ISAS - INTERNATIONAL SCHOOL FOR ADVANCED STUDIES

Thesis submitted for the degree of *Doctor Philosophiae*

Molecular basis of the splicing regulation in human CFTR exon 9

Candidate: Martina Niksic

Supervisor: Prof. Francisco Ernesto Baralle

Academic Year 1998/1999

SISSA - SCUOLA INTERNAZIONALE SUPERIORE DI STUDI AVANZATI

TRIESTE
Via Beirut 2-4

TRIESTE

ISAS - INTERNATIONAL SCHOOL FOR ADVANCED STUDIES

Thesis submitted for the degree of *Doctor Philosophiae*

Molecular basis of the splicing regulation in human CFTR exon 9

Candidate: Martina Niksic

Supervisor: Prof. Francisco Ernesto Baralle

Academic Year 1998/1999

Acknowledgements

This work was carried out in the Molecular pathology group at the International Centre for Genetic Engineering and Biotechnology.

I am deeply indebted to Prof. Francesco E. Baralle for accepting me as a Ph.D. student in this laboratory. Also, I would like to express my gratitude to him for his excellent guidance and support throughout the project.

A very special thanks goes to my co-supervisor Dr. Maurizio Romano. Without his keen interest, valuable suggestions and his unending patience, this work would not have been possible. I will always treasure his advices and his friendship.

I am grateful to Dr. E. Buratti and Dr. F. Pagani for their collaboration and unselfish help regarding the studies of trans-acting factors.

I was very fortunate to have very supportive, friendly and helpful laboratory members. It was a pleasure working with them. In particular, I would like to thank to Dr. S. Gajovic and A. Muro for passing to me their knowledge and experience in handling with embryonic stem cells and for their stimulative suggestions and encouragement during my thesis work. My deep appreciation goes to them as a scientists and as a colleagues. I will always remember Giorgia Danek, Martin Marro, Gabriela Pittis, Federico Odreman and Lorenza Ciani. Thank them for their friendship and support during these years.

Finally I would like to express my heartfelt gratitude to my family for their continuous motivation and trust in me during my studies.

This thesis is dedicated to my parents.

Table of contents

CHAPTER 1

INTRODUCTION

1. CYSTIC FIBROSIS

Cystic fibrosis (CF) is a very common single gene disorder with a complex phenotype. It affects about 1 in 2500 newborns in Northern Europe population, similar to the frequency observed among the Ashkenazi Jews (1:3300) whereas its incidence among Asians and African blacks is lower [1].

Disease is characterised by chronic obstruction and infection of the respiratory tract, exocrine pancreatic insufficiency, and its nutritional consequences, and elevated levels of sweat electrolytes (reviewed in Welsh et al. 1995) [2].

The knowledge of the pathophysiology of cystic fibrosis have occurred only in the late 80's. In 1985 it was demonstrated that the primary defect of cystic fibrosis might be in chloride transport, when it was shown a cyclic AMP-dependent transepithelial chloride current in normal but not CF epithelia [3, 4]. These conductance pathways appeared to be unique to epithelial cells in which salt and water transport rates are governed by cyclic AMP- and calcium-dependent regulatory processes [5]. Decrease in fluid and salt secretion is responsible for the blockage of exocrine outflow from the pancreas and the accumulation of heavy dehydrated mucous in the airways. In sweat glands, salt reabsorption is defective.

Following identification of the CF gene in 1989 and the characterisation of encoded protein [6], it has been proved its function as a cyclic adenosine monophosphate (cAMP) mediated Cl-channel which has been identified in several epithelia including lung, biliary tract, pancreas and vas deferens [7]. This wide distribution accounts for the multi-organ involvement seen in cystic fibrosis.

1.1. Clinical features of classic cystic fibrosis disease

Manifestations relate to the disruption of exocrine function of the pancreas, bronchial glands (chronic bronchopulmonary infection with emphysema), intestinal glands (meconium ileus), biliary tree (biliary cirrhosis) and sweat glands (high sweat electrolyte). Infertility is observed both in males and females (reviewed in Welsh et al. 1995) [2].

Respiratory tract

In the airway epithelium the defect in ion transport results in the loss or reduction of chloride secretion into the airways. Active sodium absorption is also increased and both of these ion changes increase water reabsorption from the lumen, lowering the water content of the mucus blanket coating the mucosal cells. This dehydration of the mucus layer leads to the defective mucociliary action and the accumulation of hyperconcentrated, viscid secretion that obstruct the air passages predisposing to pulmonary infections. Airway surface fluid in cystic fibrosis patients shows markedly reduced activity of antibacterial endogenous substances, possibly related to its abnormally high rates of liquid absorption. The failure to clear thickened mucus from airway surfaces likely initiates CF airways infection primarily by *Staphylococcus aureus* and *Pseudomonas aeruginosa*.

Gastrointestinal tract

In gastrointestinal tract there are several type of glands affected by CF disease: simple goblet cells, mucus glands tubular glands, Lieberkühn's crypts and Brunner glands. Changes are caused by defective production of the secret that result in forming dehydrated, sticky and incompletely digested intestinal contents. This plug cannot be removed properly and finally leads to the obstruction of intestine, especially ileum. Also, the altered functionality of

Brunner glands results in the failure of the gastric acid buffer action, causing duodenal irritation.

Pancreatic insufficiency

The pancreatic insufficiency is present in approximately 85-90% of patient. Accumulation of the mucus results in the loss of the acinar cells that are destroyed and replaced by the fibrotic tissue and fat. CF affected pancreas fails to secret digestive enzymes what makes digestion and absorption of the food difficult.

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

The vas deferens, tail and a body of the epididimys and seminal vesicles are usually atrophic, or completely absent. The pathogenesis of this structural changes relates to early intrauterine obstruction of the genital tract with inspissated secretions. About 90% female are infertile 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. Apocrine sweat glands in children are dilated and filled with retained secretions.

1.2. The CFTR protein: structure expression and function

Defect of the ion transport in CF patients is due to the malfunction or the absence of the protein called CFTR (cystic fibrosis transmembrane regulator) [8, 9].

The CF related protein, CFTR is a 169 kDa membrane-associated glycoprotein made of 1480 aminoacid residues [8]. Its aminoacid sequence shows a high homology to a superfamily of ABC (ATP-binding cassette) transporter proteins [10]. This family is specialised within the cell in uptake and secretion, intracellular transport, cell detoxification and signalling. Similarly, the CFTR protein in epithelial cells functions as the cAMP mediated chloride channel.

1.2.1. The CFTR protein structure

CFTR protein consists of 5 domains (Figure 1.1.). There are two nucleotide binding folds that bind ATP (NBF1 and NBF2) and two hydrophobic transmembrane domains that each have six membrane-spanning segments (TM1-TM12). The R-domain is a highly charged domain specific to CFTR [11] .

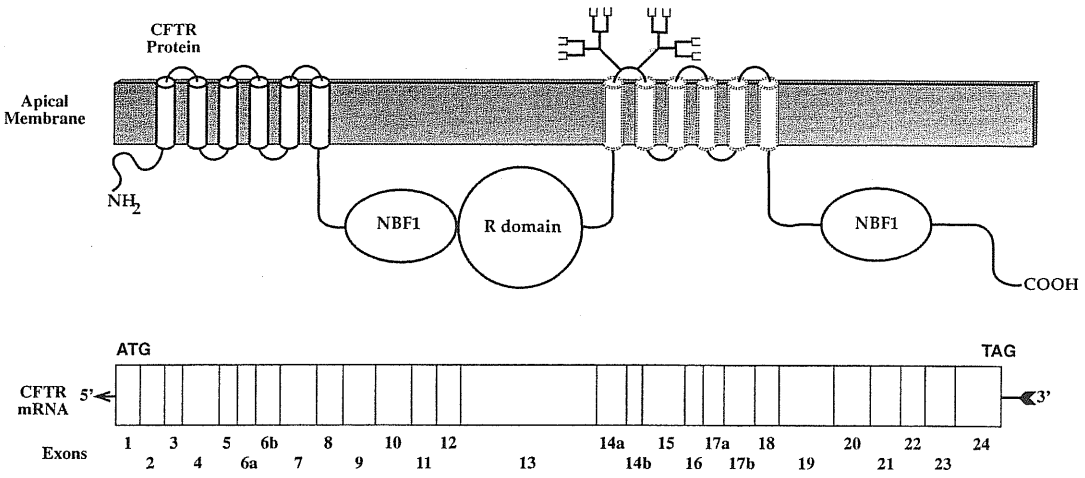


Figure 1.1. The cystic fibrosis transmembrane condctance regulator (CFTR).
a) The model of CFTR protein and its presumptive relationship with the apical membrane of secretory epithelial cells. b)The structure of the CFTR mRNA with the 27 exons.

a) transmembrane domain

CFTR is anchored in the plasma membrane by the 12 hydrophobic segments (TM 1-12), in two transmembrane domains [12]. The 12 membrane spanning segments presumably form a pore constituting the CFTR chloride channel. Different studies imply that CFTR is a multi-ion pore channel [13] and that TM1,

TM2, TM6 and TM12 are important in pore formation [14-16]. On the basis of aminoacid substitutions found in many disease causing mutations, several arginine residues appear to be particularly important for ion conductance. The current protein model predicts six extracellular loops and four cytoplasmatic loops.

b) nucleotide binding folds

The two nucleotide binding folds contain sequence motifs Walker A and Walker B, that are conserved among ATP binding proteins [11]. The gene sequence encoding NBF1 spans exons 9 through 12 and that encoding NBF2 spans exons 19 through 23. The two NBFs are believed to bind and hydrolyse ATP following AMP mediated phosphorylation of the R-domain [17]. The overall similarity between the primary structures of the two NBFs is only ~29% at the aminoacid level. The activity of the channel is regulated through the level of phosphorylation of the R domain and NBF domains. Functional studies show that NBF1 and NBF2 do not contribute to the channel regulation in the same way. It is proposed that partial phosphorylation of the R domain leads to the phosphorylation of the NBF1 and channel opening, whereas complete phosphorylation of R domain leads to the phosphorylation of NBF2 causing the channel closing [18].

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 [11]. Several phosphorylation sites seems to be important for the regulation of the channel activity. Recent studies show that phosphorylation of different sites may have either stimulatory or inhibitory effect on the CFTR channel activity [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 NBFs that, in turn, controls the channel opening and closing.

1.2.2. Expression of CFTR protein

CFTR is predominantly expressed in epithelial tissue. By Northern blot analysis CFTR mRNA is detected in lung, pancreas, sweat glands, liver, nasal polyps, salivary gland and colon [22]. The statement that CFTR is an epithelial specific protein was supported by the fact that the clinical disorder, CF, develops as a consequence of lack or nonfunctionality of the CFTR protein is specifically manifested in these tissues.

By RT-PCR, CFTR mRNA transcript was as well detected in a very low amount in nonepithelial cells, fibroblasts and lymphocytes [23]. Estimated level of CFTR gene expression in the nonepithelial cells is less than one mRNA transcript per cell. This evidence, together with the fact that promoter of the CFTR gene has characteristics of a housekeeping gene [24], implies that the expression of the CFTR gene might not be limited to the epithelial cells. However, clinical and biological significance of the CFTR in nonepithelial cells is still uncertain.

