monitoring UTR-dependent regulatory events. We verified that the SMN2 3'UTR region has a regulatory role in the mRNA stability, which results from a balance of positive and negative *cis*-acting elements. We also identified key conserved nucleotides in the 3'UTR, whose mutation resulted in increased mRNA expression levels. Analysis of the SMN2 5'UTR region indicates that this region has a negative regulatory effect on the SMN2 expression due to presence of a uORF.

We additionally performed systematic validation of novel predicted alternative SMN2 5' and 3'UTR sequences. We were able to validate the expression of an alternative 5'UTR isoform and obtained preliminary evidence for the regulation of SMN transcripts by the Nonsense Mediated Decay pathway.

Keywords: Gene expression; Post-Transcriptional Regulation; Untranslated Regions (UTR); Spinal Muscular Atrophy (SMA); SMN2 mRNA

Resumo

A quantidade de proteína produzida a partir de determinado gene resulta de um complexo balanço entre mecanismos de síntese, de modificação e de degradação. Estes mecanismos actuam em diversos níveis de regulação da expressão génica. Ao longo dos últimos anos de investigação tem-se tornado cada vez mais claro que todos os níveis envolvidos nesta regulação, desde a transcrição, processamento da molécula pré-mRNA, degradação do RNA mensageiro (mRNA), tradução, até à estabilidade da proteína, estão sujeitos a um apertado controlo, todos eles podendo ter um impacto comparável nos níveis finais de proteína.

Apesar de o controlo ao nível da transcrição ser visto como o principal nível de regulação, tem-se verificado, no entanto, que a regulação pós-transcricional tem uma profunda influência nos níveis de síntese de proteína, em particular através de mecanismos que regulam a estabilidade da molécula de mRNA e a eficiência de tradução da mesma. Envolvidos nestes mecanismos estão vários factores que são designados como efectores que, consoante os sinais celulares, actuam positiva ou negativamente na expressão de um gene. Há já algum tempo que se identificou a família das proteínas de ligação ao RNA (RNA Binding Proteins ou RBPs) como principal efectora do controlo pós-transcricional. Estas proteínas reguladoras normalmente ligam-se a sequências degeneradas, geralmente localizadas nas regiões não traduzidas (UTR) a 5' e 3' da molécula de mRNA, que funcionam como plataformas para a interacção da molécula de mRNA com a maquinaria de degradação ou de tradução. Recentemente foram identificadas pequenas moléculas de RNA não codificantes, microRNAs, que também actuam como efectores dos mecanismos de regulação ao nível do controlo pós-transcricional.

Vários anos de pesquisa permitiram reunir evidências que demonstram que uma parte significativa do controlo da expressão génica ocorre a nível pós-transcricional. A compreensão deste nível de regulação será uma ferramenta poderosa no esclarecimento da complexidade do controlo da expressão de determinado gene. Este esforço de investigação será não só importante para o conhecimento geral, mas particularmente crucial na investigação de doenças, onde a caracterização dos mecanismos regulatórios pode permitir a criação de métodos para a modulação terapêutica expressão da molécula envolvida na doença.

A Atrofia Muscular Espinhal (SMA), uma das principais doenças hereditárias causadora de morte infantil, é uma doença neuromuscular causada pela perda de função do gene Survival of Motor Neuron 1 (SMN1). Tal como acontece noutras doenças hereditárias, grandes esforços têm sido desenvolvidos para a descoberta de uma terapia eficaz de forma a aliviar os sintomas nos pacientes.

Em comparação com outras doenças causadas por perda de função de um gene, a desenvolvimento de terapias moleculares dirigidas tem sido, em parte, facilitada na doença SMA uma vez que um "alvo" principal foi já identificado. Estudos moleculares do gene envolvido na doença permitiram a identificação de uma segunda cópia, em quase tudo semelhante ao gene SMN1. Este gene SMN2 foi originado por uma duplicação invertida da porção terminal do cromossoma 5.

Os genes SMN1 e SMN2 diferem em apenas 5 nucleótidos, sendo que a região codificante não é afectada. A grande diferença entre os dois genes é uma transição silenciosa no exão 7, de um C por um T, que induz a exclusão deste exão em 75% dos transcritos produzidos a partir do gene SMN2. Este evento resulta na produção de uma versão truncada da proteína SMN que não é funcional.

Estudos efectuados apontam para que a SMA seja causada pela baixa quantidade de proteína completa e funcional, afectando principalmente as células neuronais motoras. Contudo já se verificou que há uma variabilidade natural na severidade dos sintomas, devido há presença de múltiplas cópias do gene SMN2, tendo-se verificado que quanto maior o número de cópias deste gene, maior é a compensação, nas células neuronais, da perda de função do gene principal. Estes estudos sugerem que um possível tratamento para alívio dos sintomas nos pacientes poderá passar pela indução de aumento da expressão do gene SMN2.

Embora seja cada vez mais claro que o controlo da estabilidade e eficiência de tradução da molécula de mRNA tem um grande impacto na quantidade de proteína produzida e apesar das vantagens em aumentar a expressão do gene SMN2 nos pacientes, até ao momento, nenhum estudo sistemático foi realizado no sentido de caracterizar os mecanismos pós-transcricionais que actuam sobre o mRNA do SMN2. A identificação de motivos regulatórios que possam ser utilizados para a modulação da expressão do gene SMN2, ao nível de reguladores pós-transcricionais, poderá ser vista como uma ferramenta útil para o desenvolvimento de futuras abordagens terapêuticas. Por exemplo, a caracterização de elementos na sequência que induzem a desestabilização do mRNA do SMN2 pode permitir o desenho de pequenos

oligonucleotidos específicos que bloqueiem a ligação de reguladores negativos e assim modular significativamente a quantidade de proteína SMN produzida.

Para a caracterização dos efeitos regulatórios das regiões não traduzidas do mRNA (UTRs) do SMN2 procedeu-se à identificação de elementos na sequência que possam estar envolvidos na regulação pós-transcricional. Com este objectivo, procedeu-se ao estabelecimento de um modelo experimental para estudo de efeitos regulatórios mediados pelos UTRs, envolvendo a criação de um gene repórter em que se fundiu a sequência codificante da luciferase com as regiões 5' e 3'UTR do SMN2.

Quando testadas as diferentes construções do gene repórter com as regiões UTR verificou-se que estas têm um papel na regulação pós-transcricional do mRNA do SMN2, uma vez que a quantidade de proteína produzida por cada uma é diferente em relação à obtida a partir de um vector que apenas contém a sequência da luciferase.

Em particular para o 3'UTR do gene SMN2, os resultados obtidos permitiram identificar três grandes regiões reguladoras, sendo que as extremidades têm um efeito negativo sobre a expressão do mRNA. Também se verificou que o controlo pós-transcricional promovido por esta região ocorre ao nível da regulação da estabilidade do mRNA, uma vez que para cada construção testada os níveis de mRNA são semelhantes aos de proteína. Através da análise bioinformática, identificou-se na região 3'UTR a presença de nucleótidos conservados, normalmente associados a motivos reguladores funcionais. A mutação pontual de alguns desses nucleotidos resultou em aumento dos níveis de expressão de luciferase acima dos 20%, o que pode ser explicado pela destruição de um motivo de ligação de um factor de regulação negativa ou de uma estrutura secundária desestabilizadora.

Com a clonagem da região 5'UTR do gene SMN2 no sistema repórter verificou-se uma grande redução dos níveis de expressão de luciferase. Através da análise bioinformática identificou-se a presença de uma uORF, um elemento *cis* descrito como sendo uma barreira de regulação negativa na expressão de determinados genes. Quando perturbado o reconhecimento desta uORF pela maquinaria de tradução, através da mutação pontual do uATG, verificou-se um aumento da expressão de luciferase de 50% em relação ao UTR normal.

Para além da análise dos UTRs do mRNA codificado pelo gene SMN2, tentámos proceder à validação experimental de isoformas alternativas destas sequências descritas em diversas bases de dados, uma vez que se sabe que contribuir para criar variabilidade nos mecanismos de regulação pós-transcricionais. Este estudo permitiu confirmar a

existência de uma isoforma alternativa do 5' UTR do gene SMN2, cuja função celular é desconhecida, e obter evidências preliminares para a existência de outras isoformas deste mRNA que são alvo de regulação pela via do "*non-sense mediated decay*",

Os resultados apresentados nesta tese permitiram demonstrar a existência de mecanismos de controlo pós-transcricional que regulam a expressão do mRNA do gene SMN2 através dos seus UTRs e ainda identificar algumas das sequências reguladoras envolvidas. Estes elementos identificados podem ser considerados como potenciais alvos terapêuticos para promover um aumento da produção da proteína SMN funcional nos pacientes SMA e assim aliviar a severidade da doença.

Palavras-chave: Expressão génica, regulação pós-transcricional, Regiões não traduzidas (Untranslated regions - UTR), Atrofia Muscular Espinhal (Spinal Muscular Atrophy - SMA), RNA mensageiro SMN2 (mRNA)

Abbreviations

%v/v	Percentage expressed in volume/volume
RBP	RNA Binding Protein
cDNA	Complementary DNA (to mRNA)
CDS	Coding Sequence
DNA	Deoxyribonucleic Acid
FBS	Fetal Bovine Serum
HEK 293 cell line	Human Embryonic Kidney 293 cell line
IRES	Internal Ribosome Entry Site
KDa	Kilo Dalton
min	Minute
mRNA	Messenger RNA
mRNP	Ribonucleo protein particles
NMD	Nonsense Mediated mRNA Decay
nt	Nucleotides
PCR	Polymerase Chain Reaction
RISC	RNA-induced silencing complex
RNA	Ribonucleic Acid
RT-PCR	Reverse transcriptase Polymerase Chain Reaction
sec	Second
SMA	Spinal Muscular Atrophy
SMN 1 and 2	Survival Motor Neuron 1 and 2
uATG	Upstream ATG (Initiation codon)
uORF	Upstream Open Reading Frame
UTR	Unstranslated Regions
wt	Wild-type

"Scientists are trying to unravel the complexity they pose [organisms], hoping to understand how a surprisingly small constellation of molecules can produce such variety of structure and function" *in Lodish H. et al, "Molecular Cell Biology", Third edition*

1. Introduction

The interest of Science in understanding Life and its complexity has proven to be a hard yet fascinating task.

One interesting and important part of this complexity is the study of Gene Expression. For a long time it has already understood that gene expression is regulated at many levels, requiring tight integration and coordination, to bring the genome to life [1]. When the dynamic interplay between DNA, RNA, the whole regulatory apparatus and the integration of intrinsic or environmental signals is mis-regulated, a disease can emerge.

Although the existence of several levels of gene expression regulation in eukaryots have been recognized, historically considerable relevance has been attributed to transcriptional control and little emphasis has been placed on the importance of other levels of regulation. The first level of regulation is considered to be crucial and considerable research effort has been placed in identifying DNA *cis*-elements, as promoters, enhancers, in identifying chromatin remodeling effects and its modifiers and also identifying *trans*-factors, as transcriptions factors [Reviewed by 2]. However more and more data have revealed that control of mRNA stability and translation supports the maintenance of homeostasis through rapid influence in protein synthesis levels. This regulatory level contrasts with transcriptional regulation, as it does not induce a permanent change in cell physiology or fate [3, 4].

The so-called Central Dogma of Biology – DNA specifies RNA, which in turn specifies protein – has contributed to the focus on DNA, leaving for the RNA molecule a secondary role in cell biology, of a messenger that acts only as template for synthesis of specific proteins. This thought lead to simplistic descriptions of a complex pathway that do not define the history of "the even most mundane mRNA" [5].

a. Post-transcriptional regulatory mechanisms:

The amount of protein produced from a given gene results from a complex balance of synthesis, modification and degradation acting at distinct levels in gene expression. Study of post-transcriptional regulatory mechanisms has shed light into the

complexity of gene expression control. Genetic and biochemical analysis have revealed that, rather than being independent, the different levels of regulation are intrinsically linked [2]. It has been suggested that several, if not the majority, of cellular transcripts are subject to diverse regulatory activities through their half-life. This regulation is exercised in different points of the RNA metabolism pathway, such as RNA processing (including splicing, editing, capping and 3'end processing), export, localization, turnover and translation initiation of mRNAs.

