



# **ISAS - INTERNATIONAL SCHOOL FOR ADVANCED STUDIES**

ISAS-INTERNATIONAL SCHOOL FOR ADVANCED STUDIES

*Thesis submitted for the degree of Doctor Philosophiae*

## ***Cis-Elements and Trans-Acting Factors Involved in CFTR Exon 9 Alternative Splicing***

**Candidate: Zuccato Elisabetta**

**Supervisor: Prof. Francisco Ernesto Baralle**

Academic Year 2001/2002

**SISSA - SCUOLA  
INTERNAZIONALE  
SUPERIORE  
STUDI AVANZATI**

TRIESTE  
Via Beirut 2-4

**TRIESTE**

*To my parents*

## Acknowledgments

This work was carried out in the Molecular Pathology group at the International Centre for Genetic Engineering and Biotechnology in Trieste, under the supervision of Prof. F. E. Baralle.

I am deeply indebted to Prof. Baralle for accepting me as a Ph.D student in his laboratory and for his excellent guidance and support throughout the project.

A special thanks goes to Dr. Franco Pagani for giving me the opportunity to work in this particular project and for his collaboration and discussion during this period.

I am grateful to Dr. Emanuele Buratti for his collaboration and help regarding the *in vitro* experiment part of this work. A very special thanks goes to Ms. Cristiana Stuani for her friendship, for the important moral support in all the good and bad moments of my life. I also thank her for the teaching and help in the technical part of my work.

I am grateful to all my colleagues for their friendship, support and co-operation during these years, it has been really a pleasure to work daily with them. In particular I want to thank Roberto Marcucci, Fabiola Porro, Maurizio Romano, Andres Muro and Caterina Marchetti.

Last, but not the least, thanks to my parents for their constant moral support, and to all my friends.

## TABLE OF CONTENTS

<b>Abstract</b>	p.10
<b>Introduction</b>	p.12
<b>1 Cystic Fibrosis</b>	p.13
1.1 <i>Clinical features of cystic fibrosis</i>	p.14
1.2 <i>Cystic fibrosis gene and CFTR protein</i>	p.16
1.2.1 The CFTR protein: structure and regulation	p.17
1.2.2 Expression of CFTR protein	p.20
1.2.3 CFTR function	p.20
1.3 <i>Molecular mechanisms of CFTR channel dysfunction</i>	p.21
1.3.1 Genotype-phenotype correlation	p.23
<b>2 RNA splicing</b>	p.26
2.1 <i>The spliceosome</i>	p.26
2.1.1 The splicing reaction	p.27
2.1.2 Sequence elements (cis) on the pre-mRNA required for splicing	p.27
2.2 <i>Protein components involved in splicing</i>	p.29
2.2.1 Small Nuclear Ribonucleoprotein Particles (snRNPs)	p.30
2.2.2 Non-snRNP splicing factors	p.31
2.3 <i>The spliceosomal complex</i>	p.33
2.3.1 General model for spliceosome assembly and spliceosome cycle	p.33
2.3.2 Early complex formation: intron and exon definition models	p.36
2.3.4 Splice site recognition and bridging interactions	p.40
2.4 <i>Alternative splicing</i>	p.42
2.4.1 Involvement of cis-acting elements in splice site selection	p.44

2.4.2 Trans-acting factors	p.49
2.5 <i>Regulation of the expression of the splicing factors and regulation of alternative splicing</i>	p.52
<b>3 Alternative splicing of CFTR transcripts</b>	p.53
3.1 <i>Alternative splicing of CFTR pre-mRNA and monosymptomatic forms of cystic fibrosis</i>	p.54
3.2 <i>Association studies of alternative splicing of CFTR exon 9</i>	p.55
3.3 <i>Cis-acting elements involved in the regulation of CFTR exon 9 alternative splicing</i>	p.58
3.4 <i>Trans-acting factors involved in the regulation of CFTR exon 9 alternative splicing</i>	p.58
<b>Results</b>	p.61
<b>Part 1: Functional effect of TDP43 on alternative splicing of CFTR exon 9</b>	p.62
4.1 <i>Overexpression of TDP43 in transfection assays</i>	p.62
4.1.2 <i>TDP 43 functional effect is strictly dependent on the presence of TG repeats</i>	p.63
4.1.3 <i>Dose-response effect of TDP43</i>	p.65
4.2 <i>Transient transfections of antisense oligonucleotides directed against TDP43 mRNA</i>	p.65
4.2.1 <i>Transfection of antisense PS-oligo induces CFTR exon 9 inclusion</i>	p.67
4.3 <i>Distribution of TDP43 and SF2/ASF in human tissue</i>	p.70
4.4 <i>In vivo significance of the extreme TG13T3 polymorphism</i>	p.72

<b>RESULT Part 2: CFTR intron 9 contains other splicing regulatory elements than the ISS</b>	p.76
<i>5.1 Intron 9 polypyrimidine rich elements between the 5' splice site and the ISS modulate splicing efficiency</i>	p.77
<i>5.2 In vitro Binding Experiments Shown the Ability of PTB to Interact to CFTR Intron 9</i>	p.81
5.2.1 Study of the RNA-Protein Interaction in CFTR Sequence	p.81
5.2.2 Immunoprecipitation assays show that CFTR intron 9 binds PTB	p.85
5.2.3 Recombinant PTB Interacts with CFTR Intron 9 Sequence in UV Cross-Linking Experiments	p.88
<i>5.3 Study of the functional effect of PTB on CFTR exon 9 alternative splicing</i>	p.88
5.3.1 The splicing factor PTB induces CFTR exon 9 skipping	p.88
5.3.2 The effect of PTB overexpression on the minigenes with the mutated polypyrimidine elements	p.90
5.3.3 Hybrid human/mouse minigene confirmed that the modulation of exon 9 inclusion by PTB required other elements than PY1, PY2 and PY3	p.92
<b>Part 3: CFTR exon 9 missense and nonsense substitutions can affect the splicing efficiency</b>	p.96
<i>6.1 The effect of exon 9 missense and nonsense mutations on the alternative splicing</i>	p.96
<i>6.2 Mapping of cis-acting regulatory sequences in CFTR exon 9</i>	p.100
<i>6.3 hnRNP H binds to 155G mutant</i>	p.103
<b>Discussion</b>	p.107

<b>Part 1: TDP43 binding to TG repeats at polymorphic variant at 3'end of intron 8 negatively affect exon 9 alternative splicing</b>	p.108
<i>7.1 Mechanism of TDP43-mediated CFTR exon 9 splicing inhibition</i>	p.112
<b>Part 2: PTB could be an indirect trans-acting factor involved in CFTR exon 9 alternative splicing</b>	p.115
<i>8.1 Functional Effect of PTB on Human Exon 9 Alternative Splicing</i>	p.118
<i>8.2 The PY2 element</i>	p.120
<b>Part 3: The important role of missense and nonsense mutation in CFTR exon 9 at the pre-mRNA level</b>	p.123
<i>9.1 Missense and nonsense mutations affect alternative splicing of CFTR exon 9</i>	p.123
<i>9.2 The ability of 155G mutant to bind hnRNP H</i>	p.126
<i>9.3 Clinical implications</i>	p.127
<b>MATERIALS AND METHODS</b>	p.129
<b>10.1 MATERIALS</b>	p.130
10.1.1 Chemical reagents	p.130
10.1.2 Enzymes	p.130
10.1.3 Synthetic oligonucleotides	p.130
10.1.4 Radioactive isotopes	p.131
10.1.5 Bacterial culture	p.131
10.1.6 Cell culture	p.131
<b>10.2 METHODS</b>	p.132
<i>10.2.1 Nucleic Acids Preparations</i>	p.132
10.2.1.1 Small scale preparation of plasmid DNA from bacterial cultures	p.132
10.2.1.2 Large scale preparations of plasmid DNA from bacterial cultures	p.132

10.2.1.3 Preparation of the RNA from the cultured cells	p.132
10.2.2 <i>Estimation Of Nucleic Acid Concentration</i>	p.133
10.2.2.1 Spectrophotometric	p.133
10.2.3 <i>Enzymatic Modification of DNA</i>	p.133
10.2.3.1. Restriction enzymes	p.133
10.2.3.2. Large fragment of E. Coli Polymerase I	p.134
10.2.3.3 Dephosphorylation of DNA 5' termini	p.134
10.2.3.4 T4 DNA ligase	p.135
10.2.4 <i>Elution and purification of DNA fragments from agarose gels</i>	p.135
10.2.5 <i>PCR reaction for a cloning purpose</i>	p.136
10.2.6 <i>Sequence analysis</i>	p.136
10.2.7 <i>Construction of the minigene system and site directed mutagenesis</i>	p.137
10.2.8 <i>Construction of the minigene system mutated in the polypyrimidine rich elements in intron 9</i>	p.138
10.2.9 <i>Site directed mutagenesis in exon 9</i>	p.140
10.2.10 <i>Preparation of TDP43 expression vector</i>	p.141
10.2.11 <i>Maintenance and analysis of cells in culture</i>	p.141
10.2.12 <i>Transfection of recombinant DNA into cells maintains in culture</i>	p.141
10.2.13 <i>Measurement of the transfection efficiency</i>	p.143
10.2.14 <i>mRNA Analysis by Polymerase Chain Reaction</i>	p.143
10.2.14.1 cDNA synthesis	p.143
10.2.14.2 PCR analysis	p.144
10.2.15 <i>Analysis of the PCR-Amplified Products</i>	p.144
10.2.15.1 Not radiolabelled PCR products	p.144
10.2.15.2 Radiolabelled PCR products	p.145
10.2.16 <i>Radioactive labeling of the RNA</i>	p.145
10.2.17 <i>UV crosslinking</i>	p.145
10.2.18 <i>Preparations of competitors</i>	p.146



10.2.19. <i>Cross-linking of RNA to adipic dehydrazide agarose beads for affinity purification of cellular factors</i>	p.147
10.2.20 <i>Expression of recombinant TDP-43 as GST fusion proteins</i>	p.148
10.2.21 <i>Immunoprecipitation assays with PTB and hnRNP H</i>	p.148
10.2.22 <i>Northern Blot of TDP-43 and SF2/ASF</i>	p.149
10.2.23 <i>Western Blot assay on Hep 3B protein extract</i>	p.149
10.2.23 <i>Characterisation of the TG13T3 mutation</i>	p.150
<b>References</b>	p.151

## ABSTRACT

Alternative splicing of human cystic fibrosis transmembrane conductance regulator (CFTR) exon 9 is regulated by a combination of *cis*-acting elements distributed through the exon and both flanking introns. Several studies have identified at 3' end of intron 8 a regulatory element that is composed by a polymorphic (TG)<sub>m</sub>(T)<sub>n</sub> repeated sequence. Variations at this polymorphic locus are associated with the alternative splicing of exon 9, which results in a non-functional CFTR protein. In our previous studies, we identified TDP43 as the factor that specifically binds to the (TG)<sub>m</sub> sequence. In this thesis we demonstrate, using transient transfection experiment with a minigene system, that overexpression of TDP43 results in an increase of exon 9 skipping. This effect is more pronounced with concomitant overexpression of SF2/ASF SR protein previously shown to inhibit CFTR exon 9 inclusion. Moreover, antisense inhibition of endogenous TDP43 expression results in increased inclusion of exon 9. The clinical and biological relevance of this finding *in vivo* is demonstrated by our characterization of a CF patient carrying a (TG)<sub>10</sub>T<sub>9</sub>(ΔF508)/(TG)<sub>13</sub>(T)<sub>3</sub> genotype leading to a disease-causing high proportion of exon skipping in the allele with non mutated coding sequence.

We have also previously shown the presence of a splicing inhibitory element (ISS) in the intron 9. The binding of SR proteins to the ISS modulates, together with other *trans*-acting factors, the level of exon 9 inclusion. In this thesis we studied the region of intron 9 from the 5' splice site until the ISS. This region presents a peculiar arrangement of polypyrimidine-rich elements and we demonstrated that one of these (PY2) acts as a splicing enhancer. In fact, mutations introduced in this PY2 element cause a decrease in exon 9 inclusion. While the factor that binds to this pyrimidine-rich enhancer element is still unknown, we

provide evidence for an involvement of polypyrimidine tract binding protein (PTB) on the modulation of exon 9 inclusion. The functional effect of PTB that we report is linked to the binding of the protein to elements in CFTR intron 9.

Exonic sequences are also involved in the modulation of the alternative splicing of exon 9. In fact, for the first time, we studied the effect of several natural and site-directed mutants distributed on the entire exon and we demonstrate that in some instances this mutations induce significant changes in pre-mRNA splicing, with different degree of exon skipping and inclusion. The effect of this mutations is also modulate by the length of the polymorphic (TG)<sub>m</sub>(T)<sub>n</sub> tract at 3' end of intron 9. We also identified a critical regulatory element which contains two adjacent sequences with enhancer and silencer activity. We defined this element having overlapping enhancer and silencer properties Composit Exonic Regulatory Element of Splicing (CERES). The analysis of several site directed mutants spanning the region between nucleotides 144 and 157 showed that the functional characterization of the element cannot be simply defined by the mutated position but also by the type of nucleotide substitution.

All together the results reported in this thesis shed light into the regulation of the CFTR exon 9 alternative splicing. The new cis-acting elements and trans-acting factors that we have identified point to an unexpected complexity of this system that needs further investigation. In addition, it is certain that this is not a unique situation and that exonic and intronic splicing elements widespread in all genes and currently accepted splicing models will have to be correct.

## INTRODUCTION

## 1. CYSTIC FIBROSIS

Cystic fibrosis (CF) is an autosomal recessive disorder that affects children, many of whom now live into adulthood (1). The frequency of the disease varies among ethnic groups: is highest in the Caucasian population, with 1 affected in 2500 live birth and a heterozygote frequencies reaches the value of about 1 in 25 individuals. CF is less common in other ethnic groups, but significant numbers of affected individuals are found in Southern Europe, in the Ashkenazi Jewish population and in the American blacks (2).

CF was first described by Anderson in 1938 who introduced the term "cystic fibrosis of the pancreas" to describe the destruction of pancreatic exocrine function as a result of the disease. Later in 1953, DiSant'Agnes first described the elevated electrolyte levels in CF patients' sweat, which became the hallmark for CF diagnosis (2). The clinical features of CF are dominated by the involvement of respiratory tract with chronic obstruction of the airways by sticky and thick mucus and subsequent bacterial infection, especially with *Pseudomonas* species. In most patients the obstruction of the pancreatic ducts and the subsequent destruction of exocrine function leads to pancreatic insufficiency in 85% of cases [reviewed in (1)].

Since late 80's the pathophysiology of CF was not clear. The principal ionic mechanism underlying the pathogenesis of cystic fibrosis is considered the loss of epithelial plasmamembrane chloride conductance. This conductance pathway appeared to be unique to epithelial cells in which salt and water transport rates are governed by cyclic-AMP and calcium regulatory processes (3). 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.

## 1.1 Clinical features of classic cystic fibrosis

Cystic fibrosis is characterised by a wide variety of clinical presentations and organs involvement. Usual presentations include persistent cough, recurrent refractory lung infiltrates and also gastrointestinal disturbance, including meconium ileus. Moreover, a number of individuals escape detection in the first decade or two of life, often because symptoms are unusual, subtle, or even absent. [Reviewed in (1)].

### • Respiratory Tract

Mucous obstruction and infection are the major pathologic events in the lung, but are confined, at least initially, to the conducting airways. The ion transport abnormalities in the respiratory-conducting epithelia lead to dehydration of mucous, inspissations of secretion and diminished clearance of inhaled pathogens. Another event that can contribute to the clinical manifestation is directly connected to an abnormal macromolecular composition of mucus in the CF airways, affecting both its rheologic properties and affinity for microorganism. The defective salt-sensitive antimicrobial activity in the airway surface fluid leads to the chronic respiratory infections with pathogens such as *P. aureoginosa* and *S. aureus*.

### • Gastrointestinal Tract

Changes in the gastrointestinal tract itself are not prominent. They are caused by defective production of the secret that result in forming dehydrated, sticky and incompletely digested intestinal contents. Several type of cells are affected by CF disease: Brunner's glands are hypertrophied, simple goblet-cells present hyperplasia, mucus glands, tubular glands and Lieberkun's crypts are all involved. In 10-20% of

newborn patients with CF meconium ileus occurs (4). Its pathogenesis has been ascribed to failure of pancreatic enzyme secretion and, thus, of digestion of intraluminal contents in utero.

#### • **Pancreatic Disease**

There are two groups of mutation affecting pancreatic function, one termed "severe" and the other "mild". Patients with both CF chromosomes carrying a "severe" mutation are expected to have a pancreatic insufficient (PI) phenotype (5, 6). In this situation pancreas is abnormal with areas of destruction and replacement of the exocrine tissue by fibrous tissue and fat. Pancreatic enzyme deficiency causes fat and protein maldigestion, this results in failure to gain weight and, ultimately, in a failure of linear growth. The 15% of patients that carry "mild" mutation in one or both CF chromosomes are pancreatic sufficient (PS), so they have enough pancreatic exocrine function for food digestion; however, the level of residual pancreatic enzyme activity varies among patients (7).

#### • **Hepatobiliary Disease**

Liver changes are variable. In the severe hepatic involvement (5 % of the CF patients) bile canaliculi are plugged by mucus and may develop biliary cirrhosis and hepatosplenomegaly.

#### • **Genitourinary Tract**

More than 95% of male with classical CF have altered Wolffian duct structures. The vas deferens, tail and body of the epididymis and seminal vesicles are atrophic, fibrotic or completely absent. The pathogenesis of these structural changes probably relates to early, often intrauterine, obstruction of the genital tract with inspissated secretion. Several patients may present only CBAVD, which is

considered a typical monosymptomatic form of CF.

Fertility in CF affected women is most likely greater than 10%. The infertility is due to the mucus plugged in the cervical region.

- **Sweat Glands**

The most consistent functional alteration in CF has been elevated concentrations of chloride, sodium and potassium in exocrine sweat. This is due to a failure to reabsorb chloride along the sweat gland duct and not to a structural abnormality of exocrine sweat glands. In contrast, apocrine sweat glands in children with CF are more dilated and filled with retained secretions than glands of control subjects.

## **1.2 Cystic Fibrosis Gene and CFTR Protein**

The CFTR gene was the first gene cloned by “positional cloning”. Many different molecular cloning techniques and strategies were employed in attempts to isolate the gene responsible for CF. Extensive linkage analysis provided evidence for the existence of a single CF locus on human chromosome 7 (region q31) (8) and in 1989, the gene responsible for CF was identified (6, 9). A 280-kb DNA region was isolated through “chromosome walking” and “jumping”. Small DNA segments from this region were tested for the presence of genes. Among several candidates, one was proven, largely on the basis of the discovery of a common mutation, to be the CF gene.

The encoded mRNA is about 6.5 kb-long (Figure 1.1), is constituted by 27 exons and can be detected by northern blotting in a variety of affected tissue including lung pancreas, sweat glands, liver nasal polyps, salivary glands and colon (6). Moreover the expression of CF gene has been demonstrated in non-epithelial origin cells such as leucocytes (10). CF gene codes for a membrane-associated glycoprotein named CFTR



(Cystic Fibrosis Conductance Regulator), which is constituted by 1480 aminoacid residues and has a molecular weight of 169 KDa (11).

### **1.2.1 The CFTR protein structure and regulation**

CFTR protein consists of five domains (Figure 1.1): two nucleotide binding folds that bind ATP (NBD1 and NBD2); two hydrophobic transmembrane domains (TPY1 and TPY2), each constituted by six membrane-spanning segments; the regulatory (R) domain, which contains several consensus phosphorylation sequences (6).

#### **a) Transmembrane domain**

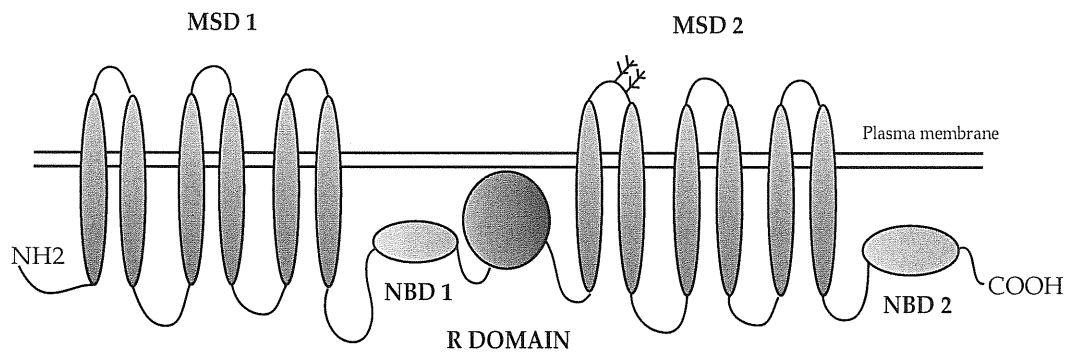
CFTR is anchored in the plasma membrane by the 12 hydrophobic segments (TPY1-12), organized in two transmembrane domains. The 12 membrane-spanning segments form an anion selective pore constituting the CFTR chloride channel. Different studies imply that CFTR is a multi-ion pore channel (12) and the TPY1, TPY2, TM6 and TPY12 are important in pore formation (13, 14). On the basis of aminoacid substitution found in many diseases causing mutation, several arginine residues appear particularly important for ion conductance. The current protein model predicts six extracellular loops and four cytoplasmic loops.

#### **b) Nucleotide binding fold**

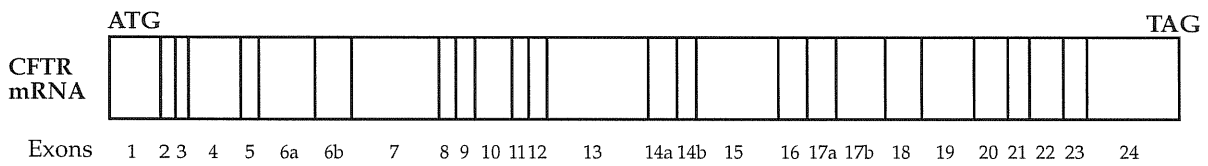
The two nucleotide binding folds contain sequence motifs Walker A and Walker B, that are conserved among ATP binding proteins (6).

The gene sequence encoding NBD1 spans exon 9 through 12 and that encoding NBD2 spans exon 19 through 23. The ATP binding and hydrolysis by the two NBDs determines the channel gating (15). The overall similarity between the primary structures of the two NBDs is only 29% at the aminoacid level. The activity of the channel is regulated through the level of phosphorylation of the R-domain and NBDs.

a)



b)



**Figure 1.1. The cystic fibrosis transmembrane conductance regulator (CFTR)**  
a) The model of CFTR protein and its relationship with the apical membrane of secretory epithelial cells. b) The structure of CFTR mRNA with 27 exons

Functional studies show that NBD1 and NBD2 do not contribute equally to the channel regulation. It is proposed that partial phosphorylation of the R-domain leads to the phosphorylation of the NBD1 and channel opening, whereas complete phosphorylation of R-domain leads to the phosphorylation of NBD2 causing channel closing (16, 17).

### **c) The R-domain**

The R-domain is a unique feature for CFTR. This region is localised on the cytoplasmatic side of the apical membrane (6). The R-domain contains several phosphorylation sites. Their phosphorylation is required for channel activity. Has been shown that phosphorylation of different sites by PKA may have either stimulatory or inhibitory effect on the CFTR channel activity (18, 19). Although cAMP activated phosphorylation of the R-domain is necessary for channel activity, it is not sufficient for the channel opening. Presumably, phosphorylation of the R-domain by protein kinase A changes its conformation and regulates the gating of the CFTR chloride channel (20, 21). The conformational alteration may provide the ATP hydrolytic ability of the NBDs that, in turn, control the channel opening and closing.

Control of CFTR chloride channel activity depends on the balance of kinases and phosphatases activity in the cell. Open CFTR channels conduct chloride anion down an electrochemical gradient from the cell to the lumen. This provides no obvious need for an external energy input, while ATP hydrolysis is actually required in the opening of the channel (22). It has been proposed a model in which ATP binds to NBD1 and the hydrolysis mediates the transition of the channel from an inactive to an active closed conformation. Binding of ATP at NBD2 leads to channel opening and the consequent hydrolysis of this ATP molecule induces a conformational change of the channel. At the end, dissociation of inorganic phosphate (Pi) or either ADP leads to channel

closure (17). More recent studies have suggested that ATP binding without hydrolysis may be sufficient to open the channel and that divalent cations increase ATP binding (23). Besides, structural studies by Karpowich *et al.* proposed that an ATP-induced alteration in the affinity of the two NBDs controls their interaction, thereby converting chemical into mechanical energy contributing to the activity of the transporter (24).

Thus in turn, the molecular mechanisms by which the ATP- and ADP-bound states of each of the two NBDs affect channel gating are not well understood.

### **1.2.2 Expression of CFTR protein**

CFTR is predominantly expressed in epithelial tissue. Northern blot analysis detected CFTR mRNA expression in lung, pancreas, intestine, salivary glands, liver, testis and endometrium (25). By RT-PCR, CFTR mRNA transcript was as well detected in a very low amount in non-epithelial cells, fibroblast and lymphocytes (10). Estimated level of CFTR gene expression in the non-epithelial cells is less than one mRNA transcript per cell. This evidence, together with the fact that the promoter of the CFTR gene has features of a housekeeping gene (10), implies that the CFTR expression might not be limited to the epithelial cells. However, clinical and biological significance of the CFTR in non-epithelial cell is still uncertain.

### **1.2.3 CFTR function**

The aminoacid sequence of CFTR protein shows a high homology to the superfamily of ABC (ATP-binding cassette) transporter proteins (26). This family is specialized within the cell in uptake and secretion, intracellular transport, cell detoxification and signaling. Similarly, the

CFTR protein in epithelial cells functions as the cAMP mediated chloride channel (27-30).

In the cell, CFTR protein is mainly located within the apical membrane where it constitutes a low conductance anion channel (31). Moreover, studies with immuno-electron photomicrographs suggest its localization also inside the cells (32), possibly within membranes of the endosomes where CFTR protein has been shown to be functionally active (33). It seems to function in parallel to H<sup>+</sup> pumps to acidify intracellular vesicles. By changing the degree of vesicles acidification, the presence or absence of CFTR chloride channels might alter pH-sensitive enzymatic processes (34). It is also possible that CFTR located either in the apical membrane or on the intracellular vesicles, may regulate vesicles endocytosis and exocytosis (35). Furthermore, CFTR can also modulate the activity of other membrane channels: CFTR is a negative regulator of amiloride-sensitive epithelial sodium channel. In fact, when CFTR function is lost, sodium conductance markedly increases in human airways (36).

### **1.3 Molecular Mechanism of CFTR Channel Dysfunction**

Mutations within CFTR gene affect the function of the CFTR protein in a number of different ways. It is not known how many mutations really exist within the CFTR gene but the fact that up to now more than 900 different mutations have been described, indicates a high mutation rate in this gene. The most important mutation responsible for CF phenotype is a deletion of a triplet (CTT) in the exon 10 resulting in a consequent deletion of a phenylalanine in the protein. This deletion is marked  $\Delta F508$  (6, 9, 37) and accounts for the 70% of disease alleles. The other mutation, with exception of G551D and G542X, are rare and affect <1% of CF alleles. The list of mutation can be found at the world web

access <http://www.genet.sickkids.on.ca>.

CFTR mutations are classified in five different classes depending on the way in which they affect normal protein function.

- **Class I: mutation affecting protein production**

There are mutations throughout the CFTR gene that produce premature termination signals because of splice site abnormalities, frame shift due to deletions or insertions, or non-sense mutations (38). In some cases, the mutation results in an unstable mRNA and no detectable protein. In other cases, a truncated protein or an aberrant protein containing deleted or novel aminoacid sequences may be produced. Such proteins are often unstable and would usually be expected to be degraded relatively rapidly or have little or no function.

- **Class II: mutations affecting protein processing**

Several CF-associated mutations fail to traffic the correct cellular location. This situation is induced by the most common mutation  $\Delta F508$ . Proteins that fail to progress through the correct biosynthetic pathway and secretion, can be followed by assessing their state of glycosilation. In recombinant cells, CFTR carrying the  $\Delta F508$  mutation fails to mature to the fully glycosylated form (39). Due to the rapid degradation of the partially glycosilated form,  $\Delta F508$  protein cannot be detect on the cell surface.

- **Class III: mutations affecting chloride channel regulation**

This class includes all the mutant proteins that appear to be fully processed and correctly exposed to the apical membrane, but are unable to function as a chloride channel. The range of effect of deregulation of the channel includes those with severe lack of function (such as G551D), reduced response to ATP stimulation (S1255P) and slight reduction of absolute activity (G551S, G1244E and G1349D) (40). Some mutations in this class appear to coincide with aminoacids

involved in ATP-binding.

- **Class IV: mutations affecting chloride channel conductance or channel gating**

A number of CF-associated mutations have been identified within the exons encoding for the multi spanning domains that contribute to the channel pore formation. Three mutations (R117H, R334W and R347P) do not affect the processing, these mutant proteins are correctly exposed in the apical membrane and a cAMP-regulated apical membrane chloride current is also generated. Nevertheless, the amount of current is reduced in respect to wild type protein and it is also different in the different mutants.

- **Class V: mutations causing reduced synthesis**

Mutations of this class include promoter mutations that reduce transcription (41), nucleotide alterations that promote alternative splicing of CFTR transcript (IVS8 T5, 3849+10kb C-T, A455E) (42, 43) and aminoacid substitution that cause inefficient protein maturation (44, 45).

These mutations would reduce relative amount of normal, functional protein. Most CFTR mutations grouped under this class are those generating alternatively spliced mRNA products. Since small amount of normal mRNA is presumably produced from these mutant alleles in addition to the various aberrantly spliced mRNA, a mild functional consequence is expected.

### **1.3.1 Genotype-Phenotype Correlation**

As described in the first section, the presentation of CF is complex. The disease affects multiple organs and the symptoms vary among different patients. In general, phenotypic features may be divided into

three categories. In the first one we can include the symptoms that are common to most CF patients, regardless of the types 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 in the disease that includes symptoms that show a good correlation between genotype and phenotype. This category is best represented by the pancreatic function of the patients (5). 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). Belonging to these modifiers are several candidate genes related to innate and adaptive immune response (46).

The clinical outcome can be explained partially by genotype, or better through the way in which a specific mutation affects the function of the protein. Thus, class I and II mutations result in the absence of protein from the correct cellular location, for this reason they would be expected to have a severe effect on function. In contrast, class III, IV and V mutants are correctly localised, but they are present in a reduced amount or show an inefficient chloride conductance activity.

In this last years, it has been reported how CFTR mutations can also be associated with monosymptomatic forms of CF. This denomination grouped several pathologic conditions characterized by a variety of isolated clinical signs such as congenital bilateral absence of vas deference (CBAVD) (47), nasal polyps (48), bronchopulmonary allergic aspergillosis (49) or idiopathic pancreatitis (50). The phenotypic variability observed in these forms of CF can be associated to mutations that result in alternative spliced forms of CFTR protein (42, 43).

Considering CBAVD, this defect is present in 97% of men with CF, but infertility due to CBAVD does not necessarily coincide with pulmonary



or pancreatic manifestations of CF (51). In this case only a low percentage of wild type CFTR transcript is present and this percentage vary among different patients (52). Extensive studies have been done on CFTR gene in CBAVD patients otherwise healthy for other CF associated symptoms. Chillon *et al.* characterised the mutations in CFTR gene in 102 patients with CBAVD. In most of them only one CF associated mutation was found on one allele while on the other allele was present the T5 variant at the polymorphic locus at 3' end of intron 8. Since the T5 allele causes reduced levels of normal CFTR mRNA, this variant would appear likely to be involved in the pathogenesis of CBAVD (47). From this study has been concluded that the combination of the 5T allele in 1 copy of the CFTR gene with a cystic fibrosis mutation in the other copy is the most common cause of CBAVD. Moreover, the 5T allele mutation has a wide range of clinical presentations, occurring in patients with CBAVD or moderate forms of cystic fibrosis and in fertile men.

## 2 RNA SPLICING

The coding sequences of most eukaryotic genes are interrupted by non-coding stretches of DNA known as intervening sequences or introns. The presence of intervening sequences was first described in non-coding region of adenovirus (53, 54), but was quickly shown to be a common feature of cellular genes. In fact in 1977, for the first time, Jeffreys and Flavell described the presence of "large insert" in the coding sequence of rabbit  $\beta$ -globin gene (55). Introns must be excised from primary transcripts and the flanking exon joined together before the mature RNA is exported from the nucleus. This process is called splicing and occurs in a macromolecular complex known as spliceosome, which consist of five small ribonuclear particles (snRNPs) and a large number of non-snRNP protein splicing factors.

Primary transcript of protein coding gene called pre-mRNA, or named also heterologous nuclear RNA (hnRNA), is synthesised by RNA polymerase II within the nuclei of eukaryotic cells. Its average size is much larger than mRNA and due to this reason it is very unstable. From the time hnRNA emerges from the transcription complex, and throughout the time it is in the nucleus, it is associated with proteins. The physical form of hnRNA is a ribonucleoprotein particle (hnRNP) in which hnRNP proteins (which comprise at least 20 distinct proteins) bind cooperatively to nascent pre-mRNA to form complexes encompassing about 500 nucleotides of RNA (56).

### 2.1 The spliceosome

The removal of introns from pre-mRNA is a critical aspect of gene expression. This reaction takes place in the spliceosome, which is formed by several RNP subunits termed uridine-rich small

ribonucleoproteins (UsnRNP), and numerous non-snRNP splicing factors (57). Each UsnRNP particle consists of a UsnRNA molecule complexed with set of eight Sm or Sm-like proteins and several particle-specific proteins (58).

The spliceosome acts through a multitude of RNA-RNA, RNA-protein and protein-protein interactions to precisely excise each intron and join the exon in the correct order (59, 60).

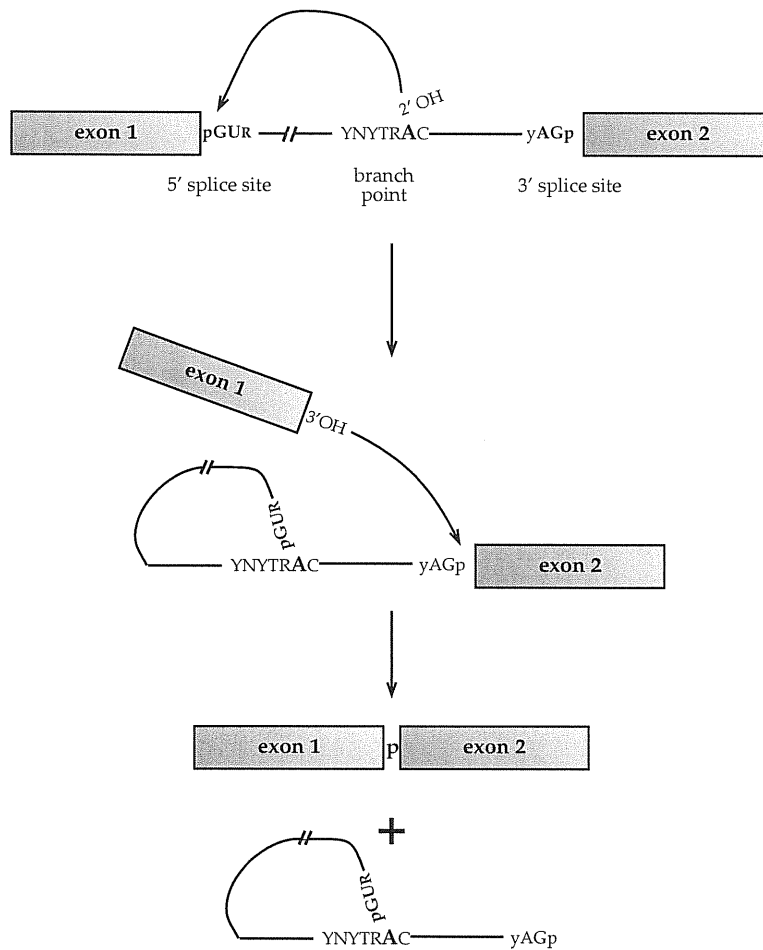
### 2.1.1 The Splicing Reaction

Splicing takes place in two catalytic steps involving two consecutive trans-esterification reactions (61). During step I, an adenosine residue generally located within 100 nucleotides of the 3' end of the intron, in a sequence element known as the branch point sequence (BPS), carries out a nucleophilic attack on the 5' splice site. This reaction generates the splicing intermediates (free exon 1 and lariat-exon 2). During step II, exon 1 attacks at the 3' splice site to generate splicing products (spliced exon and lariat intron) (Figure 2.1).

### 2.1.2 Sequence elements (cis) on pre-mRNA required for splicing

The exact sites for the trans-esterification reactions are defined by consensus sequence, primarily within the intron, around the 5' and 3' splice site (62). These sequences are highly conserved in yeast, despite the fact that only a minority of genes in *S. cerevisiae* has introns, but much less so in mammals.

The first consensus sequence is called 5' splice site and is relatively short. In mammals is AG/GURAGU (where R=purine) and the first two bases of the intron are an essential requirement for the splicing



**Figure 2.1.** Two step chemical pathway of nuclear pre-mRNA splicing. Exons are represented by orange rectangles.

reaction even if are not the only important bases (63). The second consensus at 3' splice site is defined by three separate elements: the branch site, the polypyrimidine tract and the 3' splice site (64). Together these elements make up a loosely defined 3' splice site region, which may extend up to 100 nucleotides into the intron, close to the 3' splice site. The branch site consensus is YNYURAC (Y=pyrimidine) where underlined "A" residues is that at which branch point formation occurs (65). However in higher eukaryotes there is a considerable flexibility, in these sequence requirements. The branch point is located 18-40 nucleotides upstream of the point of cleavage of the RNA. A polypyrimidine tract is usually found between the branch site and 3' splice site. This sequence is more pronounced in mammals than in yeast, where the length of this tract is variable and for this reason introns are classified in short and long polypyrimidine-tract introns, which have distinguishable properties in splicing. The 3' splice site junction is defined by YAG/G.

As previously mentioned in this section, the consensus sequences in higher eukaryotes are less conserved than in yeast. Moreover there are many sequences in mammalian transcripts that match the consensus splice sites but most of them are pseudo-splice sites. To increase the overall fidelity of the splicing reaction additional sequences are present in the exons and in the introns and they are called enhancers or silencers. Their features and role during the splicing will be described in the section 2.4.

## **2.2 Protein components involved in splicing**

In higher eukaryotes, over than 70 spliceosomal proteins have been identified thus far. Although their precise role in some cases remain unknown, spliceosomal proteins have been implicated in a growing

number of functions. Most of these proteins share very similar structural features such as a different number of RNA binding domains and/or protein binding domains.

### 2.2.1 Small Nuclear Ribonucleoprotein Particles (snRNPs)

The snRNP particles form part of the catalytic macromolecular complex of the spliceosome. Each snRNP particle consists of snRNA molecule complexed with a set of eight Sm or Sm-like proteins and several particle specific proteins (58). The major spliceosomal snRNPs U1, U2, U4, U5 and U6 are responsible for splicing the vast majority of pre-mRNA introns (so-called U2-type introns). A group of less abundant snRNPs, U11, U12, U4atac and U6atac, together with U5, are subunits of the so-called minor spliceosome that splices a rare class of pre-mRNA introns, denoted U12-type (66).

snRNAs U1, U2, U4, U5 and U6 are characterized by their small size, metabolic stability and a high degree of sequence conservation (67). Moreover, they show a high complementarity to the consensus splice sites on the pre-mRNA.