1.2.3. The CFTR function

ABC family members have different functions including ATP dependant transmembrane pumping of larger molecules, regulation of other membrane transporters and ion conductance. CFTR shares functional characteristics with these proteins and play a multiple role in epithelial transport.

In the cell, CFTR is mainly localised within the apical membrane where it constitutes a low conductance anion channel [7, 25]. CFTR takes part of the pathway by which Cl^- moves from the cell into the lumen. It uses the energy of

ATP hydrolysis to drive conducting- and nonconducting- conformations and chloride flow.

Preliminary studies with immunoelectron photomicrographs suggest its localisation also inside the cell, possibly within the membranes of the endosomes. Studies of cell expressing recombinant CFTR suggest that CFTR on endosomes is functionally active [26]. It seems to function in parallel with H+ pumps to acidify intracellular vesicles. By changing the degree of vesicle acidification, the presence or absence of CFTR Cl- channels might alter pH-sensitive enzymatic processes [27]. It is also possible that CFTR, located either in the apical membrane or on the intracellular vesicles, may regulate vesicle endocytosis and exocytosis [28].

Furthermore, CFTR can also modulate the activity of other membrane channels: CFTR is a negative regulator of amiloride-sensitive epithelial Na+ channel. In fact, when CFTR function is lost, the Na+ conductance markedly increases in human airways [29].

1.3. Genetic basis of the CFTR chloride channel dysfunction

The gene encoding the CFTR protein is located in the long arm of chromosome 7 (q31-32) [6]. It spans approximately 230 kb and consists of 27 exons ranging in size from 38 to 724 bp [11]. It is highly conserved between the different species.

Mutations within the CFTR gene affect the function of the CFTR protein. It is not known how many mutations really exist within the CFTR gene but the fact that up to now more than 700 different mutations have been described, indicate the high mutation rate in this gene (Figure 1.2.). The major mutation that results in a single aminoacid deletion (ΔF508) account for the 70% of the disease alleles. The other mutations, with the exception of G551D, G542X and 621+1 (G-A) are rare, affecting <1% of CF alleles. The list of mutations can be found at the world web access http://www.genet.sickkids.on.ca

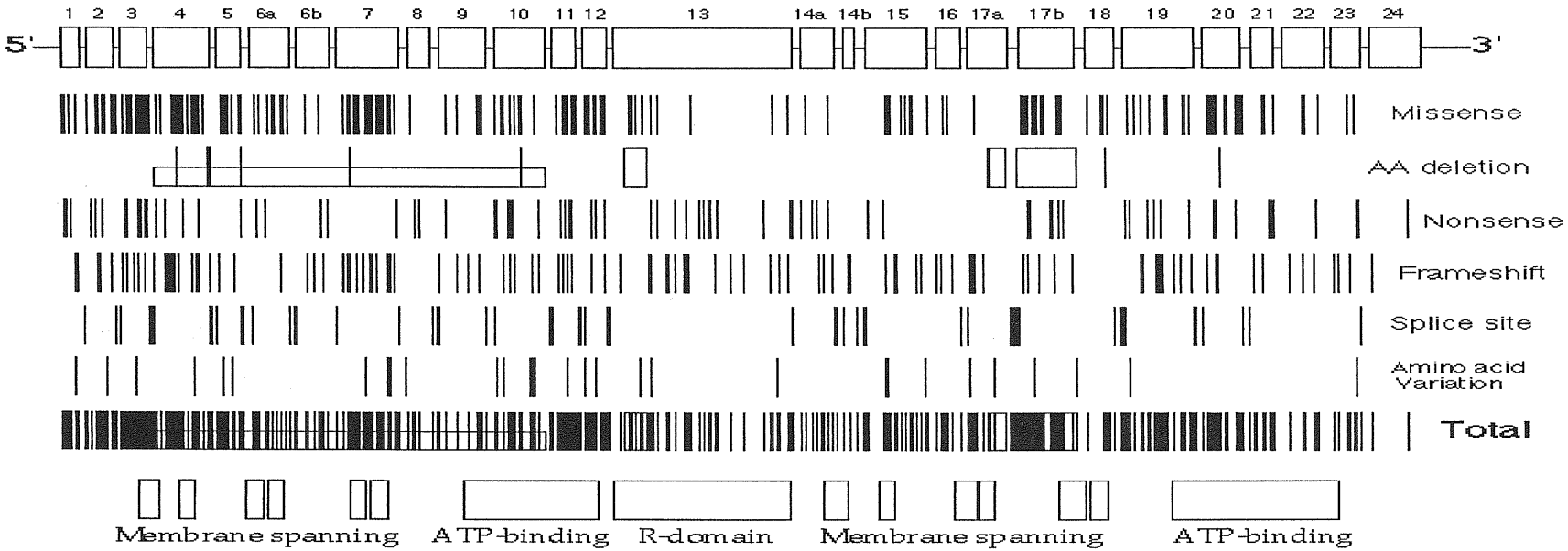


Figure 1.2. Distribution of cystic fibrosis mutations and amino acid polymorphisms.
The shematic diagram contain unpublished data from CF Genetic Analysis Consortium, July, 1995. Aminoacid variations refers to presumed benign polymorphisms.

CFTR mutations are classified in 5 different classes depending on the way in which they affect normal function of the CFTR protein .

Class I: mutations affecting protein production

There are mutations throughout the CFTR gene that produce premature termination signals because of splice site abnormalities, frameshift due to deletions or insertions, or nonsense mutations [30]. In some cases, the mutation results in an unstable mRNA and no detectable protein [31]. 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 include the most common mutation, ΔF508. The failure of proteins to progress through the biosynthetic pathway and secretion, can be followed by assessing their state of glycosylation. In recombinant cells, CFTR containing the ΔF508 mutation fails to mature to the fully glycosylated form [9]. The partially glycosylated mutant protein is rapidly degraded. Thus, ΔF508 protein cannot be detected at the cell surface.

Missense mutations in this class represent the majority of CF alleles.

Class III : mutations affecting Cl^- channel regulation

The mutant proteins of this class are those that appear to be fully processed and correctly located in the apical membrane but are unable to function as a chloride channels. The range of effects of dysregulation of the channel includes those with severe lack of function (such as that for G551D), reduced response to ATP stimulation (S1255P), and slight reduction of absolute activity (G551S, G1244E and G1349D) [32]. Some mutations in this class appear to coincide with aminoacids involved in ATP-binding.

Class IV : mutations affecting Cl^- conductance or channel gating

A number of CF-associated missense mutations have been identified within the exons coding for the membrane-spanning domains that are thought to contribute to the channel pore formation.

Examples of this group include R117H near TM2, G314E in TM5 and R334W and R347P in TM6. When these mutant CFTRs are expressed in heterologous epithelial cells, all are correctly processed, are present in the apical membrane, and generate cAMP regulated apical membrane Cl^- currents. Nevertheless, the amount of current is reduced, with wild type CFTR>R347P>R117H>R334W [16].

Class V : mutations causing reduced synthesis

Mutations of this class include promoter mutations that reduce transcription [33], nucleotide alterations that promote alternative splicing of CFTR transcript (IVS8 T5, 3849+10kb C-T)[34, 35] and aminoacid substitutions that cause inefficient protein maturation (A455E) [36]. These mutations would reduce relative amount of normal, functional CFTR protein.

Most CFTR mutations grouped under this class are those generating alternatively spliced mRNA products [37, 38]. Since a 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

Expression of the CF disease is highly variable within the different patients. Some severe affected patients fully show complete spectrum of the general symptoms described in the first chapter while others show mild form of the disease with slightly expressed symptoms. Then, some individuals do not exhibit classic CFTR symptoms but present monosymptomatic form, affecting just single organs (Table 1.3.).

Disease	Common manifestations shared with CF	Fraction of patients with at least one CFTR mutation (%)	Reference
CBAVD	absence of vas deferens (bilateral)	73/102 (73%)a	Chillon, 1995
		51/68 (75%)a	Zielenski, 1996
CUAVD	absence of vas deferens (unilateral)	6/14 (43%)b	Mickle, 1993
Diffuse Bronchiectasis	abnormal dilatation of bronchi	6/10 (60%)c	Poller, 1991
		6/48 (12.5%)d	Simon-Bouy, 1993
		9/28 (32%)a	Pignatti, 1994
Bronchiectasis with elevated sweat Cl-	abnormal dilatation of bronchi and high levels of sweat chloride	5/16(31%)e	Gervais, 1993
Allergic bronchopulmonary aspergillosis	allergic asthma tenacious sputum mucus plugs	6/11 (54%)d	Miller, 1994
Chronic Pseudomonas bronchitis	chronic sinusitis nasal polyposis	2/10 (20%)d	Miller, 1994
Chronic bronchial hypersecretion	abnormal mucous secretion	6/65 (9.2%)e	Dumur, 1990
Nasal polyposis	nasal polyps	7/112 (6.2%)f	Burger, 1991
Neonatal transitory hypertrypsinaemia	high levels of immunoreactive trypsin (IRT)	9/149(6%)e	Laroche, 1994

a-The numbers are based on comprehesive screening of CFTR mutations (including IVS8:5T) by a variety of methods; b-testing of 3 mutations ([[Delta]]F508, R117H and R75Q); c- direct sequencing of exons encoding NBF1; d- the most common CFTR mutations (unspecified); e- [[Delta]]F508 only ; f- 8 mutations ([[Delta]]F508, [[Delta]]I507, D110H, R117H, 621+1G->T, N1303K, G551D and R553X).

Table 1.3. Atypical (non-CF) diseases associated with the CFTR gene mutations (http://www.genet.sickkids.on.ca/cftr/)

The clinical variable that most readily discriminates different phenotypes is the state of pancreatic function. Most patients have pancreatic dysfunction and require supplementation with pancreatic enzymes (pancreatic insufficient) while pancreatic sufficient patients retain significant pancreatic function [39].

The clinical outcome can be explained partially by genotype, or better through the way in which a specific mutation affects function of the CFTR protein.

Thus, class I and II mutant proteins are missing from the correct cellular location and they have been associated with the severe pancreatic insufficient phenotype. In contrast, CFTR protein carrying mutations of classes III, IV or V is correctly localised but it is present in reduced amount or shows reduced Cl^- channel activity. The residual activity may be sufficient to cover the need of some organs for the CFTR encoded Cl^- channel. In fact, these mutations are usually related to the pancreatic sufficient CF patients [40, 41].

1.3.2. Complex alleles

Unfortunately, heterogeneity of the clinical symptoms is not completely determined by different genotype.

Many patients are compound heterozygous, and it is difficult to predict clinical abnormalities resulting from two CF associated mutations. Furthermore, an increasing number of reports have described the coexistence of two independent DNA alterations in the same CFTR allele [42]. Although most of these complex alleles may represent merely association of a benign sequence variation with an actual disease causing mutation in the same gene, there are examples in which a second mutation can modify the effect of the principal mutation [42]. Finally, some variabilities, especially in lung disease may result from environmental factors and variability in genetic background [43].

