Open Research Online

The Open University's repository of research publications
and other research outputs

Characterisation of a Composite Exonic Regulatory Element of Splicing (CERES) in *CFTR* Exon 12: Functional Properties and Evolutionary Constraints

Thesis

oro.open.ac.uk

Characterization of a Composite Exonic Regulatory Element of Splicing (CERES) in *CFTR* exon 12: functional properties and evolutionary constraints

Ariful Haque

A Thesis Submitted in Fulfilment of the Requirements of the Open University, (UK)

for the Degree of Doctor of Philosophy

International Centre for Genetic Engineering and Biotechnology, ICGEB

Trieste, Italy

Director of Studies: Prof. Francisco E. Baralle, M.D., Ph.D.

External Supervisor: Prof. I. C. Eperon, PhD.

July 2010

DATE OF SUBMISSION : 21 MAY 2010

DATE OF AWARD : 26 JUL 2010

Dedication

To my family

Acknowledgements

TABLE OF CONTENTS

List of figures

sequences.

ABBREVIATIONS

The standard abbreviations used in this thesis follow IUPAC rules. All the abbreviations are defined also in the text when they are introduced for the first time.

aa	Amino acid
bp	Base pairs
cDNA	Complementary DNA
CERES	Composite Exonic Regulatory Element of Splicing
CFTR	Cystic Fibrosis Transmembrane conductance Regulator
CF	Cystic Fibrosis
DAZAP1	Deleted in Azospermia associated Protein 1
ddH_2O	Double-distilled water
DNA	Deoxyribonucleic acid
dNTPs	Deoxynucleoside triphosphate (A, C, G and T)
DTT	Dithiothreitol
EDTA	Ethylenediamine tetra-acetic acid
ESE	Exonic Splicing Enhancer
ESS	Exonic Splicing Silencer
hnRNP	Heterogenous ribonuclear protein
IPTG	Isopropyl-β-d-thiogalactopyranoside
ISE	Intronic Splicing Enhancer
ISS	Intronic Splicing Silencer
kb	Kilobase
kD	Kilodalton
N	Nucleotide (A or C or G or T)
NE	Nuclear Extract
NF1	Neurofibromatosis type I gene

NMD	Nonsense-mediated decay
nt	Nucleotides
PBS	Phosphate buffer saline
PBS	Phosphate buffer saline
PCR	Polymerase Chain Reaction
Pu	Purine (G or A)
Py	Pyrimidine (T or C)
RNA	Ribonucleic acid
RNA Pol II	RNA polymerase II
RRM	RNA Recognition Motif
RT	Room Temperature
snRNA	Small nuclear RNA
snRNP	Small nuclear ribonucleoprotein particles
SRE	Splicing Regulatory Elements
SR	Arginine-serine rich protein
ss	Splice site
SDS	N-lauroylsarcosine sodium salt
SELEX	Systematic Evolution of Ligands by Exponential Enrichment
SMN	Survival of Motor Neuron
TBE	Tris-borate-EDTA (buffer)
U2AF	U2 snRNP Auxiliary Factor

ABSTRACT

Genotype screening in human disease frequently results in the identification of exon/intron sequence variations whose direct connection with occurrence of disease is often unclear, especially if they occur within exons but do not alter the amino acid coding sequence. However, it is now clear that many of these seemingly harmless changes very often can exert flaws in the splicing process by affecting splicing regulatory elements (SREs). Usually, SREs are classified based on their mode of action with regards to exon inclusion, either positive (enhancer) or negative (silencer). In addition to these clear cut definitions, using systematic site directed mutagenesis experiments in previous analyses from our lab we have identified a novel type of splicing controlling element that we called CERES (for Composite Exonic Regulatory Element of Splicing). The distinguishing feature of CERES elements resides in the fact that they represent an extreme physical overlap of enhancer and silencer sequence. As a result, the functional effect at the level of exon inclusion/skipping of a single nucleotide change in a CERES element is hard to predict.

In this study I have addressed both issues in the context of the functional CERES2 element in *CFTR* exon 12. The result show that CERES2 can bind to a number of SR (SF2/ASF and SRp55) and hnRNP (A1, A2, C2, U, DAZAP1) factors in a small stretch of RNA in close proximity to each other. In particular, one of the disease causing mutations, G48C and a synonymous substitution next to it (A49G) showed reduced binding with SF2/ASF, whereas another natural mutation, A51T showed that the SF2/ASF interaction was increased compared to the wild type exon 12 sequences. Functional assays confirmed the potential regulatory role of the SF2/ASF and hnRNP A1 interactions.

Two synonymous mouse substitutions (T40C and C52T) near the CERES2 region were observed to cause skipping in human exon 12 but had no effect if the exon was truncated in a reduced context. Restoration of the truncated sequences restored skipping of

the exon. However, if these flanking sequences were replaced with mouse sequences then no skipping occurred. This observation suggested that the human exon 12 sequences have ESS regions in both flanks of the exon whereas in the mouse sequence the flanking exon sequences contain ESE elements. Affinity purification of these flanking sequence showed that both of the mouse flanking sequences bind to SR proteins (SF/ASF, SRp 75, SRp 55 and SRp 40) but not in human. The consequences of this situation were then checked at the evolutionary level by comparing the distribution of SREs in different species. Altogether, our results suggest that in several species other than human the entire sequence of CFTR exon 12 is involved in its definition.

1   Introduction

1.1   Overview of pre-mRNA splicing process.

Converting the DNA sequence into the amino acid sequence is a fundamental process in living organisms. Prior to that, DNA sequences have to be transcribed in pre-mRNAs, processed, and finally translated according to the coding sequence. Therefore, in order to generate a flawless mRNA, suitable to be transported into the cytoplasm and used for protein synthesis, the pre-mRNA processing is an essential step for gene expression.

In higher eukaryotes, the mRNA that enters translation is much shorter than the one transcribed. Apart from a few rarities, all protein-coding genes in higher eukaryotes have relatively short coding sequences interrupted by longer non-coding sequences. The term Intron (intergenic regions) and Exons (expressed regions) were first employed by Gilbert in 1978. After the transcription, the introns are censored out of the precursor-messenger RNA (pre-mRNA) by a process called splicing. Splicing reaction is an essential RNA processing event in eukaryotes. This process is structurally and functionally associated with the nucleus and occurs in a complex called the spliceosome (Maniatis and Reed, 2002).

The description of the event itself first came from two independent groups led by Philip Sharp and Richard Roberts respectively attempting to characterize the individual gene transcripts of adenovirus 2, in 1977. This finding came out once they incubated the hexon polypeptide mRNA (a very abundant component of the viral capside) along with the viral DNA and by electron microscopy analysis observed that the RNA sequence, which was assumed to have the same base sequence of the DNA from which it was transcribed, had not. In particular, they observed stretches of DNA sequences that were not part of the viral RNA. These sequences were interspaced between coding sequences therefore interrupting the code. On this basis, they suggested the term RNA splicing to describe this process of intron removal and subsequent joining together of the exons (Berget et al., 1977; Chow et al., 1977).

Historically, the late phase of adenovirus infection was the earliest model used to study the mRNA splicing. In this way the presence of introns was first described in the mRNA segment of adenovirus, coding for hexon polypeptide, the major virion structural protein (Berget et al., 1977; Chow et al., 1977). Subsequent to the characterization of RNA splicing process in viruses, the presence of introns was also reported as a general feature of eukaryotic genes. Jeffreys and Flavell in fact described the presence of a "large insert" in the coding sequence of rabbit beta-globin gene in 1977 (Jeffreys and Flavell, 1977). Soon after, Chambon and colleagues noticed that the chicken ovalbumin gene contains seven coding sequences (exons) are interrupted by six intervening sequences (introns) and identified the precise exon-intron junctions in the ovalbumin pre-mRNA (Breathnach et al., 1978). Moreover, they found that the sequences at exon-intron boundaries carry common features, probably with the function of unique excision-ligation common points to all boundaries (Breathnach et al., 1978). Interestingly, these consensus sequences were also present in vertebrate, plant and yeast suggesting that the splicing process is evolutionarily conserved (Padgett et al., 1986).

1.2 Chemistry behind the Splicing reaction.

Splicing reaction of the nascent RNA takes place within a large macromolecular complex approximately 60S in size and commonly referred to as the "spliceosome". In order to remove the intron, several small RNAs and a large number of proteins assemble within the spliceosome. However, the reaction itself is quite simple and involves a two-step trans-esterification reaction involving two sequential nucleophilic attacks on phosphodiester bonds at the splice junctions, the concomitant formation of the spliced mRNA, and release of the excised intron(Lamond, 1993; Moore and Sharp, 1993). Conserved motifs in the nucleotide sequences, at the intron-exon boundaries, act as essential splicing signals in the chemical part of the splicing reaction. A "GU" at the exon-intron junction defines the 5'splice site (5'ss) and an "AG" at the other intron-exon junction

together with the polypyrimidine tract and the 'A' nucleotide at the branch point identify the 3' splice site (3'ss).
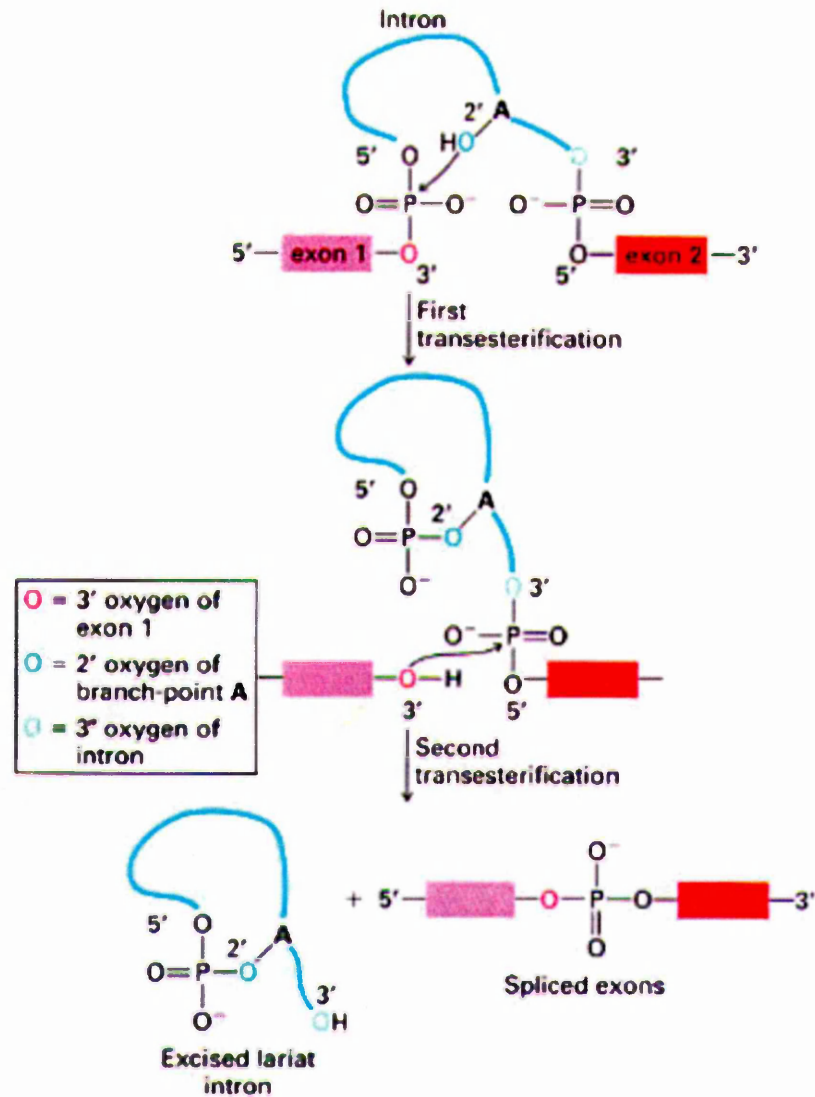
Figure1.1:

Schematic representation of the RNA splicing reaction. Pink and red boxes represent the exons. Light blue line represents the intron. **A** is the universally conserved adenine in the branch site. In the first reaction, the ester bond between the 5' phosphorous of the intron and the 3' oxygen of exon 1 is exchanged for an ester bond with the 2' oxygen (dark blue) of the branch-site A residue. In the second reaction, the ester bond between the 5' phosphorous of exon 2 and the 3' oxygen (light blue) of the intron is exchanged for an ester bond with the 3' oxygen of exon 1, releasing the intron as a lariat structure and joining the two exons. Curved arrows show where the activated hydroxyl oxygen reacts with phosphorous atoms. (Figure taken from the book Molecular cell biology (Lodish et al., 2000).

17

In the first step of primary RNA transcript splicing, the RNA forms a lariat structure, then the 2' –OH on the branch point A usually located 20-40 nucleotide upstream of the 3' splice site attacks the phosphoryl group of the G in the 5' splice site. This leads to the breaking of the phosphodiester bond between the 3' end of the exon and the 5' end of the intron and the formation of a new phosphodiester bond between the branch point A and the 5' end of the intron, which are the intermediates of the splicing reaction. More precisely, in the branch point the 5' terminal phosphate is esterified to the ribose 2'OH group and disbranches a free exon. The second trans-esterification step involves cleavage of the 3' splice site, ligation of the two exons and release of the intron (in the lariat form) (Figure1.1). A recent study has experimentally proved that under appropriate conditions both catalytic steps of the splicing reaction can be efficiently reversed (Tseng and Cheng, 2008).

1.3   The spliceosome.

The spliceosome is a dynamic complex that is composed of several small nuclear ribonucleoproteins (snRNPs) and a large number of auxiliary proteins or non-snRNP splicing factors (Jurica and Moore, 2003; Rappsilber et al., 2002). The UsnRNPs are known as snRNAs-U1, U2, U4, U5, and U6 and each of them functions as a small stable RNA bound by several proteins, plus numerous other less stably-associated splicing factors (Nielsen, 2003). Mass-spectrometry analysis of purified spliceosomes has detected hundreds of polypeptides and five uridine-rich small ribonucleoprotein particles (UsnRNPs) (U1, U2, U5, and U4/U6) each of which contains the corresponding snRNAs and a set of specific and common proteins (Staley and Guthrie, 1998; Will, 1997; Will and Luhrmann, 1997). The spliceosome acts through multiple RNA-RNA, RNA-protein and protein-protein interactions to precisely excise introns and join exons in the correct order ((Madhani and Guthrie, 1992; Nilsen, 1994). The spliceosomal snRNPs (U1, U2, U4/6 and U5) consists of one (U1, U2 and U5) or two (U4 and U6) snRNAs associated with a set of

seven Sm proteins (B/B', D1, D2, D3, E, F and G) and several specific factors (Will and Luhrmann, 2001). The U4 and U6 snRNPs are generally found as combined U4/U6 particle. The sm proteins bind to a conserved sequence called the "Sm site" which is present in all snRNAs except U6. The U6 snRNP rather contains a set of Sm-like (Lsm) proteins. Some of the proteins in the snRNPs may be directly involved in splicing whereas others may be essential in structure formation or just for assembly or interactions between the snRNP particles. The Survival Motor Neuron (SMN) protein acts as a key protein for the assembly of snRNPs (Pellizzoni, 2007).

In the spliceosome assembly process of higher eukaryotes, exons are usually assumed to be defined first because of the fact that the introns are large in size (Sterner et al., 1996). Spliceosomal assembly on exons starts with the recognition of a tripartite signal (5', 3' and branch point) recognition in an ATP independent manner, where the U1snRNP interacts with the 5', U2AF to the terminal AG of the 3' splice site as well as the polypyrimidine tract and finally the SF1 at the branch point (Berglund et al., 1997; Nelson and Green, 1989; Zamore and Green, 1989). This complex is often referred as E (Early) complex or, sometimes, the commitment complex. However, in brief, the initial signal of 5'ss definition from the U1 snRNA is mediated by RNA-RNA, as well as by protein-protein and protein-pre-mRNA interaction involving U1-70K and U1-C proteins (Will et al., 1996). Members of the SR protein family stabilize this interaction in higher eukaryotes. Another important event of the E complex, definition of the 3'ss, the AG at intron/exon junction together with the adjacent polypyrimydine tract is identified through interactions with the dimeric U2 Auxiliary Factor (U2AF). The subunit U2AF65 recognizes the polypyrimidine tract (Valcarcel et al., 1996a) and the AG dinucleotide at the 3'ss interacts with the U2AF35 subunit (Wu et al., 1999). The branch point, usually located 20-40 nucleotides (nt) upstream the 3'ss within the intron, is recognized by the branch point binding protein (BBP/SF1) (Berglund et al., 1997).
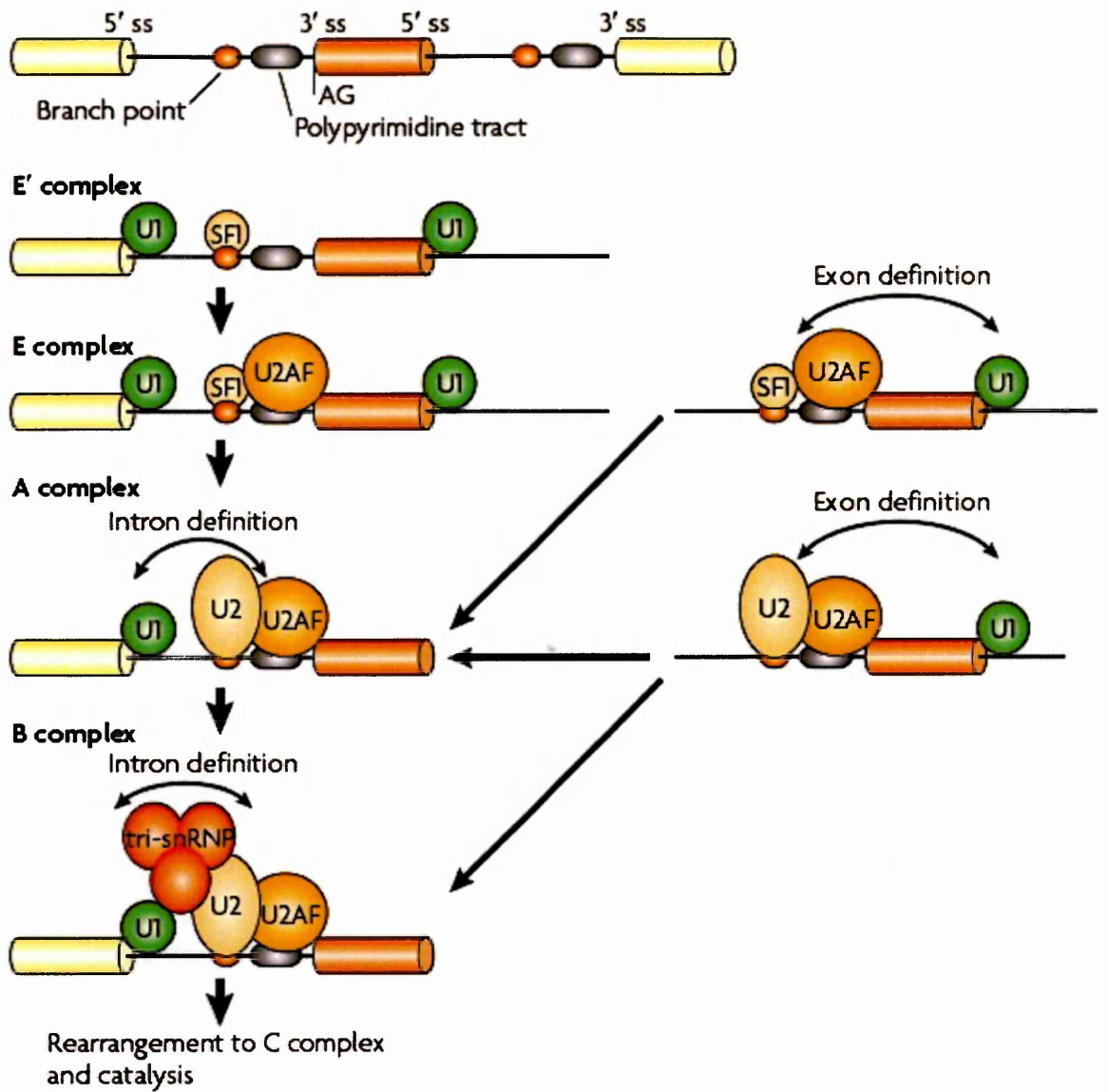
Figure1.2: Spliceosome assembly. The spliceosome assembles onto the pre-mRNA in a stepwise manner. The E complex contains U1 snRNP bound to the 5' splice site, SF1 bound to the branch point, and U2AF65 and U2AF35 bound to the pyrimidine tract and 3' splice site AG, respectively. In the A complex, SF1 is replaced by U2 snRNP at the branch point. The U4/U6/U5 tri-snRNP then enters to form the B complex. Once both exon and introns are defined then finally, a rearrangement occurs to form the catalytically active C complex, in which U2 and U6 interact, and U6 replaces U1 at the 5' splice site. Figure adopted from (Chen and Manley, 2009).

The formation of the A complex is characterized by the ATP-dependent recruitment of the U2 snRNP to the branch point sequence through replacement of the BBP/SF1 factor. This U2–branch site binding is mediated by U2 snRNA base pairing with the BPS which is

further stabilized through SF3a and SF3b subunits (Gozani et al., 1996) and also by the arginine-serine-rich domain of the U2AF65 (Valcarcel et al., 1996b). However, U2 snRNP is also identified as a component of a purified, functional E complex (Hong et al., 1997). The U2 snRNP seems to bind loosely to the pre-mRNA in the E complex via the integral U2-snRNP-associated protein SF3b, and then through an ATP-dependent process this leads to stable binding to the branch point, replacing SF1 (Das et al., 2000).Then, the ATP-dependent addition of U4/U6•U5 snRNPs in which the U4 and U6 snRNAs are base paired to the spliceosomal complex characterizes the transition from the A to B complex. However, recent studies reported that the tri-snRNPs are able to interact with the 5'ss and the upstream 5' exon at earlier step of spliceosome assembly (Maroney et al., 2000). Although B complex contains all of the snRNPs components required for splicing, it lacks a catalytic center. In order to activate the spliceosome, the complex B undergoes marked RNA-RNA rearrangements that involves the displacement of U1 by U6 snRNP via base pairing at the 5'ss through its highly conserved ACAGAG motif. At the initial step of the reaction the U4 snRNA interacts with the U6 snRNA tightly and more reluctantly with the U5 snRNA to form tri-snRNP. As the reaction proceeds further the tri-snRNP undergoes a wide structural changes, the U6 snRNA releases the U4 snRNA and interacts with the U2 snRNA. The U6 and U2 snRNA respectively binds with 5' ss and branch point and forms a lariat for the first catalytic step (Boehringer et al., 2004; Reed, 2000). All these rearrangements contribute to the fidelity of 5'ss recognition and support the contact with the branch point generating the activated B complex (Turner et al., 2004). The formation of B complex promotes towards catalytic step of splicing in order to generate the free 5' exon and the lariat-3' exon intermediates. This step is followed by the formation of the C complex, in which the second catalytic step of splicing reaction takes place. The U5 snRNA, together with the U2 and U6 snRNAs, is involved in aligning the exons for the second catalytic step through a highly conserved stem loop (O'Keefe et al., 1996). In addition the presence of a large highly conserved component of the U5 snRNP associated

protein, the Prp8 factor seems to stabilise these interactions. Prp8 had previously been shown to crosslink to both the 5'ss and 3'ss, as well as to the exons flanking these two splice sites (Umen and Guthrie, 1995). In addition, Collins and Guthrie have reported that a specific region of this factor affects a tertiary interaction between both the 5'ss and 3'ss and U6 snRNA (Collins and Guthrie, 1999).

1.4 Alternative splicing.

Alternative splicing is a major mechanism for gene regulation as well as proteomic diversity in all metazoan organisms (Maniatis and Tasic, 2002). Initially, in humans, splicing was thought to be only a minor processing pathway affecting about 5% of all genes (Sharp, 1994). At the present time, it has been estimated that about 95% of the human genes show alternative splicing and about 80% of these produce different protein sequences in a significant amount (Modrek and Lee, 2002; Wahl et al., 2009). In a typical multi-exon mRNA, because of alternative splicing decisions the splicing pattern can be altered in many ways, as the exons undergoing this process can be either spliced into the mature mRNA or skipped. Alternatively, introns that are normally excised can be retained in the mRNA or the position of either 5' or 3' splice sites can be shifted to make exons longer/shorter. A regulated exon that is sometimes included and sometimes excluded from the mRNA is usually referred to as a "cassette" exon. In some cases, multiple cassette exons are mutually exclusive producing mRNAs that always include one of few possible exon choices. All these individual patterns can be combined in a single transcription unit to produce a complex array of splice isoforms ((Black, 2003);(Smith and Valcarcel, 2000).

One of the most striking examples of alternative splicing complexity to this date is represented by the Drosophila Down Syndrome Cell Adhesion Molecule (DSCAM) that can potentially generate more than 38,000 different isoforms by alternative splicing (Schmucker et al., 2000). In humans, genomic analyses indicate that the brain has the highest frequency of alternative splicing, hinting at the correlation between alternative

splicing itself and complexity of brain function (Yeo et al., 2004). Interestingly, recent deep sequencing analysis shows that in brain alternative splicing event even differ within individuals (Wang et al., 2008).

The mechanisms that determine which splice sites are utilized and how this selection is regulated in different cell types or developmental stages have been heavily studied in recent years although it is clear that further studies will be required to fully understand these processes. In particular, much effort has been made in identifying the "combinatorial code" composed by *cis-acting* elements and trans-acting factors involved in the regulation of alternative splicing. However, high-throughput technologies like large-scale sequencing and microarrays analysis are providing opportunities to address also key questions regarding how this process is regulated at the global level (Ben-Dov et al., 2008).
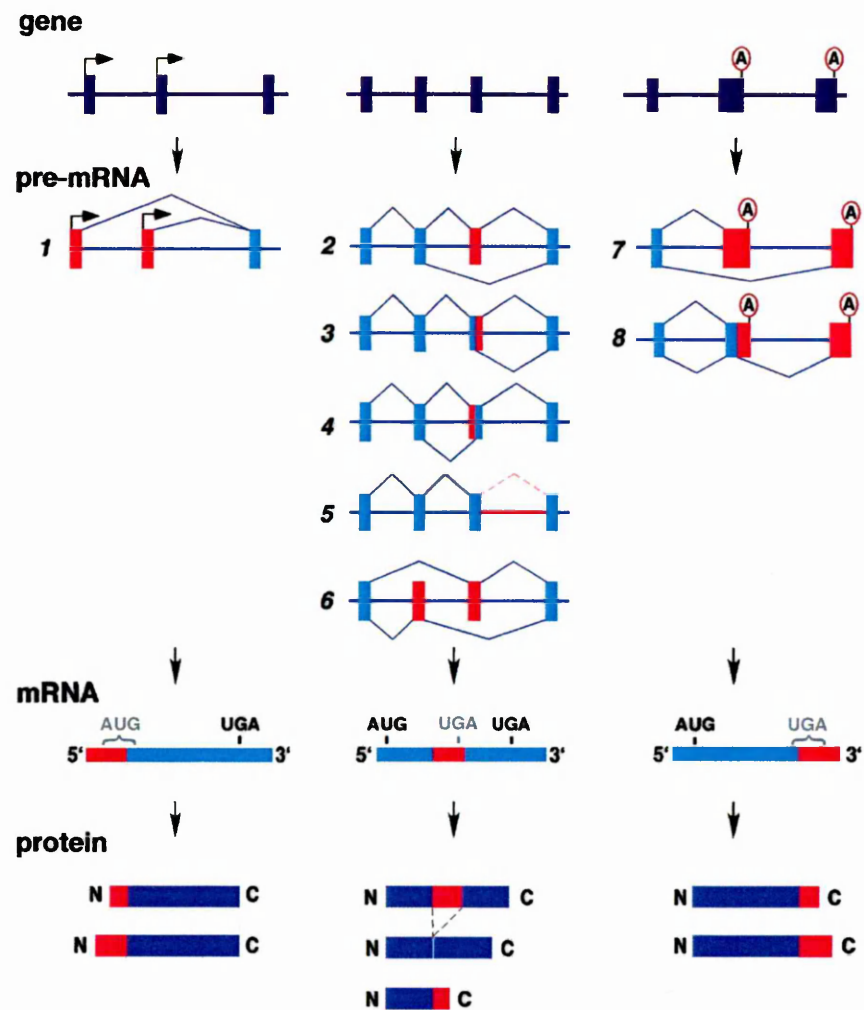
Figure1.3: Patterns of alternative splicing.

Alternative splicing generates different segments within mRNAs. Alternative promoters: selection of one of multiple first exons results in variability at the 5' end of the mRNA. Red indicates variable regions within the mRNA and encoded protein (*1*). Alternative splicing of internal exons: the alternative splicing patterns for internal exons include the cassette exon (*2*), alternative 5' splice sites (*3*), alternative 3' splice sites (*4*), intron retention (*5*), and mutually exclusive exons (*6*). Alternative terminal exons: selection of one of multiple terminal exons results from a competition between cleavage at the upstream poly(A) site or splicing to the downstream 3' splice site (*7*). There are also models of competition between a 5' splice site and a poly (A) site within an upstream terminal exon (*8*). Inconsistency at the 3' end of the mRNA produces either different C termini or mRNAs with different 3'-UTRs. Figure adapted from Faustino and Cooper (Faustino and Cooper, 2003).

24

1.5 Exon and intron architecture pose challenges to splicing.

Introns constitute a major fraction of the noncoding DNA, representing over 40% of mammalian genomes (Fedorova and Pyle, 2005). The architecture of exon and intron may be different across eukaryotic organisms. Short introns and long exons are typical of invertebrates while short exons and long introns are common in vertebrates. An standard vertebrate gene consists of multiple small exons spliced by introns that are 10 or 100 times longer (Hawkins, 1988). In all cases, however, the DNA of introns and exons represents an absolute minority in the genome composition. In humans, for example, exons account for less than the 4% of the genetic material per chromosome, whilst intronic sequences range from the 12% of chromosome Y to the 51.9% of chromosome 22. More than 90% of human genes contain multiple exons. The gene containing the highest number of exons is the TTN gene, coding for a sarcomeric structural protein named Titin, with 312 exons (Sakharkar et al., 2004).
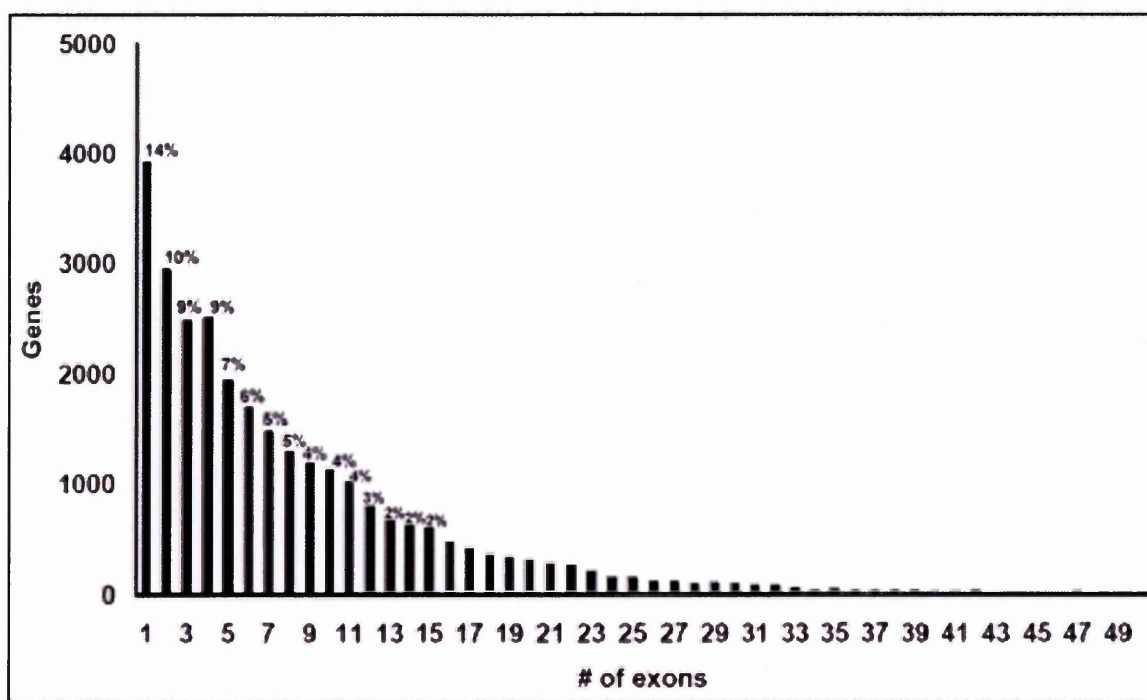
Figure1.4: Distribution of the number of exons per gene in the human genome.
The graph shows the number of genes (y) that contain a given number of exons (x). The number above the bars represents the percentage of genes having a given number of exons. (Figure taken from Sakharkar *et al.*(Sakharkar et al., 2005).

In humans, 80% of the exons are shorter than 200 nucleotides, the average size being 130 nucleotides. Exons longer than 300 or shorter than 50 seem not to be favoured by the splicing machinery (Sakharkar et al., 2004). A probable reason might be that very short exons are not efficiently recognized by the spliceosome due to physical obstruction between constitutive splicing factors whilst in long exons exon-definiton mechanisms may be excessively inefficient. In fact, in vertebrates, exons shorter than 50 nucleotides account for only 4% of the total exons (Dominski and Kole, 1991) and very rarely, and mostly in lower eukaryotes, exon length can exceed 600 base pairs (Hawkins, 1988).
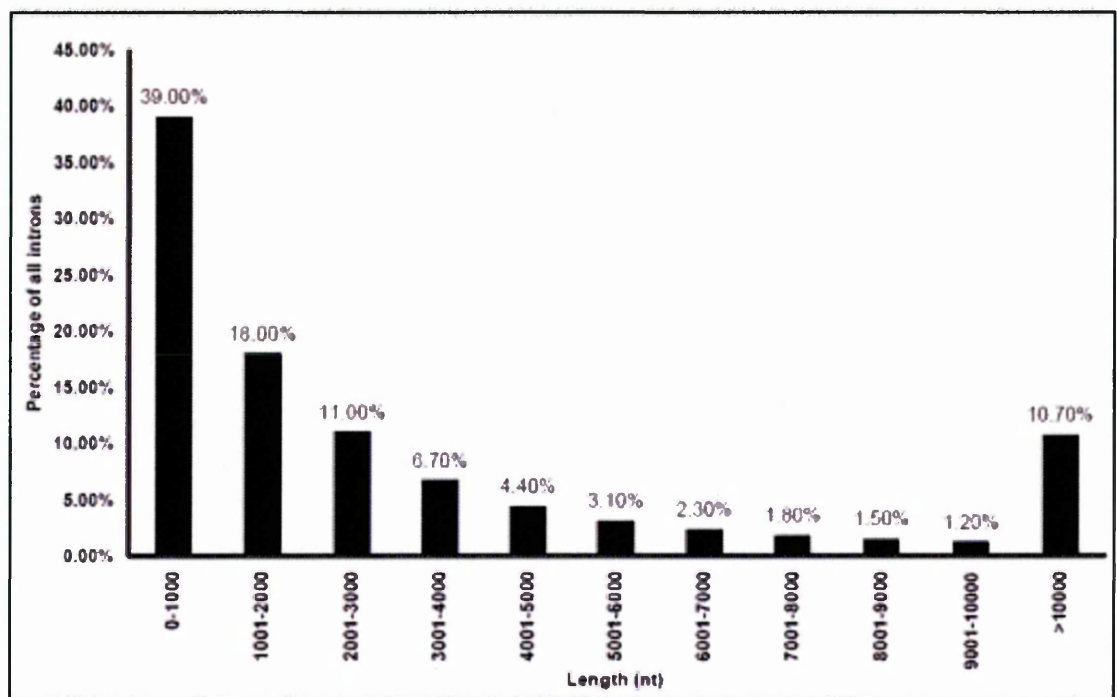
Figure1.5: Distribution of intron size in the human genome.
Almost 40% of human introns are shorter than 1 kilobase, and approximately 10% are longer than 10 kilobases. Just as a curiosity, the longest human intron spans almost 500 kilobases. Figure taken from Sakharkar *et al.*(Sakharkar et al., 2005).

The same does not hold true for introns whose length distribution is much wider. Introns can be as short as 20 bp (even if these short introns are less than 0.01% of the total) or can be very long, and more than 10% of them are longer than 11000 bp (Sakharkar et al., 2004). The enormous size of intron in humans and other vertebrates creates several drawbacks, such as: 1) considerable waste of energy during gene expression which is

26

"unwisely" spent on polymerizing extra-long intronic segments of pre-mRNA molecules; 2) delay in obtaining protein products (on average it takes about 45 min for RNA polymerase II to transcribe a 100,000 bp intron); 3) potential errors in normal splicing, since long introns contain numerous false splicing sites (so-called pseudo-exons) (Sun and Chasin, 2000). From the splicing point of view, the enormous size of the introns also pose a challenge for not only the correct recognition and pairing of the splice sites within a multitude of similar sequences but also in the positioning of these splice sites (which may lie tens of thousands of nucleotides apart) within the atomic distance that allows the trans-esterification reactions to proceed. Moreover, it is important to note that the sequences that match the 5'ss and 3'ss are very common within intron sequences and sometimes even stronger than the real splice sites. Given the huge variability in size, discrimination between exon and intron sequences is a complex task for the splicing machinery. Even more so since splicing is directed by the presence of mostly degenerate consensus sequences at the exon/intron and intron/exon junctions. In humans, an intron is generally characterized by mostly degenerate signals i.e. a GU di-nucleotide that marks the exon/intron junction at the 5' end of the intron (5' splice site or 5' ss). At the other end of the intron, the 3' splice site region comprises of three conserved elements: an A at the branch point site (BPS) located around 20-50 nucleotides upstream of the 3' ss, a poly-pyrimidine (pY) tract between the BPS and the 3'ss, and a terminal AG at the very 3' end of the intron (3'splice site or 3' ss).
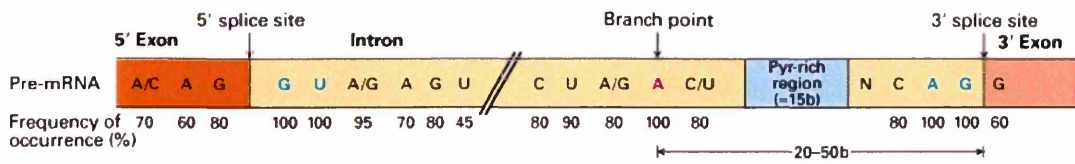
Figure1.6: Schematic representation of exon-intron boundaries.
The two exons are represented by the red and pink boxes respectively. Between them are reported the consensus sequences present within an intron. The arrows indicate the position of the 5' (GU) and 3' (AG) splice site and the branch point (A). The polypyrimidine tract, rich in pyrimidines, is highlighted by a blue rectangle. There are several conserved nucleotides near the sequences surrounding the intron-exon junctions that act as essential splicing signals. The frequency of each nucleotide in an alignment of conserved sequences from 1,683 human introns is indicated below. N means any base. To note only the universally conserved nucleotides are the dinucleotide cores of the 5' and 3' splice together with the branch point (A) showed 100% of frequency of occurrence. Figure adapted from Lodish (Lodish, 2000)

Although the above mentioned splicing signals at the boundaries are necessary, they are often also insufficient to define the correct splice sites due to their degeneracy (Fig. 1.6). Many matches to each consensus are in fact present along all pre-mRNAs but the vast majority of these sequences, known as pseudosplice sites are never selected for splicing. For example, a computer search for potential splice sites in the 42kb human *hprt* (hypoxantine phospho-ribose transferase) gene, composed of nine exons and eight introns, identifies the eight real 5' splice sites but also found over 100 5' pseudosplice sites many of which had scores higher than the lowest scoring real internal 5' splice site. The incapability of these few nucleotides to define splice sites was even more evident for the 3'ss where 683 pseudo-sites were found to have better scores than the worst scoring real site (Sun and Chasin, 2000).

Nonetheless, it is a fact that the splicing machinery is able to accurately recognise the real 5' and 3' splice sites and thereby neglect the 'pseudo' ones. It is now clear that the initial splice-site recognition across the exon is also the result of a combinatorial regulatory

mechanism (Smith and Valcarcel, 2000) that uses additional controlling elements beside the splice sites, which can act by increasing or decreasing exon recognition. These *cis*-regulatory elements are, respectively, named exonic or intronic splicing enhancer (ESE, ISE) or silencer (ESS, ISS).

1.6   *Cis* **Regulatory elements.**

As previously explained, discrimination between exon and intron sequences is a complex task for the splicing machinery. Several *cis-acting* elements, therefore, participate in this process (schematically depicted in Fig.1.7).
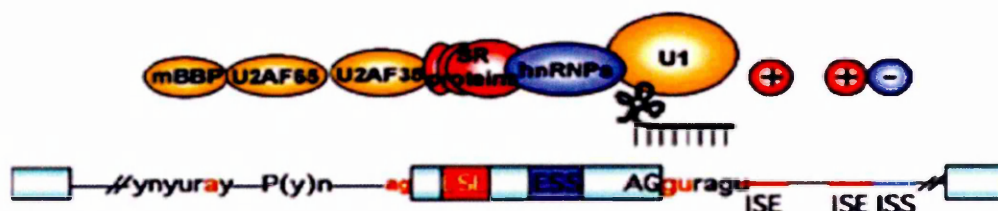
Figure1.7.: Regulatory elements in pre-mRNA splicing

The pale blue boxes correspond to exons, separated by intervening sequences (introns) shown as lines. Conserved, canonical splice signals GU//AG are present at the 5' and 3' ends of the exons respectively. These bind the U1 RNA by complementarity and the U2AF35 protein, respectively. If mutations are found in these areas, then the effect on splicing can be supposed to be aberrant. The effect of mutations on the other classical splicing signals upstream of the 3' splice site, the polypyrimidine tract and the branch point, is less certain, but many examples exist where these can cause incorrect splicing. The *trans-acting* factors that bind to these regions are U2AF35/65 and mBBP. Additional enhancer and silencer elements in the exons (ESE; ESS) and/or introns (ISE; ISS) allow the correct splice sites to be identified out of many cryptic splice sites that have identical signal sequences. *Trans-acting* splicing factors can bind with enhancers and silencers and can be subdivided into two major groups: members of the serine arginine (SR) family of proteins and heterogeneous nuclear ribonucleoproteins (hnRNPs). In general, but not exclusively, SR protein interaction at ESE facilitates exon definition whereas hnRNPs are inhibitory. Mutations in any of these sequences may have an effect on the splicing process due to disruption of the binding of these factors or indeed the creation of a binding site for them. Figure taken from (Baralle and Baralle, 2005).

1.7 The 5' splice site (5'ss).

Site directed mutagenesis suggests that the 5' splice site is moderately conserved over nine nucleotides, both upstream and downstream of the exon-intron junction (Hitomi et al., 1998; Wassarman and Steitz, 1992). This 5'ss motif in higher eukaryotes consists of nucleotides MAG/GURAGU (M indicates A or C, R indicates purines and the slash the exon-intron boundary), spanning from position -3 to +6 (Wieringa et al., 1983; Shapiro and Senapathy, 1987). In general, substitution of one of the first two bases (GU) or the last two (AG) of an intron completely abrogates splicing (Aebi et al., 1986). However, in the context of multiple introns and a complex environment, mutations at these positions can give different results: for example, a +1G→A mutation causes skipping of the associated exon, whilst a +2T→A mutation (generating a GA splice site) can lead to the accumulation of lariat-exon intermediate both *in vitro* and *in vivo* (Aebi et al., 1987). This suggests that +2 mutations are less detrimental than +1 mutation for the initial recognition of the splice site by U1 snRNP. The notion of splice site strength was first introduced and evaluated by Eperon and colleagues who designed a system for assessing 5' splice site strength in the β-globin gene. In their experiments, a Bam H1 site, 25 nucleotides upstream of a constitutively used 5' ss was used to insert oligos with various 5' splice sites that were tested for their ability to attract splicing away from the standard site (Eperon et al., 1986). Additionally, the choice of the 5' ss is also dependent on the factors that interact in the exonic sequence. For example, SF2/ASF can increase the recruitment of U1 snRNA to 5' ss, whereas the hnRNP A1 sterically blocks the occupancy of the U1 snRNA to the 5'ss (Eperon et al., 2000). Several studies have shown that the introduction of mutations that improve the match of weak splice sites to the consensus can lead to the constitutive recognition of alternatively spliced exons (Del Gatto et al., 1997; Huh and Hynes, 1993; Muro et al., 1998).