The snRNAs are transcribed by RNA polymerase II, with the only exception of U6 (and presumably U6atac snRNA) that is transcribed by RNA polymerase III and entirely assembled in the nucleus. The other pre-snRNAs must be transported to the cytoplasm where snRNPs assembly with the Sm proteins is initiated and then reimported into the nucleus thanks to the bipartite snRNP nuclear localization signal (NLS) formed by Py3cap and the Sm core domain (68). The U4 and U6 snRNAs are extensively base paired (amounting to >20 base pairs) in the U4/U6 snRNP. They associate in the nucleus forming a larger ribonucleoprotein complex. The U5 snRNA then assemble in an ATP dependent reaction with the U4/U6 snRNP giving the U4/U6•U5 three snRNPs particle (69).

The structural core of snRNPs is formed by eight proteins, called Sm proteins, B', B, D1, D2, D3, E, F and G. This class of common proteins plays an essential role in the biogenesis of the snRNPs. The Sm proteins form three distinct heteromeric complexes prior to their interaction with the highly conserved Sm site (PuAU4-6GPu flanked by two stem-loop structures) of the U1, U2, U4 and U5 snRNAs (70). Sm-like proteins belonging to the Sm protein family are specifically required for the assembly of U6 snRNA. This subclass of Sm-like proteins shares the conserved structural motif characteristic of all Sm-proteins family members (71), but can be isolated as a heteromeric complex in the absence of U6 snRNA (72).

Besides Sm proteins there are other particle-specific proteins that associate with snRNAs (58). U1-70K and U1-A proteins bind directly to the RNA and are involved in the splice site recognition and selection, while U1-C associates via protein-protein interactions with U1-70K and other Sm proteins. A subset of U2 snRNP proteins also plays a critical role in tethering the U2 snRNP to the pre-mRNA. These proteins include the heteromeric splicing factors SF3a and SF3b (73) and bind 20-nucleotide region just upstream of the branch site in a sequence-independent manner (74, 75). At least five different proteins associate with U4/U6 snRNP, including a 15.5 KDa protein, polypeptides of 20, 60 and 90 KDa that form complex with one another (76). U5 snRNP particle presents a complex protein composition. U5 220KDa protein cross-links to both 5' and 3' splice sites, as well as to the exon flanking these two splice sites (77).

### **2.2.2 Non-snRNP splicing factors**

- **U2AF**

U2 snRNP auxiliary factor (U2AF) is an essential splicing factor that is

required for the binding of U2 snRNP to the pre-mRNA (78). This factor is a heterodimer composed by two subunits named U2AF65 and U2AF35. U2AF65 contains two functional domains consisting of a sequence specific RNA binding-region composed by three canonical RNA-recognition motif (RRM) and a N-terminal short serine/arginine domain involved in protein-protein interaction (79). U2AF65 interact with the polypyrimidine tract. New findings shed light into the function of U2AF35 subunit. It functions in the recognition of the essential AG dinucleotide at 3' splice site at the earliest stage of spliceosome assembly (80, 81). From these studies has been also shown that U2AF35 is essential for splicing so-called 'AG-dependent' introns, which are introns with a weak polypyrimidine tract, but is dispensable for 'AG-independent' introns, which have a strong 3' splice site.

#### • SR proteins

SR proteins are a superfamily of highly conserved proteins that play an important role in splicing control (82, 83). The name SR proteins reflect the presence of characteristic serine/arginine domain presents in the carboxy-terminal region of these proteins. Sequence analysis of SR proteins reveled the presence of at least one amino-terminal RNA recognition motif (RRM) required for the interaction with pre-mRNA. Various human SR proteins have been identified using different techniques. Currently the SR protein family contains 10 known members: SRp20, SRp30 (SF2/ASF), SRp30b (SC35/PR264), SRp30c, SRp40, SRp46, SRp54, SRp55, SRp75, according to their apparent molecular and 9G8 (84). SR proteins are functionally redundant in splicing of some introns, but exhibit unique functions in removal of others (85, 86). Moreover, they participate both to constitutive and alternative splicing (discussed in section 2.4). It has been demonstrated that the initial binding of SR proteins with pre-mRNA is sufficient to



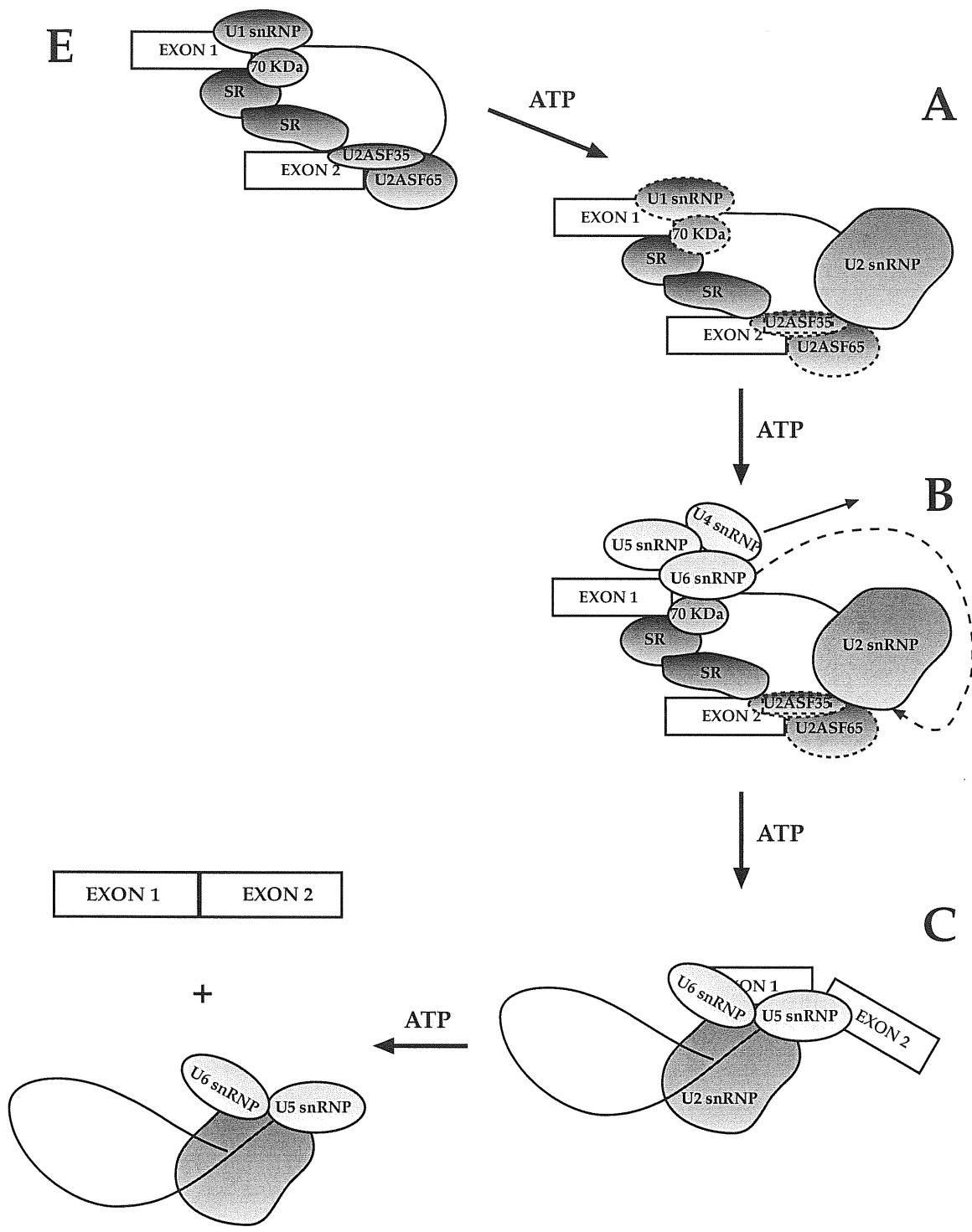
commit the splicing pathway and to facilitate spliceosomal assembly (87). The RRMs mediate RNA binding and determine substrate specificity for individual SR proteins. The hallmark of the RS domain is its significant posttranslational phosphorylation on serine residues (88, 89). The phosphorylation is required for the dissociation of the splicing factors from nuclear speckles, the nuclear compartments enriched in proteins involved in pre-mRNA splicing, and the consequent recruitment to the sites of transcription (90).

## **2.3 The Spliceosomal Complex**

Two of the main functions of the spliceosomal snRNPs are to recognise the 5' and 3' intron/exon boundaries and to assemble onto these sites the macromolecular enzyme that catalyse the splicing reaction. In mammals four distinct spliceosomal complexes, which form in the temporal order E, A, B and C, have been detected (91) (Figure 2.2). These general steps in spliceosomal assembly are now well understood, with some variations in the model proposed few years ago.

### **2.3.1 General model for spliceosome assembly and spliceosome cycle**

Assembly of the major spliceosome is initiated by the ATP-dependent recognition of the 5' splice site by the U1 snRNP, which leads to the formation of the spliceosomal complex E. This interaction is mediated by base pairing of the U1 snRNA with the 5' splice site, as well as by protein-protein and protein-pre-mRNA interaction involving U1-70K and U1-C proteins (58). Novel insights into the spliceosomal assembly are coming from studies in *S. cerevisiae* in which has been shown that the first proteins that interact with the 5' splice site during the commitment complex (E complex counterpart) formation are the cap-



**Figure 2.2.** General model for spliceosome assembly on pre-mRNA. Spliceosomal complexes E, A, B and C are shown. Dashes around U2ASF and U1 snRNP indicate that they become less tightly bound during the E to A and B translocation. Solid arrow in complex B indicates U4 snRNP dissociation and the dashed arrow on the same complex indicate the association between U2 snRNP and U6 snRNP.

binding complex protein CBC80 and seven U1 snRNP (92). Metazoan homologs of some of these proteins have been identified, indicating that similar interaction may occur in E complex. One of these proteins is Nam8 or TIA1 in metazoan, which affects splice-site recognition (93). Another important event that occurs during formation of the E complex is the recognition of the 3' end of the intron. The current model of assembly proposes that U2 snRNP first associate in an ATP-dependent manner in the A complex. However, U2 snRNP has now been identified as a component of a purified, functional E. U2 snRNP association occurs in the absence of ATP and does not required branch point sequence interaction. The interpretation of this data is that U2 snRNP first binds loosely to the pre-mRNA in the E complex via U2-snRNP associated protein SF3b and the U2AF heterodimer. Then an ATP- dependent process leads to stable binding of U2-snRNP to the branch point sequence in the A complex and this is facilitated by the presence of other proteins such as SF3A and SF1 (58). The next major event after A complex assembly is the binding of U4/U5•U6 tri-snRNPs at the 5' splice site to form the B complex. At this time many structural snRNPs rearrangements occur leading to the formation of a complex RNA-RNA network within the spliceosome. U4/U6 base pairing interaction is disrupted and U6 snRNA base pairs with U2 snRNA and also the 5' splice site. Furthermore, U5 snRNA loop I base pairs with exon sequence at the 5' splice site and later with exon sequences at 3' splice site. The 5' and 3' splice sites are also contacted by the U5-220 KDa protein (77), which has been reported to affect also the tertiary interaction between the splice sites and U6 snRNA (94). Recent studies report that tri-snRNPs interact with the 5' splice site and the upstream 5' exon at earlier step in spliceosome assembly than previously thought (95). These interactions occur in an ATP- dependent manner in the absence of stable U2 snRNP binding and appear to be guided, in part, by the U5 associated protein U5-220KDa. In this view,

formation of B complex does not reflect “recruitment” but rather a stabilisation that occurs as a consequence of interspliceosomal RNA-RNA rearrangements such as replacements of U1 by U6 and formation of U2/U6 helices that are prerequisite for the formation of the catalytically competent complex C.

An unsolved problem is the catalysis of RNA in pre-mRNA splicing, although it is widely believed that it is mediated by RNA. Spliceosome is a metallo-enzyme (96, 97) and from recent studies has been shown that U6 snRNA coordinates metal ion thus contributing to pre-mRNA splicing (98). U6 is a good candidate for catalysis because it is highly conserved through evolution and also forms intramolecular and intermolecular helices that are analogous to autocatalytic group II introns (99).

### **2.3.2 Early Complex Formation: Intron and Exon Definition Model**

Interactions between 5' and 3' splice sites and the factors that recognise them have been observed in the earliest steps of spliceosomal assembly. Two models of interaction have been proposed according to intron size (Figure 2.3).

#### **a) Intron Definition Model**

The first model suggested that the intron was the unit recognised by the splicing machinery. According to this model, a mechanism has been proposed to recognise one splice site, then, following a scanning through the intron it locates the second site (100). Such interactions have been observed with *in vitro* splicing mRNA having short introns.

- **The yeast model.**

In yeast, messengers often have unique introns and their length is usually below 100 nucleotides. This provides direct support for the model for the pairing of splice sites across introns during the first step of spliceosome assembly (101).

- Vertebrate intron/exon architecture.

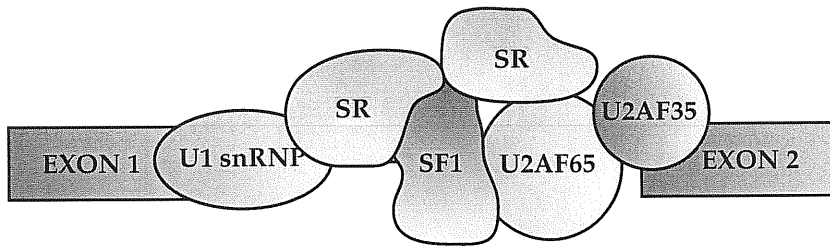
In vertebrate most exons are small and are flanked by long introns. However, some exceptionally long exon exists and when an expanded exon is placed in a gene in which the flanking introns are small, the exon is constitutively included. Experiments on a three-exon minigene containing short intronic sequences showed that an expanded size of the middle exon up to 1400 nucleotides does not lead to exon skipping .

- Drosophila intron/exon architecture.

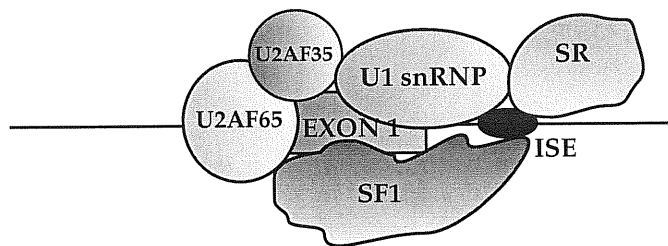
In Drosophila most exons are 100 to 180 nucleotides in length; however 15% are more than 550 nucleotides. 50% of the introns are less than 100 nucleotides and are often flanked by large exons (102). Thus there are a number of genes in Drosophila that have an inverted intron/exon architecture respect the one found in vertebrate genes. At least in three Drosophila's genes, 5' splice site mutations of small introns cause intron retention instead of exon skipping as observed in vertebrates (103, 104). As for vertebrate short exons, Drosophila short intron seems to have size limitation. Expanding the size of the short introns cause repression of splicing or activation of cryptic site. Further evidence for a peculiar mechanism in the recognition of short introns derives from the observation that spliceosomal complex A formation requires sequences at both end of the intron but not a polypyrimidine tract, while in vertebrate requires only sequence at the 3' splice site comprehensive of the polypyrimidine tract (103)

#### **b) Exon Definition Model**

In 1990 Susan Berget's group proposed a model in which the exon was the unit recognised by the splicing machinery and the identification of splice sites was facilitated by interactions across the exon (105). Basically the exon definition theory proposes that in pre-mRNAs with multiple short exons and long introns the splicing machinery search for two closely spaced splice sites. This model explain splice site selection of



**Intron definition model**



**Exon definition model**

**Figure 2.3.** Models of exon/intron definition according to exon and intron size. **A)** The intron definition model has been proposed to be used in systems in which pre-mRNA has small introns. In this cases the intron, rather than the exon, is the initial mode of pairing between splice sites. **B)** The average vertebrate gene consists of multiple small exons separated by considerably larger introns. Differently from the intron definition model, the exon definition model invokes pairing between the splice sites across an exon.

most mammalian pre-mRNA which contain very long introns and weakly conserved splice sites (102). According to this model, an interaction is first established between the 5' and 3' splice sites across the short exon, followed by an interaction across the long intron and the juxtaposition of the neighboring exons.

- Exon size requirement.

The exon length can affect splicing. Internal vertebrate exons have minimum and maximum length requirements (102).

Simultaneous recognition of splice sites bordering an exon suggests that a minimal separation between the sites might be required to prevent steric hindrance between the factors that recognise individual sites. This is demonstrated, when a constitutively recognised internal exon was internally deleted below 50 nucleotides was skipped by *in vivo* splicing machinery (106). In addition, Black et al. (107) have shown that extending the length of the N1 (neuron specific) exon to 109 nucleotides, exon is constitutively included, implying that the exon is normally skipped because it is too short to allow spliceosomes to assemble at both ends simultaneously. Instead, the expansion *in vitro* of internal exons to length above 300 nucleotides determine the activation of cryptic splice site inside the exon or exon skipping (104) demonstrating the exon length limitation for the efficient splicing. In fact, less than 1% of the known internal exons in vertebrate are longer than 400 nucleotides.

On the other hand, expansion of internal exons in vertebrate genes with moderate to large introns has two phenotypes: activation of internal cryptic sites within the expanded exons to create small exons or skipping of the entire exon (105).

- Effect of splice site mutations.

*In-vitro* (103, 106) and *in-vivo* (106, 108, 109) experiments revealed that mutations in a splice sites influence the splicing of both introns flanking the exon and not only of the intron bearing the mutated splice site, as

predicted from intron recognition oriented theories. Furthermore, mutations at the 5' splice site are suppressed by mutations that improve the consensus of the upstream 3' splice site (110).

- A network of interaction spans the exon.

Experiments carried out on the preprotachykinin pre-mRNA showed evidence for exon bridging interactions that occur between U1 snRNP bound at the 5' splice site and U2AF65 bound at the 3' splice site. U2AF65 is indeed recruited to the polypyrimidine tract by interactions with the downstream 5' splice site and the U1 snRNP (111). SR proteins have been implicated in these interactions in different experimental models (see section 2.4.2. ).

- Exon enhancer sequences.

Alternatively spliced exons with suboptimal splice sites can bear peculiar sequences that work as enhancer of splicing helping in the recruitment of splicing factors on an exon otherwise not recognised. These sequences will be explained in section 2.4.

- First and last exon recognition.

The lack of one functional splice site at the first and the last exon indicate a special mechanism for their recognition. To have an efficient removal of the first intron, the capping and the proteins that bind the cap are essential (112). Therefore both, factors that recognise the cap and the 5' splice site, are necessary to define the first exon. The last exon is usually longer than internal exon (average length 600 nucleotides versus 137). The removal of the last intron involves RNA splicing and polyadenylation factors (113).

### **2.3.4 Splice site recognition and bridging interactions**

One of the critical steps of the splicing reaction is the initial recognition of the splice site that lead to the formation of complex E. It has been proposed that splice site recognition and pairing across introns is



promoted by a network of interactions involving the SR proteins. SF2/ASF and SC35 associate via their RS domain with another RS-domain containing protein the U1 snRNP component U1-70K at 5' splice site (114-116).

SR proteins are involved also in the subsequent bridging interaction. Different models have been proposed for bridging interaction according to exon and intron size. In the intron bridging model U2ASF and SF1 cooperatively interact to recognize the branch point and the polypyrimidine tract (117). The RS domain of U2AF35 contacts SC35 that simultaneously interacts with the RS domain of U1 snRNP 70K protein and U1 snRNP is base paired to the 5' splice site (118). These U2AF65-U2AF35-SR-U1 70K interactions were also proposed to function as cross-exon contacts in constitutive and regulated mammalian splicing (119, 120).

Abovich et al. proposed a cross-intron bridging model in which the proteins involved are ubiquitous from yeast to mammals. They identified the yeast protein Prp40p, and suggested the existence of the mammalian homologue mPrp40p, which is involved in the binding of BBP (Branch Point Binding Protein), SF1 in mammals, and U1 snRNP. This model could be applied to small mammalian introns that have features similar to yeast introns (121).

Another model proposed by Kennedy et al. can also be applied to small introns. These sets of introns normally lack a classical Py-tract (122). In this case, SR54p binds to Py-tract replacing in such a way the binding of SF1. SR54p contacts the small subunit of U2AF, U2AF35, facilitating U2AF binding upstream of branch point and position of its small subunit to make optimal contact with U1-70K (123).

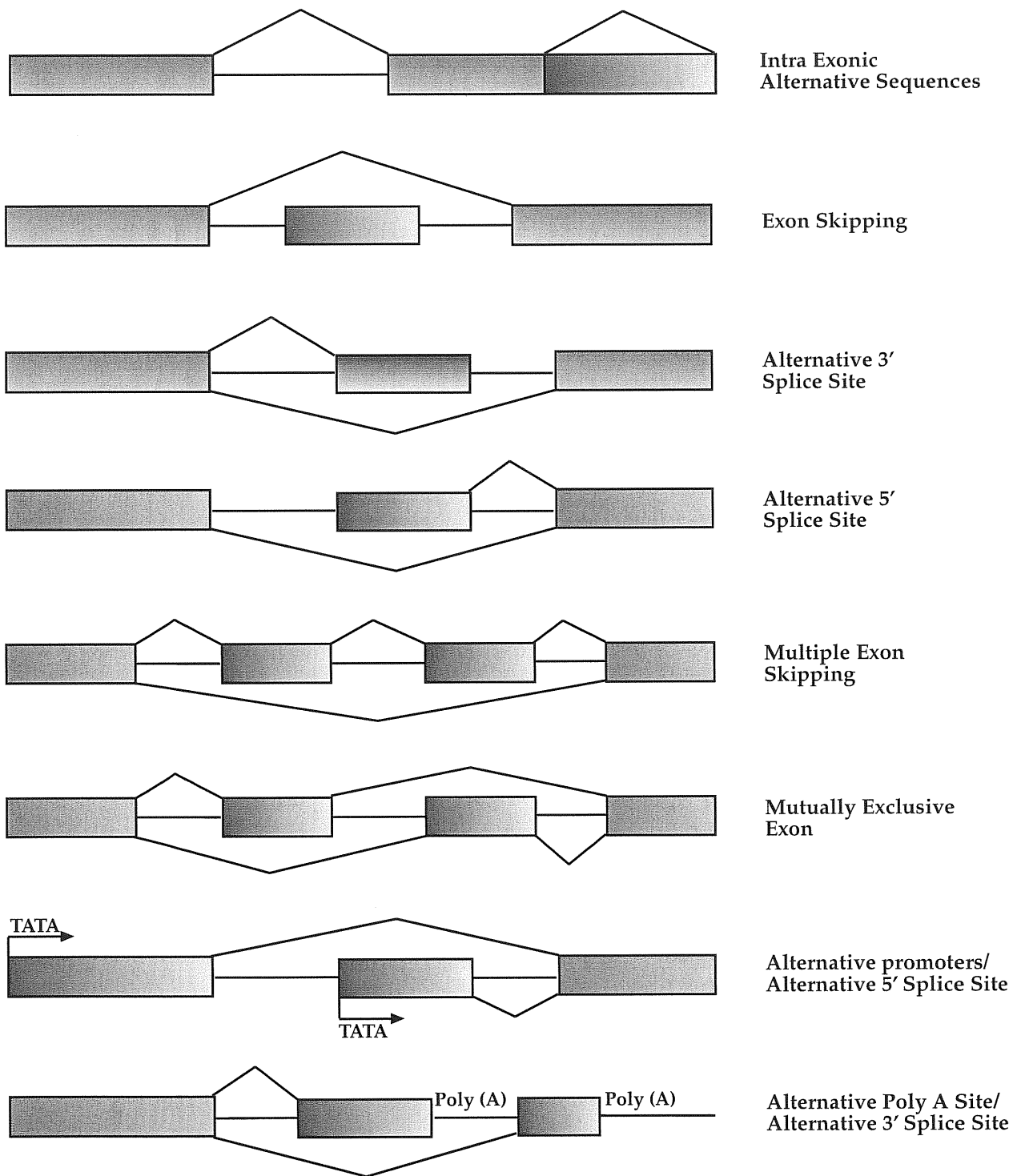
In the Exonic Splicing Enhancer (ESE)-dependent splicing there is a different class of proteins involved. These are SRPy160/300 proteins (SR-related Matrix protein) that are SR repeats containing proteins that lack the RNA recognition motif. In this case has been proposed a

splicing co-activator model for the function of SRPy160/300. SRPy160/300 promotes splice-site pairing and splicing through multiple cooperative interactions with factors bound to pre-mRNA, including SR proteins, U1 and U2 snRNPs (124, 125).

## 2.4 Alternative splicing

Alternative RNA splicing is the process that allows the selection of different combination of splice sites within precursor mRNA (Fig. 2.4). This process is seen in nearly all metazoan organisms as a means for producing functionally diverse polypeptides from a single gene (126). Alternative splicing is especially common in vertebrates. Alignment of EST sequences and mapping the resulting mRNA families to the human genome provided a minimum estimate that 35% of human genes show variably spliced products (127). However, since ESTs derive from a limited number of tissues or developmental stages, and cover a little portion of each mRNA, the true percentage is likely much higher. There are also remarkable examples of hundreds and even thousands of functionally divergent mRNAs and protein being produced from a single gene. A recent discovery in *Drosophyla* is a fascinating example of the extremely high number of proteins that can be produced from one gene. The DSCAM gene presents different exons used in a mutually exclusive manner. For each of these exons there are different numbers of alternative forms. If all the combinations of these exons were used, the single DSCAM gene would produce 38,016 different DSCAM proteins (128).

Variation in mRNA structures takes many different forms (126). Exons can be spliced into the mRNA or skipped. Introns that are normally excised can be retained in the mRNA. The position of either 5' or 3' splice sites can shift to make exons longer or shorter.



**Figure 2.4.** Types of alternative splicing. Schematic representation of different alternative splicing events that occur in nature. Constitutive exon showed as orange boxes, while alternative spliced exon showed as blue boxes. The lines above and below the boxes show possible alternative splicing events. The last two examples describe how splicing and other events, such as transcription and poly-adenylation, are connected.

In addition to these changes in splicing, alterations in transcriptional start site or polyadenylation site also allow production of multiple mRNAs. All of these changes in mRNA structure can be regulated in diverse ways, depending on sexual genotype, cellular differentiation, or the activation of particular cell signaling pathways.

The mechanisms that determine which splice site are utilized and how this selection is regulated in different cell types or developmental stages have still not been precisely defined. Much progress has been made in the identifying the cis-acting elements and the trans-acting factors involved in the alternative splicing.

#### **2.4.1 Involvement of Cis-Acting Elements in Splice Site Selection**

The question of splice-site choice is intimately connected to the problem of normal recognition of constitutive splice sites. A feature shared by both regulatory sequences and splice-site signals is that they are usually short and often degenerate. Cis-acting determinants that influence competing splicing pathways include relative strength of 5' splice site, the branch point sequence, the polypyrimidine tract, the proximity between 5' splice site and branch point, sequences between the branch point-polypyrimidine tract and 3' splice site. In some cases, the formation of pre-mRNA secondary structure participates in the regulation of splice site selection by modifying the physical distance within introns, or being involved in the definition of the exon. Moreover, additional exonic or intronic sequences are often involved in the correct recognition of splice sites, when suboptimal splice sites are present. These sequences can act increasing the recognition or decreasing it and are respectively named enhancers or silencers.

- **Exon/intron architectural requirements**

As described in the previous section, the average vertebrate genes consist of multiple small exons separated by introns that are considerably larger. The average size of exons is 137 nucleotides and exons longer than 300 nucleotides or shorter than 50 nucleotides, even existing, seems to be not favored by splicing machinery (102). Artificially shortening an internal exon leads to inefficient recognition (106, 129), presumably due to both deletion of exon accessory sequences and steric hindrance of factors such as U1 and U2 snRNPs.

Moreover, has been proposed a compensatory relationship between exon and intron size. In fact, Sterner *et al.* Observed that large internal exons are problematic for recognition if they are flanked by large introns, suggested that naturally occurring large vertebrates exons might be flanked by small introns (130).

- **The 5' and 3' splice site requirements**

The strength of splice sites is determinant in their selection. As previously described, at 3' splice site exon/intron junction consensus sequence is determined by 4 bases, while at the 5' exon/intron junction this consensus sequence is extended to 9 bases (62). The GU and AG dinucleotides at 5' and 3' intron boundaries are nearly invariant in higher eukaryotes and mutations in one of these conserved bases completely abolish splicing (131). It has been reported in different studies that mutations of a weak splice site that make it closer to the consensus sequence can lead to a constitutive recognition of an alternatively spliced (132, 133).

Mutations affecting splice sites causing aberrant splicing are observed more frequently in the 5' splice site region than in 3' splice site region (134). Furthermore, alternative-exon splice sites deviate more from the consensus: their 5' splice sites deviate from the consensus sequence

mostly at the +4 and +5 position, while adenosine is more frequently used at the -3 position of the 3' splice site (135).

- **The polypyrimidine tract**

The recognitions of the branch point sequences in the metazoan can be affected by the adjacent polypyrimidine tract (64, 136). It has been shown that progressive deletion of the polypyrimidine tract abolish lariat formation, spliceosome assembly and splicing (136, 137).

Despite the important role of polypyrimidine tract in splicing, there appears to be a great flexibility in the specific sequence of a given tract. In some cases, the introduction of purines into the tract is detrimental to splicing only if its length is shortened and if there is a reduction in the number of consecutive uridines residues (137, 138). Thus, a contiguous stretch of uridines may be optimal, but total percentage of uridines could also determine functional competitiveness. In addition, uridine and cytidine do not appear to function equivalently within a polypyrimidine tract (137, 139).

Not only the number of uridines is important in a polypyrimidine tract, but also its length and these two factors are correlated. In fact, as the number of consecutive uridines decreases, the location of pyrimidine tract becomes increasingly important showing a position dependent effect. Short tracts immediately adjacent to the 3' splice site are utilized more than twice as much as short tracts adjacent to the branch point (139).

- **The branch point sequence requirement**

While in yeast there is an invariant branch point sequence (UACUAAC) which base paired with U2 snRNA, in metazoan this sequence is not highly conserved. However, several lines of evidence have led to the belief that the mammalian branch point is specified primarily by its

proximity to the 3' splice site. Most branch points have been mapped within 18-40 nucleotides of the 3' splice site (78, 140). There are cases in which the branch point is located more than 40 nucleotides from 3' splice site, is functional and is essential for the regulation of alternative splicing. In the rat  $\alpha$ -tropomyosin gene intron 2 branch points lies 172 nucleotides upstream from 3' splice site. This proximity to 5' splice site of exon 2 enforces mutually exclusive behavior of exons 2 and 3, because the splicing factors are unable to bind productively to the two elements and form active spliceosome (141).

Generally, the mutation of the adenine residue involved in the lariat formation strongly reduces splicing efficiency of the downstream exon (142).

#### • RNA secondary structure

The cis-acting elements regulating constitutive and alternative splicing interacts with trans-acting factors. The secondary structures that pre-mRNA creates might influence directly this kind of interaction.

There are many studies that proposed the RNA secondary structure as a regulator of alternative splicing. In the case of rat calcitonin/CGRP pre-mRNA splicing, the 3' splice acceptor of exon 4 forms a stable stem-loop structure. Mutations that destabilize the stem abolish the usage of this splice acceptor site in vitro (143). Also other cis-acting elements are subjected to this kind of regulation. In fact, in fibronectin EDA exon there are an ESS and an ESE. Functional studies coupled to secondary structure analysis suggest that the role of the ESS element may be exclusively to ensure the proper RNA conformation and raise the possibility that the display of the ESE in a loop position may represent a significant feature of the exon splicing regulatory region (133). Moreover, an example of 5' splice site mutation that affects a stem-loop structure can be found in tau gene. In this case the disruption of

secondary structure increase exon 10 inclusion and this result in the production of an aberrant protein isoform that leads to neurodegeneration (144).

#### • Enhancers and silencers

As previously highlighted, higher eukaryotes have large introns defined by more degenerate splicing signals in respect to yeast genes. Additional intronic and exonic sequences are often necessary for efficient and accurate choice of the correct splice site. These sequences can either promote (enhancers) or repress (silencers) splice site usage.

SR proteins bound to splicing enhancers are thought to function by recruiting components of the splicing machinery to the nearby intron (120). Most splicing enhancers are located within 100 nucleotides of the 3' splice site but are not active when are located further away (145). However in *Drosophyla melanogaster* doublesex pre-mRNA, the *dsx* enhancer element functions when located 300 nucleotides downstream of the regulated 3' splice site (146). Recent studies show that the activity of a splicing enhancers decreases as a function of distance from the 3' splice site. Indeed, when a strong enhancer and a weak enhancer are compared, the first one has a higher ability to function at a great distance from the 3' splice site (147).

Most of the enhancer elements obtained contained extended purine-rich sequences (more than 65% purine rich), but a second, novel class of sequences lacking stretches of purines was also identified (148). In fact, Schaal and Maniatis recently identified pyrimidine-rich enhancers that are more than 67% pyrimidine-rich and function such as strong enhancers (149).

In addition to sequences that promote exon inclusion, there are sequences that inhibit splicing, so called exonic or intronic splicing silencers. The silencers are less well characterised; they can be purine or



pyrimidine-rich and bind a diverse array of proteins (150). A negative element (CAGG) involved in the down regulation of fibronectin EDA exon inclusion was identified within EDA exon (151), where it seems to be determinant for RNA conformation (133). An example of intronic splicing silencer is described for CFTR exon 9 alternative splicing. This extended region in intron 9 acts like silencer by recruiting SR proteins (152).

Negative effect of silencers can be due to the interaction with protein factors. A well-known factor that inhibits splicing is PTB (polypyrimidine tract binding protein). It recognises pyrimidine-rich elements both in introns and in exons and can function either by antagonizing U2AF65 action or creating a region of silencing across the downregulated exon (153). In this work we describe the role of PTB in the recognition of CFTR exon 9 alternative.

#### **2.4.2 Trans Acting Factors**

The trans acting factors that regulate alternative splicing are members of at least two protein families. The most important one is the SR protein family and the other is the hnRNP protein family. This last group of proteins is associated with newly synthesized pre-mRNA as well as mature mRNA during nucleo-cytoplasm transport. It seems that they can act on splicing affecting the correct spliceosomal assembly.

- **SR proteins as alternative splicing regulator**

Besides their role on constitutive splicing, SR proteins can also regulate alternative splicing either by altering the 5' or 3' splice site choice in pre-mRNA containing competing sites (154, 155). Enhancer-bound SR proteins can promote the utilization of a downstream 5' splice site and this activity may involve the recruitment of U1 snRNP to the splice site.

This model has been proposed using ASF/SF2 (114). In the presence of competing 5' splice sites, different SR proteins show distinct preferences for promoting the use of proximal or distant splice site. High concentration of ASF/SF2 and SC35 proteins promotes the use of 5' splice site proximal to the 3' splice site, while SRp40 and SRp55 promote the use of distal sites (83).

Another mechanism of action proposed for SR proteins-promoted use of nearby splice site involved U2ASF. U2AF35 subunit establishes a network of protein-protein interaction involving SR proteins. It also binds U2ASF65 subunit, which interact with polypyrimidine tract. In such a way, SR proteins promote assembly of committed complex by assisting binding of U1 snRNP (119, 156).

There are different components that determine the effect of SR proteins in the regulation of splice site choice; one of these is the level of their expression. In the case of CD45 different isoforms are expressed according to T cell activation. Specific SR proteins have antagonistic effect on this alternative splicing and also the expression pattern of SR proteins changes during the activation (157).

#### • hnRNP as a splicing regulator

The hnRNP were first described as a major group of chromatin-associated RNA-binding proteins. These proteins contain RNA-binding motifs that possess different RNA sequence binding preferences (56). In human hnRNP complexes more than 30 proteins have been identified and traditionally it was described their principal function in packaging of nascent pre-mRNAs in order to protect and organised them in a nucleosome-like structure (158). In recent years a more dynamic role of hnRNP proteins has been suggest. These protein may be viewed as a subset of the trans-acting factors involved in pre-mRNA processing and some of them are also acting in nucleus-cytoplasm mRNA export

[Reviewed in (159, 160)].

Several intronic splicing enhancers and repressors associate with hnRNPs. One of the best-analysed hnRNP protein is hnRNP A1. It was observed that in splicing extracts the relative ratio between A1 and the SR protein SF2/ASF determines the usage of duplicated 5' splice sites in a  $\beta$ -globin construct. While high relative concentrations of A1 favor the choice of distal 5' splice site, an excess of SF2/ASF result in the utilisation of proximal 5' splice site (161). The same competitive effect between A1 and SF2/ASF has been shown also in the case of two alternative 3' splice sites. In the same manner A1 promotes the use of distal 3' splice site, while SF2/ASF promotes the use of proximal 3' splice site (154).

Another well-studied protein of this family is hnRNP I, best known as PTB (Polypyrimidine Tract Binding Protein). This protein is involved in the regulation of alternative splicing of several genes [Reviewed in (162)]. PTB recognises as optimal binding site the RNA sequence UCUU within a polypyrimidine rich context (163) that sometimes overlap with U2ASF binding site and simple competition could account for the inhibitory action of PTB. A well-characterised example of a potential competition model is the repression of the 24-nucleotide exon of the GABA<sub>A</sub> $\gamma$ 2 pre-mRNA (164). In most other instances, PTB binding sites do not overlap with binding sites of general splicing factors and in this case the effect is not mediated by a direct competition. The majority of the exons silenced by PTB are flanked by PTB binding sites on both adjacent introns. Given that PTB can multimerised, it has been postulated that the interaction of PTB proteins can create a zone of silencing throughout the exon (165, 166).

In this thesis, we will show for the first time how PTB is involved also in the regulation of CFTR exon 9 alternative splicing.

## 2.5 Regulation of expression of the splicing factors and regulation of alternative splicing

As previously mentioned, alternative splicing is spatially and temporally regulated and this can be directly linked to the ratio of different trans-acting factors expressed in a tissue specific manner or in a developmentally regulated manner.

It has been shown that genes that code for SR proteins such as SC35, SRp20 and SRp40 can be up regulated by mitogenic stimuli and this can in turn affect the recruitment of different SR proteins to a transcript.

In the previous section has been reported the important balance between SF2/ASF and hnRNP A1 in the choice of competing splice sites. A natural stimulus that influences this ratio in the nucleus is, for instance, genotoxis stress which induce hnRNP A1 phosphorylation with a consequent accumulation of the protein in the cytoplasm (167). The resulting decrease in nuclear hnRNP A1 relative to SR proteins alters splice site selection.

An additional mechanism of control is correlated to the level of expression of different trans acting factors in different tissue. In rat tissue, was found that molar ratio of hnRNP and SF2/ASF can vary at least 100 fold among different tissue, supporting the notion that changes in the ratio of these proteins can affect alternative splicing (168).

In this thesis we will show how the expression of two trans acting factors regulating CFTR exon 9 alternative splicing vary among different human tissue and also is direct correlated with the organs mostly affected by the disease (169).

### 3 ALTERNATIVE SPLICING OF CFTR TRANSCRIPTS

In the first part of this chapter it has been reported the large number of mutation that affect CFTR gene and may cause cystic fibrosis. Another form of CFTR variant, the alternative splicing transcripts, has been reported in vivo both in CF patients and in normal individuals. The most common observed CFTR alternatively spliced variants are those lacking exon 4, exon 9 and exon 12. Evaluation of the relative amounts of the transcript variants indicated that the exon 4 transcript represents less than 2% of total transcript (170), exon 12 transcript represents 5-30% and transcripts lacking exon 9 represents up to 92% of total CFTR transcript (171).