1.4. Alternative splicing of CFTR transcripts

Several splicing variants have been observed in transcripts of the CFTR gene both in normal and CF individuals. The most common observed variants are these lacking exons 4, 9 or 12 which would result in transcripts with in frame deletions. They are quite abundant within the cell: 9^- transcript form is found to be present up to 92% in some individuals [44], a 12^- form is present from 5-30%, while 4^- form present less than 2% of total CFTR transcript [45].

CFTR transcript segments with other aberrant splicing patterns have also been detected by RT-PCR analysis of cell from non CF individuals; these include the skipping of exons 2/3, 3, 9/12, 10/11, 12 and part of 13, 14a, part of 15, or 23 [46]. Two examples of alternative splicing resulting in insertion of additional genomic DNA sequences in the transcript has also been reported. In one case, an insertion of a 119-nucleotide segment was detected between exons 10 and 11 [47, 48] and it is called 10b exon. Resulting transcript would truncate the normal open reading frame. The second example is the inclusion of 260 nucleotides from the 3′ end of intron 23 in about 3-16% of total CFTR transcripts named 24a exon. The presumptive CFTR product from mRNA variant with 24a sequence would be shortened by 61 amino acids at the C-terminus [49].

These alternative splicing forms observed in humans are not conserved in other species. In mice, for example, the other alternative spliced variants have been described: 2-40% of total mouse CFTR mRNA lacks exon 5 [50], and transcripts with additional 11b sequence reach a proportion of 10-15% of total CFTR mRNA, specifically in murine testis [51, 52].

It is not clear if any of those transcript has some physiological role. Up to now, the functional consequence for alternative splicing has been investigated just for exon 9- form in human and exon 5- in mouse. It is shown that those transcripts produce nonfunctional CFTR protein [50, 53]. Other transcripts are also likely to produce nonfuctional CFTR chloride channel but it is possible that some of this

alternative splicing forms of CFTR may play a role in postranscriptional regulation of gene expression.

1.4.1. Alternative splicing of CFTR pre-mRNA and monosymptomatic forms of cystic fibrosis

Recently, a growing number of diseases other than the classical CF have been correlated to mutations in the CFTR. This is the case of the congenital bilateral absence of vas deferens (CBAVD), nasal polyposis, bronchiectasis, broncho-pulmonary allergic aspergillos and chronic idiopathic pancreas [54-56]. For example, 50-82% of men with CBAVD have at least one detectable CFTR gene mutation, ~15% have two detectable CFTR mutations and at least one of the two detectable mutations appears to be of a mild form [57, 58]. Surprisingly, to make more complex the genotype-phenotype correlations, some patients present the same mutations found in healthy carriers.

In other cases, the screening did not identify any mutation within the coding region of the gene of these patients. However, PCR amplification of overlapping fragments of the CFTR cDNA has disclosed the presence of aberrantly spliced CFTR mRNA species. The further characterization of the genomic DNA spanning the exonic and the intronic regions involved in the alternative splicing has revealed the presence of mutations within the introns.

This was the case of some patients suffering for chronic lung disease with normal sweat chloride values and other patients affected by CBAVD.

Pedigree and epidemiological studies have permitted to correlate the 3849+10kb C-to-T mutation in intron 19 with the for chronic lung disease with normal sweat chloride values. The C-to-T mutation activates a cryptic splice site in intron 19 and leads to the insertion into CFTR transcripts of a new 84-bp "exon" with an in-frame stop codon between exons 19 and 20 [34].

Another example involves the polymorphic sequence in the intron 9 (IVS8-T), which is composed of poly-thymidines present in three different lengths: T9, T7 and T5. Genetic studies have permitted to associate the T5 allele of the polymorphic loci within the 3′ splice site of intron 8 with the CBAVD disease and some other monosymptomatic forms of CF. Patients carrying T5 allele were found to present a higher proportion of CFTR mRNA devoid of exon 9 [55, 59] compared to the T7 and T9 carriers. In most CBAVD patients, the T5 allele was strongly associated with the presence of a classic CF mutation in the other copy of the CFTR gene.

1.4.2. Association studies of alternative splicing of exon 9 in CFTR

The in-frame deletion of exon 9 is the most common alternatively spliced variant of the CFTR mRNA detected either in CF and in healthy individuals [44]. The mRNA devoid of exon 9 is translated in protein that is nonglycosylated and is not a functional chloride channel [50]. Even if there are suggestions that this form may have some function in the cell other than (that of) chloride channel, more likely its is nonfunctional. Its presence is without effects until the level of the 9+ form is sufficient for the normal cell function. In the healthy individuals has been found that CFTR 9- form can represent from 0 to 92% proportion of total CFTR mRNA without consequences for the normal function showing the high tolerance of the organism for production of low amounts of CFTR protein including exon 9 [44].

The fact that expression of exon 9- form is very variable among the persons but at the same time is very conserved and constant within the same individual brings to the conclusion that there are some cis-acting elements which determine the efficiency of splicing machinery to recognise and correctly splice exon 9.

An association has been established between the amount of full length CFTR transcript and nature of the IVS8-T variant, demonstrating a positive correlation

between the length of the poly-tymidine tract and the proportion of mRNA with intact exon 9 [35, 60].

Association studies have found that 84% of men with CBAVD who were heterozygous for a CF mutation carry the intron 8 polypyrimidine 5T CFTR allele on one chromosome. It has been reported that a particular mutation, R117H, found in a high proportion in CBAVD patients, can confer either a CF or CBAVD phenotype depending on which of the IVS8T alleles is present within the CFTR intron 8 [43]. R117H, in conjunction with the 5T allele, results in CF when is associated with another CF mutation [61]. Osborne et al. and Chillon et al. observed high proportion of 5T variant in the CBAVD patients having one or no CFTR mutation detected [59, 62]. Frequency of the T5 variant in these patients was 6 times higher (30%) than that of the general population (5%) [55, 59].

Together, these observations lead to the conclusion that T5 allele is associated with high levels of CFTR transcript lacking exon 9 and that ,when the production of normal functional CFTR protein is very low due to the mutations and the chloride conductance rate hardly covers the need of the cell for the CFTR protein, T5 allele and nonefficient splicing of exon 9 can be a critical parameter for developing CFTR related diseases.

Recently, it has been hypothesised that the other polymorphic locus based on TG dinucleotide repeats (9, 10, 11, 12 or 13 TG repeats) and located immediately upstream of the polypyrimidine tract (T repetition) might influence exon 9 skipping in CFTR mRNA, even independently from polyT. Cuppens et al. [63] have reported a negative influence of TG stretch on correct splicing of exon 9. T5 CFTR genes derived from CBAVD patients were found to carry a high number of TG repeats while T5 CFTR genes derived from their healthy fathers harboured a low number of TG repeats [63].

Up to now, beside the association studies no other investigations which would elucidate the molecular mechanisms involved in the alternative splicing of human CFTR exon 9 have not been performed. At a first glance, it seems that it deviates from the canonical mechanisms of the constitutive and alternative splicing in eukaryotes. We shall first review these mechanisms and then compare them with the phenomena observed in CFTR exon 9 splicing.

CHAPTER 2

2. SPLIT GENES AND RNA SPLICING
(The nuclear pre-mRNA splicing of higher eukaryotes)

Protein coding sequence in eukaryotic gene are disrupted by intervening sequence that contain very little or no information for the given protein. Those sequences called introns are copied into primary transcripts. Then, during the processing step called splicing those sequence are excised from the RNA and the information containing sequence named exons are joined together creating a mature messenger RNA that will be translated into the protein. Split genes, consisting of interspersed intronic and exonic sequences, were first discovered in adenovirus [64, 65], but were quickly shown to be a common feature of cellular genes [66]. They have been identified in all types of eukaryotic cells studied, although the proportion of genes containing introns is generally lower in unicellular organisms than in metazoa. In some cases, the same pre-mRNA sequence can be recognised either as an exon or as an intron under different conditions indicating the complexity of the splicing process.

2.1. CONSTITUTIVE SPLICING
Splicing is a part of the pre-mRNA posttranscriptional processes

Primary transcript of protein coding gene called pre-mRNA, named heterologous nuclear RNA (hnRNA), is synthesised by RNA polymerase II within the nuclei of eukaryotic cells. A mature messenger (mRNA) is formed by complex series of post-transcriptional modifications of hnRNA, that are: capping at 5′ end, polyadenilation at 3′ end and splicing. Upon modifications, formed mRNA is then transported to the cytoplasm, where is available to be translated.

From the time hnRNAs emerge from the transcription complex, and throughout the time they are in the nucleus they are associated with proteins. The most abundant cellular proteins that bind hnRNA are the heterologous nuclear

ribonucleoprotein particles (hnRNP) (reviewed in [67]). Some members of this family of proteins have helix-unwinding activity and seem to function primarily to maintain the transcript in a single-stranded state. They influence the structure of hnRNA thus facilitating the interaction of hnRNA sequences with the other components that are needed for its processing event. hnRNP proteins may also play important roles in the interaction of hnRNA with other nuclear structures, in nucleocytoplasmic transport of mRNA, and in other cellular processes [68].
The dynamic structure within the nuclei, called speckles, are places of localisation of many transcription and splicing factors. [69]. Upon the activation of the genes, those factors are recruited to the sites of transcription, bind to the hnRNA in the highly ordered cooperative way, thus driving a processing event and then return to the speckles between rounds of active pre-mRNA synthesis for recycling and reactivation [70]. The cooperativity is shown for transcription and splicing processes which occur cotemporarily and influence each other (reviewed in [71, 72]).

2.1.1. Spliceosome

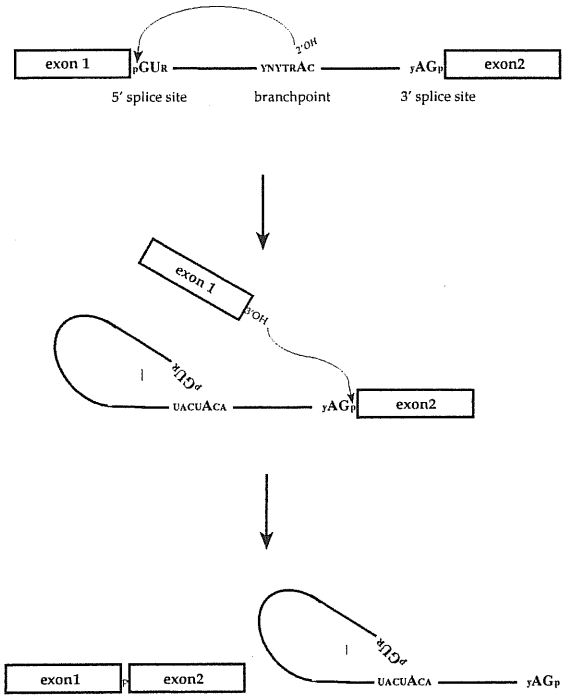
A ribonucleoprotein complex in which pre-mRNA splicing reaction occur is called spliceosome. Within the spliceosome pre-mRNA is recognized and bound by different RNA-protein particles in dynamic (association-dissociation) cycles and highly ordered way, that drive it through numerous conformational rearrangements, involving formation and disruption of numerous base pairs and finally resulting in the mature spliced mRNA.