Post-transcriptional regulation of gene expression at the level of the mRNA molecule seems to be mainly mediated by the UnTranslated Regions, or UTRs (fig.1). The eukaryotic mRNA molecule is composed by three functional regions. The central coding region is flanked by two untranslated regions, characterized by their regulatory functions, located at the 5' and 3' ends of the molecule. At the 5'end of the mRNA molecule the cap structure is found, which is critical for recognition by the ribosome and translation initiation. Additionally this structure has the function of protecting the mRNA from degradation by RNases, and can enhance mRNA export. The poly(A) tail is added at the 3' end of the mRNA during its synthesis and promotes the integrity of the molecule by protecting against exonucleolytic degradation. This structure also has sequence information for mRNA exportation from the nucleus and is functionally involved in transcription termination [6].

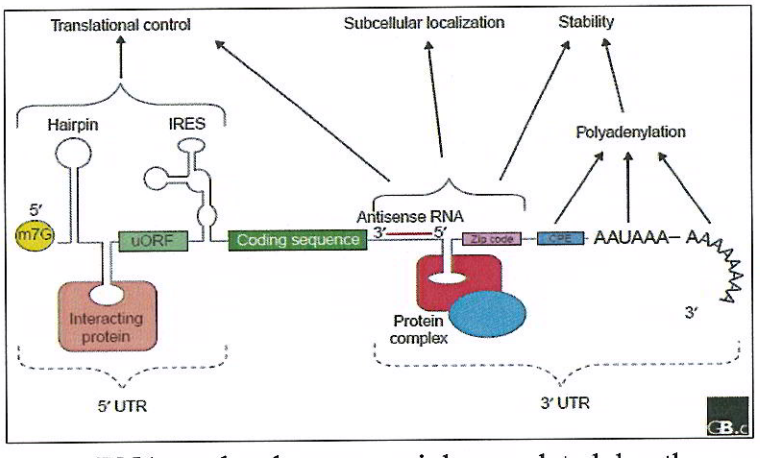


Fig. 1 – The generic structure of a eukaryotic mRNA, illustrating some post-transcriptional regulatory elements that affect gene expression [6].

mRNA molecules are mainly regulated by the combinatorial effect of RNA-Binding Proteins (RBPs) and micro-RNAs networks (microRNAs) [1, 2]. RBPs have been known for a long time to be effectors of post-transcriptional control events. These multifunctional regulatory proteins are highly expressed in cells and act through binding to degenerate sequences often located in the untranslated region of the mRNA molecule [1, 5]. They are also key regulators of alternative splicing events. More recently, microRNAs have been identified as negative post-transcriptional regulatory factors.

Interaction of these molecules with the target mRNA can promote its degradation or translation repression [7]. Together these post-transcriptional factors act in a combinatory-manner as regulatory switches that control the translation and decay machinery.

The agile cellular response to environmental changes has been proposed to occur through the coordinate binding of trans acting-factors involved in RNA regulation to cis-sequence elements that seem to be organized into RNA-operons [Reviewed by 3]. "Ribonucleoprotein structures" in which multiple mRNA encoding functionally related proteins are co-regulated because they share common RBP and microRNA target sequence elements. This leads to similar post-transcriptional regulation of related mRNAs, especially at the level of stability and translation, allowing the dynamical flow of information during cell growth, or differentiation and in response to environmental changes. Additionally, if an mRNA is involved in more than one cellular function then it would be grouped into different RNA-operons leading to differential co-regulation achieved by a coordinated combinatorial association with specific RBPs and/or microRNAs [3].

Some *trans*-acting factors remain stably associated with the target mRNA while others exchange during the RNA life-time in a dynamic way. The complex formed by the association of a mRNA molecule with its regulatory *trans*-acting factors is called a messenger ribonucleoprotein particle (mRNP) and its fate in the cell is dictated by which factors are associated as well as their binding position on the mRNA molecule. Motifs (sequence or structure) involved in *trans*-acting factors association are most often localized on the UTRs of the mRNA molecule (Fig. 1). Positive and negative regulatory-factors that regulate interactions with other cellular factors, or influence sub-cellular localization and/or stability of the target mRNA molecule have been already identified. Considering the Post-Transcriptional operon theory, when *transacting*-factor levels and activity are altered through changes in cellular conditions, different sub-groups of mRNAs would be affected and different RNA regulatory pathways would be activated, allowing, in this way, a great cellular plasticity [5]. Recent data is increasingly revealing the relevance of this regulatory level supporting the view that it may have a similar impact on gene expression outputs as the initial control at the level of mRNA synthesis.

As mentioned before, post-transcriptional regulation can be achieved through different levels of control that intercommunicate and overlap with one another. One of

these levels is the regulation of mRNA stability. Regulation of mRNA turn-over is accomplished by the presence of specific control elements, usually localized in the mRNA 3'UTR regions, which are recognized by several RBPs and microRNAs [1, 3]. To promote a correct flow of the genetic information it is important that the half-life of a specific mRNA is correctly regulated. This regulation can be modulated according to changes in intra- and extra-cellular conditions. To enable this fine regulation, an efficient degradation machinery acts in the degradation of the mRNA molecule from the 3'end, by removal of the poly-A tail. This process has been proven to be crucial for controlling mRNA stability. After this event, degradation can occur from both ends. Endonucleolytic degradation can also promote mRNA degradation, without the need for the step of deadenylation, by the recognition of sequence-specific elements in the mRNA through microRNA molecules that activate the RNA-induced silencing complex (RISC) [1, 3].

mRNAs encoding proteins that participate in the same macromolecular complex or belong to the same functional class tend to have a similar half-life. This might be explained, once again, by the post-transcriptional operons model and specifically at the level of mRNA turn-over regulation the existence of mRNA-decay regulons, formed by cooperative binding of specific RBP and microRNA complexes [1, 3, 8]. These similar turn-over patterns are found in related genes as for example, genes involved in biologic clocks [3] and seem to be conserved through-out evolution [1].

In addition to the mechanisms described above, there are others pathways that regulate mRNA decay but with a purpose of verifying the quality of the molecules and to ensure that faulty transcripts are destroyed. In order to avoid the existence of transcripts that codify for a truncated protein due to the presence of a premature stop codon, a surveillance mechanism – Nonsense-mediated Decay – is activated [9]. This surveillance mechanism is normally distinct from the pathways that control the mRNA turnover, but overlaps between components of both pathways have been found. Additionally, several lines of evidence suggest a regulatory role for NMD as a general mechanism to down-regulate gene expression and as suppressor of transposons [1].

Control at the level of translation is another important step in the post-transcriptional regulation of gene expression. Translational regulation is of great relevance as it can promote a rapid response to alterations in cellular conditions, as cellular response to stress and apoptosis or to promote growth and coordinated division. All of these situations require a fast and precise change in protein levels. On one hand

this regulation can occur at a general level, affecting simultaneously the translation of all mRNA molecules. Although it is thought that the majority of mRNAs are actively translated immediately after the export to the cytoplasm [5], translation initiation is very complex and requires assembly of multiple proteins that are present in limited amounts in the cell [1]. Phosphorylation state of the translation initiation factors and the adjustment of available ribosome in the cell are the major points of control at this level [1]. On the other hand, translation regulation can be directed to a particular group of mRNAs through specific mechanisms that are activated through differential association of RBPs and microRNAs to specific structures or control elements usually present in the UTR [1, 5]. Formation of these specific RNA-protein complexes is thought to occur after the first passage of the ribosome along the mRNA by dissociation of the nuclear RBPs. After each round of translation, deposition of specific *trans*-acting factors occurs depending on the cellular conditions, allowing for a fine control of time and rate of translation and/or a proper subcellular localization of the mRNA molecule [5]. Some of these protein-RNA complexes are thought to promote the interaction between the 5'cap and the poly(A) tail resulting in the circularization of the mRNA molecule. This circularization enables an efficient translation reinitiation by the ribosome and also enables protection of both mRNA extremities from the degradation machinery [2, 4, 5]. This close functional link connecting the 5' and 3'UTRs is an important aspect since it enables communication between the different *cis*- and *trans*-acting factors from both unstranslated regions, promoting in this way a greater range of regulatory functions, since it can enhance, inhibit or even neutralize certain regulatory signals [6, 10].

In the past years great efforts have been made in the identification and characterization of mRNA regulatory elements and in the identification of binding proteins and their target motifs. Available data suggests a crucial role for the unstranslated regions of the mRNA (5' and 3'UTR regions) at the level of post-transcriptional regulation (Fig. 1), although additional features of the mRNA can contribute to it. In particular, additional importance has been added to the UTR regions since mutations that alter these sequences have been correlated with the appearance of diseases [6, 10].

Unlike what has been described for DNA, RNA regulatory motifs result from a combination of primary and secondary structures [6]. Structural features of the 5'UTR have a major role in controlling translation initiation rates, such as 5'UTR sequence length, presence of stable secondary structures, as hair-pin structures, presence of

Internal Ribosome Entry Sites (IRES), presence of structural elements that act as platforms for interaction with translation repressors and existence of an upstream AUG (uAUG) and a upstream Open Reading Frame (uORF) [4, 6, 10].

Eukaryotic translation initiation involves binding of the translation apparatus to the 5'cap, followed by scanning for the closest AUG codon, which will be recognized as the initiation codon. Several data suggest the existence of a consensus sequence flanking the AUG initiation codon that in part can modulate the efficiency of its recognition. A preference of the 40S ribosomal subunit for the conserved GCC(A/G)CCaugG sequence, which can be observed in animals, plants and fungi has been described. However large fraction of the 5'UTR regions contain uAUGs followed by termination codons, defining uORFs upstream of the main coding sequence (ranging between 15-50% of transcripts, depending on the organism) [6]. These are particularly frequent in transcripts of oncogenes, growth factors and cellular receptors [12]. When these upstream elements are present in the 5'UTR, this region is usually described as 5'Leader, given the fact that it is not really an unstranslated region. The presence of this uAUG and uORFs is thought to allow for the fine regulation of translation. This feature inhibits the production of the main protein product [12] and is suggested to have a role in keeping the expression of specific genes at low levels [10]. Since not all protein production from the main ORF is inhibited, two mechanisms have been proposed. The first suggests that sometimes the 40S ribosome subunit can bypass the uAUG, because it is in a poor consensus context, allowing the recognition of a distal AUG more often than expected, in a mechanism called "leaky scanning" [6, 12]. The second one proposes that the ribosome remains attached to the mRNA and resumes scanning, with translation reinitiation, at a downstream AUG codon [6, 12]. Thus, these uAUG/uORFs are characterized as translation barriers, by preventing ribosome access to the downstream AUG. This mechanism seems to be relevant for the regulation of protein products known to be harmful when abundant [12].

The structures present in the 3'UTR can participate in translation control but are mainly involved in monitoring mRNA stability and sub-cellular localization. As mentioned earlier, mRNA turnover is a crucial step in post-transcriptional regulation, profoundly influencing the mRNA abundance in the cell. A well studied example is the mRNA AU-rich element, a pentanucleotide AUUUA motive recognized by specific RBPs, which promotes mRNA decay in response to specific signals. The sub-cellular localization can also be an important post-transcriptional control event, in particular

during early development. Indeed, confining a mRNA to a specific place may be more efficient than to asymmetrically localize proteins in the cell. Instead of distributing and maintaining several protein molecules inactive, all efforts can be focused in only one molecule that will serve as a template for multiple rounds of translation [6].

Existence of alternative UTRs can be considered as another post-transcriptional control, since it can represent an efficient system to modulate post-transcriptional control mechanisms by changing the inherent stability or translation efficiency properties of specific mRNA molecules. These alternative sequences are formed from the use of different transcription-start sites, polyadenylation sites or splice donor and /or accepter sites [13].

Over several years of research, an impressive body of evidence has demonstrated that a significant part of gene expression control takes place at the post-transcriptional level, in particular by determining the stability and translation efficiency of mRNA molecules. Understanding the regulatory mechanisms underlying these control events would be a powerful tool to shed light into the processes underlying coordinate regulation of gene expression in eukaryotes. This research effort would not only be important as general knowledge but it would be particularly central in disease research, where the characterization of these regulatory mechanisms could lead to the establishment of methods for therapeutic modulation of the molecule involved in the disease.

b. Spinal Muscular Atrophy and SMN 1 and 2 gene:

Spinal Muscular Atrophy (SMA) is an autossomal recessive neurodegenerative disease characterized by a progressive loss of alpha motor neurons in the spinal cord, leading to muscle weakness and atrophy [14]. Depending on the SMA type, it can manifest early in human development, representing the leading inherited cause of infant and childhood mortality [15]. SMA has an incidence of approximately of 1 in 10.000 births and a carrier frequency of 1 in 35 [16].