These are not the only splicing variants in CFTR transcript. By RT-PCR analysis from nasal epithelium of CF patients and from non-CF control subjects other four alternatively spliced forms were detected. These include transcripts lacking exon 3, exons 2-3, exon 9+12 and the final 357 base pairs of exon 15 as a result of the use of a cryptic splice site (172).

There are also two examples of alternative splicing that lead to the insertion of an intronic portion in the transcript. The first described is the insertion of 260 base pairs between CFTR exon 23 and 24. This additional sequence was named exon 24a and the transcript represents 3-16% of the total CFTR transcripts in epithelial and non-epithelial cells. The insertion of the 24a exon introduce a premature stop codon that would result in a CFTR protein shortened by 61 amino acids at the carboxyl terminal end (173). The second alternative splicing event that causes the insertion of an intronic portion involved intron 10. In this case, 119 base pairs are introduced; the resulting exon is named exon 10b and account for about 5% of total CFTR transcript. Also in this case exon 10b sequence introduces an in-frame stop codon and it would be translated in a CFTR protein truncated at the level of the first NBD

(174).

A comparison between mouse and human shows how these alternative spliced variants are not conserved. The most common CFTR alternative spliced form in mouse are the one lacking exon 5 and the one with additional exon 11b which respectively represent 20-40% and 10-15% of the total CFTR mRNA (175, 176).

Up to now it is not clear if these transcripts have a physiological role. Analysis employed to search for the presence of the mentioned truncated forms of the proteins failed. The functional consequences of alternative splicing have been investigated in human for exon 9 transcript and for mouse exon 5 transcript. In both cases the expression of the aberrant protein failed to generate phosphorylation-regulated chloride channel (175).

### **3.1 Alternative Splicing of CFTR pre-mRNA and Monosymptomatic Forms of Cystic Fibrosis**

In these last few years, it has been reported how CFTR mutations can also be associated with a variety of isolated clinical signs such as congenital bilateral absence of vas deference (CBAVD) (47), nasal polyps (48), bronchopulmonary allergic aspergillosis (49) or idiopathic pancreatitis (50). All these pathologic conditions are grouped under the name of monosymptomatic forms of CF. Considering CBAVD, this defect is present in 97% of men with CF, but infertility due to CBAVD does not necessarily coincide with pulmonary or pancreatic manifestations of CF (51). CBAVD is thus considered a primary form of CF supported by the identification of mutation in the CFTR gene in the majority of patients (47). For example, 50-82% of men with CBAVD have at least one detectable CFTR gene mutation, about 15% have two detectable mutations and at least one of the two appear to be of a mild

form (177, 178). The inability of the investigators to identify two CFTR mutations in these patients could be explained by the presence of mutations in the non-coding sequence that would lead to the production of low level of CFTR protein. This reduction may cause obstruction of vas deference, but may be insufficient to cause the disease in the other organs. The reduction in CFTR protein level could be ascribed to a decrease proportion of normal transcript caused by alternative splicing. Interestingly, a study of the exon 9 intronic junctions revealed the presence of a polymorphic TGmTn region at the 3' end of intron 8 (42, 179). Genetic studies on the polymorphic T variant permitted to associate the T5 allele with the CBAVD disease and some other monosymptomatic forms of CF. Patients with T5 allele indeed were found to present a low proportion of normal mRNA but a higher proportion of mRNA lacking exon 9 (47, 180) compared to T7 and T9 polymorphic variant alleles.

Other study showed that the level of correctly spliced mRNA in respiratory epithelial cells correlates with the severity of lung disease. Individuals with normal lung function or minimal lung disease had >25% of normal transcripts, while when the level goes down to 6-24% the severity of lung disease increases (181).

The level of normally spliced transcript has been studied in two other CFTR splicing mutations: 3849 +10 Kb C⇒T (43) and 1811+1.6 Kb A⇒G (182). In both cases low level of normal transcript, 8% and 3% respectively and in respect to normal control, is associated with severe lung disease.

### **3.2 Association Studies of Alternative Splicing of Exon 9 in CFTR**

CFTR exon 9 encodes for a region encompassing 21% of NBD1. The in-

frame skipping of exon 9 leads to the production of a non-glycosylated and non-functional chloride channel (175, 183). However, studies on CFTR transcripts lacking exon 9 in normal individual homozygous for T5 allele, showed that they can represent up to 92% of total CFTR transcripts (171). This indicates the high tolerance of the organism for the low proportion of CFTR mRNA containing exon 9.

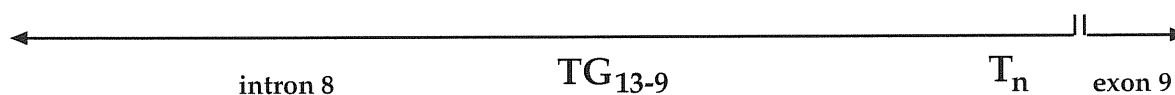
Extensive association analysis has been done between the polymorphic locus at 3' end of intron 8 and the level of exon 9' transcript (Figure 3.1). The number of TG dinucleotide repeats goes from 9 to 13 while the number of T can be 3, 5, 7 and 9. A direct correlation between the length of this poly-timidine tract and the efficiency of CFTR exon 9 inclusion exists: higher the number of T is, higher is the amount of correct transcript (42). The opposite situation is found when we consider an increase in the number of TG repeats: the amount of correct CFTR transcript decreases (184). Thus these two elements work in a cooperatively fashion to modulate the level of correct CFTR mRNA.

The proportion of CFTR alleles with each specific genotype of the TG dinucleotide repeats/or the poly-T tract varies among the non-CF and CF alleles. The most common genotypes in non-CF alleles are (TG)<sub>11</sub>T<sub>7</sub>, (TG)<sub>11</sub>, and T<sub>7</sub>. In contrast, (TG)<sub>10</sub>T<sub>9</sub>, (TG)<sub>10</sub> and T<sub>9</sub> are the most common genotypes in CF alleles (42). The frequency of T5 allele in CBAVD patients is 6 times higher (30%) than that of the general population (5%) (47).

However, the identification in our work of a CF patient carrying the TG13T3 genotype indicates that the (TG)<sub>m</sub>(T)<sub>n</sub> variability is not only associated with monosymptomatic forms of CF but that its extreme variant may also be associated with pancreatic-sufficient CF (185).



TTTGGATG TGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTG TTTTTTTTTT AACAG  
TTTTGATG TGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTG TTTTTTTTTT AACAG  
TTTTGATG TGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTG TTTTTTTTTT AACAG  
TTTTGATG TGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTG TTTTTTTTTT AACAG  
TTTTGATG TGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTG TTTTTTTTTT AACAG  
TTTTGATG TGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTG TTTTTTTTTT AACAG  
TTTTGATG TGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTG TTTTTT AACAG  
TTTTGATG TGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTG TTTTTT AACAG  
TTTTGATG TGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTG TTTT AACAG



**Figure 3.1** Schematic representation of the polymorphic tract at 3' end of intron 8 of CFTR gene. Different examples of (TG)<sub>m</sub> (in red) and (T)<sub>n</sub> (in blue) variants are reported.

### **3.3 Cis-acting Elements Involved in the Regulation of CFTR Exon 9 Alternative Splicing**

In some cases, the proportion of exon 9 skipping does not correlate to the polymorphic alleles at the 3' end of intron 8 (42), and it varies among tissue of the same subject (186). This suggests that other factors operate in conjunction with the polymorphic locus to regulate the amount of exon 9 skipping.

Multiple factors, as reported in part 2 of this introduction, are known to be involved in the regulation of alternative splicing process through a complex network of interactions between them and pre-mRNA. The characterization of splicing factors and cis-acting elements involved in the regulation of alternative splicing of CFTR exon 9 is a key element to understand the molecular basis of this skipping.

Using an *in vivo* approach we identified an intronic element different from the polymorphic locus that act as a splicing inhibitor. The experiments are based on the substantial difference between human and mouse exon 9 alternative splicing and also on the difference between human and mouse intron sequences surrounding exon 9. Hybrid human/mouse minigene system were used to identified a region of about 150 base pairs in human intron 9 that were named Intronic Splicing Silencer (ISS) since it mediates exon 9 splicing inhibition (152).

The ISS is another element that acts together with the (TG)<sub>m</sub>(T)<sub>n</sub> polymorphic locus to modulate CFTR exon 9 alternative splicing.

### **3.4 Trans-acting Factors Involved in the Regulation of CFTR Exon 9 Alternative Splicing**

In these last few years, different splicing factors belonging to the SR

protein family have been shown to regulate the alternative splicing of many pre-mRNAs (133, 187, 188). These factors interact with cis-acting elements and regulate the splicing process.

Using an *in vitro* approach consisting of UV cross-linking experiments and band shift assays, we were able to identify the proteins that interact with TG polymorphic locus and also with the ISS. While in this last case we identified SR proteins as the trans-acting factors that recognise the silencer (152), for TG locus we identified TDP43 as a novel factor that bind this region (185). This protein was described for the first time in 1995 as a factor that binds HIV-1 TAR DNA and function as a transcriptional inhibitor (189). TDP43 is indeed involved in the formation of a 50-52 KDa complex that assembles on the (TG)<sub>n</sub> elements at the 3' end of intron 8 (185).

In this thesis I have focused my attention on the functional significance of the interactions between TDP43 and the UG repeat element. Overexpression of TDP43 induces exon 9 skipping, whereas antisense inhibition by antisense oligonucleotides induces exon inclusion.

Moreover, for the first time I have reported the relevant role of intron 9 sequences close to the 5' splice site as splicing enhancers and the functional involvement of PTB in the complex regulation of CFTR exon 9 alternative splicing.

My attention has been also focused on the exon 9 sequence and on the effect that point mutations could have on alternative splicing.

## RESULTS

## Part 1: Functional Effect of TDP43 on Alternative Splicing of CFTR Exon 9

The variability at the polymorphic TG(m)T(n) locus at 3' end of CFTR intron 8 was the first element regulating the splicing efficiency of exon 9. Evaluation of the RNA binding property of this element using several *in vitro* techniques showed TDP43 as the protein that specifically interacts with the UG repeats of the polymorphic region (185). In this part of my experimental work I have evaluated the functional role of TDP43 *in vivo* on the alternative splicing of CFTR exon 9 using the hybrid minigene system. The intracellular levels of the putative regulatory splicing factor TDP43 was modulated in Hep 3B cells by increasing or reducing its concentrations with cotransfection of TDP43-coding plasmid or by antisense-mediated inhibition, respectively.

### 4.1 Overexpression of TDP43 in Transfection assays

The strategy used to study the functional role of TDP43 is based on cotransfection experiments of hybrid CFTR exon 9 minigene variants along with a TDP43 expression vector. The hybrid minigenes utilised consist of variants at the polymorphic locus with different number of TG and T repeats; we consider hybrid minigenes with four different alleles at the polymorphic locus, TG11T5, TG11T3, TG13T5 and TG13T3 (Fig. 4.1 A). The TG13T3 allele was a new allele we have found in a pancreatic-sufficient CF patient (which is described in detail in the section 4.4), which presented a high level of exon 9 skipping *in vivo*.

Transfection of the different variants showed that the percentage of exon 9 exclusion was strictly dependent on the composition of the polymorphic locus, being directly related to the TG(m) and inversely related to the T(n) repeats.

In the absence of overexpressed splicing factors, the highest level of exon exclusion was observed with higher TG(m) repeats, ranging from 55% (TG11T3) to 86% (TG13T3) (Figure 4.1). This result extended previous data reported by our group using similar hybrid minigene variants (184).

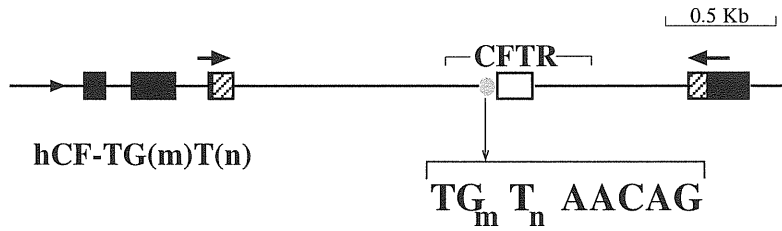
Most importantly, cotransfection of TDP43 coding-plasmid, induce an increase in the amount of exon 9- mRNA and this splicing inhibition was evident for all the minigene variants (Figure 4.1). This result indicates that TDP43, interacting with UG repeats, acts as a splicing inhibitor of CFTR exon 9 alternative splicing (Figure 4.1 B,C).

In addition we also coexpressed or in conjunction with TDP43 the regulatory splicing factor SF2/ASF. This splicing factor binds to a different region in intron 9, the ISS, acting as an exon 9 splicing inhibitors (152). In fact overexpression of SF2/ASF cause an increase in the amount of CFTR exon 9 exclusion (Figure 4.1, B) The combined overexpression of the two splicing factors along with the TG11T5 hybrid minigene resulted in an enhanced inhibitory effect (Figure 4.1 B). This indicates that the binding of the two factors on either side of exon 9 contribute independently to the splicing inhibition.

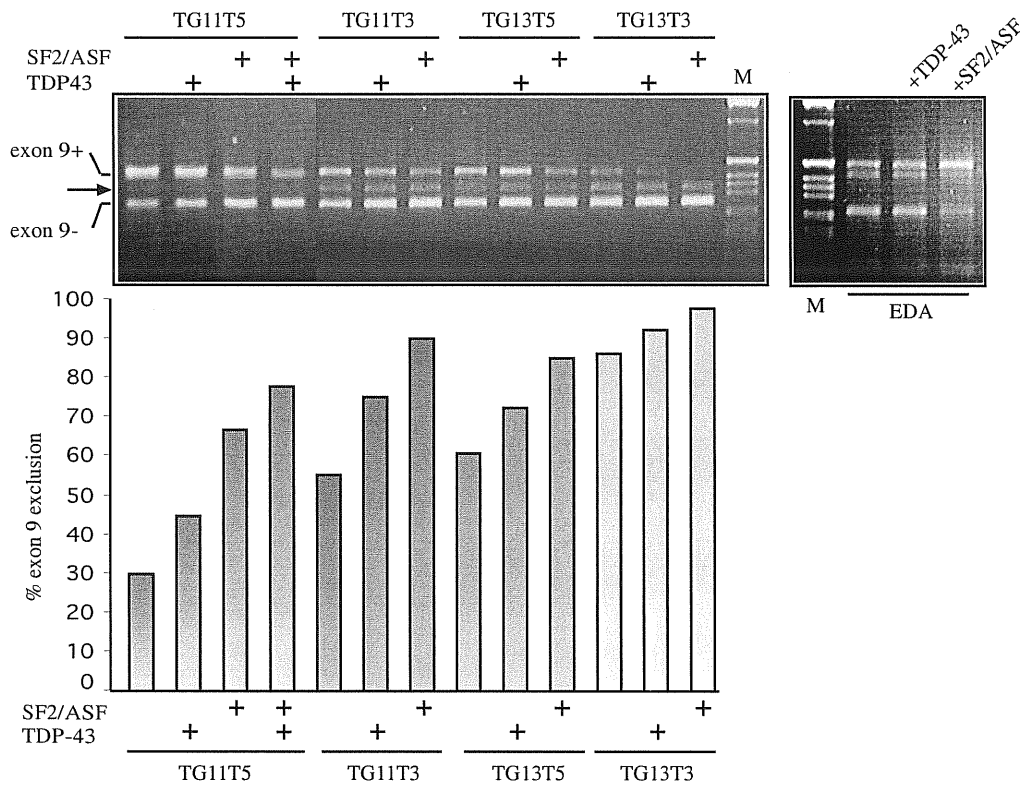
#### **4.1.2 The splicing regulatory role of TDP43 is Strictly Dependent on the Presence of UG Repeats**

TDP43 has never been reported to bind RNA and to act as a splicing factor. To check if the functional effect that we observed in the overexpression experiments is a specific function of the ability of TDP43 to bind TG repeats, we performed a cotransfection experiment on a hybrid minigene containing the fibronectin EDA exon along with the CFTR and TDP43 or SF2/ASF constructs. The alternative splicing of the EDA exon, differently from that of CFTR exon 9, is positively regulated by SF2/ASF and in the EDA intronic junctions no TG repeated tract

**A**



**B**



**Figure 4.1. A)** A schematic representation of the hybrid minigene, hCF-(TG)mTn. The minimal  $\alpha$ -globin promoter and SV40 enhancer are indicated by a small arrow at 5' end, the polymorphic locus (TG)mTn by a gray circle, and the  $\alpha$ -globin, fibronectin EDB and human CFTR by black, shades and white boxes respectively. The primers used in the RT-PCR assay are indicated by superimposed arrows. **B)** The left upper panel shows the expression of selected minigene variants in the presence of plasmids overexpressing either TDP43 or SF2/ASF. A 1.5% agarose gel is used. Exon 9+ or exon 9- bands are indicated. The percentage of exon exclusion for each construct either alone or in the presence of SF2/ASF (500 ng), TDP43 (3  $\mu$ g) or both, is reported in the lower graph. The right panel shows the effect of TDP43 and SF2/ASF overexpression on the fibronectin EDA exon. The arrow on the RT-PCR results indicates an aberrant splicing product originating from a cryptic 3' splice site located in the exon 9.



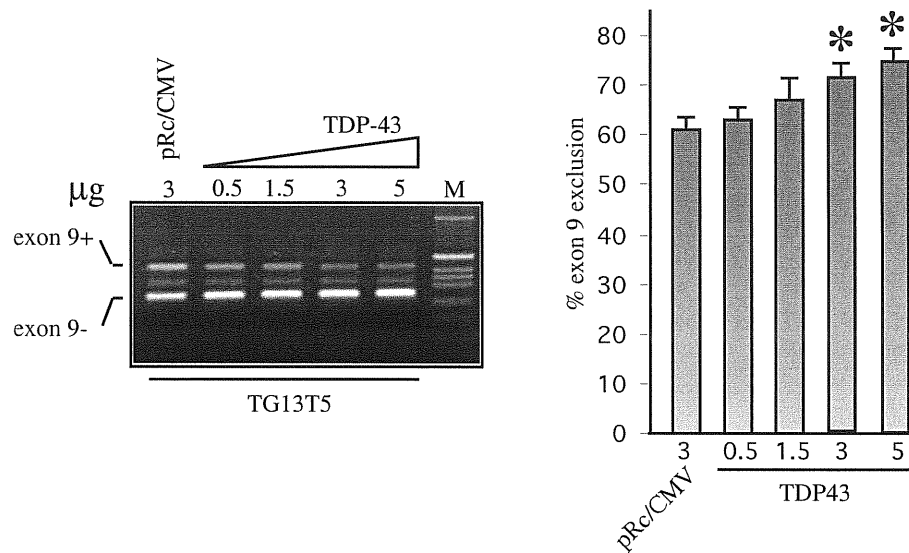
could be found (133, 190). The results of these experiments show that TDP43 does not affect the splicing pattern of EDA exon, whereas SF2/ASF produces an increase in EDA exon inclusion, accordingly on previously reported data (Figure 4.1 B, right panel).

#### **4.1.3 Dose-Response Effect of TDP43**

Once confirmed the specific inhibitory effect of TDP43, we evaluated how the exon 9 exclusion is affected by increasing amount of TDP43. The TG13T5 minigene was cotransfected with increasing amounts of TDP43 expressing vector. The results show that the inhibitory effect of the protein on exon 9 inclusion follows a dose-dependent curve that, considering the high level of basal exclusion, was statistically significant in the last two data points when 3  $\mu$ g and 5  $\mu$ g of TDP43 expressing vector were respectively used (Figure 4.2).

## **4.2 Transient Transfections of Antisense Oligos Directed Against TDP43 mRNA**

We noticed that in cotransfection experiments with CFTR minigenes the splicing inhibitory effect of TDP43 was significantly lower compared to the corresponding experiments done with the SF2/ASF splicing factors. We reasoned that this may be due to the relative high level of TDP43 in the cells, which saturate the UG binding sites. Thus we choose an alternative approach to study the TDP43 function using antisense phosphorothioate(PS)-oligonucleotides to inhibit the expression of the splicing factor.

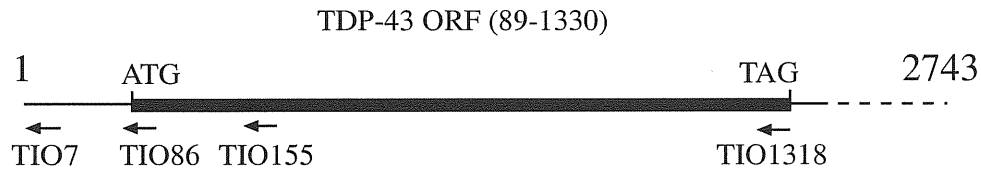
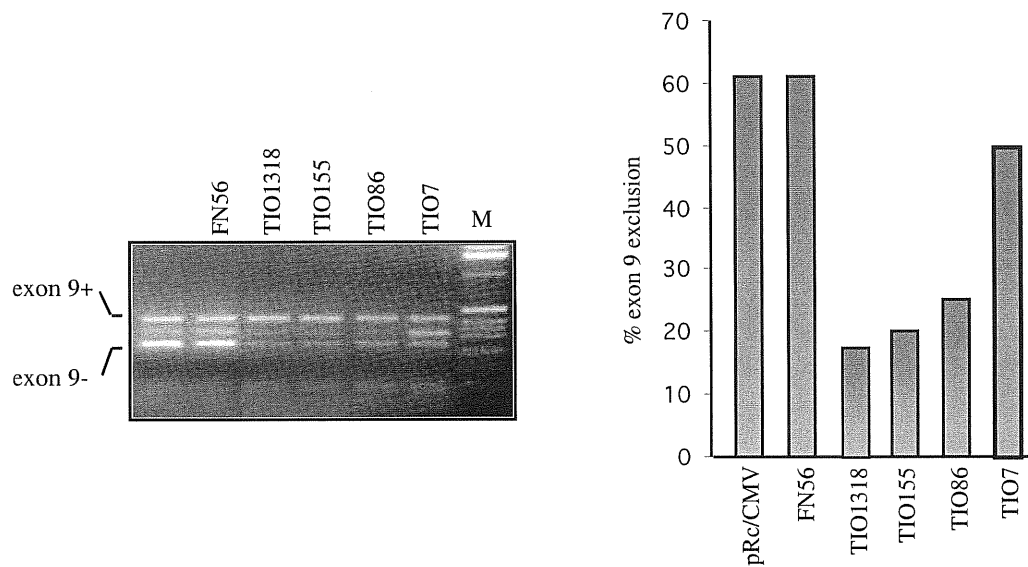
**C**

**Figure 4.2** A dose-response curve of exon 9 inclusion in the presence of increasing amounts of TDP43 expressing plasmid (from 0.5 to 5  $\mu\text{g}$ ) on TG13T5 minigene (left panel). The results shown in the right panel are the mean values from four independent transfection experiments performed as duplicates. The asterix indicate statistical significance ( $P < 0.05$ ). M is the molecular weight marker (1Kb).

#### 4.2.1 Transfection of Antisense PS-oligos Induces CFTR Exon 9 Inclusion

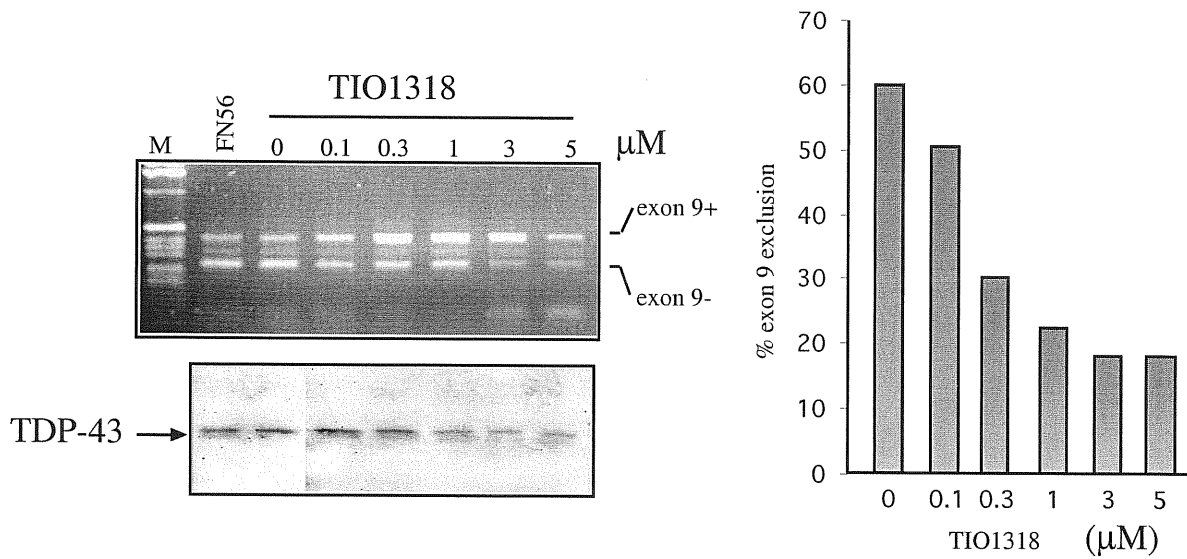
We designed four antisense PS-oligonucleotides considering several properties that have been proposed to improve their efficiency. We considered the length of the oligonucleotides, the target sequence composition on the mRNA and the presence of tetranucleotide GGGG motif in the transcript (191, 192). The first antisense PS-oligo, TIO7, was designed to anneal at the beginning of 5' the untranslated region, the second, TIO86 anneals across the ATG region, the third oligo, TIO155 anneals in the region coding for the first RNA Recognition Motif (RRM) and the last one, TIO1318 anneals across the stop codon (Figure 4.3 A). Cotransfection of with PS-oligos along with TG13T5 minigene in the Hep 3B cell line showed a consistent and significant increase of exon 9 inclusion in the processed minigene transcripts for three oligos: TIO86, TIO155 and TIO1318. On the other hand, no variations in exon 9 inclusion were observed for a control unrelated oligo, FN56 (Figure 4.3 B). Interestingly, the most efficient PS-oligos were TIO155 and TIO1318, which were designed to contain the sequence complementary to the GGGG motif proposed to be more accessible to antisense oligonucleotides than other sites.

Then we have chosen the antisense PS-oligo that worked better in the transfection assays and we performed a dose-dependence response to it. Increasing the concentration of TIO1318 we observed a concomitant increase in the proportion of transcripts containing exon 9 (Figure 4.3 C). To confirm that the effect we have observed is due to a decrease in the level of endogenous TDP43 protein, we performed a Western blot analysis on the cellular protein lysates obtained from the dose-response transfection experiments. Indeed, this analysis showed that the endogenous TDP43 level gradually decrease concomitant with the increase in TIO1318 (Figure 4.3 C).

**A****B**

**Figure 4.3.** Antisense inhibition of TDP43 in Hep 3B cell line transfected with TG13T5 minigene. **A)** A schematic diagram of four PS-oligodeoxynucleotides (TIO7, TIO86, TIO155 and TIO1318) used. Hep 3B cells were co-transfected with 3  $\mu$ g of TG13T5 minigene and each PS-oligo or a FN56 control oligo at a final concentration of 1  $\mu$ M. **B)** RT-PCR results of the transfection experiments were analysed in a 1.5% agarose gel. A TG13T5 control was also included (pRc/CMV2). The percentage of exon 9 exclusion is reported on the right panel.

C



**Figure 4.3 C** A dose-response curve with PS-oligo TIO1318. The amount of oligo used ranging from 0.1 to 5  $\mu\text{M}$  (upper panel). The negative control used in the transfection experiments is FN56. In the lower panel there is the western blot of the endogenous TDP43 level correspondent to the transfection with the antisense PS-oligo TIO1318. On the right a graph presents the percentage of exon 9 exclusion.

We have attempted to study the effect of antisense inhibition of TDP43 with TIO1318 PS-oligo on the endogenous CFTR exon 9 splicing in Hep 3B, T84 and the patient's lymphoblastoid (HA 66) cell lines. The Hep 3B endogenous CFTR gene has the TG11T7 configuration and the exon 9 is included up to 95% of the mRNA, this made very difficult to critically assess any increase in exon inclusion efficiency. On the other hand, T84 and HA 66 have a much better ratio of exon 9 exclusion, but they were very sensitive to the effective PS-oligo concentration.

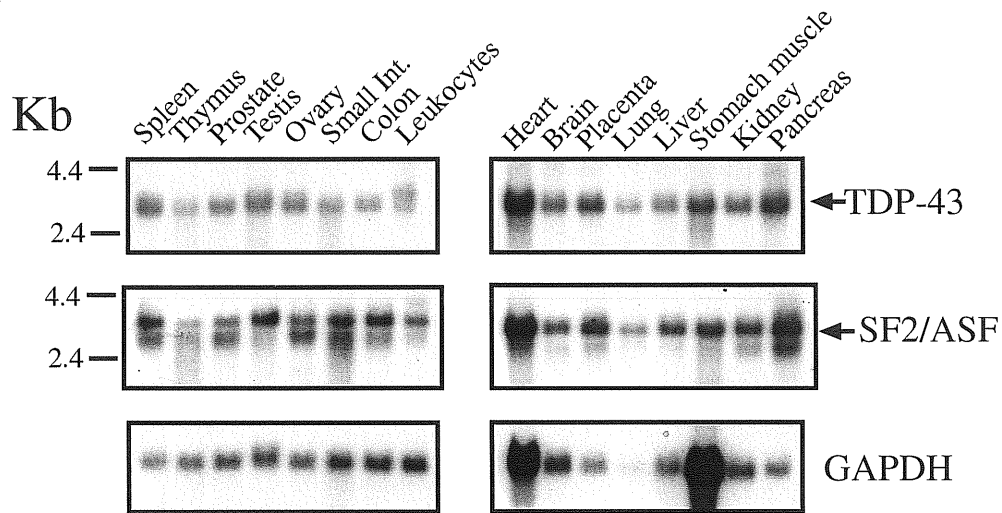
Together this results and our previous data, indicate that TDP43 binding to the polymorphic TG(m) locus at the 3' end of intron 8 inhibits CFTR exon 9 alternative splicing.

### **4.3 Distribution of TDP43 and SF2/ASF in Human Tissue**

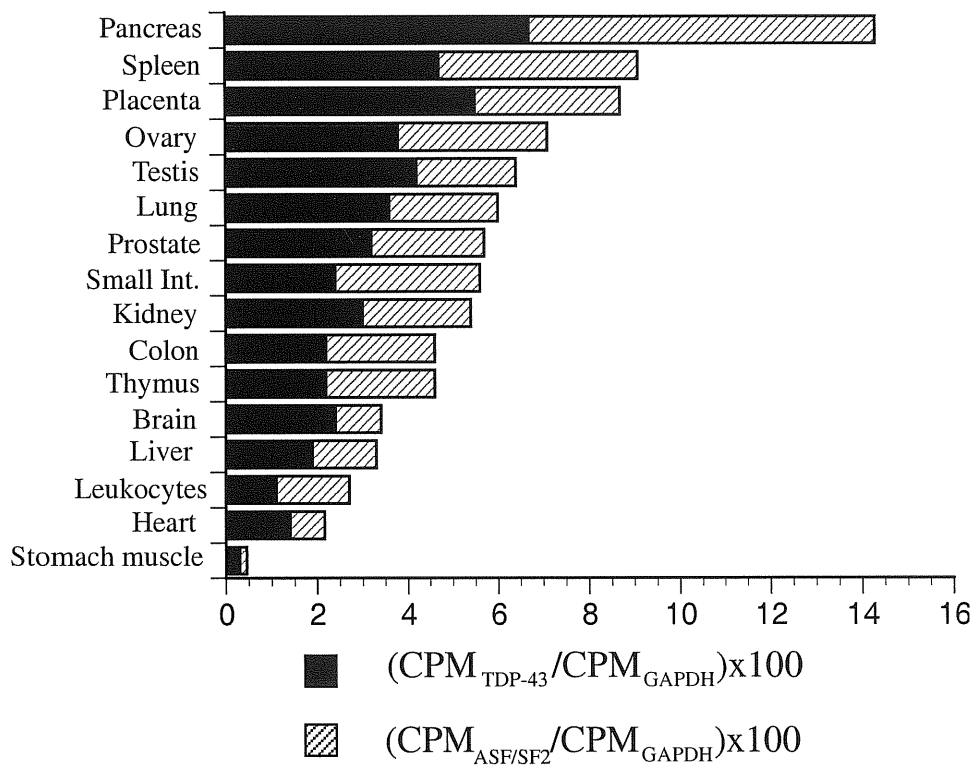
Alternative splicing is an evolutionary and tissue-dependent event found in different genes. The inclusion or the exclusion of an alternatively spliced exon in a transcript is regulated both by the cis-acting elements in the pre-mRNA sequence and trans-acting factors that interact on them. The expression of the trans-acting factors could be regulated in a tissue-dependent fashion.

The two splicing factors inhibiting CFTR exon 9 inclusion, TDP43 and SF2/ASF, have been found to be expressed in several human tissues (189, 193). In order to quantitate the expression levels and their relative concentrations we performed northern blot analysis on commercially available human polyA(+) northern blots. TDP43 and SF2/ASF transcripts were detected with variable concentrations in all sixteen human tissues analysed (Figure 4.4 A). To normalize the values we calculated the ratios of SF2/ASF and TDP43 mRNAs to GAPDH levels.

A



B



**Figure 4.4.** A) Northern blot analysis of TDP43 (upper panels), SF2/ASF (middle panels) and GAPDH (lower panels) performed on mRNA extracted from 16 different human tissues. The arrows indicate the major 2.8 kb mRNA species characteristic of TDP43 and the 3.0 kb mRNA of SF2/ASF. B) Graphical representation of the normalised TDP43 mRNA levels (black boxes) and SF2/ASF levels (shaded boxes).

The quantitation indicates that the relative level of expression of SF2/ASF and TDP43 vary considerably among different tissue (Figure 4.4 B). Interestingly, the higher expression levels were found in some of the tissues affected in CF. In fact pancreas, reproductive tissues and lung showed very high expression of these two negative regulators of CFTR exon 9 inclusion, suggesting their role in phenotypic expression.

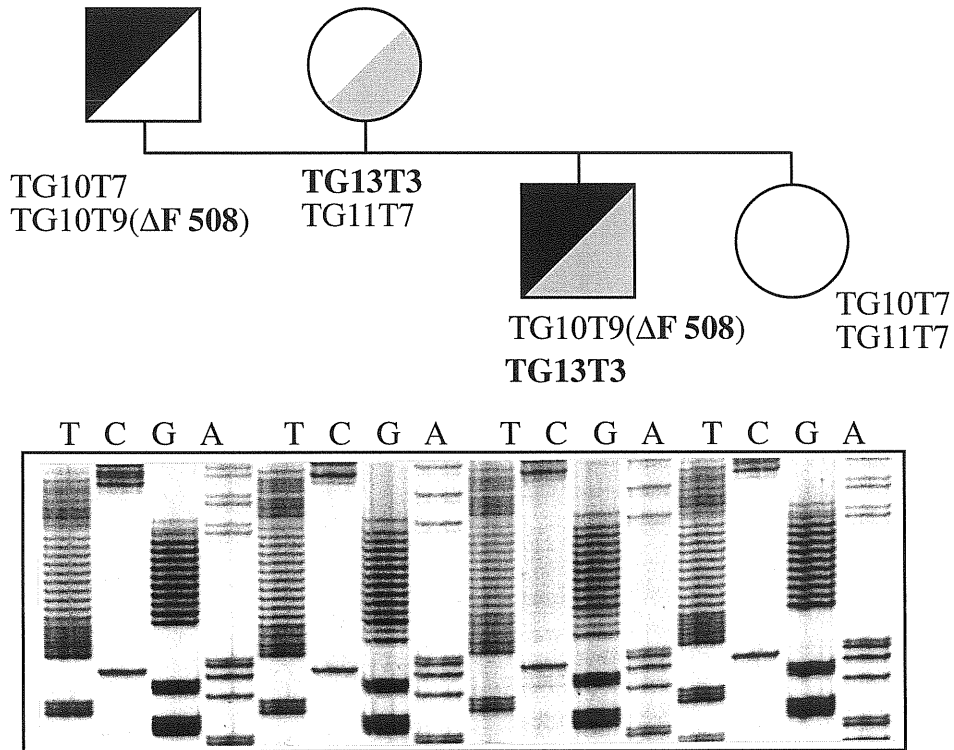
#### **4.4 In vivo Significance of the Extreme TG13T3 Polymorphism**

The (TG)<sub>m</sub>(T)<sub>n</sub> polymorphic locus at 3' end of intron 9, and in particular the T5 allele, has been associated with variable penetrance of monosymptomatic forms of CF. In figure 4.1 B we have evaluated the effect of TDP43 and SF2/ASF overexpression on the TG13T3 minigene, a CF variant found during a mutation screening which lead to a high proportion of exon skipping .

The patient carrying this particular allele is a 19-year-old male who suffered from a mild CF characterized by pulmonary symptoms with recurrent infections, elevated sweat chloride concentrations but with pancreatic sufficiency. The CFTR gene sequence of this patient shows on one allele the most frequent CFTR mutation,  $\Delta$ F508, and on the other allele TG13T3(wt) element. The  $\Delta$ F508 was in *cis* with TG10T9( $\Delta$ F508) (Figure 4.5 A). The affected patient has inherited the TG10T9( $\Delta$ F508) allele from the father and the TG13T3(wt) variant from the mother. The sister did not carry either the TG10T9( $\Delta$ F508) or the TG13T3(wt) allele. The TG13T3(wt) allele represents a rare mutational event, since it was not observed in 100 non-CF chromosomes.



A

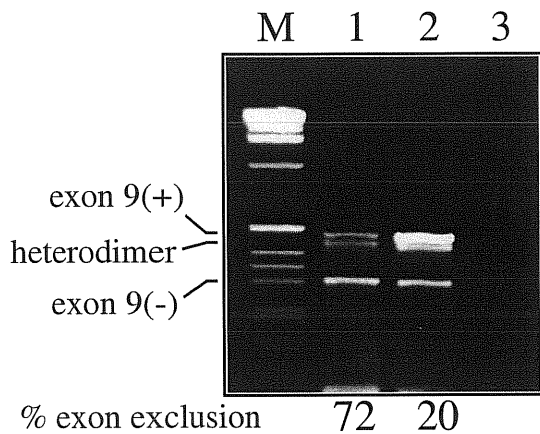
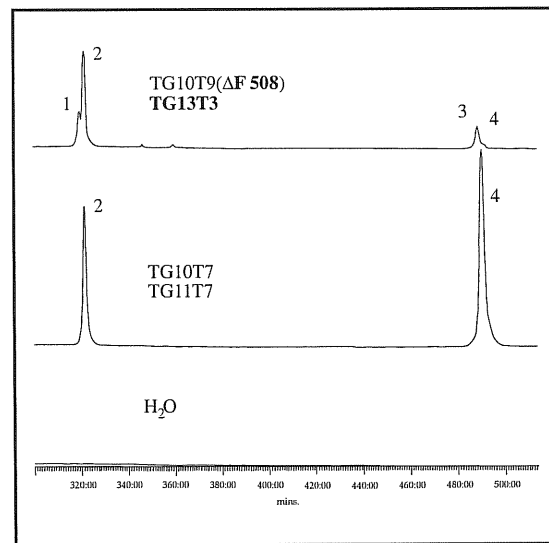


**Figure 4.5. A)** A family tree and a sequencing analysis of the pancreatic-sufficient CF patient carrying the TG13T3 allele on one chromosome and the TG10T9( $\Delta$ F508) configuration on the other chromosome

To evaluate the efficiency of exon 9 splicing the Epstein-Barr virus (EBV) transformed lymphocytes derived from the patient were analysed by RT-PCR using primers in exon 8 and 11 (Figure 4.5 B). Two PCR species were found on agarose gel electrophoresis corresponding to inclusion or exclusion of exon 9. We took the advantage of the presence of the  $\Delta F508$  mutation, which produces a 3 bp shorted transcript, to differentiate the TG10T9( $\Delta F508$ ) allele from the TG13T3(wt) allele.