2.1.1.1.The splicing reaction

Splicing takes place via a two-step mechanism, involving sequential transesterification reactions (Figure 2.1.) (reviewed in [73]). The first catalytic step

involves cleavage of the intron at the 5′ intron-exon junction and concomitant formation of so called "lariat" structure, that is a form of the intron in which 5′ terminal phosphate is esterified to the ribose 2′OH group of an adenosine residue in the intron, usually located 20-40 nucleotides upstream of the 3′ splice site. A free 5′ exon and lariat intron-3′ exon are thus generated as intermediates of the splicing reaction. The second step involves cleavage of the 3′ splice site, ligation of the two exons and release of the intron which is still in the lariat form.

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



2.1.1.2. Sequence elements (cis) on pre-mRNA required for splicing

Several sequence elements are required for the splicing reaction, and each of these elements is recognised multiple times through the spliceosome assembly.

They are placed at the intron-exon boundaries and introns, downstream or upstream to the exon.

Two sequences were found to be essential but not sufficient, for the efficient intron removal. These are the sites where trans-esterification reactions occur and are defined by consensus sequences [74]. The first sequence is called 5′ splice site junction is relatively short; (A/C)AG/GURAGU in mammals, where first two bases (GU) are essential for splicing event. An another sequence, the 3′ splice site, is defined by three separate sequence elements; the branch site, the polypyrimidine tract, and the 3′ splice site junction [75]. The 3′ splice site junction is defined by CAG/R in mammals. Branch site is located 18-40 nucleotides upstream of the point of cleavage of the RNA. However, in higher eukaryotes there is a considerable flexibility in its sequence requirements. Determined consensus sequence is UNCURAC, where underlined "A" residues is that at which branch formation occurs [76]. A polypyrimidine tract is placed between the branch site and 3′ splice site junction. It is a stretch of 10-30 bases containing pyrimidines not strictly defined.

In higher eukaryotics, beside the basic splicing elements, there are other sequences that influence pre-mRNA splicing in a positive and negative way. Sequences that promote exon definition are called enhancers and those that have an inhibitory effect are called silencers. Their characteristics and role during the splicing will be described in the section 2.2.7.

2.1.2. Protein components involved in splicing

A protein components of the splicing machinery, involved in splicing event are the small nuclear ribonucleoprotein particles (snRNPs) and other non snRNP essential splicing factors. Most of these proteins share very similar structural characteristics such as different number of RNA binding motifs and/or protein binding domains.

2.1.2.1. Small nuclear ribonucleoprotein particles (snRNP)s

The snRNP particles are functional components of the spliceosome. Within the spliceosome they are composed of RNA and protein components. snRNAs named U1,U2,U4/U6 and U5 are characterised by their small size, metabolic stability and high degree of sequence conservation [77]. They show high complementarity to the consensus splice sites on the pre mRNA. Upon their transcription by polymerase II (all but U6) they rapidly migrate to the cytoplasm where they assemble with the common core proteins. Then, the snRNPs are reimported to the nucleus. U6 snRNA is transcribed by RNA polymerase III. The U6 remains in the nucleus and assemble with the U4 snRNP to form a larger U4/U6 ribonucleoprotein complex. The U5 assemble in an ATP dependent reaction with the U4/U6 snRNP forming a [U4/U6.U5] three snRNP particle [78]. The spliceosomal snRNAs, with the exception of U6, has a common, evolutionary conserved structural motif, which consists of a single-stranded region, $PuA(U)_nGPu$ often flanked by double stranded stem. This motif serves as the association site or the common snRNP proteins.

A group of eight polypeptides (B, B′, D1, D2, D3, E, F, G) form a common core in each snRNP particle. Core proteins interact with a conserved sequence present in the U1, U2, U4 and U5 snRNA. In addition to the core proteins, the snRNPs also contain a group of specific proteins. The A, C and 70 kDa proteins are specific for the U1 snRNP. U1snRNP 70K and U1snRNPA appear to be involved in splice site recognition and selection [79]. U1snRNPA is also involved in linking the splicing and polyadenilation machinery [80]. A′ and B′ proteins are specific for the U2 snRNP. Both proteins are involved in the U2 snRNA binding specificity [81, 82]. SAP60 and SAP90 are U4/U6 specific proteins. U5 RNP particle presents a relatively complex protein composition. Eight U5 specific proteins with molecular masses ranging from 15 to 220 kDa have been identified. U4/U6/U5

complex contain five additional proteins from 15,5 to 63 kDa [83]. Some of these proteins probably mediate the association of U4/U6/U5 complex (reviewed in Will et al.) [84].

2.1.2.2. Non-snRNP protein splicing factors

a) U2AF

U2 snRNP auxiliary factor (U2AF) is an essential splicing factor that is required for the binding of U2 snRNP to the pre-mRNA [85]. It is a heterodimer composed of 65kDa and 35kDa subunit. $U2AF^{65}$ contains two functional domains consisting of a sequence specific RNA binding-region containing three RNA-recognition motif (RRM) and a N-terminal short serine/arginine rich motif essential for binding protein-protein interactions [86]. It binds specifically to the pyrimidine tract at the 3′ splice site. The $U2AF^{35}$ subunit contains at least two protein binding domains; a region required for dimerisation with $U2AF^{65}$ and an serine/arginine rich motif, required for dimerisation with other splicing factors. The lack of RNA recognition motif within the $U2AF^{35}$ implies its function in bridging two proteins [87]. In fact, in constitutive and alternative splicing it functions as a mediator between $U2AF^{65}$ and SR proteins.

b) SR proteins

SR proteins are family of highly conserved pre-mRNA splicing factors that are essential for constitutive pre-mRNA splicing (individual SR protein can restore splicing activity to cytoplasmic extract depleted of SR proteins [88]) and can also regulate alternative splicing (see section 2.2. Alternative splicing). Up to now 9 human SR proteins have been identified with different techniques by different groups. These have been named: SRp20, SRp30a (ASF/SF2), SRp30b (SC35/PR264), SR30c, SRp40, SRp55, SRp75 and 9G8 and more divergent p54, with molecular weight ranging between 20 and 75 kDa [89, 90]. These proteins are

closely related in primary structure. Analysis of SR proteins demonstrates that each family member contain one or two copies of a characteristic RNA-recognition motif (RRM) at the amino-terminal end of the protein by which it recognises certain sequence on the pre-mRNA and a serine/arginine (SR) domain at the carboxi-terminal region responsible for protein-protein interaction [91]. It has been demonstrated that the initial binding of SR protein with pre-mRNA is sufficient for commitment to the splicing pathway [92] and presents a initial step in spliceosome assembly. SR proteins are phosphorylated in vivo and in vitro at multiple serine residues within the SR domain [93]. Recent experiments showed that under in vitro conditions, both phosphorylation and dephosphorylation may be important for constitutive splicing. In fact, cycles of phosphorylation and dephosphorylation affect both spliceosome assembly and catalysis [94].

2.1.3. The spliceosomal complexes

Both, the snRNA and the spliceosomal proteins assemble on pre-mRNA in a stepwise pathway. Four distinct spliceosomal complexes (E, A, B, and C) have been detected in HeLa cell nuclear extracts. E, A and B complexes contain unspliced pre-mRNA, while the C complex contains the products of catalytic step I of the splicing reaction (i.e. the exon and lariat exon intermediate). Additional complexes, i and D are generated from the catalytic step II of the splicing reaction and contain intron and spliced exons [95, 96] (Figure 2.2.).

2.1.3.1. General model for spliceosome assembly and spliceosome cycle

Mechanism of early E complex formation introduce two independant complexes assemble, one at 5′ splice site (E5′) and the other at the 3′ splice site of the intron (E3′) [97]. The U1 snRNP binds to the 5′ splice site of pre mRNA. The binding of

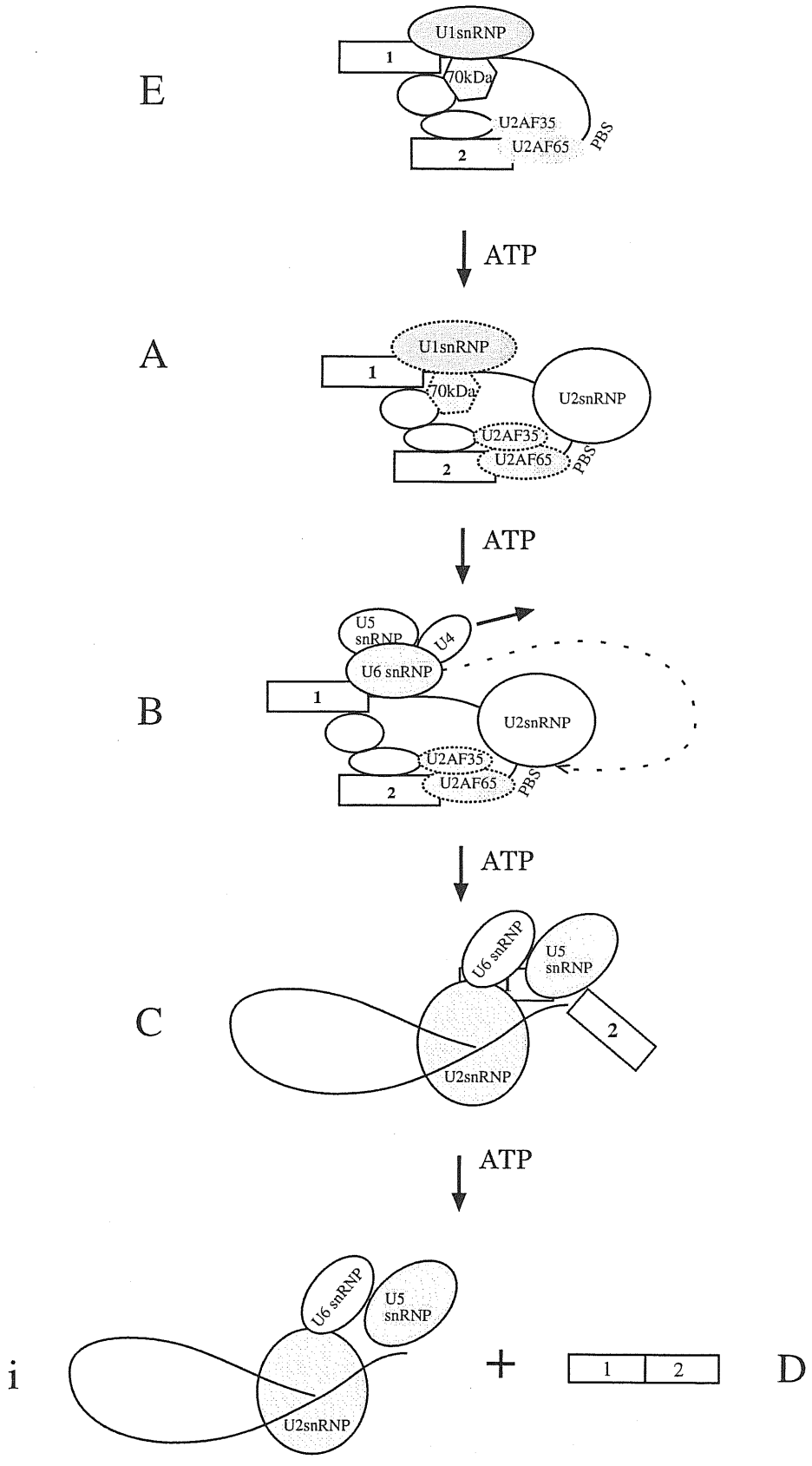


Figure 2.2. General model for spliceosome assembly on pre-mRNA. Spliceosomal complexes, E, A, B, C, D and i are shown. Dashes around U2AF and U1snRNP indicate they become less tightly bound during the E to A to B translocation. Solid arrow indicates U4 snRNP dissociation, and the dashed arrow indicates the association of U2 and U6 snRNA.