Type I SMA is the most severe form of the disease, in which patients show generalized muscle weakness very early in development, within the first 6 months after birth and have a life expectancy below 2 years of age. Type II SMA is characterized by the manifestation of symptoms after 6 months of age and these children can survive beyond their 2nd year. Type III SMA is less severe and usually patients have the first symptoms around the age of 3 and have the ability to sit and walk, showing a light form

of the disease. Type IV SMA patients are mildly affected and can have a normal motor development [16].

As with many neurodegenerative disorders, SMA affects the voluntary movement used for activities such as crawling, walking, head and neck control, and swallowing and can also compromise the respiratory muscles, leading to an increased tendency for pneumonia and lung problems, which are the main cause of early mortality. At present there is no cure for SMA and all treatment consists only of managing these symptoms and preventing further complications.

Lefebvre et al [17] reported large-scale deletions in the Survival of Motor Neuron1 (SMN1) gene as causing the SMA disorder. SMN1 is mapped in the telomeric region of the chromosome 5 and is responsible for production of the Smn protein. This 38KDa Smn protein is ubiquitously expressed and is known to have a relevant role in RNA metabolism [18]. The best characterized SMN complex is known to be involved in the biosyntheses of small nuclear ribonucleoproteins (snRNPs) complexes, crucial for the recognition of splice sites and catalytic removal of introns from the pre-mRNAs [18, 19, 20]. The reason why low levels of this protein have such dramatic effects on motor neurons it is still unknown, since the Smn protein is present in all cells types. One hypothesis is that these cells could be more sensitive to deregulation of splicing mechanisms. Another more likely hypothesis is that this protein has an additional function in alpha motor neurons. Indeed, as suggested by some authors, SMN may be involved in mRNA trafficking along the axon [18]. This is supported by the identification of Smn in dendrits and axons through immunocytochemical studies [19, 21]. When the SMN gene is knocked-out in zebrafish embryos, a significant axonal dysmorphology is observed, suggesting an important role for Smn in formation of nerve connections during development [22]. Accordingly, high levels of SMN expression have been reported in embryos, which decrease rapidly after birth [16]. Taking together these data, it has been suggested that SMA symptoms begin when Smn levels fall below a critical threshold and that severity of symptoms partly depends on the moment when this decrease occurs during development [23].

As it is the case for other neurodegenerative disorders, research efforts are being made in search for effective therapies for SMA. In this particular disease, this research effort is encouraged by an interesting molecular aspect: in 1995 a second, almost identical copy of the SMN1 gene was identified in humans, the SMN2 gene, which has originated from an event of inverted duplication [20]. All SMA patients have at least, one

copy of the SMN2 gene [15], which is located in the centromeric region of the chromosome 5 [20]. The crucial difference between these two genes is a silent mutation transition (C→ T) in exon 7. Although it doesn't affect the amino acid sequence, this silent mutation severely affects the correct splicing of exon 7, by disrupting a splicing enhancer within the SMN2 gene [15]. As a result of this mutation, skipping of exon 7 occurs in 75% of the SMN2-derived transcripts, leading to the production of a truncated and non-functional protein that is thought to be rapidly degraded (fig.2) [23].

It has been reported that regulation of inclusion of exon 7 in the SMN mRNA depends on a large number of positive-acting *cis* elements – splicing enhancers – and on negative-acting *cis* elements – splicing silencers. These *cis* elements are recognized by several *trans*-acting splicing proteins.

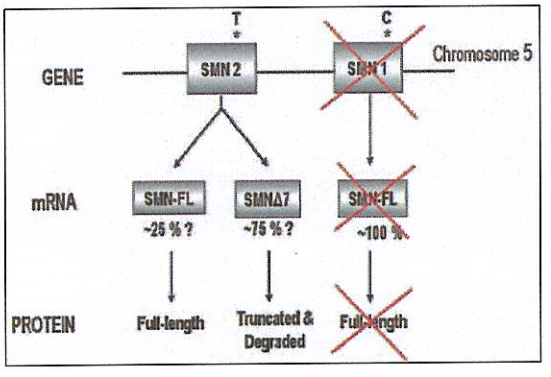


Fig. 2 – Genetics of spinal muscular atrophy [23]. Low levels of full-length mRNA and functional protein (25%) is found in cell and tissues derived from patients, however this amount is insufficient for correct function of the motor neurons.

Several researchers have described one of these splicing enhancers as being recognized by the SF2/ASF splicing factor. With the C to T silent transition this enhancer sequence is lost and SF2/ASF is no longer recruited, making exon 7 unrecognizable to the splicing machinery, resulting in the skipping of this exon [16, 24]. A purine rich splicing enhancer that promotes exon inclusion was identified in the central part of exon 7, to be recognized by the Htra2-β1 splicing factor. Hofmann et al. discovered that the interaction between this *cis* enhancer sequence and Htra2-β1 is one of several factors responsible for 25% of correctly spliced full-length SMN2 transcripts produced [25].

The copy number of the SMN2 gene varies in the population and in patients with SMA this variation modulates disease severity and the age of disease onset defining the four types of the SMA [16]. Thus, gene expression from the SMN2 gene seems to be able to rescue the SMA phenotype when sufficient levels of Smn protein synthesis are achieved. This correlation of disease symptoms and severity with SMN2 copy number is also observed in SMA mice models. When the single endogenous murine SMN gene is knocked-out, early embryonic lethally occurs, demonstrating the important role of

this gene for cell function and development. These mice can be rescued by introduction of human SMN2 in the $SMN^{-/-}$ background, and they present the SMA phenotype. Monani et al demonstrated that when only two copies of the human SMN2 gene are expressed in $SMN^{-/-}$ mice, a severe form of the disease is observed and death occurs within a week after birth. But when the number of copies is increased, rescue of the SMA symptoms is observed [20, 26]. If we take into account that 25% of the transcripts produced by the SMN2 gene are full-length (include exon 7) and that those transcripts are capable of producing a functional protein, the increase of the SMN2 gene number may alleviate the symptoms by compensating the low levels of Smn protein in motor neuron cells [23], even though a considerable amount of the delta exon 7-SMN2 mRNA is being produced. These mRNAs give rise to a truncated protein that hasn't been found to have a dominant negative effect on the functional protein. Instead, it was observed by Le et al. that high levels of the truncated protein can retain some functionality and do ameliorate disease symptoms in the SMA mouse model [27].

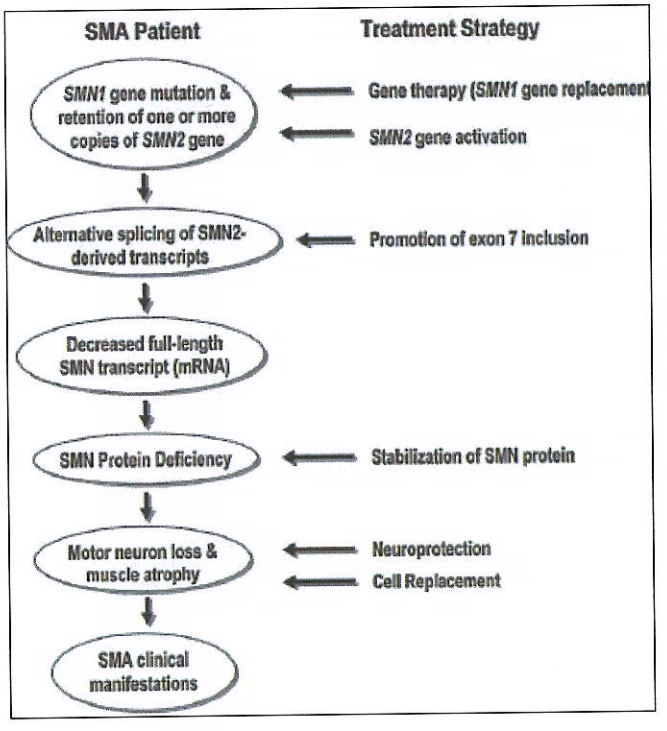


Fig. 3 – Molecular therapeutic targets proposed and tested for the SMA given the knowledge about the mechanisms regulating the SMN gene. The research focus is in the comprehension of how exon 7 inclusion can be promoted, since it is considered the most efficient therapeutic approach. However it is worth noting the visible hole on the knowledge of stability and efficiency of translation regulation of this mRNA [23].

The prospect of developing effective therapies for Spinal Muscular Atrophy is in part enhanced by the existence of this second, almost identical copy of the gene disease. Because of that, urgent understanding of the mechanisms that regulate the expression of SMN2 is needed. So far most research has been directed to understanding and characterizing mechanisms at the level of transcription, protein stability and regulation of alternative splicing (fig. 3) [14, 16, 20, 24, 25]. However, in therapeutic approaches,

targeting of the mRNA molecule seems to be more advantageous and reliable than targeting and manipulating the genome itself, since a higher number of copies of these molecules is present in cells [Reviewed by 28].

At the level of post-transcriptional regulation few studies have been developed on the SMN2 mRNA with most of the research effort focusing on the understanding of the mechanisms that regulate the alternative splicing of SMN2 transcripts, seemingly one of the most promising therapeutic approaches to the disease [Reviewed by 14 and 18]. Recent characterization of the turn-over rate of SMN1 and 2 transcripts suggests that both mRNAs have a half-life of 4-5 hours, which is significantly smaller than the half-lives of very stable cellular mRNAs, like GAPDH or beta-actin (>12h) [15]. This observation suggests that SMN mRNAs are subject to post-transcriptional control at the level of stability. Thus, finding regulatory motifs that may be used to modulate SMN2 gene expression at the post-transcriptional level could be an important basis for the development of novel therapeutic approaches. Information about sequence elements acting on mRNA regulation, which serve as binding-motifs for regulatory factors, can be used for the design of highly modified small oligonucleotides that can block these binding-sites and promote significant changes in mRNA levels, thus offering new opportunities to modulate SMN protein synthesis in SMA patients.

2. Objectives and Specific aims

The main goal of this work is to characterize the regulatory mechanisms involved in the control of SMN2 mRNA stability and translation efficiency, within the context of a long term project aiming at the development of novel approaches to increase SMN protein levels in SMA patients.

For this purpose, we propose to achieve the following aims:

1. Verify if the UTR regions of the SMN2 mRNA have any crucial role in the regulation of SMN2 gene expression.

2. Identify SMN2 UTR sequence elements relevant for mRNA stability and translation control.

3. Confirm if there are alternative sequences to the canonical SMN2 UTR regions.

3. Methods and Materials

a. Plasmid vector construction

The pCMV5-luciferase reporter vector for cloning of UTR segments was generated as follows: the Firefly luciferase CDS derived from the pGL3 vector (Promega) was introduced into the pCMV5 eukaryotic expression vector MCS by ligation into HindIII/XbaI restriction sites (Appendix, Fig. 1A), allowing for easy upstream and downstream cloning of the desired UTR sequences. The SMN2 5' and 3'UTR full-length and truncated sequences were obtained either by RT-PCR from Type I SMA B Cells or by PCR amplification of HEK 293 genomic DNA using the primer pairs described in the Appendix (Table 1). 5'UTR sequence were introduced into pCMV5-Firefly Luciferase vector by ligation into MluI/HindIII restriction sites (Appendix, Fig. 2A). 3'UTR sequence were introduced into pCMV5-Firefly Luciferase vector by ligation into XbaI/BamHI restriction sites (Appendix, Fig. 2B).

The control pCDNA3-renilla reporter vector was generated as follows: the renilla luciferase CDS originally derived from the pRL-TK vector (Promega) was introduced into the pCMV5 eukaryotic expression vector pCDNA3 by ligation into XhoI/XbaI restriction sites (Appendix, Fig. 1B).

Recombinant vectors were identified by restriction digested and the insert confirmed by sequencing.

b. Purification of nucleic acids

Small scale plasmid DNA isolations were performed with the GeneJet^{TM} Plasmid Miniprep Kit (Fermentas, Burlington, Canada). Large scale plasmid DNA purifications were performed with the PureLinkTM HiPure Plasmid DNA Purifications Kits (Invitrogen, Carlsbad, CA, USA).

Total RNA was isolated using the TRI Reagent® Solution (Applied Biosystems – Ambion, Austin, USA) according to the manufacturer's instructions.

Genomic DNA from cultured HEK 293 cells was isolated using the Wizard® Genomic DNA Purification Kit (Promega, Madison, USA) according to the manufacturer's instructions.