During E complex assembly, the 5'ss involves a nearly perfect Watson-Crick base pairing with the U1 snRNA (Horowitz and Krainer, 1994). Despite the absolute need for the base pairing between U1snRNA and the 5'ss for splicing, *in vitro* evidences of U1 independent U2 type splicing have been recently reported in intron 9 of human F1γ gene (Fukumura et al., 2009) and in NF-1 gene intron 29 (Raponi et al., 2009). Indeed, even in the absence of the 5' tail of U1snRNA, the U1snRNP particle is capable of selecting a 5'ss through the U1-C subunit. Moreover, in case of Human *FGFR* exon 10, the used splice site was found to be different from the splice site selected by U1snRNP. The Human FGFR exon 10 has two distinct splice sites with a distance of six nucleotides. One of the splice site provided the signal for presence of an exon by U1snRNP interaction but the used site was selected by U6snRNP, suggesting to the fact that U1snRNP interaction to the splice site is not always mandatory (Brackenridge et al., 2003). Finally, the observation that over-expression of SR proteins could compensate for the absence of U1snRNP *in vitro* (Tarn and Steitz, 1994); (Crispino et al., 1994) provides support to the view that the splicing process can be precise even in the absence of U1snRNP. Thus, the 5' splice site is not completely an U1snRNP dependent *cis* element, it can define the exon with the support of other factor involved in splicing (Zhang et al., 2008).

1.8 The 3' splice site (3'ss).

The 3' splice site definition comes from two distinct intronic sequence (the branch point and the polypyrimidine tract) as well as from a terminal AG (Reed, 1989). These elements on the whole contribute to 3' splice site recognition.

1.8.1 The branch site.

In lower eukaryotes, such as in yeast the branch point is highly conserved, which is UACUAAC. While the branch point in human introns is highly degenerated. Recent characterization of 20 house keeping genes in human shows that the branch point sequence is yUnAy (y= pyrimidine and n=any) and they are mapped 4-24 nucleotide downstream of

the poly pyrimidine tract (Gao et al., 2008). However, in the case of rat α-tropomyosin gene intron 2, the branch point is located 172 nt upstream from the 3'ss and nevertheless is essential for the regulation of alternative splicing (Smith and Nadal-Ginard, 1989).The proximity of this branch point to the 5' splice site of exon 2 causes a mutually exclusive regulation of exons 2 and 3, probably because the splicing factors are unable to bind productively to the two elements simultaneously and form active spliceosomes (Smith and Nadal-Ginard, 1989). The branch point is recognised by the SF1 factor during the early spliceosome assembly. Although the sequence specificity of SF1 is low, it can discriminate between sequences with general consensus branch site and sequences missing the highly conserved adenosine (Berglund et al., 1997). Possibly this is why the mutation in the branch point is not always deleterious. Once the "A " in the branch point is mutated it compensates with nearby "A" to promote splicing (Ruskin et al., 1985).

1.8.2 The polypyrimidine tract.

The polypyrimidine tract is a stretch of 5-20 pyrimidines, located between the branch site and the terminal AG at the intron/exon junction. The proximity of the polypyrimidine tract to the 3'ss is important when the pyrimidine length is limiting. Shortening the number of continuous uridines requires the localization of these uridines immediately adjacent to the 3'ss AG. Vice versa, a polypyrimidine tract containing high number of uridines is a competitive pyrimidine stretch regardless of the distance between the branch point and polypyrimidine tract itself (Coolidge et al., 1997). The polypyrimidine tract binds to several proteins, both of the constitutive spliceosomal machinery like the 65-kD subunit of U2AF (U2AF65) but also repressors of alternative splicing, for example the polypyrimidine tract binding protein (PTB) (Gooding et al., 1998);(Wagner and Garcia-Blanco, 2001). Normally, U2AF65 binds the polypyrimidine tract during the formation of the ATP independent early (E), or commitment, spliceosome complex (Kielkopf et al., 2004; Zamore et al., 1992). Conversely, the binding of PTB to the CU- rich elements in the polypyrimidine tract of a 3'ss can inhibit splicing by directly occluding the binding of

U2AF65 (Sauliere et al., 2006). Recently, PUF60 has been implicated in 3' splice site recognition that can function either cooperatively with U2AF65 or without it, and might help to recruit or stabilize U2snRNP assembly on to the pre-mRNA (Hastings et al., 2007).

1.8.3 *The terminal AG dinucleotide.*

The terminal AG dinucleotide defines the 3' border of the intron, just downstream to the polypyrimidine tract. This site is characterised by the short YAG/G sequence (Y denotes pyrimidines; the slash indicates the intron-exon boundary and the underlined nucleotides are conserved). Even if it is essential for the second step of splicing catalysis, no base-pairing interactions with snRNAs are involved in recognizing this sequence and during the early step of spliceosome assembly. The U2AF35 subunit recognizes this sites (Wu et al., 1999).

1.9 Auxiliary splicing signals: enhancers and silencers.

Classical thought of the mechanism was that the exon-intron boundary would be defined by 5' splice site, 3' splice site and a branch point. However, it soon became clear that the fidelity of the exon recognition in higher eukaryotic splicing is not only dependent on the tripartite signals. Indeed, these can provide may be half of the signals to define the authentic exon-intron junction (Lim and Hertel, 2001). Reason is the splicing signals are often in a suboptimal situation, lacking the perfect consensus for binding of the snRNPs. In fact, "false" sequences that match the consensus splice site signals as well as, or better than, natural splice sites are very common in introns. As a result, these sequences define a set of pseudo-exons that greatly outnumber genuine exons but are normally not included in the mature mRNAs (Sun and Chasin, 2000). Thus, in addition to the conventional splicing signals spanning exon-intron boundaries, higher eukaryotes are more prone to additional signal like Enhancer (Exonic Splicing Enhancers /ESEs and Intronic Splicing Enhancers /ISEs) or Silencer (Exonic Splicing Silencers/ESSs and Intronic Splicing Silencers /ISSs) (Black, 2003; Cartegni and Krainer, 2002; Ladd and Cooper, 2002). The nomenclature of

these auxiliary elements is depending on their location and their function. In spite of this, these elements are not always well defined and their functions may overlap. In fact, in some systems it may be more appropriate to talk about composite exonic regulatory elements of splicing (CERES) as described for *CFTR* exon 9 and 12 (Pagani et al., 2003a; Pagani et al., 2003b) (see below for additional details).

1.9.1 *Exonic splicing enhancer and silencers (ESEs/ESSs).*

Prior to incorporate the exon in the final transcript in a suboptimal splicing condition Exonic Splicing Enhancers (ESEs) are found to interact with specific *trans acting* factors. The majority of splicing enhancers are located within 100 nucleotides of the splice sites and they are not active when located further away (Tian and Maniatis, 1994). Using this proximity, the strength of the ESEs are measured, when a ESE has better functionality from a greater distance from the splice sites, is referred as strong ESE (Graveley et al., 1998a). The initial classification of ESEs was based on the type of nucleotides present in sequence cluster. The first ESE mapped was purine-rich, an alternate run of As and Gs, nonetheless a run of either Gs or As doesn't refer a ESEs (Marcucci et al., 2007; Tanaka et al., 1994). Through interactions with a subset of SR proteins, purine-rich ESEs recruit or strengthen the binding of basic splicing factors to suboptimal splice sites and stimulate spliceosome assembly (Graveley et al., 1998b); (Lavigueur et al., 1993); (Sun et al., 1993); (Zuo and Maniatis, 1996); (Roscigno and Garcia-Blanco, 1995). However, an exon sequence having one or more SR binding sites does not necessarily function as an ESE (Zheng et al., 1999); (Zheng et al., 1998) and some of the exonic splicing suppressors (ESSs) also bind SR proteins (Mayeda et al., 1999); (Zheng et al., 1998).

The other class of ESEs is the non-purine-rich ESE. This class comprises the exonic AC-rich enhancer and exonic pyrimidine-rich enhancer. The AC-rich enhancers were first identified by *in vivo* selection experiments and were found to stimulate splicing both *in vivo* and *in vitro* (Coulter et al., 1997). AC-rich ESEs have been shown to be involved in the regulated splicing of both viral and cellular genes (Coulter et al., 1997); (Gersappe and

Pintel, 1999); (Zheng et al., 2000). Recent studies have shown that the AC-rich ESEs interact with a cold-shock cellular protein, Y box protein 1 (YB-1), and function in a way similar to that of the purine-rich ESEs (Stickeler et al., 2001). In addition to these classes, exonic pyrimidine-rich enhancers have been described in β-globin RNA (Schaal and Maniatis, 1999b) and other mammalian RNAs (Dirksen et al., 2003); (Staffa et al., 1997). Exonic splicing enhancers (ESEs) were identified and extensively studied as regulators of alternative splicing (Black, 2003) but they have also been implicated in some constitutive splicing events (Lavigueur et al., 1993; Schaal and Maniatis, 1999a). ESEs, through SR proteins binding, drive the exon definition by recruiting splicing factors and/or by antagonizing the action of nearby splicing silencer elements (Cartegni et al., 2002).

Therefore, there is not a single consensus sequence that can describe all the ESEs, making their identification difficult through sequence comparison or even by their interacting factors.

In addition to sequences that promote exon inclusion, there are sequences that inhibit splicing called exonic or intronic splicing silencers. The silencers are less well characterized: they can be purine or pyrimidine-rich and bind a diverse array of proteins (Fairbrother and Chasin, 2000). Recent studies have suggested exon splicing silencers to have a fundamental role in preventing pseudoexon inclusion in mature transcripts (Sironi et al., 2004). Furthermore, a specific subset of ESSs were also suggested to have distinct effects on the regulation of intron retention events in alternative splicing (Wang et al., 2004). In general, splicing silencers also mediate exon skipping by binding to *trans*-acting factors that interfere with spliceosome activity mostly belonging to the hnRNP family (Fairbrother and Chasin, 2000). A well-known factor that inhibits splicing is PTB (polypyrimidine tract binding protein). It recognizes pyrimidine-rich elements both in introns and in exons and can function either by antagonizing U2AF65 action or by creating a region of silencing across the down regulated exon (Wagner et al., 1999).

1.9.2 Intronic enhancer and silencers (ISE/ISS).

Along with ESE and ESS sequences a number of intronic splicing enhancers (ISE) and silencers (ISS) are also known. Fewer large-scale screens have been conducted for intronic elements and many more intronic elements are expected to be identified in future studies. One of the best characterized is represented by G triplets (GGG) or G runs ($G_{n\geq3}$), that acts as ISE elements to enhance recognition of adjacent 5' splice sites or 3' splice sites (McCullough and Berget, 1997). For example, G runs in *THPO* genes shows that the intronic G motifs are involved with 3' splice site definition by a combinatorial effect (Marcucci et al., 2006). In this model, precisely the G7 and G10 motifs present in the intron 2 collectively facilities the recognition of proper 3'ss by interacting with hnRNP H, whereas deletion of these two G motifs activates the cryptic 3' splice site.

The most studied intronic enhancer proteins are Fox-1 and Fox-2 that act at UGCAUG motifs of the brain-enriched exons (Fagnani et al., 2007; Minovitsky et al., 2005). Fox-1 proteins regulated splicing by antagonizing the repressive effect of hnRNP proteins or by regulation of the pre spliceosomal complex formation (Zhou and Lou, 2008).

1.9.3 CERES.

The presence of the previously mentioned *cis* elements (enhancer and silencer) can certainly explain most of the impressive flexibility that is widely displayed by the splicing system. Mostly these pure ESEs/ESSs are defined by protein based score matrix and followed by *in silico* prediction or in an *in vitro* system, which hardly reflect the original cell environment. Moreover, the pure ESEs/ESSs most of the time behave pretty much like the original context even if it is in a heterologous context. As a result, it is very hard to explain the effect of overlapping ESEs/ESSs sequences on splicing regulation from a classical point of view of ESEs/ESSs. Accordingly, this new types of regulatory sequences have been renamed with the acronym of CERES (Composite Exonic Regulatory Element of Splicing). In literature, nomenclature CERES has been used so far in human *CFTR* exon 9

and 12 (Pagani et al., 2003a; Pagani et al., 2003c) and exon 10 of Luteinizing hormone receptor type 1 and 2 (Gromoll et al., 2007). However, this kind of dynamic splicing regulatory sequences or similar ones are also found in many other exons. Site directed mutations in CERES elements show that they are like an overlapping enhancer and silencer, rather than individual ESEs or ESSs. These elements are also not predictable in computer-assisted systems. Moreover functionality of these elements is not reproducible in a heterologous context. However, a little has been explored so far to understand the molecular mechanism behind the versatile nature of CERES.

1.10 Proteins involved in splicing: *trans-acting* factors.

The *trans-acting* factors that regulate alternative splicing are principally members of two protein families: Serine/Arginine-rich proteins family (SR proteins) and heterogeneous ribonuclear proteins (hnRNP).

1.10.1 *SR proteins.*

Serine/Arginine-rich proteins (SR proteins) are structurally similar RNA binding protein, highly conserved in metazoan cells (Blencowe, 2000; Graveley, 2000; Graveley et al., 1999; Huang and Steitz, 2005). The term SR proteins refers to at least ten major polypeptides ranging from 20 to 70 kD in size that share the ability to modulate the splice site choice. These proteins have a common modular structure with one or more RNA binding domains (RRMs) at the N-terminus, that provides target specificity, and a domain rich in Arg-Ser dipeptides at the C-terminus (RS domain) (Birney et al., 1993), usually involved in protein-protein interactions. However, the RS domain can also interact with the RNA. Recent evidence shows that the branch point and the RS domain interacts specifically to promote splicing (Shen et al., 2004). SR proteins have been implicated in many steps of spliceosome assembly, from recognition of the 5' splice site at the earliest step of splicing (Kohtz et al., 1994; Zuo and Manley, 1994) to binding and regulation of exonic-enhancer sequences that stimulate the usage of sub optimal splice sites. SR proteins differ for the

presence or absence of the second RNP domain and for the length of the RS domain; when present, the sequence of the second RNP domain is often divergent from the canonical consensus sequence (Fu, 1995; Manley and Tacke, 1996).

SR proteins are generally referred as a positive factor, which facilitates the exon definition. For example, once the SR protein binds to the ESE, they recruit U1snRNP for 5'ss and U2AF complex and U2snRNP to the 3'ss by interaction through RS domain (Bourgeois et al., 1999; Eperon et al., 2000; Feng et al., 2008; Graveley et al., 2001). In the regulation of some exons the two modes of action may co-exist (Cartegni et al., 2002). However, similar recruitment through SR protein has also been seen through intronic interaction. For example, TIA-1 (T-cell restricted Intracellular antigen-1), a SR family protein, binds to the downstream of the 5'ss and recruits U1snRNP (Forch et al., 2002). Similarly, Sam68 has been found in the case of CD44 V5 exon engaging U2AF to the 3'ss from its intronic position (Tisserant and Konig, 2008). SR proteins can also exert their effect by interacting with other RS domain containing proteins. In general, SR or SR related proteins contain the RS domain but can be with or without the RRM. For example, SRm160 and SRm300 have the RS domains but lack the RMM. Therefore in order to regulate splicing, these proteins have to interact with other RNA bound RS domain containing proteins (Blencowe, 2000). Similarly, for the SR proteins with the RRM domain, like Tra2 which interacts directly to the RNA, their RS domain interacts with other SR protein RS domains to stabilize binding of basic splicing factors and antagonize the negative elements nearby (Tacke and Manley, 1999).

Generally, SR proteins are considered as a positive factor, yet SR proteins act as splicing suppressor depending on their phosphorylation status. For example, SRp 38 restrains splicing in the dephosphorylated form (Shin et al., 2004), whereas it can act as splicing activator once is phosphorylated (Feng et al., 2008). In fact, the activity of SR proteins is regulated through phosphorylation/dephosphorylation cycles. SR proteins are phosphorylated *in vivo* at multiple serine residues within the RS domain (Stamm, 2008).

Phosphorylation and dephosphorylation processes seem to be crucial to define the organization of splicing factors inside the cell nucleus by affecting the RNA-binding activity and sub nuclear localization of RS-domain containing proteins (Misteli et al., 1997). Phosphorylation is important for specific RNA recognition, since the high positive charge of unphosphorylated RS domain masks the specificity of the RNP domains and enhances non-specific binding (Tacke et al., 1997; Xiao and Manley, 1997). Protein-protein interactions are enhanced by phosphorylation (Xiao and Manley, 1997), which also affects the subnuclear localization of splicing factors, causing their release from storage site of splicing components (Colwill et al., 1996; Xiao and Manley, 1997).

The interactions between pre-mRNA sequences and SR proteins seem to regulate constitutive as well as alternative splicing (Sanford et al., 2005; Tacke et al., 1997). The SR proteins have flexibility regarding the RNA targets and their binding specificities and consequent differential affinity may play a role both in constitutive and alternative splicing. From and evolutionary point of view, the functional interaction between SR proteins and exons place a significant constraint on the type of RNA sequences that can be present in coding exons, because codons and SR proteins binding sites overlap. It has been suggested that the degenerate binding specificity of SR proteins have evolved to cope this functional conflict (Graveley, 2000).
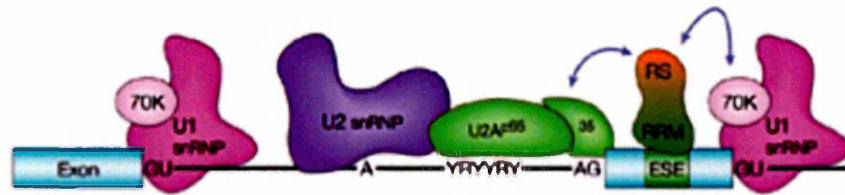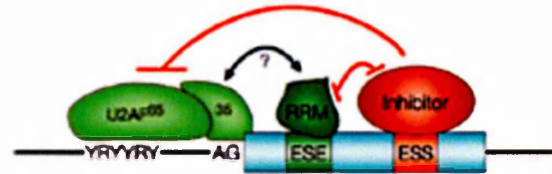
Figure1.8: Models of SR protein action in exonic-splicing-enhancer-dependent splicing.

a) RS-domain-dependent mechanism. An SR protein binds to an exonic splicing enhancer (ESE) through its RNA-recognition motifs (RRM) and contacts the splicing factor U2AF35 and/or U1-70K at the adjacent splice sites through its RS domain. $U2AF^{65}$ interacts with the polypyrimidine (Y) tract, which here is interrupted by purines (R) and is therefore part of a weak 3' splice site. U2AF65 also encourages binding of U2 snRNP to the branch site. The U1 snRNP particle binds to the upstream and downstream 5' splice sites through base paring of the U1 snRNA. The three sets of splicing-factor–pre-mRNA interactions (U2AF–3' splice site, U1 snRNP–5' splice site and SR protein–ESE) are strengthened by the protein–protein interactions (blue arrows) that are mediated by the RS domain. b) RS-domain-independent mechanism. Here, the main role of the SR protein that is bound to an ESE is to antagonize the negative effect on splicing of an inhibitory protein that is bound to a juxtaposed exonic splicing silencer (ESS). The SR protein is shown lacking its RS domain, although this domain is generally present and might still promote U2AF binding, or other domains might be involved in protein–protein interactions. Inhibitory interactions are shown (red), as is a putative stimulatory binding (double-headed arrow). These models are not mutually exclusive, and the splicing of some introns might involve a combination of these mechanisms. Figure adapted from Cartegni et al. (Cartegni L, 2002).

1.10.1.1   SF2/ASF.

One of the best-known SR proteins is the alternative splicing factor/splicing factor 2 (SF2/ASF). This 27kD protein consists of a RS domain and two RRM domains. However, several groups have highlighted the two basic properties of this SR protein. SF2/ASF was described as an essential splicing factor necessary for the early step of splicing (Krainer et al., 1990) and was also characterized as an alternative splicing factor able to drive splice site selection (Ge and Manley, 1990). ASF/SF2 can promote the recruitment of U1 snRNP to the 5'ss (Eperon et al., 2000; Kohtz et al., 1994) to help 5'ss and 3'ss bridging across introns (Wu and Maniatis, 1993), and plays a role in splicing regulation, through binding to exonic splicing enhancers (Sun et al., 1993). The function of SF2/ASF in pre-mRNA splicing depends on the context of the pre-mRNA sequence to which it binds, as shown by the fact that SF2/ASF inhibits adenovirus IIIa pre-mRNA splicing by providing sterical block to the U2snRNP recruitment at the BPS when bound to or near it at an intronic repressor element (Kanopka et al., 1996).

ASF/SF2, together with other SR proteins, is involved in additional roles in gene expression. For example, ASF/SF2 remains associated with the spliced mRNA and is able to shuttle between the nucleus and the cytoplasm (Caceres et al., 1998), suggesting a role in mRNA export (Huang and Steitz, 2005). In addition ASF/SF2 seems to regulate the mRNA stability by binding to the 3'UTR and enhancing RNA degradation in the cytoplasm (Lemaire et al., 2002). ASF/SF2 can also stimulate translation of reporter mRNAs by associating with translating ribosomes (Sanford et al., 2004).

Despite these advances in understanding the functions of ASF/SF2 less is known about the physiological roles of this protein. Depletion of ASF/SF2 by RNAi resulted in lethality in *C. elegans* (Longman et al., 2000) and tissue-specific deletion in mice resulted in defects in the developing heart (Xu et al., 2005). ASF/SF2 showed also an unexpected role in maintaining genomic stability by protecting cells from the deleterious effects of R-loop formation (Li and Manley, 2005). In addition a recent work found that ASF/SF2 is an

41

oncoprotein with roles in both the establishment and the maintenance of cell transformation (Karni et al., 2007). In this respect, SF2/ASF has been found to control alternative splicing of the oncogene *Ron* which modulates cell motility (Ghigna et al., 2005). Nonetheless, posttranslational modification (PTM) like arginine methylation of the SF2/ASF RS domain (R93, R97 and R109) has been shown have effect on it's subcellular localization (Sinha et al., 2010). Additive effect of this methylation provides a significant control to SF2/ASF on cellular localization as well as on multiple functions in different cellular compartments. In addition to this SF2/ASF has been also found in the splicing independent pathway of miR7 microRNA processing(Wu et al., 2010).

1.10.2 hnRNP proteins.

The hnRNP proteins were first described as a major group of chromatin-associated RNA-binding proteins. Initially, nearly 30 proteins of this class were identified by two-dimensional gel electrophoresis of human hnRNP complexes with molecular weight ranging from 34 (hnRNP A1) to 120 kD (hnRNP U) (Dreyfuss et al., 2002; Dreyfuss et al., 1993). The structure of hnRNP proteins is modular and consists of one or more RNA binding domains associated with an auxiliary domain often involved in protein-protein interactions (Dreyfuss et al., 1993). For instance, the hnRNP A/B proteins contain two RNP domains at the N-terminus and a Gly-rich auxiliary domain at the carboxy end whilst hnRNP E1-E2 proteins contain three KH domains (Ostareck-Lederer et al., 1998). Likewise, the hnRNP H family members contain two (2H9) or three (H, H' and F) quasi RNA recognition motifs (qRRMs) and one or two glycine rich auxiliary domains (Honore et al., 1995). In addition, several of these proteins have multiple isoforms produced by alternative splicing processes and this diversity can be further increased by post-translational modifications of potential physiological significance, including phosphorylation, arginine methylation and SUMOylation (Dreyfuss et al., 2002; Martinez-Contreras et al., 2007). Finally, their abundance can also vary considerably within cells, with some hnRNPs being highly expressed, whilst others are present in lower amount

depending on different tissues (Kamma et al., 1995; Markovtsov et al., 2000).

In general, the proposed mode of action for hnRNP proteins in splicing regulation is through competition with SR proteins. In brief, hnRNPs block the possession of the snRNPs or other positive factors to the splice site or ESEs. For example, polypyrimidine track binding protein (PTB or hnRNP I) can antagonize U2AF65 activity in the α-tropomyosin and the GABA(A) receptor γ2 genes (Lin and Patton 1995; Ashiya and Grabowski 1997). Another example is hnRNPA1 binding to the upstream silencer sterically blocks the U2snRNP interaction in HIV *Tat* exon 3 and interferes with the branch point selection (Tange et al., 2001). However, apart from this local competing mode, hnRNPs can also function from a distance. For example, hnRNP H can bind on both side of an exonic sequence and "loop it out" from the splicing queue (Chabot et al., 1997). In this way, although the splice site might be defined the presence of the loop sterically blocks any further spliceosomal complex formation (Nasim et al., 2002).

From an RNA binding point of view, hnRNP proteins usually do not recognize specific sites exclusively but distinguish different RNAs with a wide spectrum of affinities. Preferred sequences tend nevertheless to coincide with sites of functional significance in pre-mRNA processing, suggestive of that hnRNP proteins may form specialized complexes or indirectly recruit other factors to such sites. In addition, RNA binding is further modulated by mutual protein-protein interactions but even so the array of hnRNP proteins bound to a given hnRNA is determined by the RNA sequence (Dreyfuss et al., 1993).

1.11 Combinatorial and Position effect of the *trans acting* factors.

Whether an exon is included or not it is often determined by the combinatorial effect of the bound positive and negative factors. This combinatorial model is supported by the observation that *in vivo* and *in vitro* the counteracting activities of multiple antagonistic factors (generally SRs and hnRNPs) can regulate alternative splicing, suggesting that the physiological concentration of competing splicing factors is important for regulation of

splice site selection (Caceres et al., 1994; Hanamura et al., 1998). For example, in α-tropomyosin exon 2, SR protein 9G8 and hnRNP H and hnRNP F compete to bind same element within the exon for definition (Zhu and Krainer, 2000). Similar kind of arrangement has been shown within hnRNP A1, SC35 and SF2/ASF in β-tropomyosin exon 6B (Expert-Bezancon et al., 2004). Initially, this combinatorial effects of the *trans-acting* factor was thought to be prevalent for most exons but a recent microarray analysis has suggested that antagonizing effect of the SR and hnRNPs on the same sequence is active on not more than 5% gene in *D. melanogaster* (Blanchette et al., 2009).

The binding position with respect to the exon may influence their mode of action as enhancer or silencer. This has been well studied for hnRNPs H and L that seem to repress splicing when bound in an exon, but activate it from the intron downstream (Black, 2003; House and Lynch, 2006; Hui et al., 2005). In addition, it has recently been shown that some proteins (NOVA1, NOVA2, FOX1, FOX2 and hnRNP F) can have double role on splicing depending on their binding position. For example hnRNP H can down regulate the exon inclusion by specifically interacting with G rich sequence, if it is present within the exon (Caputi and Zahler, 2001), whereas an opposite effect can be seen when the interaction is in intron near the 5'ss (Schaub et al., 2007). However, using techniques like CLIP and high throughput sequencing a wide spread view of the position effect on splicing have been also described for proteins like NOVA1, NOVA2, and FOX2 (Ule et al., 2006; Yeo et al., 2009).

1.12 RNA secondary structure.

It is now also widely accepted that the local structure of the pre-mRNA can affect *cis-acting* elements accessibility to *trans-acting* factors. In particular, secondary structure can act by hiding or unmasking splice sites, enhancers, and silencers, with the obvious relative effects (Buratti and Baralle, 2004). Mostly these effects of RNA structure on splicing are based on *in silico* prediction. However, a number of experiments have been

carried out recently combining the *in silico* prediction and wet lab work to prove the importance of RNA structure in splicing (Buratti et al., 2007a; Buratti et al., 2004b; Dhir et al.). For example, a comparative analysis of splicing behaviour of the human and mouse fibronectin EDA orthologous exon revealed that the role of the ESS element was to stabilize the secondary structure of the ESE in such way as to promote binding of SR proteins (Buratti et al., 2004b; Muro et al., 1999).

However, the principal limitation in identifying the role of secondary structure is that our predictive abilities are still rather limited and safe judgement can be made only through implementation of additional functional studies and experimental probing.

1.13 Splicing is part of a co-transcriptional process.

A potential link between splicing and transcription was first proposed by showing that the folding capacity of the pre mRNA after transcription may expose it's SREs differently and affect splicing (Eperon et al., 1988). However, it is now well known that splicing is a cotranscriptional mechanism. Besides the role of transcribing the RNA, polymerase II (Pol II) is involved in cellular mechanism like RNA capping, polyadenylation and splicing (Bentley, 2005; Kornblihtt et al., 2004). A recent microarray analysis shows that 482 genes were affected in Hep3B cell due to UV irradiated Pol II manipulation, in fact there were changes in 1408 alternative events within those genes (Munoz et al., 2009). However, two models have been proposed in these regards. Firstly, the "recruitment model" derived from the observation that several *trans-acting* factors can interact directly or indirectly with the RNA Pol II and other transcription factors (Bentley, 2005; Das et al., 2007; Moldon et al., 2008). The carboxy-terminal domain (CTD) of the RNA Pol II has been shown to play a central role in coupling transcription to pre-mRNA processing acting as assembly platform for proteins involved both in transcription and pre-mRNA process regulation (Bentley, 2005). In fact, in the efficiency of splicing in this model proposed to depend on the factors recruited by the transcription factors. For

example, in the FN1 pre mRNA differentially recruits SF2/ASF to regulate it's splicing based on the nature of the promoter (Cramer et al., 1999). In fact, recruitment does not only appear with well known splicing factors like SR proteins, transcription co-activators can also be recruited and play a role in splicing. For example, transcription co-activator like CAPER and COAA are structurally similar to SR and hnRNP proteins, besides the Pol II, recruitment of transcription coactivators like CAPER and COAA have been shown to regulate alternative splicing in CD44 pre mRNA (Auboeuf et al., 2004). Additionally, the presence of enhancer sequences, like SV40 next to the promoter can stimulate RNA Pol II elongation while the deletion of this enhancer causes a reduction in exon skipping (Kadener et al., 2002).

Secondly, the "kinetic model", is most likely a race against transcription time and spliceosomal assembly (Kornblihtt et al., 2004). This model was proposed by experiments in which RNA Pol II pausing sites were artificially introduced into a gene, delaying the transcription of a splicing inhibitory element and therefore resulting in higher inclusion levels of an alternative exon (Roberts et al., 1998). Further evidence supported this model indicating that transcription can affect splicing acting at different *cis-* and *trans-*acting levels (Kornblihtt et al., 2004). For instance differences in promoter architecture have been shown to affect the subsequent selection of the fibronectin EDA alternative exon and the *CFTR* exon 9 (Cramer et al., 1999; Cramer et al., 1997; Pagani et al., 2003d). However, *in vivo* most of the genes have a single promoter and the regulation of splicing through transcription is more likely to occur through the binding of different transcription factors. In line with this view, it has been described that transcriptional activators can affect alternative splicing (Nogues et al., 2002).

A more direct proof for the kinetic model is derived from studies on RNA Pol II elongation rate (de la Mata et al., 2003; Roberts et al., 1998). A slow Pol II and/or the presence of internal transcriptional stalling sites, results in an increased inclusion of alternative exon harbouring a weak 3'ss. By contrast, when a highly processive RNA Pol II

transcribes the same pre-mRNA, the weak alternative splice site is unable to compete with the stronger downstream 3'ss, which results in skipping of the alternative exon (de la Mata et al., 2003). Recently a reciprocal coupling between splicing and transcriptional elongation has also been reported. In fact, splicing proteins have been involved in transcriptional elongation *in vitro* (Fong and Zhou, 2001) and specific depletion of SC35 showed RNA Pol II accumulation and attenuated elongation *in vivo* (Lin et al., 2008). Furthermore, it was also described that an efficient RNA Pol II transcription is strictly connected with the presence of intronic sequences (Furger et al., 2002). Taking these evidences together, a complex view has emerged from the studies focused on the coupling between transcription and pre-mRNA processing suggesting that both recruitment of factors to the CTD and RNA Pol II kinetic are involved in this connection (Kornblihtt et al., 2004).
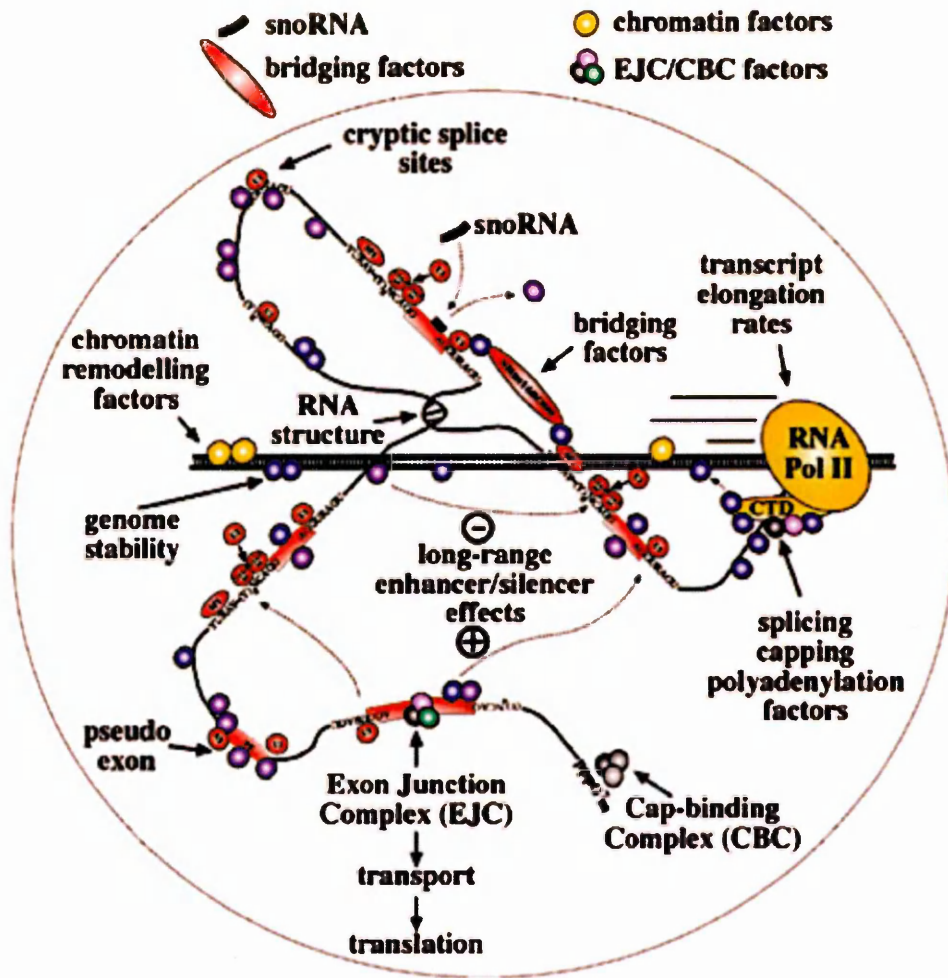
Figure1.9:   Schematic representation of the complex network during mRNA biogenesis.

The diagram shows the complexity of the regulatory elements that participate in splicing control. Figure taken from (Buratti et al., 2006).

1.14 Defective splicing and disease.

In the past, the initial survey by Krawzak and colleagues originally estimated that at least 15% mutation that involved genetic disease caused aberrant pre-mRNA splicing (Krawzak et al.,1992). Over recent years, the pathological alterations that can be directly linked with aberrant splicing processes have grown exponentially, and the study of the complex network interactions between defective splicing and occurrence of disease has become a central issue in the medical research field (Faustino and Cooper, 2003; Garcia-Blanco et al., 2004; Nissim-Rafinia and Kerem, 2005).

In fact, it is now estimated that up to half of the mutation found in exons are anticipated to be affecting the exon inclusion (Lopez-Bigas et al., 2005). Indeed, in the *NF-1* gene, genomic variations that affect splicing may represent up to 50% of all the mutations that cause gene dysfunctions (Ars et al.,2000;Teraoka et al.,1999). Interestingly, within disease causing mutations only 9-10% of the single-point mutations affect the standard consensus splicing signals (Wang and Cooper, 2007)

Therefore, a large number of disease causing mutations must disrupt auxiliary splicing regulatory elements as demonstrated by recent studies (Pagani et al., 2003a). At the moment, the unambiguous identification of this kind of splicing mutations is still hampered by the fact that an ever growing variety of splicing regulatory sequences and proteins are being uncovered every year, making even normal splicing pathways almost too complex to predict on the basis of the primary sequence alone.

In recent years, the development of numerous methodologies has increased our knowledge between splicing and disease. For example, the refinement of minigene-base technology has allowed a relatively fast approach to identify splicing mutation and to study their functional behaviour (Baralle and Baralle, 2005; Cooper, 2005). In addition, SELEX methodologies to characterize binding specificity of *trans-acting* factors involved in splicing regulation (Buratti and Baralle, 2005; Fu, 2004) represent a good starting point for preliminary identification of general splicing regulatory sequences. However, the true molecular mechanism underlying many splicing mutations are still poorly understood often due to difficulties in modelling the splicing systems, highlighting the complexity of splicing (Baralle and Baralle, 2005).

Although this growing complexity makes for very fascinating basic science, in the diagnostic and practical clinical setting it is also the cause of several problems. These range from the simple identification of putative splicing mutations to the exact evaluation of their potential impact upon the normal splicing processing pathways (and thus their effect on disease). The result of all these complexity is that today's task of identifying splicing

spoilers from a background of harmless polymorphisms has become very difficult and all that can really be emphasized is the need to consider any genomic variation, even those that occur deep within intronic regions, as a potential splicing mutation.

1.15 Correlation of synonymous and non-synonymous mutation and evolution.

In the recent past, evolution was especially concerned with the biochemical properties of the protein, dealing with the conception that the domain architecture of the protein should be untouched to avoid selection pressure and proteins with dense functional domain will evolve slowly than the one with less domains (Pal et al., 2006). However, the selection pressure at the protein level comes to consideration once the pre-mRNA is processed flawlessly. More precisely, soon after the transcription pre-mRNA is subjected to splicing, transportation and finally translation. The initial concept of connecting splicing with evolution came from Gilbert's proposal that the same exon can be present in multiple transcripts, which now recognized as alternative splicing (Gilbert, 1978). That concept has been delineated by proving that the same exons are not necessarily be spliced constitutively in all species (Gilbert, 1978; Pan et al., 2005). In fact, genome wide sequencing as well as comparative genomics has made it possible to look back the evolutionary history of exons and their mode of splicing (Kan et al., 2005; Nagasaki et al., 2005). As a consequence, alternative splicing as opposed to constitutive splicing has been recognized as an accelerating force to the rate of evolution. The reason is that, insertion of new constitutive exons within the transcript is expected to bring up negative selection pressure (unless the new transcript has a positive phenotypic effect). On the other hand, if the new exon is alternatively spliced, the selection pressure is more relaxed, because the cell retains a fraction of the original transcript.

Nucleotide changes in the coding region that do not change the codon usage are referred as synonymous (or silent) mutation. In contrast, nucleotide changes that change the amino acid are known as non-synonymous mutations. Non-synonymous mutations can be

50

either missense if they change the codon amino acid specificity or non-sense, if they introduce a premature stop codon. Synonymous changes in the coding sequence were thought to be benign polymorphism and functionally neutral in clinical diagnosis. The only consideration regarding their association with pathology was if the synonymous change was affecting the splice sites (Li et al., 1995; Richard and Beckmann, 1995). It is now clear that before natural selection can act at the translation level by favouring new amino acid sequences, the splicing process has to be preserved. The other way round, a nucleotide change that might cause exon skipping would never be selected even if the function of the resulting protein might be improved with respect to the original (Pagani and Baralle, 2004; Pagani et al., 2005).

In general, the rate of evolution is measured by calculating the sequence divergence within the common ancestors as well as the time has past since the divergence of speciation. Most frequently used metrics for measuring evolution at the molecular level was based on the ratio of Ka (non- synonymous change) over Ks (synonymous changes) in the same protein. A Ka/Ks ratio lower than 1 indicates purifying selection, while a Ka/Ks ratio higher than 1 indicates positive selective pressure. Ka/Ks metrics has been widely used to measure the selection pressure on amino acids (Yang and Nielsen, 2000). However, for measuring sequence divergence, Ka/ Ks ratio often produces ambiguous results in alternatively spliced exons (Hurst and Pal, 2001). In fact, alternative exons in the alternatively spliced genes show higher Ka/Ks ratio in sequence comparison, which rather indicates the importance of local segments like SREs of the gene for selection (Xing and Lee, 2005).

In recent years, the presumed neutrality of synonymous mutations has been challenged because of the presence of splicing regulatory elements overlapping with the amino acid code. Several recent bioinformatics analyses have confirmed this view by showing a reduction in the rate of synonymous evolution in regions that contain an ESE (Hurst & Pal, 2001; Orban and Olah, 2001). Moreover, it has recently been observed that

51

synonymous single nucleotide polymorphisms (SNPs) that disrupt ESEs have also been selected against other SNPs (Carlinai and Genut, 2006; Fairbrother et al., 2004). Altogether, evolutionary studies represent an important field for investigation of the elements involved in splicing regulation as they can give us the chance of finally understand the splicing mechanism itself.
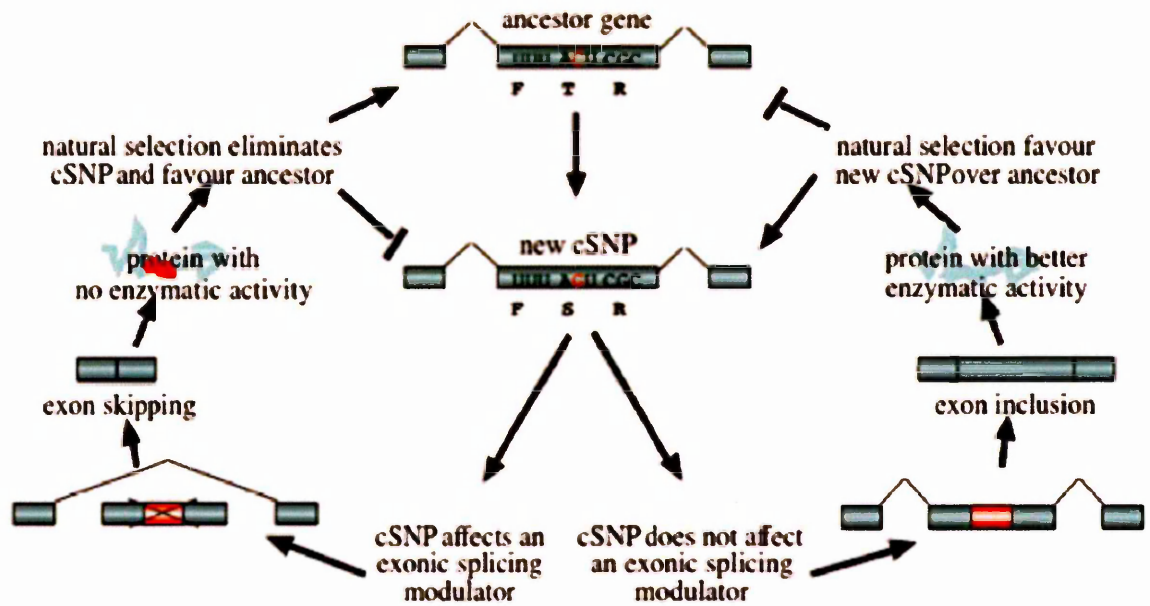
Figure1.10: The primary selective pressure on exons is for their inclusion in mRNA. The selection of a new amino acid that leads to a better enzyme can occur only if the codon substitution caused by a single nucleotide polymorphism (cSNP) and does not affect an exonic regulatory element. In this model, assumed that a protein had amino acid changes at the catalytic site that will produce a more active enzyme that has a selective advantage. In order to gain that, first of all the nucleotide substitution has to be compatible with the splicing machinery that identifies the exon. If the inclusion is guaranteed, then the amino acid change is favoured. If not, exon skipping will result in a non functional enzyme. Adopted from (Pagani and Baralle, 2004)

1.16 Cystic fibrosis (CFTR).

Cystic fibrosis (CF), also known as mucoviscoidosis is an autosomal recessive genetic disease that mostly affects the entire body causing progressive disability and early death. The frequency of the disease differs among ethnic groups. However, it is more common in the Caucasian with a incidence of 1 in every 3300 whilst within Hispanics the occurrence is 1 in 9500 and it is an even rarer disorder in native Africans and native Asians, where it is estimated to occur in less than 1 in 50000 individuals (Goss and Rosenfeld, 2004).