U1 involves the formation of Watson-Crick base pairing between the 5′ splice site junction ((A/C)AG/GUAAGU) and a complementary region at the 5′ end of the U1 snRNA [98]. In addition, U2AF with its 65kDa subunit binds to the polypyrimidine tract at the 3′ splice site [99]. In the A complex, U2 snRNP binds to the branch point site (BPS) [100]. U2 snRNP binding requires ATP hydrolysis and a base-pairing interaction between the branch point sequence and a complementary sequence in the U2 snRNA [101]. The base paring interaction between U2 snRNA and the BPS functions to bulge the branch site adenosine, specifying it as the nucleophile attack on the 5′ splice site. U1 snRNP and U2AF become less tightly bound to pre mRNA during the ATP dependent transition from E to A complex [102]. An additional set of polypeptides associated with the U2 snRNP are needed for A complex assembly and the following splicing steps, among these the better characterised are SF1 and SF3 [103]. The B complex is formed subsequently by the binding of U4/U6/U5 snRNP, and when U5 and U6 snRNP base paring interactions replace U1 snRNA at the 5′ splice site [78]. At some point, a major conformational change occurs, as the U4/U6 snRNA helix is disrupted, U4 snRNP dissociates, and an interaction between U2 and U6 snRNAs is established [104]. These rearrangements result in the close juxtapositioning of the 5′ splice site-U6 duplex and BPS-U2 duplex, and these RNA-RNA interactions may establish the active site of the spliceosome (catalytic step I). The C complex contains the products of the catalytic step I of the splicing reaction. U5 snRNA interacts with both exons in the C complex and may function to hold them together for ligation. After catalytic step II, the spliced exons are released from the spliceosome and are detected in a discrete complex of unknown composition (possibly containing SR proteins). The snRNPs are detected bound to the excised lariat intron complex i. The lariat-intron is rapidly degraded while the spliced product will end its processing pathway and will be transferred to the cytoplasm throughout the nuclear pores. It is not well known the fate of the single snRNP when the spliceosome is disassembled but since they have a long

half life it is likely that they reassemble in a new spliceosome (rewieved in [84, 95]).

2.1.3.2. Early (E) complex formation (intron and exon definition model)

There are two principal models of interaction of the two splice sites.

a) Intron Definition Model

The first model suggested that the intron was the unit recognised by the splicing machinery and proposed a scanning mechanism in which the recognition of one site was followed by scanning through the intron to locate the second site [105]. Experimentally, such interactions have been observed with in vitro splicing mRNA having short introns.

i) The yeast model.

In yeast, where messengers often have unique introns and their length is usually below 100 nucleotides, direct supports the the model for the pairing of splice sites across introns during the first step of spliceosome assembly [106].

ii) Vertebrate intron/exon architecture.

In vertebrate most exon 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 [107].

iii) 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 [108]. 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 genes, 5′ splice site mutations of small introns cause intron retention instead of exon skipping as observed in

vertebrates [109, 110]. 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 comprensive of the polypyrimidine tract [109].

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 [111]. 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 most mammalian pre-mRNA which contain very long introns and weakly conserved splice sites [108]. 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 neighbouring exons.

i) Exon size requirement.

The exon length can affect splicing. Internal vertebrate exons have minimum and maximum length requirements [108].

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 bellow 50 nucleotides was skipped by in vivo splicing machinery [112]. In addition, Black et all. [113] have shown that extending the length of the N1 (neuron specific) exon to 109 nucleotides, exon is costitutively 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 [110] 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 [111].

ii) Effect of splice site mutations.

In-vitro [112, 114, 115] and in-vivo [112, 116, 117] 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 [118].

iii) 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 [119]. SR proteins have been implicated in these interactions in different experimental models (see section 2.1.4.2.).

iv) 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. This sequences will be explained in a following chapters.

v) First and last exon recognition.

Since the lack of one functional splice site the first and the last exon should require a special mechanism for their recognition. To have an efficient first intron removal, the capping and the proteins that bind the cap are essential [120]. 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 polyadenilation factors [80].

2.1.3.3. Coexistence and Co-Operation of Intron- and Exon-Based Models

Intron and exon definition mechanism are both validated from the experimental evidences. Depending on the kind of intron/exon architecture, the intron or the exon and the flanking splice sites are the unit recognised by the splicing. In vertebrate, there is a predominant recognition of short exons interspersed between longer introns. In yeast short and often unique introns are recognised. In Drosophila, although most genes fit neatly into the two categories, genes with short introns and large exons and genes with short exons and large introns there are a considerable number of genes in which a mixed architecture is present and in which both mechanisms could operate on the same pre-mRNA.

2.1.3.4. Splice site recognition and bridging interactions

Critical step to the splicing reaction is the initial recognition of the splice sites during the early E complex formation. SR proteins has important role in both, E5′ and E3′ formation, and that is to promote initial recognition of 3′ and of 5′ splice sites interacting directly with the pre-mRNA. Thus, ASF/SF2, SC35 and SRp20 interact with the RNA in the E5′ complex at the 5′ splice site junction sequence [121, 122]. ASF/SF2 and SC35 have been shown to promote U1 snRNP binding to the 5′ splice site via protein-protein interaction with the U1 snRNP component U1 70K [123, 124]. ASF/SF2 and SC35 stimulate E3′ complex

formation as well, binding specifically the 3′ splice site and via a direct protein-protein interaction with the 35 kDa subunit of the splicing factor U2AF [124].
Beside their role in the initial recognition of the splice sites, SR proteins are also proposed to be involved in cross-exon contact and formation of commitment complex. In exon bridging model, $U2AF^{65}$ is bound to the polypyrimidine stretch and contacts $U2AF^{35}$ which is in direct interaction with the SR proteins. This protein forms a simultaneous contact with the SR tail of the U1snRNP 70K protein that is base pared to the 5′ splice site (Figure 2.3.).
The similar protein interactions $U2AF^{65}$-U2AF35-SR-U170K are proposed for the cross-intron reactions [123, 125, 126]. The findings that the pre-binding of SR protein with pre-mRNA is sufficient for commitment to the splicing pathway [92, 97] in an initial step of the spliceosome assembly supports the hypothesis.

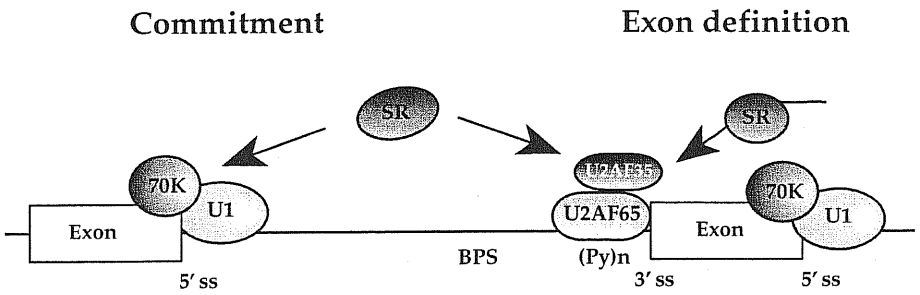


Figure 2.3. Early steps of splice site recognition [127].
5′ splice site is bound to U1snRNPs, 3′ splice site is bound to U2AF65, which exists in association with U2AF35. SR proteins are proposed to mediate commitment between pairs of splice sites through exon and intron bridging reactions.

The findings of Kennedy [128] further supports this model, but suggests that for small introns, the intron bridging is due to the direct U2AF^{65}-SR binding, without involvement of U2AF^{35}. SR proteins involved in intron-bridging could be a members of group of matrix SR proteins (group of proteins which lacks an RNA recognition motif) (Figure, 2.4.) [129].

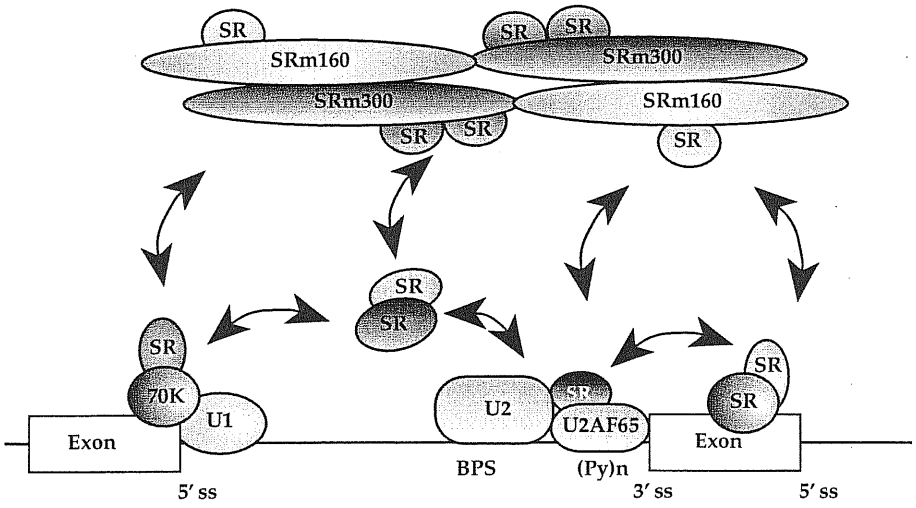


Figure 2.4. Intron bridging model proposed by Blencowe et all.[129]. SR(matrix) proteins promote splice site paring and splicing through multiple cooperative interactions with factors bound to pre-mRNA, including SR proteins, U1 and U2snRNPs and U2AF65.

Recently, Abovich and Rosbash have proposed a model for cross-intron interaction in mammals involving a very different set of the proteins [130]. This model relies on interaction of a mammalian protein SF1, associated with the U2snRNP during the A complex assembly. In addition, SF1 binds to the U2A^{65} and thus promotes its binding to the not well defined 3′ splice site pre-mRNA. The existance of the splicing factor named mPrp40 (its name suggests that it could be a mammalian analog of yeast Prp40), is suggested also for mammals and this factor, might be associated with U1snRNP. Finally, the bridge among mPrp40-SF1-U2AF would be an explanation for the intron bridging interaction (Figure 2.5.) [130] of small mammalian introns.

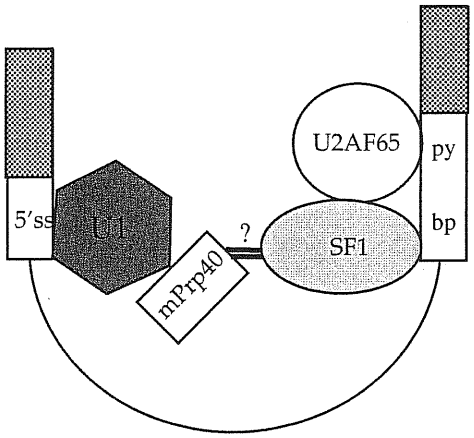


Figure 2.5. Model of cross-intron reaction (Abovich and Rosbash, 1997).
bp, branchpoint; py, polypyrimidine tract; mPrp40p, mammalian analog for yeast Prp40p. In this model, SF1 is proposed to bind brachpoint and promote binding ofU2AF65 to weak 3′ splice site.Its binding to other factors, including yet nonidentified mammalian analog of yPrp40p bound to U1snRNP would form cross-intron reaction.