Nucleic acid concentration was determined by measurement in the Nanodrop Spectrophotometer ND-1000.

c. Site directed mutagenesis

Point mutations were introduced into the Luciferase-SMN 5'UTR or Luciferase-SMN 3'UTR, using the KOD1 HotStart Taq DNA polymerase (Novagene, Darmstadt, Germany) with complementary pairs of custom designed HPLC-purified mutagenic primers (Stabvida, Portugal) (Appendix, Table 1). Primers were designed according to guidelines of the QuickChange® Site-Directed Mutagenesis Kit (Stratagene). PCR amplification conditions were as follows: 95°C, 2min; [95°C, 20sec; 55°C, 10sec; 72°C, 3min] X 23 cycles. Mutagenic PCR amplification was confirmed by agarose gel electrophoresis and the resulting plasmid was digested with DpnI (Fermentas, Burlington, Canada) for removal of parental bacterial DNA. After bacterial transformation and plasmid DNA isolation, the presence of each mutation was verified by DNA sequencing.

d. Reverse Transcription and PCR

Before reverse transcription, RNA samples were treated with DNAse I (Fermentas) according to the manufacturer's instructions. The first strand cDNA synthesis from total RNA samples was performed using the RevertAidTM H Minus First Strand cDNA Synthesis Kit (Fermentas) according to the manufacturer's instructions using 1μg of RNA and oligo dT priming. PCR reactions were performed with the DreamTaqTM DNA Polymerase (Fermentas) according to the manufacturer's procedure. Standart thermal cycler conditions for PCR reaction were as follows: 95°C, 2min; [95°C, 30sec; 60°C, 30sec; 72°C, 1min] X 40 cycles; 72°C, 5min.

To confirm the existence of alternative isoforms of the SMN2 UTR regions, previously predicted by bioinformatic analysis, isoform specific sets of primers were designed (Appendix, Fig. 5 and Table 1).

e. Cell culture and transfection assays

HEK 293 cells were grown in Minimum Essential Medium (MEM) + Glutamax-I supplemented with 10% (v/v) Fetal Bovine Serum (FBS), 1% (v/v) PenStrep and with 1% (v/v) Non-Essential Amino Acids and subcultured every 3/4 days until a maximum of X passages.

HEK 293 cells were transiently transfected, using TurboFect™ *in vitro* Transfection Reagent (Fermentas, Burlington, Canada), following the manufacturer's instructions.

Transfection conditions (reporter and DNA concentration, cell density and amount of transfection reagent) were optimized for the specific experimental conditions. For transient transfection assays, HEK 293 cells were seeded 24h before at a cell density of $2X10^4$cells/well (for 96-well plates) or $5X10^5$cells/well in 6-well plates.

For luminescent reporter assays, the Firefly and Renilla reporter vectors were premixed in a ratio of 1:1 using the pSK plasmid as a filling vector in the following proportion/96 well transfection: 10ng of Firefly Luciferase vector, 10ng of Renilla Luciferase vector and 180ng of pSK.

Transfections were performed with a mix of 200ng of DNA, 20µl Opti-MEM and 0,3µl TurboFect™ per well for 96 well plates, or 1,5µg of DNA, 150µl Opti-MEM and 1,5µl TurboFect™ per well for 6 well plates.

f. Luminescence Assays

Luciferase activity was determined using the Dual-Glo® Luciferase Assay System (Promega, Madison, USA) and the Glomax-96 microplate Luminometer after 24 hours of transfection, following the manufacturer's procedure.

g. Quantitative RT-PCR

Total RNA was extracted from HEK 293 cells transfected with the Luciferase reporter vectors as described above. cDNA samples were diluted 5 times before PCR and negative control (RT-) reactions were tested for genomic DNA contamination.

qPCR was performed with the Power SYBR® Green PCR Master Mix (Applied Biosystems) in the CFX Real Time System C1000 Thermal Cycle (BioRad) following manufacturer's instructions. Firefly and renilla luciferase sequences were amplified using specific primer pairs (Appendix, Table 1) and mRNA levels were determined form a standard amplification curve made from serial 1:5 dilutions of the corresponding plasmid DNA, from an initial concentration of 1pg/µl.

h. cDNAs samples

cDNA samples from GM03813B Fibroblasts and GM10684A B-Lymphocytes SMA patients cell lines, from Coriell Institute, were used for amplification of the SMN2 UTR sequences. Commercial cDNA samples from Ambion were used for detection of alternative UTR isoforms expression.

4. Results

Post-transcriptional control events have proven, throughout the years, to play a crucial role in the regulation of gene expression. Many of these control mechanisms are achieved by sequence elements present in the non-coding regions of the transcripts.

The main goal of this project was to identify sequence elements in the SMN2 mRNA UTR regions that might be relevant for regulatory control of SMN2 gene expression. To achieve that, we took advantage of a reporter system based on a vector containing the Firefly Luciferase ORF and the SMN2 5' or 3'UTR regions. Possible effects of these regions in SMN2 expression will be inferred from monitoring their effects on Luciferases expression levels.

1. Characterization of the 3'UTR region of the SMN2 mRNA

Our major interest is to identify and then characterize the effect of the SMN2 3'UTR motifs involved in the control of stability of the mRNA molecule. For this purpose it was necessary to clone the 3'UTR region in the pCMV-luciferase reporter vector, which was built for easy subcloning of UTR segments upstream and downstream of the luciferase coding sequence (Appendix, Fig. 1A and 2B).

Although according to nucleotide database there are no difference between the SMN1 and SMN2 UTR sequences, we isolate cDNA from SMA patient-derived cell lines, which do not express SMN1 transcripts.

For PCR amplification of this region it was necessary to design a specific set of primers. The 3'UTR terminal region of the SMN2 mRNA is a U-rich motif sequence, which difficults the design of a reverse primer for complete amplification of the UTR. For this reason, a first set of primers was designed for SMN2 3'UTR amplification from SMA fibroblast and B cell lines cDNA, upstream of this 60 nucleotide 3'end region. After the cloning process, sequencing of our constructions revealed that the reverse primer mis-hybridized with different motifs along the UTR, enabling the construction of a panel of truncated forms of the 3'UTR. Additionally, we confirmed that the isolated sequences were identical to the SMN1 and SMN2 UTR sequences present in nucleotide databases. However complete sequence of the SMN2 3'UTR was still necessary to infer correctly the regulatory effect of the downstream U-rich region. Sequence analysis of the SMN2 genomic region allowed the identification of a downstream region from the gene that was used for full-length amplification of the 3'UTR from HEK293 genomic

DNA. This panel of full-length and truncated 3'UTR constructs (Appendix, Fig. 3) was used for the characterization of potential regulatory regions within the SMN 3'UTR.

1.1. The SMN2 3'UTR has a regulatory effect on mRNA expression

First of all, it is necessary to confirm that the 3'UTR region of the SMN2 gene has any regulatory effect on controlling SMN2 gene expression. After transfection of HEK 293 with the different Luciferase-SMN 3'UTR constructions, Luciferases expression levels were measured and normalized to Renilla expression levels. Similar Luciferase protein level is observed in the empty, full-length and 512nt truncated 3'UTR constructions. Compared to these constructions, the 296nt truncated 3'UTR shows the highest protein levels and the 150nt truncated construction, the lowest (Fig. 4).

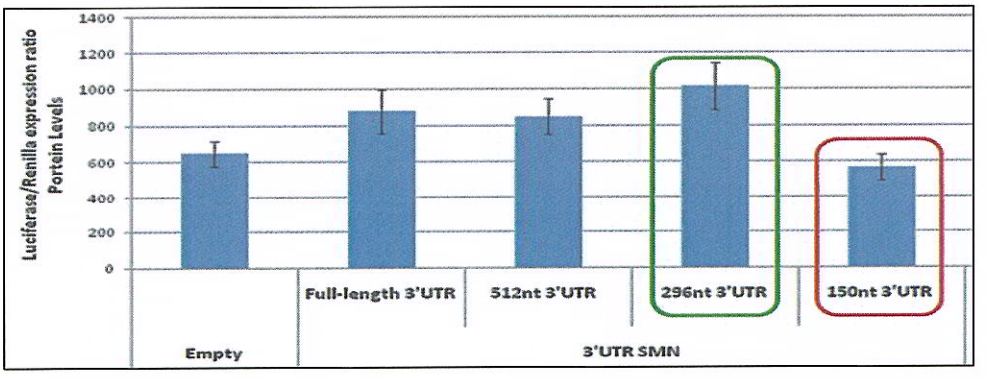


Fig. 4 – Normalized luciferase expression for the panel of Luciferase-SMN-3'UTRs constructions. Cloning of the 3'UTR region of the SMN2 into the pCMV-firefly luciferase reporter vector was used for characterization of the SMN2 3'UTR regulatory effects. Using Dual-Glo Luciferase Reagent®, Firefly Luciferase expression levels were measured and normalized to a control Renilla Luciferase reporter. Compared to the empty and full-length SMN2 3'UTR construction, the 296nt 3'UTR construction shows the highest expression levels and the 150nt construction, the lowest. These results provide evidence for post-transcriptional control events promoted by the SMN2 3'UTR region.

Although, comparing to the empty construct a slight increase in luciferase expression is observed for the full-length construct, the different levels of luciferase protein production from the truncated constructions suggest that the 3'UTR region has some regulatory effect over the reporter gene expression. These preliminary results suggest the existence of post-transcriptional regulatory events promoted by the SMN2 3'UTR region that control SMN2 gene expression.

Additionally, it can be observed that the truncated forms of the 3'UTR region have different regulatory effects on the luciferase expression (Appendix, Fig. 4). The 296nt truncated construction has a positive effect on the luciferase protein production since higher ratio is obtained between the firefly and renilla luciferase's activity

measurement. In contrast, the 150nt truncated construction has the lowest levels of normalized luciferase activity, suggesting a negative effect from this portion of the UTR region on reporter gene expression.

Through the evaluation of luciferase expression in the panel of Luciferase-SMN 3'UTR reporter vectors it is possible to identify three majors regions of regulation in the UTR (Appendix, Fig. 4). When cloning the first portion (5'end) of the 3'UTR in the reporter system – 150nt 3'UTR truncation – low levels of Luciferase protein is observed, due to a negative regulatory effect by this portion. When the middle portion of the 3'UTR is added – 296nt 3'UTR truncation – recovery of the Luciferase protein production is observed, suggestion the presence of positive sequence elements in this fragment. With the addition of the 3'end of the UTR region to the reporter vector – full-length and 512nt 3'UTR truncation – reduction of luciferase expression levels is again observed. This reduction, however is not as severe as the one observed for the first portion of the 3'UTR, suggesting the presence of negative-acting sequence elements in this region.

With these observations, three regulatory boundaries can be identified in the SMN2 3'UTR region, where the terminal ends of the SMN 3'UTR have negative effects on gene expression. The balance between these three regions, between the positive and negative acting-sequence elements, dictates the overall expression rate of the luciferase-SMN2 mRNA.

1.2. Post-transcriptional regulation by the SMN2 3'UTR occurs at the level of mRNA stability

After verification that the SMN2 3'UTR region is involved in post-transcriptional regulatory events, it is essential to characterize the underlying mechanism. For this purpose it is necessary to determine whether an increase or decrease in the luciferase signals is due to differential stability of the mRNA molecule or to differential levels of translation.

HEK 293 cell were transiently transfected with the full-length and truncated forms of the firefly luciferase-3'UTR reporter vectors and the control renilla luciferase vector and the mRNA levels produced from each construction was measured by qRT-PCR.

Data analysis from the normalized mRNA levels indicates that the empty vector, the full-length and 512nt 3'UTR truncated constructions have very similar mRNA levels. Additionally, in comparison to all constructions, the 296nt 3'UTR construction

shows the highest Luciferase mRNA levels, while the 150nt 3'UTR construction shows the lowest (Fig. 5). The distribution of the different luciferase-SMN 3'UTR mRNA levels is thus very similar to the distribution observed for luciferase activity (Fig. 4). This similarity between the mRNA and protein levels suggests that the post-transcriptional events are taking place at the level of mRNA stability. If the post-transcriptional control was occurring at the translation efficiency level, similar mRNA levels would be observed for all of the Luciferase-SMN 3'UTR reporter vectors.