Cystic fibrosis (CF) is an extremely heterogeneous disease both for the age of onset and for the clinical features. A number of individuals escape detection in the first decade or two of life, often because symptoms are mild. However, the typical patient with CF generally shows symptoms like thick production of mucus, which causes an obstructive chronic lungs disease, exocrine pancreatic insufficiency, intestinal obstruction of ileum in the newborns (meconium ileus) and sterility in 95% of men and in 10% of women (Kerem and Kerem, 1996; Koch and Hoiby, 1993). Although several organs are affected in CF, the underlying principle mechanism of pathogenesis is considered to be the loss of epithelial plasmamembrane chloride conductance. Decrease in salt and fluid secretion is responsible for the blockage of exocrine outflow from the pancreas and the accumulation of heavy and dehydrated mucus in the airways. Later onset, due to the favourable condition or loss antimicrobial activity in the airway surface, causes respiratory infections by pathogens like *Pseudomonas aeruginosa*, *Staphylococcus aureus* and *Aspergillus fumigatus* (Goldman et al., 1997).

The gene *CFTR* (Cystic Fibrosis Transmembrane conductance Regulator) that encodes the protein is located at the human chromosome 7 (region q31). This was the first gene identified by positional cloning (Riordan et al., 1989). The mRNA encoded from the gene is of 6kb and contains 27 exons. CFTR is a membrane associated glycoprotein of 169 kD in size and consists of 1480 amino acid (Gregory et al., 1990). Structurally the protein

has two nucleotide binding domains (NBD1 and NBD2), one regulatory domain (R) and two hydrophobic transmembrane domain (each consist of six membrane spanning segments) (Riordan et al., 1989). The NBD1 subdivision of the protein is encoded by exon 9 through exon 12. In particular, exon 12 encodes amino acid spanning 560 to 588 towards the C terminal of the NBD1 (Lewis et al., 2005). In brief, the protein functions to regulate the channel through phosphorylation of the R and NBDs. The proposed mechanism is that partial phosphorylation of the R domain cause the dimerization with NBD1, which leads to the opening of the gate at the transmembrane domain. Whereas, complete phosphorylation of the R domain cause the closing of the gate by interacting with the NBD2 (Bompadre et al., 2005). Mutation to any of these functional domains produces defective proteins and leads to disease.

(A)

(B)
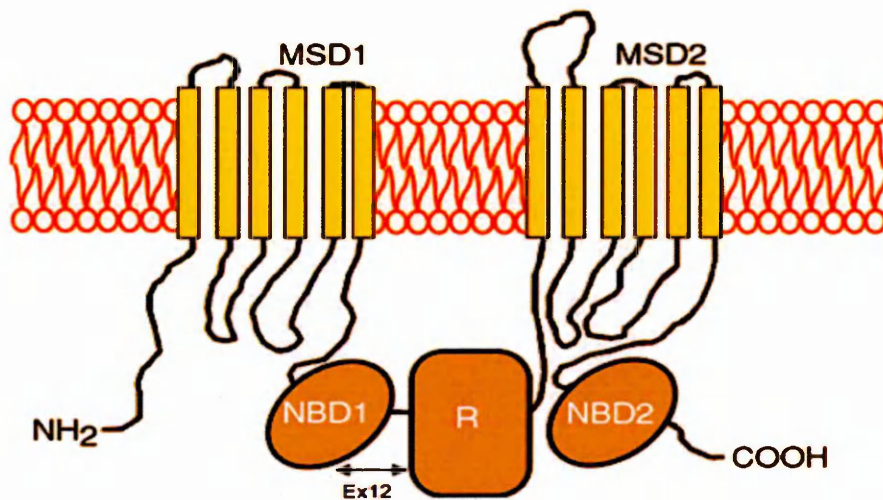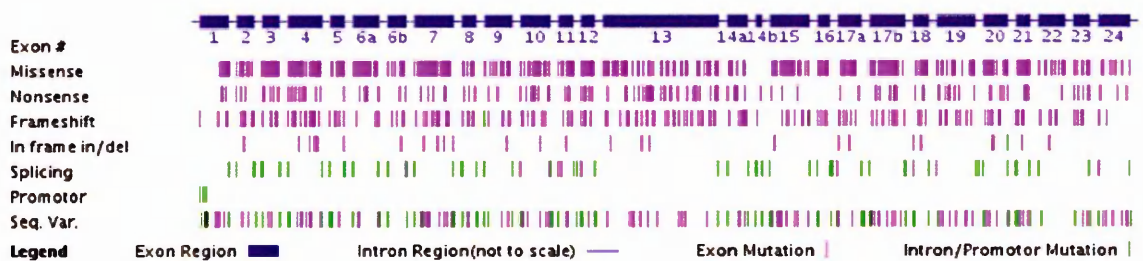
Figure1.11: *CFTR* **protein and affected regions in the gene.**

The *CFTR* gene is located on the long arm of chromosome 7 in the q31 region, it is 250 kilobases long and it has 27 exons. Its transcription and translation produce a 6129 nucleotides long mRNA and a 1480 amino acids long protein respectively. (A) The protein

54

is inserted in the membrane and it has five domains: two transmembrane domains (MSD1 and MSD2), two nucleotide binding domains (NBD1 and NBd2) and a regulatory domain (R domain). (B) The figure shows the frequency and pattern of the mutations causing cystic fibrosis in the *CFTR* gene. Adopted from http://www.genet.sickkids.on.ca .

The symptoms among CF patients differ a lot and it is hard correlate the phenotypic symptoms with the genotype. However, phenotypic symptoms can be categories in three groups. In the first one, we can include the symptoms that are common to most CF patients, regardless of the type of mutations. An example is the abnormal electrolyte composition of sweat that is common to virtually all patients with classical CF. In the second category there are features which show a good correlation between genotype and disease phenotype. This category is best represented by the pancreatic function of the patients. The third category of phenotypic features includes symptoms that do not show significant correlation with genotype, such as the pulmonary status of CF patients, in which the severity of the disease is strongly affected by environmental and secondary genetic factors (CF modifiers) (Zielenski, 2000; Zielenski and Tsui, 1995).

The number of mutations in the *CFTR* gene causing these symptons is growing every day. The complete list of mutation identified up to now can be found at the world web access http://www.genet.sickkids.on.ca. In most of the cases, the pathological effect of single base substitutions in *CFTR* gene were considered to be based on the change in coding sequence (or were mis-classified as benign polymorphism databases). Several examples have been shown that many of these changes affect splicing pattern of the gene instead changing the specific amino acids (Pagani et al., 2003a; Pagani et al., 2003c; Zielenski et al., 1995).

1.17 Final introductory considerations.

This introduction has attempted to put together a brief summary of RNA processing mechanisms and to introduce a model system that is extremely rich in mechanistic and functional outcomes of alternative spicing. There are plenty of data focusing on specific regulatory elements in disease genes and on the effect that mutations in their sequence may have on the splicing outcome. However, very few of these works have attempted a systematic study on a specific exon to classify all the enhancers and silencers that participate in its definition. The chosen *CFTR* exon 12 is an excellent base to attempt this, because of the abundant information on mutation analysis(Pagani et al., 2005; Pagani et al., 2003b), illustrated RNA secondary structure constrains (Meyer and Miklos, 2005), presence of distinctive *cis* element CERES (Pagani et al., 2003b) and obviously its association with disease. As it will be clear in the results section, even a well studied exon such as this one displays evident complex changes in the SREs distribution that define its inclusion when looked across species evolution.

1.18 Aim of the project.

A general view of alternative splicing implies focusing on specific sequence elements, some interacting predominantly with SR proteins that act in a way as to include the exon (splicing enhancers) and some with hnRNPs that promote exon skipping (splicing suppressors). In theory, this should be a rather simple task of mutagenizing systematically the sequence of the exon of interest and simply classify the resulting elements in enhancers/silencers depending on their effects upon exon inclusion following mutagenesis. However, this strategy often produces ambiguous results, particularly in the situation where the cis elements are extremely overlapped like the CERES2 element in CFTR exon 12. Until now, the functional role of CERES elements was only limited to systematic mutation analysis and the molecular mechanisms that explained its action still largely remained unexplored.

To find some of the answers to these questions, this thesis was aimed to address following issues:

- Correlation between codon and amino acid changes with splicing efficiency.

- Regulation of CFTR exon 12 CERES2 in terms of trans-acting factors binding to this sequence.

- Effect of sequence variation across species on splicing regulation.

2 Results

2.1 Effect of the pathological missense and synonymous mutation on *CFTR* exon 12 in the CERES2 element and nearby regions.

In normal conditions, the level of *CFTR* exon 12 skipping from the full mRNA transcript is variable and ranges from 5%-30% (Hull et al., 1994). Loss of this exon removes the highly conserved region from the NBD1 and produces non-functional CFTR protein. Previous work from our lab showed that two missense mutations (D565G/A15G and G576A/G48C) in the *CFTR* exon 12 causes cystic fibrosis by affecting the splicing of the exon and Y577F/A51T by changing the amino acid (Pagani et al., 2003b). Initially these pathological mutations were considered as polymorphism in the CFTR mutation database.

In order to understand the role of these mutations in splicing, Pagani et al at first analyzed the RNA from patient nasal epithelial tissue and then carried the analysis into a minigene system to better understand their effects on splicing regulation. In the minigene system, along with exon 12 they inserted 333nt from intron 11 and 270nt from intron 12 to obtain a reliable splicing pattern that mimicked the endogenous splicing pattern. The result of the minigene system showed that the G576A/G48C mutation caused 93% exon skipping. On the other hand, the adjacent Y577F/A51T mutation resulted in complete inclusion. Moreover, the transfection of these minigenes in different cell lines (HeLa, COS, T84, NT2, CFPAC and Hep3B) yielded slightly different levels of inclusion suggesting that there might be a differential expression of tissue specific splicing factors. Further, using site-directed mutagenesis to better analyze the effects of splicing of these mutations they characterized two segments within this exon and showed that they are important for *CFTR* exon 12 splicing, naming them CERES 1 and CERES 2. The important feature of these

CERES elements was represented by the observation that they did not behave as classical enhancer or silencer elements, but rather as a composite sequence. Accordingly, it was very difficult to predict the effect on the splicing process of single-nucleotide substitutions introduced in these regions.

In this study, we have focused on understanding the molecular mechanism of CERES2. We have selected four mutations (G48C, A49G, A51T and C52T) that are present within the CERES2 regions, and a nearby mutation (T40C) for further detail analysis of their effect and interactions. The missense mutations G576A/G48C Y577F/A51T were naturally found in patients and S573S/T40C, G576G/A49G and Y577Y/C52T that are synonymous substitution. The T40C and C52T substitutions are particularly interesting form an evolutionary point of view as they are naturally present in the mouse *CFTR* exon 12 sequence (without affecting the exon inclusion) and C52T has also been reported as a human polymorphism/possible mutation in the Cystic Fibrosis Mutation Database (www.genet.sickkids.on.ca). The localization of these mutations in the pTB minigene are shown in Figure 2.1A.
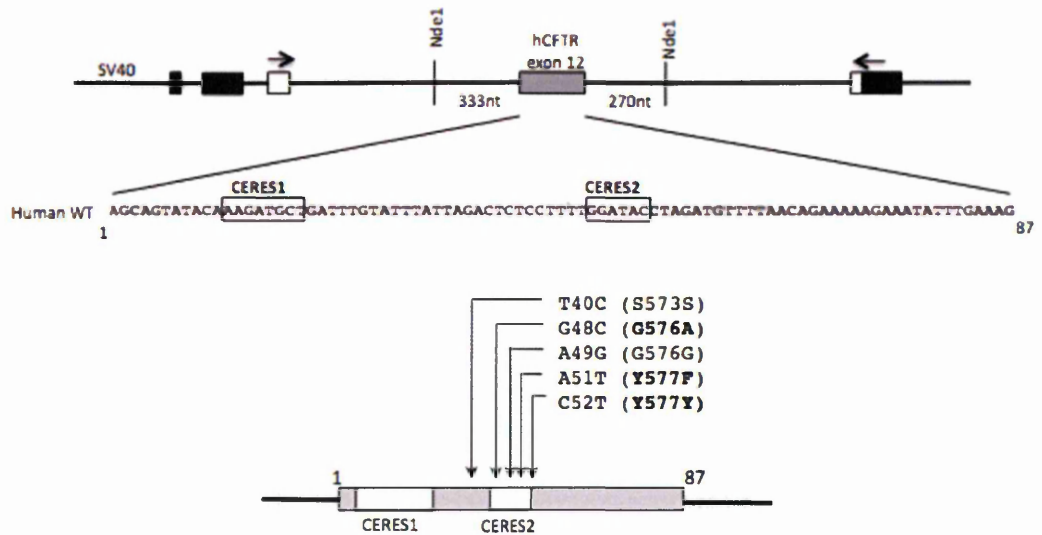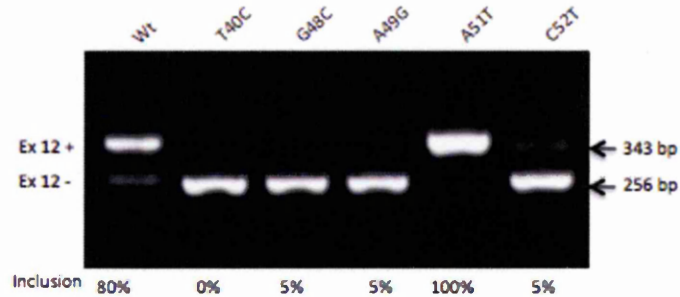
Figure 2.1: Transient transfection to HeLa cells of CFTR exon 12 minigenes. (A). Schematic presentation of the hybrid minigene used in transfection experiments. The α-globin, fibronectin EDB, and human *CFTR* Exon 12 are shown as black, white and gray boxes, respectively. The sequence of *CFTR* exon 12 and position of the CERES1 and CERES2 elements is reported in full. (**B).** Schematic diagram of *CFTR* Ex 12 construct used in the analysis. The vertical superimposed arrows indicate the locations of both natural and synonymous mutations. The amplified RT–PCR products stained with ethidium bromide are shown in the bottom panels. Spliced transcripts are shown with Ex.12+ for inclusion and Ex.12- for exclusion of the exon.

Transfection in HeLa cells of these constructs produced the following results (as previously reported by Pagani et al., 2003b): the T40C, G48C, A49G and C52T mutation carrying minigenes caused exon skipping in the full exon 12 context whilst A51T caused to full exon inclusion(Pagani et al., 2005; Pagani et al., 2003b).

2.2 Functional role of CERES2 in a heterologus context.

In ordered to test the activity of the CERES2 element, I cloned it within a heterologous context. To do this we used the dsx-XH vector. This vector is derived from *D. melanogaster dsx* and contains exon 3 and 4 with a suboptimal 3' splice site (Tanaka et al., 1994). Recognition of exon 4 is dependent on presence of an ESE (gray box, Fig.2.2) localized in its exonic sequence. We therefore replaced the natural ESE of exon 4 with the CERES2 sequence either in the wild-type form or carrying the A51T and G48C mutations. As positive control sequence, we used the previously well studied ASLV (avian sarcoma-leukosis virus) ESE whilst as negative control we used a well known ISS sequence from the *CFTR* exon 9 gene (AS3 ISS) (Bruzik and Maniatis, 1995; Buratti et al., 2007b).

(A)

(B)
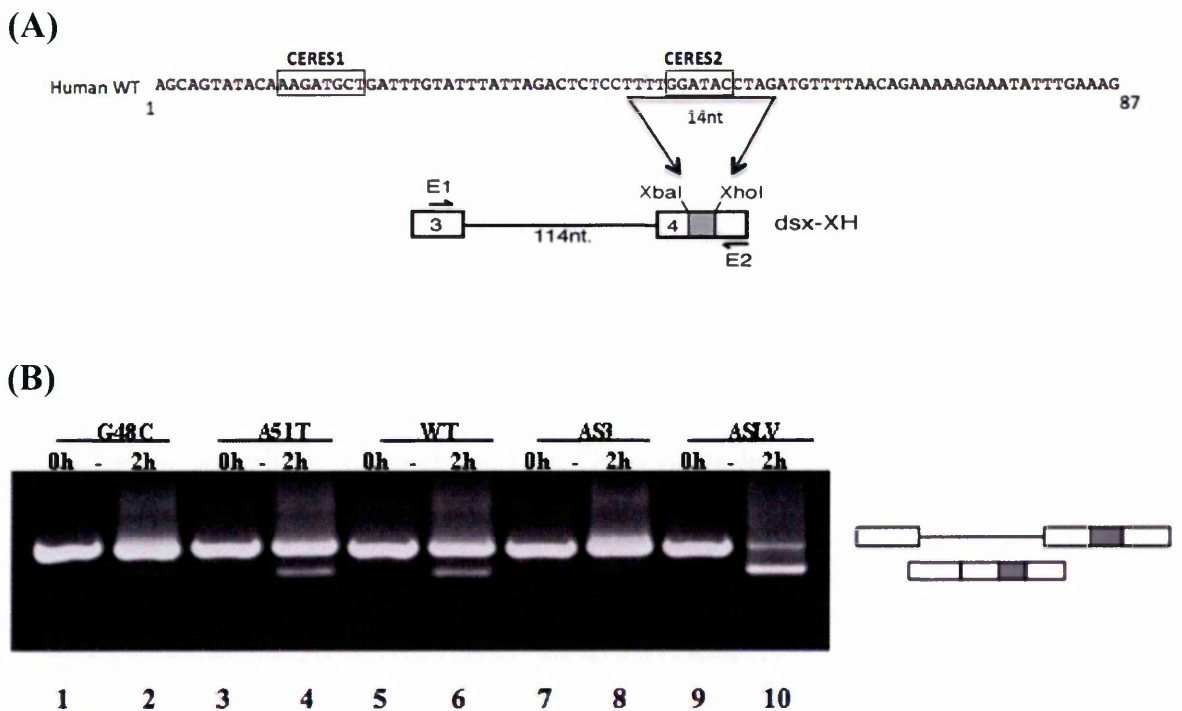
Figure 2.2: HeLa cell nuclear extract based splicing assay of CERES2 in a pdsx-HX system. (A). Schematic presentation of the dsx-XH construct used in transfection experiments. The gray box indicates the location where 14nt of CERES2 were cloned. Horizontal arrows with the letter E1 and E2 are positioned for sense and antisense primer respectively. **(B).** The amplified RT–PCR products after splicing were stained with ethidium bromide and shown in the bottom panels. Spliced transcripts (box for exons) and unspliced pre mRNA (box for exon and line for intron) are shown at the side.

PCR amplification of the splicing assay shows that both WT and A51T are capable of promoting splicing with very low efficiency but G48C doesn't splice at all. If we compare these results with those obtained with the ASLV ESE (lanes 9-10), it is apparent that in a heterologous context the CERES2 element possesses a low enhancing ability, which is abolished by the G48C mutation and is not enhanced by A51T. Taken together, these results suggest that the activity of CERES2 is strongly context-dependent and for this reason we did not pursue this approach any further.

2.3   Functional role of CERES2 in a shortened *CFTR* 12 exonic context.

For this reason, we therefore wanted to see if all these substitutions were dependent on the context provided by the rest of the exon sequence. To analyze this, the *CFTR* exon 12 was shortened by removing the flanking regions near the 3 and 5's but maintaining 4 and 3 nucleotides close to the 3' intron-exon and exon-5' intron junctions, respectively. This minigene construct was called "mini" exon 12 (Figure2.3 A).

(A)
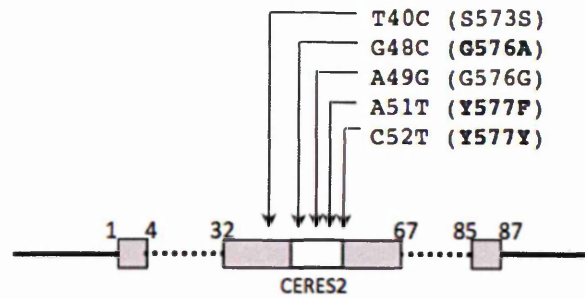
(B)
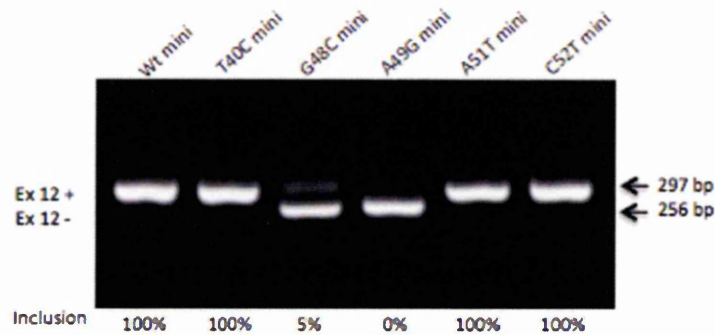
Figure 2.3: Transient transfection to HeLa cells of nucleotide substitutions in a shortened *CFTR* exon 12 minigene. **(A).** Schematic diagram of *CFTR* mini Ex 12 constructs used in the analysis. Dotted line indicates the exonic sequence which was removed. The vertical superimposed arrows indicate the locations of both natural and synonymous mutations. The amplified RT–PCR products stained with ethidium bromide are shown in the bottom panels **(B)**. Spliced transcripts are shown with Ex.12+ for inclusion and Ex.12- for exclusion of the exon.

When all the previously analyzed mutations are placed in this reduced context, both the G48C and A49G were observed to cause the exon skipping observed in the full length exon, whereas in A51T it was not possible to distinguish for any enhancing effect due to the fact that the wild-type mini exon 12 is fully included in the spliced transcript (as opposed to only 80% inclusion of the full length exon 12) (Figure2.3 B). Interestingly, in the case of the two mouse-specific synonymous substitutions, T40C and C52T, the mini-exons carrying these mutations were completely unaffected. This observation contrasts the complete inhibitory effect of these two substitutions in the full *CFTR* exon 12 minigene context (Fig. 2.1). Notably both T40C and C52T are naturally present in the mouse *CFTR*

exon 12 sequence. One possibility was that their silencing effect in the mini-exon 12 context might have been influenced by mouse-specific splicing factors. In fact, it has been previously shown that in many genes, splicing can also be regulated in a tissue specific as well as in a species specific manner due to presence of distinct tissue specific factors (Caceres et al., 1994; van der Voort et al.; Venables et al., 2004). For example, T-STAR is one of three members of the SAM68 (SR) family protein. Human T-STAR, like its rodent orthologues can influence splice site choice and that human, but not mouse, T-STAR-dependent alternative splicing is modulated by SIAH1. SIAH-mediated down regulation of alternative splicing may be an important developmental difference between highly conserved T-STAR proteins (Venables et al., 2004).

2.4 Testing the eventual presence of mouse cell-line specific effects of mouse substitutions in human *CFTR* exon 12.

In order to evaluate, if there is any species specific factor present in mouse but absent in humans that can modulate the splicing of T40C and C52T, the full *CFTR* exon 12 minigene constructs were transfected in mouse Hepatocyte cell line N-Muli. As shown in Figure 2.4, however, transfection of these constructs in the mouse cell line didn't make any difference with regards to exon skipping and the result was exactly similar to the one observed in HeLa cells (Fig. 2.1). This suggests that there are no mouse specific factors, which can affect the way these two mutations act on the *CFTR* exon 12 splicing process.

(A)
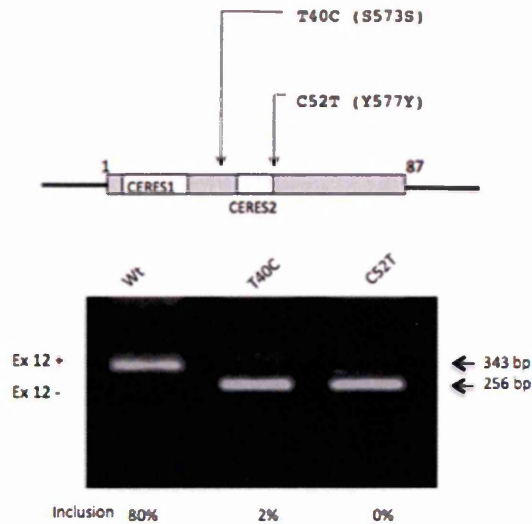
Figure 2.4: Transient transfection to mouse Hepatocyte cells of the mouse base substitutions in human context. (A). Schematic diagram of *CFTR* Ex 12 constructs used in the analysis. The vertical superimposed arrows indicate the locations of two synonymous mouse substitutions. **(B)** Along with the human *CFTR* 12 WT, T40C and C52T constructs were transfected in to the mouse hepatocyte cell line N-Muli. The amplified RT–PCR products were stained with ethidium bromide are shown in the bottom panels. Spliced transcripts are shown with Ex.12+ for inclusion and Ex.12- for exclusion of the exon.

2.5 Identifying the *trans-acting* factors whose binding may be affected by pathological missense mutation (G48C and A51T).

Previously, only the effects of splicing factor hnRNP A1 and SF2/ASF overexpression were shown to affect *CFTR* exon 12 splicing in general (Pagani et al., 2003c).

For the scope of my thesis, therefore, it was decided to better characterize these observations in terms of binding to the CERES2 element of a wider range of splicing factors. To achieve this, a pulldown system was used. This particular methodology has been previously optimized in our lab to identify specific RNA binding proteins in a variety of exonic/intronic contexts (Buratti et al., 2004a; Buratti et al., 2001) and is described in detail in the Material and methods section. Briefly, in this pulldown affinity assay the RNA of interest was bound to adipic acid dehydrazide beads and was incubated with nuclear extract to allow the binding of all interacting proteins. Following extensive washings, the beads

were collected and the bound proteins were separated using SDS-PAGE electrophoresis. In order to normalize the pulldown assay, as well as a positive control for the experimental procedure, all the RNAs bound to the beads were also tagged with $(UG)_8$ tails. TDP43 has been previously shown to be very specific for UG repeats (at least 6 repeats) (Buratti and Baralle, 2001). Once these RNAs carrying $(UG)_8$ tails were incubated in NE, the TDP43 of the HeLa NE were bound to the 3' end of the RNA tail. Further in the elution process of the pulldown (see details in 4.30 of materials and methods section) these proteins were also harvested along with the other proteins. This additional binding of TDP43 allowed us to measure each RNA bound protein band intensity in westernblot analysis and normalize them according to the binding levels of the TDP-43 protein of that particular experiment (Fig.2.5A).

(A)

```
        32                    CERES2              67
WT  : UUAGACUCUCCUUUU GGAUAC CUAGAUGUUUUAACA-----(UG)8

G48C : UUAGACUCUCCUUUU GCAUAC CUAGAUGUUUUAACA------(UG)8

A51T : UUAGACUCUCCUUUU GGAUUC CUAGAUGUUUUAACA-----(UG)8
```

(B)

(C)
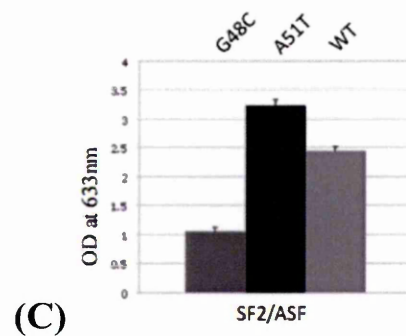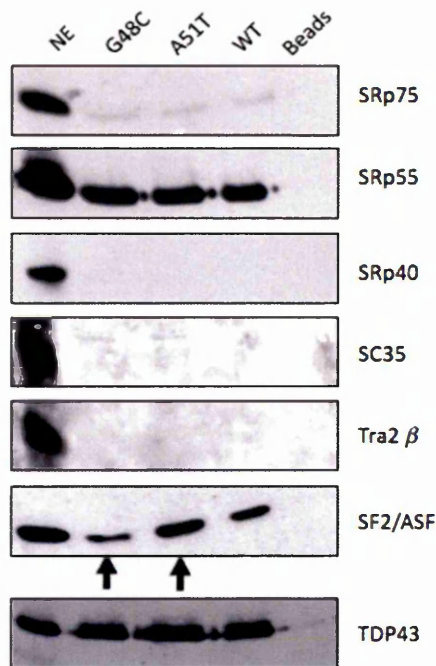
Figure2.5: **Pulldown analysis of SR proteins for pathological substitutions.** Western blot analysis of recovered proteins after pulldown of two naturally occurring non-sense *CFTR* mutations (G48C/G576A and A51T/Y577F) compared with *CFTR* Ex 12 wild type. *In vitro* transcribed RNA was used for analysis. **(A).** Transcribed RNA sequences used for pulldown analysis with the mutations (underlined). A (UG)8 repeat specific for TDP43 was added at the 3'end of each RNA to normalize the data after western blot. **(B)** Affinity assay for binding of the following SR factors: SRp75, SRp55, SRp40, SC35, Tra2β and SF2/ASF. Detection of all these proteins was performed by Western blot using specific antibodies. **(C)** Three independent experiments for SF2/ASF have been quantified relative to TDP-43 using an Ultro Scan XL, Pharmacia LKB - laser densitometer at 633nM wavelength according to manufacturer's instructions and presented.

To perform this experiment, I have used *in vitro* transcribed RNA sequences equal in length to the wild-type mini *CFTR* exon 12 sequence and two versions of this sequence carrying the two missense mutations G48C and A51T. Using Western blot, I then tested all these RNAs for binding to the following proteins: SRp75, SRp55, SRp40, SF2/ASF and

Tra2b (Figure 2.5) and hnRNP U, PTB, hnRNP H, DAZAP1, hnRNP C2, A1, A2 (Figure 2.6).

The result of this analysis shows that there was no binding for the hnRNP H and PTB, either in its wild-type form or carrying the G48C or the A51T mutations. Similarly, no binding could be observed for SRp75, SRp40, SC35, and Tra2b proteins in the same stretch of sequences. On the other hand, hnRNP U, hnRNP A1, hnRNPA2, DAZAP1 and SRp55 could bind all these sequences to approximately the same levels, irrespectively of the presence or absence of mutations.
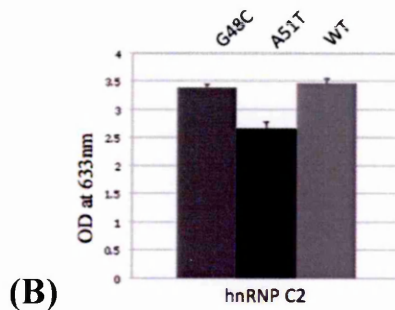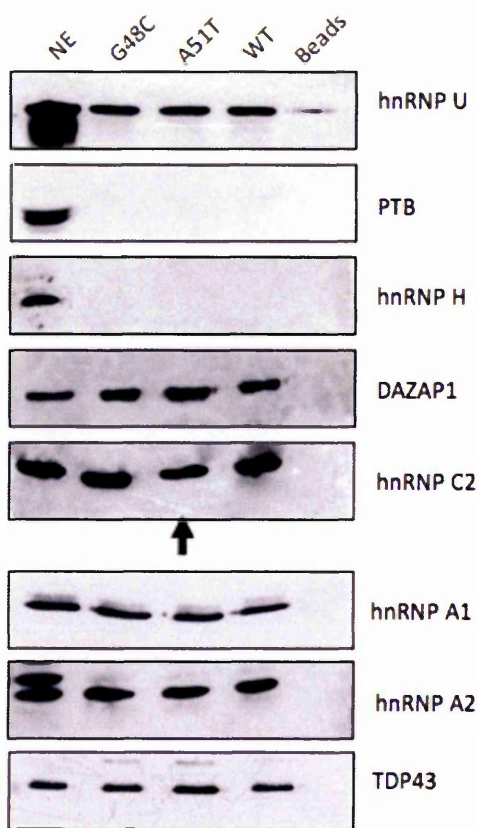
Figure 2.6: Pulldown analysis of hnRNPs for pathological substitutions. Western blot analysis of recovered proteins after pulldown of two naturally occurring non-sense *CFTR* mutations (G48C/G576A and A51T/Y577F) compared with *CFTR* Ex 12 wild type. Previously mentioned *in vitro* transcribed RNA was used for this analysis. **(A)** Affinity assay for binding of the following hnRNPs: U, PTB, H, DAZAP1, C2, A1 and A2. Detection of all these proteins was performed by Western blot using specific antibodies. **(B)** Three independent experiments for hnRNP C2 have been quantified relative to TDP-43 using an Ultro Scan XL, Pharmacia LKB - laser densitometer at 633nM wavelength according to manufacturer's instructions and presented.

Interestingly, however, two factors displayed a differential binding ability in the wild-type sequence with respect to these mutations.

In particular, for SR proteins the most prominent change could be seen for the SF2/ASF protein that bound more efficiently to the A51T mutant with respect to the wild-type sequence. In addition, this factor also showed less affinity to the G48C mutant compared to the wild-type (Figure 2.5).

In case of the hnRNPs, the hnRNP C2 factor was observed to bind less efficiently to the A51T mutant with respect to wild-type.

Taken together, these SR and hnRNP profiles were very much consistent with the observed minigene results. In fact, the G48C mutation that caused exon skipping in the reduced context displayed a lower binding efficiency for SF2/ASF (a positive splicing factor in most contexts). Whereas in A51T the SF2/ASF bound more and hnRNPC2 (a negative splicing factor (Venables et al., 2008) marginally less compared to the wild-type.

2.6 Identifying the *trans-acting* factors whose binding is affected by synonymous substitutions (T40C, A49G, and C52T).

Next we investigated the three synonymous mutations using as template the sequence shown in Fig. 2.7A. In these cases, the pulldown profiles showed a similar result to that observed with the non-synonymous mutations. In fact, no binding to any RNA could be observed for SRp75, SRp40, SC35, and Tra2b (Figure2.7) and hnRNP H and PTB (Figure2.8).

(A)

```
            32                        CERES2              67
WT     UUAGACUCUCCUUUU GGAUAC CUAGAUGUUUUAACA-------(UG)8

T40C   UUAGACUCCCCUUUU GGAUAC CUAGAUGUUUUAACA-------(UG)8

A49G   UUAGACUCUCCUUUU GGGUAC CUAGAUGUUUUAACA-------(UG)8

C52T   UUAGACUCUCCUUUU GGAUAU CUAGAUGUUUUAACA-------(UG)8
```

(B)

(C)
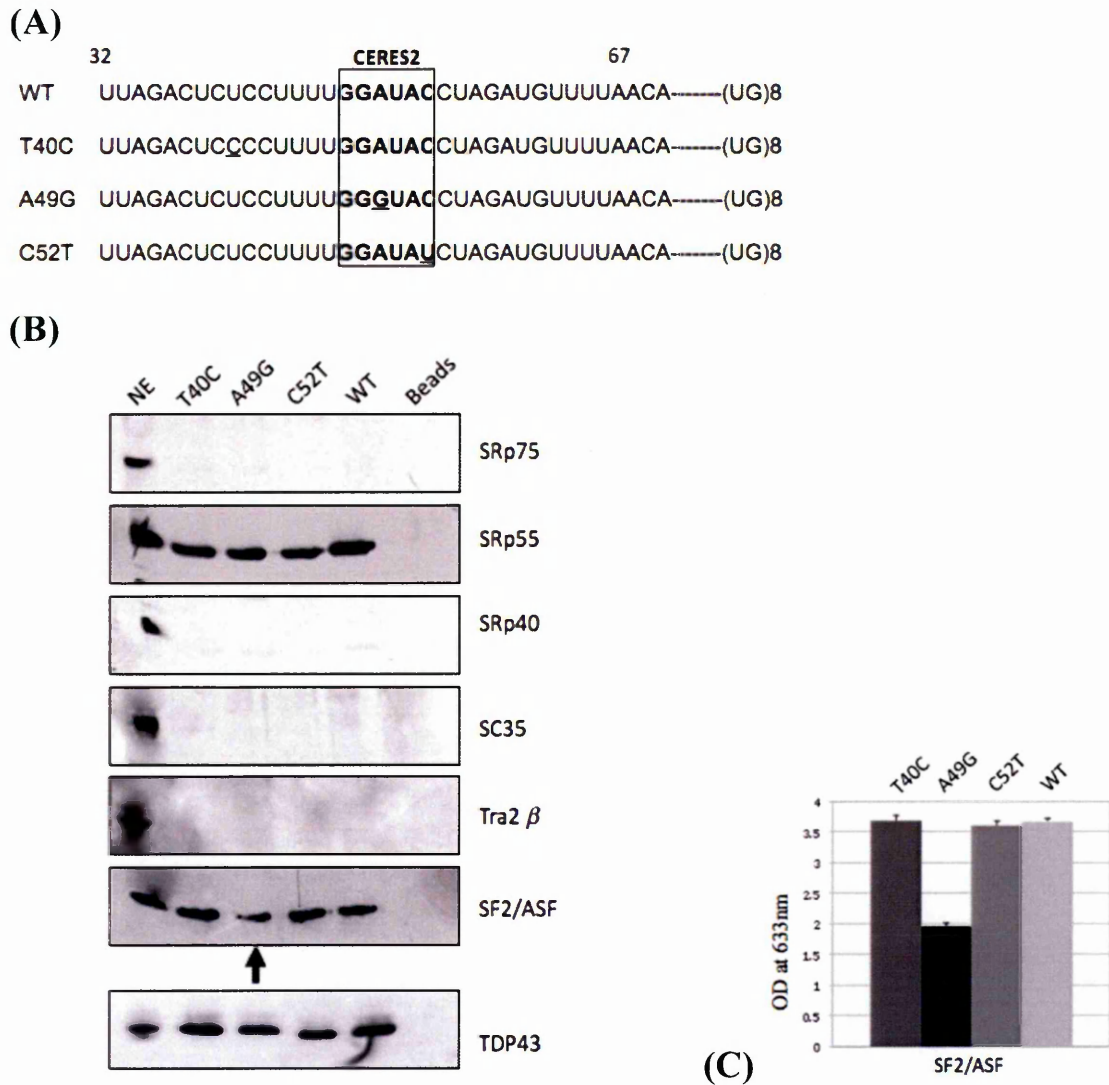
Figure 2.7: Pulldown analysis of SR proteins for synonymous substitutions. Affinity pulldown performed for the three synonymous mutations (T40C,A49G and C52T) in *CFTR* exon 12. **(A).** Transcribed RNA sequences used for pulldown analysis along with their mutations (underlined). A (UG)8 repeat specific for TDP43 was added at the 3'end of each RNA for pulldown normalization. **(B)** Affinity assay for binding of the following SR factors: SRp75, SRp55, SRp40, SC35, Tra2β and SF2/ASF. Detection of all these proteins was performed by Western blot using specific antibodies. **(C)** Three independent experiments for SF2/ASF have been quantified relative to TDP-43 using an Ultro Scan XL, Pharmacia LKB - laser densitometer at 633nM wavelength according to manufacturer's instructions and presented.

Only in case of the A49G mutation less binding of the SF2/ASF protein was observed than in the wild-type sequence, a situation like G48C and consistent with the observation that also this mutation has an inhibitory effect on *CFTR* exon 12 splicing both

in the full and the mini-context. Interestingly, no changes could be observed in the binding profiles of RNAs carrying the T40C and C52T substitutions with respect to the wild-type sequence. In this respect, the observation that no changes could be observed for any of these proteins was also consistent with the functional assays demonstrating that these two substitutions were neutral in the human mini context (see Figure2.3).

(A)
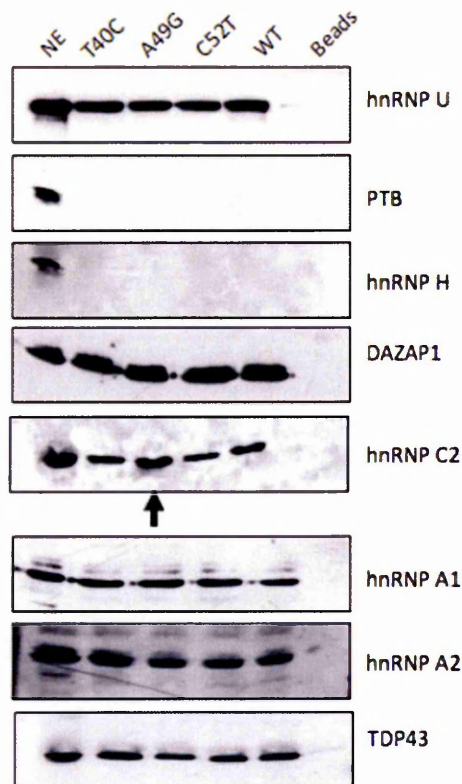
(B)
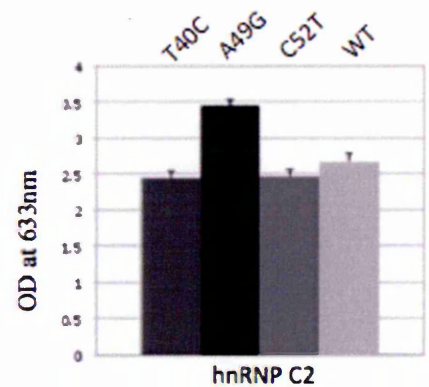
Figure 2.8: Pulldown analysis of hnRNPs for synonymous substitutions. Affinity pulldown performed for the three synonymous mutations (T40C,A49G and C52T) in *CFTR* exon 12. Previously described in Figure 2.7 (A), *in vitro* transcribed RNA sequences used for this pulldown analysis along with their mutations (underlined). A (UG)8 repeat specific for TDP43 was added at the 3'end of each RNA for pulldown normalization. **(A)** Affinity assay for binding of the following hnRNPs: U, PTB, H, DAZAP1, C2, A1 and A2. Detection of all these proteins was performed by Western blot using specific antibodies. **(B)** Three independent experiments for hnRNP C2 have been quantified relative to TDP-43 using an Ultro Scan XL, Pharmacia LKB - laser densitometer at 633nM wavelength according to manufacturer's instructions and presented.

2.7 Validating the role played by SR factors in *CFTR* exon 12 splicing.

As previously shown in the pulldown assays (Fig. 2.5 and 2.7), two positive splicing factors SRp55 and SF2/ASF were found to interact with wild-type and mutant CERES2 regions of *CFTR* exon 12. In order to validate the functional role played by these two SR proteins in splicing, we tested the effects of their overexpression for the G48C and A49G minigenes of mini-exon 12 sequences. Together with SF2/ASF and SRp55, we also tested SC35 (as an example of a SR protein not interacting with the mini exon). The empty pCG plasmid was also used as a negative control.