Analysis of the RT-PCR products, separated on a denaturing polyacrilamide gel, showed in the patient the presence of four transcripts TG10T9( $\Delta F508$ ) 9(-), TG13T3 9(-), TG10T9( $\Delta F508$ ) 9(+), and TG13T3 9(+) (Fig 4.5C, peaks 1 to 4). Semi quantitative analysis showed that the proportions of these four species were 11:61:24:4, indicating that the TG13T3 allele in the heterozygous patient allows for some 6% normal splicing in his lymphoblasts. This result is entirely consistent with the results from transfection experiments (Figure 4.5 C).

This result highlight how the amount of exon 9 skipping observed with the naturally occurring TG13T3 variant is the highest thus far reported and this explain the CF phenotype of this patient.

**B****C**

**Figure 4.5. B)** RT-PCR products spanning exons 8-11 of the CFTR cDNA, obtained from the CF patient compound heterozygous for the TG13T3 mutation and the TG10T9( $\Delta$ F508) allele (lane 1) and from a control individual compound heterozygous for the TG11T7 and TG10T& alleles (lane 2), separated on a 2% agarose gel. The percentage of exon 9 exclusion in each CFTR transcript for the two alleles, as determined after denaturing PAGE separation, is given below. Lane 3, no template; M, 1kb marker. **C)** A semiquantitative analysis of exon 9 skipping. The RT-PCR products from the CF patient (upper profile), a control individual (middle profile) and a negative control (lower profile), were separated on a denaturing polyacrylamide gel using an ALF sequence. Fluorescence signals were quantified using the Fragment Manager software. Peaks 1 and 3 correspond to the amplified 9(-) and 9(+) fragments from the TG10T9( $\Delta$ F508) allele, whilst peaks 2 and 4 correspond to amplified 9(-) and 9(+) fragments from non- $\Delta$ F508 alleles.

## RESULT Part 2: CFTR intron 9 contains other splicing regulatory elements than the ISS

Our previous molecular studies have shown that more than one player determines the pattern of CFTR exon 9 alternative splicing. In fact we have identified the following cis-acting elements involved in this modulation: the polymorphic (TG)<sub>m</sub>(T)<sub>n</sub> variant in the intron 8, the enhancer and silencer elements in the exon 9 and the ISS in the intron 9. On the other hand, the trans-acting factors that interact with these elements in the pre-mRNA are the TDP43 protein, SF2/ASF and probably other members of the SR proteins family (152, 185).

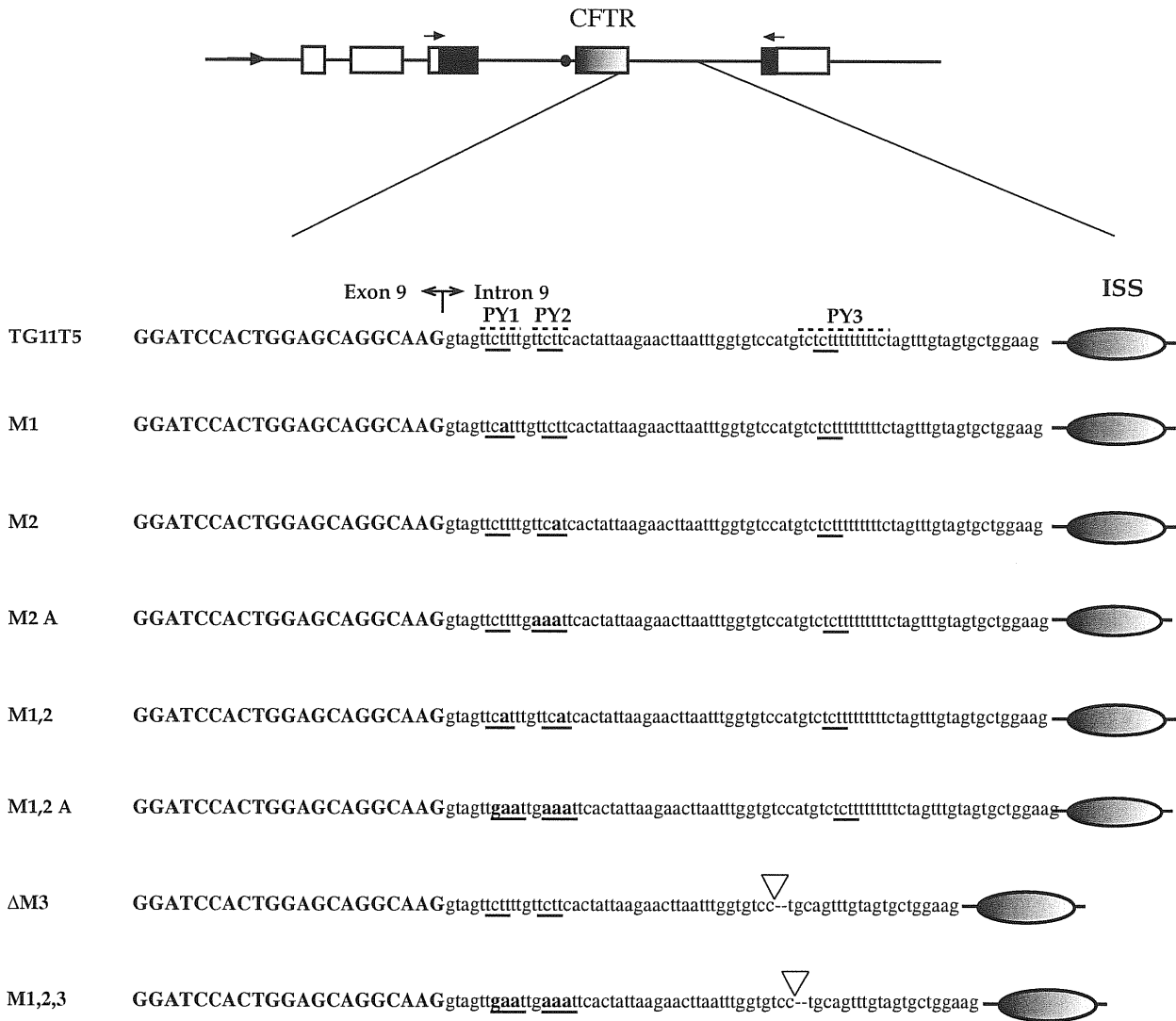
For the recruitment of the spliceosomal components not only the 5' splice site, 3' splice site and the branch point sequence are important. In fact, as described in the introduction, other sequences close to the junctions or in the exon are involved. At 3' end of CFTR intron 8 we identified a particular arrangement at the polymorphic locus that is recognized by TDP43 and is involved in exon 9 splicing modulation.

The sequences between the 5' end of intron 9 and the ISS have not been previously studied and their role in the regulation of exon 9 alternative splicing is not known. This region has some interesting structural features. In particular, a peculiar arrangement of polypyrimidine rich elements. Uridine stretches are known to be a common target of RNA-binding proteins; among these we found hnRNP C (194), Sex-lethal protein of *Drosophila* (195, 196), the essential splicing factor U2AF65 (79) and PTB (197, 198).

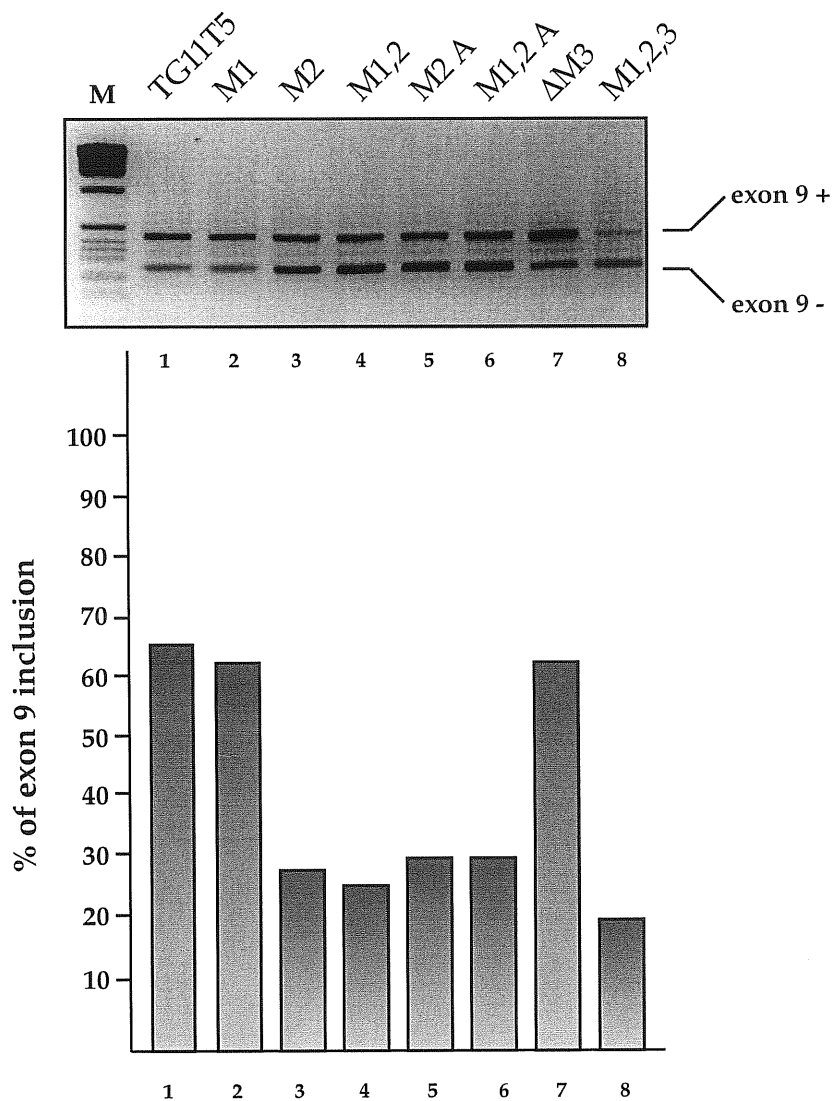
Due to the fact that several proteins involved in splicing recognise pyrimidine rich elements, we decided to study if these elements located at the 5' end of CFTR intron 9 are also involved in the regulation of exon 9 alternative splicing.

## 5.1 Intron 9 polypyrimidine rich elements between the 5' splice site and the ISS modulate splicing efficiency

Between the 5' splice site and the ISS element close to the exon 9/ intron 9 junction, there are three polypyrimidine rich elements that we named PY1, PY2 and PY3 (Figure 5.1). The two adjacent PY1 and PY2 sites are located near the 5' splice site, spanning the nucleotides from 6 to 17 of intron 9. PY3 is a 15 bp-long pyrimidine tract located about 50 bases downstream the splice site. These polypyrimidine-rich elements contain TCTT motifs, which have been characterised as PTB core recognition sequence (163). To investigate their role in mediating the CFTR exon 9 recognition we prepared minigene constructs in which we disrupted these polypyrimidine-rich sequences. This was obtained introducing mutations or deletions in these sites either alone or in combination. The TCTT sequence of PY1 and PY2 was changed to TCAT and in some cases to GAAT or AAAT, respectively. The 15 bp-long pyrimidine tract of PY3 was deleted (Figure 5.1). The minigenes with the mutations in the intron 9 pyrimidine rich elements (Table 5.1) were transfected in Hep3B cells and the pattern of splicing analysed with specific primers. Wild type TG11T5 showed 65% of exon 9 inclusion and a similar level of exon inclusion (63%) was obtained using M1 (Figure 5.2, lanes 1 and 2 respectively) suggesting that PY1 is not involved in the regulation of CFTR exon 9 alternative splicing. On the contrary, deletion of the PY2 motif (M2 and M2A) showed a reduced inclusion of exon 9 in comparison to the WT minigene. In fact, the substitution of an A (M2 minigene, TCTT⇒TCAT) or the mutation of three bases (M2A minigene, TCTT⇒GAAT) showed 26% and 29% of exon inclusion, respectively (Figure 5.2, lane 3 and 5).



**Figure 5.1** Schematic representation of mutant minigene construct intron 9. The sequence starts at the Bam HI site in the exon 9. Exon/intron junction is indicated by opposite arrows. PTB-binding sites are underlined and the mutations in each construct are in bold. PY1, PY2 and PY3 (dashed lines) indicate respectively the polypyrimidine-rich elements. The ISS is schematised by an ellipse. An inverted triangle represents the deletion of 15 nucleotides in PY3 element.



**Figure 5.2** RT-PCR assay on transfection experiments using minigene constructs. The minigenes analysed in the experiments present mutations on the polypyrimidine-rich motifs in the intron 9. Transfection experiments were carried out using 3  $\mu\text{g}$  of each minigene. The samples were loaded on a 1.5% agarose gel (Upper panel). In the lower panel a graphical representation of the percentage of exon 9 inclusion obtained using the different minigene construct. The RT-PCR products were loaded on a 8% denaturing polyacrilamide gel and the percentage of the two splicing products were quantitated using a Cyclone.

Mutations in both PY1 and PY2 motifs resulted also in reduced amount of exon inclusion. In fact, the double mutants M1,2 and M1,2A gave respectively 22% and 28% of exon 9 inclusion (Figure 5.2, lane 4 and 6).

CONSTRUCT	PY1	PY2	PY3
hTG11T5	TCTT	TCTT	Normal
M1	TCAT	TCTT	Normal
M2	TCTT	TCAT	Normal
M2A	TCTT	<b>AAAT</b>	Normal
M1,2	TCAT	TCAT	Normal
M1,2A	<b>GAAT</b>	<b>AAAT</b>	Normal
$\Delta$ M3	TCTT	TCTT	<b>Deleted</b>
M1,2,3	<b>GAAT</b>	<b>AAAT</b>	<b>Deleted</b>

Table 5.1 Schematic representation of the mutations in the three polypyrimidine-rich regions located within the first 77 bp of intron 9. Mutated bases are in bold. In regard to the PY3 motive a deletion of 15 base pairs has been done.

The deletion of 15 base pairs in the PY3 motif ( $\Delta$ M3) did not significantly affect the pattern of alternative splicing (62% of exon 9 included) (Figure 5.2, lane 7). Deletion of all the PY1, PY2 and PY3 motifs (M1,2,3) showed the lowest inclusion level of exon 9 with about 20% of exon included (Figure 5.2, lane 8).

These data suggest that the polypyrimidine regions in intron 9 are involved in the modulation of CFTR exon 9 alternative splicing. In particular, the PY2 element seems to behave as an Intronic Splicing Enhancer (ISE); in fact, the mutations introduced in this element caused a significant reduction in the exon 9+ form.



## 5.2 *In vitro* Binding Experiments Shown the Ability of PTB to Interact to CFTR Intron 9

The results presented in the previous section showed that the polypyrimidine rich elements in intron 9 are involved in the regulation of CFTR exon 9 alternative splicing. As previously reported, PTB is able to bind to pyrimidine rich motif and induce exon skipping, so we have tested the ability of PTB to interact with these elements present in intron 9. To study this hypothesis we performed *in vitro* binding experiments using HeLa nuclear extract.

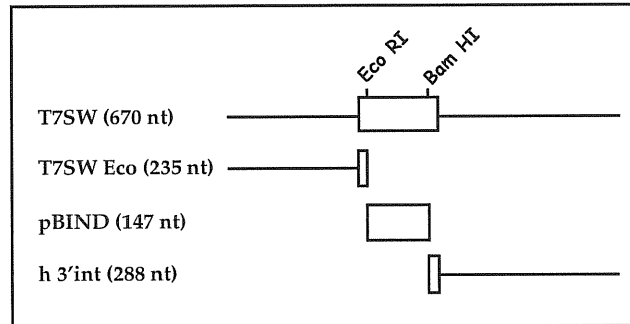
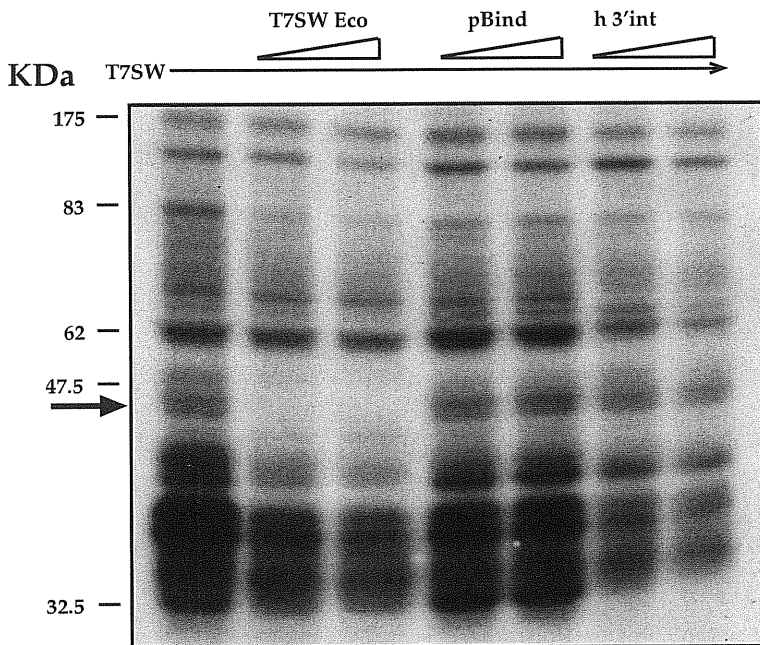
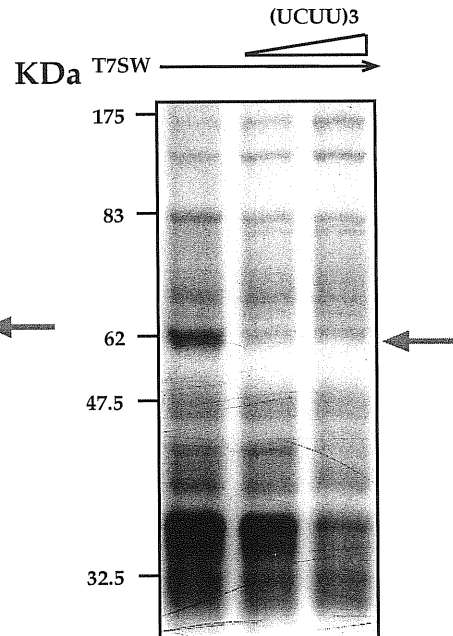
### 5.2.1 Study of the RNA-Protein Interaction in CFTR Sequence

To study the proteins that potentially could interact with CFTR intron 9 sequences we performed UV cross-linking assays using HeLa nuclear extract. For each experiment we incubated the radiolabeled probe alone or with one of the cold competitor in the presence of 30  $\mu$ g of HeLa nuclear extract.

The entire intron 8/exon 9/intron 9 sequence (670 base pairs) was cloned under the control of T7 promoter and *in vitro* transcribed in the presence of [<sup>32</sup>P]UTP to obtain the radiolabeled probe T7SW (Figure 5.3 A). The UV cross-linked proteins to this probe were competed with increasing amounts of the following cold competitors: the intron 8 probe Eco of 235 nucleotides (T7SW), the exon 9 probe of 147 nucleotides (pBIND) and the intron 9 probe of 288 nucleotides (h3'int) (Figure 5.3 A). Also these sequences were cloned under the control of T7 promoter and *in vitro* transcribed in the presence of non-labeled nucleotides. Among the different UV cross-linked proteins by the T7SW RNA we observed a specific and reproducible competition of a band of molecular weight of approximately 60 KDa in the presence of h3'int

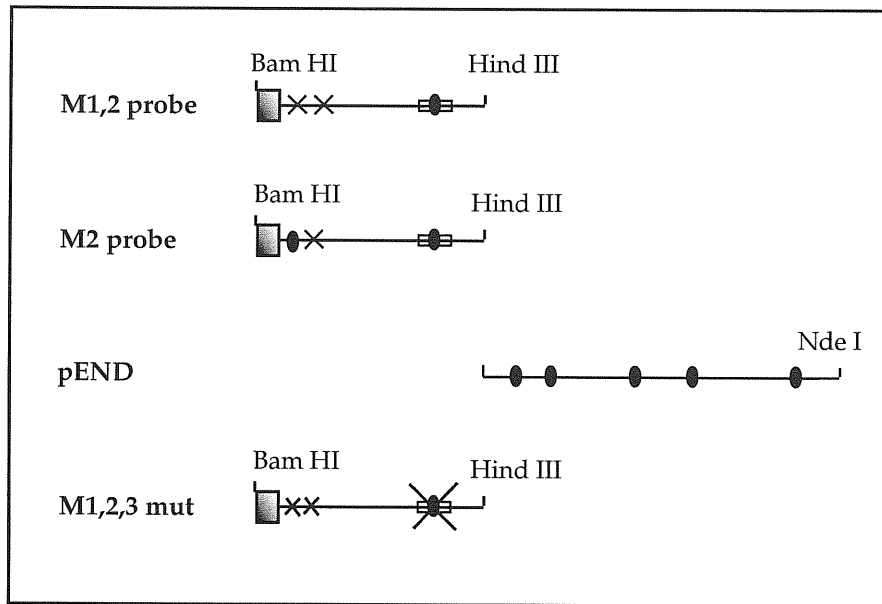
competitor (Figure 5.3 B, right arrow). In some experiments, due to a better electrophoretic resolution, the 60 KDa complex appeared like a doublet (see experiment in figure 5.4, B). Strikingly, this 60 KDa doublet was also competed by (UCUU)<sub>3</sub> cold RNA, a result which suggested that it contains PTB. In addition, the competition performed with intron 8 (T7SW Eco probe) showed the disappearance of a band of about 50 KDa (Figure 5.3 B, left arrow). This band was previously shown to be TDP43. This confirms our previous result on the binding of this protein to UG polymorphic region at 3' end of intron 8 (185).

To better characterise the binding property of the 60 KDa band in intron 9 we performed additional UV cross-linking experiments. The entire intron 9 sequences were used as a probe (h 3' int) and the resulting UV cross-linked pattern competed with different amount of sequences correspondent to portion of the same intron. The competitors used are shown in (Figure 5.4, A). The M2, M1,2 and M1,2,3 mut correspond to the first 77 base pairs of intron 9: M2 contains a TCTT⇒AAAT substitution in the second polypyrimidine rich element, in M1,2 both the first and second elements were mutated as follow TCTT⇒GAAT/TCTT⇒AAAT, and in the M1,2,3 mut, all the three polypyrimidine elements were disrupted. We also tested a pEND cold competitor, which covers the region from nucleotide +77 from exon 9/intron 9 junction until the NdeI site in intron 9 and it is approximately 190 base pairs long (Figure 5.4 A). The results of these competition assays showed that only M1,2,3mut did not compete for the binding of the UV cross-linked protein with apparent molecular weight of 60 KDa. All the other competitors tested, which conserved at least one pyrimidine rich element, were able to compete the 60 KDa band (Figure 5.4 B-C).

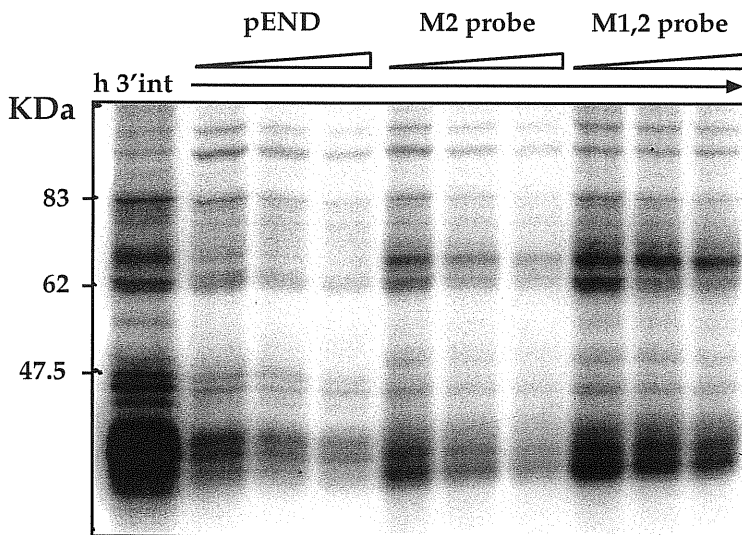
**A****HUMAN PROBES****B****C**

**Figure 5.3** UV cross-linking assay. **A)** A schematic representation of human CFTR sequences cloned under the control of T7 promoter. The T7SW probe was in vitro transcribed in the presence of [<sup>32</sup>P]UTP. T7SW Eco, pBind and h3'int RNAs were in vitro transcribed in the absence of labelled nucleotides and used as cold competitors. **B)** Increasing amounts of competitor were added to the reaction. In the first competition were used the same amount of labelled probe and cold competitor, in the second competition a double amount of cold competitor in respect to the labelled probe were used. The red arrow is consistent with the PTB doublet. These bands are competed away when cold h3'int RNA is used. **C)** The 60 KDa doublet is competed also by (UCUU)<sub>3</sub> RNA which has been shown to be the typical PTB-binding motif. The black arrow indicates the competition of the TDP43 band by T7SW Eco cold probe. Molecular weight markers are on the left of the panel.

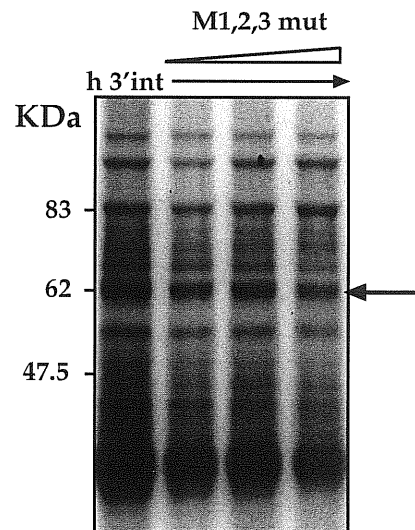
A



B



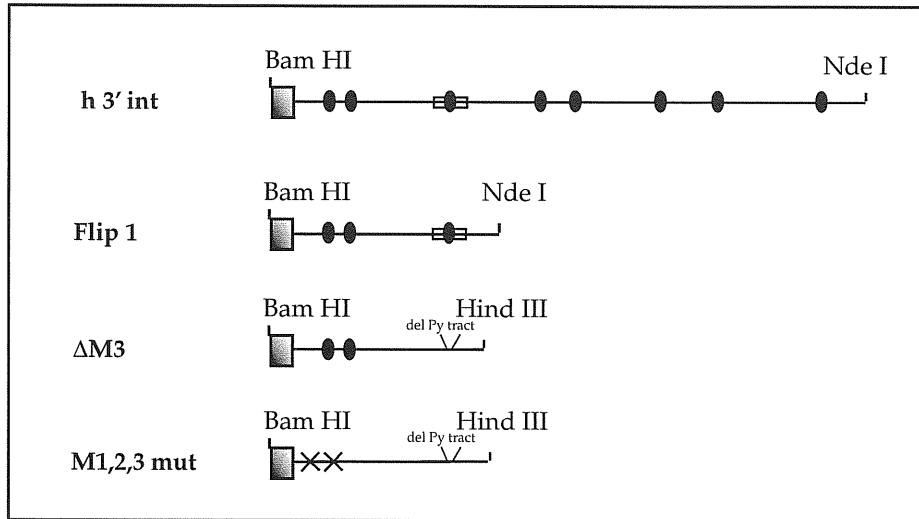
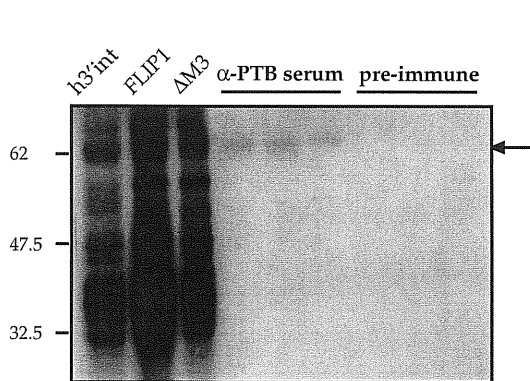
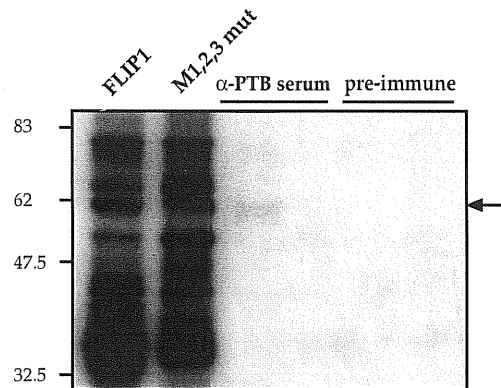
C



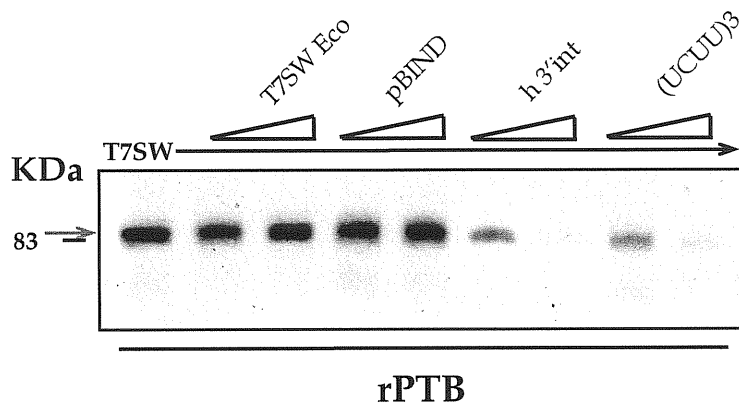
**Figure 5.4** UV cross-linking assays on CFTR intron 9. **A)** Schematic representation of the probes used as cold competitors. M2 probe, M1,2 probe and M1,2,3 mut probe covered the first 77 nucleotides of intron 9. pEND probe is approximately 190 bp long and spanned the region downstream the first 77 nucleotides. **B)** Competitive UV cross-linking assay using as radiolabeled RNA the h3'int. Three increasing amounts of each cold RNA were used to compete the binding to h3'int. The three RNAs used in this panel are pEND, M2 probe and M1,2 probe. All these probes were able to compete the binding of the protein of about 60 KDa. The arrow indicates the band corresponding to this protein. **C)** Competitive UV cross-linking assay using as radiolabeled probe h3'int and as cold competitor M1,2,3 mut probe. In this case the competitor was not able to compete the binding of the protein of about 60 KDa. All the samples were loaded on a 10% SDS-PAGE gel. Molecular weight marker is reported on the left of each panel.

## 5.2.2 Immunoprecipitation assays show that CFTR intron 9 binds PTB

The apparent molecular weight of the 60 KDa UV cross-linked band (Figure 5.3 B) and the presence of UCUU core sequence in the pyrimidine rich elements in intron 9 suggest that PTB might be involved in the regulation of CFTR exon 9. To study protein binding in intron 9, we performed immuno-precipitation (IP) experiments on the UV cross-linked reactions using an antibody specific for PTB. We prepared four different radiolabeled probes: the entire intron 9 fragment (h 3'int), the Flip 1 probe that correspond to the first 77 nucleotides of intron, the  $\Delta$ M3 probe that differs from the previous one for the deletion PY3 element and the last probe is M1,2,3 mut, in which PY1, PY2 and also PY3 were disrupted (Figure 5.5 A). These sequences, cloned under the control of T7 promoter, were *in vitro* transcribed in the presence of [<sup>32</sup>P]UTP. After UV cross-linking with HeLa nuclear extract we performed IP experiments using an anti-PTB anti-serum and its pre-immunised serum as control. As shown in figure 5.5 B, a band of 57 KDa, at the same electrophoretic mobility of the competed band in the UV cross-linking assays, can be specifically immuno-precipitated by the anti-PTB antiserum following the UV cross-linking of three out of four of these probes. In fact, anti-PTB antiserum was unable to immunoprecipitate PTB using M1,2,3 mut probe, where the three polypyrimidine rich elements were disrupted (Figure 5.5, C). This suggests that PTB could not bind to the first 77 nucleotides of intron 9 when all the pyrimidine-rich elements (PY1, PY2 and PY3) are disrupted.

**A****B****C**

**Figure 5.5** Immuno-precipitation experiments using  $\alpha$ -PTB serum. **A)** Schematic representation of the CFTR intron 9 sequences used as radio-labeled probes in the IP experiments. **B)** The radio-labeled probes were UV cross-linked with 30  $\mu$ g of HeLa nuclear extract, this assay were done in triplicate for each probe. One aliquote was immuno-precipitated using  $\alpha$ -PTB serum, the other one using pre-immune serum as control. The samples were loaded on a 10% SDS-PAGE gel. h 3'int probe, Flip 1 probe and  $\Delta$ M3 probe were all able to immuno-precipitate PTB (shown by the arrow). No PTB corresponding band were immunoprecipitated by the pre-immune serum. **C)** The same experiment reported in B) were repeated using h 3'int probe and 3bs MUT probe. This last probe does not contain the three polypyrimidine rich region PY1, PY2 and PY3. As shown in the panel M1,2,3 mut RNA is not able to immuno-precipitate PTB. Molecular markers are on the left side of each panel.



**Figure 5.6** UV cross-linking experiment with recombinant PTB (rPTB). The binding of rPTB to T7SW labeled RNA is competed with the cold RNAs schematised in figure 5.3 A. (UCUU)<sub>3</sub> probe is used as positive control of PTB binding as it represent the core optimal PTB-binding sequence. The red arrow indicates the binding of rPTB. The amount of each cold competitor is the same of the labeled T7SW in the first lane and double in the second lane of each competition experiment. The samples are loaded on a 10% SDS-PAGE gel. Molecular weight marker is on the left.

### **5.2.3 Recombinant PTB Interacts with CFTR Intron 9 Sequence in UV Cross-Linking Experiments**

To validate the results obtained from the IP assays, we prepared recombinant GST-PTB fusion protein and we performed binding experiments. These studies were done using UV cross-linking assays with the probes used in the experiments described in the section 5.2.1. The (UCUU)<sub>3</sub> RNA probe is used as positive control as it represents the optimal PTB-binding sequence. The binding of recombinant PTB (0.6 µg each UV cross-linking experiment) on the CFTR sequence (T7SW) was competed only by the h3'int cold competitor (Figure 5.6). This result confirms the ability of PTB to interact to CFTR intron 9.

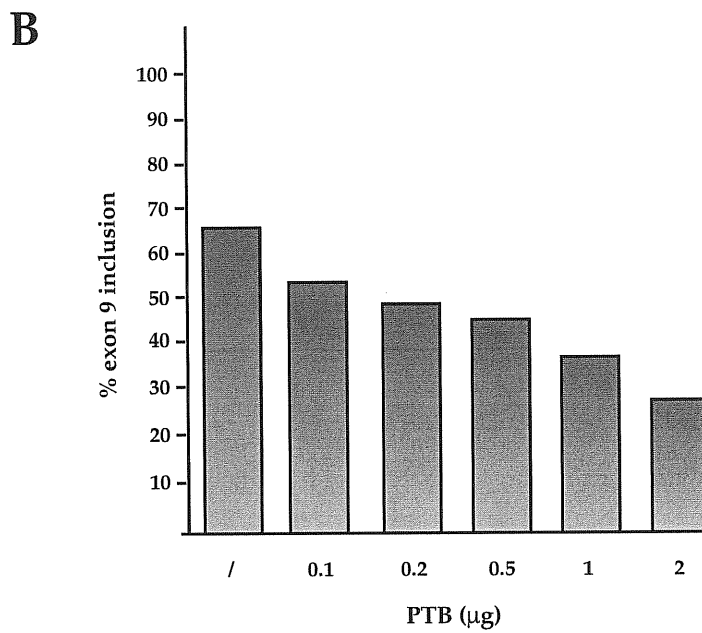
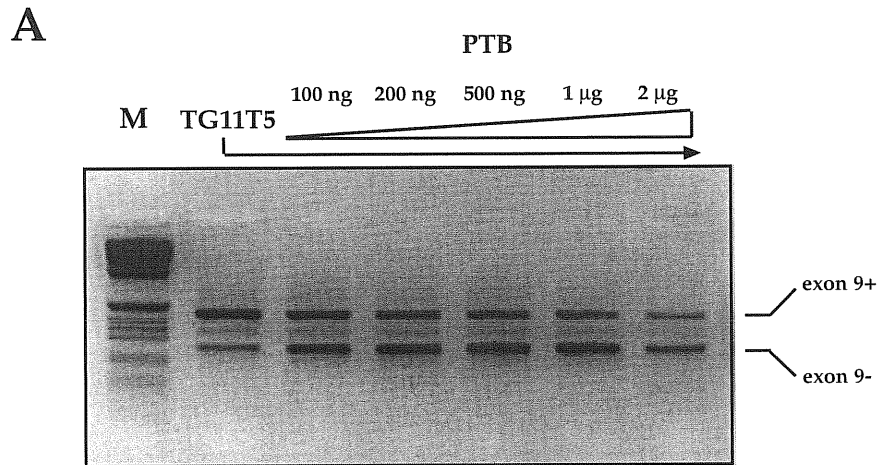
### **5.3 Study of the functional effect of PTB on CFTR exon 9 alternative splicing**

The results described in the previous section suggest that PTB can interact with CFTR intron 9 sequences. To test the functional effect of these interactions we performed minigene transient transfection assay in Hep3B cell line.

#### **5.3.1 The splicing factor PTB induces CFTR exon 9 skipping**

To study of the effect of PTB on human CFTR exon 9 alternative splicing we performed transient transfection experiments on Hep3B cells using the WT CFTR exon 9 construct, the TG11T5 hybrid minigene. This variant has about 65% of CFTR exon 9 inclusion (see figure 5.2, lane 1). We cotransfected this minigene with increasing amounts of PTB expressing vector (from 0.1 µg to 2 µg) and the splicing pattern was





**Figure 5.7** Effect of PTB overexpression on the alternative splicing of CFTR exon 9. **A)** Minigene TG11T5 variant (3  $\mu$ g) was cotransfected along with increasing amounts of PTB expressing plasmid (from 0.1  $\mu$ g to 2  $\mu$ g) in Heb 3B cell line. RT-PCR products were analysed in a 1.5% agarose gel. As negative control pTracer/CMV2 (3  $\mu$ g) is cotransfected together with TG11T5 minigene variant (lane 1). M, is the molecular weight 1kb marker. **B)** Graphical representation of exon 9 inclusion when PTB is overexpressed.