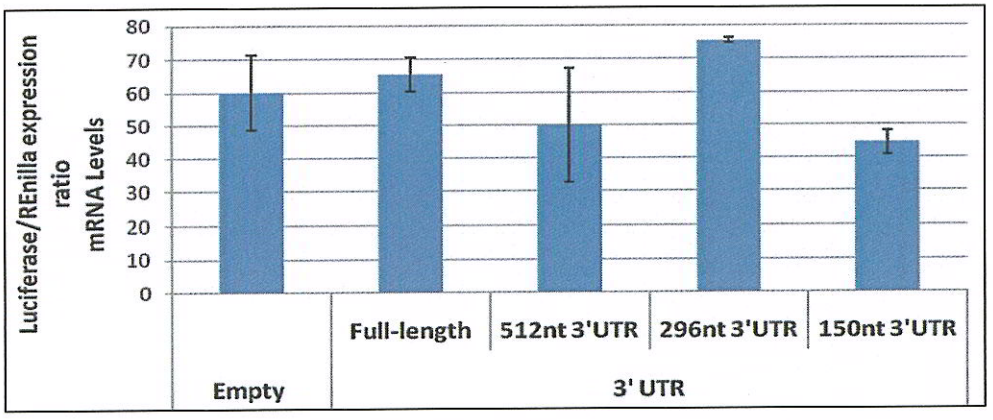


Fig. 5 – Analysis of Luciferase-SMN 3'UTR mRNAs levels by qPCR. Transfection of HEK 293 cell line with the all Luciferase-SMN 3'UTR reporter vectors and the renilla control was performed. After mRNA extraction with TRI Reagent® Solution (Applied Biosystems), quantitative RT-PCR was performed using specific primers for amplification of Firefly and Renilla Luciferase mRNAs and the ratio between the two was determined. Normalized mRNA levels for each luciferase reporter follow the same pattern as the distribution of protein levels observed earlier (Fig. 4). These results suggest that the post-transcriptional event promoted by the 3'UTR region is regulating the stability of the mRNA molecule.

1.3. Conserved nucleotides in the SMN2 3'UTR have a negative effect on luciferase expression

After verification that the 3'UTR region of the SMN 2 gene is involved in post-transcriptional regulation of gene expression by influencing mRNA stability, it would be important to identify specific UTR sequence elements influencing SMN2 mRNA expression. In particular, conserved nucleotides in this region and regions already identified or predicted to be binding-sites for regulatory RBP or microRNAs are good candidates for key regulatory elements (Appendix, Fig. 5).

Conservation of non-coding regions across species is a hallmark for the presence of regulatory motifs. Although SMN2 is found only in humans, the fact that is almost identical to SMN1 allows for a cross-species comparison of the UTR sequence. Previous bioinformatic analysis performed in the lab across five mammalian species revealed that the 3'UTR of the SMN1/2 mRNAs has several nucleotides with a high

degree of conservation (Fig. 6), suggesting the presence of regulatory sequence elements.

Fig. 6 – Point mutations carried out in the conserved sequences in the SMN 3'UTR. Conserved nucleotides in the SMN 3'UTR between human, cow, dog, elephant and armadillo sequences are represented by capital letters. The color code depicts the panel of seven mutagenized vectors derived from the Luciferase-SMN 3'UTR reporter vector. The mutations introduced in each construct are superscripted.

To obtain evidence for a regulatory role of conserved nucleotides in the SMN 3' UTR we attempted to disrupt conserved sequence stretches by introducing 2 or 3 purine-pyrimidine transitions by site directed mutagenesis (Fig. 6).

Seven mutagenized vectors derived from the Luciferase-SMN 3'UTR reporter were tested for differential Luciferase expression ratio. Data from the different mutant constructs was normalized to the expression levels of the wild-type 3'UTR reporter (Fig. 7). Only relative changes in luciferase expression above 20% of the wild-type reporter were considered as significant. Considering this threshold, we find that three out of seven mutagenized regions induced an increase in luciferase expression, whereas the others showed no significant effects. It is very striking to observe such changes in reporter gene expression when, in each construction, only two or three conserved residues were mutated. These results suggest that these residues may have an important role as part of a regulatory sequence element. Mutagenesis of these conserved residues can disrupt the target motif for RBP or microRNA binding involved in negative regulation of the SMN2 expression. Indeed, a bioinformatic analysis of the SMN 3'UTR performed in our lab using algorithms for the prediction of such binding sites identified several motifs overlapping with the introduced mutations (Appendix, Fig. 5). Alternatively, the mutagenesis of the conserved nucleotides may disrupt a regulatory secondary structure element.

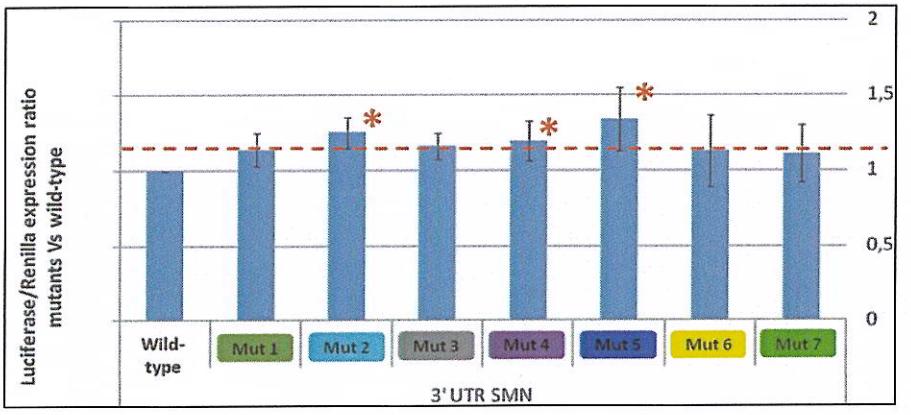


Fig. 7 – Regulatory effects of conserved nucleotides in the SMN2 3'UTR - Luciferase expression levels of the mutant reporters. Luciferase expression levels of the mutated constructs were determined, using Dual-Glo Luciferase Reagent®. Data is normalized to the wild-type reporter. Mutants 2, 4 and 5 show an increase in expression above 20% (orange line) when compared to the wild-type construct. Such significant changes in the reporter gene expression resulting from mutation of two conserved residues are most likely explained as a consequence of disrupting the target motif for an RBP or microRNA involved in negative regulation of SMN2 mRNA expression.

2. Characterization of the 5'UTR region of the SMN2 mRNA

To determine if the SMN2 5'UTR region plays a role in the post-transcription control of gene expression, it was initial necessary to obtain this region for cloning into the Luciferase reporter vector. For this purpose, a pair of specific primers was designed in order to amplify this region from SMA B cell line cDNA. However, this amplification was very inefficient and after analysis of the 5'UTR region it was found that it has a very high GC content that is likely to interfere with the reverse transcription reaction. For this reason, the 5'UTR segment was obtained from genomic DNA amplification from HEK 293 cells.

2.1. SMN2 5'UTR region has a negative regulatory effect on the SMN2 expression

To confirm if the SMN2 5'UTR region has any post-transcriptional regulatory effect on the SMN2 gene expression, co-transfection of HEK 293 cells with the Luciferase-SMN 5'UTR or the luciferase empty vector and the renilla luciferase vector was performed.

In contrast to the Luciferase-SMN 3'UTR constructions (Fig. 4), the reporter with the 5'UTR shows the lowest levels of luciferase expression, suggesting a strong inhibitory effect by this unstranslated portion of the SMN2 mRNA (Fig. 8A).

To understand why we observed a great inhibitory effect in protein production, bioinformatic analysis of the 5'UTR sequence was performed, in order to identify any

regulatory element as a secondary structure or the presence of a uATG or a uORF. Upon sequence analysis, presence of a uORF was identified (Fig. 8B). This regulatory element has been described to have a negative effect over the translation initiation control [11]. However not all protein production is inhibited, since some Luciferase activity is still observed (Fig. 8A).

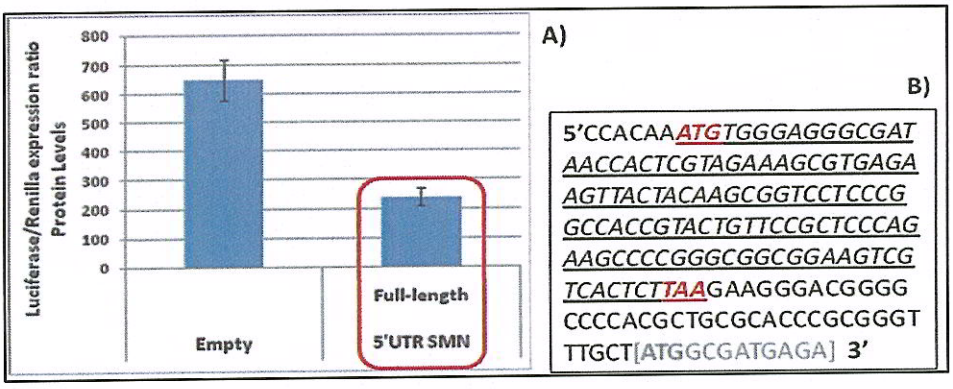


Fig. 8 – SMN2 5'UTR regulatory effect on the Luciferase reporter system. A) Luciferase expression ratio in Luciferase-SMN 5'UTR construction. For characterization of the 5'UTR regulatory effects, upstream cloning of the SMN2 5'UTR region in the Firefly Luciferase reporter system was performed. Luciferase expression levels were measured as described before. Compared to the empty construction, the Luciferase-SMN 5'UTR construction shows low Luciferase expression levels. These suggest a great inhibitory effect by this UTR region. **B) Sequence of the 5'UTR region of the SMN2 gene.** 5'UTR sequence analysis revealed the presence of a uORF (signalized here in underline).

2.2. The SMN2 5'UTR inhibitory effect is due to presence of a uORF

To determine whether the inhibitory effect of the SMN2 5'UTR sequence is due to the presence of the identified uORF it is necessary to disrupt the uATG. For this purpose, a specific set of primers was designed to perform a point mutation of the uATG in the Luciferase-SMN 5'UTR construction into a TTG.

HEK 293 cells were transiently transfected with the wild-type and the uATG mutated Luciferase-SMN 5'UTR constructions (Fig. 9). Compared to the wild-type, we observe a 50% of recovery in the luciferase protein levels when the uATG is mutated, thereby confirming that the inhibitory effect of 5'UTR on luciferase expression is due to the presence of the uORF in this unstranslated region of the SMN2 mRNA.

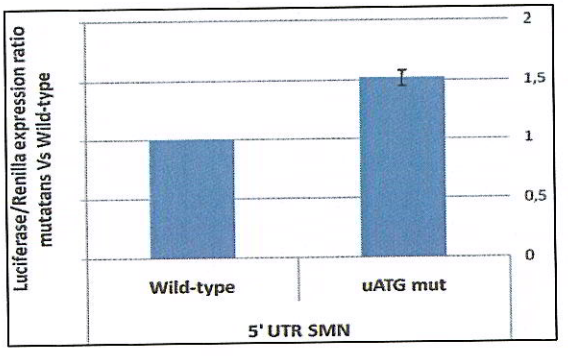


Fig. 9– Luciferase expression levels of the 5'UTR uATG mutant reporter. Luciferase-SMN 5'UTR wild-type was point mutated to disrupt the uATG. Comparing to the wild-type sequence, the mutated construct displays a 50% recovery in luciferase expression. This suggests that the inhibitory effect observed in the Luciferase-SMN 5'UTR construction is due to the presence of a uORF.

3. Search for alternative sequences to the canonical SMN2 UTR regions

Characterization of alternative UTRs present in mRNA molecules indicates that these elements act as modulators of post-transcriptional mechanisms [13]. Using the ExonMine algorithm and database [29], two alternative 5' and 3'UTR isoforms of SMN1/2 transcripts were identified. The significance of these isoforms for SMN2 gene expression is currently unknown. However it is first necessary to prove experimentally the existence of these alternative sequences. For this purpose specific sets of primers were design to search for the presence of transcripts containing these alternative UTRs across different human tissue samples by RT-PCR (Appendix, Fig. 6).

3.1. SMN2 5'UTR isoform is expressed in normal and SMA cell lines

For detection of the alternative 5'UTR, HEK 293 and SMA B derived cDNAs were tested by RT-PCR. Expression of the alternative SMN2 5'UTR (150bp amplicon) is detected in the all RT+ templates tested (Fig. 10A). Further analysis is necessary to characterize the role of this alternative sequence in SMN2 expression.