2.2. ALTERNATIVE SPLICING

Alternative splicing is the process that allows the selection of different combinations of splice sites within a pre-mRNA. Regulated alternative splicing can lead to the production of different proteins from a single pre-mRNA or can function as an on/off control of gene expression regulating the production of a functional or non-functional protein [131-133]. It is a part of the expression program of a large number of genes implicated in the cell growth and differentiation. Some alternative splicing events appear to be unregulated, with mRNA variants coexisting at constant ratios in the same organism where the others can be alternatively spliced in tissue specific (α and β tropomyosin, c-src, fibronectin EDA and EDB) or developmental fashion (retroviruses and adenoviruses). Patterns of alternative splicing can be very complex as a result of use alternative 5′ splice sites, alternative 3′ splice sites, optional exons, mutually excluded exons and retained introns (Figure 2.6.). With the exception of intron

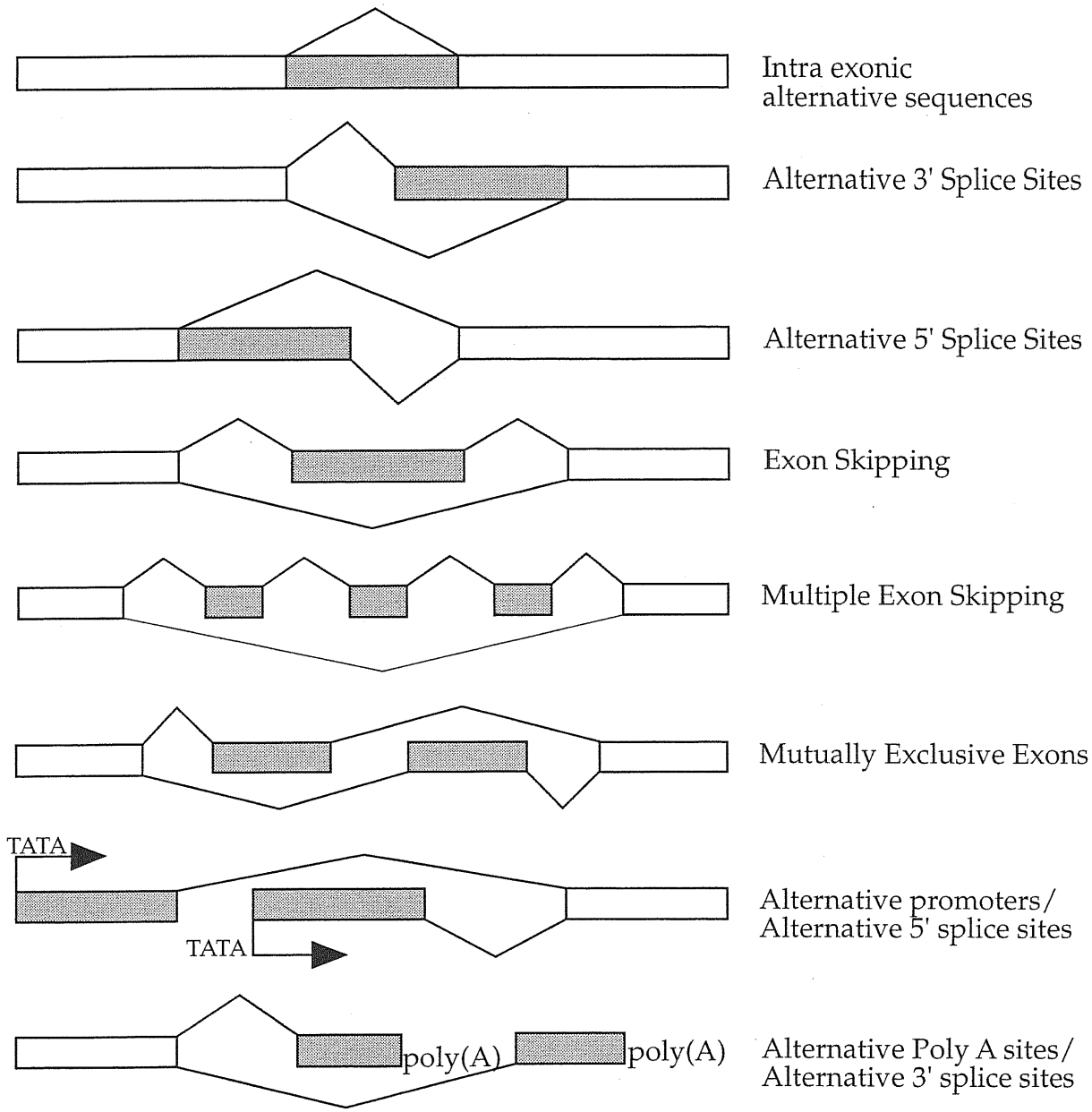


Figure 2.6. Various types of alternative splicing events.
Schematic representation of various alternative splicing events that occur in vivo. Open boxes represent constitutive exons, shadowed boxes represent alternatively spliced exons. The last two examples describe how splicing and other processes are connected. Alternative 5′ splice sites can give alternative 5′ ends and thus alternative initiation of translation. Alternative 3′ splice sites can give alternative polyadenilation sites.

retention, most alternative splicing decisions involve the competition among potential splice sites; thus, splicing patterns can be controlled by any mechanism that alters the relative rates of splice site recognition. Apparently similar patterns can involve fundamentally different pathways.

Mutations within the splice site can lead to nonefficient splicing.

A survay of mammalian mutations available in the database of 1994 indicated over 100 splice site mutations have been characterized in disease gene DNA. Exon skipping is the preferred, (51%), although activation of cryptic splicing sites (32%), creation of a pseudo exon within an intron (11%) and intron retention (6%) are also found [134].

The mechanisms that determine which splice sites are utilised and how this selection is regulated in different cell types or developmental stages have still not been precisely defined. Much progress have been made in identifying the cis-acting elements involved in alternative splicing, and the roles of some general factors have been demonstrated.

2.2.1. Involvement of cis-Acting elements in splice site selection

Cis-acting determinants that influence competing splicing pathways include the relative strength of the competing 5′ splice sites, the branch point sequences, 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 RNA secondary structure participates in the regulation of splice site selection by modifying the physical distances within introns, or being involved in the definition of the exons. Furthermore, an additional intronic and exonic sequences are often necessary for correct splice site selection when there are suboptimal splice sites. Depending on the direction in which they drive recognition of splice sites they are called splicing enhancers or splicing silencers.

2.2.1.1. Exon/intron architectural requirements

As was described in the previous sections, the exon size should respect certain rules in order to be well recognised. The length more than 300bp and less than 50 bp seems to be nonfavoured by the splicing machinery resulted in activation of the cryptic splice sites [112, 116].

Moreover, Sterner et al. 1996 showed importance of the intron size for an efficient splicing. In vivo experiments with APTR gene showed different ability of this exon to be included when flanked by introns of different size. In the case of short exon, the splicing reaction will occur regardless the length of the intron. On the other hand, when exons expanded to suboptimal size are surrounded with large introns, the splicing efficiency decreases [135].

2.2.2.2. The 5′ and 3′ splice sites requirements

Splice site strength is an important determinant in splice site selection. Consensus sequences are limited to 4 bases at the 3′ intron/exon junction and to 9 bases at the 5′ exon/intron junction [74]. It is shown that mutation one of first two bases (AG) in intron, immediately downstream of 3′ splice site (PyAG/G), and as well as GU in intron, immediately upstream of 5′ splice site ((A/C)AG/GUPuAGU) completely abolish splicing [136]. Studies of alternative splicing have shown that mutations of a weak splice sites that make them more close to the consensus can led to the constitutive recognition of alternatively skipped exon [137-139] or to the prevalent recognition of an alternative 5′ or 3′ splice site [140]. Mutations of the splice sites give a significant contribution to the human genetic disease: approximately 15% of point mutations that cause genetic disease affect pre-mRNA splicing [141]. In addition, mutations causing aberrant splicing are more frequently found at the 5′ splice site than at the 3′ splice site.

However, the consensus sequences are not sufficient to account for the observed high specificity of splice site selection.

2.2.2.3. The polypyrimidine tract

The polypyrimidine tract is one of the important cis-acting sequence elements directing intron removal in pre-mRNA splicing. It is known that polypyrimidine tract deletion or mutation decreases the splicing efficiency as well as the elongated length of pyrimidine in row can led to improved efficiency of splicing [142].

Further studies by Coolidge, 1997 demonstrated by cis-competition assay that polypyrimidine strength is not determinated solely by its length but also by its composition [143]. The best polypyrimidine tract is composed by a poly-uridine tract, whereas, a poly-cytidine tract is not functional [142]. Then, proximity of the polypyrimidine tract to the 3′ splice site becomes important when the pyrimidine length is limiting. In fact, limiting number of continuous uridines to six demands that these uridines are placed immediately adjacent to the 3′ splice site AG. On the other hand, a polypyrimidine tract with 11 uridine is a high competitive pyrimidine stretch regardless of distance between the branch point and polypyrimidine tract itself [143]. It is shown that functional pyrimidine tracts do not absolutely require continuous uridines. In constitutively included exon 3 of the human Apolipoprotein, Apo AII polypyrimidine tract is composed of a $(TG)_{16}$ dinucleotide repeats [144]. In this case it is likely that total uridine content may relate the splicing efficiency.

2.2.2.4. The branch point sequence requirement

The branch site is a poorly conserved in metazoan. However, a consensus sequence, (YNCURAC), is a required for lariat formation during the spliceosome

reaction. Generally, an adenine residue is the site of branch formation. Mutations of these residues strongly reduce splicing efficiency of the downstream exon [145]. The sequence is usually located between 18 and 40 nucleotides upstream of the 3′ splice site [146]. On the other hand, several studies of the viral and cellular genes have demonstrated that branch point placed at more than 40 nt from the 3′ splice junctions are functional and are essential for the regulation of alternative splicing. One example of distant branchpoints which regulates alternative splicing is the human α tropomyosin gene where the branch point of intron 2 is located more than 100 nucleotides upstream of the of 3′ splice junction and plays a direct role in preventing the two mutually exclusive exons 2 and 3 from being spliced together [147]. Also there are minimal distance requirements between the branchpoint and 3′ splice site junction. For retroviruses the minimal distance is found to be 16 bp [146].

2.2.2.5. Relative strength and cooperation between signals

Several studies have disclosed the synergy existing either between the 5′ and 3′ splice sites and between the polypyrimidine tract and the branch point.
It has been demonstrated that strong sequence within the 5′ splice site of an exon can promote the use of its own 3′ splice site [148]. On the other hand, sequences, upstream 3′ splice site of an exon can facilitate the use of a downstream 5′ splice site [149].
The polypyrimidine tract determines the location of branch point sequence and indirectly the 3′ splice site (that is usually the first AG downstream the branch point). It is also shown that a strong polypyrimidine tract can partially replace a weak but not nonfunctional branch point sequence. As well, as a strong branch point site can partially replace a weak polypyrimidine tract [150]. Finally, also distance between the branch point sequence and the pyrimidine tract is critical for efficient lariat formation [151].