(A)
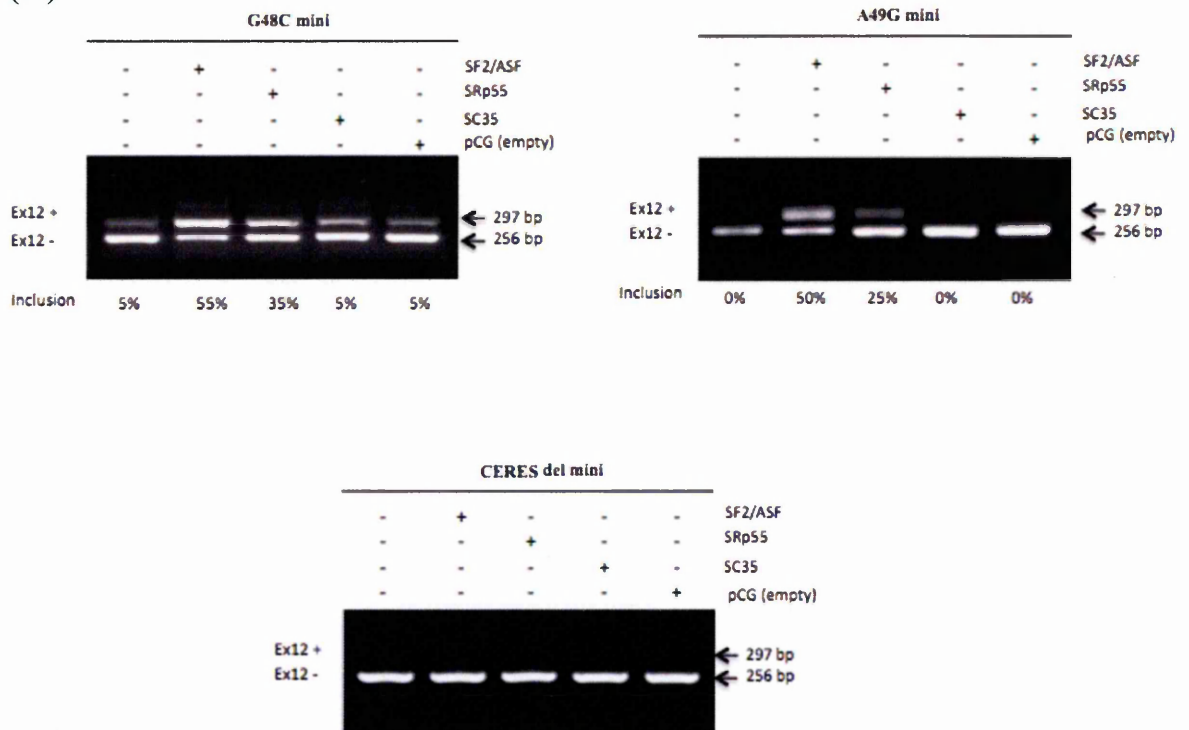
Figure 2.9: Effect of overexpression of SR proteins in G48C and A49G mutation.
(A). Analysis showing the overexpression of SR proteins (SRp55, SC35 and SF2/ASF) to rescue G48C and A49G *CFTR* exon 12 mutant minigenes in their mini context. In the case of the CERESdel mini construct, 6 nt of CERES2 (GGATAC) was removed. The amplified RT- PCR product of the spliced/unspliced mRNAs are stained using ethidium bromide and run in an agarose 1.8% gel. Exon inclusion and skipping are shown by Ex12+ and Ex12-, respectively.

The results shown in Figure 2.9 demonstrate that both SF2/ASF and SRp55 consistently have a higher enhancing effect on the mini exon 12 inclusion levels than SC35,

suggesting that direct interaction provides an advantage over the well known generalized exon inclusion enhancing effect of SR proteins. Interestingly, however, the enhancement observed for the two mutants was not the same, with A49G being less responsive especially for SRp55 overexpression than G48C. Finally, it should be noted that deletion of the central CERES2 region also abolished completely the response of the mini-exon 12 to all SR protein overexpression, demonstrating that their action in the mini-exon context is mediated only through the CERES2 sequence. However, we haven't measured the expression level of our exogenous SR proteins after transfection due to the limitation of differentiating them from the endogenous protein in a westernblot analysis. This is why in the future tagging the exogenous SR proteins (i.e. Flag or Myc) and correlating them according to the exon inclusion level will be required to strengthen the evidence.

2.8 Identification of the domains of SF2/ASF involved in affecting *CFTR* exon 12 inclusion levels.

In addition to the experiments described above, in order to rule out the non-specific effect of SF2/ASF overexpression we also performed overexpression analysis with a series of SF2/ASF deletion mutants.

(A)
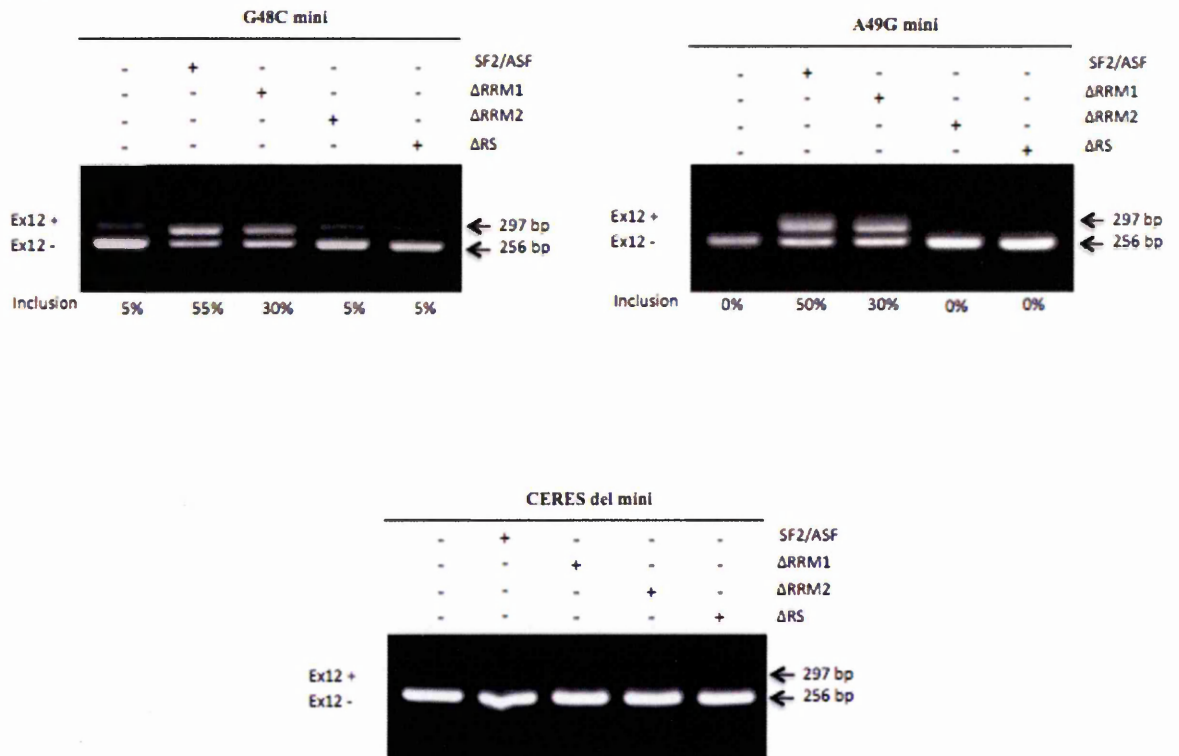
Figure 2.10: Effect of overexpression of Δ SF2/ASF in G48C and A49G mutation.

(A). Analysis showing the overexpression of wild-type SF2/ASF and a series of SF2/ASF mutants (ΔRRM1, ΔRRM2 and ΔRS) in the presence *CFTR* exon 12 mutant minigenes in their mini context. The amplified RT- PCR products of the spliced/unspliced mRNAs are stained using ethidium bromide and run in an agarose 1.8% gel. Exon inclusion and skipping are shown by Ex12+ and Ex12-, respectively.

In this experiment, mutants lacking either the RRM2 region (ΔRRM2) or the RS domain (ΔRS) could not enhance inclusion (Figure 2.10). The positive effect seen in case of SF2/ASF carrying the ΔRRM1 deletion suggested that only RRM2 was the RNA interacting domain of this protein in the case of *CFTR* exon 12. This is consistent with the previous observation where it has been shown that the RRM2 of the SF2/ASF is

74

responsible for its enhancing ability (Chiodi et al., 2004). Taken together, the results of this overexpression experiment further supports the specific role played by SF2/ASF binding to the CERES2 sequence.

2.9 Validating the role played by hnRNP factors in *CFTR* exon 12 splicing.

Several hnRNP proteins identified in our screening analysis (such as A1, A2, C and DAZAP1) are well known for being capable of modulating the pre-mRNA splicing process in many exons (Goina et al., 2008; Venables et al., 2008). In order to test their functional role in *CFTR* exon 12 splicing we first attempted a similar approach used for the SR proteins (overexpression). However, no satisfactory results were obtained due to their abundance in the nucleolus (data not shown). For this reason, in order to test the functional effects of the hnRNP interactors found in pulldown analysis, we performed individual siRNA-mediated knockdown of each of these well known hnRNP proteins.

(A)
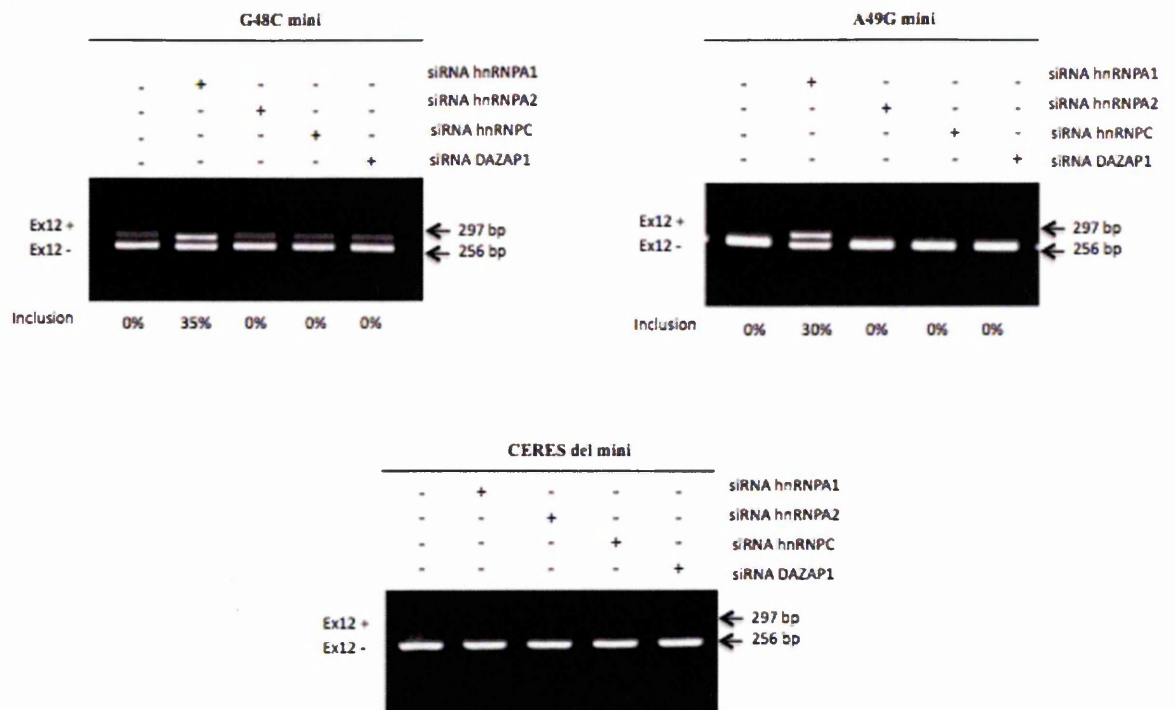
(B)
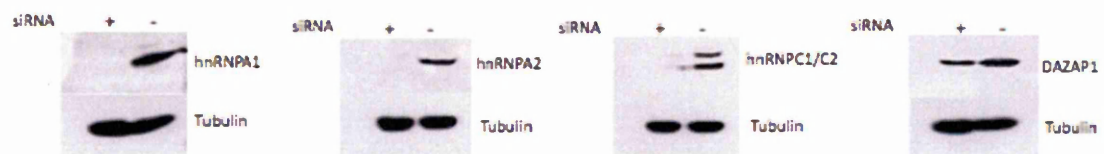
Figure 2.11: Effect of hnRNPs depletion in G48C and A49G mutation. Analysis showing that depletion of hnRNP A1 in HeLa cells rescues *CFTR* exon 12 in mini context, irrespective of the G48C and A49G mutations. HeLa cells were treated (shown as +) with siRNAs against hnRNP A1, A2, C and DAZAP1 and then were transfected with minigene plasmids. Luciferase siRNA treated cells have been used as a control **(A)** RT PCR of the spliced product on 1.8 % agarose gel stained with ethidium bromide. CERES2 Del minigene was also used for transfection as a control where the mutations are not present due to 6nt deletion **(B)** Endogenous levels of hnRNP A1, A2, C, DAZAP1 siRNA treated (shown as +) and luciferase treated (shown as -) HeLa cells, tubulin control were measured by western blotting 48 h after transfection.

As shown in Figure 2.11, the only siRNA knockdown that could rescue both G48C and A49G mini-exons inclusion was hnRNP A1. Most importantly, knockdown of this protein had no effect on the CERESdel minigene, demonstrating that rescue ability in the

76

mini exon 12 context was dependent on the presence of this sequence. However, no effect could be detected following the hnRNP A2, C2 and DAZAP1 knockdowns. This was rather surprising considering that the general role played by these proteins (especially hnRNP A2) in the regulation of splicing control. For this reason, simultaneous depletion of hnRNP A1, A2, C and DAZAP1 were carried out in different combinations to rule out any combinatorial effect between these proteins. However, none of these experiments could confirm their role in the splicing regulation of *CFTR* exon 12 (data not shown). It should be noted, however, that these results do not necessarily mean that only hnRNP A1 can modulate *CFTR* exon 12 splicing. It may well be, in fact, that some of these proteins may play an active role in the presence of reduced amounts of positive factors (ie. SF2/ASF or SRp55). Moreover, siRNA against DAZAP1 failed to deplete the protein completely in HeLa cells. Complete depletion of DAZAP1 or depletion in combination with hnRNP A1/A2 and C2 might have an effect on splicing of *CFTR* exon 12. Additional experiments should be carried out to further clarify this issue.

2.10 Recovery of the T40C and C52T inhibitory action through the add back of human and mouse flanking *CFTR* exon 12 sequences.

As discussed in Figure 2.3, both T40C and C52T substitution had no down regulatory effect in the mini exon context, leading to the hypothesis that either or both human flanking sequences were involved in mediating their effect on human *CFTR* exon 12. In order to better characterize this issue the human deleted sequences were added back selectively both in their human and mouse forms.

(A)

```
Human WT  GCAGTATACAAAGATGCTGATTTGTATTTATTAGACTCTCCTTTTGGATACCTAGATGTTTTAACAGAAAAAGAAATATTTGAAAG
Mouse WT  GCAGTATA AAAGATGCTGATTTGTAC TATTAGA TC CCTTTTGGATA CTAGATGTTT ACT GAA AA AA TATTTGAAAG
       A  GCAGTATACAAAGATGCTGATTTGTATTTATTAGACTCTCCTTTTGGATACCTAGATGTTTTAACA                        AAG
       B  GCAGTATACAAAGATGCTGATTTGTATTTATTAGACTC CCTTTTGGATACCTAGATGTTTTAACA                        AAG
       C  GCAGTATACAAAGATGCTGATTTGTATTTATTAGACTCTCCTTTTGGATA CTAGATGTTTTAACA                        AAG
       D  GCAGTATA AAAGATGCTGATTTGTAC TATTAGACTCTCCTTTTGGATACCTAGATGTTTTAACA                        AAG
       E  GCAGTATA AAAGATGCTGATTTGTAC TATTAGACTC CCTTTTGGATACCTAGATGTTTTAACA                        AAG
       F  GCAGTATA AAAGATGCTGATTTGTAC TATTAGACTCTCCTTTTGGATA CTAGATGTTTTAACA                        AAG
```

(B)
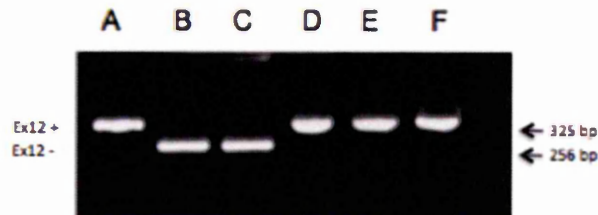
(C)
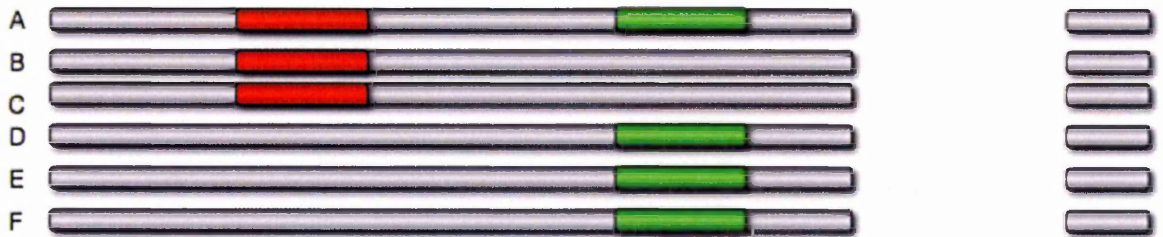
Figure 2.12: HeLa cell specific splicing assays of add back constructs (A to F). (A).
Schematic diagram of the wild-type and hybrid mouse and human *CFTR* exon 12
sequences used to construct a series of minigenes (labeled A to f) were used for 5'
regulatory sequence identification. Nucleotide differences in the mouse sequence with
respect to the human are boxed in red. Blank spaces in the alignment represent the
sequences removed form the exon. **(B).** Results of the transfection analysis of the
minigenes labeled A to F following transfection in HeLa cells. Exon inclusion is shown by
Ex12+ and skipping Ex12-. RT-PCR samples are stained with ethidium bromide and run on
a 1.8% agarose gel. **(C)** A schematic presentation of ESE and ESS distributions, where
green box presents ESE and red box presents ESS.

As shown in Figure 2.12, the wild type exon 12 constructs that contain the human
and mouse upstream regions vice-versa displays full inclusion (construct A). However,
when the upstream human sequence was inserted back at the presence of the T40C and

78

C52T mutations the inhibitory effect could be detected (constructs B and C respectively). No skipping could also be detected when the mouse upstream sequence was inserted back in the wild-type human mini-exon context (construct D). Interestingly, this inhibitory effect could also not be detected when the added back upstream sequence was the mouse *CFTR* exon 12 sequences but in the presence of the T40C and C52T mutations (constructs E and F).

(A)

G	GCA	TTAGACTCTCCTTTTGGATACCTAGATGTTTTAACAGAAAAAGAAATATTTGAAAG
H	GCA	TTAGACTCCCCTTTTGGATACCTAGATGTTTTAACAGAAAAAGAAATATTTGAAAG
I	GCA	TTAGACTCTCCTTTTGGATACCTAGATGTTTTAACAGAAAAAGAAATATTTGAAAG
J	GCA	TTAGACTCTCCTTTTGGATACCTAGATGTTTTAACAGAACAACAACTATTTGAAAG
K	GCA	TTAGACTCCCCTTTTGGATACCTAGATGTTTTAACAGAACAACAACTATTTGAAAG
L	GCA	TTAGACTCTCCTTTTGGATACCTAGATGTTTTAACAGAACAACAACTATTTGAAAG

(B)

(C)

Figure 2.13: HeLa cell specific splicing assays of add back constructs (G to H). (A). Schematic diagram of the hybrid mouse and human *CFTR* exon 12 sequences used to construct a series of minigenes (labeled G to L) were used for 3' regulatory sequence identification. Nucleotide differences in the mouse sequence with respect to the human are boxed in red. Blank spaces in the alignment represent the sequences removed form the exon. **(B).** Results of the transfection analysis of the minigenes labeled G to L following transfection in HeLa cells. Exon inclusion is shown by Ex12+ and skipping Ex12-. RT-PCR samples are stained with ethidium bromide and run on a 1.8% agarose gel. **(C)** A schematic presentation of ESE and ESS distribution, where green box presents ESE and red box presents ESS.

A similar situation to that observed with the upstream regions of *CFTR* exon 12 could also be observed with the downstream regions. In fact, Figure 2.13 shows that wild-type exon 12 sequences with the added-back human and mouse downstream region display full inclusion (constructs G and J, respectively). However, when T40C and C52T are inserted back the inhibitory effect could be detected only in the constructs with the human but not with the mouse downstream sequence (compare constructs H-I with K-L).

Taken together, these results suggest that the human and mouse flanking regions have different splicing regulatory properties. In human *CFTR* 12, both upstream and downstream flanking sequences are predominantly inhibitory, whilst the mouse upstream and downstream sequences seem to enhance exon recognition. This assumption is graphically presented in both Figure 2.12 and Figure 2.13 C.

(A)

```
M   GCA                  TTAGACTCTCCTTTTGGATA CTAGATGTTTTAACAGAA AAGAAATATTTGAAAG
N   GCA                  TTAGACTCTCCTTTTGGATA CTAGATGTTTTAACAGAA AA AAATATTTGAAAG
O   GCA                  TTAGACTCTCCTTTTGGATA CTAGATGTTTTAACAGAA AAGAA TATTTGAAAG
```

(B)
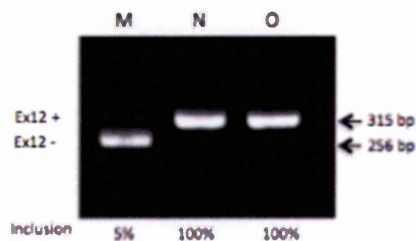
Figure 2.14: HeLa cell specific splicing assays of add back constructs (M to O). (A). Schematic diagram of the hybrid mouse and human *CFTR* exon 12 sequences used to construct a series of minigenes (labeled M to O) in aspect of identification of the specific nucleotide responsible ESE activity. Nucleotide differences in the mouse sequence with respect to the human are boxed in red. Blank spaces in the alignment represent the sequences removed form the exon. **(B).** Results of the transfection analysis of the minigenes labeled M to O following transfection in HeLa cells. Exon inclusion is shown by Ex12+ and skipping Ex12-. RT-PCR samples are stained with ethidium bromide and run on a 1.8% agarose gel.

With regards to the upstream *CFTR* exon 12 regions in humans and mouse it should be noted that the splicing-enhancing ability of mouse CC at positions 28 and 29 with respect to the TT nucleotides in human positions 28 and 29 was previously described (Pagani et al., 2005). Indeed, in the presence of the C52T substitution it was already reported that a T28C substitution alone could recover exon inclusion up to 90%, thus disrupting the silencing effect of human upstream sequence. Therefore, in this case, the enhancing ability of the mouse upstream sequence was well characterized.

In order to better characterize the mouse polypurinic GAAGAACAAG motif present in the downstream region (Figure 2.14) we then performed a further mutation analysis. In these experiments, the results that could be noted were that the majority of substitutions that tended to restore the mouse sequence could successfully resist the inhibitory action of the C52T substitution (constructs N-O as opposed to M).

2.11 Evidence of ESEs in both of flanking sequence.

In order to further validate the hypothesis that the flanking upstream and downstream mouse sequences act as enhancers in their native context we cloned the entire mouse *CFTR* exon 12 sequence in the pTB minigene system (Figure 2.15).

(A)

(B)   **(C)**
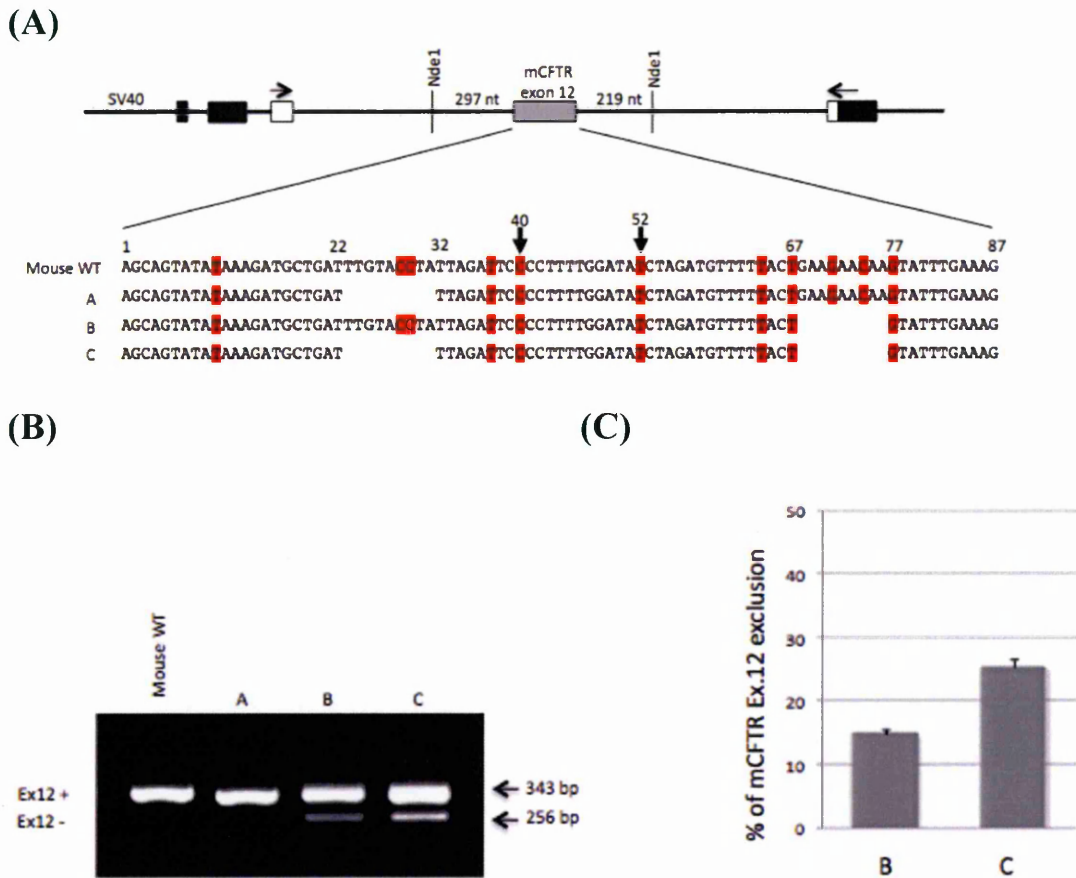
Figure 2.15: HeLa cell specific splicing assays of mouse *CFTR* ex 12 and ESEs deleted constructs. Analysis of predicted ESEs in Mouse *CFTR* exon 12. **(A)** Mouse sequence is aligned along with Human *CFTR* Ex 12. Blank space in the alignment is for sequence deleted in the minigene. Sequence A, B and C is nucleotide deletion in Mouse *CFTR* ex 12 spanning from 21 to 30, 66 to 75 and both sequence deletions respectively. **(B)** Mouse *CFTR* exon 12 minigenes (A, B and C) were transfected in HeLa cells for splicing assay. Amplified RT PCR products of the splicing assay is presented on 1.8% agarose gels. Labels on the top of the gel (A, B and C) represents its correspondent sequences as presented at the top (Figure 2.15 A). Exon inclusion is shown by Ex12 + and skipping Ex12-. **(C)** Three independent experiments were quantified and presented.

In the mouse *CFTR* exon 12 sequence then the upstream (23-31nt) and downstream (68-76nt) were deleted either separately or in combination (Figure2.15 A mutants A-C). The ESE activity at the functional level reported in Figure2.15 B showed that deleting only the upstream sequence (mutant A) had no effect of mouse *CFTR* exon 12 inclusion levels. On the other hand, deletion of the downstream sequence (mutant B) resulted in approximately 15% exon skipping. Interestingly, if both regions were deleted at the same time (mutant C), the levels of exon skipping increased to 25%, indicating that also the polypurinic downstream sequence can function as an ESE once the upstream ESE sequence is absent.

2.12 *Trans*-acting factors binding to the human and mouse 17-38 and 63-81 sequences.

It is well known that the ESEs exert their effect by interacting with specific *trans acting* factors (Graveley, 2000). Based on the deletion mutants in mouse *CFTR* exon 12, it was therefore expected that the mouse flanking sequences would bind a different set of protein compared to humans. For this reason, we selected the 17-38 and 63-81 mouse and human regions (Figure2.16 A) to perform pulldown assays (as previously described for the CERES2 region) with synthetic RNA oligos.

(A)

(B)
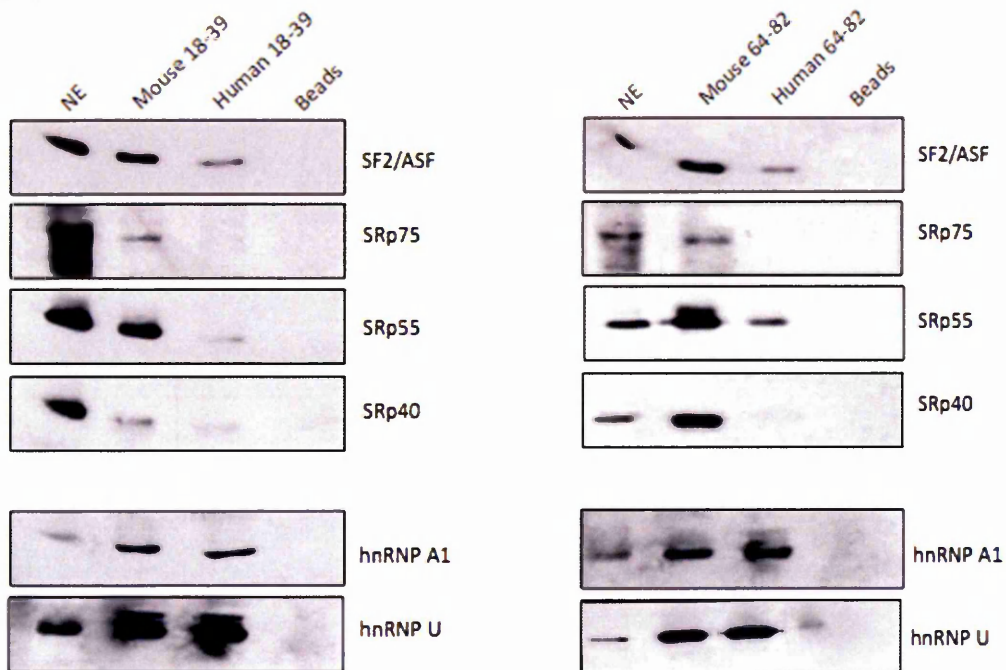
Figure 2.16: Pulldown analysis for *trans-acting* factors at the flanking SREs of mouse and human *CFTR* ex 12. Comparative analysis of Human and Mouse splicing factor affinity. **(A)** Both Mouse and Human synthetic RNA Sequences are positioned under whole exon sequence. Mouse distinct nucleotides compared to human are labeled with red boxes. RNA oligos were immobilized with adipic acid dehydrazide-agarose beads and used for affinity purification. **(B)** Both Human and Mouse *CFTR* exon 12 sequence spanning from 17 to 38 were comparatively analyzed for splicing factors. Harvested proteins were targeted for SR proteins (SRp 75, SRp 55, SRp 40, SF2/ASF) in western blot analysis. Membrane was later probed with hnRNP A1 and U antibody. **(C)** Similar comparative analysis for splicing factors was done for both Human and mouse *CFTR* exon 12 sequences spanning from 63 to 81 and normalized against hnRNP A1 and hnRNP U instead of TDP43.

The Western blot analysis to check for SR protein binding showed that both mouse sequences could bind SF2/ASF, SRp75, and SRp55 more efficiently than the respective human sequences. In addition, mouse nucleotides 64-82 are also capable of binding SRp40

whilst the human 64-82 sequences could not (Figure 2.16 B). With regards to hnRNP protein binding, the recognition with antibodies against hnRNP proteins A1 and U showed that both these factors could bind equally well with the mouse and human sequences, substituting TDP-43 as a pulldown and loading control (these oligos were synthetic sequences and lacked the UG-terminal tail).

2.13 Possible *Trans*-acting factor binding to the human 17-38 ESS sequences.

Unfortunately, selective candidate approach (pulldown followed by western blot analysis) did not allow us to detect *trans-acting* factors that might be responsible for the ESS activity in human 18-39 and 64-82 sequences. The reason probably lies in the fact that using this approach only a fraction of potential candidates can be tested for, as there are many more potential known or unknown proteins that can be interacting in this region. In addition, the pulldown assays were checked in a comassie stained gel to look at the complete protein profile but unfortunately, again, no differences could be detected (data not shown). In order to find the *trans acting* factor binding in both human ESSs, therefore, the synthetic RNA oligos were labelled with p^{32}, incubated in NE at splicing condition, cross linked by UV, and finally separated in SDS-PAGE gel (without being subjected by RNAse treatment as not to remove the 5' radioactive phosphate). Using this experiment, we therefore wanted to check all direct interactions of the 18-39 and 64-82 sequences with nuclear extract proteins.

In our UV crosslinking experiments there were no visible difference in case of 64-82 sequences (data not shown). However, a band of 32kD was detected in case of 18-39 sequence. This suggests that there is a factor of more or less 25 kD (after deducting 7 kD of oligo size) interacting with human ESS directly. Competition with unlabeled mouse oligos confirmed the specificity of the factor that has not yet been identified.

(A)
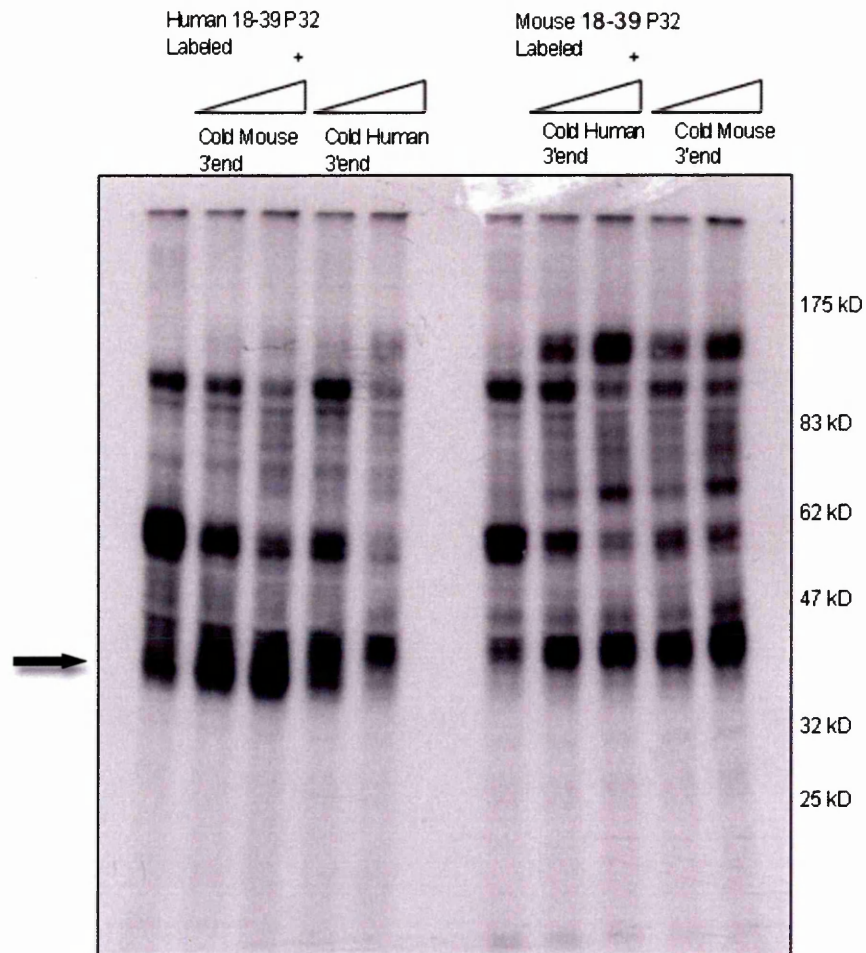
Figure 2.17: UV crosslinking analysis of human and mouse 17-38 sequences. (A) UV crosslinking analysis of interacting proteins, compared between human and mouse *CFTR* Ex 12 sequence (18-39 nt). Used Mouse and Human synthetic RNA Sequences are positioned under whole exon sequence. Mouse distinct nucleotides compared to human are labeled in red box. Both Mouse and Human synthetic RNA oligos were labeled with γ-^{32}P-ATP, incubated in NE and separated in a 10% SDS gel.

3 Discussion

Historically, in genotypic screening studies, the pathologic effect of an apparently benign polymorphism such as codon neutral variations that did not change the amino acid sequence were hardly assessed. Indeed only small subset of apparently harmless SNPs were referred to as disease-causing mutations but these were only based on phenotypic correlation. However, the functional impact of these inherited variations within the population and their potential association with the disease often remained unclear. More recently, it has been shown that apparently harmless nucleotide changes can substantially affect a whole range of different regulatory pathways such as interaction with microRNAs binding sites, RNA secondary structures, and especially splicing regulatory elements. In particular, with regards to splicing, it is now an obvious consideration that if an apparently silent nucleotide change occurs in a splicing regulatory region, it is then likely to affect the splicing process of the exon by enhancing or inhibiting its inclusion in the final mRNA. At the molecular level, the effects of this mutation can be explained either by the disruption of a splice site or creation of a cryptic splice site or by creating/destroying the binding site for an auxiliary *trans-acting* factor (Cartegni et al., 2002). In our lab, these kinds of investigations have been performed in the past by focusing on alternative splicing events described to occur in exons 9 and 12 of the *CFTR* gene, in consideration its importance in the development of cystic fibrosis.

Cystic fibrosis is the most frequent genetic disease within the Caucasian population. In particular, the inactivation of a highly conserved region of the NBD1 of the CFTR protein due to missense mutations (encoded also by exon 12 sequences) has been previously described to be a causative agent of CFTR (Delaney et al., 1993). Interestingly, among all the disease-causing mutations in *CFTR* exon 12, several have been shown to induce

skipping of the exon through alterations in the splicing process (Hull et al., 1993; Strong et al., 1992; Zielenski et al., 1995). In this study, we have combined the functional study of *CFTR* exon 12 natural pathogenic mutations on the splicing process and complemented the results with a comparison between human and mouse *CFTR* exon 12 sequence. This sort of combinatorial analysis has facilitated not only the exploration of the splicing regulatory elements involved in the pathogenic mutation, but it has also provided an overview of the dynamic interplay of various splicing regulatory elements during the course of evolution.

3.1 Regulation of CERES2.

So far, several checkpoints have been described in literature for splicing regulation and especially the extensive role played by ESEs and ESSs sequences (Chen and Manley, 2009). In splicing, overlapping ESEs and ESSs are often illustrated as two independent functional events that act in concert rather than considering them as a potentially single sequence. Using systematic site directed mutagenesis experiments, we have previously identified a new kind of exonic regulatory element in *CFTR* exon 12 and described as a distinct class of *cis* element (Pagani et al., 2003). This element was designated as CERES (for Composite Exonic Regulatory Element of Splicing). Elements of this kind have the characteristic of possessing overlapping enhancer and silencer properties. In practice, what this means is that mutation of single adjacent nucleotides (or in some cases even within the same nucleotide position) could either behave as having an enhancing or silencing effect on exon inclusion. In keeping with this complexity, splicing outcomes could not be predicted by available computer-assisted analysis of potential SR binding sites (ESEfinder) or enhancer/silencer elements (Rescue-ESE), further highlighting the functional distinction of CERES elements from "pure" enhancer and silencer sequences. Another distinguishing characteristic of the CERES elements identified so far seems to be their strong dependence on the context for their proper function and, accordingly, the composite characteristics of these elements may also extend to the flanking nucleotides.

In human *CFTR* exon 12, two such sequences have been previously characterized at the *cis-acting* level: CERES 1 (spanning nt. 12-18 of this exon) and CERES 2 (spanning nt. 47-52). However, in these studies the characterization of these elements was limited to mapping their *cis-acting* sequences. Therefore, in order to understand the possible molecular mechanism of CERES mediated splicing regulation we carried out an analysis of the *trans-acting* factors that bind the CERES2 element localized in *CFTR* exon 12. Initially, the focus was placed on characterizing its binding properties with regards to the most common splicing regulators, and in particular to those belonging to either the SR or the hnRNP class of *trans-acting* factors. The RNA affinity purification assay shows that in wild type condition, human *CFTR* exon 12 CERES2 can bind to a considerable number of these factors, which may be rather surprising considering that the CERES2 comprises a very short stretch of RNA sequence (<10 nucleotides). The protein binding profile included SF2/ASF and SRp55 within the SR protein family, specific binding could be detected for hnRNP A/B family members with regards to the hnRNP proteins. Interestingly, differential interaction was identified for SF2/ASF and hnRNPC2 in case of pathological missense mutations (G576A /G48C and Y577F /A51T) and a synonymous substitution (A49G) that were already known to affect *CFTR* exon 12 inclusion levels. From a basic RNA binding protein point of view this finding highlights the great flexibility provided by RRM motifs that can recognize a few specific bases at selected positions using their main chains and then use side-chain interactions to stabilize binding (Auweter et al., 2006). This probably explains why the CERES2 sequence rather than functioning as a binding site for a single protein only can function as a kind of aggregation site for many of these SR/hnRNP factors. Evidence of a functional interaction was confirmed only for SF2/ASF, SRp55, and hnRNP A1. However, it should be noted that these experiments were performed in a severely reduced context (mini-exon 12) and that in a more natural setting many of the additional binding of proteins might be able to play a role. Moreover, using our approach we have tested only the most common splicing regulatory proteins and hence cannot rule out the

presence of additional yet unknown factors that might also contribute to define/hinder *CFTR* exon 12.
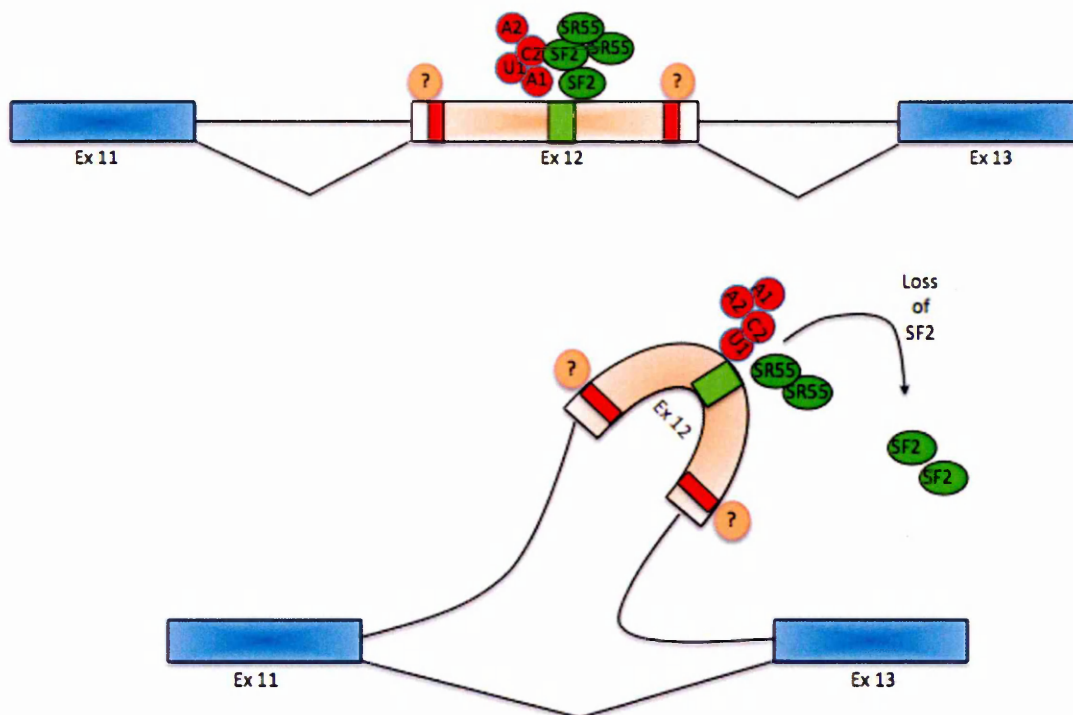
Figure 3.1: Regulation of *CFTR* exon 12. Schematic representation of wild type *CFTR* exon 12 splicing regulation is shown at the top. CERES2 element in exon 12 seems to be a splicing-factor rich region in which relatively many different proteins are presumably competing with each other on a very narrow stretch of sequence. Mutation into CERES2 (green box) that causes loss of SF2/ASF, in effect, increases the antagonistic effect of hnRNPs locally. On the other hand two ESSs (red box) present in flanking sequence inhibit to this exon definition. In total, the entire exon 12 sequence participate in its definition.

Taken together, our results suggest that it is this crowding together of many different factors (both positive and negative in terms of their effect on splicing) may explain why single-point substitutions within these CERES sequences have such an unpredictable effect on the exon recognition level. Therefore, and on a slightly wider scale, the CERES sequence is a situation similar to what has already been found in several small exons, such as *SMN* exon 7, where trans acting factors (SF2/ASF and hnRNP A1) binding to the same exonic region contribute to exon inclusion/skipping (Cartegni et al., 2006; Kashima et al., 2007). Actually, this particular layer of A1 and SF2 competition in *SMN 7*

is collaboratively effective with the support of adjacent downstream ESE, which exerts its effect by interacting with Tra2 β directly and then further interaction of other proteins with *Tra2 β* (Hofmann et al., 2000; Hofmann and Wirth, 2002; Young et al., 2002). Indeed, for *SMN* exon 7 it has been hypothesized the existence of an Extended Inhibitory Context (Exinct) that is caused by overlapping regulatory motifs not all of which have still been characterized in depth (Singh et al., 2004). Other examples of very complex systems include the human *c-src* exon N1 (Chou et al., 2000; Sharma et al., 2005), *CD44* exon v5 (Cheng and Sharp, 2006; Matter et al., 2002; Matter et al., 2000), *HipK3* "T" exon (Venables et al., 2005), and chicken *cTNT* exon 5 (Ladd et al., 2005) where numerous factors have been shown to participate in splicing regulation in close spatial proximity to each other. In conclusion, therefore, it is very probable that there are numerous CERES-type splicing regulatory sequences existing in human exons. Future research might therefore benefit from the identification of these elements and comparative analyses of their behavior/protein binding profiles. The results, besides enhancing our current understanding of the splicing process will also be very useful in improving bioinformatics approaches that attempt to predict splicing outcomes.