3.2. Expression of the alternative SMN2 3'UTR isoform was not detected in human cell line or tissue samples

Expression of the alternative SMN2 3'UTR sequence was not observed in the same cDNAs tested for the alternative 5'UTR region. To determine if this was due to t tissue specific expression, the presence of alternative 3'UTR isoforms was performed on commercial cDNA samples from several different human tissues (Fig. 10B). As a control, RT-PCR for the SMN coding sequence was performed on the same samples. As shown in Fig. 10, although the SMN mRNA could be easily amplified in all samples tested, we obtained no evidence that confirms the expression of the 3'UTR alternative isoform.

These preliminary results do not confirm the existence of an alternative 3'UTR for the SMN1/2 transcripts. Nevertheless, the inability to detect this sequence could be due to an active degradation of this mRNA isoform, as a gene expression control mechanism. In fact, this mRNA is predicted to have a splice junction downstream of the Stop codon that could make it a target for the Nonsense Mediated Decay (NMD) pathway.

In order to test this hypothesis, we used cDNAs from HeLa cells where NMD was inactivated by siRNA against the UPF1 protein and the corresponding siRNA control

experiment (a kind gift from Rute and Luísa Romão) to detect the presence of the canonical and alternative 3'UTR sequences using specific primers (Fig. 11). Expression of the alternative SMN2 3'UTR it is still not detected neither in either sample, suggesting that this isoform may be extremely rare or specific or have been predicted from an aberrant database entry.

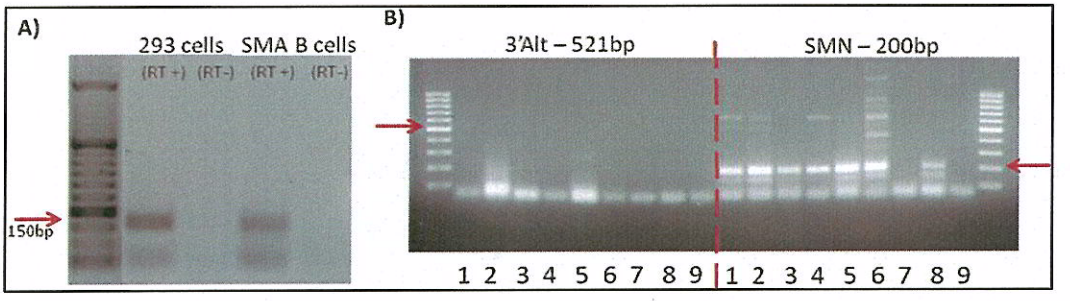


Fig. 10 – Identification of alternative SMN2 UTR isoforms. A) RT-PCR assay for the alternative SMN2 5'UTR isoform in cultured cells. HEK293 and SMA B cell line cDNA were tested by PCR for amplification of 5'UTR alternative isoform. The predicted PCR product (150bp) is detected in RT+ templates and not in the RT- control. **B) RT-PCR assays for the alternative SMN2 3'UTR isoform and SMN mRNA in human tissue cDNA samples.** PCR amplification of the 3'UTR alternative isoform (left) and SMN mRNA (right) was performed on the following human tissue and cell line cDNA samples: 1-kidney, 2-heart, 3-brain, 4-testis, 5-liver, 6-HEK 293 cell line, 7- negative control; 8-SMA fibroblast cell line, 9- negative control The expected PCR product for the alternative 3' UTR (521bp) is not detected in any of the samples, whereas the SMN mRNA (200bp) was detected in all samples tested.

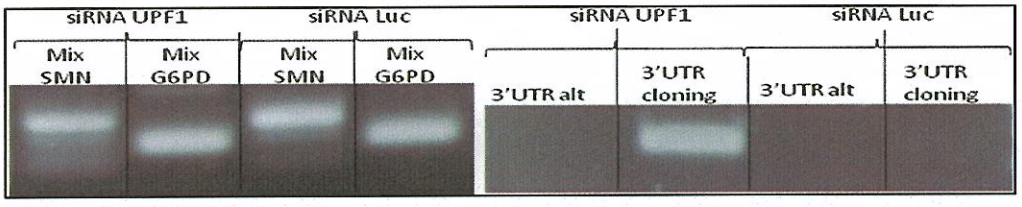


Fig. 11 – Detection of expression of SMN2 alternative 3'UTR in HeLa UPF1 cDNA. Since it was not possible to detect the expression of the alternative 3'UTR isoform in the several cDNAs tested, we hypothesed that the transcript may be degraded by the Nonsense Mediate Decay mechanism. To confirm this, cDNA from HeLa cells treated with UPF1or control siRNA was tested by PCR with the appropriate set of primers. Amplification of the G6PD mRNA (132bp) was used as an endogenous control. Expression of the SMN coding sequence with a primer set which amplifies the internal exons (Mix SMN – 200bp) and of the normal (3'UTR cloning – 512bp) and the alternative (3'UTR alt – 521bp) 3'UTRs was performed. Expression of the alternative 3'UTR was not detected in any of the cDNAs tested. However, expression of the normal 3'UTR is greatly increased in the HeLa-UPF1 cDNA but not in the HeLa-Luc cDNA, suggesting that the canonical SMN2 mRNA is normally regulated by NMD pathway.

Quite surprisingly, we found that the expression of the normal 3'UTR was significantly increased in the HeLa-UPF1 cDNA when compared to the HeLa-Luc cDNA, suggesting that the canonical SMN2 mRNA is normally regulated by NMD pathway. Further studies are necessary to confirm this observation and determine its significance for SMN2 gene expression.

5. Discussion and Perspectives

The main goal of this project was to characterize the post-transcriptional mechanisms defined by the UTR regions involved in the regulation of SMN2 gene expression. Our results strongly suggest that both SMN2 UTRs have a role in post-transcriptional regulation, as significant positive and negative regulatory effects were observed when these regions were cloned into luciferase reporter vectors.

In particular, for the SMN2 3'UTR, analysis of the 3'UTR truncated constructions allowed the definition of three regulatory boundaries. The 150nt 3'UTR truncated construction showed the lowest protein expression levels, suggesting a strong inhibitory effect by this portion of the SMN2 3'UTR. When analyzing the 296nt truncated construction we observed a release of the inhibitory effect exerted by the first UTR segment. Recovery of luciferase levels suggests the presence of positive *cis*-acting elements in this intermediate region of the 3'UTR. However when the effect determined by the terminal portion of the SMN2 3'UTR was analyzed (full-length construction), an inhibitory effect was observed again, although less strong than the one exerted by the first 150nt nucleotides of the 3'UTR. The balance of the negative-acting sequence elements present in the terminal ends of the SMN2 mRNA with the positive-acting sequence elements present in the middle portion of the UTR regions dictates the expression output of the SMN mRNA. The communication between the different regions within the 3'UTR results from the binding of positive and negative-acting *trans*-factors to specific motifs present in this region. However the balance of the positive and negative effects can change through time according to specific extra and intracellular signals by changes in the *trans*-factors associated to the 3'UTR region.

The binding of these regulatory switches, RBPs and microRNAs, control the gene expression at the post-transcriptional level, by determining the stability and translation efficiency of the mRNA. In order to characterize what post-transcriptional mechanisms are acting through the 3'UTR on the control of the SMN2 mRNA, we tried to understand if the changes in luciferase levels were due to changes in the translation rate or stability of the mRNA molecule. Analysis of mRNA levels for each construct suggests that the regulation is carried out through control of molecule stability, as the levels of the mRNA molecules vary in a very similar way as the protein levels. If the observed post-transcriptional regulation was acting through control of translation efficiency, the observation of similar mRNA levels in all constructions would be

expected. These results suggest the presence of binding-motifs for regulatory factors that intercommunicate with each other to regulate the rate of mRNA turn-over.

For a fine characterization of regulatory motifs active on the SMN2 3'UTR region, we relied on a previous bioinformatic analysis performed in our lab in order to identify the presence of conserved nucleotides and putative RBP and microRNA target sites (unpublished). This study revealed a high degree of conservation in the 3'UTR and identified several candidate regulatory RBPs and microRNAs. In order to identify if these specific UTR sequence motifs influence SMN2 expression, a panel of mutant reporters was produced. We observed that disruption of some of these conserved nucleotides resulted in increased luciferase expression, possibly because the conserved motif that was disrupted was serving as a platform for binding of negative-acting *trans*-factors. These are very promising data, since when we introduce small changes likely to disturb the binding of regulatory factors to some conserved nucleotides, a significant increase in the amount of protein synthesis is achieved.

To complement these data, it would be important to also characterize the post-transcriptional mechanisms acting on these mutant reporters, although we expected it to be at the level of mRNA stability as it was seen for the other SMN 3'UTR constructions. It would also be interesting to combine some of the mutations performed to observe if there is a synergistic effect on gene expression, thus providing evidence in support of our hypothesis of combinatorial control involving different 3'UTR segments.

Study of the SMN2 5'UTR revealed a strong inhibitory effect by this region in the SMN2 expression due to presence of a sequence element that is expected to lead to disturbance of start codon recognition by the ribosome – a previously unidentified uORF. This regulatory element has been described as a barrier that regulates translation initiation in genes that are critical in the cell. We have confirmed that the inhibitory effect observed in the 5'UTR was due to presence of a uORF because when we disrupted the recognition of the uATG by the ribosome, a 50% increase in luciferase levels was observed. This data is very interesting and therapeutically promising. If possible, it would be interesting the design of complementary modified small oligonucleotide that prevents the recognition of this uATG by the ribosome and abolishes the uORF effect.

To gain further mechanistic insight into the post-transcriptional control by the SMN2 UTR regions, it would be important to test a reporter system with both UTR regions, in order to determine if there is intercommunication between these two

regulatory regions. It has been described that, upon circularization of the mRNA molecule, the UTR regions communicate with each other and promote or neutralize some regulatory effects. Indeed, it has been already described that 3'UTR sequence can alleviate the inhibitory effect of the uORF through different regulatory steps [12]. Thus, to fully understand the post-transcriptional events acting on the SMN2 mRNA and identify those that are valuable for therapeutic modulation, construction of such a reporter is absolutely essential.

Alternative isoforms of the SMN2 UTR regions were previously predicted by bioinformatic analysis. These sequence elements have been described as modulators of the post-transcriptional mechanisms that contribute to cellular response plasticity [13]. Expression of the alternative 5'UTR isoform was detected, but not the predicted alternative 3'UTR isoform, despite testing for possible differential expression depending on cell type and tissue. We hypothesized that the inability to detect this alternative isoform could be because this transcript is regulated by the Nonsense Mediated mRNA Decay mechanism. Indeed, this alternative isoform is generated by a splicing event downstream of the stop codon, which should be recognized as a premature termination codon, committing this transcript to rapid decay. However when this hypothesis was tested with cDNA from HeLa cells with UPF1 interference, expression of a transcript with the alternative 3'UTR was still not detected. But when we tested the expression of the canonical 3'UTR as a control reaction, an interesting result was obtained. Expression of this sequence was greatly increased in the NMD-inactive samples, suggesting that the normal SMN2 gene product may be regulated by this surveillance mechanism. Although further studies are necessary, it would be very interesting to understand if the activation of this regulatory mRNA decay for the canonical SMN2 mRNA could be influenced by the presence of the uORF as suggested by some studies [9, 10].

In order to fully understand the regulatory events acting on the SMN1/2 mRNAs, this work needs to be complemented with the characterization of *trans*-acting factors involved in SMN post-transcriptional regulation. For this purpose, the vectors generated during this study will be used to perform a large scale shRNA screening with a library covering 430 human RNA binding and splicing associated proteins [30], as part of a larger collaborative project. The feasibility of this screening approach depends on a robust reporter model that does not introduce significant variations in the experimental readouts. While developing this study aiming at the identification of regulatory

sequence elements present in the SMN UTR, several HEK293 cell lines stably expressing the luc-SMN3'UTR and renilla mRNAs derived from our plasmid constructs were established and characterized. Thus, in addition to the results described in detail in this thesis, we have built a model system that will allow for future high-throughput screening of regulators of SMN expression acting through 3'UTR sequence elements.

In conclusion, in this work we studied the post-transcriptional regulatory effects of the SMN2 UTR regions, in order to better understand the mechanisms that regulate the stability and translation efficiency of this molecule. By characterizing these effects, we expect to provide novel insights that will help the design of novel therapeutic approaches to modulate SMN gene expression, which may ultimately contribute to improve life quality and expectancy of SMA patients.