2.2.2.6. RNA secondary structure

The RNA secondary structure has been suggested to play a role in regulation of alternative splicing.

The cis-acting elements of alternatively spliced exons must be exposed in order to be recognised by the different trans acting cellular factors. In the case of the chicken tropomyosin gene, the secondary structure of the primary transcript regulates alternative splicing by completely hiding the muscle specific exon [152]. It was also shown that the Drosophila SR proteins Tra/Tra2 and B52 recognise specific hairpin loop structures in the target RNA [153]. Furthermore, Muro et al., have demonstrated the peculiar secondary structure of the fibronectin EDA exon that localises the enhancer element in an exposed position as a part of a loop region. Then, in the same EDA exon, a silencer element lies in a stem structure and is a determinant element for RNA structure and the exposure of cis elements to trans-acting factors [154].

Altogether, these findings indicate that pre-mRNA structure play a role in alternative splicing either by sequestering specific splice sites or by providing a binding site for cellular factors.

2.2.2.7. Enhancers and silencers

The poorly conserved splice sites and branch site in higher eukaryotic pre-mRNA do not provide sufficient information for the exon recognition. Additional intronic and exonic sequences are often necessary for efficient and accurate choice. These sequences can either promote (enhancers) or repress (silencers) splice site usage.

Enhancers are mostly located close to the splice site that they activate, and changing their location can alter their dependence on particular trans-acting factors [155] determine whether they activate 5′ or 3′ splice sites, or even

transform them into the negative regulatory elements [156]. By in vitro selection procedure, 6 different group of splicing enhancer elements on shared sequence similarity have been identified. Majority of the enhancers are enriched in purines (more than 67% pyrimidine content) [157]. In fact, in vivo many of exonic enhancers share a general consensus sequence (GAR)n and are located in the exons where in general they activate splicing of the upstream intron by promoting use of a weak 3′ splice site [158, 159]. However, beside the (GAR)n-based motifs, there are other sequences found to function as splicing enhancers [160]. For example, A/C rich motif selected by in vitro SELEX procedure is demonstrated to have ability to enhance splicing both in vitro and in vivo [161]. There are examples of more enhancers regulating splicing efficiency of one exon. The effect of more exonic enhancers does not seem to be synergistic, but increases the probability that an enhancing effect might occur [162].

Intronic sequences have also been found to function as enhancers. For example, the neurone specific splicing of γ2 pre-mRNA requires several nucleotides in both regulated exon and the adjacent intron [163]. Also muscle specific inclusion of cTNT pre-mRNA exon 5 depends on four sequence elements located within the upstream and downstream introns [164].

The localisation of the enhancer within the mRNA can be determinant for its function. Some types of enhancers are inactive or require additional trans acting factors, when located more than 300 nucleotide downstream from 3′ splice site [155].

Exonic and intronic sequences which repress the use of splice sites have also been identified. Some of these sequences are involved in formation of mRNA secondary structure limiting the accessibility of splice sites [154], while others give priority to the aberrant splicing.

An example of splicing silencer is the negative regulator of splicing (RNs) element within the gag gene of Rous sarcoma virus. NRS is a bipartite element that lies ~400 nucleotide downstream of viral 5′ splice site. It is likely that NRS

inhibition involves U1snRNP and it functions by interacting nonproductively with the 3′ splice site [165]. It is shown that NRS can inhibit splicing of heterologous introns in vivo [166].

A negative element (CAAGG) involved in the down regulation of EDA exon inclusion was identified within the EDA exon [167] where it seems to be determinant for the RNA conformation [154]. Recently, an additional splicing silencer within the same exon was mapped [168].

Negative effect of silencers can be result of their interactions with protein factors, usually hnRNPs. Binding of the hnRNPs to some sequences in pre-mRNA can interfere with the access of the splicing factors to the splicing sites. The UCUU sequence (binding site for PTB) and TAGGGCTTAGGGT (consensus sequence for the hnRNPA1) are found to be functional silencers [169, 170] .

2.3. Trans acting factors

There are at least two protein families involved in the regulation of alternative splicing. In the mammalian cells, the most analysed trans-acting factors that regulate alternative splicing is the family of SR proteins. Another family of the proteins with a potential role in the splicing regulation are hnRNP proteins. They are associated with newly synthesised mRNA during its processing as well as with the mature mRNA during nucleocytoplasmic transport. They are not directly involved in splicing but their presence on the pre mRNA during spliceosome assembly can influence splicing.

These factors modulate the utilisation of the specific 3′ and 5′ splice sites particularly when splice site sequences and intron size are nonoptimal.

2.3.1. SR proteins as a alternative splicing regulators

SR proteins are required for constitutive splicing but may also regulate alternative splicing patterns or by altering the 5′ splice site choice in pre-mRNAs containing competing 5′ splice sites or by promoting alternative exon inclusion [123, 171]. In 5′ splice site competition assays, different SR proteins show distinct preferences for promoting the use of distal or proximal splice sites. Thus, higher concentrations of ASF/SF2 or SC35 proteins promote the use of 5′ splice site proximal to the 3′ splice site, while SRp40 and SRp55 promote the use of the distal sites [91]. It has been proposed that ASF/SF2 affects 5′ splice site selection by increasing the occupancy by U1 snRNP of all the competing sites [123, 172]. An intrinsic tendency to splice closest neighbours bound by splicing components would then favour selection of the proximal site.

SR proteins recognise and bound enhancer sequences that are often purine rich. The proposed mechanism for promoting the use of nearby splice site by SR proteins is the following; the $U2AF^{35}$ subunit establishes a protein-protein interaction with certain SR proteins. Also, it binds to $U2AF^{65}$ and in this way forms a bridge between $U2AF^{65}$ and SR protein facilitating the cooperative binding of SR protein to enhancer sequences and $U2AF^{65}$ subunit to the polypyrimidine tract [124]. This theory is supported with the observation that an increase in the polypyrimidine content of the some regulated 3′ splice sites make them enhancer-independent.

In some cases, SR proteins may act as a splicing repressors, either by binding to sites that sterically occlude spliceosome assembly, or by blocking the binding of more active SR proteins.

A purine rich sequence located just upstream of a regulated 3′ splice site in the adenovirus late transcript inhibited use of this site when an excess ASF/SF2 protein was added to the nuclear extract [156]. This observed inhibition is the result of the interference with recruitment of U2 snRNP to the nearby branch

site. Interestingly, the same purine rich sequence that was target binding site for ASF/AS2, was found to function as a splicing enhancer, when placed in the downstream exon. Thus, it is not a SR protein itself one that will determine positive or negative function but the position of the sequence on the pre-mRNA that the protein will recognise.

2.3.2. hnRNP as a splicing regulators

The hnRNP comprise a large family of proteins (reviewed in [67, 173-175]). At least 30 major hnRNP proteins are described for mammalian cells, ranging in molecular weight from 34 to 120 kDa. hnRNP proteins share structural and functional features with protein components of stable spliceosomal particles (snRNPs) or splicing regulators; they contain different numbers of domains with common RNA binding motifs and auxiliary domains that mediate protein-protein interaction. The multiple isoforms of single hnRNP are produced either by alternative splicing of encoding genes or by the postranscriptional modifications.

One of the best analysed hnRNP proteins is hnRNPA1. Several intronic splicing enhancers and repressors associate with hnRNPA1. A1 protein has activity that promotes the annealing of complementary RNAs suggesting that it may be able to alter the structure of pre-mRNA and its interaction with other splicing factors. A1 protein has been shown to regulate 5′ splice site choice of pre-mRNAs containing alternative 5′ splice sites [176]. The relative concentration of A1 protein and the splicing factors ASF/SF2 or SC35 determines whether the proximal or the distal 5′ splice site is used [176-178]. The two protein are antagonistic. High level of A1 protein favour distal 5′ splice site choice while high levels of ASF/SF2 cause the selection of the proximal 5′ splice site. The mechanism of action is not completely understood but it is demonstrated that the presence of a strong hnRNPA1 binding site between the competing 5′ splice

sites shift selection towards the distal site but does not modify U1 snRNP binding to any of the 5′ splice sites [170].

The hnRNPAI exhibits a binding preference for the 3′ end of the introns [179]. Mutations of the conserved AG dinucleotide at the 3′ splice site impair the binding of A1 protein to pre-mRNA in vitro [179]. Recently, it has been proposed that hnRNPA1 can promote, as well, distal 3′ splice site in calcitonin CGRO transcript in vivo by antagonising the proximal site favoured by ASF/SF2 [178].

Another member of this family is hnRNPI, named also the polypyrimidine binding protein (PTB). Initial reports showing correlation between PTB binding and 3′ splice site selection suggested that it may function as an activator of branch point/polypyrimidine tract [180-182]. Later work suggested that PTB might function as a negative regulator of splicing [183] by selectively inhibiting the in vitro splicing of introns containing strong PTB binding site.

PTB recognises specifically the UCUU sequence within the polypyrimidine rich context where it competes with the binding site for $U2AF^{65}$ [184]. Therefore, excess of PTB inhibits the binding of the $U2AF^{65}$ to the polypyrimidine tract, blocking the assembly of functional A complexes.

In $GABA_A$ receptor $\gamma 2$ subunit has a specific processing pattern in the neural cells which include exons that in other cell types are not present. PTB is shown to bind upstream of a neural specific exon repressing both, the 3′ splice and also 5′ splice site [185]. The effect of on the 5′ splice site also involves a PTB-binding element within the exon. In this way PTB may help to form a large complex involving all elements that represses the 3′ and 5′ splice sites selection simultaneously.

Recently, it has been shown positive regulation by the PTB protein. In human calcitonin gene PTB binds to a pyrimidine rich tract of enhancer located adjacent to an alternative 3′ terminal exon. Inclusion of the alternative exon increases as the level of PTB in vivo increase [186]. Also, altering the enhancer pyrimidine tract to a consensus sequence for the binding of U2AF eliminated enhancement of exon inclusion in vivo and exon polyadenylation in vitro [186].

PTB has important role in the regulation of the tissue specific splicing.

Beside these two hnRNP for which is shown that they are involved in regulation of alternative splicing, there are other hnRNP which potential role in splicing regulation is less studied. Different hnRNP vary in their RNA binding specificity. Thus, hnRNP F/H, E and U proteins recognise poly G, the hnRNPC recognise poly U sequence (reviewed in [173]. For hnRNPH protein is demonstrated its involvement in splicing regulation of exon 7 of rat β-tropomyosin by binding to its exonic silencer [187]. Furthermore, a heterodimer formed by hnRNPH and hnRNPF is shown to be positively involved in neuronal specific splicing of c-src N1 exon [188, 189].

The hnRNPC which binds preferably to the U rich polypyrimidine region around the 3′ splice site of pre-mRNAs and in downstream regions important for 3′-end cleavage and polyadenilation [179, 190-192]. Mutations of the AG dinucleotide at the 3′ end of the intron strongly inhibit A1 binding but do not impair the binding of the C protein [179]. There are some evidence that this protein can be involved in alternative splicing but the precise role is still not known.