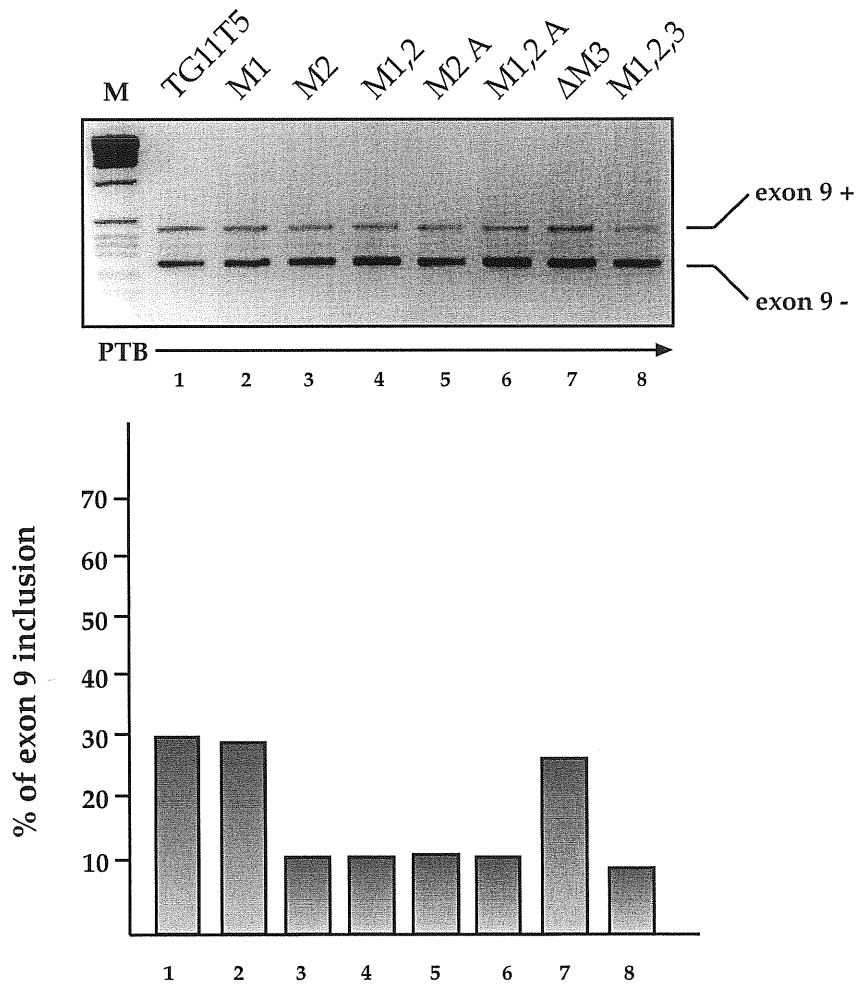
analysed by RT-PCR amplification with specific primers. The results showed that increasing the amount of PTB expressing plasmid in transfection experiments resulted in a progressive reduction in the CFTR exon 9 inclusion, following a dose response curve. The 65% percentage of exon 9 inclusion in WT CFTR became 53%, 30% and 28% using 0.1  $\mu$ g 1  $\mu$ g and 2  $\mu$ g of PTB coding plasmid, respectively (Figure 5.7 A and B). These results show for the first time that PTB has a negative effect on CFTR exon 9 alternative splicing.

### **5.3.2 The effect of PTB overexpression on the minigenes with the mutated polypyrimidine elements**

We then want to study the effect of PTB overexpression in transfection experiments using the mutated constructs at the polypyrimidine-rich elements.

Overexpression of PTB along with the wild type TG11T5 minigene resulted in a reduction of exon 9 inclusion from 65% to a level of 30% (Figure 5.8, lane 1). This splicing inhibition is similar in M1 and  $\Delta$ M3 which, from a basal level of 63%, reached a value of exon 9 inclusion of 28% and 27% respectively (Figure 5.8, lanes 2 and 7). A significant decrease in the level of exon 9 inclusion was observed when M2, M1,2, M2A, M1,2A and M1,2,3 were cotransfected along with PTB expressing vector. In fact, the disruption of the PY2 element either alone (M2 and M2 A minigene constructs) or in combination with in PY1 or PY3 (M1,2, M1,2A and M1,2,3 minigene constructs) led to a strong splicing inhibition that is lowest (8%) when all the pyrimidine-rich elements were disrupted (Figure 5.8, lanes 3-6,8).

These data indicate that PTB overexpression induces a reduction of the level of exon 9+ form in all the normal and mutated constructs analysed independently from the mutations introduced in the polypyrimidine-



**Figure 5.8** RT-PCR assay on co-transfection experiments of minigene constructs and PTB expressing plasmid. The minigenes analysed in the experiments present mutations on the polypyrimidin-rich motifs in the intron 9. Transfection experiments were carried out using 3  $\mu$ g of each minigene and 1  $\mu$ g of PTB expressing plasmid. The samples were loaded on a 1.5% agarose gel (Upper panel). In the lower panel is shown a graphical representation of the percentage of exon 9 inclusion obtained using the different minigene construct. The RT-PCR products were loaded on a 8% denaturing polyacrilamide gel and the percentage of the two splicing products were quantitated using a Cyclone.

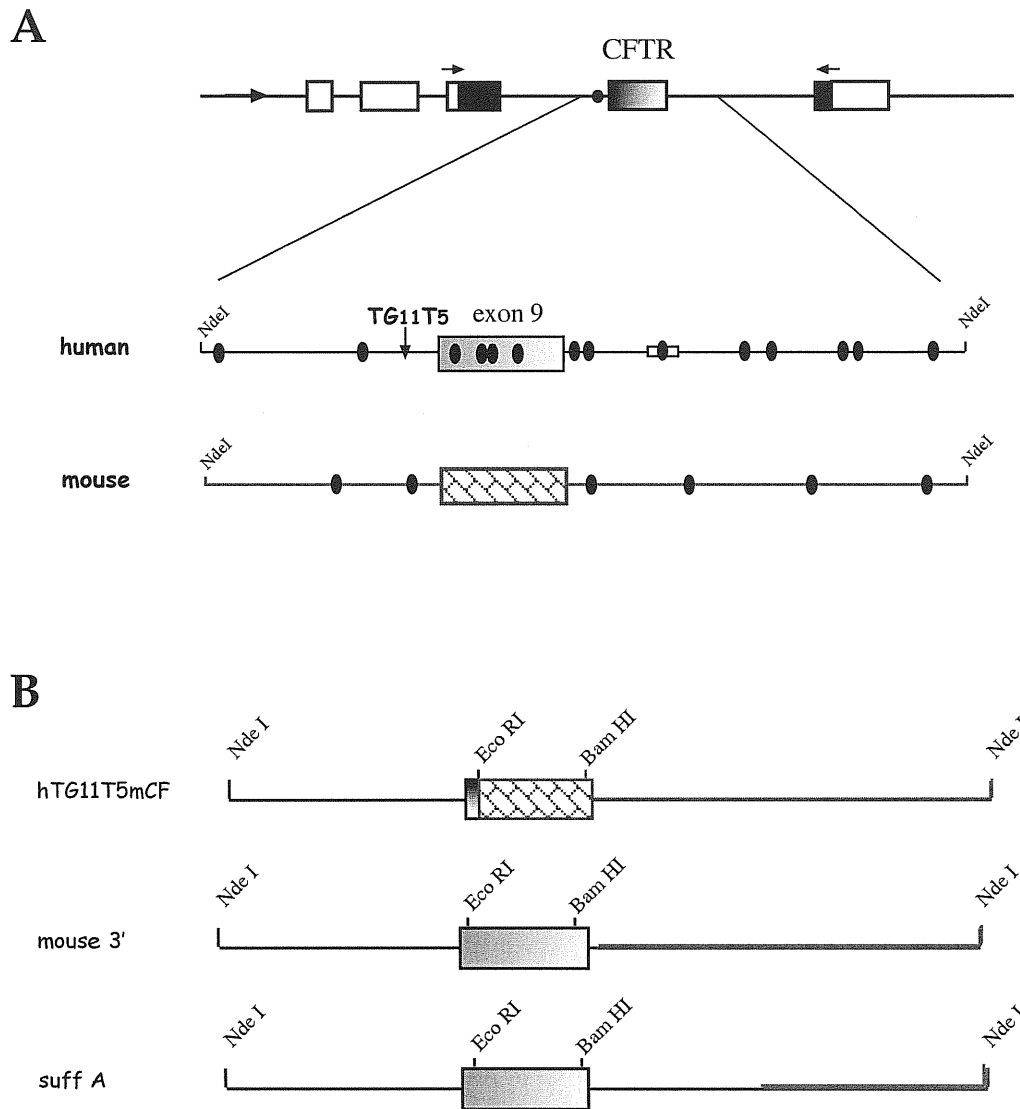
rich elements located between the 5' splice site and the ISS element. PTB binding to both the polypyrimidine-rich elements close to the junction of CFTR exon 9 was shown in one *in vitro* experiment (Figure 5.4, B). However, the functional studies do not seem to confirm a functional role for this binding. Taken together, these results suggest that indeed PTB has a functional effect on the modulation of CFTR exon 9 inclusion, but this does not seem to be directly related to the ability of the protein to bind to the intron 9 region closer to the 5' splice site junction.

### **5.3.3 Hybrid human/mouse minigene confirmed that the modulation of exon 9 inclusion by PTB required other elements than PY1, PY2 and PY3**

To test the possibility that the ISS region could modulate the splicing inhibition induced by PTB, we studied the effect of PTB overexpression on hybrid human/mouse constructs.

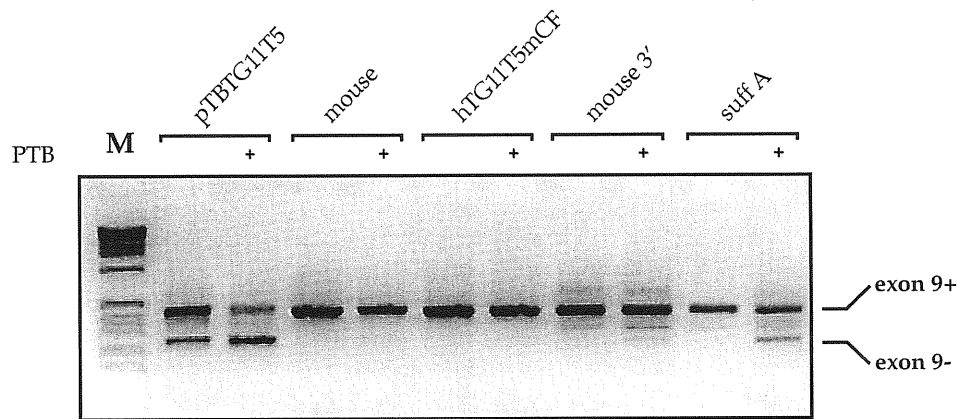
As previously reported, the mouse gene is significantly different from the human one within the flanking introns of exon 9. Both the TG and T repeats are absent at the intron 8/exon 9 junction, and in addition there are substantial sequence differences in intron 9, including the ISS element binding SR proteins (152, 184). We have analysed a hybrid minigene containing the mouse CFTR exon 9 and its flanking sequences for the response to PTB (Figure 5.9 A). Contrary to the human counterpart, the mouse CFTR exon 9 was not negatively regulated by PTB (Figure 5.9 B), indicating that specific sequences present in the human DNA may mediate the PTB response. To confirm that PTB effect is mediated by human intron 9 sequences, we prepared hybrid minigene constructs with parts of the human homologue inserted in the mouse context (Figure 5.10). These human/mouse hybrids were transiently transfected in Hep3B cell line along with PTB expressing

plasmid. The human intron 8 with the variant TG11T5 at the polymorphic locus (Figure 5.9 B) and the human exon 9 (Figure 5.9 B) did not confer a PTB mediated inhibitory effect to the mouse exon, indicating that the human sequences present in intron 8, including the polymorphic TG and T repeats, and the exon itself are not the target of this splicing factor. When the first 77 nucleotides of intron 9 were substituted to the corresponding intronic sequences of mouse, overexpression of PTB induced exon 9 skipping (Figure 10). However, the PTB-mediated splicing inhibition in the Suff A minigene is significantly lower than the effect of the same protein on the wild type TG11T5 minigene that includes the ISS element. Indeed, PTB can interact also with the ISS element (see section 5.2.1, pEND competitor) in *in vitro* binding assays and this may be due to the presence of several TCTT motifs. This data suggest that multiple elements located in the entire CFTR intron 9 we have analysed are involved in the modulation of splicing efficiency via PTB. Moreover, the fact that Flip 1 (which contains only PY1, PY2 and PY3), Suff A (in which the ISS was replaced by mouse sequence) and M1,2,3 (with the polypyrimidine-rich elements disrupted but with the ISS), respond to PTB suggest that cooperative binding of PTB on different elements distributed in intron 9 is responsible for the observed splicing inhibition.



**Figure 5.9** Hybrid mouse/human minigene constructs. **A)** Schematic view of the Nde/Nde CFTR sequence cloned in the minigene. Human sequence is in black and mouse sequence in red. Dots represent TCTT sequence (optimal PTB binding sites). The open box in the human intron 9 represent a ransfection experiments. **B)** Hybrid human/mouse minigene constructs. In this panel a less detailed pictures are used, but the elements present in the mouse and human sequences (the PTB-binding sites and the polypyrimidine-rich tract) directly correspond to that on panel A).

C



**Figure 5.10** Cotransfection experiments of hybrid human/mouse minigene constructs (3  $\mu$ g) along with PTB expressing plasmid (1  $\mu$ g). The RT-PCR products were analysed on a 1.5% agarose gel. PTB overexpression affect TG11T5 minigene variant alternative splicing leading to a 35% exon 9 inclusion (lane 2). No exon 9- form can be detected using either mouse minigene construct (lanes 3 and 4) or the first two human/mouse hybrid constructs (lanes 5-8). PTB negative effect is rescued only when the first 80 bp of human intron 9 replace the mouse sequence (lanes 9-10). M. molecular weight marker 1 kb.

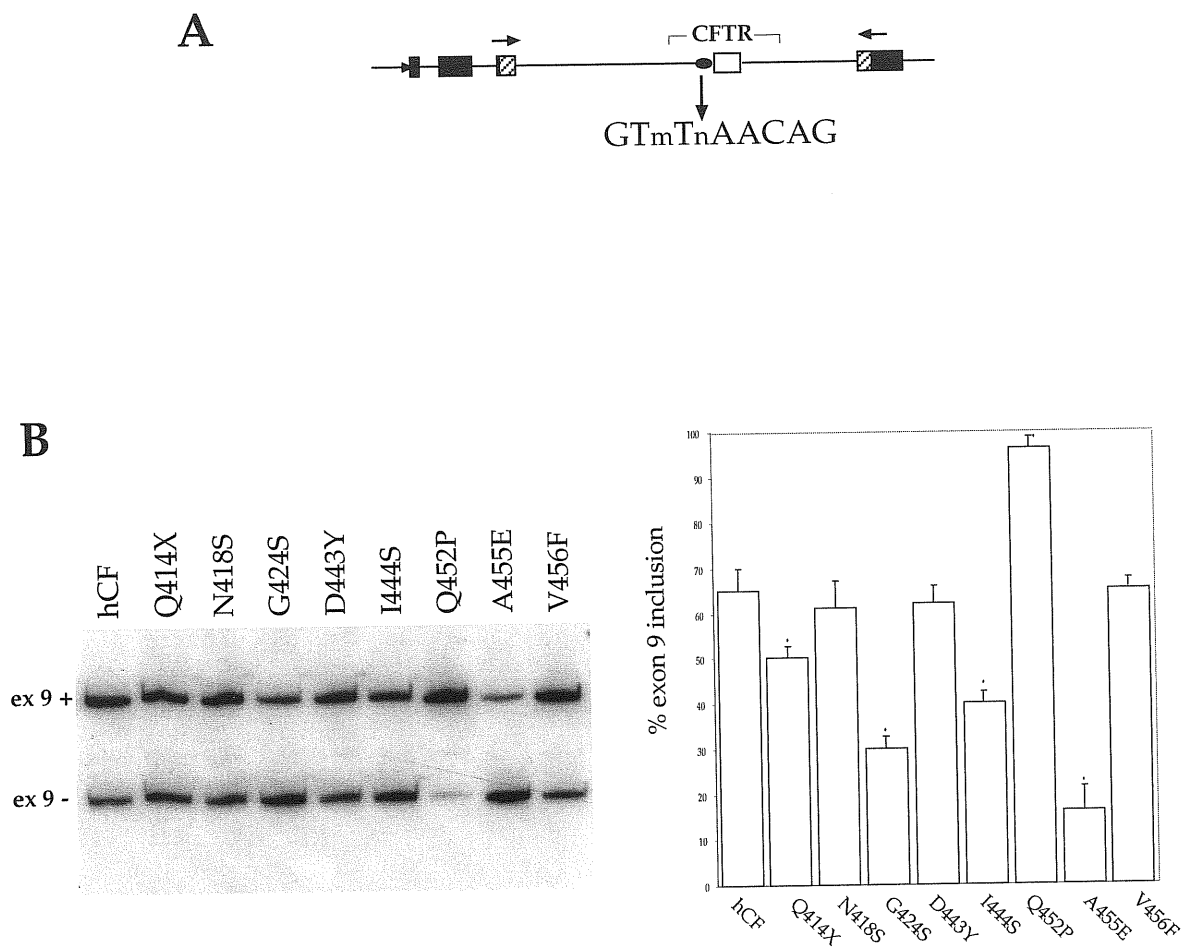
### **Part 3: CFTR exon 9 missense and nonsense substitutions can affect the splicing efficiency**

We have previously shown that deletions of discrete regions of CFTR exon 9 can result in both increase or reduction of the splicing efficiency indicating the presence in this exon of sequences with both enhancer and silencer activity (152). However, we considered that multiple base deletions might cause gross alterations in the recognition of the whole exon different from single nucleotide substitutions causing missense or nonsense mutations that can affect the splicing pattern in a more subtle way. To evaluate the contribution of exonic elements in the regulation of CFTR exon 9 alternative splicing we studied seven missense and one nonsense mutations.

#### **6.1 The effect of exon 9 missense and nonsense mutations on the alternative splicing**

Some of these missense mutations are particularly interesting as they have been associated with CF phenotypes of different severity with a tissue specific involvement. Normal or mutated exon sequences were inserted in the hybrid minigene that has been widely utilized to study alternative splicing regulation of different genes (Figure 6.1 A) and in the case of the CFTR exon 9 it has been shown to mimic the endogenous pattern (152, 185). Minigenes with a fixed number of (TG) and T repeat at the 3' end of intron 8 locus (TG11T5 minigene) were transiently transfected into Hep3B cell line and the splicing pattern analysed by RT-PCR amplification. Compared with wild type TG11T5 minigene construct, which showed 65 % of exon 9 inclusion, some of the missense substitutions significantly modify the splicing pattern (Figure 6.1 B).

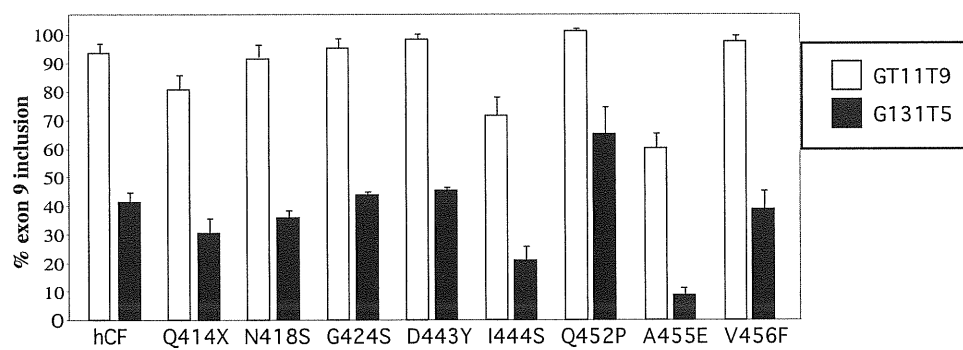




**Figure 6.1** Effect of natural missense and nonsense mutations on CFTR exon 9 alternative splicing. **A**) A schematic representation of the hybrid minigene, hCF-(TG)m(T)n. The detailed description of the minigene has been done on figure 4.1 A). **B**) Autoradiography of exon 9+ and exon 9- splicing variants obtained from RT-PCR assays performed on transfection experiments of Hep3B cell line using minigene (TG)11(T)5 minigene variant carrying the reported missense and nonsense mutations (left panel). The RT-PCR products are resolved on 6% polyacrilamide gel electrophoresis. The percentage of the two splicing products were quantitated using a Cyclone and is shown in the right panel. Each bar represent the mean $\pm$  SD of three independent experiments done in duplicate. For the comparison to hCF the asterisks indicates  $p < 0.01$

Three of the missense mutations, G424S, I444S and A455E decreased significantly exon 9 inclusion to 30%, 40%, and 16%, respectively and a modest decrease was evident for N418S. The nonsense Q414X substitution also induces a significant reduction of the exon 9 inclusion to 48%. The missense mutations D443Y and V456F did not affect significantly the splicing pattern whereas the Q452P caused an almost complete inclusion on the exon (96 %) (Figure 6.1 B).

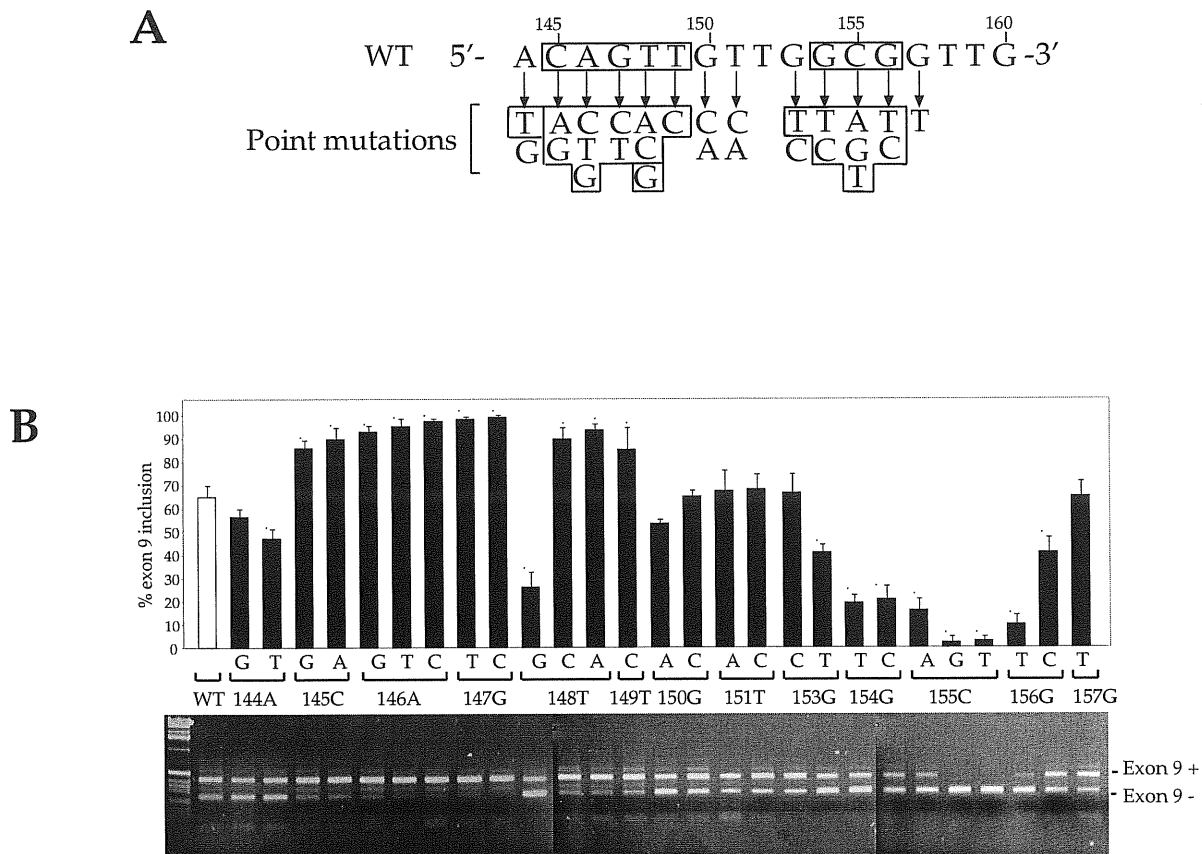
Our previous data (184) and that we reported in the first part of the results have shown that the number of the polymorphic TG and T repeats located near the 3' splice site in intron 8 modulate the efficiency of exon 9 splicing. In fact the increase in the number of the TG repeats and/or the decrease in the length of the T tract reduces the exon 9 inclusion. We thus evaluated the changes in the splicing efficiency induced by the exonic substitutions in relation to the composition of this polymorphic region. The missense and nonsense variations, analysed in figure 6.1 B within the TG11T5 minigene, were introduced in the TG11T9 and TG13T5 minigene contexts. These two polymorphic variants cause respectively increase and decrease in the splicing efficiency. Transfection of the different missense and nonsense minigenes showed that the proportion of exon 9+ transcript was modulated by the composition of the polymorphic locus (Figure 6.2) indicating that the exonic elements affected by the missense and nonsense mutations modulate the efficiency of splicing in conjunction with the 3' splice site variability. Because exon 9 skipping produce a non-functional CFTR protein (175, 183), some of these missense mutations analysed, in particular the A455E, may exert part of their phenotypic expression and disease variability affecting the splicing pattern, in particular when associated to an unfavorable polymorphic variant near the 3' splice site.



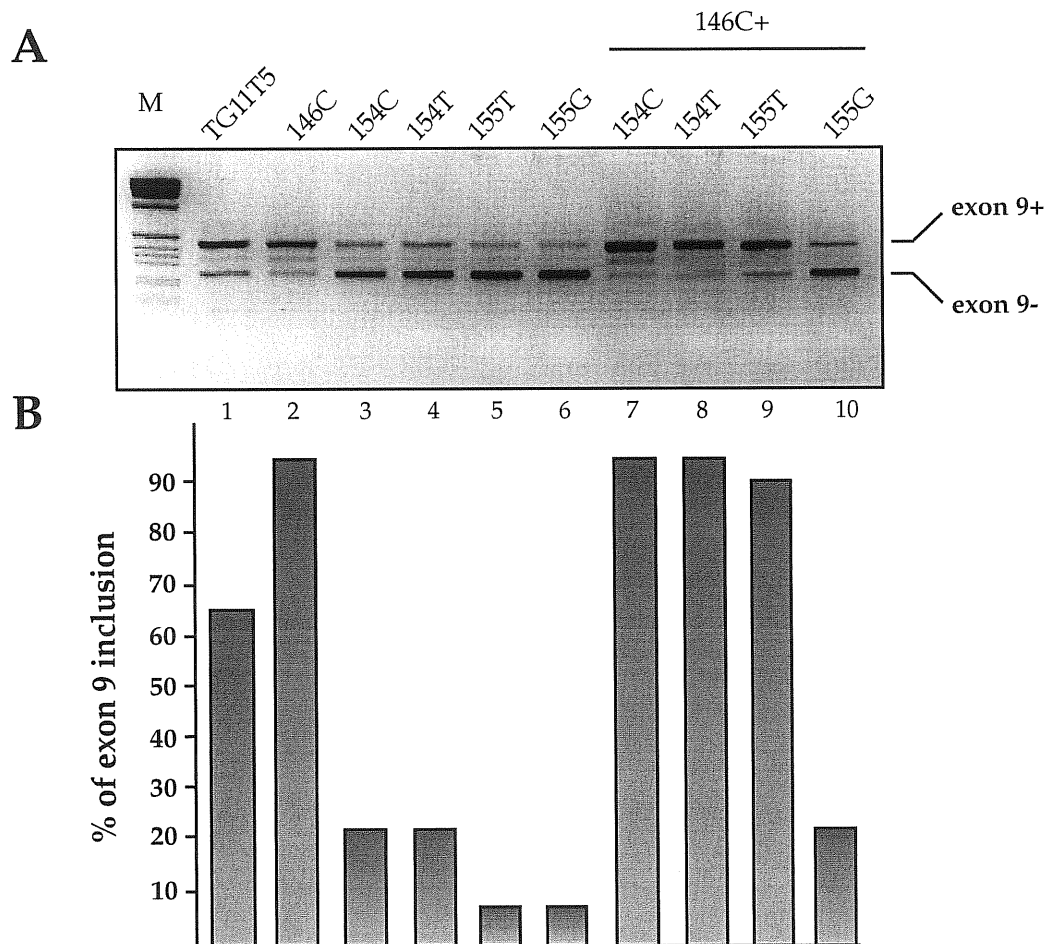
**Figure 6.2** Relationship between the exonic substitution and two different polymorphic variants. The missense and nonsense mutations reported in figure 5.1 B, were introduced in the (TG)11(T)9 and (TG)13(T)5 minigene variants. This graphic shows the percentage of exon 9 inclusion detected by RT-PCR assays on transfection experiments. The bar represent the mean $\pm$ SD of two independent experiments done in duplicate.

## 6.2 Mapping of cis-acting regulatory sequences in CFTR exon 9

To better characterise the mechanism by which some of CFTR exon 9 mutations alter splicing we analysed the effect of several site-directed mutants distributed in a short region of 15 bp (position 144 -157) which include three missense mutations (Figure 6.3 A) with completely different effect on the splicing efficiency: Q452P (A146->C) which induces exon inclusion, A455E (C155->A) causing exon exclusion and V456F (G157->T) which has no effect. The substitution of wild-type C in position 155, which corresponds to the base affected by A455E, to either A (the natural mutation), G or T induce exon skipping, indicating the critical role in this position of the cytosine residue in the enhancing function (Figure 6.3 B). A reduction in the percentage of exon 9 inclusion is also observed as a result of the adjacent positions 154 and 156 point substitutions and for one of the mutant in position 153, 153G->T. On the contrary the 153G-C and the 157G-T (V456F) variants did not significantly affect the splicing pattern. Extension of the mutagenesis in 5' direction including the Q452P (A146->C) variant showed that all mutants from position 145 to 149, with the notable exception of the 148T->G, induce exon inclusion (Figure 6.3 B). The 148T-G substitution in fact, contrary to the other nearby mutations produce significant exon skipping. It is interesting to note that this particular mutation create an overlapping and adjacent consensus sequence for 5' and 3' splice sites consisting of CAGGTG. These results identify two adjacent short flanking sequences with mainly enhancer and silencer specificity, a 5' silencer CAGTT element and a 3' enhancer GGCCG element.



**Figure 6.3** Mapping of exonic splicing enhancer and silencer elements in CFTR exon 9. **A)** CFTR exon 9 sequence from nucleotide 144 to 160 showing the single nucleotide substitutions analysed. The natural mutations Q456P, A455E and V456F correspond to A146C, C155A and G157T, respectively. **B)** RT-PCR products from splicing assays from the single nucleotide substitutions. 3  $\mu$ g of each minigene were transfected on Hep3B cell line and the RT-PCR products were analysed on a 1.5% agarose gel (lower panel). The upper panel shows the percentage of exon inclusion detected by radioactive PCR and quantitated using a Cyclone. Each bar represent the mean of three independent experiments. Error bars are the standard deviations. For the comparison of normal exon 9 to mutations the asteriks indicates  $p < 0.01$ .



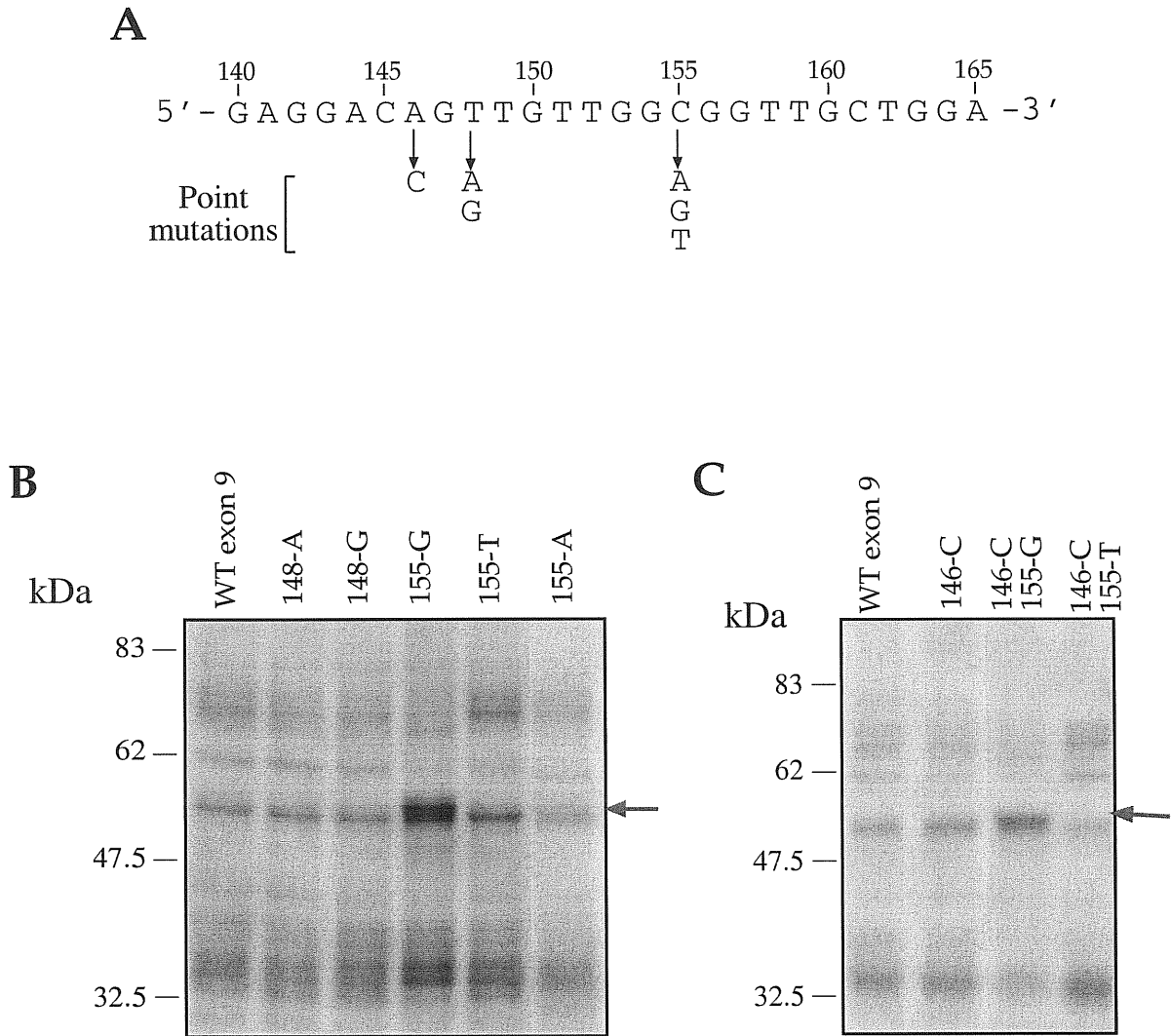
**Figure 6.4** Effect of double mutant on CFTR exon 9 inclusion. **A)** In lanes 2,3,4,5 and 6 of the panel are shown the RT-PCR results of the point mutation in the position 146, 154 and 155 of the exon 9. The percentage of exon 9 inclusion is about 95% for 146C mutant, but is decreased to 20% and 5% when mutants 154C and 154T, and 155T and 155G, respectively, were used. In lanes 7-10, the same point mutations were studied in association with 146C substitution. Only the double mutant 146C/155G induced a significant exon skipping. The samples were loaded in a 1.5% agarose gel. M, is 1Kb molecular weight marker. **B)** Graphical representation of the percentage of exon 9 inclusion obtained from the experiment described in A).

These elements were defined as a Composit Exonic Regulatory Elements of Splicing (CERES) with overlapping enhancer and silencer properties.

To study the nature of sequence we evaluated the effect on the splicing efficiency of double mutants affecting simultaneously the two elements. The 146C substitution in the ESS, which presents 95% of exon inclusion, was analysed in association with the nearby exon skipping mutations in position 154 (C or T) and 155 (G or T). These four substitutions gave a similar percentage of exon 9 inclusion, which is about 20 % for 154C and 154T and less then 5% for 155G and 155T. However, when they were analysed in association with the 146C substitution, only the 155G was able to induce significant exon (Figure 6.4) indicating a strict functional interdependence between the two juxtaposed elements.

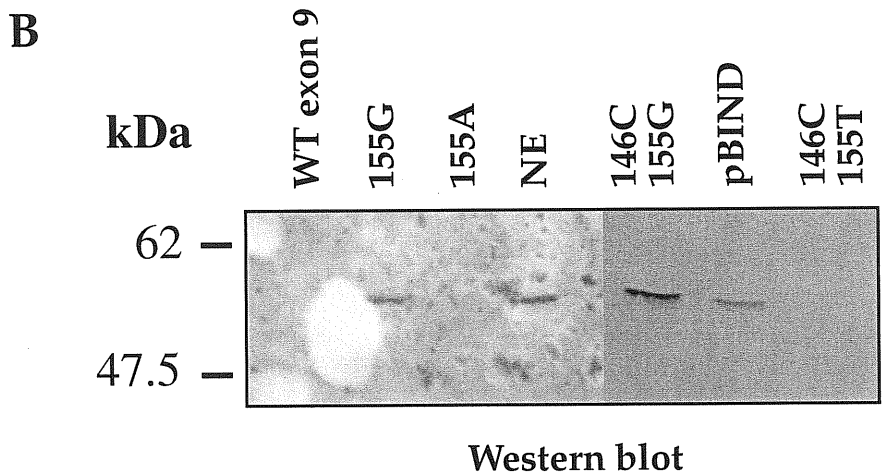
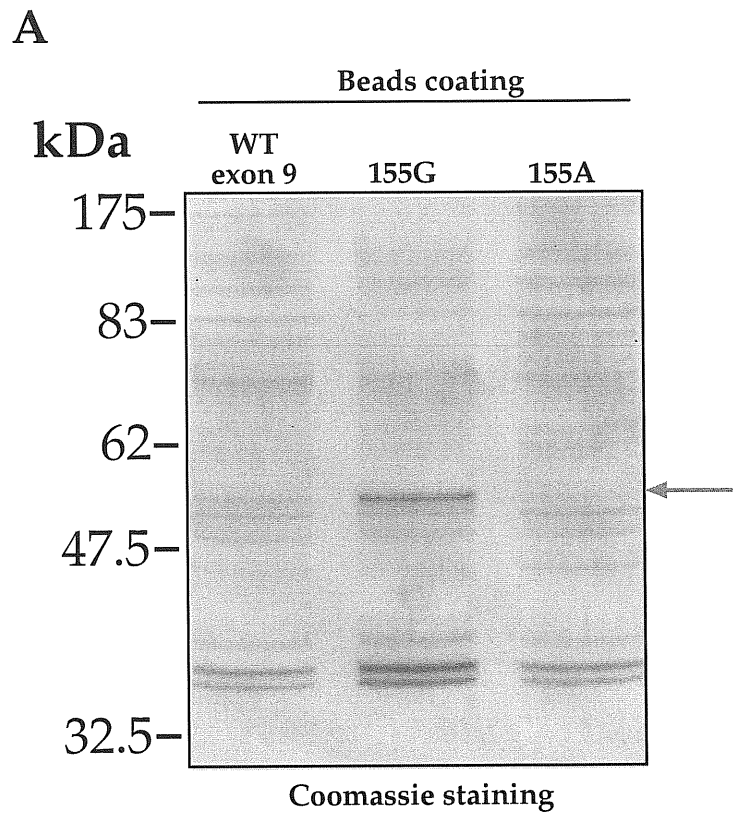
### 6.3 hnRNP H binds to 155G mutant

Mutations at position 155, G and T, may induce exon skipping with different mechanisms. In fact, both cause severe exon skipping defect alone, but only 155G maintains a strong splicing inhibition in association to 146C, as is shown by the different splicing efficiency of 146C-155G and 146C-155T double mutants (Figure 6.4). This may be due to changes in binding properties or affinity of diverse *trans*-acting factors. To identify *trans*-acting factors possibly regulating the CFTR exon 9 splicing at the CERES that could explain this difference we performed UV cross-linking experiments on the 3' portion of the CFTR exon 9 (Figure 6.5). We considered not only the double mutants at position 146 and 155 but also other mutants causing exon skipping or exon inclusion (Figure 6.5 A and B). The 155G and 146C-155G substitutions showed in the UV cross-linking experiments an additional band of about 50 KDa.