6. References

[1] Mata J, Marguerat S, Bähler J: **Post-transcriptional control of gene expression: a genome-wide perspective.** *Elsevier 2005*, **30**:506-514

[2] Komili S, Silver PA: **Coupling and coordination in gene expression processes: a systems biology view.** *Nature Reviews 2008*, **9**:38-48

[3] Keene JD: **RNA regulons: coordination of post-transcriptional events.** *Nature Reviews 2007*, **8**:533-543

[4] Sonenberg N, Hinnebusch AG: **Regulation of translation initiation in Eukaryotes: mechanisms and biological targets.** *Cell 2009*, **136**:731-745

[5] Moore MJ: **From birth to death: the complex lives of eukaryotic mRNAs.** *Science 2005*, **309**:1514-1518

[6] Migone F, Gissi C, Liuni S, Pesole G: **Unstranslated regions of mRNA.** *Genome Biology Reviews 2002*, **3**:

[7] Bartel DP: **MicroRNAs: genomics, biogenesis, mechanisms and fuction.** Cell 2004, 116:281-297

[8] Wang Y, Liu CL, Storey JD, Tibshirani RJ, Herschlag D, Brown PO: **Precision and functional specificity in mRNA decay.** *PNAS 2002*, **99**:5860-5865

[9] Silva AL, Romão L: **The mammalian nonsence-mediated mRNA decay pathway: to decay or not to decay! Which players make the decision?.** *FEBS 2009*, **583**:499-505

[10] Pesole G, Migone F, Gissi C, Grillo G, Licciulli F, Liuni S: **Structural and functional features of eukaryotic mRNA unstranslated regions.** *Gene 2001*, **276**:73-81

[11] Wilkie GS, Dickson KS, Gray NK: **Regulation of mRNA translation by 5'- and 3'-UTR-binding factors.** *TRENDS Biochemical Sciences 2003*, **28**: 182-188

[12] Sachs MS, Geballe AP: **Downstream control of upstream open reading frames.** *Gene & Development 2006*, **20**:915-921

[13] Hughes TA: **Regulation of gene expression by alternative untranslated regions.** TRENDS in Genetics 2006, 22: 119-122

[14] Sumner CJ: **Therapeutics development for spinal muscular atrophy.** *NeuroRx 2006*, **3**:235-245.

[15] Heier CR, Gogliotti RG, DiDonato CJ: **SMN transcripts stability: could modulation of messenger RNA degradation provide a novel therapy for Spinal Muscular Atrophy?.** *J. Child. Neurol. 2007*, 22:1013-1018

[16] Wirth B, Brichta L, Hahnen E: **Spinal Muscular Atrophy: From gene to therapy.** *Elsevier – Seminars in Pediatric Neurology 2006*, 121-131

[17] Lefebvre S, Birglen L, Reboullet S, Clermont O, Burlet P, Viollet L, Benichou B, Cruaud C, Millasseau P, Zeviani M, Paslier DL, Frézal J, Cohen D, Weissenbach J, Munnich A, Melki J: **Identification and Characterization of a Spinal Muscular Atrophy-Determining Gene.** *Cell Press 1995*, **80**:155-165

[18] Burghes AHM, Beattie CE: **Spinal Muscular Atrophy: why the low levels of survival motor neuron protein make motor neurons sick?**. *Nature Reviews – Neuroscience 2009*, **10**:597-609

[19] Monani UR: **Spinal Muscular Atrophy: A deficiency in a Ubiquitous protein; a motor neuron-specific disease**. *Elsevier 2005*, **48**:885-896

[20] Wirth B: **Spinal Muscular Atrophy: state-of-the-art and therapeutic perspectives**. *ALS and other motor neuron disorders 2002*, **3**:87–95

[21] Pagliardini S, Giavazzi A, Setola V, Lizier C, Luca MD, DeBiasi S, Battaglia G: **Subcellular localization and axonal transport of the survival motor neuron (SMN) protein in the developing rat spinal cord**. *Hum Mol Genet 2000*, **9**:47-56

[22] McWhorter ML, Monani UR, Burghes AHM, Beattie CE: **Knockdown of the survival motor neuron (Smn) protein in zebrafish causes defects in motor axon outgrowth and pathfinding**. *The Journal of Cell Biology 2003*, **162**:919–931

[23] Sumner, CJ: **Molecular Mechanisms of Spinal Muscular Atrophy**. *J. Child. Neurol. 2007*, 22:979-989

[24] Cartegni L, Hastings ML, Calarco JA, de Stanchina E, Krainer AR: **Determinants of Exon 7 Splicing in the Spinal Muscular Atrophy Genes, *SMN1* and *SMN2***. *Am. J. Hum. Genet. 2006*;**78**:63–77

[25] Hofmann Y, Lorson CL, Stamm S, Androphy EJ, Wirth B: **Htra2-b1 stimulates an exonic splicing enhancer and can restore full-length SMN expression to survival motor neuron 2 (SMN2)**. *PNAS 2000*, **97**:9618–9623

[26] Monani UR, Sendtner M, CooRevt DD, Parsons DW, Andreassi C, Le TT, Jablonka S, Schrank B, Rossol W, Prior TW, Morris GE, Burghes AHM: **The human centromeric survival motor meuron gene (SMN2) rescues embryonic lethality in *Smn-/-* mice and results in a mouse with spinal muscular atrophy**. *Hum Mol Genet 2000*, **9**:333-339

[27] Le TT, Pham LT, Butchbach MER, Zhang HL, Monani UR, CooRevt DD, Gavrilina TO, Xing L, Bassell GJ, Burghes AHM: **SMND7, the major product of the centromeric survival motor neuron (SMN2) gene, extend survival in mice with spinal muscular atrophy and associates with full-length SMN**. *Hum Mol Genet 2005*, **14**:845–857

[28] Wood M, Yin H, McClorey G: **Modulating the expression of disease genes with RNA-based therapy**. *PLoS 2007*, **3**:845-854

[29] Mollet IG, Ben-Dov C, Felício-Silva D, Grosso AR, Eleutério P, Alves R, Staller R, Silva TS, Carmo-Fonseca M.: **Unconstrained mining of transcript data reveals increased alternative splicing complexity in the human transcriptome**. *Nucleic Acids Res. 2010*, 1-15

[30] Oberdoerffer S, Moita LF, Neems D, Freitas RP, Hacohen N, Rao A: **Regulation of CD45 alternative splicing by heterogeneous ribonucleoprotein, hnRNPLL.** *Science* 2008, **321**:686-691.

Appendix

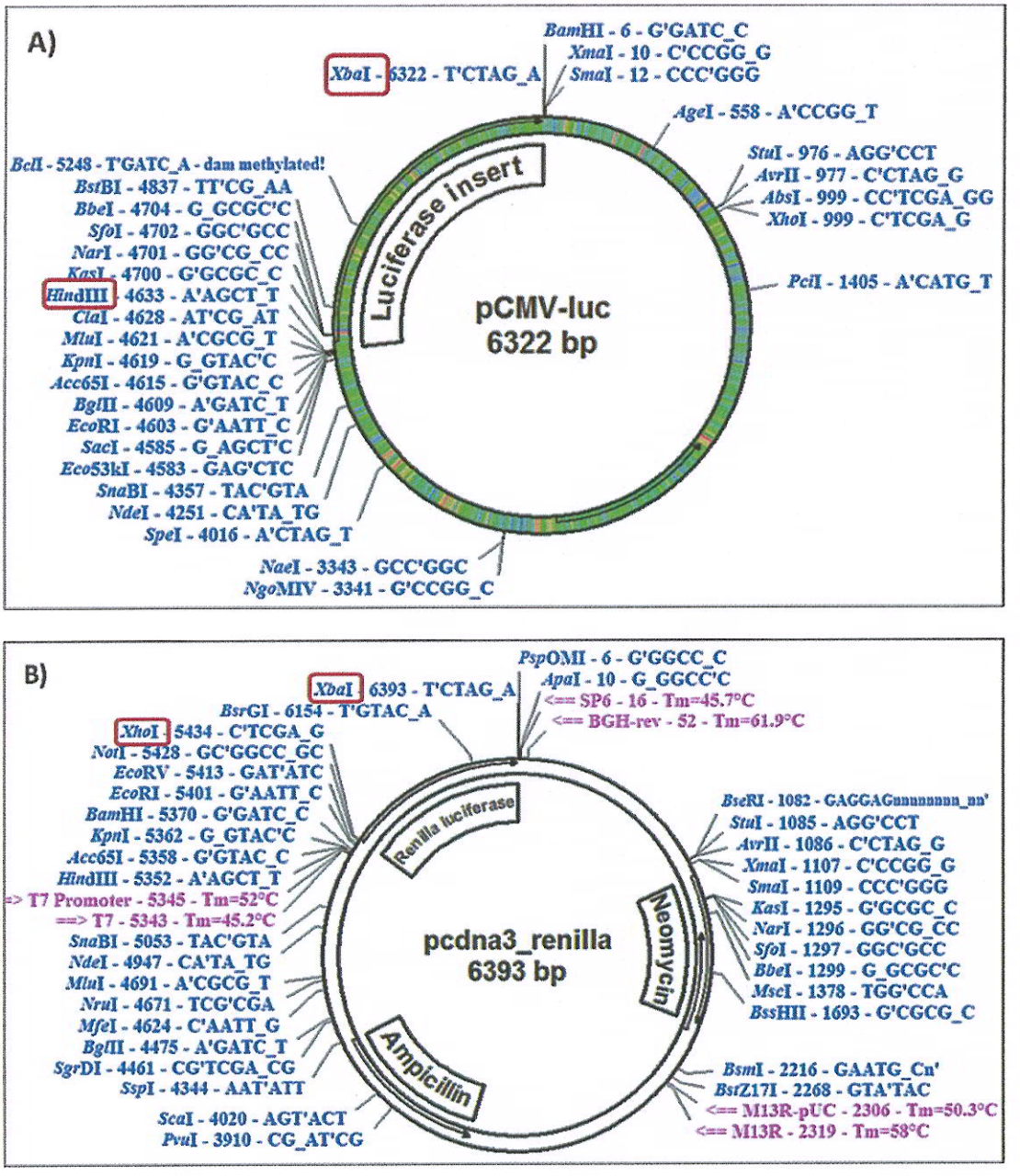


Fig. 1 – Reporter system used in the luminescence assays. The Firefly Luciferase reporter vector was used for studying the SMN UTR function and the Renilla Luciferase reporter vector was used to infer the basal expression in each cell population tested in each assay. **A) Firefly Luciferase expression vector.** The Firefly Luciferase CDS originally derived from the pGL3 vector was introduced into the pCMV5 eukaryotic expression vector MCS by ligation into HindIII/XbaI restriction sites, allowing for easy upstream and downstream cloning of the desired UTRs in frame with the reporter gene. **B) Renilla Luciferase expression vector.** The Renilla Luciferase CDS originally derived from the pRL-TK vector was introduced into the pCMV5 eukaryotic expression vector pCDN3 by ligation into HindIII/XbaI restriction sites.