3.2 Comparative analysis of SREs at the flanking sequence of Human and Mouse *CFTR.*

It is clear that alternative splicing is a balance between positive and negative elements once the entire exon definition is taken in account (Zhang et al., 2009). Breaking exons into small segments for plainness of exon definition may go against its entire natural counterparts. Moreover, search of the SREs existence of these numerous SRE sequences co-existing together on the same exon has very important implications with regards to evolutionary constraints in codon composition.

Our results have shown very clearly that the inhibitory effect on human *CFTR* exon 12 inclusion levels of some synonymous nucleotide substitutions, T40C and C52T,

naturally occurring in the mouse sequence, can be explained by the distribution of different splicing regulatory properties of the human and mouse flanking exonic sequences. In particular, human flanking sequence have been shown to contain ESS sequences in both the 5'ss and 3'ss flanking regions whereas in mouse these sequences are ESEs. This finding is also supported by previous work where it was shown that the presence of two mouse substitutions (28C and 29C) near the 3'ss alone can compensate for the loss of CERES2 caused by mutations like C52T and T40C (Pagani et al., 2005). In addition, in this study we have shown that both the mouse ESEs interact with some of the SR proteins (SRp55 and SF2/ASF in 17-38 and SRp 40, 55 and SF2/ASF in 63-81), a predominant nature of ESEs. Presently, although we couldn't pull out any specific *trans-acting* factor for human ESSs but UV crosslinking experiments showed a direct interaction of an unknown factor at the 17-81 nucleotides stretch. In order to identify the trans acting factor as well as overcome this technical shortcoming, a mass spectrometry base sequencing of the whole harvested proteins can be carried out after the pulldown (Schmidt et al. 2010).

Figure3.2: Sequence comparison for SREs in different species.

Schematic comparison and conservation of the splicing regulatory elements (ESEs in green, ESSs in red, CERES1 in violet and CERES2 in blue) in the Mouse, Ground Squirrel, Guinea pig, Rabbit, Cow, Pig, Horse, Lemur and Human *CFTR* exon 12 sequences based on our functional analysis. The comparison is done here is by correlating the functional data of this thesis, as well as the functional data available in the literature (Pagani et al., 2005). In this comparison CERES1 was conserved in all the nine species whereas CERES2 carries C52T substitution in case mouse, cow and rabbit. The ESE at the upstream flanking sequences of the exon is present in all species apart from human. However, in case of guinea pig the C28T could be ESE or a neutral sequence because the T29C substitution in human exonic context can compensate with mutation like C52T (Pagani et al., 2005). Almost all the non synonymous changes (amino acids labeled in yellow) compared to human are centered in the downstream polypurinic ESE present in the mouse. These synonymous and non-synonymous substitutions created ESSs in guinea pig, rabbit, horse, lemur and human, whereas in cow, pig and ground squirrel possibly neutralized the ESE or created an ESS. However, the comparison shows that the *CFTR* exon 12 requires at least one enhancer to get included in the transcript. Moreover, nucleotide substitution can take place, as long the there is enough signal exon definition.

From an evolutionary point of view, by comparing the ESE, ESS, and CERES elements in different species, some important considerations can be made. Looking at the sequence comparison in Fig. 3.2, in fact, it can be hypothesized that the creation of the

CERES2 element in *CFTR* exon 12 has relieved the selection pressure to keep the two ESE elements loosely localized in the 17-38 and 63-81 flanking regions that are present in the mouse sequence. Only after the creation of CERES2 element these two regions could thus accumulate nucleotide substitutions that either weakened these elements (ie. in the Ground squirrel and Guinea pig 63-81 region, see Figure 3.2) or even changed them to functionally silencer sequences (in human 17-38 and 63-81). It is also interesting to note that a comparison of these sequence elements also in other species does not contradict the conclusions we have drawn from mouse versus humans. In fact, for example, both the cows and rabbit sequences that do not contain the CERES2 element have absolutely conserved ESE sequences. Of course, additional experiments will need to be performed before we can draw firm conclusions on this issue. Nonetheless, in our work, we show that an integrated analysis of *cis-* and *trans-acting* factors binding to exonic elements can also provide a substantial wealth of information on potential evolutionary mechanisms.

3.3   Selection pressure on CFTR exon 12 sequence due to synonymous and non synonymous substitutions.

The consequence of sequence evolution in splicing is usually considered more with the exonic regulatory elements rather than the tri partite splicing signals (Fairbrother et al., 2004; Fairbrother et al., 2002; Wang et al., 2004; Zhang et al., 2009). Possibly the reason is that the splicing in metazoans is more dependent on the exonic definition rather than the Intronic signal and tends to keep the biased splice sites (although this is not exclusively constrained) and several of the trans acting factor have equally evolved over time (Ram and Ast, 2007; Warnecke et al., 2008). Moreover, dependency on the exonic sequence is a way of avoiding large sets of amino acid changes (Warnecke et al., 2008). Several bioinformatic and comparative genomic approaches have been carried out to understand the role of exonic regulatory elements in evolutionary mechanisms (Chamary et al., 2006; Parmley et al., 2006; Parmley and Hurst, 2007b). Instead of being fully dependent of *in silico*

predictions, here in this study we took the initiative of understanding the effect of synonymous and nonsynonymous substitutions in exon selection by a wet lab approach in *CFTR* exon 12 (Fairbrother et al., 2004; Parmley and Hurst, 2007a). However, our result suggests that new genomic variants that are selected need to be compatible with the splicing process. This is especially true also for changes that should be neutral with regards to the coding capacity.

Selection towards codon optimization has been shown to act on silent sites in bacteria since a long time. Around 47% of silent substitutions in bacteria can be accounted for by codon usage whilst 14% by their distance from Ori C (Sharp et al., 1989). In bacteria, the basis of this selection appears to be the optimization of translation that is largely achieved by the use of codons for particular tRNAs that are most/least abundant or with certain binding characteristics. Of course, this may not be the only reason, and the potential also exist for silent selection to act on other features such as DNA folding or RNA secondary structures which may account for the remaining 39% of cases. In mammals, the evidence for any of these kinds of selection with respect to codon bias is more difficult to obtain and a few studies have tended to downplay this possibility (Eyre-Walker, 1991; Smith et al., 1999) with the notable exception of the maintenance of conserved RNA secondary structures at the start of mammalian genes (Nick and Hurst, 1999). However, a frequently used method for investigating sequence evolution was based on the Ka/Ks ratio (non synonymous/ synonymous) due to their correlation with protein function. Ka/Ks ratio in CFTR exon 12 is relatively high because synonymous changes are expected to be translationally silent. Comparative analysis within human and mouse CFTR exon 12 shows that a high large number of synonymous differences (9 positions within 87nt) are present but not all of them are silent in aspect of splicing regulation. This sort of inappropriate Ka/Ks ratio has been also reported for alternatively spliced BRCA1 (Hurst and Pal, 2001). In fact, alternative exons in the alternatively spliced genes show higher Ka/Ks ratio, which indicates the importance of local segments of the gene (Xing and Lee, 2005). Moreover,

this observation is consistent with the genome wide analysis showing that most of the synonymous substitutions are ESE neutral but the substitutions that create or remove ESEs are less frequent (Carlini and Genut, 2006). Besides, in sequence evolution synonymous substitutions in ESEs were either counted as a loss of functional ESE or gain of a new ESE (Carlini and Genut, 2006; Fairbrother et al., 2004; Parmley et al., 2006; Parmley and Hurst, 2007b). Our functional analysis adds a caution to this fact by showing that two synonymous changes (28th and 29th nt) in fact created ESS. It is true that ESEs are more vulnerable to synonymous changes in the alternatively spliced exon, which might bring in selection pressure on the exon (Parmley et al., 2006). Our result suggests synonymous changes can also be either disruptive to the ESSs or can create new ESSs.

Non synonymous substitution are largely expected to be deleterious because of their ability to affect proteins (Fay and Wu, 2003), although 20-45% of the non synonymous changes have been suggested as advantageous changes (Bierne and Eyre-Walker, 2004). In fact protein evolution is more concern with the factors like expression, protein importance, pleiotropy, structural constrain rather than genomic influence. Within human and mouse CFTR exon 12 there are four amino acid changes, out of which three amino acids are neutral in their side chain charge. One possibility could be that over time these amino acids substitution within neutral charge optimized the metabolic efficiency rather than protein structural optimization. Indeed, in bacteria this type of selection pressure based on metabolic efficiency of the amino acid exits (Rocha and Danchin, 2004). Interestingly a negatively charged glutamic acid in mouse is replaced by positively charged lysine (E584K) in human might have given a better structural stability or function to human CFTR, in fact this change is also conserved in other species we tested for, except mouse (see fig. 3.2). This change might be a little contribution to a large protein like CFTR when the whole protein is considered but one has to wonder that the way the splicing regulation and protein evolution is interlinked. Moreover, this picture of sequence evolution becomes more complex once concomitantly the evolution of trans acting factor is accounted (Ram

and Ast, 2007).

However, the amino acid change for fine-tuning of the protein might be gradually acclimatized under positive selection but before that persistence of the splicing regulatory sequence is essential for exon inclusion. This means, the amino acid change must not cause the exon skipping which will create a non-functional protein, otherwise they will be subjected to negative selection. Novel amino acids can be integrated in the protein for better function only if the selection of the exon is preserved.

3.4 Conclusion.

In conclusion, it is clear from this work that the *CFTR* exon 12 sequence is literally covered by regulatory elements that we probably still consider (rather mistakenly) as separate elements. Proof of this is the observation that the activity of many of these splicing regulatory elements (especially the CERES) cannot be exported in different contexts (Pagani et al., 2003c). Indeed, our results point towards a situation where in exons like *CFTR* exon 12 we should virtually consider every nucleotide as potentially capable of affecting splicing inclusion levels.

The analysis carried out in my work focussed only on a selected number of *trans acting* factors interacting with the exonic sequences with respect to the mutations in a minigene system *in vitro*. Our "selective candidate" approach to study the trans acting factors for the CFTR exonic sequence limits our view on the other splicing factors that might play role in splicing of the exon. Indeed, the effect of many other factors on the intronic regulatory sequence has not been analyzed. This is why in the future a possible follow up of the study might consider a mass spectrometry based proteome analysis to identify all the binding factors. Also the RNA-based pull-down analysis I have carried out in this thesis has some technical pitfalls because the nuclear extracts from HeLa cells often doesn't reflect the exact binding conditions for the splicing factors since most of the splicing factors interact with the RNA with a very low affinity. Therefore, any variation of the concentration of the factors in the extract may change their binding affinity *in vitro*. In

order to overcome this limitation, a RNA IP in different cell lines can be performed against the known factors we have already defined.

Splicing is extremely cell-type and tissue-type dependent. Therefore it will be important to extend our observations in a more physiological context (Barash et al., 2010). Indeed, several of the splicing factors and their posttranslational modifiers are expressed differently in different tissues. In this thesis we have performed cell based splicing assays that constrained our understanding of the effect of mutations *in vivo*. In order to overcome this limitation our analyzed mutations may be studied in an animal model. Previously, the effect of the SMN splicing mutation has been successfully studied in a mouse model. However, this approach for studying CFTR will be rather challenging due to the fact that mouse has an additional Calcium dependent ion channel. This calcium dependent ion channel can compensate the conductance regulatory mechanism in the cell in absence of chloride channel. Possibly a double knock out mouse model will overcome this limitation. On the other hand the effect of general splicing factors (i.e. SRs and hnRNPs) in an animal model has not been carried successfully till now. There are several reason for this but in general a splicing factor interacts with RNA with low affinity and depletion or low expression of a splicing factor in an animal model often shows no effect because the function of that particular *trans acting* factor is replaced by other factors with same functionality. Nevertheless, animal models with the CFTR exon 12 mutations will be useful for therapeutic purposes.

From a clinical point of view, increased importance will have to be given at analyzing RNA transcripts directly from patient tissues or, routinely, through a minigene based systems that will mimic this kind of global splicing regulatory networks (Baralle et al., 2009). This will be essential for further therapeutic development. Many approaches have been explored so far and many more can be envisioned to modify the splicing pattern of a mutant pre mRNA or eliminate an mRNA that bears a disease-causing mutation to achieve therapy. Increasing knowledge of RNA biology and chemistry is stimulating efforts to

target the RNA itself or the splicing and translational machinery as entry points for therapeutic intervention. For example strategies like antisense oligonucleotides, antisense snRNAs, trans-splicing, and small molecules have been proposed to be effective for diseases like DMD, SMN, CF and ALS (Bonetta, 2009). Although, our incomplete knowledge of splicing regulation limits our understanding of several factors, such as individual disease susceptibilities or therapeutic responses in the presence of local nucleotide polymorphic differences (Cooper et al., 2009; Wang et al., 2008), a complete interpretation of the splicing code in future will pave the way to the personalization of the drug.

Secondly, this increased awareness will be useful for the development of novel bioinformatics methods aimed at predicting splicing outcomes that, until now, have been primarily focused at considering enhancer and silencer elements as distinct entities, with rather limited success (Hartmann et al., 2008; Houdayer et al., 2008). Moreover, the tissue specific splicing prediction has been poorly understood. Recently a splicing code predicting approach has been carried out to clarify this issue by analyzing transcripts from different mouse tissues and considering all possible factors involved in tissue specific splicing (Barash et al., 2010). However, this newly developed code prediction is rather limited due to the fact of evolutionary conservation of alternative-splicing regulation (estimated to be around 20% between humans and mice) and opens up the question of species-specific codes (Tejedor and Valcarcel).

Finally, our results should help to gradually shift our view of splicing from one in which exon inclusion levels are viewed as the straightforward algebraic sum of enhancer/silencer elements to one where exons are rather considered as an integrated unit, where silencing and enhancing functions may functionally overlap to a degree that has often been underestimated.

4   Material and Methods

4.1   Chemical reagents.

Commonly used chemicals were purchased from Sigma Chemical Co., Merck, Gibco BRL, Boehringer Mannheim.

4.2   Standard solutions.

Frequently used solutions were made according to the following recipes:

a) TE: 10 mM Tris-HCl (pH 7.4), 1 mM EDTA (pH 7.4)

b) PBS: 137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO4, 1.8 mM KH_2PO4, pH 7.4

c) 10X TBE: 108 g/l Tris, 55 g/l Boric acid, 9.5 g/l EDTA

d) 6X DNA sample buffer: 0.25 % w/v bromophenol blue, 0.25 % w/v xylene cyanol FF, 30 % v/v glycerol in H_20.

e) 10X protein sample buffer: 20 % w/v SDS, 1 M DTT, 0.63 M Tris-HCl (pH 7), 0.2% w/v bromophenol blue, 20 % v/v glycerol, 10 mM EDTA (pH 7).

4.3   Enzymes.

Restriction enzymes were purchased from New England Biolabs, Inc (NEB). DNA modifying enzymes such as Taq Polymerase, DNaseI RNase free, and T4 DNA ligase were obtained from Roche Diagnostic and NEB. Klenow fragment of *E. coli* DNA polymerase I and T4 polynucleotide Kinase were from New England Biolabs, Inc. RNase A was purchased from Sigma Chemicals Ltd. A 10 mg/ml solution of RNase A was prepared in sterile water and boiled for 10 minutes to reduce the trace amounts of DNase activity. All enzymes were used following manufacturer's instructions.

4.4 Synthetic DNA oligonucleotides.

Synthetic DNA oligonucleotydes were purchased from Sigma-Genosys Ltd and MWG Biotech (Germany).

4.5 Radioactive isotopes.

Radioactive [γ-^{32}P]ATP nucleotides (each with 10 mCi/ml) 3000 Ci/mmol were supplied by Amersham U.K. Ltd.

4.6 Nucleotides .

Nucleoside-5'-Triphosphate (ATP, CTP, GTP, UTP; 100 mM each and dATP, dCTP, dGTP, dTTP; 100mM each). m7G(5')ppp(5')G cap (7-Monomethyl-diguanosine Triphosphate). Enlisted nucleotides were provided from Promega (USA), m7G (5')ppp(5')G cap was supplied by New England BioLabs (UK).

4.7 Bacterial culture.

The *E. coli* K12 strain DH5α was transformed with the plasmids described in this study and used for their amplification. Plasmids were maintained in the short term as single colonies on agar plates at 4 °C but for long term storage they were kept on glycerol stocks made by adding sterile glycerol to a final 30% v/v concentration to liquid bacterial cultures. Glycerol stocks were stored at –80°C. When necessary, from the glycerol stocks an overnight culture of bacteria was grown in Luria-Bertani medium [LB medium: per litre: 10 g Difco Bactotryptone, 5 g Oxoid yeast extract, 10 g NaCl, (pH 7.5)]. LB media was sterilized before use by autoclaving. When appropriate, ampicillin was added to the media at a final concentration of 200 µg /ml.

4.8 DNA preparation .

4.8.1 Small scale preparation of plasmid DNA from bacterial cultures.

Rapid purification of small amounts of recombinant plasmid DNA was performed with the method previously described by Sambrook (Sambrook et al., 1989). Briefly, alkaline lysis of recombinant bacteria was performed by resuspending the bacterial pellet in 200 µl of ddH$_2$O; 150 µl of solution II (0.2 M NaOH, 1 % w/v SDS) were then added and the contents mixed by inversion. 250µl of solution III (3 M potassium acetate pH 5.2) were then added and the contents mixed by inversion. The bacterial lysate was then centrifuged in an Eppendorf microcentrifuge at maximum speed and the supernatant transferred to a new tube. An equal volume of 1:1 v/v phenol:chloroform solution was added to the supernatant. The tube was then vortexed and centrifuged as above. The aqueous phase containing the DNA was transferred to a new tube. An equal volume of chloroform was added to the supernatant. The tube was then vortexed and centrifuged as above. The aqueous phase containing the DNA was then recovered and the DNA pelleted by ethanol precipitation. The final pellet was resuspended in 50 µl of ddH$_2$O and 5 µl of such preparation were routinely taken for analysis by restriction enzyme digests. Along with it, very often Promega MiniPrep kit was used according to manufacturer's protocols.

4.8.2 Large scale preparations of plasmid DNA from bacterial cultures.

In order to get a good amount of plasmid, a single colony was grown in 50 ml of TB medium at 37°C for overnight. Following morning, bacterial cells were pelleted down at 4000rpm for 5 minute. Then JetStar purification kit (Genomed) was used according to the manufacturer's instructions.

4.8.3 Genomic DNA isolation.

In order to get clean genomic DNA for further amplification in PCR reaction, cell were treated with lysis buffer (0.32M sucrose, 10mM Tris HCl p 7.5, 5mM MgCl2 and 1% Tritorn X100). After lysis supernatant were incubated with proteinase K for over night.

Following morning a phenol chloroform purification step was carried out using standard protocol.

4.9 RNA preparation from cultured cells.

Cultured cells were washed with PBS and then 500 μl of TRIreagent, purchased from Ambion was added for every p35 mm plates. Then, 100 μl of chloroform was added and incubated for 10 minutes followed by addition of RNAse free water 250μl. Supernatant was precipitated with 500μl of isopropanol. The pellet was resuspended in 100 μl of ddH2O and digested with 1U of DNase RNase free (Invitrogen). The mix was incubated at 37 °C for 30 minutes, and then the RNA was purified by acid phenol extraction. The final pellet was resuspended in 20 μl of ddH2O and frozen at –80 °C. The RNA quality was checked by electrophoresis on 1% agarose gels or on a RNA gel if required.

4.10 Quantification of nucleic acid concentration.

An optical density of 1.0 at 260 nm is usually taken to be equivalent to a concentration of 50 μg/ml for double stranded DNA, 40 μg/ml for single stranded DNA and RNA, and approximately 20 μg/ml for single-stranded oligonucleotides samples. The ratio of values for optical densities measured at 260 nm and 280 nm is considered as 1.8 for pure sample of DNA and 2 for RNA and these are reduced by protein contaminants (Sambrook et al., 1989). Therefore, these values were used to determinate not only the concentration but also the purity of the samples.

4.11 Restriction enzymes.

Restriction endonucleases were used in the construction and analysis of recombinant plasmids. Each restriction enzyme functions optimally in a buffer of specific ionic strength. Restriction enzymes were purchased from NEB UK. The company supplied the enzymes with the buffers and were used according to the manufacturer's instruction. For analytical

digests 100-500 ng of DNA were digested in a volume of 100 μl containing 10 U of the appropriate restriction enzyme. The reaction was incubated for 2-3 hours at 37 °C. In case of NdeI, a 2x buffer condition were proved to be efficient regardless of manufacturer's instruction. Prior to inactivate enzymatic activity, digested product was purified by phenol chloroform extraction or gel purification kit (JetStar, Genomed).

4.12 Large fragment of E. coli Polymerase I and T4 Polynucleotide Kinase .

These enzymes were used to treat PCR products for blunt-end ligation during construction of recombinant plasmids. The large fragment of DNA Polymerase I (Klenow) is a proteolytic product of E. coli DNA Polymerase I. It retains polymerization and 3'->5' exonuclease activity, but has lost 5'->3' exonuclease activity. This was useful for digesting specific residues added by Taq DNA polymerase at the 3' terminus to create compatible ends for ligation. T4 Polynucleotide Kinase catalyses the transfer of phosphate from ATP to the 5' hydroxyl terminus of DNA. It was used for example in the addition of 5'-phosphate to PCR products to allow subsequent ligation. Klenow fragment (2.5 U) was added to 23 μl of PCR product in 5 mM MgCl₂ buffer. The mixture was incubated at room temperature for 10 minutes. EDTA to a final concentration of 0.2 mM, ATP to a final concentration of 1 mM, 10 U of T4 Polynucleotide Kinase and the proper quantity of Kinase buffer were added to the above mixture and incubated at 37 °C for 30 min. The enzymes were inactivated by incubation at 80 °C for 20 min.

4.13 T4 DNA ligase.

T4 DNA ligase catalyses the formation of a phosphodiester bond between adjacent 3' hydroxyl and 5' phosphoryl termini in DNA, requiring ATP as a cofactor in this reaction. This enzyme was used to join double stranded DNA fragments with compatible sticky or blunt ends, during generation of recombinant plasmid DNAs.

20 ng of linearized vector were ligated with a 5-10 fold molar excess of insert in a total volume of 20 µl containing 1X ligase buffer and 1U of T4 DNA ligase. Reaction was carried out at room temperature for 6-12 hours or at 37°C for an hour.

4.14 Agarose gel electrophoresis of DNA.

DNA samples were size fractionated by electrophoresis in agarose gels ranging in concentrations from 0.8 % w/v (large fragments) to 2 % w/v (small fragments). The gels contained ethidium bromide (0.5 µg /ml) and 1X TBE. Horizontal gels were routinely used for fast analysis of DNA restriction enzyme digests, estimation of DNA concentration, or DNA fragment separation prior to elution from the gel. Samples of 20 µl containing 1X DNA loading buffer were loaded into submerged wells. The gels were electrophoresed at 50-80 mA in 1X TBE running buffer for a time depending on the fragment length expected and gel concentration. DNA was visualized by UV transillumination and the result recorded by digital photography.

4.15 Denaturing polyacrylamide gel electrophoresis (SDS-PAGE).

Protein samples were added to protein sample buffer (2X final). Conventional slab gel SDS PAGE (Laemmli, 1970) was performed in vertical gels with the required percentage of polyacrylamide (37,5:1 acrylamide:bis-acrylamide, ProtoGel, National Diagnostics), depending on each case. The gels were run at 40 mA in 1X SDS-PAGE running buffer (50 mM Tris, 0.38 M glycine, 0.1 % w/v SDS). After running, gels were either stained with coomassie Blue R250 in methanol-water-acetic acid 45:45:10 (v/v/v).

4.16 Elution and purification of DNA fragments from agarose gels.

This protocol was used to purify small amounts (less than 1 µg) of DNA for sub-cloning. The DNA samples were electrophoresed onto an agarose gel as described previously. The DNA was visualized with UV light and the required DNA fragment band

was excised from the gel. This slab was cut into pieces, and the JETquick Spin Column Technique (Genomed) was used according to the manufacturer's instructions. Briefly, 600 μl of gel solubilisation solution L1 (NaClO$_4$, Na acetate and TBE) were added for each 100 mg of the gel slice pieces and incubated at 55 °C for 15 min vortexing every 5 min. The mixture was loaded into a prepared JETquick column and it was centrifuged at maximum speed for 1 min. The flowthrough was discarded. 700 μl of washing and reconstituted solution L2 (ethanol, NaCl, EDTA and Tris-HCl) were added into the spin column and after 5 min, the column was centrifuged in the same conditions twice. The flowthrough was again discarded both times. To elute the bound DNA, 30-50 μl of pre-warmed sterile water was added onto the centre of the silica matrix of the spin column and the system was centrifuged for 2 min. The amount of DNA recovered was approximately calculated by UV fluorescence of intercalated ethidium bromide in an agarose gel electrophoresis.

4.17 Preparation of bacterial competent cells

Bacterial competent cells were prepared following the method described by Chung and Niemela (Chung et al., 1989). *E. coli* strains were grown overnight in 3 ml of LB at 37°C. The following day, 300 ml of fresh LB were added and the cells were grown at room temperature for 4-5 h until the OD$_{600}$ was 0.3-0.4. The cells were then put in ice and centrifuged at 4 °C and 1000g for 15 min. The pellet was resuspended in 30 ml of cold TSS solution (10% w/v PEG, 5% v/v DMSO, 35mM MgCl$_2$, pH 6.5 in LB medium). The cells were aliquoted, rapidly frozen in liquid nitrogen and stored at –80°C. Competence was determined by transformation with 0.1 ng of pUC19 and was deemed satisfactory if this procedure resulted in more than 100 colonies.

4.18 Transformation of bacteria

Transformations of ligation reactions were performed using 1/2 of the reaction volume. Transformation of clones was carried out using 20 ng of the plasmid DNA. The

DNA was incubated with 60 µl of competent cells for 20 min on ice and at 42°C for 1.5 minutes. At this point 60 µl of LB were added and the bacteria allowed to recover for 10 min at 37 °C. The cells were then spread onto agarose plates containing the appropriate antibiotic. The plates were then incubated for 12-15 hours.

When DNA inserts were cloned into β-galactosidase-based virgin plasmids, 25 µl of IPTG 100 mM and 25 µl of X-Gal (4 % w/v in dimethylformamide) were spread onto the surface of the agarose before plating to facilitate screening of positive clones (white colonies) through identification of β-galactosidase activity (blue colonies).

4.19 Amplification of selected DNA fragments.

The polymerase chain reaction was performed on genomic or plasmidic DNA following the basic protocols of the Roche Diagnostic Taq DNA Polymerases. The volume of the reaction was 50 µl. The reaction buffer was: 1X Taq buffer, dNTP mix 200 µM each, oligonucleotide primers 1 nM each, Taq DNA Polymerase 2.5 U. As DNA template, 0.1 ng of plasmid or 100-500 ng of genomic DNA were used for amplification. The amplification conditions are described for each particular PCR. The amplifications were performed on a Cetus DNA Thermal Cycler (Perkin Elmer) or on a Gene Amp PCR System (Applied Biosystems).

4.20 The pTB minigene system.

A mutation that affects splicing can be identified using transient transfection of minigenes or *in vitro* splicing assays, comparing the splicing patterns of mutant and wild type exons.

A minigene, as its name indicates, is a simplified version of a gene. In this thesis the minigene used is a hybrid construct containing exons from α-globin and fibronectin, under the control of the α-globin promoter. The intronic region between the two fibronectin exons contains a unique NdeI site in which the exon along with its flanking regions with and

without the nucleotide substitution under study can be inserted. In fact this system allows us to insert a single exon with its intronic flanking regions, which could contain splicing regulatory elements, and analyse its splicing outcome. In the presence of a wild type context the exons of the minigene together with the one under study should be correctly spliced. On the other hand in presence of a mutation which affect the 5' splice site of the exon we could basically have four different outcomes: normal splicing pattern if the mutation is neutral; exon skipping, intron retention or activation of a cryptic 5' splice site if the mutation affects the mRNA processing.
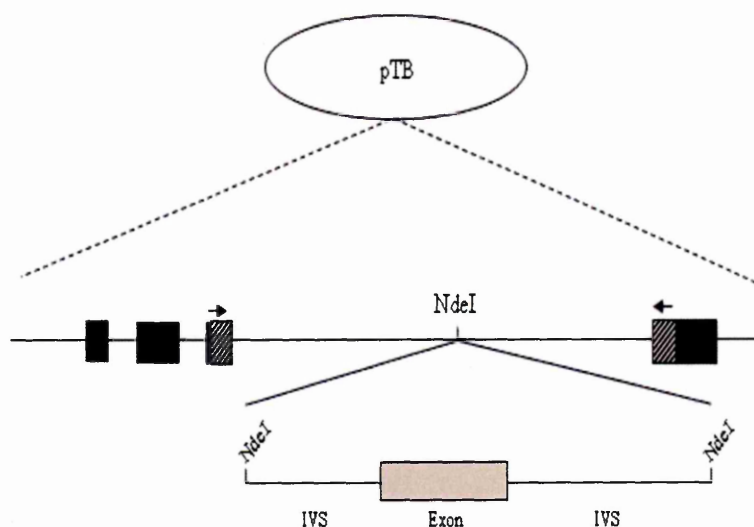
Figure4.1: Schematic representation of minigene splicing assay. It contains at the 5' end a α-globin gene promoter and SV40 enhancer sequences to allow polymerase II transcription in the transfected cell lines. The reporter gene is composed by α-globin (black boxes) and fibronectin exons (shaded boxes) while at the 3' end a functionally competent polyadenylation site, derived from the α-globin gene, is present.

4.21   Generation of minigenes and Expression Vector for SR proteins.

In this thesis several minigenes have been created Human *CFTR* exon 12 minigene constructs (T40C, G48C, A49G, A51T, C52T and WT) have been previously described (Pagani et al., 2005; Pagani et al., 2003c). Further modification of the exon were introduced by PCR-directed mutagenesis using specific primers and cloned inside the NdeI restriction site of the pTB plasmid. The DNA fragments were cloned into pUC19 plasmid using the

blunt SmaI enzyme restriction site and through sequencing using oligonucleotides Universal For and Rev we checked for the absence of any other nucleotide variation in the entire amplified fragments. Then, in order to study the splicing outcome we subcloned the two DNA fragments into the pTB minigene using the NdeI enzyme restriction site. The orientation of the inserted fragment was checked through colony PCR using, respectively, specific oligonucleotides heads and pTB1950as that recognized respectively the inserted fragment and the pTB plasmid; thus the amplification of the product could have been possible only in presence of a fragment with the correct orientation. Subsequent positive clones were further controlled through sequencing using oligonucleotide pTB1698s. The splicing assay was performed by transfecting each minigene into HeLa cells.

Mouse *CFTR* Exon 12 along with the mouse introns (297 nt of intron 11 and 219 nt of intron 12) was amplified from genomic DNA (*Mus musculas* strain C57BL/6) using primers MCF12 F and MCF12 R. Mouse *CFTR* exon 12 modifications were also performed by PCR-directed mutagenesis with specific primers.

The amplifications for the generation of the fragments to clone in the minigene were performed using the following oligonucleotides.

Name of the primer	Sequence of the primer 5' to 3'
MINIEX12WTAS	TAGGTATCCAAAAGGAGAGTCTAATGCTCTAAAAAGA AAATGG
MINIEX12WTS	ACTCTCCTTTTGGATACCTAGATGTTTTAACAAAGGTA TGTTC
MINIEX12G48CAS S	TAGGTATGCAAAAGGAGAGTCTAATGCTCTAAAAAGA AAATGG
MINIEX12G48C	ACTCTCCTTTTGCATACCTAGATGTTTTAACAAAGGTA

	TGTTC
MINIEX12A51TAS	TAGGAATCCAAAAGGAGAGTCTAATGCTCTAAAAAGA AAATGG
MINIEX12A51TS	ACTCTCCTTTTGGATTCCTAGATGTTTTAACAAAGGTAT GTTC
MINIEX12T40CAS	TAGGTATCCAAAAGGGGAGTCTAATGCTCTAAAAAGA AAATGG
MINIEX12T40CS	ACTCCCCTTTTGGATACCTAGATGTTTTAACAAAGGTA TGTTC
MINIEX12A49GAS	TAGGTACCCAAAAGGAGAGTCTAATGCTCTAAAAAGA AAATGG
MINIEX12A49GS	ACTCTCCTTTTGGGTACCTAGATGTTTTAACAAAGGTA TGTTC
MINIEX12C52TAS	TAGATATCCAAAAGGAGAGTCTAATGCTCTAAAAAGA AAATGG
MINIEX12WTS	ACTCTCCTTTTGGATATCTAGATGTTTTAACAAAGGTA TGTTC
CERES DEL F	AGACTCTCCTTTTCTAGATGTTTTAA
CERES DEL R	TTAAAACATCTAGAAAAGGAGAGTCT
MCF12 F: 5'	GGCTCCAGGCTTGAGCATATGTACTAATCTG
MCF12 R: 5'	CAAGAAATTGGCTTCATATGTGATCATCAGA
MCF12 5 DEL F	CTAGATGTTTTTACTGTATTTGAAAGGT

MCF12 5 DEL F	ACCTTTCAAATACAGTAAAAACATCTAG
MCF12 3 DEL F	CAAAGATGCTGATTTAGATTCCCCTTT
MCF12 3 DEL R	AAAGGGGAATCTAAATCAGCATCTTTG
PTB2400 S	GTGCTTTTGAATGGCACAT
PTB3161 AS	CTACACACAGTACTGTTAGGC

Plasmids containing the cDNA of SRp 55 and SC35 were previously cloned in to pCG by Cristiana Stuani in the Lab. Plasmid encoding the SF2/ASF, ΔRRM1, ΔRRM2 and ΔRS were kind gifts from Dr. J Caceres.

4.22 Maintenance and analysis of cells in culture.

Both HeLa and N-muli cell line was grown in Dulbecco's Mem with Glutamax I (Gibco) (Dulbecco's modified Eagle's medium with glutamine, sodium pyruvate, pyridoxine and 4.5 g/l glucose) supplemented with 10% foetal calf serum (Euro Clone) and Antibiotic Antimycotic (Sigma) according to the manufacturer's instructions.

Plates containing a confluent monolayer of cells were treated with 0.1% w/v trypsin as follows. Cells washed with PBS solution, were incubated at 37° C with 1-2 ml of PBS/EDTA/trypsin solution (PBS containing 0.04% w/v EDTA and 0.1% w/v trypsin) for 2 minutes or until cells were dislodged. After adding 10 ml of media, cells were pelleted by centrifugation and resuspended in 5 ml pre-warmed medium. 1-2 ml of this cell suspension was added to 10 ml medium in a fresh plate and was gently mixed before incubation.

4.23 Plasmid Transfections.

The DNA used for transfections were purified using phenol-choloroform extraction followed by a sephacryl S-400 (GE healthcare) column purification step. Liposome-mediated transfections of 80-90% confluent plates were performed using Effectene reagent (Quiagen). 0.5 µg of construct DNA was mixed with 4 µl of Enhancer for each transfection and the mixture was incubated at room temperature for 5 minutes to allow the condensation of the DNA. Then, 5 µl of Effectene were added to the mixture and an incubation of 10 minutes has been performed. After the addition of 500 µl of complete culture medium the mixture was added to the cells in 2 ml of the as described above.

4.24 Small interfering RNA (siRNA) transfections.

siRNA transfections were performed in HeLa cells using Oligofectamine Reagent (Invitrogen). One day before siRNA transfection, 0.7×10^5 cells were plated to achieve 30-40% confluences. The next day, 3 µl Oligofectamine (Invitrogen) was combined with 15 µl of Opti-MEM medium (Invitrogen) and 3µl of 40 µM siRNA duplex oligonucleotides was diluted in a final volume of 180 µl of Opti-MEM medium. The two mixtures were combined and left for 20 min at room temperature. Finally, this mixture was added to the cells, which were mantained in 0.9 ml of Opti-MEM only. After 6 hour 500 µl of 30% FBS (Foetal bovine serum, Invitrogen) was added. Six to eight hours later Opti-MEM was exchanged with Dulbecco's modified Eagle medium and the cells were transfected with the minigene of interest (500ng) using Qiagen Effectene transfection reagents. On the third day, HeLa cells were harvested for protein and RNA extractions. RT-PCR from total RNA was performed as for the transfection protocol described above. Whole-protein extracts were obtained by cell sonication in lysis buffer (1X PBS and 1X Protease inhibitor cocktail) and analyzed for hnRNPA1, A2, C1/C2 and DAZAP1 endogenous protein expression by immunoblotting using the antibodies described above. Tubulin was used as total protein loading control. Sequence of the siRNA is given bellow

Human hnRNPA1-	cagcugaggaagcucuuca (Sigma)
Human hnRNPA2-	ggaacaguuccguaagcuc (Sigma)
Human hnRNP C1/C2-	gcaaacaagcaguagagau (Sigma)
Human DAZAP1-	gagacucugcgcagcuacu (Dharmacon)
Luciferase no. 2	gccauucuauccucuagaggaug (Dharmacon).

4.25 cDNA synthesis.

In order to synthesize cDNA, the 1 μg of total RNA extracted from cells were mixed with random primers (Pharmacia) in a final volume of 20 μl. After denaturation at 70°C the RNA and the primer were incubated for 1 hour at 37 °C in the following solution: 1X First Strand Buffer (Gibco), 10 mM DTT, 1 mM dNTPs, RNase inhibitor 20 U (Ambion) and Moloney murine leukemia virus reverse transcriptase 100 U (Gibco). 3μl of the cDNA reaction mix was used for the PCR analysis.

4.26 cDNA analysis

PCR analysis of cDNA was carried out for 35 cycles (94 °C 1:30 min, 55 °C 1:30 min, 72 °C 1:30 min) in 50 μl reaction volumes using the following oligonucleotides which recognize specific regions of the minigene.

Name of the primer	Sequence of the primer 5' to 3'
ALFA-2,3	CAACTTCAAGCTCCTAAGCCACTGC
BRA-2	TAGGATCCGGTCACCAGGAAGTTGGTTAAATCA

4.27 *In vitro* Splicing assay

Splicing reactions were performed *in vitro* using m7G capped, SP-6 transcribed cold RNAs. Standard reactions were carried out in a 25 µl volume at 30°C for 2 hours. Each reaction contained 15ml of Nuclear Extract from HeLa cells (CilBiotech, Mons, Belgium, approx. concentration 10 µg/µl), 5 µl of 13% (w/v) polyvinyl alcohol, 1µl of 80mM $MgCl_2$, 1µl of 12.5 mM ATP, 1 µl of 0.5 M creatine phosphate and 1.25 µl of 0.4 M Hepes-KOH pH=7.3 and 2 µl of *in vitro* transcribed pre-mRNA at 200mg/ml. Therefore, the final concentrations of the various components in a standard processing reaction were as follows:- 3.2mM $MgCl_2$, 500mM ATP, 20mM creatine phosphate, 2.7%(w/v) PVA, 20mM Hepes (pH=7.3), 6µg/µl of Hela nuclear extract, and 16mg/ml of *in vitro* transcribed pre-mRNA.

The processed RNAs were then extracted from the reaction mix using acid phenol: chloroform (1:1) and chloroform only. The precipitated RNAs were analyzed by RT-PCR using the antisense primer followed by PCR with primers at the beginning and end of the exons. The primers used were as follows:

Name of the primer	Sequence of the primer 5' to 3'
E1	CTATTAGAAAATTCCGCTATCCTTG
E2	CACATACGATTTAGGTGACAC

4.28 *In vitro* transcription and synthetic RNA oligo.

RNA was transcribed from PCR templates amplified from the respective plasmids. A T7 promoter sequence was added towards the 5' end of the template using primer carrying T7 sequence and, similarly, a (TG)8 repeated sequence was used to tag the 3' end. Every

time, 5 µg of DNA template was used in a 60 µl T7 polymerase (Stratagene) transcription reaction. The RNA was then purified using standard Acid-Phenol purification method followed by Ethanol precipitation. Synthetic RNA oligos corresponding to Mouse 17-38, Human 17-38, Mouse 63-81 and Human 63-81 sequences were purchased from Eurofins MWG Operon, Germany.

Name of the primer	Sequence of the primer 5' to 3'
T7EX12S	TAATACGACTCACTATAGGTTAGACTCTCCTTTTG
Ex12ASTag	CACACACACACACACATGTTAAAACATCTA
T7EX12T40CS	TAATACGACTCACTATAGGTTAGACTCCCCTTTTG
3' T7 S	TAATACGACTCACTATAGGGTATACAAAGATGCTGA
3' Tag AS	CACACACACACACACAGGAGAGTCTAATAAATAC
3' tag T40C AS	CACACACACACACACAGGGGAGTCTAATAAATAC
5' T7 S	TAATACGACTCACTATAGGTGTTTTAACAGAAAAAG
5' Tag AS	CACACACACACACACATCAAATATTTCTTTTTCT

Name of the RNA oligo	Sequence of the RNA oligo 5' to 3'

115

Mouse 18-39	CUGAUUUGUACCUAUUAGAUUC
Human 18-39	CUGAUUUGUAUUUAUUAGAUUC
Mouse 64-82	ACUGAAGAACAAGUAUUU
Human 64-82	ACAGAAAAAGAAAUAUUU

4.29 UV Cross linking

RNA oligos were labeled with γ- P^{32} in an T4 polynucleotide kinase reaction. After acid phenol standard purification step labeled RNA oligos were incubated in splicing mix (final concentration of 3.2mM $MgCl_2$, 500mM ATP, 20mM creatine phosphate, 2.7%(w/v) PVA, 20mM Hepes (pH=7.3) and 6mg/ml of Hela nuclear extract). Splicing mix were then UV cross linked at 0.800 KHtz. Cross linked samples were then separated in an 10% SDS gel and dried on a blotting paper. Dry gel were then exposed to X-Ray film for signal.