**Figure 6.5.** UV cross-linking experiments on the exon 9 sequence from nucleotides 140 to nucleotide 165. **A)** Representation of the point mutations introduced in the wild-type exon 9 sequence. Mutation A->C in position 146 corresponds to the natural mutant Q452P, mutation T->A in position 148 corresponds to I444S natural mutant and the mutation C->A in position 155 corresponds to A455E natural mutant. **B)** UV cross-linking experiments using as labeled probes the wild-type exon 9, the following mutant sequences: 148-A, 148-G, 155-G, 155-T and 155-A. The red arrow indicates the additional band that only the mutant 155-G is able to bind. **C)** UV cross-linking experiments using as labeled probes the wild-type exon 9 and the following mutant sequences: 146-C, 146-C/155-G double mutant and 146-C/155-T double mutant. As in the left panel, the arrow indicates the additional band that only the double mutant 146-C/155-G is able to bind. The samples were loaded on a 10% polyacrilamide gel. Molecular weight marker is reported on the left of each panel.





**Figure 6.6** Binding of hnRNP H to 155G mutant. A) Pull-down assay using beads coated with the *in vitro* transcribed RNA, indicated on the top of the panel, in the presence of labeled UTP. The samples were loaded on a 10% SDS-PAGE gel and stained with Coomassie blue. The red arrow indicates a protein of about 50 kDa that was pull-downed only using 155C-G probe. C) A Western blot assay performed on the pull-down assay. The antibody against hnRNP H has been used to confirm that the protein bound to 155C-G mutant is hnRNP H. The samples were loaded on 10% SDS-PAGE. The molecular weight marker are reported on the left of the panels.

No significant changes in the UV cross-linked pattern was observed in the other mutants, which can be possibly due to the fact that this experimental assay can detect predominantly changes in binding properties of a relatively abundant splicing factors. To identify the splicing factor binding at 155G, we performed pull-down experiments with WT exon 9, 155G and 155A variants. Again, in the 155G mutant a protein of about 50 KDa was identified from the proteins eluted from the beads.

hnRNP H is an abundant splicing factor with the molecular weight of 49.2 KDa that binds to G-rich sequences and it has been shown to inhibit splicing of several exons (199-203). The 155G mutant creates a stretch of five G which can be indeed the specific target for this inhibitory splicing factor. Using a specific antibody against hnRNP H we analysed by Western blotting the proteins eluted from the pull down experiments (Figure 6.6 A). Immunoreactive hnRNP H was found in 155G mutant and in the double mutant 146C-155G, but not in 155A and in 146C-155T mutants. Only small amount of immunoreactive hnRNP H was observed when the entire exon 9 sequence was analysed. This could be due to the presence of G stretches at 5' end of the exon (Figure 6.6 B). This data suggest that the five G produced by the by 155G variant creates a new exonic splicing silencer element with specific binding to hnRNP H.

## DISCUSSION

## **Part 1: TDP43 binding to TG repeats at the polymorphic variant at 3' end of CFTR intron 8 negatively affect exon 9 alternative splicing**

The importance of the (TG)<sub>m</sub> sequence and of the adjacent (T)<sub>n</sub> polymorphic region in the regulation of CFTR exon 9 alternative splicing has been the subject of several genetic and molecular studies, owing to the discovery of the association of certain alleles of this characteristic polymorphism with monosymptomatic forms of CF (42, 47, 181). At this locus a variable number of dinucleotide TG repeats (from 9 to 13) followed by a T repeat (T5, T7, T9) can be found in the normal population. In addition we describe a patient with a peculiar composition at this locus composed by (TG)<sub>13</sub>(T)<sub>3</sub>, which is affected by a more severe CF phenotype. In general a high number of TG repeats and a low number of T repeats have been shown to favor the exclusion of CFTR exon 9 in mRNA (Figure 7.1). Exon 9 sequence codes for the first part of NBD1 and the skipping of this exon leads to the production of a non-functional CFTR protein (175, 183). However the proportion of exon skipping does not correlate in some cases to the polymorphic alleles at the 3' end of intron 8 and it varies among the tissue of the same subject, suggesting that other factors operate in conjunction with the polymorphic locus to regulate alternative splicing of CFTR exon 9.

We have identified TDP43 (189) as a trans-acting factor binding to the (TG)<sub>m</sub> polymorphic repeat region near the 3' splice site of human CFTR exon 9 (185). To evaluate the functional significance of the TDP-43/(UG)<sub>m</sub> interaction on CFTR exon 9 alternative splicing we used CFTR exon 9 hybrid minigene variants with different polymorphic repeats at the intron 8/exon 9 junction including the TG<sub>13</sub>T<sub>3</sub> allele found in a pancreatic sufficient CF patient.

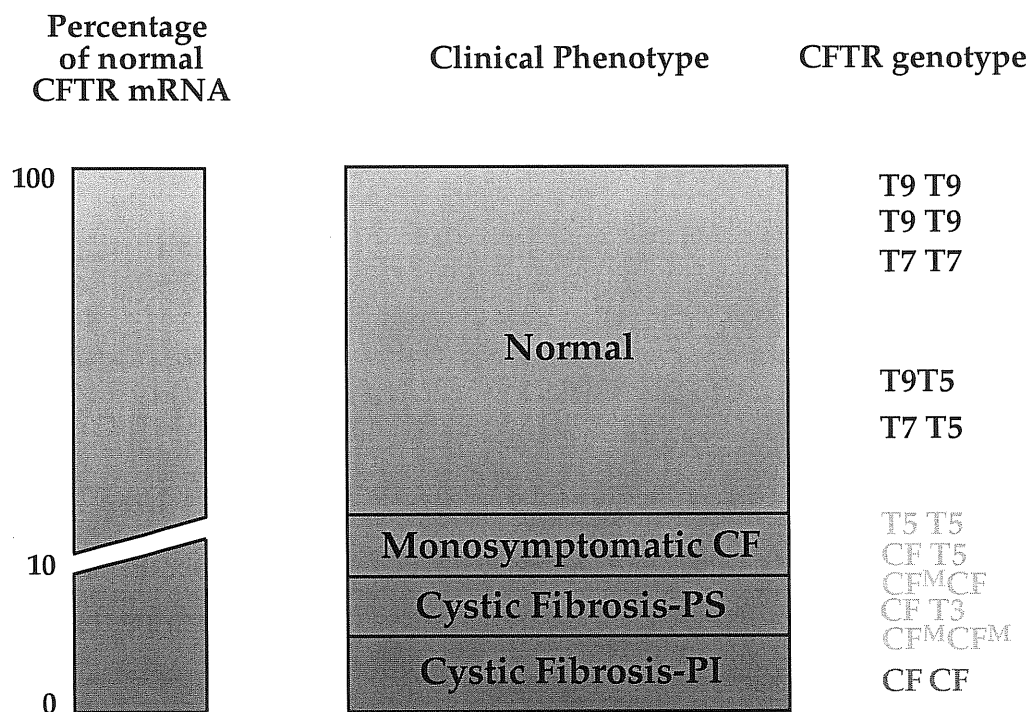
These minigenes were transfected in Hep3B cells with plasmids coding

for TDP-43 or for the splicing factor SF2/ASF that has been previously shown to induce CFTR exon 9 skipping. The results showed that the proportion of exon 9 exclusion was strictly dependent on the composition of the polymorphic locus, being directly related to the (TG)<sub>m</sub> polymorphic number and inversely related to the (T)<sub>n</sub> polymorphic number. Overexpression of TDP-43 resulted in a greater amount of CFTR exon 9(-) mRNA in all the four variants tested (Figure 4.1, B and C). To better define the CFTR exon 9 regulation we also co-expressed SF2/ASF, which binds to a different region in intron 9, the ISS (152). Overexpression of SF2/ASF and TDP-43 with the hybrid minigene that presented the lowest rate of exon exclusion (TG11T5) resulted in an enhanced inhibitory effect, indicating that binding of the two factors on either side of the exon has an additive effect. Co-transfection of increasing amounts of TDP-43 along with the TG13T5 minigene confirms that the inhibitory effect on the inclusion of exon 9 follows a dose-dependent curve which, although considering the high level of basal exclusion, was statistically significant in the 3 $\mu$ g and 5 $\mu$ g data points (Figure 4.2).

We noticed that the inhibition of CFTR exon 9 inclusion after TDP 43 overexpression was less abundant than that we observed when the SR protein SF2/ASF was overexpressed. The lower effect of TDP 43 could be due to the high endogenous level of this protein. To avoid this problem we inhibit TDP 43 expression in culture cells using PS-antisense oligo-deoxy-nucleotides (Figure 4.3). The PS-oligos were transfected in Hep3B along with the TG13T5 minigene. Among the four antisense PS-oligonucleotides tested three showed a consistent and reproducible effect on the CFTR exon 9 splicing. The analysis of the most efficient PS-oligonucleotide TIO1318 in a dose response study showed by Western blotting a concomitant reduction in the amount of the endogenous protein. Reduction in endogenous TDP 43 level resulted in

a significant increase in the percentage of exon 9 inclusion. At the highest dose of antisense the percentage of exon inclusion from the basal level of 40% reached 80% (Figure 4.3 B). This result is particularly important because TDP43 inhibition might represent a novel therapeutic target to correct CFTR exon 9 aberrant splicing in some patients.

TDP43 and SF2/ASF represent important splicing factors with inhibitory activity on CFTR exon 9 alternative splicing, which variable tissue specific concentrations may participate in the phenotypic expression of the disease. To understand the contribution of these splicing factors *in vivo* we performed the measurement of the levels of TDP 43 and SF2/ASF mRNAs in different human tissues. The results showed that mRNAs of both proteins are abundant in pancreas, lung and genital tract, three organs that are known to be particularly affected by CF (Figure 4.4). The fact that in transient transfection experiments, the overexpression of SF2/ASF together with TDP 43 results in an additive inhibitory effect on exon 9 recognition, makes it tempting to speculate that in patients these two factors might act in concert to lower the exon 9+ transcript. This concerted inhibitory effect would be more deleterious in those organs where these two factors are most abundantly expressed. The potential variability of TDP 43 and SF2/ASF concentrations also among the same tissue of different individuals provides one possible molecular basis to explain the variable phenotypic expression of the T5 allele, and why genital tract, pancreas and lung involvement varies in CF patients even when they have the same mutations.



**Figure 7.1** Comparison of percentage of normal CFTR mRNA, clinical phenotypes and CFTR genotypes. A high number of T at the polymorphic locus at 3' end of intron 8 is associated with a high percentage of normal CFTR mRNA. CF<sup>M</sup> indicates mild mutation.

## 7.1 Mechanism of TDP43-mediated CFTR exon 9 splicing inhibition

Human CFTR exon 9 is fairly unique in mammals of being subject to alternative splicing, in fact, our previous studies have shown that mouse exon 9 is a constitutive exon (184). The unusual behavior of the human exon 9 may be due to an evolutionary accident like a transposon hit that left sequences like the ISS in the intron 9, which is the main responsible of SF2 inhibition (152) and the UG repeats in the intron 8, target of the TDP43 interaction.

TDP43 has never been described before to affect the splicing process and its cellular function remained elusive (189). TDP43 is a cellular protein originally described to bind polypyrimidine rich region of the HIV-1 TAR DNA element and in such a way acting as a transcriptional inhibitor. Even if TDP 43 was described as a DNA binding protein, the presence of two RNA recognition motifs (residues 106-175 and 193-257) and several experimental evidences clearly indicate that TDP 43 binds RNA (169, 185).

As summarized in the introduction, the constitutive and the alternative splicing of pre-mRNA are highly regulated processes. This regulation involves different elements which act together to recognize the 5' splice site, 3' splice site and the branch point sequence, basic elements involved in the two trans-esterification reaction that occur in the spliceosome. TDP 43 interacts to (TG)<sub>n</sub> repeats in proximity to the 3' splice, but the effect of TDP43 may not be necessarily mediated by simple binding competition with 3' splice site recognition factors, but could also be linked to the still unknown function of this protein. On the other hand, TDP43 may bind to some cellular proteins that disrupt the recognition of exon 9 by the splicing machinery when positioned next to it. The ability of TDP 43 to form protein-protein interaction is



suggested by its glycine rich domain. The binding of TDP 43 to (TG)<sub>m</sub> repeats might influence splicing assembly in two ways. In the first case, the binding might indirectly mask the adjacent (T)<sub>n</sub> tract modulating the binding efficiency of U2AF65 to the polypyrimidine tract. Alternatively, TDP 43 could be directly involved in the formation of new non-favored protein-protein interaction with other proteins of the splicing machinery thus affecting exon 9 recognition.

The presence of (TG)<sub>m</sub> repeats at the 3' end of intron 8 could be considered as a disturbing element interfering with the maturation process of the CFTR pre-mRNA. A comparison with CFTR exon 9 and intronic sequences from mouse, in which we could not observe alternative splicing, reveal a great difference. In mouse intron 8 the polymorphic (TG)<sub>m</sub> repeat is absent and this suggest that the introduction of this region could be part of a more complex event that placed this (TG)<sub>m</sub> element adjacent to the 3' splice site of human exon 9 early during the course of evolution (204, 205). This hypothesis is supported also from our recent sequencing of mouse introns flanking mouse exon 9 which shown that they lack the two intronic regulatory elements and are also very different in length compared with the human introns (184). The difference in human and mouse sequences suggests that the disturbing influence of (TG)<sub>m</sub>-TDP 43 complex on the 3' splice site can be rescued by a long poly-pyrimidine (T)<sub>n</sub> tract, which might distance the (TG)<sub>m</sub> repeat from the 3' splice site. The *in vivo* importance of this 'masking' effect aggravated by lower numbers of T repeats is evident from the association of the T5 alleles with certain clinical entities such CBAVD or milder forms of CF. Moreover, this effect is underscored by the report presented in this work of a TG13T3 allele in a pancreatic-sufficient CF patient. In fact, our data show that the novel TG13T3 allele leads to the largest extent of exon 9 skipping reported thus far for a natural occurring (TG)<sub>m</sub>(T)<sub>n</sub> variant, sufficiently high to explain the CF phenotype of this patient in which the only

CFTR mutation reported is the  $\Delta F508$  on the other allele.

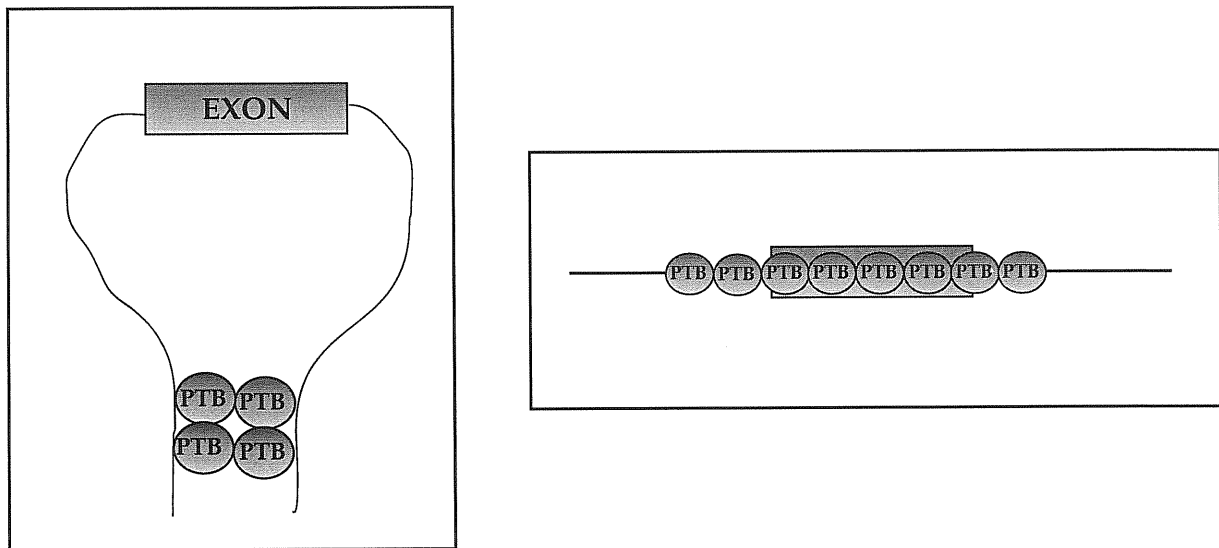
## Part 2: PTB could be another trans-acting factor involved in CFTR exon 9 alternative splicing

Several eukaryotic RNA binding proteins preferentially interact with uridine-rich sequences and thus have been classified as Py-tract-binding proteins (57). The most important Py-tract-binding proteins identified are: U2AF65, *Drosophila* SXL, PTB and recently TIA1. U2AF65 is the essential splicing factor that recognises a wide variety of Py-tracts and it mainly interacts with polypyrimidine tracts near the 3' splice sites (79). *Drosophila* SLX regulates the 3' splice site switching of *tra* pre-mRNA and exon skipping of its own pre-mRNA (195). PTB, also known as hnRNP I, was originally identified by its binding to the Py-tracts of adenoviral major late promoter (Ad ML) and  $\alpha$ -tropomyosin pre-mRNA (198). TIA1, the homologue of *S. cerevisiae* Nam8p protein, is involved in the splicing regulation of different genes (206, 207) and it has been recently reported to interact with RNAs containing short stretches of uridines near the 5' splice sites (208).

The CFTR intron 9 region between the 5' splice site and the ISS element, contains a particular arrangement three polypyrimidine-rich elements (PY1, PY2 and PY3). By mutagenesis we created and tested several mutants at these three elements using hybrid minigene expression assay. This analysis clearly show that one of these elements (PY2) behaves like an Intronic Splicing Enhancers (ISE), as its deletion causes a significant reduction in the percentage of exon inclusion (Figure 5.2). In fact, mutation in the first TCTT motif located in position +6 from the junction or mutation in the third motif did not affect the splicing pattern, while mutants in the second TCTT motif either alone or in combination with mutations on the other elements strongly reduced the level of exon 9 inclusion from 65% in the wild type construct, to 25-30% in the mutants (Figure 5.2).

The mutations introduced in the polypyrimidine elements might disrupt the ability of these sequences to interact with specific trans-acting factors. To study this hypothesis we performed UV cross-linking experiments in which we competed the binding of nuclear proteins to the intron8/ exon 9/ intron 9 CFTR sequence with competitors correspondent to the intron 8, the exon and the intron 9 sequences. Using the intron 9 as RNA competitor, we observed a specific competition in the binding of a protein with an apparent molecular weight of 60 KDa, which, in better electrophoretic resolution appears as a doublet (Figure 5.4, B). These bands were also competed by the specific (UCUU)<sub>3</sub> PTB-binding RNA (Figure 5.3, C) and the doublet is also characteristic of PTB splicing factor. By immuno-precipitation assays using anti-PTB antiserum, we found that PTB is able to bind CFTR intron 9 sequence and this binding is lost when the three polypyrimidine-rich elements are disrupted (Figure 5.5). PTB presents three well characterized isoforms (PTB1, PTB2 and PTB4) derived from pre-mRNA alternative splicing (209). We have used recombinant PTB1 isoform for *in vitro* binding experiments where we showed an efficient binding of rPTB to (UCUU)<sub>3</sub> sequence (Figure 5.6). Immuno-precipitation experiments and the competition with (UCUU)<sub>3</sub> RNA in UV cross-linking assays, indicate that the 60 KDa doublet in UV cross-linking contains PTB. Several proteins have been shown to bind to polypyrimidine-rich regions and the presence of a protein in the molecular weight range of PTB, with similar binding properties but causing exon inclusion (and which overlap with the PTB bands in UV cross-linking experiments) cannot be completely excluded. The identity of this putative protein will be followed up using mass spectrometry sequence determination of the pull-down material.

PTB was first identified ~10 years ago (197) and then was found to have features of an hnRNP protein (209). It was only recently that PTB was recognised as an important player in alternative splicing (162).



**Figure 8.1** Schematic representation of two potential mechanisms of PTB-mediated exon silencing. In the left panel the binding of PTB on intronic sequences and the subsequent multimerisation of PTB molecules forms a bridge-structure that sequesters the exon. In the right panel, PTB can oligomerise across the exon resulting in its coating and in such a way antagonising the definition.

Nevertheless, the precise mechanism by which PTB influences splicing is still unclear, even though it has been implicated in the alternative splicing of a number of genes. PTB-binding sites sometimes overlap binding sites for U2AF65, and simple competition could account for the inhibitory action of PTB (210). In most other instances, PTB-binding sites do not directly overlap with the binding sites of general splicing factors such as U2AF65. Therefore, in this case, PTB-mediated exon silencing is clearly not caused by direct competition with general splicing factors. Two models have been proposed to explain PTB involvement in different alternative splicing examples (Figure 8.1). The majority of exons silenced by PTB are flanked by PTB-binding sites on both adjacent introns. Given that PTB can multimerise, it has been postulated that PTB proteins can interact across the exon (165, 166, 211). The PTB sites flanking silenced exons could define, in such a way, a repressive zone within pre-mRNA.

A very similar interaction was proposed for hnRNP A1 proteins binding on either side of a regulated exon of the hnRNP A1 pre-mRNA (212). Alternatively PTB can oligomerise across an exon, resulting in the coating of the exon itself, thus antagonising its definition.

## **8.1 Functional Effect of PTB on Human Exon 9 Alternative Splicing**

The functional effect of PTB on CFTR exon 9 alternative splicing has been studied using a minigene system in Hep 3B cell line. From our experiments we observed that the overexpression of PTB causes a reduction of exon 9 inclusion in a dose-dependent manner (Figure 5.7). Once defined the functional effect of PTB overexpression on the CFTR wild type minigene, we studied its effect on the minigene constructs in which the polypyrimidine rich region were disrupted. PTB is a splicing

inhibitor and there are several examples of PTB-regulated alternative splicing where disruption of PTB-binding sites leads to an increase in the exon inclusion. In  $\alpha$ -tropomyosin gene, exons 2 and 3 are mutually exclusive exons and in the intron 2 there is a regulatory element able to bind PTB. It has been shown that mutations introduced in the UCUU motif within this element caused a misregulation of alternative 3' splice site in vivo (213). PTB is also involved in the regulation of FGF-R2 IIIb exon alternative splicing via the interaction with a regulatory element located in the intron downstream. Mutation in this element caused a derepression in the exon IIIb alternative splicing (214). In our system, overexpression of PTB using minigene constructs with disrupted PTB-binding sites still resulted in an inhibitory effect on exon 9 inclusion (Figure 5.8). This behavior could be due to a cooperative binding of PTB in different sites. In fact, there are many putative PTB-binding sites in the intron 9 portion included in our minigene and we also observed that intron 9 sequence containing at least one PTB-binding sites (Figure 5.4) could compete for the binding of PTB to the wild type intron 9. This kind of modulation between different PTB binding sites has been reported for the regulation of c-src neuron-specific N1 exon in which PTB molecules flanking the exon are likely to interact cooperatively to repress N1 splicing in non-neuronal extract (166). In the CFTR intron 9 sequence that we analysed there are several optimal PTB-binding sites TCTT (or UCUU in the pre-mRNA). When we performed overexpression of PTB along with hybrid human/mouse minigenes, the response to PTB was rescued when first 77 nucleotides of intron 9 from the junction were substituted to mouse sequence (Figure 5.9). Indeed this effect is lower than that obtained from wild type TG11T5 minigene. This suggest that also the PTB-binding sites present in the ISS are important in the modulation of PTB response and this is confirmed by the results obtained from transfection experiments in which we use

the M1,2,3 minigene. In fact, in this case the basal level of exon 9 inclusion was approximately 20% (Figure 5.2) and PTB overexpression lowered it to an 8% (Figure 5.8), indicating that the functional effect of the protein is shifted to the ISS-TCTT-containing region.

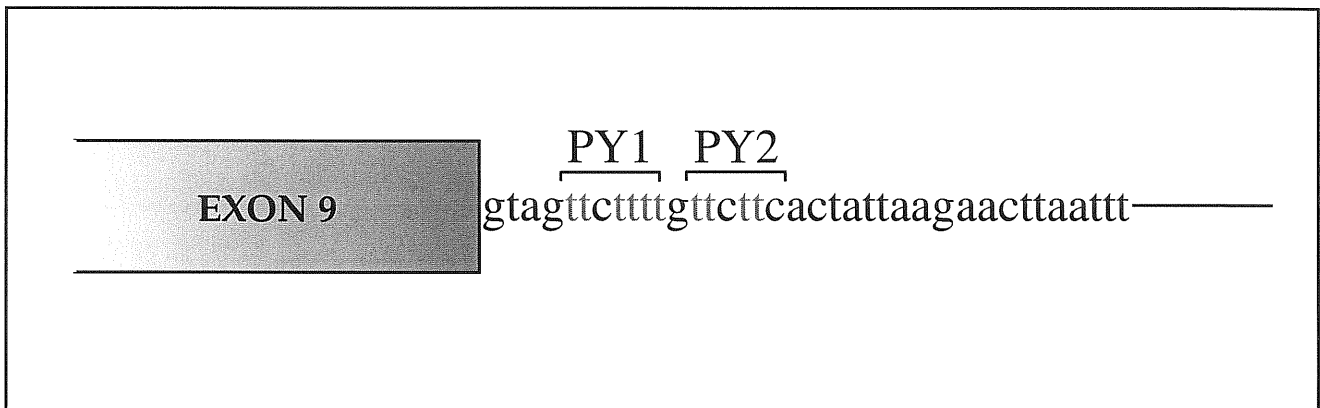
## 8.2 The PY2 element

Interestingly the mutation in PY2 resulted in a great reduction in exon 9 inclusion both considering the basal level and after PTB overexpression (Figure 5.2 and 5.8). When the mutation introduced in the PY2 element is associated with mutations either in PY1 or in PY3 the level of exon 9 inclusion is similar to that obtained when only PY2 mutation is present. This suggests that PY2 element might have a dominant effect on the other two elements.

Moreover, the fact that minigene constructs in which these mutations have been introduced still responded to PTB overexpression, suggests that other mechanisms could be involved in the modulation of exon 9 inclusion via the PY2 site. One possibility could be represented by TIA-1 protein. This protein is the homologue of *S. cerevisiae* Nam8p protein and is involved in the splicing regulation of different genes (206, 207). It has been reported that TIA-1 interacts with RNAs containing short stretches of uridine (208) and these sequences are frequently found in regulatory regions of RNAs such as the 3' splice site of introns, and in 5' and 3' untranslated regions. In different studies has been demonstrated that the interaction of TIA-1 to pre-mRNA facilitates the binding to U1 snRNP to weak 5' splice site and in a such a way this protein acts as a positive splicing regulator (206, 207). In the rat fibroblast growth factor receptor 2 (FGFR-2) the alternative exons K-SAM and BEK are alternatively spliced. In the intron downstream the K-SAM exon there is a polypyrimidine activator sequence named IAS1 (215) located at 8



nucleotides from the 5' splice site junction. This element contains 16 non-contiguous uridines and it has been demonstrated that is able to activate splicing of an heterologous exon in vitro. The IAS1 element functions independently from other regulatory elements to enhance the 5' splice site recognition via TIA-1 binding (206), this activation occurs only if the element is located close to the exon/intron junction. In the CFTR intron 9 sequence, PY1 and PY2 polypyrimidine elements start at position +5 from the 5' splice site and are separated by a guanine residue. Together this two elements form a stretch that contains 12 pyrimidine residues, 10 of which are uridines (Figure 8.2). Our results suggested that PTB is not directly involved in the modulation of the exon 9 inclusion acting on PY2 since mutations on this element did not enhance the recognition of the exon. Indeed, we can speculate that TIA1 might bind to this element, which can function as a splicing activator. This activatory role is consistent with the fact that mutation which disrupted the PY2 element caused a decrease in the level of exon 9 inclusion.



**Figure 8.2** Schematic representation of the polypyrimidine rich elements close to 5' splice site. PY1 element starts at base +5 from the junction while PY2 starts at base +13. Together these two elements contain 13 pyrimidines. The 10 uridine residues are in red.

## **Part 3: The important role of missense and nonsense mutation in CFTR exon 9 at the pre-mRNA level**

CFTR exon 9 alternative splicing presents a complex regulation. As we previously reported in our studies (152, 185), the (TG)<sub>m</sub>(T)<sub>n</sub> polymorphic region at 3' end of intron 8 and the ISS element in the intron 9 bind to TDP43 and SR proteins, respectively, and in such a way modulate this process.

Variable skipping of CFTR exon 9 has been associated to CF phenotypes of different severity. This phenotypic variability is one of the hallmarks of CF. In fact, patients with the same CF mutation might present different organs involvement and clinical situation.

In our previous experiments we focused our attention on the intronic sequences flanking exon 9. Indeed, the introns are not the only sequences of the pre-mRNA involved in the binding of constitutive and alternative splicing factors. Many studies have shown that also exonic sequences contain informations not only for codon specification, but also cis-acting elements involved in the regulation of the splicing process.

Due to this reason we have analysed the role that exonic sequences might have in the alternative splicing, taking advantage of missense and nonsense mutations naturally occurring in the exon 9 and extending these by extensive site-directed mutagenesis.

### **9.1 Missense and nonsense mutations affect alternative splicing of CFTR exon 9**

In our work we have considered eight natural point mutations distributed in the entire exon 9 (Figure 9.1).



Five of these mutants, G424S, I444S, A455E, Q141X and Q452P, caused a modification in the splicing pattern (Figure 6.1 B). The first four mutants decrease the splicing efficiency while the last one cause an increase in the exon inclusion to a value that is approximately 98%. These mutants were also studied in relation to different polymorphic variant at the (TG)<sub>m</sub>(T)<sub>n</sub> locus. The effect of the point mutations was independent but modulated by the composition of the polymorphic tract (Figure 6.2). This suggest that part of the phenotypic expression and disease variability associated to some of these mutants may result from splicing alterations, in particular when linked to unfavorable composition of the polymorphic variant.

To better analyse the mechanism involved in exon skipping or inclusion caused by the missense and nonsense mutations, we have done a detailed site-directed mutagenesis experiments on a critical region of about 15 nucleotides, which includes three missense mutations (Q452P, A455E and V456F) with different and opposite effect on the exon definition (Figure 6.3). We used site-directed mutant experiments because they can give more detail on the mechanism involved in the natural substitutions and also because discrete deletions can potentially result in gross alterations of exon definition. These results identify two adjacent short flanking sequences with both enhancer and silencer properties, a 5' silencer CAGTT element and a 3' enhancer GGCGG element. This elements could be defined Composit Exonic Regulatory Element of Splicing (CERES).

From *in vitro* and *in vivo* studies using minigenes, exonic enhancer and silencer elements have been defined as those elements that when mutated cause exon exclusion and inclusion, respectively. On the other hand, a particular alteration of the splicing pattern could also be due to the presence of a new element introduced by a mutation and not to a disruption of exonic enhancers or silencers. The Composit Exonic Regulatory Element of Splicing (CERES) that we have identified, is an

example of this last situation.

## 9.2 The ability of 155G mutant to bind hnRNP H

Mutations in the CERES sequence differently affected exon 9 inclusion. To study the nature of this regulation we have analysed the effect of double mutants on this region on the alternative splicing. The 146A $\Rightarrow$ C substitution presents 95% of exon inclusion. This mutation was analysed in association with the mutations which induce skipping 154G $\Rightarrow$ C, 154G $\Rightarrow$ T, 155C $\Rightarrow$ G and 155C $\Rightarrow$ T. Interestingly, only the double mutant 146A $\Rightarrow$ C/155C $\Rightarrow$ G showed a significant decrease in exon 9 inclusion (Figure 6.4), indicating a peculiar property of the 155C $\Rightarrow$ G substitution. This mutation introduced a G in the exon 9 sequence and the resulting motif is represented by a stretch of five consecutive Gs. *In vitro* studies were used to detect possible changes in the binding pattern of these double mutants. We identified hnRNP H as the protein that specifically bind to 155C $\Rightarrow$ G mutants. hnRNP H has been implicated in several diverse aspects of pre-mRNA processing including splicing stimulation (216, 217) and splicing repression (199). This factor belong to a protein family which members are hnRNP H, H',F and 2H9. The core sequence required for the assembly of hnRNPs of the H family onto RNA is GGGA, whereas a run of five Gs followed by a C promotes only hnRNP H and H' binding (202). There is strong evidence of a role for GGG triplets in splicing regulation. In fact, GGG triplets have been found associated with the intronic portion of 5'- and 3'- splicing sites (218, 219) and have been shown to positively regulate exon selection when positioned in short introns (220). Furthermore, sequences enriched in G triplets placed in an alternative exon have been

identified as elements capable to promoting exon skipping (150). These results suggest that GGG sequences act in an opposite way according to their location either in exons or in introns.

Several possible mechanisms may account for how hnRNP H blocks the inclusion of CFTR exon 9 when bound to 155C $\Rightarrow$ G mutant. For example, the binding of hnRNP H to the mutant may sterically interfere with the binding of SR proteins with enhancing properties to the same elements. Another speculation could involve an interaction between hnRNP H and other factors such as TDP43 or SF2/ASF that could “mask” the recognition of the exon 9.

### 9.3 Clinical implications

An important implication of our work is related to the fact that missense mutations in exon 9, like the majority of CFTR missense mutations, which have been considered mainly from the point of view of their effect at the protein level, can affect splicing regulation. Our work shows for the first time that some missense mutations in the CFTR exon 9 coding sequence can affect exonic regulatory elements causing a defect in exon recognition.

An interesting situation is presented by the missense mutations A455E, a rather frequent mutation in some populations (221, 222) that appears to be able to achieve adequate levels of chloride conduction at the cell surface (44, 223), and it has been suggested to cause only a partial CFTR protein processing defect (224). Our data strongly suggest that the poor CFTR activity observed “*in vivo*” in patients carrying the A455E mutation may originate from the combination of these two molecular defects: the partial early biogenesis block in protein processing and the splicing defect. The modulation by the concentration of splicing factors, which have an inhibitory effect on the CFTR exon 9 (152, 225) and a

specific and possibly individual variation in their distribution, can provide a possible explanation for the phenotypic variability and tissue specific defects in CF patients, particularly in those carrying the A455E substitution, which has been found associated with very different phenotypes (226).

The finding that missense mutations in CFTR exon 9 can cause splicing defects may be extended to mutations located in other exons of the CFTR gene or in other gene which could affect splicing if they involve a regulatory element. Indeed exonic mutations not involving conserved obvious canonical splice sites have been recently shown to cause splicing defects in different human genetic disease (227-229). The fact that different nucleotide changes in the same position, like the ones described above for 155C $\Rightarrow$ G and 155C $\Rightarrow$ T, can cause very different effect on splicing must reinforce our attention to functional studies of mutations and avoid to draw conclusions simply on the codon changes.



## **MATERIALS AND METHODS**

## **10.1 MATERIALS**

### **10.1.1 Chemical reagents**

General chemicals were purchased from Sigma Chemical Co., or MERK FR, Germany.

### **10.1.2 Enzymes**

Restriction enzymes were from Pharmacia Biotech, Sweden or New England Biolabs, Inc (USA). DNA modifying enzymes such as Taq Polymerase, Klenow fragment of *E.coli* DNA polymerase I, T4 DNA Polymerase and T4 DNA ligase were obtained from Roche Mannheim GmbH (Germany). T4 polynucleotide Kinase was from New England Biolabs, Inc. and Calf Intestinal Alkaline Phosphatase was from Pharmacia Biotech (Sweden) RNAase A was purchased from Sigma Chemicals Ltd. A 10 mg/ml solution of RNAase A was prepared in sterile water and boiled for 10 minutes to destroy trace amounts of DNAase activity. All enzymes were used following manufacturer instructions.

### **10.1.3 Synthetic oligonucleotides**

Synthetic DNA and RNA oligonucleotides were purchased from Roche Diagnostic.

The PS-antisense-deoxy-oligonucleotides were purchased from MWG.

#### 10.1.4 Radioactive isotopes

Radioactive  $\alpha$ - $^{32}\text{P}$  dCTP,  $\gamma$ - $^{32}\text{P}$  dATP,  $^{35}\text{S}$ ,  $\alpha$ - $^{32}\text{P}$ -UTP were supplied by Amersham U.K. Ltd.

#### 10.1.5 Bacterial culture

The K12, *E. coli* family strain DH5 $\alpha$  was used for transformation by plasmid and growth of the plasmid.

They were maintained in short term as single colonies on agar plates at 4°C and for longer term storage were kept on glycerol stocks, made by adding sterile glycerol to a final 15% v/v t. Glycerol stocks were stored at -20°C. An overnight culture of bacteria was grown in LB medium.

LB: Luria-Bertani medium per litre:

Difco Bactotryptone 10 g, Oxoid yeast extract 5 g, NaCl 10 g, pH 7.5.

Bacterial growth media were sterilised before using by autoclaving. Appropriate antibiotic, ampicillin, was added to media to a final concentration of 200  $\mu\text{g}/\text{ml}$ .

#### 10.1.6 Cell culture

The following cell line was used:

Hep3B; human, hepatocellular carcinoma, fibroblast-like.

## **10.2 METHODS**

### **10.2.1 Nucleic Acids Preparations**

#### **10.2.1.1 Small-scale preparation of plasmid DNA from bacterial cultures**

Rapid purification of small amounts of recombinant plasmid DNA was performed using the method based on alkaline lysis of recombinant bacteria and described in Sambrook et al. (230). The final pellet was resuspended in 50  $\mu$ l of dH<sub>2</sub>O and 5  $\mu$ l of such preparation were routinely taken for analysis by restriction enzyme digests.

#### **10.2.1.2 Large-scale preparations of plasmid DNA from bacterial cultures**

For large-scale preparations of plasmid DNA that was usually used for the transfection experiments were used JetStar purification kit (Genomed) following instructions of the producer. In order to get a huge amount of plasmid, we were dealing with 300- 500 ml of overnight inoculum using LB or TB medium.

#### **10.2.1.3 Preparation of the RNA from the cultured cells**

Cultured cells were washed two times with PBS and then RNazol B provided from TEL-TEX inc. was added. The chloroform extraction was performed one times. Supernatant was then precipitated with 70% isopropanole. The final pellet was then resuspended in 50  $\mu$ l of ddH<sub>2</sub>O and frozen at -80°C. The RNA quality was checked by electrophoresis on 0.8% agarose gels.

## 10.2.2 Estimation Of Nucleic Acid Concentration

### 10.2.2.1 Spectrophotometric

An optical density of 1.0 at 260 nm is roughly equivalent to a concentration of 50  $\mu\text{g}/\mu\text{l}$  for double stranded DNA, 33  $\mu\text{g}/\mu\text{l}$  for single stranded DNA and 40  $\mu\text{g}/\mu\text{l}$  for RNA samples. the ratio of values for optical densities measured at 260 nm and 280 nm is 1.8 for pure sample of DNA and 2 for RNA, these are reduced by protein contaminants and therefore the values were used to asses both the concentration and the purity of the samples.

## 10.2.3 Enzymatic Modification of DNA

### 10.2.3.1. Restriction enzymes

Restriction endonucleases recognise and cut within specific sequences of double stranded DNA leaving blunt ends, 5' and 3' protruding ends. These were used in the construction and analysis of recombinant plasmids. Each restriction enzyme functions optimally in a buffer of specific ionic strength. All buffers were supplied by the same company that supplied the enzymes and were used according with the manufacturer's instructions.

For analytical digests 100-500 ng DNA were digested in a volume of 10-20  $\mu\text{l}$  containing 5 U of the appropriate restriction enzyme per  $\mu\text{g}$  DNA.

The digest was incubated for 3-6 hours at the optimal temperature required by the enzyme used.

Preparative digest were made of 1-20  $\mu\text{g}$  DNA using the above conditions but a larger reaction volume. Enzymatic activity was then

removed either by incubation at 75°C for 20 minutes or phenol-chloroform extraction.