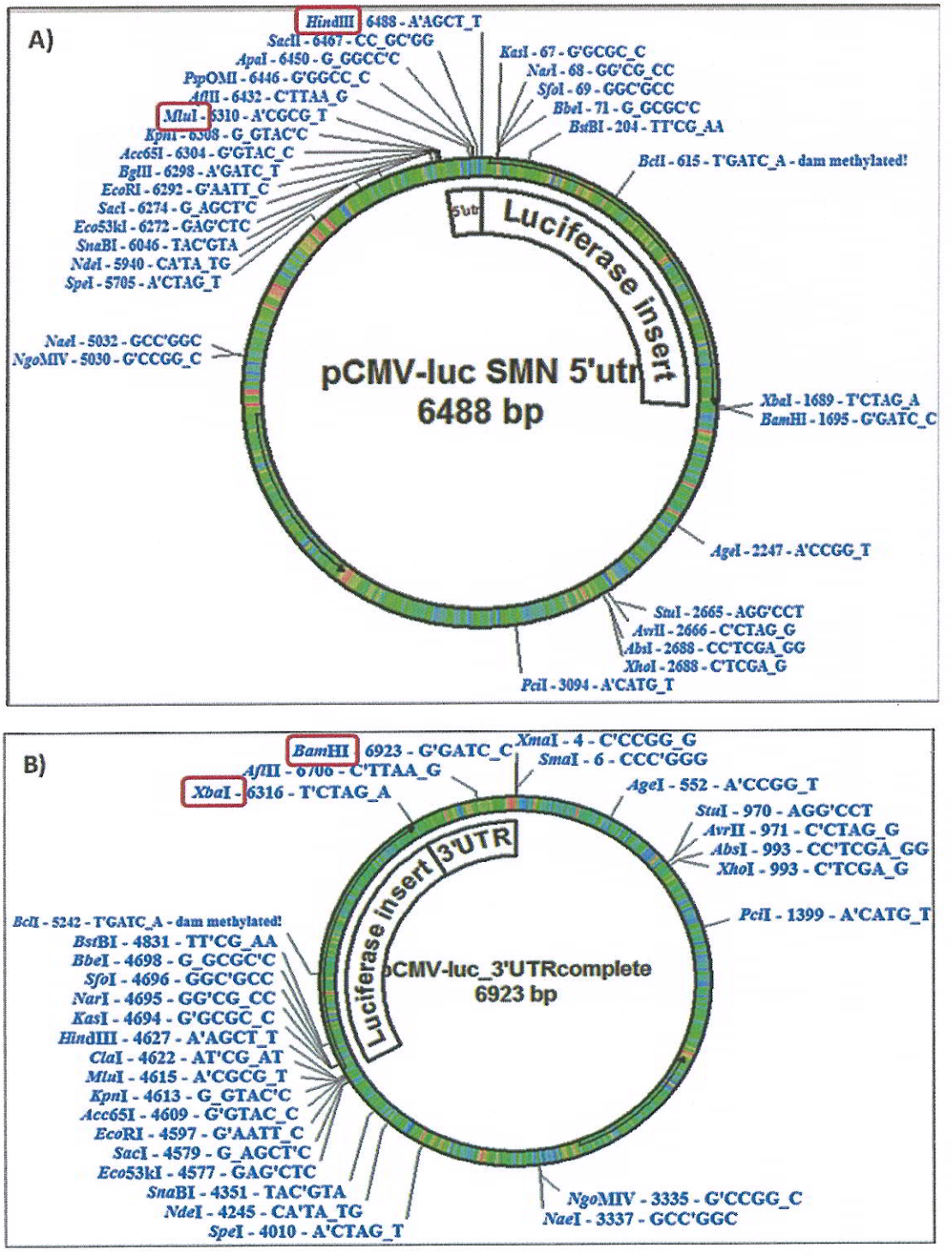


Fig. 2 – Luciferase-SMN UTR expression vectors. The SMN2 5' and 3'UTR sequences were obtained by RT-PCR from Type I SMA B Cells and SMA fibroblasts. However, whenever this amplification was not successful, these UTR sequences were obtained by PCR amplification on genomic DNA originated from HEK 293 cultured cells. **C) Luciferase-SMN 5'UTR expression vector.** 5'UTR sequence were introduced into pCMV5-Firefly Luciferase vector by ligation into MluI/HindIII restriction sites. **D) Luciferase-SMN 3'UTR expression vector.** 3'UTR sequence were introduced into pCMV5-Firefly Luciferase vector by ligation into XbaI/BamHI restriction sites.

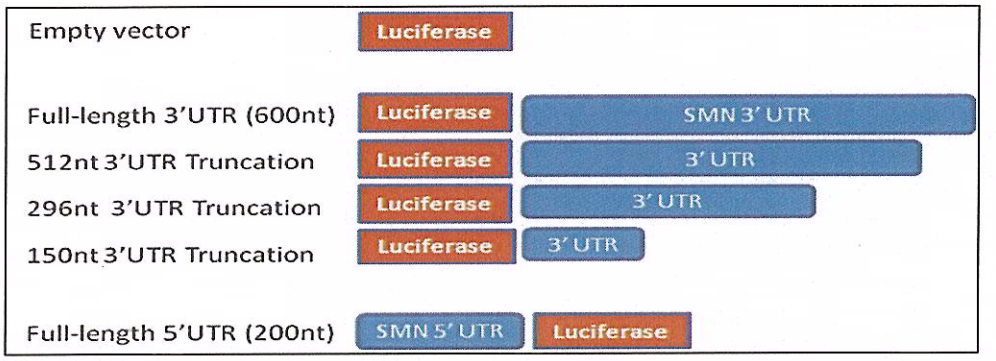


Fig. 3 – Panel of Luciferase-SMN UTR constructions. These constructions were used to evaluate the regulatory effects of the UTR regions through monitoring the Firefly Luciferase expression levels. After subcloning of the SMN2 UTR regions into the pCMV-Luciferase reporter vector, all constructions were confirmed by sequencing. The empty vector (construction with the reporter gene Luciferase only) serves as control for normalization of expression of the others vectors. A panel of Luciferase-SMN 3'UTR constructions were obtained due to mis-priming of the reverse primer at permissive hybridization temperatures, which originated a 512nt, 296nt and a 150nt truncation of the 3'UTR of the SMN2. The full-length SMN2 5' and 3'UTR were obtained by PCR amplification of HEK 293 cell line

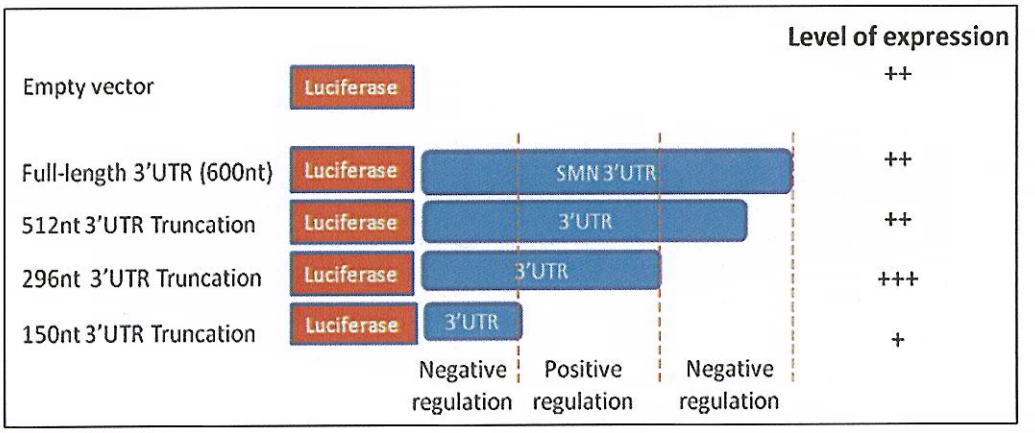


Fig. 4 – Identification of regulatory regions in the 3'UTR of the SMN2 mRNA. Downstream cloning of the full-length and truncated versions of the SMN 3'UTR in a Luciferase reporter vector, as represented, was used to infer the SMN-3'UTR dependet regulatory effects. Normalized luciferase expression levels determined as described before are qualitatively represented as + signals. This study identifies three major regions of regulation, with the terminal ends of the SMN 3'UTR having negative effects on gene expression, as indicated.

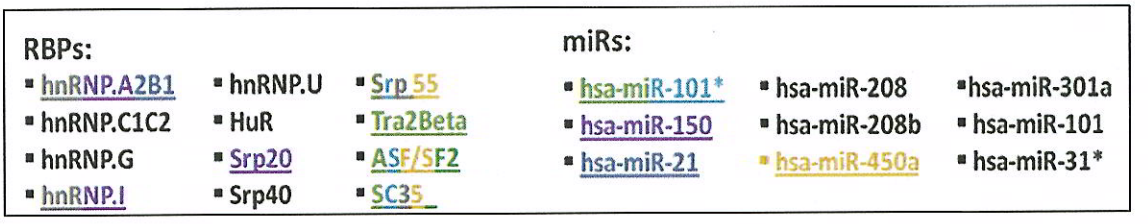


Fig. 5 – RNA Binding Proteins (RBPs) and microRNAs (miRs) predicted to bind to motifs in the SMN2 3'UTR sequence. The RBPs and miRs listed here were identified in bioinformatic analysis of the conserved nucleotides using the Splicing Rainbow and Target Scan algorithms (Gama-Carvalho, unpublished). Color codes highlight regulators that bind to mutagenized elements in Fig. 7.

Table 1 - List of Primers

Cloning Primers	
SMN 5'UTR Primer	5'ataacgcgtacaaatgtgggagggcgata 3' (Contains MluI site)
	5'tctaagcttctcatcgccatagcaaacc 3' (Contains HindIII site)
SMN 3'UTR Primer	5'atctctagaggagaaatgctggcatagag3' (Contains XbaI site)
	5'atcggatcccaatgaacagccatgtccac 3' (Contains BamHI site)
SMN 3'UTR genomic Primer	5'tcggatccaaatacaatcacagatacagct 3' (Contains BamHI site)

qPCR Primers	
Luciferase Primer	5'caactgcataaggctatgaagaga 3'
	5'atttgtattcagcccatatcgtt 3'
Renilla Primer	5'aacgcggcctcttcttattt 3'
	5'accagatttgcctgatttgc 3'

Mutagenesis Primers	
Mut uATG	5' ggtaccacgcgtacaattgtgggagggcgataac 3'
	5' gttatcgccctcccacaattgtacgcgtggtacc 3'
Mut 1	5' ctggaatgtgaagcgttatacatgataactggcctcatttcttc 3'
	5' gaagaaatgaggccagttatcatgtataacgcttcacattccag 3'
Mut 2	5'gaagcgttatagaagagatctggcctcatttcttc 3'
	5'gaagaaatgaggccagatctcttctataacgcttc 3'
Mut3	5'cgttatagaagataactggcatgatttgtacataatatcaagtgttgg 3'
	5'ccaacacttgatattatgtacaaatcatgccagttatcttctataacg 3'
Mut4	5'gataactggcctcatttgtacataatatcaagtgttggg 3'
	5' cccaacacttgatattatgtacaaatgaggccagttatc 3'
Mut5	5'ggataattattggtgatgttatgagctgtgagaagggtg 3'
	5'caccettctcacagctcataacatcaccaataattatcc 3'
Mut6	5'gtgaggcgtatgtggcataatcatacagaatctaactgg 3'
	5'ccagttagattctgtatgattatgccacatacgcctcac 3'
Mut 7	5'gtggcaaaatgttacagtatgtaactggtggacatg 3'
	5'catgtccaccagttacatactgtaacattttgccac 3'

Alternative UTR Primers	
5'UTR Alt Primer	5'tgcctctcagagaaagcaga 3'
	5'cccaaatgtcagaatcatcg 3'

3'UTR Alt Primer (V1)	5'accacctcccatatgtccag 3'
	5'ggtggtcgaatcacttgagg 3'
SMN Primer	5'ccttctggaccaccaataattc 3'
	5'ctgctctatgccagcatttct 3'
G6PD Primer	5'gaacgtgaagctccctgacg 3'
	5'tcaatctggtgcagcagtgg 3'

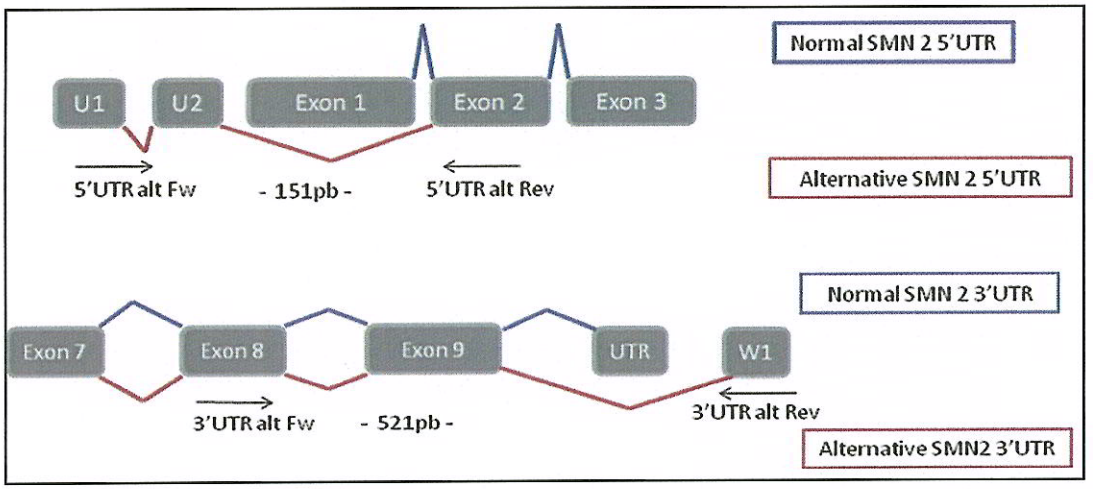


Fig. 6 – Representation of the alternative UTR sequences predicted by ExonMine Database and representation of the primers set used to confirm expression of these alternative isoforms. Alternative UTR sequences identified with the ExonMine Tool (http://www.imm.fm.ul.pt/exonmine/).