4.30 Affinity purification of RNA binding proteins and Western Blot analysis.

10 μg of synthetic RNA oligo or 15 μg of transcribed RNA were oxidized in the dark for an hour with sodium m-periodate in a 400 μl reaction mixture (100 mM Sodium acetate pH 5.2 and 5 mM Sodium m-periodate). RNA was then ethanol precipitated and resuspended in 200 μl of 100 mM sodium acetate. Approximately 350 μl of prewashed equilibrated adipic acid dehydrazide-agarose beads (50% slurry; Sigma) were added to each oxidized RNA volume and placed in the rotor at 4°C for overnight incubation. This RNA-Bead covalent link formation was also performed in the dark. The immobilized RNA were then washed once with 1 ml of 2 M NaCl and twice using washing buffer (5.2 mM HEPES pH 7.5, 1 mM MgCl2, 0.8 mM Mg acetate). Meanwhile 200 μl of Nuclear extract was mixed with 900 μl RNAse free water, 1X binding buffer (5.2 mM HEPES pH 7.9, 1 mM

116

MgCl2, 0.8 mM Mg acetate, 0.52 mM dithiothreitol, 3.8% glycerol, 0.75 mM ATP, 1 mM GTP) and Heparin at the final concentration of 0.5 µg/µl). The RNA-bound beads were then equilibrated in 300 µl of NE mix and incubated for 25 min on a rotor at room temperature. Beads were then washed four times with 1.5 ml washing buffer. In every washing step beads were gently precipitated by gravity on ice. Finally, 50µl of 3X SDS loading buffer was added and samples were heated for 5 min before loading on a 10% SDS-polyacrylamide denaturing gel.
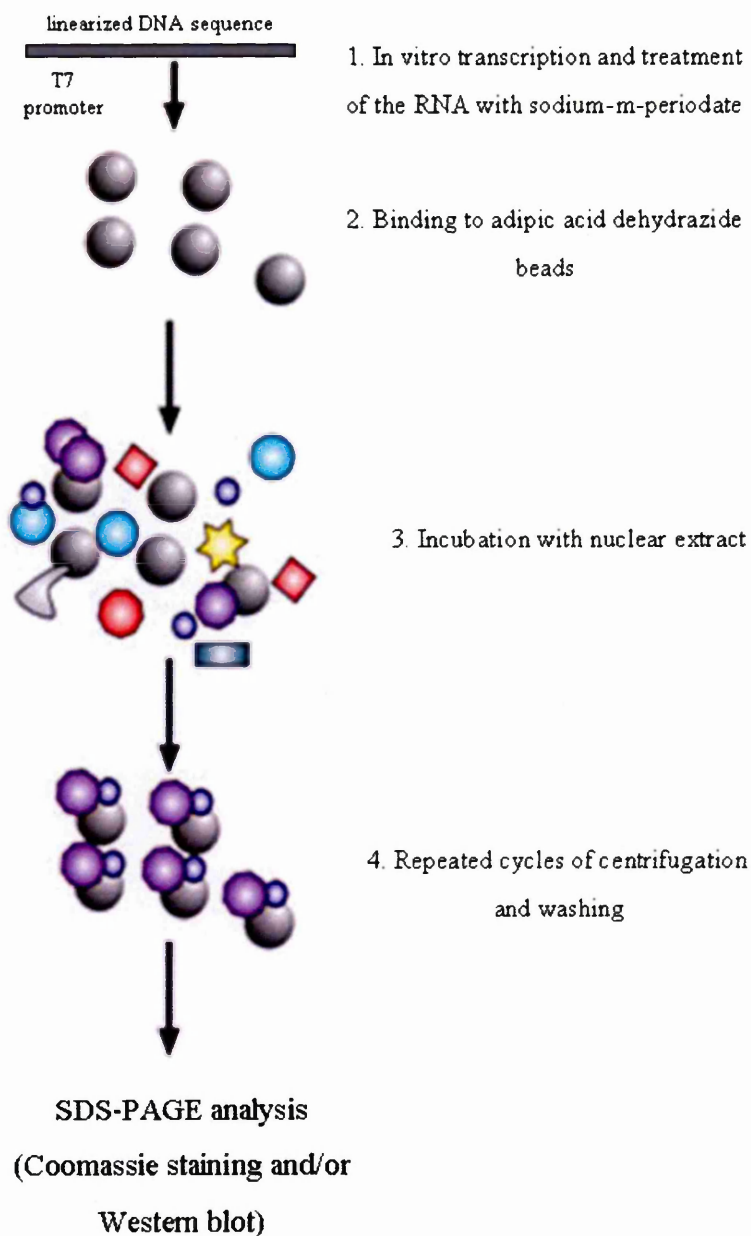
Figure4.2: Schematic representation of the pull down analysis. All the steps of the experimental procedure are specified.

The gel was then electroblotted onto a polyvinylidene difluoride membrane according to standard protocols (Amersham Biosciences) and blocked with 10% skimmed milk (Non fat dry milk in 1x PBS). Membranes targeted for SR protein recognition were blocked using Western blocking reagent (Roche) in order to detect the phosphorylated form of the protein. All the antibodies were diluted in to 1:1000 ratio in their blocking solution at 0.5X concentration. After incubation for 1 hour in room temperature with the antibodies and 3 constitutive washes with 1X PBS, proteins in the membrane were detected with a chemiluminescence kit (ECL; Pierce Biotechnology). Purified glutathione S-transferase (GST)- hnRNP A1/A2 protein was used to immunize 2 rabbit (New Zealand strain) according to standard protocols to obtain polyclonal anti-hnRNP A1/A2 antibodies. Polyclonal antibodies for hnRNP C1/C2 and TDP43, H, I/PTB and DAZAP1 was previously produced in the lab by Dr. E. Buratti, Dr. R. Marcucci, Dr. M. Romano and Dr. N. Soko respectively. Antibodies against hnRNP U and Tra2 β were kind gifts from G. Dreyfuss and I.C Eperon, respectively. Monoclonal Anti-ASF/SF2, SC35 and 1H4 (against SRp 75, 55, 40) antibodies were purchased from Zymed Laboratories Inc.

5   Reference

Aebi, M., Hornig, H., Padgett, R.A., Reiser, J., and Weissmann, C. (1986). Sequence requirements for splicing of higher eukaryotic nuclear pre-mRNA. Cell *47*, 555-565.

Aebi, M., Hornig, H., and Weissmann, C. (1987). 5' cleavage site in eukaryotic pre-mRNA splicing is determined by the overall 5' splice region, not by the conserved 5' GU. Cell *50*, 237-246.

Auboeuf, D., Dowhan, D.H., Kang, Y.K., Larkin, K., Lee, J.W., Berget, S.M., and O'Malley, B.W. (2004). Differential recruitment of nuclear receptor coactivators may determine alternative RNA splice site choice in target genes. Proc Natl Acad Sci U S A *101*, 2270-2274.

Auweter, S.D., Oberstrass, F.C., and Allain, F.H. (2006). Sequence-specific binding of single-stranded RNA: is there a code for recognition? Nucleic Acids Res *34*, 4943-4959.

Baralle, D., and Baralle, M. (2005). Splicing in action: assessing disease causing sequence changes. J Med Genet *42*, 737-748.

Baralle, D., Lucassen, A., and Buratti, E. (2009). Missed threads. The impact of pre-mRNA splicing defects on clinical practice. EMBO Rep *10*, 810-816.

Barash, Y., Calarco, J.A., Gao, W., Pan, Q., Wang, X., Shai, O., Blencowe, B.J., and Frey, B.J. (2010). Deciphering the splicing code. Nature *465*, 53-59.

Ben-Dov, C., Hartmann, B., Lundgren, J., and Valcarcel, J. (2008). Genome-wide analysis of alternative pre-mRNA splicing. J Biol Chem *283*, 1229-1233.

Bentley, D.L. (2005). Rules of engagement: co-transcriptional recruitment of pre-mRNA processing factors. Curr Opin Cell Biol *17*, 251-256.

Berget, S.M., Moore, C., and Sharp, P.A. (1977). Spliced segments at the 5' terminus of adenovirus 2 late mRNA. Proc Natl Acad Sci U S A *74*, 3171-3175.

Berglund, J.A., Chua, K., Abovich, N., Reed, R., and Rosbash, M. (1997). The splicing factor BBP interacts specifically with the pre-mRNA branchpoint sequence UACUAAC. Cell *89*, 781-787.

Bierne, N., and Eyre-Walker, A. (2004). The genomic rate of adaptive amino acid substitution in Drosophila. Mol Biol Evol *21*, 1350-1360.

Black, D.L. (2003). Mechanisms of alternative pre-messenger RNA splicing. Annu Rev Biochem *72*, 291-336.

Blanchette, M., Green, R.E., MacArthur, S., Brooks, A.N., Brenner, S.E., Eisen, M.B., and Rio, D.C. (2009). Genome-wide analysis of alternative pre-mRNA splicing and RNA-binding specificities of the Drosophila hnRNP A/B family members. Mol Cell *33*, 438-449.

Blencowe, B.J. (2000). Exonic splicing enhancers: mechanism of action, diversity and role in human genetic diseases. Trends Biochem Sci *25*, 106-110.

Boehringer, D., Makarov, E.M., Sander, B., Makarova, O.V., Kastner, B., Luhrmann, R., and Stark, H. (2004). Three-dimensional structure of a pre-catalytic human spliceosomal complex B. Nat Struct Mol Biol *11*, 463-468.

Bompadre, S.G., Ai, T., Cho, J.H., Wang, X., Sohma, Y., Li, M., and Hwang, T.C. (2005). CFTR gating I: Characterization of the ATP-dependent gating of a phosphorylation-independent CFTR channel (DeltaR-CFTR). J Gen Physiol *125*, 361-375.

Bonetta, L. (2009). RNA-based therapeutics: ready for delivery? Cell *136*, 581-584.

Bourgeois, C.F., Popielarz, M., Hildwein, G., and Stevenin, J. (1999). Identification of a bidirectional splicing enhancer: differential involvement of SR proteins in 5' or 3' splice site activation. Mol Cell Biol *19*, 7347-7356.

Brackenridge, S., Wilkie, A.O., and Screaton, G.R. (2003). Efficient use of a 'dead-end' GA 5' splice site in the human fibroblast growth factor receptor genes. Embo J *22*, 1620-1631.

Breathnach, R., Benoist, C., O'Hare, K., Gannon, F., and Chambon, P. (1978). Ovalbumin gene: evidence for a leader sequence in mRNA and DNA sequences at the exon-intron boundaries. Proc Natl Acad Sci U S A *75*, 4853-4857.

Bruzik, J.P., and Maniatis, T. (1995). Enhancer-dependent interaction between 5' and 3' splice sites in trans. Proc Natl Acad Sci U S A *92*, 7056-7059.

Buratti, E., and Baralle, F.E. (2001). Characterization and functional implications of the RNA binding properties of nuclear factor TDP-43, a novel splicing regulator of CFTR exon 9. J Biol Chem *276*, 36337-36343.

Buratti, E., Baralle, M., De Conti, L., Baralle, D., Romano, M., Ayala, Y.M., and Baralle, F.E. (2004a). hnRNP H binding at the 5' splice site correlates with the pathological effect of two intronic mutations in the NF-1 and TSHbeta genes. Nucleic Acids Res *32*, 4224-4236.

Buratti, E., Dhir, A., Lewandowska, M.A., and Baralle, F.E. (2007a). RNA structure is a key regulatory element in pathological ATM and CFTR pseudoexon inclusion events. Nucleic Acids Res *35*, 4369-4383.

Buratti, E., Dork, T., Zuccato, E., Pagani, F., Romano, M., and Baralle, F.E. (2001). Nuclear factor TDP-43 and SR proteins promote in vitro and in vivo CFTR exon 9 skipping. Embo J *20*, 1774-1784.

Buratti, E., Muro, A.F., Giombi, M., Gherbassi, D., Iaconcig, A., and Baralle, F.E. (2004b). RNA folding affects the recruitment of SR proteins by mouse and human polypurinic enhancer elements in the fibronectin EDA exon. Mol Cell Biol *24*, 1387-1400.

Buratti, E., Stuani, C., De Prato, G., and Baralle, F.E. (2007b). SR protein-mediated inhibition of CFTR exon 9 inclusion: molecular characterization of the intronic splicing silencer. Nucleic Acids Res *35*, 4359-4368.

Caceres, J.F., Screaton, G.R., and Krainer, A.R. (1998). A specific subset of SR proteins shuttles continuously between the nucleus and the cytoplasm. Genes Dev *12*, 55-66.

Caceres, J.F., Stamm, S., Helfman, D.M., and Krainer, A.R. (1994). Regulation of alternative splicing in vivo by overexpression of antagonistic splicing factors. Science *265*, 1706-1709.

Caputi, M., and Zahler, A.M. (2001). Determination of the RNA binding specificity of the heterogeneous nuclear ribonucleoprotein (hnRNP) H/H'/F/2H9 family. J Biol Chem *276*, 43850-43859.

Carlini, D.B., and Genut, J.E. (2006). Synonymous SNPs provide evidence for selective constraint on human exonic splicing enhancers. J Mol Evol *62*, 89-98.

Cartegni, L., Chew, S.L., and Krainer, A.R. (2002). Listening to silence and understanding nonsense: exonic mutations that affect splicing. Nat Rev Genet *3*, 285-298.

Cartegni L, C.S., Krainer A (2002). Listening to silence and understanding nonsense: exonic mutations that affect splicing. Nat Rev Genet *3*, 285-298.

Cartegni, L., Hastings, M.L., Calarco, J.A., de Stanchina, E., and Krainer, A.R. (2006). Determinants of exon 7 splicing in the spinal muscular atrophy genes, SMN1 and SMN2. Am J Hum Genet *78*, 63-77.

Cartegni, L., and Krainer, A.R. (2002). Disruption of an SF2/ASF-dependent exonic splicing enhancer in SMN2 causes spinal muscular atrophy in the absence of SMN1. Nat Genet *30*, 377-384.

Chamary, J.V., Parmley, J.L., and Hurst, L.D. (2006). Hearing silence: non-neutral evolution at synonymous sites in mammals. Nat Rev Genet *7*, 98-108.

Chen, M., and Manley, J.L. (2009). Mechanisms of alternative splicing regulation: insights from molecular and genomics approaches. Nat Rev Mol Cell Biol *10*, 741-754.

Cheng, C., and Sharp, P.A. (2006). Regulation of CD44 alternative splicing by SRm160 and its potential role in tumor cell invasion. Mol Cell Biol *26*, 362-370.

Chiodi, I., Corioni, M., Giordano, M., Valgardsdottir, R., Ghigna, C., Cobianchi, F., Xu, R.M., Riva, S., and Biamonti, G. (2004). RNA recognition motif 2 directs the recruitment of SF2/ASF to nuclear stress bodies. Nucleic Acids Res *32*, 4127-4136.

Chou, M.Y., Underwood, J.G., Nikolic, J., Luu, M.H., and Black, D.L. (2000). Multisite RNA binding and release of polypyrimidine tract binding protein during the regulation of c-src neural-specific splicing. Mol Cell *5*, 949-957.

Chow, L.T., Gelinas, R.E., Broker, T.R., and Roberts, R.J. (1977). An amazing sequence arrangement at the 5' ends of adenovirus 2 messenger RNA. Cell *12*, 1-8.

Chung, C.T., Niemela, S.L., and Miller, R.H. (1989). One-step preparation of competent Escherichia coli: transformation and storage of bacterial cells in the same solution. Proc Natl Acad Sci U S A *86*, 2172-2175.

Collins, C.A., and Guthrie, C. (1999). Allele-specific genetic interactions between Prp8 and RNA active site residues suggest a function for Prp8 at the catalytic core of the spliceosome. Genes Dev *13*, 1970-1982.

Coolidge, C.J., Seely, R.J., and Patton, J.G. (1997). Functional analysis of the polypyrimidine tract in pre-mRNA splicing. Nucleic Acids Res *25*, 888-896.

Cooper, T.A., Wan, L., and Dreyfuss, G. (2009). RNA and disease. Cell *136*, 777-793.

Coulter, L.R., Landree, M.A., and Cooper, T.A. (1997). Identification of a new class of exonic splicing enhancers by in vivo selection. Mol Cell Biol *17*, 2143-2150.

Cramer, P., Caceres, J.F., Cazalla, D., Kadener, S., Muro, A.F., Baralle, F.E., and Kornblihtt, A.R. (1999). Coupling of transcription with alternative splicing: RNA pol II promoters modulate SF2/ASF and 9G8 effects on an exonic splicing enhancer. Mol Cell *4*, 251-258.

Cramer, P., Pesce, C.G., Baralle, F.E., and Kornblihtt, A.R. (1997). Functional association between promoter structure and transcript alternative splicing. Proc Natl Acad Sci U S A *94*, 11456-11460.

Crispino, J.D., Blencowe, B.J., and Sharp, P.A. (1994). Complementation by SR proteins of pre-mRNA splicing reactions depleted of U1 snRNP. Science *265*, 1866-1869.

Das, R., Yu, J., Zhang, Z., Gygi, M.P., Krainer, A.R., Gygi, S.P., and Reed, R. (2007). SR proteins function in coupling RNAP II transcription to pre-mRNA splicing. Mol Cell *26*, 867-881.

Das, R., Zhou, Z., and Reed, R. (2000). Functional association of U2 snRNP with the ATP-independent spliceosomal complex E. Mol Cell *5*, 779-787.

de la Mata, M., Alonso, C.R., Kadener, S., Fededa, J.P., Blaustein, M., Pelisch, F., Cramer, P., Bentley, D., and Kornblihtt, A.R. (2003). A slow RNA polymerase II affects alternative splicing in vivo. Mol Cell *12*, 525-532.

Del Gatto, F., Plet, A., Gesnel, M.C., Fort, C., and Breathnach, R. (1997). Multiple interdependent sequence elements control splicing of a fibroblast growth factor receptor 2 alternative exon. Mol Cell Biol *17*, 5106-5116.

Delaney, S.J., Rich, D.P., Thomson, S.A., Hargrave, M.R., Lovelock, P.K., Welsh, M.J., and Wainwright, B.J. (1993). Cystic fibrosis transmembrane conductance regulator splice variants are not conserved and fail to produce chloride channels. Nat Genet *4*, 426-431.

Dhir, A., Buratti, E., van Santen, M.A., Luhrmann, R., and Baralle, F.E. The intronic splicing code: multiple factors involved in ATM pseudoexon definition. EMBO J *29*, 749-760.

Dirksen, W.P., Mohamed, S.A., and Fisher, S.A. (2003). Splicing of a Myosin Phosphatase Targeting Subunit 1 Alternative Exon Is Regulated by Intronic Cis-elements and a Novel Bipartite Exonic Enhancer/Silencer Element. J Biol Chem *278*, 9722-9732.

Dominski, Z., and Kole, R. (1991). Selection of splice sites in pre-mRNAs with short internal exons. Mol Cell Biol *11*, 6075-6083.

Dreyfuss, G., Kim, V.N., and Kataoka, N. (2002). Messenger-RNA-binding proteins and the messages they carry. Nat Rev Mol Cell Biol *3*, 195-205.

Dreyfuss, G., Matunis, M.J., Pinol-Roma, S., and Burd, C.G. (1993). hnRNP proteins and the biogenesis of mRNA. Annu Rev Biochem *62*, 289-321.

Eperon, I.C., Makarova, O.V., Mayeda, A., Munroe, S.H., Caceres, J.F., Hayward, D.G., and Krainer, A.R. (2000). Selection of alternative 5' splice sites: role of U1 snRNP and

models for the antagonistic effects of SF2/ASF and hnRNP A1. Mol Cell Biol *20*, 8303-8318.

Eperon, L.P., Estibeiro, J.P., and Eperon, I.C. (1986). The role of nucleotide sequences in splice site selection in eukaryotic pre-messenger RNA. Nature *324*, 280-282.

Eperon, L.P., Graham, I.R., Griffiths, A.D., and Eperon, I.C. (1988). Effects of RNA secondary structure on alternative splicing of pre-mRNA: is folding limited to a region behind the transcribing RNA polymerase? Cell *54*, 393-401.

Expert-Bezancon, A., Sureau, A., Durosay, P., Salesse, R., Groeneveld, H., Lecaer, J.P., and Marie, J. (2004). hnRNP A1 and the SR proteins ASF/SF2 and SC35 have antagonistic functions in splicing of beta-tropomyosin exon 6B. J Biol Chem *279*, 38249-38259.

Eyre-Walker, A.C. (1991). An analysis of codon usage in mammals: selection or mutation bias? J Mol Evol *33*, 442-449.

Fagnani, M., Barash, Y., Ip, J.Y., Misquitta, C., Pan, Q., Saltzman, A.L., Shai, O., Lee, L., Rozenhek, A., Mohammad, N., *et al.* (2007). Functional coordination of alternative splicing in the mammalian central nervous system. Genome Biol *8*, R108.

Fairbrother, W.G., and Chasin, L.A. (2000). Human genomic sequences that inhibit splicing. Mol Cell Biol *20*, 6816-6825.

Fairbrother, W.G., Holste, D., Burge, C.B., and Sharp, P.A. (2004). Single nucleotide polymorphism-based validation of exonic splicing enhancers. PLoS Biol *2*, E268.

Fairbrother, W.G., Yeh, R.F., Sharp, P.A., and Burge, C.B. (2002). Predictive identification of exonic splicing enhancers in human genes. Science *297*, 1007-1013.

Faustino, N.A., and Cooper, T.A. (2003). Pre-mRNA splicing and human disease. Genes Dev *17*, 419-437.

Fay, J.C., and Wu, C.I. (2003). Sequence divergence, functional constraint, and selection in protein evolution. Annu Rev Genomics Hum Genet *4*, 213-235.

Fedorova, O., and Pyle, A.M. (2005). Linking the group II intron catalytic domains: tertiary contacts and structural features of domain 3. EMBO J *24*, 3906-3916.

Feng, Y., Chen, M., and Manley, J.L. (2008). Phosphorylation switches the general splicing repressor SRp38 to a sequence-specific activator. Nat Struct Mol Biol *15*, 1040-1048.

Fong, Y.W., and Zhou, Q. (2001). Stimulatory effect of splicing factors on transcriptional elongation. Nature *414*, 929-933.

Forch, P., Puig, O., Martinez, C., Seraphin, B., and Valcarcel, J. (2002). The splicing regulator TIA-1 interacts with U1-C to promote U1 snRNP recruitment to 5' splice sites. EMBO J *21*, 6882-6892.

Fu, X.D. (1995). The superfamily of arginine/serine-rich splicing factors. RNA *1*, 663-680.

Fukumura, K., Taniguchi, I., Sakamoto, H., Ohno, M., and Inoue, K. (2009). U1-independent pre-mRNA splicing contributes to the regulation of alternative splicing. Nucleic Acids Res *37*, 1907-1914.

Furger, A., O'Sullivan, J.M., Binnie, A., Lee, B.A., and Proudfoot, N.J. (2002). Promoter proximal splice sites enhance transcription. Genes Dev *16*, 2792-2799.

Gao, K., Masuda, A., Matsuura, T., and Ohno, K. (2008). Human branch point consensus sequence is yUnAy. Nucleic Acids Res *36*, 2257-2267.

Garcia-Blanco, M.A., Baraniak, A.P., and Lasda, E.L. (2004). Alternative splicing in disease and therapy. Nat Biotechnol *22*, 535-546.

Ge, H., and Manley, J.L. (1990). A protein factor, ASF, controls cell-specific alternative splicing of SV40 early pre-mRNA in vitro. Cell *62*, 25-34.

Gersappe, A., and Pintel, D.J. (1999). CA- and purine-rich elements form a novel bipartite exon enhancer which governs inclusion of the minute virus of mice NS2-specific exon in both singly and doubly spliced mRNAs. Mol Cell Biol *19*, 364-375.

Ghigna, C., Giordano, S., Shen, H., Benvenuto, F., Castiglioni, F., Comoglio, P.M., Green, M.R., Riva, S., and Biamonti, G. (2005). Cell motility is controlled by SF2/ASF through alternative splicing of the Ron protooncogene. Mol Cell *20*, 881-890.

Gilbert, W. (1978). Why genes in pieces? Nature *271*, 501.

Goina, E., Skoko, N., and Pagani, F. (2008). Binding of DAZAP1 and hnRNPA1/A2 to an exonic splicing silencer in a natural BRCA1 exon 18 mutant. Mol Cell Biol *28*, 3850-3860.

Goldman, M.J., Anderson, G.M., Stolzenberg, E.D., Kari, U.P., Zasloff, M., and Wilson, J.M. (1997). Human beta-defensin-1 is a salt-sensitive antibiotic in lung that is inactivated in cystic fibrosis. Cell *88*, 553-560.

Gooding, C., Roberts, G.C., and Smith, C.W. (1998). Role of an inhibitory pyrimidine element and polypyrimidine tract binding protein in repression of a regulated alpha-tropomyosin exon. Rna *4*, 85-100.

Goss, C.H., and Rosenfeld, M. (2004). Update on cystic fibrosis epidemiology. Curr Opin Pulm Med *10*, 510-514.

Gozani, O., Feld, R., and Reed, R. (1996). Evidence that sequence-independent binding of highly conserved U2 snRNP proteins upstream of the branch site is required for assembly of spliceosomal complex A. Genes Dev *10*, 233-243.

Graveley, B.R. (2000). Sorting out the complexity of SR protein functions. Rna *6*, 1197-1211.

Graveley, B.R., Hertel, K.J., and Maniatis, T. (1998a). A systematic analysis of the factors that determine the strength of pre- mRNA splicing enhancers. Embo J *17*, 6747-6756.

Graveley, B.R., Hertel, K.J., and Maniatis, T. (1998b). A systematic analysis of the factors that determine the strength of pre-mRNA splicing enhancers. Embo J *17*, 6747-6756.

Graveley, B.R., Hertel, K.J., and Maniatis, T. (1999). SR proteins are 'locators' of the RNA splicing machinery. Curr Biol *9*, R6-7.

Graveley, B.R., Hertel, K.J., and Maniatis, T. (2001). The role of U2AF35 and U2AF65 in enhancer-dependent splicing. Rna 7, 806-818.

Gregory, R.J., Cheng, S.H., Rich, D.P., Marshall, J., Paul, S., Hehir, K., Ostedgaard, L., Klinger, K.W., Welsh, M.J., and Smith, A.E. (1990). Expression and characterization of the cystic fibrosis transmembrane conductance regulator. Nature 347, 382-386.

Gromoll, J., Lahrmann, L., Godmann, M., Muller, T., Michel, C., Stamm, S., and Simoni, M. (2007). Genomic checkpoints for exon 10 usage in the luteinizing hormone receptor type 1 and type 2. Mol Endocrinol 21, 1984-1996.

Hanamura, A., Caceres, J.F., Mayeda, A., Franza, B.R., Jr., and Krainer, A.R. (1998). Regulated tissue-specific expression of antagonistic pre-mRNA splicing factors. Rna 4, 430-444.

Hartmann, L., Theiss, S., Niederacher, D., and Schaal, H. (2008). Diagnostics of pathogenic splicing mutations: does bioinformatics cover all bases? Front Biosci 13, 3252-3272.

Hastings, M.L., Allemand, E., Duelli, D.M., Myers, M.P., and Krainer, A.R. (2007). Control of pre-mRNA splicing by the general splicing factors PUF60 and U2AF65. PLoS ONE 2, e538.

Hawkins, J.D. (1988). A survey on intron and exon lengths. Nucleic Acids Res 16, 9893-9908.

Hitomi, Y., Sugiyama, K., and Esumi, H. (1998). Suppression of the 5' splice site mutation in the Nagase analbuminemic rat with mutated U1snRNA. Biochem Biophys Res Commun *251*, 11-16.

Hofmann, Y., Lorson, C.L., Stamm, S., Androphy, E.J., and Wirth, B. (2000). Htra2-beta 1 stimulates an exonic splicing enhancer and can restore full-length SMN expression to survival motor neuron 2 (SMN2). Proc Natl Acad Sci U S A *97*, 9618-9623.

Hofmann, Y., and Wirth, B. (2002). hnRNP-G promotes exon 7 inclusion of survival motor neuron (SMN) via direct interaction with Htra2-beta1. Hum Mol Genet *11*, 2037-2049.

Hong, W., Bennett, M., Xiao, Y., Feld Kramer, R., Wang, C., and Reed, R. (1997). Association of U2 snRNP with the spliceosomal complex E. Nucleic Acids Res *25*, 354-361.

Honore, B., Rasmussen, H.H., Vorum, H., Dejgaard, K., Liu, X., Gromov, P., Madsen, P., Gesser, B., Tommerup, N., and Celis, J.E. (1995). Heterogeneous nuclear ribonucleoproteins H, H', and F are members of a ubiquitously expressed subfamily of related but distinct proteins encoded by genes mapping to different chromosomes. J Biol Chem *270*, 28780-28789.

Horowitz, D.S., and Krainer, A.R. (1994). Mechanisms for selecting 5' splice sites in mammalian pre-mRNA splicing. Trends Genet *10*, 100-106.

Houdayer, C., Dehainault, C., Mattler, C., Michaux, D., Caux-Moncoutier, V., Pages-Berhouet, S., d'Enghien, C.D., Lauge, A., Castera, L., Gauthier-Villars, M., *et al.* (2008).

Evaluation of in silico splice tools for decision-making in molecular diagnosis. Hum Mutat *29*, 975-982.

House, A.E., and Lynch, K.W. (2006). An exonic splicing silencer represses spliceosome assembly after ATP-dependent exon recognition. Nat Struct Mol Biol *13*, 937-944.

Huang, Y., and Steitz, J.A. (2005). SRprises along a messenger's journey. Mol Cell *17*, 613-615.

Huh, G.S., and Hynes, R.O. (1993). Elements regulating an alternatively spliced exon of the rat fibronectin gene. Mol Cell Biol *13*, 5301-5314.

Hui, J., Hung, L.H., Heiner, M., Schreiner, S., Neumuller, N., Reither, G., Haas, S.A., and Bindereif, A. (2005). Intronic CA-repeat and CA-rich elements: a new class of regulators of mammalian alternative splicing. EMBO J *24*, 1988-1998.

Hull, J., Shackleton, S., and Harris, A. (1993). Abnormal mRNA splicing resulting from three different mutations in the CFTR gene. Hum Mol Genet *2*, 689-692.

Hull, J., Shackleton, S., and Harris, A. (1994). Analysis of mutations and alternative splicing patterns in the CFTR gene using mRNA derived from nasal epithelial cells. Hum Mol Genet *3*, 1141-1146.

Hurst, L.D., and Pal, C. (2001). Evidence for purifying selection acting on silent sites in BRCA1. Trends Genet *17*, 62-65.

Jeffreys, A.J., and Flavell, R.A. (1977). The rabbit beta-globin gene contains a large large insert in the coding sequence. Cell *12*, 1097-1108.

Jurica, M.S., and Moore, M.J. (2003). Pre-mRNA splicing: awash in a sea of proteins. Mol Cell *12*, 5-14.

Kadener, S., Fededa, J.P., Rosbash, M., and Kornblihtt, A.R. (2002). Regulation of alternative splicing by a transcriptional enhancer through RNA pol II elongation. Proc Natl Acad Sci U S A *99*, 8185-8190.

Kamma, H., Portman, D.S., and Dreyfuss, G. (1995). Cell type-specific expression of hnRNP proteins. Exp Cell Res *221*, 187-196.

Kan, Z., Garrett-Engele, P.W., Johnson, J.M., and Castle, J.C. (2005). Evolutionarily conserved and diverged alternative splicing events show different expression and functional profiles. Nucleic Acids Res *33*, 5659-5666.

Kanopka, A., Muhlemann, O., and Akusjarvi, G. (1996). Inhibition by SR proteins of splicing of a regulated adenovirus pre- mRNA. Nature *381*, 535-538.

Karni, R., de Stanchina, E., Lowe, S.W., Sinha, R., Mu, D., and Krainer, A.R. (2007). The gene encoding the splicing factor SF2/ASF is a proto-oncogene. Nat Struct Mol Biol *14*, 185-193.

Kashima, T., Rao, N., David, C.J., and Manley, J.L. (2007). hnRNP A1 functions with specificity in repression of SMN2 exon 7 splicing. Hum Mol Genet *16*, 3149-3159.

Kerem, E., and Kerem, B. (1996). Genotype-phenotype correlations in cystic fibrosis. Pediatr Pulmonol *22*, 387-395.

Kielkopf, C.L., Lucke, S., and Green, M.R. (2004). U2AF homology motifs: protein recognition in the RRM world. Genes Dev *18*, 1513-1526.

Koch, C., and Hoiby, N. (1993). Pathogenesis of cystic fibrosis. Lancet *341*, 1065-1069.

Kohtz, J.D., Jamison, S.F., Will, C.L., Zuo, P., Luhrmann, R., Garcia-Blanco, M.A., and Manley, J.L. (1994). Protein-protein interactions and 5'-splice-site recognition in mammalian mRNA precursors. Nature *368*, 119-124.

Kornblihtt, A.R., de la Mata, M., Fededa, J.P., Munoz, M.J., and Nogues, G. (2004). Multiple links between transcription and splicing. RNA *10*, 1489-1498.

Krainer, A.R., Conway, G.C., and Kozak, D. (1990). The essential pre-mRNA splicing factor SF2 influences 5' splice site selection by activating proximal sites. Cell *62*, 35-42.

Ladd, A.N., and Cooper, T.A. (2002). Finding signals that regulate alternative splicing in the post-genomic era. Genome Biol *3*, reviews0008.

Ladd, A.N., Stenberg, M.G., Swanson, M.S., and Cooper, T.A. (2005). Dynamic balance between activation and repression regulates pre-mRNA alternative splicing during heart development. Dev Dyn *233*, 783-793.

Lamond, A.I. (1993). The spliceosome. Bioessays *15*, 595-603.

Lavigueur, A., La Branche, H., Kornblihtt, A.R., and Chabot, B. (1993). A splicing enhancer in the human fibronectin alternate ED1 exon interacts with SR proteins and stimulates U2 snRNP binding. Genes Dev 7, 2405-2417.

Lemaire, R., Prasad, J., Kashima, T., Gustafson, J., Manley, J.L., and Lafyatis, R. (2002). Stability of a PKCI-1-related mRNA is controlled by the splicing factor ASF/SF2: a novel function for SR proteins. Genes Dev 16, 594-607.

Lewis, H.A., Zhao, X., Wang, C., Sauder, J.M., Rooney, I., Noland, B.W., Lorimer, D., Kearins, M.C., Conners, K., Condon, B., et al. (2005). Impact of the deltaF508 mutation in first nucleotide-binding domain of human cystic fibrosis transmembrane conductance regulator on domain folding and structure. J Biol Chem 280, 1346-1353.

Li, X., and Manley, J.L. (2005). Inactivation of the SR protein splicing factor ASF/SF2 results in genomic instability. Cell 122, 365-378.

Li, X., Park, W.J., Pyeritz, R.E., and Jabs, E.W. (1995). Effect on splicing of a silent FGFR2 mutation in Crouzon syndrome. Nat Genet 9, 232-233.

Lim, S.R., and Hertel, K.J. (2001). Modulation of survival motor neuron pre-mRNA splicing by inhibition of alternative 3' splice site pairing. J Biol Chem 276, 45476-45483.

Lin, S., Coutinho-Mansfield, G., Wang, D., Pandit, S., and Fu, X.D. (2008). The splicing factor SC35 has an active role in transcriptional elongation. Nat Struct Mol Biol 15, 819-826.

Lodish, H., Berk, A., Zipursky, S.L., Matsudaira, P., Baltimore, D., and Darnell, J. (2000). Molecular Cell Biology, 4th edn (W. H. Freeman).

Lodish, H., Berk, A., Zipursky, S. L., Matsudaira, P., Baltimore, D., & Darnell, J. E. (2000). Molecular cell biology (4th edition).

Longman, D., Johnstone, I.L., and Caceres, J.F. (2000). Functional characterization of SR and SR-related genes in Caenorhabditis elegans. Embo J *19*, 1625-1637.

Lopez-Bigas, N., Audit, B., Ouzounis, C., Parra, G., and Guigo, R. (2005). Are splicing mutations the most frequent cause of hereditary disease? FEBS Lett *579*, 1900-1903.

Madhani, H.D., and Guthrie, C. (1992). A novel base-pairing interaction between U2 and U6 snRNAs suggests a mechanism for the catalytic activation of the spliceosome. Cell *71*, 803-817.

Maniatis, T., and Reed, R. (2002). An extensive network of coupling among gene expression machines. Nature *416*, 499-506.

Maniatis, T., and Tasic, B. (2002). Alternative pre-mRNA splicing and proteome expansion in metazoans. Nature *418*, 236-243.

Manley, J.L., and Tacke, R. (1996). SR proteins and splicing control. Genes Dev *10*, 1569-1579.

Marcucci, R., Baralle, F.E., and Romano, M. (2006). Complex splicing control of the human Thrombopoietin gene by intronic G runs. Nucleic Acids Res.

Marcucci, R., Baralle, F.E., and Romano, M. (2007). Complex splicing control of the human Thrombopoietin gene by intronic G runs. Nucleic Acids Res *35*, 132-142.

Markovtsov, V., Nikolic, J.M., Goldman, J.A., Turck, C.W., Chou, M.Y., and Black, D.L. (2000). Cooperative assembly of an hnRNP complex induced by a tissue-specific homolog of polypyrimidine tract binding protein. Mol Cell Biol *20*, 7463-7479.

Maroney, P.A., Romfo, C.M., and Nilsen, T.W. (2000). Functional recognition of 5' splice site by U4/U6.U5 tri-snRNP defines a novel ATP-dependent step in early spliceosome assembly. Mol Cell *6*, 317-328.

Martinez-Contreras, R., Cloutier, P., Shkreta, L., Fisette, J.F., Revil, T., and Chabot, B. (2007). hnRNP proteins and splicing control. Adv Exp Med Biol *623*, 123-147.

Matter, N., Herrlich, P., and Konig, H. (2002). Signal-dependent regulation of splicing via phosphorylation of Sam68. Nature *420*, 691-695.

Matter, N., Marx, M., Weg-Remers, S., Ponta, H., Herrlich, P., and Konig, H. (2000). Heterogeneous ribonucleoprotein A1 is part of an exon-specific splice-silencing complex controlled by oncogenic signaling pathways. J Biol Chem *275*, 35353-35360.

Mayeda, A., Screaton, G.R., Chandler, S.D., Fu, X.D., and Krainer, A.R. (1999). Substrate specificities of SR proteins in constitutive splicing are determined by their RNA recognition motifs and composite pre-mRNA exonic elements. Mol Cell Biol *19*, 1853-1863.

McCullough, A.J., and Berget, S.M. (1997). G triplets located throughout a class of small vertebrate introns enforce intron borders and regulate splice site selection. Mol Cell Biol *17*, 4562-4571.

Meyer, I.M., and Miklos, I. (2005). Statistical evidence for conserved, local secondary structure in the coding regions of eukaryotic mRNAs and pre-mRNAs. Nucleic Acids Res *33*, 6338-6348.

Minovitsky, S., Gee, S.L., Schokrpur, S., Dubchak, I., and Conboy, J.G. (2005). The splicing regulatory element, UGCAUG, is phylogenetically and spatially conserved in introns that flank tissue-specific alternative exons. Nucleic Acids Res *33*, 714-724.

Modrek, B., and Lee, C. (2002). A genomic view of alternative splicing. Nat Genet *30*, 13-19.

Moldon, A., Malapeira, J., Gabrielli, N., Gogol, M., Gomez-Escoda, B., Ivanova, T., Seidel, C., and Ayte, J. (2008). Promoter-driven splicing regulation in fission yeast. Nature *455*, 997-1000.

Moore, M.J., and Sharp, P.A. (1993). Evidence for two active sites in the spliceosome provided by stereochemistry of pre-mRNA splicing. Nature *365*, 364-368.

Munoz, M.J., Perez Santangelo, M.S., Paronetto, M.P., de la Mata, M., Pelisch, F., Boireau, S., Glover-Cutter, K., Ben-Dov, C., Blaustein, M., Lozano, J.J., *et al.* (2009). DNA damage regulates alternative splicing through inhibition of RNA polymerase II elongation. Cell *137*, 708-720.

Muro, A.F., Caputi, M., Pariyarath, R., Pagani, F., Buratti, E., and Baralle, F.E. (1999). Regulation of fibronectin EDA exon alternative splicing: possible role of RNA secondary structure for enhancer display. Mol Cell Biol *19*, 2657-2671.

Muro, A.F., Iaconcig, A., and Baralle, F.E. (1998). Regulation of the fibronectin EDA exon alternative splicing. Cooperative role of the exonic enhancer element and the 5' splicing site. FEBS Lett *437*, 137-141.

Nagasaki, H., Arita, M., Nishizawa, T., Suwa, M., and Gotoh, O. (2005). Species-specific variation of alternative splicing and transcriptional initiation in six eukaryotes. Gene *364*, 53-62.

Nasim, F.U., Hutchison, S., Cordeau, M., and Chabot, B. (2002). High-affinity hnRNP A1 binding sites and duplex-forming inverted repeats have similar effects on 5' splice site selection in support of a common looping out and repression mechanism. Rna *8*, 1078-1089.

Nelson, K.K., and Green, M.R. (1989). Mammalian U2 snRNP has a sequence-specific RNA-binding activity. Genes Dev *3*, 1562-1571.

Nilsen, T.W. (1994). RNA-RNA interactions in the spliceosome: unraveling the ties that bind. Cell *78*, 1-4.

Nissim-Rafinia, M., and Kerem, B. (2005). The splicing machinery is a genetic modifier of disease severity. Trends Genet *21*, 480-483.

Nogues, G., Kadener, S., Cramer, P., Bentley, D., and Kornblihtt, A.R. (2002). Transcriptional activators differ in their abilities to control alternative splicing. J Biol Chem *277*, 43110-43114.

O'Keefe, R.T., Norman, C., and Newman, A.J. (1996). The invariant U5 snRNA loop 1 sequence is dispensable for the first catalytic step of pre-mRNA splicing in yeast. Cell *86*, 679-689.

Ostareck-Lederer, A., Ostareck, D.H., and Hentze, M.W. (1998). Cytoplasmic regulatory functions of the KH-domain proteins hnRNPs K and E1/E2. Trends Biochem Sci *23*, 409-411.

Padgett, R.A., Grabowski, P.J., Konarska, M.M., Seiler, S., and Sharp, P.A. (1986). Splicing of messenger RNA precursors. Annu Rev Biochem *55*, 1119-1150.

Pagani, F., and Baralle, F.E. (2004). Genomic variants in exons and introns: identifying the splicing spoilers. Nat Rev Genet *5*, 389-396.

Pagani, F., Buratti, E., Stuani, C., and Baralle, F.E. (2003a). Missense, nonsense, and neutral mutations define juxtaposed regulatory elements of splicing in cystic fibrosis transmembrane regulator exon 9. J Biol Chem *278*, 26580-26588.

Pagani, F., Raponi, M., and Baralle, F.E. (2005). Synonymous mutations in CFTR exon 12 affect splicing and are not neutral in evolution. Proc Natl Acad Sci U S A *102*, 6368-6372.

Pagani, F., Stuani, C., Tzetis, M., Kanavakis, E., Efthymiadou, A., Doudounakis, S., Casals, T., and Baralle, F.E. (2003b). New type of disease causing mutations: the example

of the composite exonic regulatory elements of splicing in CFTR exon 12. Hum Mol Genet *12*, 1111-1120.

Pagani, F., Stuani, C., Zuccato, E., Kornblihtt, A.R., and Baralle, F.E. (2003c). Promoter architecture modulates CFTR exon 9 skipping. J Biol Chem *278*, 1511-1517.

Pal, C., Papp, B., and Lercher, M.J. (2006). An integrated view of protein evolution. Nat Rev Genet *7*, 337-348.

Pan, Q., Bakowski, M.A., Morris, Q., Zhang, W., Frey, B.J., Hughes, T.R., and Blencowe, B.J. (2005). Alternative splicing of conserved exons is frequently species-specific in human and mouse. Trends Genet *21*, 73-77.

Parmley, J.L., Chamary, J.V., and Hurst, L.D. (2006). Evidence for purifying selection against synonymous mutations in mammalian exonic splicing enhancers. Mol Biol Evol *23*, 301-309.

Parmley, J.L., and Hurst, L.D. (2007a). How common are intragene windows with KA > KS owing to purifying selection on synonymous mutations? J Mol Evol *64*, 646-655.

Parmley, J.L., and Hurst, L.D. (2007b). How do synonymous mutations affect fitness? Bioessays *29*, 515-519.

Pellizzoni, L. (2007). Chaperoning ribonucleoprotein biogenesis in health and disease. EMBO Rep *8*, 340-345.

Ram, O., and Ast, G. (2007). SR proteins: a foot on the exon before the transition from intron to exon definition. Trends Genet *23*, 5-7.

Raponi, M., Buratti, E., Dassie, E., Upadhyaya, M., and Baralle, D. (2009). Low U1 snRNP dependence at the NF1 exon 29 donor splice site. Febs J *276*, 2060-2073.

Rappsilber, J., Ryder, U., Lamond, A.I., and Mann, M. (2002). Large-scale proteomic analysis of the human spliceosome. Genome Res *12*, 1231-1245.

Reed, R. (1989). The organization of 3' splice-site sequences in mammalian introns. Genes Dev *3*, 2113-2123.

Reed, R. (2000). Mechanisms of fidelity in pre-mRNA splicing. Curr Opin Cell Biol *12*, 340-345.

Richard, I., and Beckmann, J.S. (1995). How neutral are synonymous codon mutations? Nat Genet *10*, 259.

Riordan, J.R., Rommens, J.M., Kerem, B., Alon, N., Rozmahel, R., Grzelczak, Z., Zielenski, J., Lok, S., Plavsic, N., Chou, J.L., *et al.* (1989). Identification of the cystic fibrosis gene: cloning and characterization of complementary DNA. Science *245*, 1066-1073.