#### **10.2.3.2. Large fragment of E. Coli Polymerase I**

The large fragment of DNA Polymerase I (Klenow) lacks the 5' to 3' exonuclease activity of the intact enzyme, but retains the 5' to 3' exonuclease activity and 3' to 5' exonuclease activities. It is used to catalyse the polymerisation of deoxyribonucleotide triphosphates in a 5' to 3' direction on a template of double stranded DNA with a recessed 3' hydroxyl and protruding 5' phosphoryl terminus to generate a flush-ended DNA molecule. This was useful for creating compatible ends for ligation during construction of recombinant plasmids, to digest aspecific A residues added by Taq DNA polymerase at the 5' terminus and for radiolabelling of DNA fragments by random priming.

Klenow fragment was used with the proper buffer supplied by New England Biolabs Inc. at a final concentration of 5U per µg DNA. When a "fill-in" was required (DNA fragments with protruding 3' ends) dNTP for a final concentration of 0.5 mM were added. The mixture was incubate at 37°C for 20 minutes.

#### **10.2.3.3 Dephosphorylation of DNA 5' termini**

Calf intestinal phosphatase catalyses the removal of 5' terminal phosphate groups from linear DNA molecules. This is used to reduce self ligation of vector DNA during generation of recombinant clones, thus increasing the proportion of resulting recombinants containing the required DNA inserts.

This reaction was carried out in a final volume of 50-100 µl using 1U of enzyme per 0.5 µg DNA incubating for 1 hour at 37°C. The enzyme has

then be inactivated by adding 1 mM EDTA and incubating for 20 minutes at 75°C.

#### **10.2.3.4 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 linearised vector were ligated with a 5-10 fold molar excess of insert in a total volume of 20 µl containing 1x ligase buffer and 1 U T4 DNA ligase. Reaction was carried out at room temperature for 3-4 hours.

In some reactions synthetic oligonucleotide were included in the reaction. In these cases amounts added to each reaction to obtain inclusion of oligonucleotides in the resulting plasmid were about 100 fold molar excess over the DNA vector.

#### **10.2.4 Elution and purification of DNA fragments from agarose gels**

This protocol was used to purify small amounts (less than 1 µg) of DNA for subcloning. The DNA samples were loaded onto an agarose minigel and electrophoresed as described above.

The DNA was visualised with UV light and the required DNA fragment band was excised from the gel. This slab was put into the minimum length of dialysis tubing with as little gel buffer as possible and was electrophoresed in a minigel apparatus at 150 V for the time required for the DNA sample to migrate out of the gel slice. At this stage the

current was reversed for 30 seconds to elute any bound DNA from the dialysis tubing.

Buffer containing DNA was removed to a microcentrifuge tube and the tubing was washed twice with 100  $\mu$ l of gel buffer. Aliquots were pooled and DNA recovered by ethanol precipitation. To remove small gel fragments from the aqueous solution was centrifuged for 10 minutes at 4°C and the supernatant was transferred to a fresh tube.

### **10.2.5 PCR reaction for a cloning purpose**

The polymerase chain reaction was performed on genomic or plasmid DNA following the basic protocols of the Boehringer and Promega Taq DNA Polymerases. The volume of the reaction was 50-100  $\mu$ l with 1x Taq buffer, dNTP mix 200  $\mu$ M each, oligonucleotide primers 100nM each, Taq DNA Polymerase 2.5-5 U. 0.1 ng for plasmid and 100 ng of genomic DNA were used for amplification.

The amplification conditions are described for each particular PCR.

The amplifications were performed on a Perkin Elmer Cetus DNA Thermo Cycler.

### **10.2.6 Sequence analysis**

Sequence analysis of plasmid DNA were performed when was necessary by the dideoxynucleotide chain termination method using a T7-based DNA sequencing kit (Amersham-Pharmacia). Approximately, 1  $\mu$ g of purified DNA was subjected to the sequencing reaction. The samples were then loaded to a 8% acrylamide gel and run at 50W. The gel is then dried and exposed to kodak x-omat films for 12 hours.



### 10.2.7 Construction of the minigene system and site directed mutagenesis

The human CFTR genomic region including intron 8 (221 bp at 3' end), exon 9 (183 bp) and intron 9 (269 bp at 5' end) was amplified by PCR (94°C 30 sec, 60°C 30 sec, 72°C 60 sec, 35 cycles) using the following oligos, that include Nde I target site: hcfIVS8 dir, 5'-tttcatatggggccgctctaggacttgataatgggcaaatactta-3'; hcfIVS9 rev, 5'-cccctgacatatgctcgccatgtgcaagatacag-3'. PCR product was NdeI cut and ligated into a pBluescript KS plasmid (Stratagene, La Jolla, Ca), previously mutagenised by deletion of the XbaI/XhoI fragment and insertion of NdeI site through two complementary synthetic oligonucleotides within NotI site (pBSCFNde construct).

To permit subsequent cloning, Eco RI target site was introduced through a two-step PCR overlap extension method (231) by replacing an adenine with a cytosine at position +15 within exon 9. Sequencing excluded the presence of mutations within the insert.

To create Tn and (TG)<sub>m</sub> alleles, a XbaI-EcoRI cassette was generated by PCR using a common sense primer (5'-catatggggccgctctagga-3') and antisense primer (5'-aaagaattcccaaatcctgttaaaaaaacacacacacacacacacacacatcaaaaataaaagatgagtt-3') where A and/or (CA) number was changed according to the desired genotype. Identity of each construct was confirmed by sequencing.

To generate expression vectors, pBluescript human exon 9 inserts were NdeI cut and ligated within the EDB-intron of the h-CFTG11T5 NdeI-digested vector, previously described (232). The h-CFTG11T5 minigene consists of a modified version of the  $\alpha$ -globin-fibronectin EDB minigene which transcription is driven from the minimal  $\alpha$ -globin promoter and SV40 enhancer (152).

The mouse CFTR genomic region including intron 8 flanking sequences (157 bp at 3' end), exon 9 (183 bp) and intron 9 flanking sequences (209 bp at 5' end) was amplified by PCR (94°C 30 sec, 60°C 30 sec, 72°C 60 sec, 35 cycles) using the following oligos: mCF8i dir, 5'-tttcatatgtctagaaccatgtgctttatagt-3', that include Nde I and Xba I target sites; mCF9i rev, 5'-aaaacatatgataggttatccaatcttaagtgcagttctaaacacgtgta-3', that include Nde I target site.

PCR product was NdeI cut and ligated into previously described pBluescript NdeI-digested. Eco RI target site was introduced also in mouse minigene through a two-step PCR overlap extension by replacing an adenine with a cytosine at position +15 within exon 9 using oligonucleotides 5'-ccaggaattcccaaatccctattcata-3' and 5'-atagggatttggggaattcctggaga-3'.

Mouse minigene was NdeI excised from pBluescript and transferred into the EDA-intron of the pSv-EDA NdeI-digested vector. Human/mouse hybrids were generated or by exchanging human/mouse XbaI-EcoRI cassettes or through a two-step PCR overlap extension using previous constructs as template and oligonucleotides h/Py3'dir, 5'-tgttgcggttgctggatctactggatcaggaaaggact-3', h/Py3'rev 5'-agtaccttctgatccagtagatccagcaaccaccaaca-3'. Before expression the identity of all constructs was checked by sequencing.

### 10.2.8 Construction of the minigene system mutated in the polypyrimidine rich elements in intron 9

The mutations introduced in the polypyrimidine elements in intron 9 were obtained using PCR reaction with different oligonucleotides sets and using different templates in regard to the mutations inserted. The resulted products were digested Bam HI/ Kpn I and cloned in the

CFTR intron 9 sequence contained in the NdeI insert of the modified pBluescript described above. The mutagenised NdeI fragment was then subcloned in the h-CFTG11T5 NdeI digested minigene. The BamHI-KpnI cassettes were created by PCR-mediated site-directed mutagenesis using different primers following primers:

M1: CF9junA dir (5'-ctggatccactggagcaggcaaggtagttcattg-3');

UNI REV (5'-ggaaacagctatgacatg-3')

M2: CFA/3-2bs dir(5'-ctggatccactggagcaggcaaggtagttctttgttcac-3')

UNI REV (5'-ggaaacagctatgacatg-3')

M1,2: the same primers of M1ISS but on the M2ISS template

M2A: CF9/2bsmut(5'-ctggatccactggagcaggcaaggtagttctttgtaaac-3')

UNI REV (5'-ggaaacagctatgacatg-3')

M1,2A:CF9/1,2bsmut(5'-ctggatccactggagcaggcaaggtagtgattgtaaac-3')

UNI REV (5'-ggaaacagctatgacatg-3')

ΔM3: a two-step overlap extension method was used. For the first amplification the two sets of primers used are the following:

hCFex9/in9 dir (5'-ctggatccactggagcaggcaaggtagttctttg-3') and ΔY rev (5'-caaactgcaggacaccaaattaagtcttaac-3');

ΔY dir (5'-gtgtcctgcagttgtagtgctggaaggtat-3') and UNI REV

For the second amplification the hCFex9/in9 and UNI REV set was used

M1,2,3:CF9/1,2bsmut(5'-ctggatccactggagcaggcaaggtagtgattgtaaac-3')

and UNI REV on ΔM3 template.

For the minigene constructs Flip 1 were used the following primers:

hCFex9/in9 dir (5'-ctggatccactggagcaggcaaggtagttctttg-3');

FLIP 1 (5'-atggtaccatagccagcactacaaactaga-3') as reverse primer.

## 10.2.9 Site directed mutagenesis in exon 9

The missense and artificial point mutations were introduced in the original hCF-TG11T5 minigene in the central region of wild-type exon 9 between the EcoRI (which was previously created by site-directed mutagenesis) and BamHI sites which were substituted with the appropriate EcoRI-BamHI cassettes created by PCR-mediated site directed mutagenesis using the following oligonucleotides:

G424S dir (5'-aacttctaatagtgatgacag-3');

G424S rev (5'-ctgtcatcactattagaagtt-3');

I444S dir (5'-ctgtcctgaaagatagtaatt-3');

I444 rev (5'-aattactatctttcaggacag-3');

D443Y dir (5'-ctgtcctgaaatatattaatt-3');

D443Y rev (5'-aattaatatatttcaggacag-3');

A455E rev (5'-gtggatccagcaacctccaacaac-3');

V456F rev (5'-gtggatccagcaarcgccaacaac-3');

153 R rev (5'-gtggatccagcaaccgraacaac-3');

154 R rev (5'-gtggatccagcaaccgraacaac-3');

155 M rev (5'-gtggatccagcaaccmccaacaac-3');

156 R rev (5'-gtggatccagcaacrgccaacaac-3').

The preparation of the I444S and A445E minigenes with different number of TG and T repeats at the polymorphic locus was done by inserting the EcoRI/BamHI cassette of these two mutants in the hybrid minigenes hCF-TG13T5 and hCF-TG11T7 (152). The exon 9 sequences and flanking intronic junctions of all hybrid minigenes were verified by sequence analysis.

### **10.2.10 Preparation of TDP43 expression vector**

Mammalian expression vector for TDP-43 was obtained by subcloning in pRc-CMV vector the 2743 bp TDP-43 cDNA sequence (Resource Center and Primary Database, Berlin, Germany). The cDNA was extracted from pAMP1 vector using NotI-ApaI restriction enzymes.

### **10.2.11 Maintenance and analysis of cells in culture**

Hep3B cells line were maintained in 100x20 mm Falcon tissue culture plates, incubated at 37°C and with 5% carbon dioxide. This cell line were grown in Dulbecco's modified Eagle's medium supplemented with 4.5gr/l Glucose, 10% foetal calf serum, 50 µg/ml gentamicin and 4 mM glutamine.

Plates containing a confluent monolayer of cells were passaged 1 in 3-5 with 1% trypsin as follows. Cells were washed with 5 ml PBS solution, then were incubated at room temperature with 1 ml PBS/EDTA/trypsin solution (PBS containing 0.02% w/v EDTA and 2% v/v trypsin solution) for 2 minutes or until cells were dislodged. After adding 10 ml of media cells were pelleted by centrifugation in polypropylene tubes and were resuspended in 5 ml pre-warmed medium. 1-2 ml of this cell suspension was added to 10 ml medium in fresh plate and was gently mixed before incubation. This procedure was required on average once every three days.

### **10.2.12 Transfection of recombinant DNA into cells maintained in culture**

Cells were passaged as above into 6 wells tissue culture plates and grown to a confluency of 40-70%.

The transfection was performed by DOTAP Liposomal Transfection Reagent (Roche Diagnostics GmbH, Mannheim, Germany); 3 $\mu$ g of expression plasmid was mixed with 20 $\mu$ g of lipofectin and the mixture was incubated at room temperature for 15 min. to allow the formation of DNA-liposome complexes. The mixture was added to the cells in 2 ml of serum free culture medium and incubated at 37° for up to 48 hours. The medium was then collected and the cells were washed with PBS. RNA isolation followed as described.

In cotransfection experiments, cells were transfected with 3 $\mu$ g of the minigene variants and different amount (indicated in each figure) of the plasmid carrying cDNA coding for ASF/SF2 (this plasmid was kindly provided by dr. J. Caceres), TDP43 and PTB (this plasmid was kindly provided by dr. M. G. Romanelli) .

In each transfection experiment as a control were used 250 ng of the empty vectors corresponding to each factor (SF2/ASF, TDP43 and PTB).

Antisense Phosphorothioate oligodeoxynucleotides (PS) were synthesized by MWG-Biotech: TIO7 (5'-accaagcgcagcccagcca-3'); TIO86 (5'-ccgaatatattcagacatct-3'); TIO155 (5'-gagagcagcaccgtcccatc-3'); TIO1318 (5'-ctgtctacattcccagcca-3') and FN56 (5'-gtcaccgcactcgatatccag-3'. Hep3B 70-80% confluent cells were cotransfected with 3  $\mu$ g of minigenes hCF-TG13T5 and different amounts of PS oligodeoxynucleotide using DOTAP. 28h. post-transfection total RNA was extracted and protein lysates (30  $\mu$ g) were used in Western blot assay.

Each transfection experiment was repeated at least 3 times.

Also, 200 ng of vector carrying the gene encoding the human growth hormone (hGH) was cotransfected in each transfection experiment,

whose expression was quantitatively measured by hGH ELISA and used to normalise the transfection efficiency.

### **10.2.13 Measurement of the transfection efficiency**

The 48 hours after transfection, the medium was collected and colorimetric enzyme assay for the quantitative determination of secreted human growth hormone was performed using hGH ELISA reagent kit provided by (Roche Diagnostics GmbH, Mannheim, Germany). The 200 $\mu$ l of diluted medium was loaded to the microtiter plate and incubated for 1 hour at 37°C to allow binding of the protein to the plate. The unbound proteins were then removed by five washings with washing buffer. A 200 $\mu$ l of working dilution of anti-hGH-DIG was added in each well following incubation of 1h at 37°. After five washings, 200 $\mu$ l of anti-DIG-POD antibody dilution was added and incubated for 1h at 37°. After repeated washings, POD substrate was added and Photometric measurements were performed at 405nm.

### **10.2.14 mRNA Analysis by Polymerase Chain Reaction**

#### **10.2.14.1 cDNA synthesis**

In order to synthesise cDNA, the 3  $\mu$ g of total RNA from cells or 5  $\mu$ g of total RNA from tissues were mixed with 100ng random primer mixture and a sterile water in final volume of 20  $\mu$ l. After denaturation at 95°C specific buffer (10mM Tris-HCL (pH8.4), 50mM KCl, 2.5mM, MgCl<sub>2</sub>) 10mM DTT, 1mM of each dNTPs, RNAsin 0.1U/ $\mu$ l and 200U of

Moloney murine leukemia virus reverse transcriptase (GIBCO Brl.) were added to reaction mixture. The reaction was incubated for 1 hour at 37°C. 1.5-3 µl of the cDNA reaction mix was used for the PCR analysis.

#### **10.2.14.2 PCR analysis**

PCRs were carried out for 35 cycles (45 sec at 94°C, 45 sec at 56°C, 45 sec at 72°C) in 50 µl reaction volumes. For radioactive amplifications, a 0.1µl of  $\alpha$ -<sup>32</sup>P-dCTP (1 µCi) was added to PCR reaction. Oligonucleotides used were those specific for the minigene system, placed at the a-globin EDB-1 and a-globin EDB+1 junctions (152): the oligo 2-3 a 5'-CAACTTCAAGCTCCTAAGCCACTGC-3' and B2 5'-TAGGATCCGGTCACCAGGAAGTTGGTTAAATCA-3'

### **10.2.15 Analysis of the PCR-Amplified Products**

#### **10.2.15.1 Not radiolabelled PCR products**

3 µl of each amplification reaction were then analysed by electrophoresis on a 1.5% (w/v) agarose gel, which was stained with ethidium bromide and subject to densitometric analysis. Densitometric analysis of PCR amplicons was carried out using Macintosh version of the public domain NIH Image 1.62 program (developed at the U.S. National Institutes of Health and available on the Internet at <http://rsb.info.nih.gov/nih-image>). The optical density assigned to each band was then normalised on the basis of the length of the product. Densitometry data were confirmed by Phosphoimager quantitation of radioactive PCRs.



#### **10.2.15.2 Radiolabelled PCR products**

PCRs were optimized to remain in the exponential range of amplification.

Aliquots of PCR samples after 27, 30, 33, 35 and 38 cycles were collected and loaded on the 8% denaturing polyacrilamide gel. The gel was dried and then exposed signals measured by the PhosphorImager (Instant Imager, Packard Instrument Co., Meriden, CT) At any time, the ratio 9+/9- was constant, so confirming that the amplification of splice variants is not affected by the size of the amplicons and that, at cycle number 35. The PCRs were still in the exponential phase.

#### **10.2.16 Radioactive labeling of the RNA**

Plasmids used for the UV cross-linking were linearised by digestion with appropriate restriction enzyme. Transcription with T7 RNA polymerase (Stratagene) was performed using 1-2  $\mu\text{g}$  of linearised plasmid, in the presence of  $\alpha\text{-}^{32}\text{P}\text{-UTP}$ , following the manufacturers instructions. Labeled RNA was then purified on a Nick column (Pharmacia), precipitated and resuspended in RNase-free water. The specific activities were in the range of  $4 \times 10^6$  c.p.m./ $\mu\text{g}$  of RNA.

#### **10.2.17 UV cross-linking**

HeLa nuclear extract were prepared from HeLa cells grown in suspension with 10% calf serum as described (233). Total protein concentration was then measured by using Bio-Rad protein assay. The UV cross-linking assays were performed by adding  $\alpha\text{-}^{32}\text{P}\text{-UTP}$  labeled RNA probes ( $1 \times 10^6$  cpm per incubation) in a water bath for 15 min at

30°C with 20 µg of the protein extracts in 30µl of final volume. Final binding conditions were 20mM Hepes pH7.9, 72mM KCl, 1.5mM MgCl<sub>2</sub>, 0.78mM magnesium acetate, 0.52mM DTT, 3.8% glycerol, 0,75mM ATP and 1mM GTP and 2µg of E.coli tRNA as a non-specific competitor. In competition experiments cold RNA (20 and 40 fold molar amount) was also added as a competitor 5 min. before addition of the labeled RNAs. Samples were then transferred in the wells of an HLA plate (Nunc, InterMed) and irradiated with UV light on ice (at 800 000kJ for 5 min) using NIO-LINK (Euroclone). Unbound RNA was then digested with 30 µg of RNaseA1 by the incubation at 37°C for 30min in a water bath. Samples were then analysed by 10% sodium-dodecyl-sulfate polyacrylamide gel electrophoresis (SDS-PAGE) followed by autoradiography.

### 10.2.18 Preparations of competitors

The T7SW Eco competitor was prepared cutting and religating the pBSCFNde construct with Eco RI and Kpn I enzymes. pBIND competitor was prepared cloning the Eco RI-Bam HI fragment of exon 9 in pBS construct under the control of T7 promoter. The h3'int competitor was prepared cutting and religating pBSCFNde construct with Sac I and Bam HI enzymes. The competitors in the intron 9 that span the first 77 nucleotides were prepared directly cloning the PCR product (direct primer: hCFex9/in9 dir; reverse primer IN9Hind/Kpn: 5'-ggaagcttccaaaagcttcagcac-3') cut with Bam HI and Hind III enzymes in PBS vector under the control of T7 promoter. The pEND competitor was prepared digesting ΔM3 PCR product with Pst I and Kpn I and cloning the insert in pBS under the control of T7 promoter. T7SW Eco was linearised with Eco RI enzyme. PBIND was was linearised with

Hind III enzyme. h3'int was linearised with Nde I enzyme. All the competitor spanning the first 77 nucleotides of intron 9 were linearised with Hind III enzyme. PEND competitor was linearised with Nde I enzyme. Transcription with T7 RNA polymerase (Stratagene) was performed using 1-2  $\mu\text{g}$  of linearised plasmid.

### **10.2.19 Cross-linking of RNA to adipic dehydrazide agarose beads for affinity purification of cellular factors**

500 pmoles of RNA (approx. 15  $\mu\text{g}$  of a 100mer RNA) were placed in a 400  $\mu\text{l}$  reaction mixture containing 100 mM NaOAC pH 5.0 and 5 mM sodium m-periodate (Sigma). The reaction mixtures were incubated for 1 hour in the dark at room temperature. The RNA was then ethanol precipitated and resuspended in 500  $\mu\text{l}$  of 0.1 M NaOAC, pH 5.0. Then, 400  $\mu\text{l}$  of adipic acid dehydrazide agarose bead 50% slurry (Sigma) were washed 4 times in 10 ml of 0.1 M NaOAC pH 5.0 and pelleted after each wash at 300 rpm for 3 minutes in a clinical centrifuge. After the final wash, 300  $\mu\text{l}$  of 0.1 M NaOAC pH 5.0 were added to the beads. The slurry was then mixed with the periodate-treated RNA and incubated for 12 hours at 4°C on a rotator. The beads with the bound RNA were pelleted and washed 3 times in 2ml of 2M NaCl and 3 times in 3ml of buffer D (20 mM HEPES-KOH, pH 7, 6.5% v/v glycerol, 0.1 M KCl, 0.2 mM EDTA, 0.5 mM DTT). The beads containing immobilized RNA were incubated with a protein mixture containing 0.6 mg of HeLa cell nuclear extract for 20 minutes at 30°C in 650  $\mu\text{l}$  final volume. Beads were then pelleted by centrifugation at 1000 rpm for 3 minutes and washed 5 times with 5ml of buffer D containing 4 mM  $\text{MgCl}_2$ . After the final centrifugation the proteins bound to the RNA on the beads were eluted by addition of 60  $\mu\text{l}$  of protein sample buffer and heating for 5 minutes

at 90°C. Proteins were then separated at the 10% SDS-PAGE. Band corresponding to studied protein was cut and sent for a sequencing.

### **10.2.20 Expression of recombinant TDP-43 as GST fusion proteins**

The full coding sequence of TDP-43 was amplified by PCR following RT-PCR from HeLa total RNA using the two forward and reverse primers: 5'atgtctgaatatattcgggtaaccga3' and 5'ctacattccccagccagaagactta3'. The central portion of TDP-43 (aa. 101 to 261) was amplified using the following reverse and forward oligos: 5'cagaaaacatccgattta3' and 5'ctattcggcattggatatatg3'. Both PCR products were cloned into the SmaI site of plasmid pGEX-3X (Pharmacia) and the resulting recombinant proteins were expressed and purified with Glutathione S Sepharose 4B beads (Pharmacia) according to manufacturer's instructions.

### **10.2.21 Immunoprecipitation assays with PTB and hnRNPH**

Polyclonal antiserum against PTB and hnRNP H were obtained by immunizing a 3 months old rabbit (New Zealand strain) according to standard protocols (Kindly provided by E. Buratti). Antibodies were subsequently bound to protein A/G PLUS-Agarose beads (Santa Cruz Biotechnology) following 1 hour incubation at RT in IP buffer (20mM Tris pH=8.0, 300mM NaCl, 1 mM EDTA, 0.25% NP-40). The beads were then divided in 20µl aliquots and incubated for 2 hours at RT with different UV cross-linked samples (in a final volume of 1ml IP buffer).

Each sample was washed 5 times with 1ml IP buffer before addition of SDS-loading buffer and loaded on a 10% SDS-PAGE gel

#### **10.2.22 Northern Blot of TDP-43 and SF2/ASF**

Northern blot analyses were performed on commercially available human poly(A)<sup>+</sup> Northern Blots (Clontech). The probe for TDP-43 was a uniformly labelled EcoRI-NdeI DNA fragment corresponding to residues 59 to 165 of its coding region. The probe for SF2/ASF was obtained by amplifying the entire full cDNA coding sequence. Each blot was hybridised at 68°C and washed under standard conditions. A Camberra Packard Instant Imager was then used to quantify the mRNA levels and a GAPDH probe was used to normalise for different loadings of RNA in each lane.

#### **10.2.23 Western Blot Assay on Hep 3B protein extract**

Cells were washed two times with PBS and collected from the plates. Cell lysates were obtained by sonication using a Soniprep 150 sonicator. Each sample was sonicated two times at 10 amplitude microns. Protein extracts were quantitated by Bradford method. 20 µg of protein extract was added to protein sample buffer (2X final concentration), boiled 5 min, chilled and run on a 10% SDS-PAGE gel. The proteins were then transferred to nitrocellulose membranes for 2.5 hours at 200mA in the presence of 25 mM Tris-HCl (pH6.8), 192 mM glycine, 5% methanol. Following a blocking step of 1 h in the appropriate buffer, the membranes were washed and incubated for 1 h with the primary antibody. Then the blots were washed again and incubated 1 h with the secondary antibody conjugated with

horseradish peroxidase (1:2000), washed again and developed using the Enhanced Chemi-Luminescent (ECL) system (Amersham Pharmacia Biotech, Sweden). All washes were carried out during 10 min (2-3 solution changes).

#### **10.2.24 Characterisation of the TG13T3 mutation**

Genomic DNA was extracted from EDTA blood samples of patient and family members by a routine salting out procedure. Mutation screening of the CFTR gene was performed by SSCP analysis and direct sequencing as described (234). Lymphoblastoid cell lines were established by EBV transformation. RT-PCR was carried out using 5 $\mu$ g of total RNA with fluorescein-labelled primers: forward 5'-cagaagtagtgatggagaatgtaac-3' (exon 8), reverse 5'-gttgacctcactcagtgattc-3' (exon 11) or 5'-ttcatcataggaacaccaaag-3' (exon 10, codon 508 specific). 35 cycles of PCR were performed with an annealing at 62°C for 1 min, extension at 72°C for 75 sec and denaturation at 94°C for 45 sec. For semiquantitative analysis of exon 9 skipping, RT-PCR products were separated on either 8 % or 10 % denaturing polyacrylamide gels using an A.L.F. sequencer and fluorescence signals were quantified using the Fragment Manager software (Pharmacia).

## REFERENCES

1. Welsh M., T. L.-C., Boat T., Beaudet A. 1995. Cystic fibrosis. *In* The metabolic bases of inherited disease. M. G. Hill, editor, New York.
2. Tsui, L. C., and M. Buchwald. (1991). Biochemical and molecular genetics of cystic fibrosis. *Adv Hum Genet.* **20**:153-266.
3. Kopelman, H., P. Durie, K. Gaskin, Z. Weizman, and G. Forstner. (1985). Pancreatic fluid secretion and protein hyperconcentration in cystic fibrosis. *N Engl J Med.* **312**:329-34.
4. Kerem, E., M. Corey, B. Kerem, P. Durie, L. C. Tsui, and H. Levison. (1989). Clinical and genetic comparisons of patients with cystic fibrosis, with or without meconium ileus. *J Pediatr.* **114**:767-73.
5. Kerem, B. S., J. A. Buchanan, P. Durie, M. L. Corey, H. Levison, J. M. Rommens, M. Buchwald, and L. C. Tsui. (1989). DNA marker haplotype association with pancreatic sufficiency in cystic fibrosis. *Am J Hum Genet.* **44**:827-34.
6. Riordan, J. R., J. M. Rommens, B. Kerem, N. Alon, R. Rozmahel, Z. Grzelczak, J. Zielenski, S. Lok, N. Plavsic, J. L. Chou, and et al. (1989). Identification of the cystic fibrosis gene: cloning and characterization of complementary DNA. *Science.* **245**:1066-73.
7. Boat, T. F., and P. W. Cheng. (1989). Epithelial cell dysfunction in cystic fibrosis: implications for airways disease. *Acta Paediatr Scand Suppl.* **363**:25-9.
8. Buckle, V. J., P. J. Scambler, and B. J. Wainwright. (1987). Localisation of a sequence, 7C22, showing close linkage to the cystic fibrosis locus. *Cytogenet Cell Genet.* **44**:41-2.
9. Kerem, B., J. M. Rommens, J. A. Buchanan, D. Markiewicz, T. K. Cox, A. Chakravarti, M. Buchwald, and L. C. Tsui. (1989). Identification of the cystic fibrosis gene: genetic analysis. *Science.*



245:1073-80.

10. Yoshimura, K., H. Nakamura, B. C. Trapnell, W. Dalemans, A. Pavirani, J. P. Lecocq, and R. G. Crystal. (1991). The cystic fibrosis gene has a "housekeeping"-type promoter and is expressed at low levels in cells of epithelial origin. *J Biol Chem.* **266**:9140-4.
11. Gregory, R. J., S. H. Cheng, D. P. Rich, J. Marshall, S. Paul, K. Hehir, L. Ostedgaard, K. W. Klinger, M. J. Welsh, and A. E. Smith. (1990). Expression and characterization of the cystic fibrosis transmembrane conductance regulator. *Nature.* **347**:382-6.
12. Hyde, S. C., P. Emsley, M. J. Hartshorn, M. M. Mimmack, U. Gileadi, S. R. Pearce, M. P. Gallagher, D. R. Gill, R. E. Hubbard, and C. F. Higgins. (1990). Structural model of ATP-binding proteins associated with cystic fibrosis, multidrug resistance and bacterial transport. *Nature.* **346**:362-5.
13. Anderson, M. P., D. P. Rich, R. J. Gregory, A. E. Smith, and M. J. Welsh. (1991). Generation of cAMP-activated chloride currents by expression of CFTR. *Science.* **251**:679-82.
14. Anderson, M. P., R. J. Gregory, S. Thompson, D. W. Souza, S. Paul, R. C. Mulligan, A. E. Smith, and M. J. Welsh. (1991). Demonstration that CFTR is a chloride channel by alteration of its anion selectivity. *Science.* **253**:202-5.
15. Berger, H. A., M. P. Anderson, R. J. Gregory, S. Thompson, P. W. Howard, R. A. Maurer, R. Mulligan, A. E. Smith, and M. J. Welsh. (1991). Identification and regulation of the cystic fibrosis transmembrane conductance regulator-generated chloride channel. *J Clin Invest.* **88**:1422-31.
16. Kartner, N., J. W. Hanrahan, T. J. Jensen, A. L. Naismith, S. Z. Sun, C. A. Ackerley, E. F. Reyes, L. C. Tsui, J. M. Rommens, C. E. Bear, and et al. (1991). Expression of the cystic fibrosis gene

in non-epithelial invertebrate cells produces a regulated anion conductance. *Cell*. **64**:681-91.

17. Tabcharani, J. A., J. M. Rommens, Y. X. Hou, X. B. Chang, L. C. Tsui, J. R. Riordan, and J. W. Hanrahan. (1993). Multi-ion pore behaviour in the CFTR chloride channel. *Nature*. **366**:79-82.

18. Akabas, M. H., C. Kaufmann, T. A. Cook, and P. Archdeacon. (1994). Amino acid residues lining the chloride channel of the cystic fibrosis transmembrane conductance regulator. *J Biol Chem*. **269**:14865-8.

19. McDonough, S., N. Davidson, H. A. Lester, and N. A. McCarty. (1994). Novel pore-lining residues in CFTR that govern permeation and open-channel block. *Neuron*. **13**:623-34.

20. Travis, S. M., M. R. Carson, D. R. Ries, and M. J. Welsh. (1993). Interaction of nucleotides with membrane-associated cystic fibrosis transmembrane conductance regulator. *J Biol Chem*. **268**:15336-9.

21. Carson, M. R., S. M. Travis, and M. J. Welsh. (1995). The two nucleotide-binding domains of cystic fibrosis transmembrane conductance regulator (CFTR) have distinct functions in controlling channel activity. *J Biol Chem*. **270**:1711-7.

22. Gunderson, K. L., and R. R. Kopito. (1995). Conformational states of CFTR associated with channel gating: the role ATP binding and hydrolysis. *Cell*. **82**:231-9.

23. Winter, M. C., and M. J. Welsh. (1997). Stimulation of CFTR activity by its phosphorylated R domain. *Nature*. **389**:294-6.

24. Welsh, M. J., and A. E. Smith. (1993). Molecular mechanisms of CFTR chloride channel dysfunction in cystic fibrosis. *Cell*. **73**:1251-4.

25. Dulhanty, A. M., and J. R. Riordan. (1994). Phosphorylation by cAMP-dependent protein kinase causes a

conformational change in the R domain of the cystic fibrosis transmembrane conductance regulator. *Biochemistry*. **33**:4072-9.

26. Gadsby, D. C., and A. C. Nairn. (1994). Regulation of CFTR channel gating. *Trends Biochem Sci*. **19**:513-8.

27. Cantiello, H. F., A. G. Prat, I. L. Reisin, L. B. Ercole, E. H. Abraham, J. F. Amara, R. J. Gregory, and D. A. Ausiello. (1994). External ATP and its analogs activate the cystic fibrosis transmembrane conductance regulator by a cyclic AMP-independent mechanism. *J Biol Chem*. **269**:11224-32.

28. Ikuma, M., and M. J. Welsh. (2000). Regulation of CFTR Cl- channel gating by ATP binding and hydrolysis. *Proc Natl Acad Sci U S A*. **97**:8675-80.

29. Karpowich, N., O. Martsinkevich, L. Millen, Y. R. Yuan, P. L. Dai, K. MacVey, P. J. Thomas, and J. F. Hunt. (2001). Crystal structures of the MJ1267 ATP binding cassette reveal an induced-fit effect at the ATPase active site of an ABC transporter. *Structure (Camb)*. **9**:571-86.

30. Trezise, A. E., and M. Buchwald. (1991). In vivo cell-specific expression of the cystic fibrosis transmembrane conductance regulator. *Nature*. **353**:434-7.

31. Barinaga, M. (1992). Novel function discovered for the cystic fibrosis gene [news; comment]. *Science*. **256**:444-5.

32. Drumm, M. L., D. J. Wilkinson, L. S. Smit, R. T. Worrell, T. V. Strong, R. A. Frizzell, D. C. Dawson, and F. S. Collins. (1991). Chloride conductance expressed by delta F508 and other mutant CFTRs in *Xenopus* oocytes. *Science*. **254**:1797-9.

33. Rommens, J. M., M. C. Iannuzzi, B. Kerem, M. L. Drumm, G. Melmer, M. Dean, R. Rozmahel, J. L. Cole, D. Kennedy, N. Hidaka, and et al. (1989). Identification of the cystic fibrosis gene: chromosome walking and jumping. *Science*. **245**:1059-65.

34. Denning, G. M., L. S. Ostedgaard, S. H. Cheng, A. E.

- Smith, and M. J. Welsh. (1992). Localization of cystic fibrosis transmembrane conductance regulator in chloride secretory epithelia. *J Clin Invest.* **89**:339-49.
35. Puchelle, E., D. Gaillard, D. Ploton, J. Hinrasky, C. Fuchey, M. C. Boutterin, J. Jacquot, D. Dreyer, A. Pavirani, and W. Dalemans. (1992). Differential localization of the cystic fibrosis transmembrane conductance regulator in normal and cystic fibrosis airway epithelium. *Am J Respir Cell Mol Biol.* **7**:485-91.
36. Lukacs, G. L., X. B. Chang, N. Kartner, O. D. Rotstein, J. R. Riordan, and S. Grinstein. (1992). The cystic fibrosis transmembrane regulator is present and functional in endosomes. Role as a determinant of endosomal pH. *J Biol Chem.* **267**:14568-72.
37. Barasch, J., B. Kiss, A. Prince, L. Saiman, D. Gruenert, and Q. al-Awqati. (1991). Defective acidification of intracellular organelles in cystic fibrosis. *Nature.* **352**:70-3.
38. Bradbury, N. A., T. Jilling, G. Berta, E. J. Sorscher, R. J. Bridges, and K. L. Kirk. (1992). Regulation of plasma membrane recycling by CFTR. *Science.* **256**:530-2.
39. Ismailov, II, M. S. Awayda, B. Jovov, B. K. Berdiev, C. M. Fuller, J. R. Dedman, M. Kaetzel, and D. J. Benos. (1996). Regulation of epithelial sodium channels by the cystic fibrosis transmembrane conductance regulator. *J Biol Chem.* **271**:4725-32.
40. Tsui, L. C. (1992). The spectrum of cystic fibrosis mutations. *Trends Genet.* **8**:392-8.
41. Cheng, S. H., R. J. Gregory, J. Marshall, S. Paul, D. W. Souza, G. A. White, C. R. O'Riordan, and A. E. Smith. (1990). Defective intracellular transport and processing of CFTR is the molecular basis of most cystic fibrosis. *Cell.* **63**:827-34.
42. Anderson, M. P., and M. J. Welsh. (1992). Regulation by ATP and ADP of CFTR chloride channels that contain mutant

nucleotide-binding domains. *Science*. 257:1701-4.

43. Romey, M. C., C. Guittard, S. Carles, J. Demaille, M. Claustres, and M. Ramsay. (1999). First putative sequence alterations in the minimal CFTR promoter region. *J Med Genet*. 36:263-4.

44. Chu, C. S., B. C. Trapnell, S. Curristin, G. R. Cutting, and R. G. Crystal. (1993). Genetic basis of variable exon 9 skipping in cystic fibrosis transmembrane conductance regulator mRNA. *Nat Genet*. 3:151-6.

45. Chiba-Falek, O., E. Kerem, T. Shoshani, M. Aviram, A. Augarten, L. Bentur, A. Tal, E. Tullis, A. Rahat, and B. Kerem. (1998). The molecular basis of disease variability among cystic fibrosis patients carrying the 3849+10 kb C-->T mutation. *Genomics*. 53:276-83.

46. Fulmer, S. B., E. M. Schwiebert, M. M. Morales, W. B. Guggino, and G. R. Cutting. (1995). Two cystic fibrosis transmembrane conductance regulator mutations have different effects on both pulmonary phenotype and regulation of outwardly rectified chloride currents. *Proc Natl Acad Sci U S A*. 92:6832-6.

47. Zielenski, J., D. Markiewicz, S. P. Lin, F. Y. Huang, T. L. Yang-Feng, and L. C. Tsui. (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*. 47:125-32.

48. Zielenski, J. (2000). Genotype and phenotype in cystic fibrosis. *Respiration*. 67:117-33.

49. Chillon, M., T. Casals, B. Mercier, L. Bassas, W. Lissens, S. Silber, M. C. Romey, J. Ruiz-Romero, C. Verlingue, M. Claustres, and et al. (1995). Mutations in the cystic fibrosis gene in patients with congenital absence of the vas deferens. *N Engl J Med*.

332:1475-80.