Roberts, G.C., Gooding, C., Mak, H.Y., Proudfoot, N.J., and Smith, C.W. (1998). Co-transcriptional commitment to alternative splice site selection. Nucleic Acids Res *26*, 5568-5572.

Rocha, E.P., and Danchin, A. (2004). An analysis of determinants of amino acids substitution rates in bacterial proteins. Mol Biol Evol *21*, 108-116.

Roscigno, R.F., and Garcia-Blanco, M.A. (1995). SR proteins escort the U4/U6.U5 tri-snRNP to the spliceosome. Rna *1*, 692-706.

Ruskin, B., Greene, J.M., and Green, M.R. (1985). Cryptic branch point activation allows accurate in vitro splicing of human beta-globin intron mutants. Cell *41*, 833-844.

Sakharkar, M.K., Chow, V.T., and Kangueane, P. (2004). Distributions of exons and introns in the human genome. In Silico Biol *4*, 387-393.

Sakharkar, M.K., Perumal, B.S., Sakharkar, K.R., and Kangueane, P. (2005). An analysis on gene architecture in human and mouse genomes. In Silico Biol *5*, 347-365.

Sanford, J.R., Ellis, J., and Caceres, J.F. (2005). Multiple roles of arginine/serine-rich splicing factors in RNA processing. Biochem Soc Trans *33*, 443-446.

Sanford, J.R., Gray, N.K., Beckmann, K., and Caceres, J.F. (2004). A novel role for shuttling SR proteins in mRNA translation. Genes Dev *18*, 755-768.

Sauliere, J., Sureau, A., Expert-Bezancon, A., and Marie, J. (2006). The polypyrimidine tract binding protein (PTB) represses splicing of exon 6B from the beta-tropomyosin pre-mRNA by directly interfering with the binding of the U2AF65 subunit. Mol Cell Biol *26*, 8755-8769.

Schaal, T.D., and Maniatis, T. (1999a). Multiple distinct splicing enhancers in the protein-coding sequences of a constitutively spliced pre-mRNA. Mol Cell Biol *19*, 261-273.

Schaal, T.D., and Maniatis, T. (1999b). Selection and characterization of pre-mRNA splicing enhancers: identification of novel SR protein-specific enhancer sequences. Mol Cell Biol *19*, 1705-1719.

Schaub, M.C., Lopez, S.R., and Caputi, M. (2007). Members of the heterogeneous nuclear ribonucleoprotein H family activate splicing of an HIV-1 splicing substrate by promoting formation of ATP-dependent spliceosomal complexes. J Biol Chem *282*, 13617-13626.

Schmidt, C., Lenz, C., Grote, M., Luhrmann, R., and Urlaub, H. Determination of protein stoichiometry within protein complexes using absolute quantification and multiple reaction monitoring. Anal Chem *82*, 2784-2796.

Schmucker, D., Clemens, J.C., Shu, H., Worby, C.A., Xiao, J., Muda, M., Dixon, J.E., and Zipursky, S.L. (2000). Drosophila Dscam is an axon guidance receptor exhibiting extraordinary molecular diversity. Cell *101*, 671-684.

Sharma, S., Falick, A.M., and Black, D.L. (2005). Polypyrimidine tract binding protein blocks the 5' splice site-dependent assembly of U2AF and the prespliceosomal E complex. Mol Cell *19*, 485-496.

Sharp, P.A. (1994). Split genes and RNA splicing. Cell *77*, 805-815.

Sharp, P.M., Shields, D.C., Wolfe, K.H., and Li, W.H. (1989). Chromosomal location and evolutionary rate variation in enterobacterial genes. Science *246*, 808-810.

Shen, H., Kan, J.L., and Green, M.R. (2004). Arginine-serine-rich domains bound at splicing enhancers contact the branchpoint to promote prespliceosome assembly. Mol Cell *13*, 367-376.

Shin, C., Feng, Y., and Manley, J.L. (2004). Dephosphorylated SRp38 acts as a splicing repressor in response to heat shock. Nature *427*, 553-558.

Singh, N.N., Androphy, E.J., and Singh, R.N. (2004). An extended inhibitory context causes skipping of exon 7 of SMN2 in spinal muscular atrophy. Biochem Biophys Res Commun *315*, 381-388.

Sinha, R., Allemand, E., Zhang, Z., Karni, R., Myers, M.P., and Krainer, A.R. (2010). Arginine methylation controls the subcellular localization and functions of the oncoprotein splicing factor SF2/ASF. Mol Cell Biol *30*, 2762-2774.

Sironi, M., Menozzi, G., Riva, L., Cagliani, R., Comi, G.P., Bresolin, N., Giorda, R., and Pozzoli, U. (2004). Silencer elements as possible inhibitors of pseudoexon splicing. Nucleic Acids Res *32*, 1783-1791.

Smith, C.W., and Nadal-Ginard, B. (1989). Mutually exclusive splicing of alpha-tropomyosin exons enforced by an unusual lariat branch point location: implications for constitutive splicing. Cell *56*, 749-758.

Smith, C.W., and Valcarcel, J. (2000). Alternative pre-mRNA splicing: the logic of combinatorial control. Trends Biochem Sci *25*, 381-388.

146

Smith, N.G., Knight, R., and Hurst, L.D. (1999). Vertebrate genome evolution: a slow shuffle or a big bang? Bioessays *21*, 697-703.

Staffa, A., Acheson, N.H., and Cochrane, A. (1997). Novel exonic elements that modulate splicing of the human fibronectin EDA exon. J Biol Chem *272*, 33394-33401.

Staley, J.P., and Guthrie, C. (1998). Mechanical devices of the spliceosome: motors, clocks, springs, and things. Cell *92*, 315-326.

Stamm, S. (2008). Regulation of alternative splicing by reversible protein phosphorylation. J Biol Chem *283*, 1223-1227.

Sterner, D.A., Carlo, T., and Berget, S.M. (1996). Architectural limits on split genes. Proc Natl Acad Sci U S A *93*, 15081-15085.

Stickeler, E., Fraser, S.D., Honig, A., Chen, A.L., Berget, S.M., and Cooper, T.A. (2001). The RNA binding protein YB-1 binds A/C-rich exon enhancers and stimulates splicing of the CD44 alternative exon v4. Embo J *20*, 3821-3830.

Strong, T.V., Smit, L.S., Nasr, S., Wood, D.L., Cole, J.L., Iannuzzi, M.C., Stern, R.C., and Collins, F.S. (1992). Characterization of an intron 12 splice donor mutation in the cystic fibrosis transmembrane conductance regulator (CFTR) gene. Hum Mutat *1*, 380-387.

Sun, H., and Chasin, L.A. (2000). Multiple splicing defects in an intronic false exon. Mol Cell Biol *20*, 6414-6425.

Sun, Q., Mayeda, A., Hampson, R.K., Krainer, A.R., and Rottman, F.M. (1993). General splicing factor SF2/ASF promotes alternative splicing by binding to an exonic splicing enhancer. Genes Dev 7, 2598-2608.

Tacke, R., Chen, Y., and Manley, J.L. (1997). Sequence-specific RNA binding by an SR protein requires RS domain phosphorylation: creation of an SRp40-specific splicing enhancer. Proc Natl Acad Sci U S A 94, 1148-1153.

Tacke, R., and Manley, J.L. (1999). Functions of SR and Tra2 proteins in pre-mRNA splicing regulation. Proc Soc Exp Biol Med 220, 59-63.

Tanaka, K., Watakabe, A., and Shimura, Y. (1994). Polypurine sequences within a downstream exon function as a splicing enhancer. Mol Cell Biol 14, 1347-1354.

Tange, T.O., Damgaard, C.K., Guth, S., Valcarcel, J., and Kjems, J. (2001). The hnRNP A1 protein regulates HIV-1 tat splicing via a novel intron silencer element. EMBO J 20, 5748-5758.

Tarn, W.Y., and Steitz, J.A. (1994). SR proteins can compensate for the loss of U1 snRNP functions in vitro. Genes Dev 8, 2704-2717.

Tejedor, J.R., and Valcarcel, J. (2010). Gene regulation: Breaking the second genetic code. Nature 465, 45-46.

Tian, M., and Maniatis, T. (1994). A splicing enhancer exhibits both constitutive and regulated activities. Genes Dev 8, 1703-1712.

Tisserant, A., and Konig, H. (2008). Signal-regulated Pre-mRNA occupancy by the general splicing factor U2AF. PLoS One *3*, e1418.

Turner, I.A., Norman, C.M., Churcher, M.J., and Newman, A.J. (2004). Roles of the U5 snRNP in spliceosome dynamics and catalysis. Biochem Soc Trans *32*, 928-931.

Ule, J., Stefani, G., Mele, A., Ruggiu, M., Wang, X., Taneri, B., Gaasterland, T., Blencowe, B.J., and Darnell, R.B. (2006). An RNA map predicting Nova-dependent splicing regulation. Nature *444*, 580-586.

Umen, J.G., and Guthrie, C. (1995). The second catalytic step of pre-mRNA splicing. Rna *1*, 869-885.

Valcarcel, J., Gaur, R.K., Singh, R., and Green, M.R. (1996a). Interaction of U2AF65 RS region with pre-mRNA branch point and promotion of base pairing with U2 snRNA [corrected]. Science *273*, 1706-1709.

Valcarcel, J., Gaur, R.K., Singh, R., and Green, M.R. (1996b). Interaction of U2AF65 RS region with pre-mRNA branch point and promotion of base pairing with U2 snRNA [corrected] [published erratum appears in Science 1996 Oct 4;274(5284):21]. Science *273*, 1706-1709.

van der Voort, R., Verweij, V., de Witte, T.M., Lasonder, E., Adema, G.J., and Dolstra, H. An alternatively spliced CXCL16 isoform expressed by dendritic cells is a secreted chemoattractant for CXCR6+ cells. J Leukoc Biol.

Venables, J.P., Bourgeois, C.F., Dalgliesh, C., Kister, L., Stevenin, J., and Elliott, D.J. (2005). Up-regulation of the ubiquitous alternative splicing factor Tra2beta causes inclusion of a germ cell-specific exon. Hum Mol Genet *14*, 2289-2303.

Venables, J.P., Dalgliesh, C., Paronetto, M.P., Skitt, L., Thornton, J.K., Saunders, P.T., Sette, C., Jones, K.T., and Elliott, D.J. (2004). SIAH1 targets the alternative splicing factor T-STAR for degradation by the proteasome. Hum Mol Genet *13*, 1525-1534.

Venables, J.P., Koh, C.S., Froehlich, U., Lapointe, E., Couture, S., Inkel, L., Bramard, A., Paquet, E.R., Watier, V., Durand, M., *et al.* (2008). Multiple and specific mRNA processing targets for the major human hnRNP proteins. Mol Cell Biol *28*, 6033-6043.

Wagner, E.J., Carstens, R.P., and Garcia-Blanco, M.A. (1999). A novel isoform ratio switch of the polypyrimidine tract binding protein. Electrophoresis *20*, 1082-1086.

Wagner, E.J., and Garcia-Blanco, M.A. (2001). Polypyrimidine tract binding protein antagonizes exon definition. Mol Cell Biol *21*, 3281-3288.

Wahl, M.C., Will, C.L., and Luhrmann, R. (2009). The spliceosome: design principles of a dynamic RNP machine. Cell *136*, 701-718.

Wang, E.T., Sandberg, R., Luo, S., Khrebtukova, I., Zhang, L., Mayr, C., Kingsmore, S.F., Schroth, G.P., and Burge, C.B. (2008). Alternative isoform regulation in human tissue transcriptomes. Nature *456*, 470-476.

Wang, G.S., and Cooper, T.A. (2007). Splicing in disease: disruption of the splicing code and the decoding machinery. Nat Rev Genet *8*, 749-761.

Wang, Z., Rolish, M.E., Yeo, G., Tung, V., Mawson, M., and Burge, C.B. (2004). Systematic identification and analysis of exonic splicing silencers. Cell *119*, 831-845.

Warnecke, T., Parmley, J.L., and Hurst, L.D. (2008). Finding exonic islands in a sea of non-coding sequence: splicing related constraints on protein composition and evolution are common in intron-rich genomes. Genome Biol *9*, R29.

Wassarman, D.A., and Steitz, J.A. (1992). Interactions of small nuclear RNA's with precursor messenger RNA during in vitro splicing. Science *257*, 1918-1925.

Will, C.L. (1997). snRNP structure and function. In Eukaryotic mRNA processing, e. O.I.P. Krainer AR, ed. (Oxford), pp. 130-173.

Will, C.L., and Luhrmann, R. (1997). Protein functions in pre-mRNA splicing. Curr Opin Cell Biol *9*, 320-328.

Will, C.L., and Luhrmann, R. (2001). Spliceosomal UsnRNP biogenesis, structure and function. Curr Opin Cell Biol *13*, 290-301.

Will, C.L., Rumpler, S., Klein Gunnewiek, J., van Venrooij, W.J., and Luhrmann, R. (1996). In vitro reconstitution of mammalian U1 snRNPs active in splicing: the U1-C protein enhances the formation of early (E) spliceosomal complexes. Nucleic Acids Res *24*, 4614-4623.

Wu, H., Sun, S., Tu, K., Gao, Y., Xie, B., Krainer, A.R., and Zhu, J. (2010). A splicing-independent function of SF2/ASF in microRNA processing. Mol Cell *38*, 67-77.

Wu, J.Y., and Maniatis, T. (1993). Specific interactions between proteins implicated in splice site selection and regulated alternative splicing. Cell *75*, 1061-1070.

Wu, S., Romfo, C.M., Nilsen, T.W., and Green, M.R. (1999). Functional recognition of the 3' splice site AG by the splicing factor U2AF35. Nature *402*, 832-835.

Xing, Y., and Lee, C. (2005). Evidence of functional selection pressure for alternative splicing events that accelerate evolution of protein subsequences. Proc Natl Acad Sci U S A *102*, 13526-13531.

Xu, X., Yang, D., Ding, J.H., Wang, W., Chu, P.H., Dalton, N.D., Wang, H.Y., Bermingham, J.R., Jr., Ye, Z., Liu, F., et al. (2005). ASF/SF2-regulated CaMKIIdelta alternative splicing temporally reprograms excitation-contraction coupling in cardiac muscle. Cell *120*, 59-72.

Yang, Z., and Nielsen, R. (2000). Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. Mol Biol Evol *17*, 32-43.

Yeo, G., Holste, D., Kreiman, G., and Burge, C.B. (2004). Variation in alternative splicing across human tissues. Genome Biol *5*, R74.

Yeo, G.W., Coufal, N.G., Liang, T.Y., Peng, G.E., Fu, X.D., and Gage, F.H. (2009). An RNA code for the FOX2 splicing regulator revealed by mapping RNA-protein interactions in stem cells. Nat Struct Mol Biol *16*, 130-137.

Young, P.J., DiDonato, C.J., Hu, D., Kothary, R., Androphy, E.J., and Lorson, C.L. (2002). SRp30c-dependent stimulation of survival motor neuron (SMN) exon 7 inclusion is facilitated by a direct interaction with hTra2 beta 1. Hum Mol Genet *11*, 577-587.

Zamore, P.D., and Green, M.R. (1989). Identification, purification, and biochemical characterization of U2 small nuclear ribonucleoprotein auxiliary factor. Proc Natl Acad Sci U S A *86*, 9243-9247.

Zamore, P.D., Patton, J.G., and Green, M.R. (1992). Cloning and domain structure of the mammalian splicing factor U2AF. Nature *355*, 609-614.

Zhang, X.H., Arias, M.A., Ke, S., and Chasin, L.A. (2009). Splicing of designer exons reveals unexpected complexity in pre-mRNA splicing. Rna *15*, 367-376.

Zhang, Z., Lotti, F., Dittmar, K., Younis, I., Wan, L., Kasim, M., and Dreyfuss, G. (2008). SMN deficiency causes tissue-specific perturbations in the repertoire of snRNAs and widespread defects in splicing. Cell *133*, 585-600.

Zheng, Z.M., He, P.J., and Baker, C.C. (1999). Function of a bovine papillomavirus type 1 exonic splicing suppressor requires a suboptimal upstream 3' splice site. J Virol *73*, 29-36.

Zheng, Z.M., Huynen, M., and Baker, C.C. (1998). A pyrimidine-rich exonic splicing suppressor binds multiple RNA splicing factors and inhibits spliceosome assembly. Proc Natl Acad Sci U S A *95*, 14088-14093.

Zheng, Z.M., Reid, E.S., and Baker, C.C. (2000). Utilization of the bovine papillomavirus type 1 late-stage-specific nucleotide 3605 3' splice site is modulated by a novel exonic bipartite regulator but not by an intronic purine-rich element. J Virol 74, 10612-10622.

Zhou, H.L., and Lou, H. (2008). Repression of prespliceosome complex formation at two distinct steps by Fox-1/Fox-2 proteins. Mol Cell Biol 28, 5507-5516.

Zhu, J., and Krainer, A.R. (2000). Pre-mRNA splicing in the absence of an SR protein RS domain. Genes Dev 14, 3166-3178.

Zielenski, J. (2000). Genotype and phenotype in cystic fibrosis. Respiration 67, 117-133.

Zielenski, J., Markiewicz, D., Lin, S.P., Huang, F.Y., Yang-Feng, T.L., and Tsui, L.C. (1995). Skipping of exon 12 as a consequence of a point mutation (1898 + 5G-->T) in the cystic fibrosis transmembrane conductance regulator gene found in a consanguineous Chinese family. Clin Genet 47, 125-132.

Zielenski, J., and Tsui, L.C. (1995). Cystic fibrosis: genotypic and phenotypic variations. Annu Rev Genet 29, 777-807.

Zuo, P., and Maniatis, T. (1996). The splicing factor U2AF35 mediates critical protein-protein interactions in constitutive and enhancer-dependent splicing. Genes Dev 10, 1356-1368.

Zuo, P., and Manley, J.L. (1994). The human splicing factor ASF/SF2 can specifically recognize pre-mRNA 5' splice sites. Proc Natl Acad Sci U S A 91, 3363-3367.

Functional properties and evolutionary splicing constraints on a composite exonic regulatory element of splicing in CFTR exon 12

Ariful Haque, Emanuele Buratti and Francisco E. Baralle*

International Centre for Genetic Engineering and Biotechnology (ICGEB), 34149 Trieste, Italy

ABSTRACT

In general, splicing regulatory elements are defined as Enhancers or Silencers depending on their positive or negative effect upon exon inclusion. Often, these sequences are usually present separate from each other in exonic/intronic sequences. The Composite Exonic Splicing Regulatory Elements (CERES) represent an extreme physical overlap of enhancer/silencer activity. As a result, when CERES elements are mutated the consequences on the splicing process are difficult to predict. Here, we show that the functional activity of the CERES2 sequence in CFTR exon 12 is regulated by the binding, in very close proximity to each other, of several SR and hnRNP proteins. Moreover, our results show that practically the entire exon 12 sequence context participate in its definition. The consequences of this situation can be observed at the evolutionary level by comparing changes in conservation of different splicing elements in different species. In conclusion, our study highlights how it is increasingly difficult to define many exonic sequences by simply breaking them down in isolated enhancer/silencer or even neutral elements. The real picture is close to one of continuous competition between positive and negative factors where affinity for the target sequences and other dynamic factors decide the inclusion or exclusion of the exon.

INTRODUCTION

Only a few years ago, mutations in the protein-coding section of genes that did not affect amino acid coding capacities were considered to be neutral with regards to the protein functional properties (and thus the evolutionary fitness of the gene). Since then, advances in both the pre-mRNA splicing and the translational research fields have severely challenged this assumption, especially with regards to its implications in the occurrence of human disease and on evolutionary mechanisms in general (1–3). With regards to the pre-mRNA alternative splicing process (4), this assumption was first challenged by the discovery that splicing regulatory elements (SREs) could be found embedded within exonic coding sequences both in alternative and constitutive splicing (5,6) and that these elements and the strength of the basic splicing consensus motifs (7) could regulate exon inclusion levels. More than 20 years since these discoveries, the importance of splicing regulatory regions within exon coding sequences has, if anything, greatly increased. In fact, global analyses of splicing events (8,9) and the search for these splicing regulatory motifs embedded within exons has received considerable attention, especially through the use of high-throughput and bioinformatics approaches (10,11). The results of all these analyses clearly indicate that SRE elements represent important players in controlling both alternatively and constitutively splicing processes (12,13).

In general, SRE elements are individually referred to as Exonic and Intronic Splicing Enhancers (ESEs and ISEs) or Exonic and Intronic Splicing Silencers (ESSs and ISSs) sequences depending on their localization and functional effects (14–18). The way in which SRE elements exert their action is through the binding of *trans*-acting factors, predominantly belonging to the SR and hnRNP protein families (19–23). It should be noted, however, that the list of proteins capable of modulating splicing is growing every year (24) and much still remains to be uncovered in this area of research. Nonetheless, aside from individual identities, it has been determined long ago that the balance between antagonistic factors binding to a particular SRE element is one key factor in discriminating exon inclusion/exclusion levels (25,26).

*To whom correspondence should be addressed. Tel: +39 040 3757337; Fax: +39 040 3757361; Email: baralle@icgeb.org

The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

The possibility of controlling this balance has provided a great advantage to the eukaryotic cell because, whenever necessary, it can be shifted in one direction or the other through several mechanisms. In fact, beside the intrinsic affinity of splicing factors for their respective cis-targets, the binding level of each factor can be easily modified by variations in its relative expression levels (27–29), posttranslational modifications (30–33) or local RNA folding arrangements that limit/enhance their availability (34–39).

Studying these issues also makes for fascinating insights with regards to the potential relationships between coding and splicing regulatory regions during the course of evolution. It is now clear, in fact, that synonymous or even advantageous substitutions at the protein level may still be selected against if they end up to be harmful with regards to splicing decisions. On the other hand, suboptimal codon usage arrangements may be maintained to preserve correct splicing functionality (1,40). In keeping with this concept, recent analyses have uncovered the presence of extensive purifying selection against substitutions in ESE elements as determined by the reduced single nucleotide polymorphism density in these regions (41) and by the fact that codon usage at the exon–intron boundaries may be considerably affected by the need to maintain SRE sequences (42). It should also be mentioned, however, that these kind of analyses are still hampered by our limited knowledge of SRE sequences and caution should be employed when making these kind of comparisons on the basis of bioinformatic studies (43). For this reason, it is advisable to back up any eventual conclusion with functional experiments that might either support or not the bioinformatics considerations.

Among the known SRE sequences, a particularly interesting class of elements is represented by the discovery of the Composite Exonic Regulatory Elements of Splicing (CERES), that were first identified in human CFTR exon 12 and exon 9 (44,45). Unlike the classical exonic regulatory elements, that tend to predominantly possess either enhancing or silencing properties, the effect of mutagenesis in CERES elements is very unpredictable with regards to splicing outcomes. This makes it very difficult to evaluate the potential pathologic effect of apparently benign substitutions in these regions. In this work, we have combined the analysis of natural pathogenic mutations with a comparative human–mouse genomic approach to better understand the characteristics of one of these CERES elements in CFTR exon 12 (CERES2). The results of this analysis have reinforced the emerging concept that in many cases dividing exonic sequences in well defined enhancer/silencer or neutral splicing regulatory elements does not satisfactorily explain anymore the effects of artificial and natural substitutions. It is only by adopting a more global view of splicing regulatory codes that will allow us to understanding many dynamic sequence interplays aimed at preserving splicing definition of eukaryotic exons.

MATERIALS AND METHODS

Hybrid minigene constructs

Human CFTR exon 12 minigene constructs (T40C, G48C, A49G, A51T, C52T and WT) have been previously described (40,44). Further modification of the exon were introduced by PCR-directed mutagenesis using specific primers and cloned inside the NdeI restriction site of the pTB plasmid. Primer sequences for each described mutants can be provided up on request. Mouse CFTR Exon 12 along with the mouse introns (297 nt of intron 11 and 219 nt of intron 12) was amplified from genomic DNA (*Mus musculas* strain C57BL/6) using primers MCF12 F: 5′-ggctccaggcttgagcatatgtactaatctg-3′ and MCF12 R: 5′-caagaaattggcttcatatgtgatcatcaga-3′. Mouse CFTR exon 12 modifications were also performed by PCR-directed mutagenesis with specific primers.

CFTR exon 12 sequences of different animals were recovered using NCBI BLASTN search. Accession numbers of the sequences are Human (Gene Bank accession number NM_000492.3 *Homo sapiens*), Guinea pig (Gene Bank accession number AF133216.1 *Cavia porcellus*), Ground Squirrel (Gene Bank accession number AC184040.3 *Spermophilus tridecemlineatus*), Mouse (Gene Bank accession number NM_021050.2 *Mus musculus*), Rabbit (*Oryctolagus cuniculus*) GenBank accession no. NM_001082716, Cow (*Bos tortus*) GenBank accession no. NM_174018, Pig (*Sus scrofa*) GenBank accession no. AY585334, Horse (*Equus caballus*) GenBank accession no. NM_001110510, Lemur (*Lemur catta*) GenBank accession no. AC123543.

Cell culture, transfections and RT–PCR analysis

HeLa cells were cultured in Dulbecco's modified Eagle's medium with Glutamax (Invitrogen) in standard conditions. The minigenes used for transfection were purified using phenol–choloroform extraction followed by a sephacryl S-400 (GE healthcare) column purification step. HeLa cells were plated at a concentration of 2.8×10^5 to achieve 80–90% confluence. The following day, 500 ng of plasmid DNA were transfected in the cells using Effectene transfection reagents (Qiagen). In case of *in vivo* overexpression of SR proteins (a kind gift from J. Caceres), 1 μg of expression vector was mixed with 500 ng of minigenes. Finally, after 24 h total RNA was extracted using TRIreagent solution (Ambion). One microgram of total RNA was used in the retrotranscription reaction with random primers and Moloney murine leukemia virus enzyme (Invitrogen). Spliced products from the transfected minigene were obtained using primers Bra2 5′-taggatccggtcaccaggaagttggttaaatca-3′ and α 2–3 5′-caacttcaagctcctaagccactgc-3′. PCR conditions were the following: 94°C for 5 min.; 94°C for 30 s, 55°C for 30 s, and 72°C for 30 s for 30 cycles; and 72°C for 7 min for the final extension. PCRs were optimized to be in the exponential phase of amplification and products were routinely fractionated in 1.8% (wt/vol) agarose gels.

In vitro transcription and synthetic RNA oligo

RNA was transcribed from PCR templates amplified from the respective plasmids. A T7 promoter sequence was added towards the 5′-end of the template using primer carrying T7 sequence and, similarly, a (TG)8 repeated sequence was used to tag the 3′-end. Every time, 5 µg of DNA template was used in a 60 µl T7 polymerase (Stratagene) transcription reaction. The RNA was then purified using standard Acid–Phenol purification method followed by Ethanol precipitation. Synthetic RNA oligos corresponding to Mouse 17–38, Human 17–38, Mouse 63–81 and Human 63–81 sequences were purchased from Eurofins MWG Operon, Germany.

Affinity purification of RNA binding proteins and western blot analysis

Ten micrograms of synthetic RNA oligo or 15 µg of transcribed RNA were oxidized in the dark for an hour with sodium *m*-periodate in a 400 µl reaction mixture (100 mM Sodium acetate pH 5.2 and 5 mM Sodium m-periodate). RNA was then ethanol precipitated and resuspended in 200 µl of 100 mM sodium acetate. Approximately 350 µl of prewashed equilibrated adipic acid dehydrazide-agarose beads (50% slurry; Sigma) were added to each oxidized RNA volume and placed in the rotor at 4°C for overnight incubation. This RNA-Bead covalent link formation was also performed in the dark. The immobilized RNA were then washed once with 1 ml of 2 M NaCl and twice using washing buffer (5.2 mM HEPES pH 7.5, 1 mM $MgCl_2$, 0.8 mM Mg acetate). Meanwhile, 200 µl of nuclear extract was mixed with 900 µl RNAse free water, 1× binding buffer (5.2 mM HEPES pH 7.9, 1 mM $MgCl_2$, 0.8 mM Mg acetate, 0.52 mM dithiothreitol, 3.8% glycerol, 0.75 mM ATP, 1 mM GTP and Heparin at the final concentration of 0.5 µg/µl). The RNA-bound beads were then equilibrated in 300 µl of NE mix and incubated for 25 min on a rotor at room temperature. Beads were then washed four times with 1.5 ml washing buffer. In every washing step beads were gently precipitated by gravity on ice. Finally, 50 µl of 3× SDS loading buffer was added and samples were heated for 5 min before loading on a 10% SDS–polyacrylamide denaturing gel. The gel was then electroblotted onto a polyvinylidene difluoride membrane according to standard protocols (Amersham Biosciences) and blocked with 10% skimmed milk (Non fat dry milk in 1× PBS). Membranes targeted for SR protein recognition were blocked using Western blocking reagent (Roche). Proteins were probed with different antibodies and detected with a chemiluminescence kit (ECL; Pierce Biotechnology). Antibodies against hnRNP U and Tra2 β were kind gifts from G. Dreyfuss and I.C Eperon, respectively. Monoclonal Anti-ASF/SF2, SC35 and 1H4 (against SRp 75, 55, 40) antibodies were purchased from Zymed Laboratories Inc. Changes of protein binding levels have been quantified relative to TDP-43 using an Ultro Scan XL, Pharmacia LKB—laser densitometer at 633 nM wavelength according to manufacturer's instructions.

siRNA knockdown of splicing factors

siRNA sense sequences used for silencing the different target proteins were the following: human hnRNPA1—cagcugaggaagcucuuca (Sigma), and human hnRNPA2—ggaacaguuccguaagcuc (Sigma); human hnRNP C1/C2-gcaaacaagcaguagagau (Sigma), human DAZAP1-gagacucugcgcagcuacu (Dharmacon) and luciferase no. 2 gene control, gccauucuauccucuagaggaug (Dharmacon). HeLa cells were plated at 0.7×10^5 cells per well in 35-mm plates to achieve 30–40% confluence. The next day, 3 µl Oligofectamine (Invitrogen) was combined with 15 µl of Opti-MEM medium (Invitrogen) and 3 µl of 40 µM siRNA duplex oligonucleotides was diluted in a final volume of 180 µl of Opti-MEM medium. The two mixtures were combined and left for 20 min at room temperature. Finally, this mixture was added to the cells, which were mantained in 0.9 ml of Opti-MEM only. After 6 h, 500 µl of 30% FBS (Foetal bovine serum, Invitrogen) was added. Six to eight hours later Opti-MEM was exchanged with Dulbecco's modified Eagle medium and the cells were transfected with the minigene of interest (500 ng) using Qiagen Effectene transfection reagents. On the third day, HeLa cells were harvested for protein and RNA extractions. RT–PCR from total RNA was performed as for the transfection protocol described above. Whole-protein extracts were obtained by cell sonication in lysis buffer (1× PBS and 1× Protease inhibitor cocktail) and analyzed for hnRNPA1, A2, C1/C2 and DAZAP1 endogenous protein expression by immunoblotting using the antibodies described above. Tubulin was used as total protein loading control.

RESULTS

In order to better characterize the CERES2 element in CFTR exon 12 we selected two pathological missense mutations (G48C, A51T) and three samesense substitutions (T40C, A49G and C52T) that were already known to affect CFTR exon 12 splicing when inserted in the pTB minigene (Figure 1). In particular, transfection in HeLa cells of the G48C and A49G substitutions caused exon skipping in the full exon 12 context whilst A51T caused full exon inclusion, as previously reported (44). In addition, to widen the number of mutations under study we also chose the T40C and C52T synonymous substitutions that were already known to cause total exon skipping when inserted in the human context (40). These substitutions are particularly interesting form an evolutionary point of view as they are naturally present in the mouse CFTR exon 12 sequence and C52T has also been reported as a human polymorphism/possible mutation in the Cystic Fibrosis Mutation Database (www.genet.sickkids.on.ca). The position and consequences of all these substitutions when inserted in a CFTR exon 12 minigene are summarized in Figure 1A and B, respectively.

First of all, we were interested to see whether the functional effects of these substitutions were dependent on the context provided by the rest of the exon sequence. To study this, we have analyzed their functional effect in
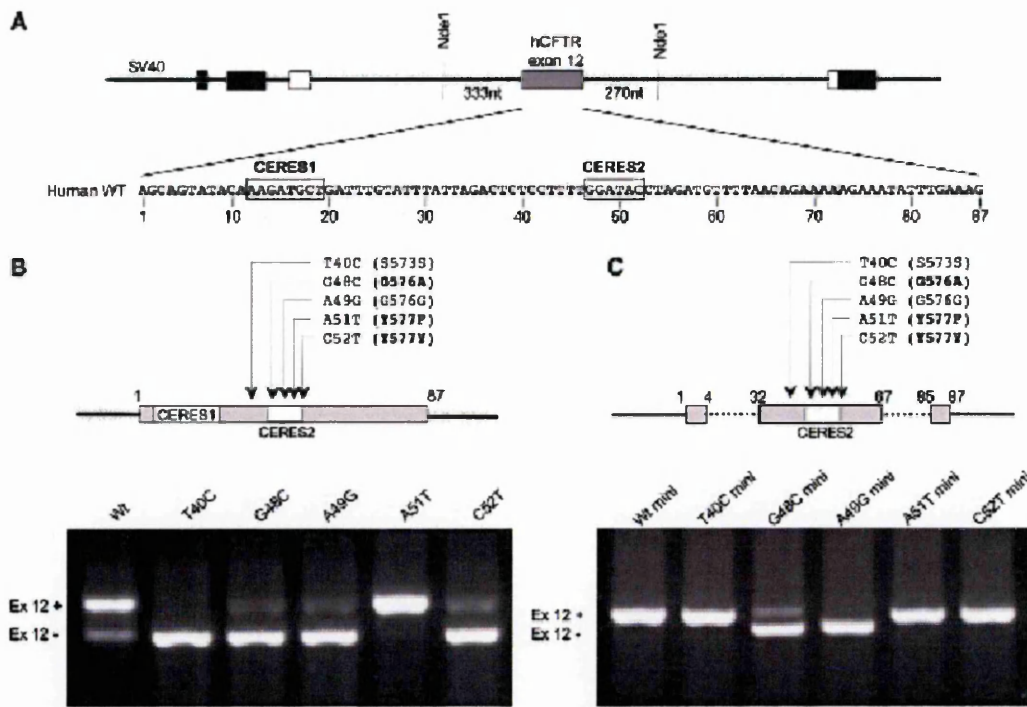
Figure 1. (**A**) Schematic presentation of the hybrid minigene used in transfection experiments. The α-globin, fibronectin EDB, and human CFTR exon 12 are shown as black, white and gray boxes, respectively. The sequence of CFTR exon 12 and position of the CERES1 and CERES2 elements is reported in full. (**B** and **C**) Schematic diagram of both full and mini CFTR Ex. 12 constructs used in the analysis. The vertical superimposed arrows indicate the locations of both natural and synonymous mutations. The amplified RT–PCR products stained with ethidium bromide are shown in the bottom panels. Spliced transcripts are shown with Ex.12+ for inclusion and Ex.12− for exclusion of the exon.

a shortened CFTR exon 12 sequence obtained by removing the regions near the 3 and 5's and downstream regions but maintaining 4 and 3 nts close to the 3′ intron-exon and exon-5′ intron junctions, respectively. We called this construct 'mini' exon 12 (Figure 1C, upper panel). When all the mutations analyzed in Figure 1B were inserted in this reduced context both the G48C and A49G were still capable of inducing exon skipping as observed in the full length exon 12, whilst the enhancing effect of A51T could not be observed owing to the fact that the wild-type mini exon 12 is fully included in the spliced transcript (as opposed to only 80% inclusion of the full length exon 12) (Figure 1C, lower panel). Interestingly, both mouse-specific T40C and C52T substitutions lost the ability to induce exon skipping, suggesting that their effect on the CFTR exon 12 splicing process necessitated the presence of either one or both human flanking regions (Figure 1C, lower panel and Figure 6).

Identifying the trans-acting factors whose binding is affected by these substitutions

Considering that overexpression of the classical splicing factors hnRNP A1 and SF2/ASF was already described to affect CFTR exon 12 splicing (44) it was decided to better characterize the effect of these substitutions in terms of binding to a wide range of SR and hnRNP

splicing factors. To achieve this, we have used a pulldown system previously used in our lab to identify specific RNA binding proteins in a variety of exonic/intronic contexts (46,47). The transcribed RNAs carry a $(UG)_8$ tail that functions as a loading control for the TDP-43 protein (48) (Figure 2A).

In the first analysis, we tested the mini wild-type CFTR exon 12 sequence and two versions carrying the two missense mutations G48C and A51T for binding to the following proteins: hnRNP U, PTB, hnRNP H, DAZAP1, hnRNP C2, A1, A2 and SRp75, SRp55, SRp40, SF2/ASF and Tra2β. The results of this analysis are reported in Figure 2B. This figure shows that no binding could be observed for the hnRNP H, PTB, SRp75, SRp40, SC35 and Tra2β proteins to the mini-exon 12 sequence, both in its wild-type form and carrying either the G48C or the A51T mutations. On the other hand, some of the proteins tested could bind all these sequences, irrespectively of the presence or absence of mutations (hnRNP U and SRp55). Interestingly, a few displayed a differential binding ability in the wild-type sequence with respect to these mutations. In particular, the most striking change could be observed for the SF2/ASF protein that could better bind the A51T mutant with respect to the wild-type sequence. At the same time, it could also bind less efficiently to the G48C mutant than to the wild-type. Finally, hnRNP C2 could also bind less efficiently to the A51T mutant. It is worth noting that
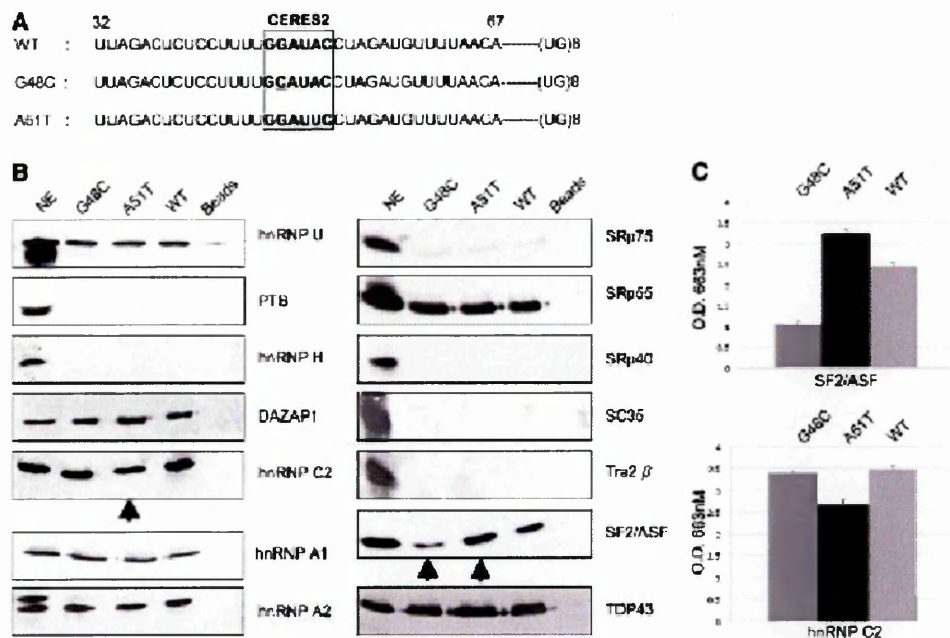
Figure 2. Western blot analysis of recovered proteins after pulldown of two naturally occurring nonsense CFTR mutations (G48C/G576A and A51T/Y577F) compared with CFTR Ex. 12 wild type. *In vitro* transcribed RNA was used for analysis. (**A**) Transcribed RNA sequences used for pulldown analysis with the mutations (underlined). A (UG)$_8$ repeat specific for TDP43 was added at the 3′-end of each RNA to normalize the data after western blot. (**B**) Affinity assay for binding of the following hnRNPs: U, PTB, H, DAZAP1, C2, A1 and A2. (**C**) Affinity assay for binding of the following SR factors: SRp75, SRp55, SRp40, SC35, Tra2β and SF2/ASF. Detection of all these proteins was performed by western blot using specific antibodies. Quantification of SF2/ASF and hnRNP C2 binding levels as determined by densitometric analysis are reported in Figure 2C (normalized by TDP-43). Standard deviation values from three independent experiments are shown.

these changes are not all or nothing effects, but they are small changes then amplified by the combinatorial effect of all the other elements involved. On the other hand, the A51T mutant that in the whole exon 12 context has an exon inclusion enhancing effect, displayed increased SF2/ASF binding levels than the wild-type sequence. Quantification of hnRNP C2 and SF2/ASF binding levels in these experiments (normalized against TDP-43) are reported in Figure 2C as determined using densitometric analysis from three independent experiments.

In the case of the three synonymous mutations (T40C, A49G and C52T), the pulldown experiments yielded less varied results (Figure 3B). In fact, no changes could be observed in the binding profiles of RNAs carrying the T40C and C52T substitutions with respect to the wild-type sequence. However, in the case of the A49G mutation we observed less binding of the SF2/ASF protein than in the wild-type sequence and increased binding of hnRNP C2 (quantification of these proteins are reported in Figure 3C), a situation that made the effects of this mutation very similar to those observed for G48C (Figure 2B). Furthermore, it was also consistent with its inhibitory effect in the mini-exon 12 minigene (Figure 1C). In this respect, the observation that no changes could be observed for any of these proteins in the case of T40C and C52T was also consistent with the functional assays demonstrating that these two

substitutions were neutral in the human mini-context (Figure 1C).