50. Irving, R. M., R. McMahon, R. Clark, and N. S. Jones. (1997). Cystic fibrosis transmembrane conductance regulator gene mutations in severe nasal polyposis. *Clin Otolaryngol.* **22**:519-21.
51. Cockrill, B. A., and C. A. Hales. (1999). Allergic bronchopulmonary aspergillosis. *Annu Rev Med.* **50**:303-16.
52. Cohn, J. A., K. J. Friedman, P. G. Noone, M. R. Knowles, L. M. Silverman, and P. S. Jowell. (1998). Relation between mutations of the cystic fibrosis gene and idiopathic pancreatitis. *N Engl J Med.* **339**:653-8.
53. Barreto, C., L. M. Pinto, A. Duarte, J. Lavinha, and M. Ramsay. (1991). A fertile male with cystic fibrosis: molecular genetic analysis. *J Med Genet.* **28**:420-1.
54. Chiba-Falek, O., R. B. Parad, E. Kerem, and B. Kerem. (1999). Variable levels of normal RNA in different fetal organs carrying a cystic fibrosis transmembrane conductance regulator splicing mutation. *Am J Respir Crit Care Med.* **159**:1998-2002.
55. Berget, S. M., C. Moore, and P. A. Sharp. (1977). Spliced segments at the 5' terminus of adenovirus 2 late mRNA. *Proc Natl Acad Sci U S A.* **74**:3171-5.
56. Chow, L. T., R. E. Gelinis, T. R. Broker, and R. J. Roberts. (1977). An amazing sequence arrangement at the 5' ends of adenovirus 2 messenger RNA. *Cell.* **12**:1-8.
57. Dreyfuss, G., M. J. Matunis, S. Pinol-Roma, and C. G. Burd. (1993). hnRNP proteins and the biogenesis of mRNA. *Annu Rev Biochem.* **62**:289-321.
58. Moore MJ, Q. C., Sharp PA. 1993. Splicing of precursor to mRNA by the spliceosome. *In* RNA world. A. J. Gesteland RF, editor. Cold Spring Harbor Laboratory Press. 303-357.
59. Will CL, L. R. 1997. snRNP structure and function. *In*

Eukaryotic mRNA processing. O. I. P. Krainer AR, editor, Oxford. 130-173.

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

61. Madhani, H. D., and C. Guthrie. (1994). Dynamic RNA-RNA interactions in the spliceosome. *Annu Rev Genet*. **28**:1-26.

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

63. Shapiro, M. B., and P. Senapathy. (1987). RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. *Nucleic Acids Res*. **15**:7155-74.

64. Aebi, M., H. Hornig, and C. Weissmann. (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**:237-46.

65. Reed, R. (1989). The organization of 3' splice-site sequences in mammalian introns. *Genes Dev*. **3**:2113-23.

66. Nelson, K. K., and M. R. Green. (1989). Mammalian U2 snRNP has a sequence-specific RNA-binding activity. *Genes Dev*. **3**:1562-71.

67. Burge CB, T. T., Sharp PA. 1999. Splicing of precursor to mRNAs by the spliceosome. *In* The RNA World, edn. 2. C. T. Gesteland RF, Atkins JF. New York: Cold Spring Harbor Laboratory Press, editor.

68. Kambach, C., S. Walke, and K. Nagai. (1999). Structure and assembly of the spliceosomal small nuclear ribonucleoprotein particles. *Curr Opin Struct Biol*. **9**:222-30.

69. Fischer, U., V. Sumpster, M. Sekine, T. Satoh, and R. Luhrmann. (1993). Nucleo-cytoplasmic transport of U snRNPs: definition of a nuclear location signal in the Sm core domain that

binds a transport receptor independently of the Py3G cap. *Embo J.* **12**:573-83.

70. Konarska, M. M., and P. A. Sharp. (1987). Interactions between small nuclear ribonucleoprotein particles in formation of spliceosomes. *Cell.* **49**:763-74.

71. Raker, V. A., G. Plessel, and R. Luhrmann. (1996). The snRNP core assembly pathway: identification of stable core protein heteromeric complexes and an snRNP subcore particle in vitro. *Embo J.* **15**:2256-69. [abs.html](#).

72. Mayes, A. E., L. Verdone, P. Legrain, and J. D. Beggs. (1999). Characterization of Sm-like proteins in yeast and their association with U6 snRNA. *Embo J.* **18**:4321-31.

73. Achsel, T., H. Brahm, B. Kastner, A. Bachi, M. Wilm, and R. Luhrmann. (1999). A doughnut-shaped heteromer of human Sm-like proteins binds to the 3'- end of U6 snRNA, thereby facilitating U4/U6 duplex formation in vitro. *Embo J.* **18**:5789-802.

74. Brosi, R., H. P. Hauri, and A. Kramer. (1993). Separation of splicing factor SF3 into two components and purification of SF3a activity. *J Biol Chem.* **268**:17640-6.

75. Gozani, O., R. Feld, and R. Reed. (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**:233-43.

76. Champion-Arnaud, P., and R. Reed. (1994). The prespliceosome components SAP 49 and SAP 145 interact in a complex implicated in tethering U2 snRNP to the branch site. *Genes Dev.* **8**:1974-83.

77. Teigelkamp, S., T. Achsel, C. Mundt, S. F. Gothel, U. Cronshagen, W. S. Lane, M. Marahiel, and R. Luhrmann. (1998). The 20kD protein of human [U4/U6.U5] tri-snRNPs is a novel cyclophilin that forms a complex with the U4/U6-specific 60kD



- and 90kD proteins. *Rna*. 4:127-41.
78. Umen, J. G., and C. Guthrie. (1995). The second catalytic step of pre-mRNA splicing. *Rna*. 1:869-85.
79. Ruskin, B., J. M. Greene, and M. R. Green. (1985). Cryptic branch point activation allows accurate in vitro splicing of human beta-globin intron mutants. *Cell*. 41:833-44.
80. Zamore, P. D., and M. R. Green. (1991). Biochemical characterization of U2 snRNP auxiliary factor: an essential pre-mRNA splicing factor with a novel intranuclear distribution. *Embo J*. 10:207-14.
81. Wu, S., C. M. Romfo, T. W. Nilsen, and M. R. Green. (1999). Functional recognition of the 3' splice site AG by the splicing factor U2AF35. *Nature*. 402:832-5.
82. Merendino, L., S. Guth, D. Bilbao, C. Martinez, and J. Valcarcel. (1999). Inhibition of msl-2 splicing by Sex-lethal reveals interaction between U2AF35 and the 3' splice site AG. *Nature*. 402:838-41.
83. Fu, X. D. (1995). The superfamily of arginine/serine-rich splicing factors. *Rna*. 1:663-80.
84. Zahler, A. M., W. S. Lane, J. A. Stolk, and M. B. Roth. (1992). SR proteins: a conserved family of pre-mRNA splicing factors. *Genes Dev*. 6:837-47.
85. Graveley, B. R. (2000). Sorting out the complexity of SR protein functions. *Rna*. 6:1197-211.
86. Kawano, T., M. Fujita, and H. Sakamoto. (2000). Unique and redundant functions of SR proteins, a conserved family of splicing factors, in *Caenorhabditis elegans* development. *Mech Dev*. 95:67-76.
87. Longman, D., I. L. Johnstone, and J. F. Caceres. (2000). Functional characterization of SR and SR-related genes in *Caenorhabditis elegans*. *Embo J*. 19:1625-37.

88. Fu, X. D. (1993). Specific commitment of different pre-mRNAs to splicing by single SR proteins. *Nature*. **365**:82-5.
89. Gui, J. F., W. S. Lane, and X. D. Fu. (1994). A serine kinase regulates intracellular localization of splicing factors in the cell cycle. *Nature*. **369**:678-82.
90. Roth, M. B., A. M. Zahler, and J. A. Stolk. (1991). A conserved family of nuclear phosphoproteins localized to sites of polymerase II transcription. *J Cell Biol*. **115**:587-96.
91. Misteli, T., J. F. Caceres, J. Q. Clement, A. R. Krainer, M. F. Wilkinson, and D. L. Spector. (1998). Serine phosphorylation of SR proteins is required for their recruitment to sites of transcription in vivo. *J Cell Biol*. **143**:297-307.
92. Reed R, P. L. 1997. Spliceosome assembly. *In* Eukaryotic mRNA Processing. O. I. P. Krainer AR, editor. 103-129.
93. Zhang, D., and M. Rosbash. (1999). Identification of eight proteins that cross-link to pre-mRNA in the yeast commitment complex. *Genes Dev*. **13**:581-92.
94. Puig, O., A. Gottschalk, P. Fabrizio, and B. Seraphin. (1999). Interaction of the U1 snRNP with nonconserved intronic sequences affects 5' splice site selection. *Genes Dev*. **13**:569-80.
95. Collins, C. A., and C. Guthrie. (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**:1970-82.
96. Maroney, P. A., C. M. Romfo, and T. W. Nilsen. (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**:317-28.
97. Steitz, T. A., and J. A. Steitz. (1993). A general two-metal-ion mechanism for catalytic RNA. *Proc Natl Acad Sci U S A*. **90**:6498-502.

98. Sontheimer, E. J., S. Sun, and J. A. Piccirilli. (1997). Metal ion catalysis during splicing of premessenger RNA. *Nature*. **388**:801-5.
99. Yean, S. L., G. Wuenschell, J. Termini, and R. J. Lin. (2000). Metal-ion coordination by U6 small nuclear RNA contributes to catalysis in the spliceosome. *Nature*. **408**:881-4.
100. Nielsen, T. (1998). RNA structure and function.
101. Lang, K. M., and R. A. Spritz. (1983). RNA splice site selection: evidence for a 5' leads to 3' scanning model. *Science*. **220**:1351-5.
102. Goguel, V., and M. Rosbash. (1993). Splice site choice and splicing efficiency are positively influenced by pre-mRNA intramolecular base pairing in yeast. *Cell*. **72**:893-901.
103. Hawkins, J. D. (1988). A survey on intron and exon lengths. *Nucleic Acids Res*. **16**:9893-908.
104. Talerico, M., and S. M. Berget. (1994). Intron definition in splicing of small *Drosophila* introns. *Mol Cell Biol*. **14**:3434-45.
105. Berget, S. M. (1995). Exon recognition in vertebrate splicing. *J Biol Chem*. **270**:2411-4.
106. Robberson, B. L., G. J. Cote, and S. M. Berget. (1990). Exon definition may facilitate splice site selection in RNAs with multiple exons. *Mol Cell Biol*. **10**:84-94.
107. Dominski, Z., and R. Kole. (1991). Selection of splice sites in pre-mRNAs with short internal exons. *Mol Cell Biol*. **11**:6075-83.
108. Black, D. L. (1991). Does steric interference between splice sites block the splicing of a short c-src neuron-specific exon in non-neuronal cells? *Genes Dev*. **5**:389-402.
109. Xu, R., J. Teng, and T. A. Cooper. (1993). The cardiac troponin T alternative exon contains a novel purine-rich positive splicing element. *Mol Cell Biol*. **13**:3660-74.
110. Dominski, Z., and R. Kole. (1994). Identification and

characterization by antisense oligonucleotides of exon and intron sequences required for splicing. *Mol Cell Biol.* **14**:7445-54.

111. Carothers, A. M., G. Urlaub, D. Grunberger, and L. A. Chasin. (1993). Splicing mutants and their second-site suppressors at the dihydrofolate reductase locus in Chinese hamster ovary cells. *Mol Cell Biol.* **13**:5085-98.

112. Hoffman, B. E., and P. J. Grabowski. (1992). U1 snRNP targets an essential splicing factor, U2AF65, to the 3' splice site by a network of interactions spanning the exon. *Genes Dev.* **6**:2554-68.

113. Izaurralde, E., J. Lewis, C. McGuigan, M. Jankowska, E. Darzynkiewicz, and I. W. Mattaj. (1994). A nuclear cap binding protein complex involved in pre-mRNA splicing. *Cell.* **78**:657-68.

114. Gunderson, K. L., and R. R. Kopito. (1994). Effects of pyrophosphate and nucleotide analogs suggest a role for ATP hydrolysis in cystic fibrosis transmembrane regulator channel gating. *J Biol Chem.* **269**:19349-53.

115. Eperon, I. C., D. C. Ireland, R. A. Smith, A. Mayeda, and A. R. Krainer. (1993). Pathways for selection of 5' splice sites by U1 snRNPs and SF2/ASF. *Embo J.* **12**:3607-17.

116. Kohtz, J. D., S. F. Jamison, C. L. Will, P. Zuo, R. Luhrmann, M. A. Garcia-Blanco, and J. L. Manley. (1994). Protein-protein interactions and 5'-splice-site recognition in mammalian mRNA precursors. *Nature.* **368**:119-24.

117. Zuo, P., and J. L. Manley. (1994). The human splicing factor ASF/SF2 can specifically recognize pre-mRNA 5' splice sites. *Proc Natl Acad Sci U S A.* **91**:3363-7.

118. Berglund, J. A., K. Chua, N. Abovich, R. Reed, and M. Rosbash. (1997). The splicing factor BBP interacts specifically with the pre-mRNA branchpoint sequence UACUAAC. *Cell.* **89**:781-7.

119. Wu, J. Y., and T. Maniatis. (1993). Specific interactions

between proteins implicated in splice site selection and regulated alternative splicing. *Cell*. 75:1061-70.

120. Zuo, P., and T. Maniatis. (1996). The splicing factor U2AF35 mediates critical protein-protein interactions in constitutive and enhancer-dependent splicing. *Genes Dev.* 10:1356-68.

121. Reed, R. (1996). Initial splice-site recognition and pairing during pre-mRNA splicing. *Curr Opin Genet Dev.* 6:215-20.

122. Abovich, N., and M. Rosbash. (1997). Cross-intron bridging interactions in the yeast commitment complex are conserved in mammals. *Cell*. 89:403-12.

123. Mount, S. M., C. Burks, G. Hertz, G. D. Stormo, O. White, and C. Fields. (1992). Splicing signals in *Drosophila*: intron size, information content, and consensus sequences. *Nucleic Acids Res.* 20:4255-62.

124. Kennedy, C. F., A. Kramer, and S. M. Berget. (1998). A role for SRp54 during intron bridging of small introns with pyrimidine tracts upstream of the branch point. *Mol Cell Biol.* 18:5425-34.

125. Eldridge, A. G., Y. Li, P. A. Sharp, and B. J. Blencowe. (1999). The SRPy160/300 splicing coactivator is required for exon-enhancer function. *Proc Natl Acad Sci U S A.* 96:6125-30.

126. Blencowe, B. J., R. Issner, J. A. Nickerson, and P. A. Sharp. (1998). A coactivator of pre-mRNA splicing. *Genes Dev.* 12:996-1009.

127. Lopez, A. J. (1998). Alternative splicing of pre-mRNA: developmental consequences and mechanisms of regulation. *Annu Rev Genet.* 32:279-305.

128. Croft, L., S. Schandorff, F. Clark, K. Burrage, P. Arctander, and J. S. Mattick. (2000). ISIS, the intron information system, reveals the high frequency of alternative splicing in the human

genome. *Nat Genet.* **24**:340-1.

129. Schmucker, D., J. C. Clemens, H. Shu, C. A. Worby, J. Xiao, M. Muda, J. E. Dixon, and S. L. Zipursky. (2000). Drosophila Dscam is an axon guidance receptor exhibiting extraordinary molecular diversity. *Cell.* **101**:671-84.

130. Dominski, Z., and R. Kole. (1992). Cooperation of pre-mRNA sequence elements in splice site selection. *Mol Cell Biol.* **12**:2108-14.

131. Sterner, D. A., T. Carlo, and S. M. Berget. (1996). Architectural limits on split genes. *Proc Natl Acad Sci U S A.* **93**:15081-5.

132. Langford, C. J., F. J. Klinz, C. Donath, and D. Gallwitz. (1984). Point mutations identify the conserved, intron-contained TACTAAC box as an essential splicing signal sequence in yeast. *Cell.* **36**:645-53.

133. Huh, G. S., and R. O. Hynes. (1993). Elements regulating an alternatively spliced exon of the rat fibronectin gene. *Mol Cell Biol.* **13**:5301-14.

134. Muro, A. F., M. Caputi, R. Pariyarath, F. Pagani, E. Buratti, and F. E. Baralle. (1999). Regulation of fibronectin EDA exon alternative splicing: possible role of RNA secondary structure for enhancer display. *Mol Cell Biol.* **19**:2657-71.

135. Nakai, K., and H. Sakamoto. (1994). Construction of a novel database containing aberrant splicing mutations of mammalian genes. *Gene.* **141**:171-7.

136. Stamm, S., J. Zhu, K. Nakai, P. Stoilov, O. Stoss, and M. Q. Zhang. (2000). An alternative-exon database and its statistical analysis. *DNA Cell Biol.* **19**:739-56.

137. Mullen, M. P., C. W. Smith, J. G. Patton, and B. Nadal-Ginard. (1991). Alpha-tropomyosin mutually exclusive exon selection: competition between branchpoint/polypyrimidine

- tracts determines default exon choice. *Genes Dev.* 5:642-55.
138. Roscigno, R. F., M. Weiner, and M. A. Garcia-Blanco. (1993). A mutational analysis of the polypyrimidine tract of introns. Effects of sequence differences in pyrimidine tracts on splicing. *J Biol Chem.* 268:11222-9.
139. Norton, P. A. (1994). Polypyrimidine tract sequences direct selection of alternative branch sites and influence protein binding. *Nucleic Acids Res.* 22:3854-60.
140. Coolidge, C. J., R. J. Seely, and J. G. Patton. (1997). Functional analysis of the polypyrimidine tract in pre-mRNA splicing. *Nucleic Acids Res.* 25:888-96.
141. Reed, R., and T. Maniatis. (1985). Intron sequences involved in lariat formation during pre-mRNA splicing. *Cell.* 41:95-105.
142. Smith, C. W., E. B. Porro, J. G. Patton, and B. Nadal-Ginard. (1989). Scanning from an independently specified branch point defines the 3' splice site of mammalian introns. *Nature.* 342:243-7.
143. Reed, R., and T. Maniatis. (1988). The role of the mammalian branchpoint sequence in pre-mRNA splicing. *Genes Dev.* 2:1268-76.
144. Coleman, T. P., and J. R. Roesser. (1998). RNA secondary structure: an important cis-element in rat calcitonin/CGRP pre-messenger RNA splicing. *Biochemistry.* 37:15941-50.
145. Grover, A., H. Houlden, M. Baker, J. Adamson, J. Lewis, G. Prihar, S. Pickering-Brown, K. Duff, and M. Hutton. (1999). 5' splice site mutations in tau associated with the inherited dementia FTDP-17 affect a stem-loop structure that regulates alternative splicing of exon 10. *J Biol Chem.* 274:15134-43.
146. Tian, M., and T. Maniatis. (1994). A splicing enhancer exhibits both constitutive and regulated activities. *Genes Dev.*

8:1703-12.

147. Tian, M., and T. Maniatis. (1993). A splicing enhancer complex controls alternative splicing of doublesex pre-mRNA. *Cell*. **74**:105-14.
148. Graveley, B. R., K. J. Hertel, and T. Maniatis. (1998). A systematic analysis of the factors that determine the strength of pre-mRNA splicing enhancers. *Embo J*. **17**:6747-56.
149. Tian, H., and R. Kole. (1995). Selection of novel exon recognition elements from a pool of random sequences. *Mol Cell Biol*. **15**:6291-8.
150. Schaal, T. D., and T. Maniatis. (1999). Selection and characterization of pre-mRNA splicing enhancers: identification of novel SR protein-specific enhancer sequences. *Mol Cell Biol*. **19**:1705-19.
151. Fairbrother, W. G., and L. A. Chasin. (2000). Human genomic sequences that inhibit splicing. *Mol Cell Biol*. **20**:6816-25.
152. Caputi, M., G. Casari, S. Guenzi, R. Tagliabue, A. Sidoli, C. A. Melo, and F. E. Baralle. (1994). A novel bipartite splicing enhancer modulates the differential processing of the human fibronectin EDA exon. *Nucleic Acids Res*. **22**:1018-22.
153. Pagani, F., E. Buratti, C. Stuardi, M. Romano, E. Zuccato, M. Niksic, L. Giglio, D. Faraguna, and F. E. Baralle. (2000). Splicing factors induce cystic fibrosis transmembrane regulator exon 9 skipping through a nonevolutionary conserved intronic element. *J Biol Chem*. **275**:21041-7.
154. Wagner, E. J., R. P. Carstens, and M. A. Garcia-Blanco. (1999). A novel isoform ratio switch of the polypyrimidine tract binding protein. *Electrophoresis*. **20**:1082-6. [pii].
155. Bai, Y., D. Lee, T. Yu, and L. A. Chasin. (1999). Control of 3' splice site choice in vivo by ASF/SF2 and hnRNP A1. *Nucleic Acids Res*. **27**:1126-34.



156. Krainer, A. R., G. C. Conway, and D. Kozak. (1990). The essential pre-mRNA splicing factor SF2 influences 5' splice site selection by activating proximal sites. *Cell*. **62**:35-42.
157. Staknis, D., and R. Reed. (1994). SR proteins promote the first specific recognition of Pre-mRNA and are present together with the U1 small nuclear ribonucleoprotein particle in a general splicing enhancer complex. *Mol Cell Biol*. **14**:7670-82.
158. ten Dam, G. B., C. F. Zilch, D. Wallace, B. Wieringa, P. C. Beverley, L. G. Poels, and G. R. Screaton. (2000). Regulation of alternative splicing of CD45 by antagonistic effects of SR protein splicing factors. *J Immunol*. **164**:5287-95.
159. MAfee JG, H. M., Soltaninassab S, Rech JE, Iyengar S, LeStourgeon WM. 1997. THE PACKAGING OF PRE-MRNA. In Eukaryotic mRNA processing. K. A. I. p. a. O. U. Press, editor. 68-102.
160. Krecic, A. M., and M. S. Swanson. (1999). hnRNP complexes: composition, structure, and function. *Curr Opin Cell Biol*. **11**:363-71.
161. Weighardt, F., G. Biamonti, and S. Riva. (1996). The roles of heterogeneous nuclear ribonucleoproteins (hnRNP) in RNA metabolism. *Bioessays*. **18**:747-56.
162. Mayeda, A., S. H. Munroe, J. F. Caceres, and A. R. Krainer. (1994). Function of conserved domains of hnRNP A1 and other hnRNP A/B proteins. *Embo J*. **13**:5483-95.
163. Wagner, E. J., and M. A. Garcia-Blanco. (2001). Polypyrimidine tract binding protein antagonizes exon definition. *Mol Cell Biol*. **21**:3281-8.
164. Singh, R., J. Valcarcel, and M. R. Green. (1995). Distinct binding specificities and functions of higher eukaryotic polypyrimidine tract-binding proteins. *Science*. **268**:1173-6.
165. Zhang, L., M. Ashiya, T. G. Sherman, and P. J. Grabowski.

- (1996). Essential nucleotides direct neuron-specific splicing of gamma 2 pre- mRNA. *Rna*. 2:682-98.
166. Perez, I., J. G. McAfee, and J. G. Patton. (1997). Multiple RRM's contribute to RNA binding specificity and affinity for polypyrimidine tract binding protein. *Biochemistry*. 36:11881-90.
167. Chou, M. Y., J. G. Underwood, J. Nikolic, M. H. Luu, and D. L. Black. (2000). Multisite RNA binding and release of polypyrimidine tract binding protein during the regulation of c-src neural-specific splicing. *Mol Cell*. 5:949-57.
168. van der Houven van Oordt, W., M. T. Diaz-Meco, J. Lozano, A. R. Krainer, J. Moscat, and J. F. Caceres. (2000). The MKK(3/6)-p38-signaling cascade alters the subcellular distribution of hnRNP A1 and modulates alternative splicing regulation. *J Cell Biol*. 149:307-16.
169. Hanamura, A., J. F. Caceres, A. Mayeda, B. R. Franza, Jr., and A. R. Krainer. (1998). Regulated tissue-specific expression of antagonistic pre-mRNA splicing factors. *Rna*. 4:430-44.
170. Buratti, E., and F. E. Baralle. (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*. 276:36337-43.
171. Bremer, S., T. Hoof, M. Wilke, R. Busche, B. Scholte, J. R. Riordan, G. Maass, and B. Tummler. (1992). Quantitative expression patterns of multidrug-resistance P-glycoprotein (MDR1) and differentially spliced cystic-fibrosis transmembrane-conductance regulator mRNA transcripts in human epithelia. *Eur J Biochem*. 206:137-49.
172. Chu, C. S., B. C. Trapnell, S. M. Curristin, G. R. Cutting, and R. G. Crystal. (1992). Extensive posttranscriptional deletion of the coding sequences for part of nucleotide-binding fold 1 in respiratory epithelial mRNA transcripts of the cystic fibrosis

transmembrane conductance regulator gene is not associated with the clinical manifestations of cystic fibrosis. *J Clin Invest.* **90**:785-90.

173. Hull, J., S. Shackleton, and A. Harris. (1994). Analysis of mutations and alternative splicing patterns in the CFTR gene using mRNA derived from nasal epithelial cells. *Hum Mol Genet.* **3**:1141-6.

174. Yoshimura, K., C. S. Chu, and R. G. Crystal. (1993). Alternative splicing of intron 23 of the human cystic fibrosis transmembrane conductance regulator gene resulting in a novel exon and transcript coding for a shortened intracytoplasmic C terminus. *J Biol Chem.* **268**:686-90.

175. Melo, C. A., C. Serra, V. Stoyanova, C. Aguzzoli, D. Faraguna, A. Tamanini, G. Berton, G. Cabrini, and F. E. Baralle. (1993). Alternative splicing of a previously unidentified CFTR exon introduces an in-frame stop codon 5' of the R region. *FEBS Lett.* **329**:159-62.

176. Delaney, S. J., D. P. Rich, S. A. Thomson, M. R. Hargrave, P. K. Lovelock, M. J. Welsh, and B. J. Wainwright. (1993). Cystic fibrosis transmembrane conductance regulator splice variants are not conserved and fail to produce chloride channels. *Nat Genet.* **4**:426-31.

177. Trezise, A. E., M. Buchwald, and C. F. Higgins. (1993). Testis-specific, alternative splicing of rodent CFTR mRNA. *Hum Mol Genet.* **2**:801-2.

178. Mercier, B., C. Verlingue, W. Lissens, S. J. Silber, G. Novelli, M. Bonduelle, M. P. Audrezet, and C. Ferec. (1995). Is congenital bilateral absence of vas deferens a primary form of cystic fibrosis? Analyses of the CFTR gene in 67 patients. *Am J Hum Genet.* **56**:272-7.

179. Culard, J. F., M. Desgeorges, P. Costa, M. Laussel, G.

- Razakatzara, H. Navratil, J. Demaille, and M. Claustres. (1994). Analysis of the whole CFTR coding regions and splice junctions in azoospermic men with congenital bilateral aplasia of epididymis or vas deferens. *Hum Genet.* 93:467-70.
180. Chu, C. S., B. C. Trapnell, J. J. Murtagh, Jr., J. Moss, W. Dalemans, S. Jallat, A. Mercenier, A. Pavirani, J. P. Lecocq, G. R. Cutting, and et al. (1991). Variable deletion of exon 9 coding sequences in cystic fibrosis transmembrane conductance regulator gene mRNA transcripts in normal bronchial epithelium. *Embo J.* 10:1355-63.
181. Zielenski, J., P. Patrizio, M. Corey, B. Handelin, D. Markiewicz, R. Asch, and L. C. Tsui. (1995). CFTR gene variant for patients with congenital absence of vas deferens. *Am J Hum Genet.* 57:958-60.
182. Rave-Harel, N., E. Kerem, M. Nissim-Rafinia, I. Madjar, R. Goshen, A. Augarten, A. Rahat, A. Hurwitz, A. Darvasi, and B. Kerem. (1997). The molecular basis of partial penetrance of splicing mutations in cystic fibrosis. *Am J Hum Genet.* 60:87-94.
183. Chillon, M., T. Dork, T. Casals, J. Gimenez, N. Fonknechten, K. Will, D. Ramos, V. Nunes, and X. Estivill. (1995). A novel donor splice site in intron 11 of the CFTR gene, created by mutation 1811+1.6kbA-->G, produces a new exon: high frequency in Spanish cystic fibrosis chromosomes and association with severe phenotype. *Am J Hum Genet.* 56:623-9.
184. Strong, T. V., D. J. Wilkinson, M. K. Mansoura, D. C. Devor, K. Henze, Y. Yang, J. M. Wilson, J. A. Cohn, D. C. Dawson, R. A. Frizzell, and et al. (1993). Expression of an abundant alternatively spliced form of the cystic fibrosis transmembrane conductance regulator (CFTR) gene is not associated with a cAMP-activated chloride conductance. *Hum Mol Genet.* 2:225-30.

185. Niksic, M., M. Romano, E. Buratti, F. Pagani, and F. E. Baralle. (1999). Functional analysis of cis-acting elements regulating the alternative splicing of human CFTR exon 9. *Hum Mol Genet.* 8:2339-49.
186. Buratti, E., T. Dork, E. Zuccato, F. Pagani, M. Romano, and F. E. Baralle. (2001). Nuclear factor TDP-43 and SR proteins promote in vitro and in vivo CFTR exon 9 skipping. *Embo J.* 20:1774-84.
187. Teng, H., M. Jorissen, H. Van Poppel, E. Legius, J. J. Cassiman, and H. Cuppens. (1997). Increased proportion of exon 9 alternatively spliced CFTR transcripts in vas deferens compared with nasal epithelial cells. *Hum Mol Genet.* 6:85-90.
188. Caceres, J. F., S. Stamm, D. M. Helfman, and A. R. Krainer. (1994). Regulation of alternative splicing in vivo by overexpression of antagonistic splicing factors. *Science.* 265:1706-9.
189. Zahler, A. M., K. M. Neugebauer, W. S. Lane, and M. B. Roth. (1993). Distinct functions of SR proteins in alternative pre-mRNA splicing. *Science.* 260:219-22.
190. Ou, S. H., F. Wu, D. Harrich, L. F. Garcia-Martinez, and R. B. Gaynor. (1995). Cloning and characterization of a novel cellular protein, TDP-43, that binds to human immunodeficiency virus type 1 TAR DNA sequence motifs. *J Virol.* 69:3584-96.
191. Cramer, P., J. F. Caceres, D. Cazalla, S. Kadener, A. F. Muro, F. E. Baralle, and A. R. Kornblihtt. (1999). Coupling of transcription with alternative splicing: RNA pol II promoters modulate SF2/ASF and 9G8 effects on an exonic splicing enhancer. *Mol Cell.* 4:251-8.
192. Agrawal, S., and E. R. Kandimalla. (2000). Antisense therapeutics: is it as simple as complementary base recognition? *Mol Med Today.* 6:72-81.

193. Tu, G. C., Q. N. Cao, F. Zhou, and Y. Israel. (1998). Tetranucleotide GGGGA motif in primary RNA transcripts. Novel target site for antisense design. *J Biol Chem.* **273**:25125-31.
194. Ge, H., P. Zuo, and J. L. Manley. (1991). Primary structure of the human splicing factor ASF reveals similarities with *Drosophila* regulators. *Cell.* **66**:373-82.
195. Gorlach, M., C. G. Burd, and G. Dreyfuss. (1994). The determinants of RNA-binding specificity of the heterogeneous nuclear ribonucleoprotein C proteins. *J Biol Chem.* **269**:23074-8.
196. Sosnowski, B. A., J. M. Belote, and M. McKeown. (1989). Sex-specific alternative splicing of RNA from the transformer gene results from sequence-dependent splice site blockage. *Cell.* **58**:449-59.
197. Valcarcel, J., R. Singh, P. D. Zamore, and M. R. Green. (1993). The protein Sex-lethal antagonizes the splicing factor U2AF to regulate alternative splicing of transformer pre-mRNA. *Nature.* **362**:171-5.
198. Patton, J. G., S. A. Mayer, P. Tempst, and B. Nadal-Ginard. (1991). Characterization and molecular cloning of polypyrimidine tract-binding protein: a component of a complex necessary for pre-mRNA splicing. *Genes Dev.* **5**:1237-51.
199. Garcia-Blanco, M. A., S. F. Jamison, and P. A. Sharp. (1989). Identification and purification of a 62,000-dalton protein that binds specifically to the polypyrimidine tract of introns. *Genes Dev.* **3**:1874-86.
200. Chen, C. D., R. Kobayashi, and D. M. Helfman. (1999). Binding of hnRNP H to an exonic splicing silencer is involved in the regulation of alternative splicing of the rat beta-tropomyosin gene. *Genes Dev.* **13**:593-606.
201. Fogel, B. L., and M. T. McNally. (2000). A cellular protein, hnRNP H, binds to the negative regulator of splicing element

- from Rous sarcoma virus. *J Biol Chem.* **275**:32371-8.
202. Jacquenet, S., A. Mereau, P. S. Bilodeau, L. Damier, C. M. Stoltzfus, and C. Branlant. (2001). A second exon splicing silencer within human immunodeficiency virus type 1 tat exon 2 represses splicing of Tat mRNA and binds protein hnRNP H. *J Biol Chem.* **276**:40464-75.
203. Caputi, M., and A. M. Zahler. (2001). Determination of the RNA binding specificity of the heterogeneous nuclear ribonucleoprotein (hnRNP) H/H'/F/2H9 family. *J Biol Chem.* **276**:43850-9.
204. Caputi, M., and A. M. Zahler. (2002). SR proteins and hnRNP H regulate the splicing of the HIV-1 tev-specific exon 6D. *Embo J.* **21**:845-55.
205. Rozmahel, R., H. H. Heng, A. M. Duncan, X. M. Shi, J. M. Rommens, and L. C. Tsui. (1997). Amplification of CFTR exon 9 sequences to multiple locations in the human genome. *Genomics.* **45**:554-61.
206. Kazazian, H. H., Jr., and J. V. Moran. (1998). The impact of L1 retrotransposons on the human genome. *Nat Genet.* **19**:19-24.
207. Ghetti, A., S. Pinol-Roma, W. M. Michael, C. Morandi, and G. Dreyfuss. (1992). hnRNP I, the polypyrimidine tract-binding protein: distinct nuclear localization and association with hnRNAs. *Nucleic Acids Res.* **20**:3671-8.
208. Lin, C. H., and J. G. Patton. (1995). Regulation of alternative 3' splice site selection by constitutive splicing factors. *Rna.* **1**:234-45.
209. Oh, Y. L., B. Hahm, Y. K. Kim, H. K. Lee, J. W. Lee, O. Song, K. Tsukiyama-Kohara, M. Kohara, A. Nomoto, and S. K. Jang. (1998). Determination of functional domains in polypyrimidine-tract-binding protein. *Biochem J.* **331**:169-75.
210. Blanchette, M., and B. Chabot. (1999). Modulation of exon

- skipping by high-affinity hnRNP A1-binding sites and by intron elements that repress splice site utilization. *Embo J.* **18**:1939-52.
211. Del Gatto-Konczak, F., C. F. Bourgeois, C. Le Guiner, L. Kister, M. C. Gesnel, J. Stevenin, and R. Breathnach. (2000). The RNA-binding protein TIA-1 is a novel mammalian splicing regulator acting through intron sequences adjacent to a 5' splice site. *Mol Cell Biol.* **20**:6287-99.
212. Forch, P., O. Puig, N. Kedersha, C. Martinez, S. Granneman, B. Seraphin, P. Anderson, and J. Valcarcel. (2000). The apoptosis-promoting factor TIA-1 is a regulator of alternative pre-mRNA splicing. *Mol Cell.* **6**:1089-98.
213. Dember, L. M., N. D. Kim, K. Q. Liu, and P. Anderson. (1996). Individual RNA recognition motifs of TIA-1 and TIAR have different RNA binding specificities. *J Biol Chem.* **271**:2783-8.
214. Chou, M. Y., N. Rooke, C. W. Turck, and D. L. Black. (1999). hnRNP H is a component of a splicing enhancer complex that activates a c-src alternative exon in neuronal cells. *Mol Cell Biol.* **19**:69-77.
215. Markovtsov, V., J. M. Nikolic, J. A. Goldman, C. W. Turck, M. Y. Chou, and D. L. Black. (2000). Cooperative assembly of an hnRNP complex induced by a tissue-specific homolog of polypyrimidine tract binding protein. *Mol Cell Biol.* **20**:7463-79.
216. Engelbrecht, J., S. Knudsen, and S. Brunak. (1992). G+C-rich tract in 5' end of human introns. *J Mol Biol.* **227**:108-13.
217. Nussinov, R. (1988). Conserved quartets near 5' intron junctions in primate nuclear pre-mRNA. *J Theor Biol.* **133**:73-84.
218. McCullough, A. J., and S. M. Berget. (1997). G triplets located throughout a class of small vertebrate introns enforce intron borders and regulate splice site selection. *Mol Cell Biol.* **17**:4562-71.
219. Rozen, R., M. De Braekeleer, J. Daigneault, L. Ferreira-



Rajabi, M. Gerdes, L. Lamoureux, G. Aubin, F. Simard, T. M. Fujiwara, and K. Morgan. (1992). Cystic fibrosis mutations in French Canadians: three CFTR mutations are relatively frequent in a Quebec population with an elevated incidence of cystic fibrosis. *Am J Med Genet.* **42**:360-4.

220. Gan, K. H., H. J. Veeze, A. M. van den Ouweland, D. J. Halley, H. Scheffer, A. van der Hout, S. E. Overbeek, J. C. de Jongste, W. Bakker, and H. G. Heijerman. (1995). A cystic fibrosis mutation associated with mild lung disease. *N Engl J Med.* **333**:95-9.

221. Sheppard, D. N., L. S. Ostedgaard, M. C. Winter, and M. J. Welsh. (1995). Mechanism of dysfunction of two nucleotide binding domain mutations in cystic fibrosis transmembrane conductance regulator that are associated with pancreatic sufficiency. *Embo J.* **14**:876-83.

222. Van Oene, M., G. L. Lukacs, and J. M. Rommens. (2000). Cystic fibrosis mutations lead to carboxyl-terminal fragments that highlight an early biogenesis step of the cystic fibrosis transmembrane conductance regulator. *J Biol Chem.* **275**:19577-84.

223. Nissim-Rafinia, M., O. Chiba-Falek, G. Sharon, A. Boss, and B. Kerem. (2000). Cellular and viral splicing factors can modify the splicing pattern of CFTR transcripts carrying splicing mutations. *Hum Mol Genet.* **9**:1771-8.

224. Claustres, M., C. Guittard, D. Bozon, F. Chevalier, C. Verlingue, C. Ferec, E. Girodon, C. Cazeneuve, T. Bienvenu, G. Lalau, V. Dumur, D. Feldmann, E. Bieth, M. Blayau, C. Clavel, I. Creveaux, M. C. Malinge, N. Monnier, P. Malzac, H. Mitre, J. C. Chomel, J. P. Bonnefont, A. Iron, M. Chery, and M. D. Georges. (2000). Spectrum of CFTR mutations in cystic fibrosis and in congenital absence of the vas deferens in France. *Hum Mutat.* **16**:143-56.

225. Ars, E., E. Serra, J. Garcia, H. Kruyer, A. Gaona, C. Lazaro, and X. Estivill. (2000). Mutations affecting mRNA splicing are the most common molecular defects in patients with neurofibromatosis type 1. *Hum Mol Genet.* 9:237-47.
226. D'Souza, I., P. Poorkaj, M. Hong, D. Nochlin, V. M. Lee, T. D. Bird, and G. D. Schellenberg. (1999). Missense and silent tau gene mutations cause frontotemporal dementia with parkinsonism-chromosome 17 type, by affecting multiple alternative RNA splicing regulatory elements. *Proc Natl Acad Sci U S A.* 96:5598-603.
227. Lorson, C. L., E. Hahnen, E. J. Androphy, and B. Wirth. (1999). A single nucleotide in the SMN gene regulates splicing and is responsible for spinal muscular atrophy. *Proc Natl Acad Sci U S A.* 96:6307-11.
228. Muro, A. F., A. Iaconcig, and F. E. Baralle. (1998). Regulation of the fibronectin EDA exon alternative splicing. Cooperative role of the exonic enhancer element and the 5' splicing site. *FEBS Lett.* 437:137-41.