Validating the role played by SR factors in CFTR exon 12 splicing

In order to validate the role played by the SR proteins, we tested the response of both the G48C and A49G minigenes to overexpression of the specific interactors of the mini-exon 12 sequence, SF2/ASF and SRp55 and of SC35 (as an example of a SR protein not interacting with the mini exon). The results shown in Figure 4A demonstrate that both SF2/ASF and SRp55 consistently have a higher enhancing effect on the mini exon 12 inclusion levels than SC35, suggesting that direct interaction provides an advantage over the well known generalized enhancing effect of SR proteins. Interestingly, however, the enhancement observed for the two mutants was not the same, with A49G being less responsive especially for SRp55 overexpression than G48C. Finally, it should be noted that deletion of the central CERES2 region also abolished completely the response of the mini-exon 12 to all SR protein overexpression, demonstrating that their action in the mini-exon context is mediated only through the CERES2 sequence. In parallel, to further rule out non-specific effects of SF2/ASF overexpression, we also performed overexpression analysis of a series of deletion mutants (Figure 4B). Also in this case, mutants lacking either the RRM2 region (ΔRRM2) or the RS domain
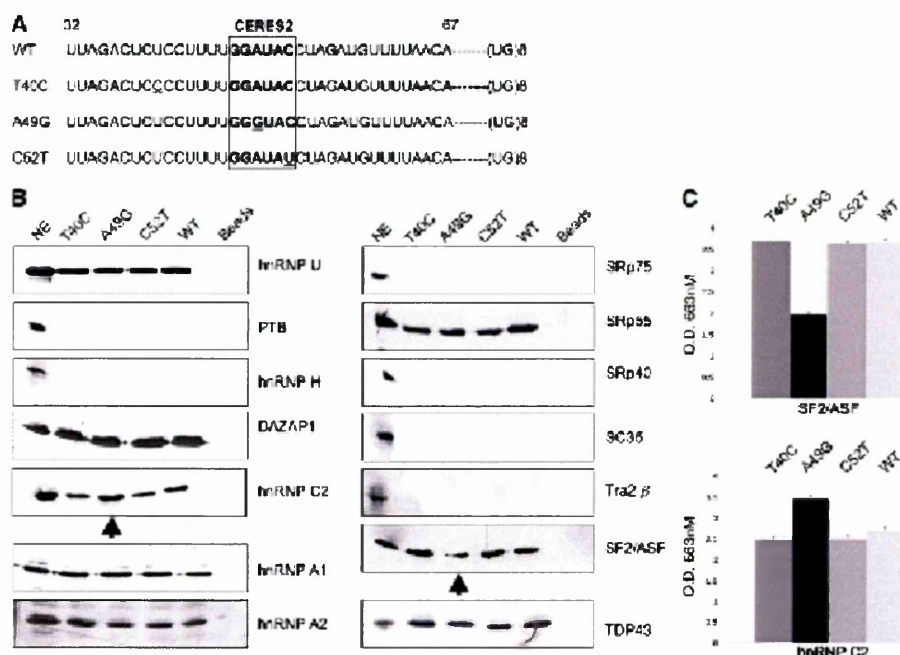
Figure 3. Affinity pulldown performed for the three synonymous mutations (T40C, A49G and C52T) in CFTR exon 12. (A) Transcribed RNA sequences used for pulldown analysis along with their mutations (underlined). A (UG)8 repeat specific for TDP43 was added at the 3′-end of each RNA for pulldown normalization. (B) Affinity pulldown assay for hnRNPs U, PTB, H, DAZAP1, C2, A1 and A2. (C) Affinity pulldown assay for SR proteins SRp75, SRp55, SRp40, SC35, Tra2 β and SF2/ASF. Detection of all these proteins was performed by western blot using specific antibodies. Quantification of SF2/ASF and hnRNP C2 binding levels as determined by densitometric analysis is reported in Figure 3C (normalized by TDP-43). Standard deviation values from three independent experiments are shown.
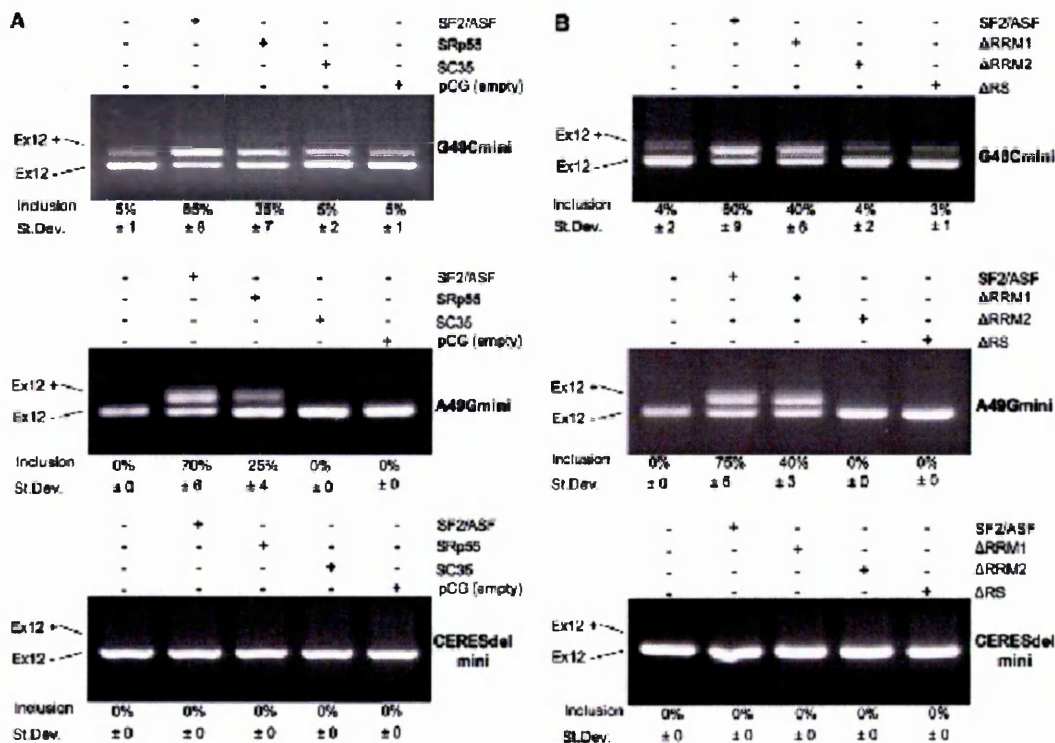

Figure 4. (A) Analysis showing the *in vivo* overexpression of SR proteins (SRp55, SC35 and SF2/ASF) to rescue CFTR exon 12 mutant minigenes in their mini context. In the case of the CERESdel mini construct, 6 nt of CERES2 (GGATAC) was removed. (B) Overexpression of wild-type SF2/ASF and a series of SF2/ASF mutants (ΔRRM1, ΔRRM2 and ΔRS) in the presence CFTR exon 12 mutant minigenes in their mini context. The amplified RT-PCR product of the spliced/unspliced mRNAs are stained using ethidium bromide and run in an agarose 1.8% gel. Exon inclusion and skipping are is shown by Ex12+ and Ex12−, respectively.
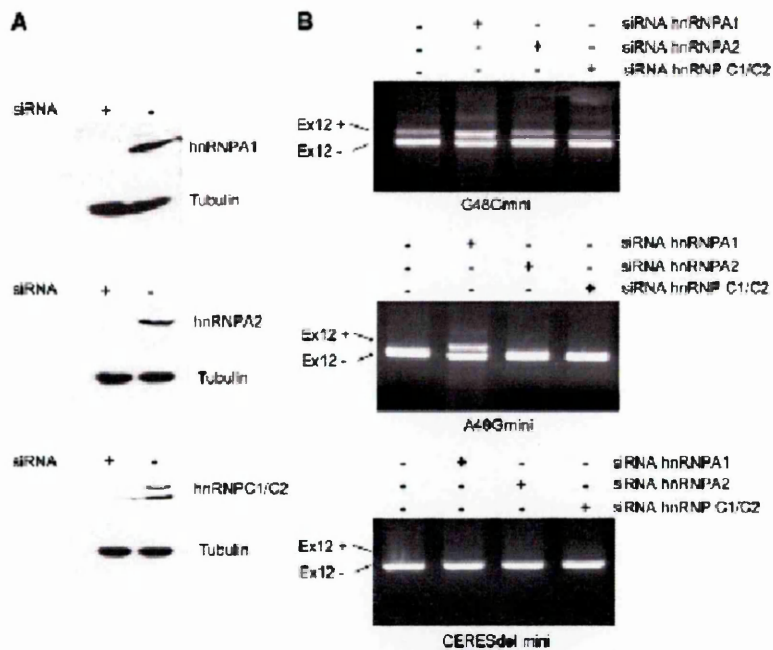
Figure 5. (A) Knockdown of hnRNP factors in HeLa cells using siRNAs (+) against hnRNP A1, A2 and C2. Luciferase siRNA treated cells have been used as a control and correct depletion was determined by western blot assay (extracts have been normalized using tubulin) (B) Effects of transfecting the G48C, A49G and CERESdel minigenes in HeLa cells selectively depleted of each of these hnRNP proteins. RT-PCR samples are stained with ethidium bromide and run on a 1.8% agarose gel. Exon inclusion and skipping are shown by Ex12+ and Ex12–, respectively.

(ΔRS) could not enhance inclusion, further supporting the specificity of this SR protein functional interaction with the CERES2 sequence.

Validating the role played by hnRNP factors in CFTR exon 12 splicing

Because of their abundance in the nuclear extract, overexpression studies for the different hnRNP proteins did not yield satisfactory results (data not shown). For this reason, in order to test effectively the functional effects of the hnRNP interactors found in pulldown analysis we performed individual siRNA-mediated knock-down of the well known hnRNP A1, A2 and C2 proteins (Figure 5A). As shown in Figure 5B, the only siRNA knockdown that could rescue both the G48C and A49G mini-exons inclusion was hnRNP A1. Importantly, knockdown of this protein had no effect on the CERESdel minigene. Finally, no effect could be detected following hnRNP C2 knockdown which is rather surprising considering that the role of this protein in the regulation of splicing control has been recently re-evaluated in high-throughput studies (23). In addition, as there are many hnRNPs with redundant functions we have also tried the simultaneous depletion of A1, A2 and C2 in different combinations but could not confirm their role in the splicing regulation of CFTR exon 12 (data not shown). It should be noted, however, that these results do not mean that only hnRNP A1 can modulate CFTR exon 12 splicing. In fact, some of these proteins could still play an active role in the presence of reduced amounts of positive factors (i.e. SF2/ASF or SRp55) and further work will be needed to clarify this issue.

Recovery of the T40C and C52T inhibitory action through the add back of human and mouse flanking CFTR exon 12 sequences

In order to better characterize the mode of action of these substitutions to the mini-exon 12 minigene, we selectively added back the missing upstream and downstream sequences (both in their human and mouse forms) and observed which mutant was able to recover the inhibitory effect of the T40C/C52T substitutions (Figure 6A). As shown in Figure 6B, right panel, the wild type exon 12 constructs that contain the added-back human and mouse upstream regions display full inclusion (constructs A and D). However, when we inserted back the T40C and C52T mutations the inhibitory effect could be detected only in the constructs with the added back human sequence but not with those with the mouse sequence (compare constructs B–C with E–F). A similar situation could also be observed with the downstream regions. In fact, Figure 6B, middle panel, shows that wild-type exon 12 sequences with the added-back human and mouse downstream region display full inclusion (constructs G and J). However, when T40C and C52T were inserted back it was observed that the inhibitory effect could be detected only in the constructs with the human but not with the mouse downstream sequence (compare constructs H–I with K–L). Interestingly, the integrity of maintaining the mouse polypurinic GAAGA ACAAG motif present in the mouse sequence (underlined
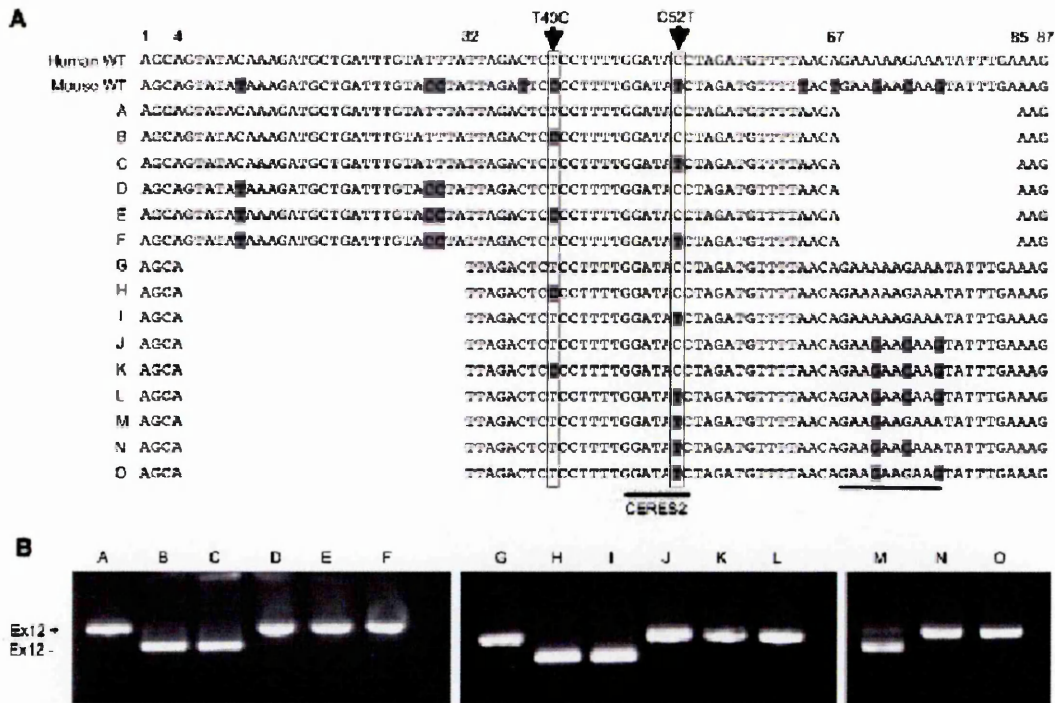
Figure 6. (A) Schematic diagram of the wild-type and hybrid mouse and human CFTR exon 12 sequences used to construct a series of minigenes (labeled A to O). Nucleotide differences in the mouse sequence with respect to the human are boxed. Blank spaces in the alignment represent the sequences removed form the exon. (B) Results of the transfection analysis of the minigenes labeled A to O following transfection in HeLa cells. Exon inclusion is shown by Ex12+ and skipping Ex12−. RT–PCR samples are stained with ethidium bromide and run on a 1.8% agarose gel.

in Figure 6A, bottom) is particularly important. In fact, the majority of substitutions that tend to restore the mouse sequence can successfully withstand the inhibitory action of the C52T substitution (construct N-O as opposed to M).

Taken together, these results suggest that the human and mouse flanking regions have different splicing regulatory properties: human sequences, both upstream and downstream of the central exon 12 region containing CERES2, may be predominantly inhibitory. On the other hand, the mouse upstream and downstream sequences seem to enhance exon recognition.

The hypothesis that mouse sequences enhanced exon recognition was thus tested at the functional level by amplifying a mouse CFTR exon 12 sequence and inserting it in the pTB minigene system (Figure 7A). In this sequence, we then deleted the upstream and downstream regions, either separately or in combination (mutants A–C). The results of this analysis are reported in Figure 7B and show that deleting only the upstream sequence (mutant A) had no effect of mouse CFTR exon 12 inclusion levels. On the other hand, deletion of the downstream sequence (mutant B) resulted in ~15% exon skipping. Interestingly, if both regions are deleted at the same time (mutant C) the levels of exon skipping increase to 25%, indicating that also the polypurinic downstream sequence can function as an ESE once the upstream ESE sequence is absent.

Trans-acting factors binding to the human and mouse 17–38 and 63–81 sequences

Based on these results, it was thus likely that mouse and human flanking sequences could bind a different set of proteins. For this reason, we selected the 17–38 and 63–81 mouse and human regions (Figure 8A) to perform pulldown assays as previously described for the central region (Figures 2–3). The Western blot analysis to check for SR protein binding showed that both mouse sequences could bind SF2/ASF, SRp75 and SRp55 more efficiently than the respective human sequences. In addition, mouse nucleotide stretch 64–82 is also capable of binding SRp40 whilst the human 64–82 sequence is not (Figure 8C). As control, the recognition with antibodies against hnRNP proteins A1 and U showed that both these factors could bind with equally well with the mouse and human sequences. Unfortunately, using this approach we were unable to determine which factors are responsible for the ESS activity of the human 18–39 and 64–82 sequences. Of course, the most probable reason is that we have tested only a fraction of the candidates and there are many potential known or unknown proteins that could exert an effect on splicing regulation. Further work is currently under way to better define this point. In any case, these results were consistent with the _in vivo_ results that detected different functional properties between the mouse and human flanking regions.
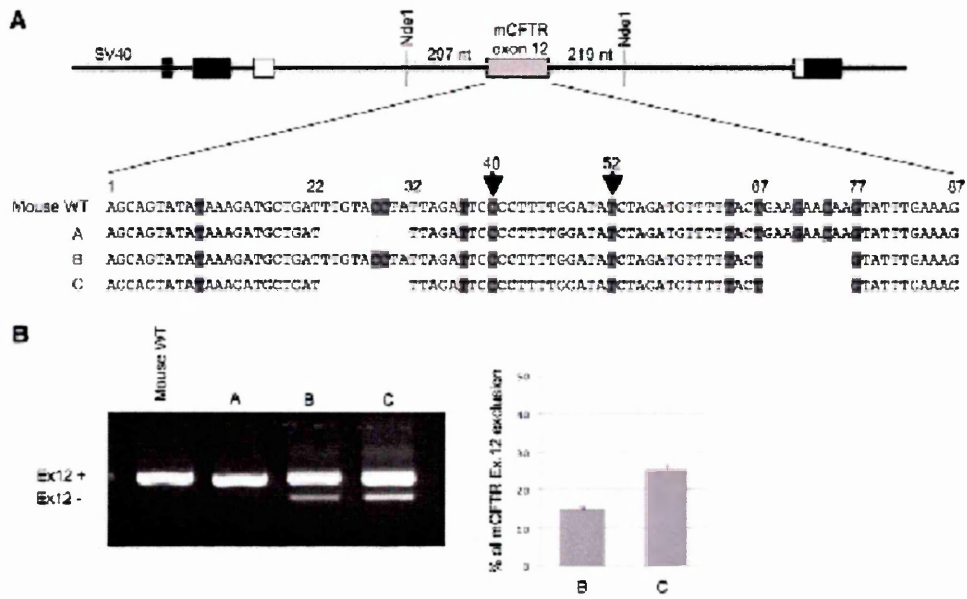
Figure 7. (A) Schematic analysis of the pTB minigene containing the mouse CFTR exon12 sequence (reported in full). Mouse specific nucleotides are boxed. Mutants A, B and C show the combination of nucleotide deletions spanning from nucleotides 23 to 31 and 68 to 76 included (B) Mouse CFTR exon 12 minigenes (wild-type, A, B and C) were transfected in HeLa cells for *in vivo* splicing assay. The amplified RT–PCR products of this splicing assay are stained by ethidium bromide and run on a 1.8% agarose gels. Exon inclusion is shown by Ex12+ and skipping by Ex12−.
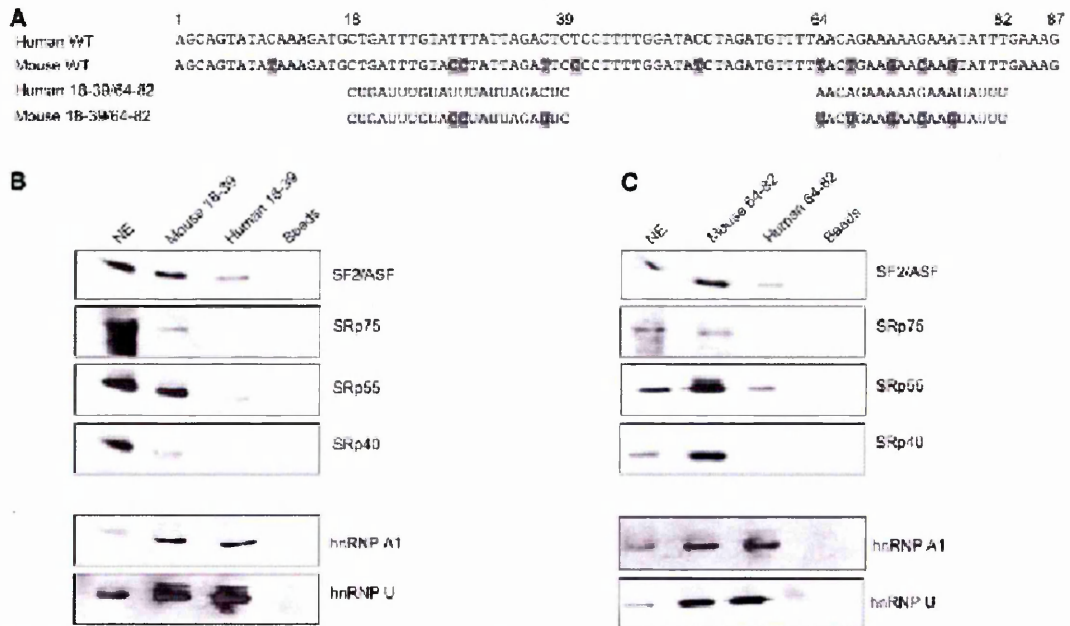

Figure 8. (A) Mouse and Human synthetic RNA sequences used in a affinity purification assay. Mouse specific nucleotides are boxed. (B) Pulldown analysis performed for human and mouse CFTR exon 12 sequences spanning from nucleotide 18–39 were comparatively analyzed for splicing factors. Harvested proteins were targeted for SR proteins (SRp 75, SRp 55, SRp 40, SF2/ASF) in western blot analysis. The membrane was later probed with hnRNP A1 and U antibody. (C). Similar comparative analysis for splicing factors was performed for both the human and mouse CFTR exon 12 sequences spanning nucleotides 64–82.

DISCUSSION

Missense mutations in human CFTR exon 12 have been described to be the causative agent of Cystic Fibrosis through the inactivation of a highly conserved region that encodes part of the first nucleotide binding fold of the protein (49). In particular, among all disease-causing mutations known to affect this exon, several have been described to induce its skipping during the splicing
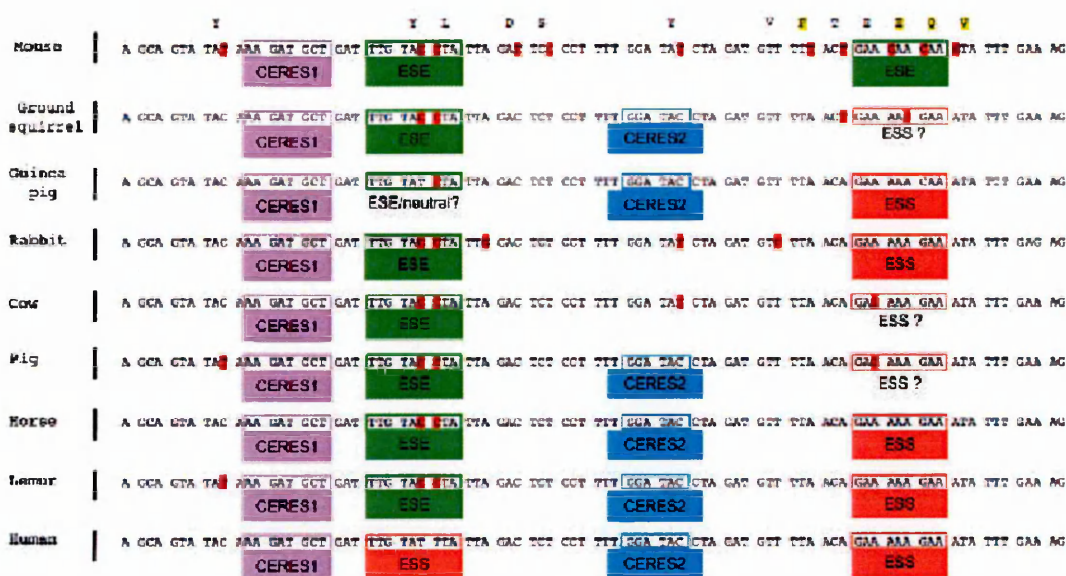
Figure 9. Schematic comparison and conservation of the splicing regulatory elements (ESEs, ESSs, and CERES) in different species based on our functional analysis.

process (50–52). In our lab, we originally defined within CFTR exon 12 two regions, named CERES1 and CERES2 (Figure 9), that functioned in a highly context-dependent manner to regulate the splicing process of this exon (44). In fact, the effect of natural and artificial mutations within these regions could not be predicted easily using current bioinformatics approaches (44), highlighting recent recommendations that these programs should be used with caution when they are used as a diagnostic tool (53,54). Indeed, successive experimental comparison between human and mouse CFTR exon 12 sequences demonstrated that about one quarter of all artificial combinations of mouse–human same sense substitutions resulted in exon skipping (40). Taken together, these findings suggested that the whole coding sequence of CFTR exon 12 is under strong selective pressures not only for functional reasons at the protein level, but also for the maintenance of proper exon recognition by the splicing machinery.

Up to now, however, the detailed molecular bases of CERES action were not known. We have performed an analysis of the trans-acting factors that bind the CERES2 element localized in CFTR exon 12. Our analysis was preliminarily focused at characterizing the binding properties of the most common splicing regulators, and in particular those belonging to either the SR or the hnRNP class of *trans*-acting factors. The results of our analysis have demonstrated that in normal conditions human CERES2 can bind a substantial number of these proteins, something that may be rather surprising since the core CERES2 element represents a very short stretch of RNA sequence (<10 nucleotides). Among SR proteins, we have found SF2/ASF and SRp55 whilst regarding hnRNP proteins specific binding could be detected for most hnRNP A/B family members. Most importantly, the

relative binding capacity of some of the factors was modified following the introduction of disease-associated missense mutations or of samesense substitutions that were already known to affect CFTR exon 12 inclusion levels. From a basic RNA binding protein point of view, this finding highlights the great flexibility provided by RRM motifs that can recognize a few specific bases at selected positions using their main chains and then use side-chain interactions to stabilize binding (55). This probably explains why the CERES2 sequence rather than functioning as a binding site for a single protein only can function as a kind of aggregation site for many SR/hnRNP factors. Evidence of a functional interaction was confirmed only for SF2/ASF, SRp55 and hnRNP A1. However, it should be noted that these experiments were performed in a severely reduced context (mini-exon 12) and that in a more natural setting many of these proteins will be able to play a role, especially considering the fact that the exonic flanking regions also can bind several SR/hnRNP factors (see below). At the moment, of course, we have tested only few regulatory proteins and hence cannot rule out the presence of additional yet unknown factors that might also contribute to define/hinder CFTR exon 12. Another possibility is that our substitutions may affect the RNA secondary structure of CFTR exon 12. However, an evolutionary-based model of CFTR exon 12 RNA secondary structure has already been previously published by Meyer and Miklos (56). In their work, they have also analyzed the potential impact of some splicing mutations including substitutions in the 40T and 52C positions. The conclusion is that there are only marginally significant changes in the RNA structure of the mutants with respect to the predicted wild-type sequence. It is for this reason that we decided to concentrate on analyzing *trans*-acting factors rather than taking

into consideration structural changes following our substitutions.

Taken together, our results suggest that this crowding together of many different proteins (both positive and negative in terms of their effect on splicing) may explain why single-point substitutions within the CERES2 element have such an unpredictable effect on the exon recognition levels. Up to now, on a slightly wider scale the CERES sequence situation is similar to what has already been found in several small exons, such as SMN exon 7, where *trans*-acting factors (SF2/ASF and hnRNP A1) binding to the same exonic region contribute to exon inclusion/skipping (57,58). Other examples of very complex systems include the human *c-src* exon N1 (59,60), *CD44* exon v5 (61–63), *HipK3* 'T' exon (64), and chicken *cTNT* exon 5 (65) where numerous factors have been shown to participate in splicing regulation in close spatial proximity to each other. Indeed, for SMN exon 7 it has been hypothesized the existence of an Extended Inhibitory Context (Exinct) that is caused by overlapping regulatory motifs not all of which have still been characterized in depth (66).

On a more general note, the existence of these numerous splicing factor binding sites co-existing together on the same exon has very important implications with regards to evolutionary constraints in codon composition. For example, our results have shown very clearly that the inhibitory effect of some synonymous nucleotide substitutions (A40T and C52T) naturally occurring in the mouse sequence can be explained by the different splicing regulatory properties of the human and mouse flanking exonic sequences (summarized in Figure 9). In particular, by comparing the ESE, ESS, and CERES elements several considerations can be made with regards to the sequence changes that have occurred in the mouse to human transition during the course of evolution. In fact, as shown in Figure 9 it can be hypothesized that the creation of the human CERES2 element in CFTR exon 12 has relieved the pressure to keep the two weak ESE elements loosely localized in the mouse 18–39 and 64–82 flanking regions. Only after the creation of CERES2 element these two regions could thus undergo nucleotide substitutions that either weakened these elements (i.e. in the Ground squirrel and Guinea pig 64–82 region) or even changed them to functional silencer sequences (in human 18–39 and 64–82 region). The advantage of these coding region changes can be only hypothesized at this stage, but it may involve further steps beyond mRNA processing such as enzymatic activity or protein stability. Irrespective of their significance, however, the important issue is that sequence changes could only be introduced through the creation of the CERES2 element (the importance of which is highlighted by the observation that many of the mouse to human substitutions analyzed in this study lead directly to total exon skipping). It is also interesting to note that a comparison of these sequence elements also in other species (Figure 9) does not contradict the conclusions we have drawn from mouse versus humans. In fact, for example, both cows and rabbits that do not contain the CERES2 element have absolutely conserved

ESE sequences. Of course, additional experiments will need to be performed before we can draw firm conclusions on this issue. Nonetheless, in our work, we show that an integrated analysis of *cis*- and *trans*-acting factors binding to exonic elements can provide a substantial wealth of information on potential evolutionary mechanisms.

Looking at the splicing regulatory elements in Figure 9 it is also possible to draw some additional conclusions with regards to our understanding of splicing regulation in general. It is clear that, the CFTR exon 12 sequence is literally covered by regulatory elements that we probably still consider (rather mistakenly) as separate elements. Proof of this is the observation that the activity of many of these splicing regulatory elements (especially CERES) cannot be exported in different contexts (44). Indeed, our results point towards a situation where in exons like CFTR exon 12 we should virtually consider every nucleotide as potentially capable of affecting splicing inclusion levels. The only question that might then remain to us is the direction of this change (whether increased inclusion/skipping) and its extent. It is very probable that as our knowledge of splicing systems increases this kind of situations will be more and more common. From a practical point of view, this will have several consequences. From a clinical point of view, increased importance have to be given at analyzing RNA transcripts directly from patient tissues or, routinely, through minigene based systems that will mimic this kind of global splicing regulatory networks (67). Second, this increased awareness will be useful for the development of novel bioinformatics methods aimed at predicting splicing outcomes that, until now, have been primarily focused at considering enhancer and silencer elements as well distinct entities with rather limited success (53,54). Finally, it should gradually shift our view of splicing where exon inclusion levels should not always be viewed as the straightforward algebraic sum of enhancer/silencer elements but as rather as an integrated unit, where silencing and enhancing functions may functionally overlap to a degree that has often been underestimated.

FUNDING

Conflict of interest statement. None declared.

REFERENCES

1. Pagani,F. and Baralle,F.E. (2004) Genomic variants in exons and introns: identifying the splicing spoilers. *Nat. Rev. Genet.*, 5, 389–396.
2. Parmley,J.L. and Hurst,L.D. (2007) How do synonymous mutations affect fitness? *Bioessays*, 29, 515–519.
3. Chamary,J.V., Parmley,J.L. and Hurst,L.D. (2006) Hearing silence: non-neutral evolution at synonymous sites in mammals. *Nat. Rev. Genet.*, 7, 98–108.

4. Black,D.L. (2003) Mechanisms of alternative pre-messenger RNA splicing. *Annu. Rev. Biochem.*, **72**, 291–336.
5. Mardon,H.J., Sebastio,G. and Baralle,F.E. (1987) A role for exon sequences in alternative splicing of the human fibronectin gene. *Nucleic Acids Res.*, **15**, 7725–7733.
6. Reed,R. and Maniatis,T. (1986) A role for exon sequences and splice-site proximity in splice-site selection. *Cell*, **46**, 681–690.
7. Breathnach,R., Benoist,C., O'Hare,K., Gannon,F. and Chambon,P. (1978) Ovalbumin gene: evidence for a leader sequence in mRNA and DNA sequences at the exon-intron boundaries. *Proc. Natl Acad. Sci. USA*, **75**, 4853–4857.
8. Ben-Dov,C., Hartmann,B., Lundgren,J. and Valcarcel,J. (2008) Genome-wide analysis of alternative pre-mRNA splicing. *J. Biol. Chem.*, **283**, 1229–1233.
9. Moore,M.J. and Silver,P.A. (2008) Global analysis of mRNA splicing. *RNA*, **14**, 197–203.
10. Buratti,E. and Baralle,F.E. (2005) Another step forward for SELEXive splicing. *Trends Mol. Med.*, **11**, 5–9.
11. Chasin,L.A. (2007) Searching for splicing motifs. *Adv. Exp. Med. Biol.*, **623**, 85–106.
12. Wang,J., Smith,P.J., Krainer,A.R. and Zhang,M.Q. (2005) Distribution of SR protein exonic splicing enhancer motifs in human protein-coding genes. *Nucleic Acids Res.*, **33**, 5053–5062.
13. Pozzoli,U. and Sironi,M. (2005) Silencers regulate both constitutive and alternative splicing events in mammals. *Cell. Mol. Life Sci.*, **62**, 1579–1604.
14. Cartegni,L., Chew,S.L. and Krainer,A.R. (2002) Listening to silence and understanding nonsense: exonic mutations that affect splicing. *Nat. Rev. Genet.*, **3**, 285–298.
15. Blencowe,B.J. (2000) Exonic splicing enhancers: mechanism of action, diversity and role in human genetic diseases. *Trends Biochem. Sci.*, **25**, 106–110.
16. Venables,J.P. (2007) Downstream intronic splicing enhancers. *FEBS Lett.*, **581**, 4127–4131.
17. Wang,Z., Rolish,M.E., Yeo,G., Tung,V., Mawson,M. and Burge,C.B. (2004) Systematic identification and analysis of exonic splicing silencers. *Cell.*, **119**, 831–845.
18. Wang,Z. and Burge,C.B. (2008) Splicing regulation: from a parts list of regulatory elements to an integrated splicing code. *RNA*, **14**, 802–813.
19. Lin,S. and Fu,X.D. (2007) SR proteins and related factors in alternative splicing. *Adv. Exp. Med. Biol.*, **623**, 107–122.
20. Huang,Y. and Steitz,J.A. (2005) SRprises along a messenger's journey. *Mol. Cell.*, **17**, 613–615.
21. Long,J.C. and Caceres,J.F. (2009) The SR protein family of splicing factors: master regulators of gene expression. *Biochem. J.*, **417**, 15–27.
22. Martinez-Contreras,R., Cloutier,P., Shkreta,L., Fisette,J.F., Revil,T. and Chabot,B. (2007) hnRNP proteins and splicing control. *Adv. Exp. Med. Biol.*, **623**, 123–147.
23. Venables,J.P., Koh,C.S., Froehlich,U., Lapointe,E., Couture,S., Inkel,L., Bramard,A., Paquet,E.R., Watier,V., Durand,M. *et al.* (2008) Multiple and specific mRNA processing targets for the major human hnRNP proteins. *Mol. Cell. Biol.*, **28**, 6033–6043.
24. Jurica,M.S. and Moore,M.J. (2003) Pre-mRNA splicing: awash in a sea of proteins. *Mol. Cell.*, **12**, 5–14.
25. Zhu,J., Mayeda,A. and Krainer,A.R. (2001) Exon identity established through differential antagonism between exonic splicing silencer-bound hnRNP A1 and enhancer-bound SR proteins. *Mol. Cell.*, **8**, 1351–1361.
26. Eperon,I.C., Makarova,O.V., Mayeda,A., Munroe,S.H., Caceres,J.F., Hayward,D.G. and Krainer,A.R. (2000) Selection of alternative 5′ splice sites: role of U1 snRNP and models for the antagonistic effects of SF2/ASF and hnRNP A1. *Mol. Cell. Biol.*, **20**, 8303–8318.
27. Caceres,J.F., Stamm,S., Helfman,D.M. and Krainer,A.R. (1994) Regulation of alternative splicing in vivo by overexpression of antagonistic splicing factors. *Science*, **265**, 1706–1709.
28. Relogio,A., Ben-Dov,C., Baum,M., Ruggiu,M., Gemund,C., Benes,V., Darnell,R.B. and Valcarcel,J. (2005) Alternative splicing microarrays reveal functional expression of neuron-specific regulators in Hodgkin lymphoma cells. *J. Biol. Chem.*, **280**, 4779–4784.

29. Grosso,A.R., Gomes,A.Q., Barbosa-Morais,N.L., Caldeira,S., Thorne,N.P., Grech,G., von Lindern,M. and Carmo-Fonseca,M. (2008) Tissue-specific splicing factor gene expression signatures. *Nucleic Acids Res.*, **36**, 4823–4832.
30. Shin,C., Feng,Y. and Manley,J.L. (2004) Dephosphorylated SRp38 acts as a splicing repressor in response to heat shock. *Nature*, **427**, 553–558.
31. Blaustein,M., Pelisch,F. and Srebrow,A. (2007) Signals, pathways and splicing regulation. *Int. J. Biochem. Cell. Biol.*, **39**, 2031–2048.
32. Huang,Y., Yario,T.A. and Steitz,J.A. (2004) A molecular link between SR protein dephosphorylation and mRNA export. *Proc. Natl Acad. Sci. USA*, **101**, 9666–9670.
33. Misteli,T., Caceres,J.F., Clement,J.Q., Krainer,A.R., Wilkinson,M.F. and Spector,D.L. (1998) Serine phosphorylation of SR proteins is required for their recruitment to sites of transcription in vivo. *J. Cell. Biol.*, **143**, 297–307.
34. Buratti,E., Dhir,A., Lewandowska,M.A. and Baralle,F.E. (2007) RNA structure is a key regulatory element in pathological ATM and CFTR pseudoexon inclusion events. *Nucleic Acids Res.*, **35**, 4369–4383.
35. Buratti,E., Muro,A.F., Giombi,M., Gherbassi,D., Iaconcig,A. and Baralle,F.E. (2004) RNA folding affects the recruitment of SR proteins by mouse and human polypurinic enhancer elements in the fibronectin EDA exon. *Mol. Cell. Biol.*, **24**, 1387–1400.
36. Hiller,M., Zhang,Z., Backofen,R. and Stamm,S. (2007) Pre-mRNA secondary structures influence exon recognition. *PLoS Genet.*, **3**, e204.
37. Shepard,P.J. and Hertel,K.J. (2008) Conserved RNA secondary structures promote alternative splicing. *RNA*, **14**, 1463–1469.
38. Cheah,M.T., Wachter,A., Sudarsan,N. and Breaker,R.R. (2007) Control of alternative RNA splicing and gene expression by eukaryotic riboswitches. *Nature*, **447**, 497–500.
39. Graveley,B.R. (2005) Mutually exclusive splicing of the insect Dscam pre-mRNA directed by competing intronic RNA secondary structures. *Cell.*, **123**, 65–73.
40. Pagani,F., Raponi,M. and Baralle,F.E. (2005) Synonymous mutations in CFTR exon 12 affect splicing and are not neutral in evolution. *Proc. Natl Acad. Sci. USA*, **102**, 6368–6372.
41. Parmley,J.L., Chamary,J.V. and Hurst,L.D. (2006) Evidence for purifying selection against synonymous mutations in mammalian exonic splicing enhancers. *Mol. Biol. Evol.*, **23**, 301–309.
42. Parmley,J.L. and Hurst,L.D. (2007) Exonic splicing regulatory elements skew synonymous codon usage near intron-exon boundaries in mammals. *Mol. Biol. Evol.*, **24**, 1600–1603.
43. Irimia,M., Rukov,J.L. and Roy,S.W. (2009) Evolution of alternative splicing regulation: changes in predicted exonic splicing regulators are not associated with changes in alternative splicing levels in primates. *PLoS ONE*, **4**, e5800.
44. Pagani,F., Stuani,C., Tzetis,M., Kanavakis,E., Efthymiadou,A., Doudounakis,S., Casals,T. and Baralle,F.E. (2003) New type of disease causing mutations: the example of the composite exonic regulatory elements of splicing in CFTR exon 12. *Hum. Mol. Genet.*, **12**, 1111–1120.
45. Pagani,F., Buratti,E., Stuani,C. and Baralle,F.E. (2003) Missense, nonsense, and neutral mutations define juxtaposed regulatory elements of splicing in cystic fibrosis transmembrane regulator exon 9. *J. Biol. Chem.*, **278**, 26580–26588.
46. Buratti,E., Baralle,M., De Conti,L., Baralle,D., Romano,M., Ayala,Y.M. and Baralle,F.E. (2004) hnRNP H binding at the 5′ splice site correlates with the pathological effect of two intronic mutations in the NF-1 and TSHbeta genes. *Nucleic Acids Res.*, **32**, 4224–4236.
47. Buratti,E., Dork,T., Zuccato,E., Pagani,F., Romano,M. and Baralle,F.E. (2001) Nuclear factor TDP-43 and SR proteins promote in vitro and in vivo CFTR exon 9 skipping. *EMBO J.*, **20**, 1774–1784.
48. Buratti,E. and Baralle,F.E. (2001) Characterization and functional implications of the RNA binding properties of nuclear factor TDP-43, a novel splicing regulator of CFTR exon 9. *J. Biol. Chem.*, **276**, 36337–36343.
49. Delaney,S.J., Rich,D.P., Thomson,S.A., Hargrave,M.R., Lovelock,P.K., Welsh,M.J. and Wainwright,B.J. (1993) Cystic fibrosis transmembrane conductance regulator splice variants are

not conserved and fail to produce chloride channels. *Nat. Genet.*, **4**, 426–431.

50. Hull,J., Shackleton,S. and Harris,A. (1993) Abnormal mRNA splicing resulting from three different mutations in the CFTR gene. *Hum. Mol. Genet.*, **2**, 689–692.

51. Zielenski,J., Markiewicz,D., Lin,S.P., Huang,F.Y., Yang-Feng,T.L. and Tsui,L.C. (1995) Skipping of exon 12 as a consequence of a point mutation (1898 + 5G→T) in the cystic fibrosis transmembrane conductance regulator gene found in a consanguineous Chinese family. *Clin. Genet.*, **47**, 125–132.

52. Strong,T.V., Smit,L.S., Nasr,S., Wood,D.L., Cole,J.L., Iannuzzi,M.C., Stern,R.C. and Collins,F.S. (1992) Characterization of an intron 12 splice donor mutation in the cystic fibrosis transmembrane conductance regulator (CFTR) gene. *Hum. Mutat.*, **1**, 380–387.

53. Hartmann,L., Theiss,S., Niederacher,D. and Schaal,H. (2008) Diagnostics of pathogenic splicing mutations: does bioinformatics cover all bases? *Front Biosci.*, **13**, 3252–3272.

54. Houdayer,C., Dehainault,C., Mattler,C., Michaux,D., Caux-Moncoutier,V., Pages-Berhouet,S., d'Enghien,C.D., Lauge,A., Castera,L., Gauthier-Villars,M. *et al.* (2008) Evaluation of in silico splice tools for decision-making in molecular diagnosis. *Hum. Mutat.*, **29**, 975–982.

55. Auweter,S.D., Oberstrass,F.C. and Allain,F.H. (2006) Sequence-specific binding of single-stranded RNA: is there a code for recognition? *Nucleic Acids Res.*, **34**, 4943–4959.

56. Meyer,I.M. and Miklos,I. (2005) Statistical evidence for conserved, local secondary structure in the coding regions of eukaryotic mRNAs and pre-mRNAs. *Nucleic Acids Res.*, **33**, 6338–6348.

57. Cartegni,L., Hastings,M.L., Calarco,J.A., de Stanchina,E. and Krainer,A.R. (2006) Determinants of exon 7 splicing in the spinal muscular atrophy genes, SMN1 and SMN2. *Am. J. Hum. Genet.*, **78**, 63–77.

58. Kashima,T., Rao,N., David,C.J. and Manley,J.L. (2007) hnRNP A1 functions with specificity in repression of SMN2 exon 7 splicing. *Hum. Mol. Genet.*, **16**, 3149–3159.

59. Sharma,S., Falick,A.M. and Black,D.L. (2005) Polypyrimidine tract binding protein blocks the 5′ splice site-dependent assembly of U2AF and the prespliceosomal E complex. *Mol. Cell.*, **19**, 485–496.

60. Chou,M.Y., Underwood,J.G., Nikolic,J., Luu,M.H. and Black,D.L. (2000) Multisite RNA binding and release of polypyrimidine tract binding protein during the regulation of c-src neural-specific splicing. *Mol. Cell.*, **5**, 949–957.

61. Matter,N., Herrlich,P. and Konig,H. (2002) Signal-dependent regulation of splicing via phosphorylation of Sam68. *Nature*, **420**, 691–695.

62. Matter,N., Marx,M., Weg-Remers,S., Ponta,H., Herrlich,P. and Konig,H. (2000) Heterogeneous ribonucleoprotein A1 is part of an exon-specific splice-silencing complex controlled by oncogenic signaling pathways. *J. Biol. Chem.*, **275**, 35353–35360.

63. Cheng,C. and Sharp,P.A. (2006) Regulation of CD44 alternative splicing by SRm160 and its potential role in tumor cell invasion. *Mol. Cell. Biol.*, **26**, 362–370.

64. Venables,J.P., Bourgeois,C.F., Dalgliesh,C., Kister,L., Stevenin,J. and Elliott,D.J. (2005) Up-regulation of the ubiquitous alternative splicing factor Tra2beta causes inclusion of a germ cell-specific exon. *Hum. Mol. Genet.*, **14**, 2289–2303.

65. Ladd,A.N., Stenberg,M.G., Swanson,M.S. and Cooper,T.A. (2005) Dynamic balance between activation and repression regulates pre-mRNA alternative splicing during heart development. *Dev. Dyn.*, **233**, 783–793.

66. Singh,N.N., Androphy,E.J. and Singh,R.N. (2004) An extended inhibitory context causes skipping of exon 7 of SMN2 in spinal muscular atrophy. *Biochem. Biophys. Res. Commun.*, **315**, 381–388.

67. Baralle,D., Lucassen,A. and Buratti,E. (2009) Missed threads. The impact of pre-mRNA splicing defects on clinical practice. *EMBO Rep.*, **10**, 810–816.