2.4. Relative concentration of splicing factors and regulation of alternative splicing

In some cases it is shown that different level of certain splicing factors can affect splice site choice thus affecting the expression of a wide number of genes. In the more complex systems determinant factor for gene expression is the concentration of one factor relative to the level of factors with opposing effect. Studies of 5′ splice site competition show that activity of SR protein in promoting the use of proximal 5′ splice sites was antagonised by hnRNPAI and related proteins [89, 127, 193]. It has been shown that variations in the level of different SR proteins as well as the level of hnRNP A/B vary between different cell types. In rat, it was found that molar ratio of hnRNP and ASF/SF2 can vary at least 100

fold among different tissues, indicating that the concentration of splicing factors could be a force able to drive a switch between some alternative splice sites [194].

There is an antagonistic effect also among the different SR proteins. This is the case of chicken β tropomyosin gene that contains a pair of mutually exclusive exons; exon 6A present in smooth muscle cells and fibroblast and 6B specific to a skeletal muscle. A pyrimidine rich element (S4) in the downstream intron is essential for recognition of the 5′ splice site of exon 6A. ASF/SF2 binds to S4 element and stimulates splicing of 6A while SC35 can antagonises the S4 stimulatory effect of ASF/SF2 [195]. The fact that the ratio of SC35 to ASF/SF2 is at least two fold higher in SR preparations of muscle cells than in those from HeLa supports the hypothesis that variations of ratio of these two proteins might influence the splicing efficiency of exon 6A in vivo [196].

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

Ratio of the different proteins can be controlled in the tissue specific of developmentally regulated manner by regulation of the expression of their genes either at the transcriptional or postranscriptional level.

It has been shown that genes which code for SR proteins such as a SC35, SRp20 and SRp40 can be upregulated by mitogenic stimuli and this in turn influence the alternative splicing by affecting the recruitment of different SR proteins to a transcript. Then, analysis of the hnRNPAI gene has revealed a complex promoter, consistent with the modulation of its expression in response to different cellular and physiological stimuli [197] indicating the similar way of modulation control of hnRNP for its binding to pre-mRNA.

Postranscriptional regulation of SR protein gene expression is establish at the level of splicing. Several alternative spliced mRNA variants of SR proteins have been found in cDNA libraries prepared from various tissues [198]. It is shown

that ASF2 and ASF3 alternative form of ASF/SF2 protein can downregulate the expression of ASF/SF2 but at the same time are very unstable and cannot replace ASF/SF2 in control of alternative splicing. Different transcription forms of hnRNP AI have also been detected and it seems that are much less active than AI, in vivo and in vitro [199].

Additional levels of control for trans acting elements occur at the postranslational level. Both, SR and hnRNP are phosphorylated in vivo and the phosphorylation state is important for splicing. It has been shown that phosphorylation state of the SR proteins can control their subnuclear localisation and trafficking between speckles and sites of transcription. When proteins are dephosphorylated they tend to self associate, so contributing to the speckles forming, while phosphorylation weakness this interaction and facilitate SR protein dispersion in the nucleus and also their function in the regulation of splicing [200].

2.6. The Drosophila sex determination pathway

(the best characterised example of alternative splicing)

The sex determination pathway in Drosophila melanogaster is the best characterised example of alternative pre-mRNA splicing regulation [132, 201] and deserve to be discussed.

The primary sex determining signal is the number of X chromosomes in relation to the number of sets of autosomes in the genome (X:A ratio). The products of several X-linked genes (*sis-a, sis-b, run*) and at least one autosomal gene (*dpn*) determine the function of the gene Sxl (sex-lethal) that is responsible for the different morphology of the two sexes. Sxl, that is a RNA-binding protein controls three alternative splicing decisions. In females, it activates its transcription from the (early) promoter located downstream from the promoter that is used in both sexes and in its own RNA represses inclusion of a default

exon. The way of action of the sxl in females is following: it binds to the polypyrimidine tract of exon 3 and in this way it competes with U2AF "blocking" this splice site and promoting the utilisation of the downstream 3′ splice site flanking exon 4 [202]. Exon 3 contains several stop codons and its skipping allow the synthesis of functional sxl protein. In this way, it initiates a positive feedback loop that maintains functional Sxl expression in females.

With the same mechanism the sxl protein mediates exclusion of exon 2 of the transformer (tra) mRNA [131]. Tra pre-mRNA exon 2 also contain stop codons and its exclusion results in the production of the female specific tra protein. Tra protein together with transformer 2 (tra2), a protein expressed in both male and female, controls the splicing of the alternative splicing of doublesex (dsx) pre-mRNA. The male (default) splicing of dsx pre-mRNA produces an mRNA consisting of exons 1, 2, 3, 5 and 6 encoding a factor necessary for male sexual differentiation. The female splicing produces a mRNA consisting of exons 1, 2, 3 and 4, which encodes a factor necessary for female differentiation [203]. Tra and tra2 activate the 3′ splice site upstream exon 4 which has a low consensus sequence and is not recognised by the splicing machinery. Tra2 binds directly to a regulatory element (enhancer) located 300 nucleotides downstream the female-specific 3′ splice site characterised by the presence of 6 repeated 13-nucleotides repeated elements and a polypurine sequence, named the dsxRE element [204-206]. Tra, tra2 and SR proteins assemble in a multiprotein complex on the dsxRE element and are essential to commit dsx pre-mRNA to the female pathway of splicing [205, 207]. Similar to the SR proteins tra2 contains both a RRM and a SR domain while tra contains only a SR domain. Tra and tra2 interact with each other, with SR proteins and with other general splicing factor via their SR domain [125, 208]. A multiprotein complex assembled on the dsxRE facilitate the assembly of spliceosomal components on the weak female-specific 3′ splice site through a network of protein-protein interactions similar to the one proposed for pre-mRNAs containing weak splicing sites and ESE sequences . Dsx is thought to be the final regulatory gene in the pathway. It binds directly to gene enhancer sequences and has been proposed to be a transcriptional repressor [209].

Finally, the dosage compensation in Drosophila requires the male-specific lethal (msl) proteins to make a gene expression from a single male X chromosome equivalent to that from both female X chromosomes. Expression of the msl2 in females is repressed post-transcriptionally by the binding of Sxl to the multiple sites in the 5′ and 3′ untranslated regions of the msl12. By its binding, Sxl block splicing of an intron in the 5′ untranslated region and thus preventing the hypertranscription of the X-linked genes [210].

CHAPTER 3

RESULTS

3.1. Analysis of the human CFTR exon 9 splice sites

Up to now, no structural and functional studies of the cis-acting elements and trans-acting factors involved in the regulation of the human CFTR exon 9 alternative splicing have been performed.

In our study, as a first approach, we have carried out the analysis of the primary sequences of human CFTR exon 9 splicing sites (including the 3′ and 5′ splice junctions, the polypyrimidine tract and the branchpoint) (Figure 3.1).

The comparison of the 5′ splice junction with the consensus sequence has shown that it differs in position +4, where consensus base A is replaced by another purine, G, and in position +5, where the consensus purine G is replaced by a pyrimidine T. On the other hand, the analysis of the 3′ splice site has shown a perfect match with the consensus.

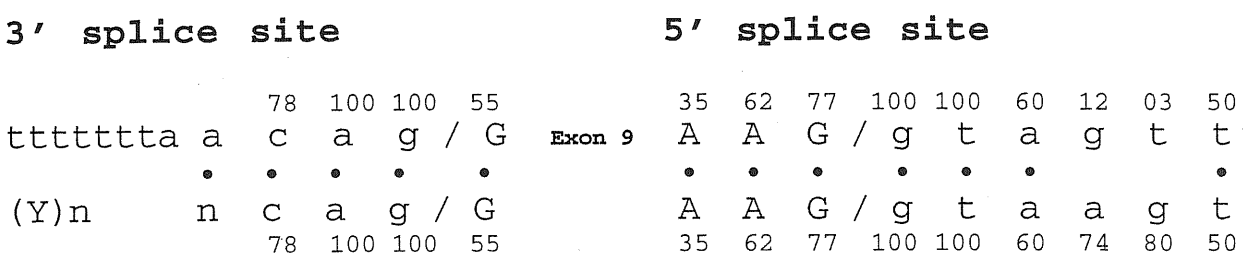
Then, the region between the branchpoint and the 3′ splice junction contains two polymorphic stretches: a polypyrimidine tract presenting three allelic variants (with 5, 7 or 9 thymidines) and, immediately upstream of the poly-T, another polymorphic tract containing an alterned pyrimidine/purine TG dinucleotide repetition with 5 allelic variants (with 9, 10, 11, 12, or 13 TG repeats).

Finally, two potential candidates for the branchpoint sites were selected using FindPattern program (GCG package). A criteria for the selection were used a good match of the sequence with the YNYURAC branchpoint consensus together with its relative position in respect with the exon 9 3′ splice site junction (first downstream AG).

The first candidate sequence was found to be placed at around 60 bases upstream of the intron/exon junction (Figure 3.1.). The second, instead, is placed at around

40 bases upstream of the intron/exon junction (the distance varies depending of the length of the TG/T stretches).

a



b

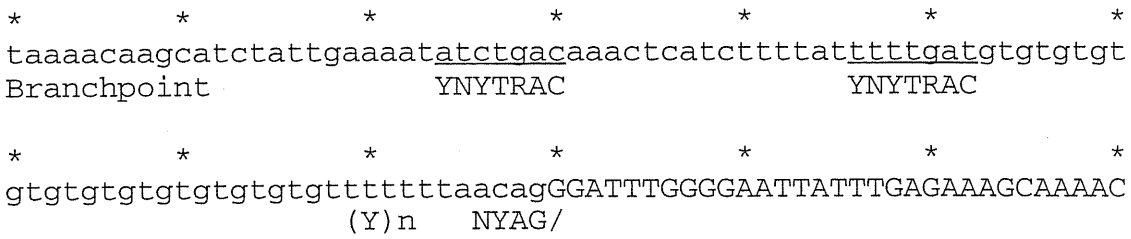


Figure 3.1. Comparison of splice sites from the human CFTR exon 9 with consensus sequences for 5′ and 3′ splice sites. a) 3′- and 5′-splice sites of human CFTR exon 9 (upper line) are compared with consensus sequences (lower line). Points indicate homology between the two sequences. Respective base frequencies are also indicated for both, consensus sequences and exon 9 splice site junctions. N indicates any nucleotide, Y pyrimidine and R purine. Upper case, exon 9; lower case, introns. b) Organisation of the 3′ splice site of human CFTR exon 9 showing the two potential branch points (underlined, consensus YNYTRAC), the polypyrimidine tract, (Y)n, and the 3′ splice junction, (consensus, NYAG/) .

3.2. Cis-acting elements and their involvement in alternative splicing

Our strategy to study the cis acting elements that influence the alternative splicing of human CFTR exon 9 was the following: A CFTR minigene system (named pSVCF) was created by cloning the CFTR fragment of interest, encompassing exon 9 and its flanking intronic sequences (221 bp of intron 8 and 269 bp of intron 9), within the the pSV-mEDAΔNdeI vector. This vector derived from the pSV-mEDA vector by deletion of 1066 bp NdeI fragment containing EDA exon together with the part of the flanking introns carrying the cis-elements

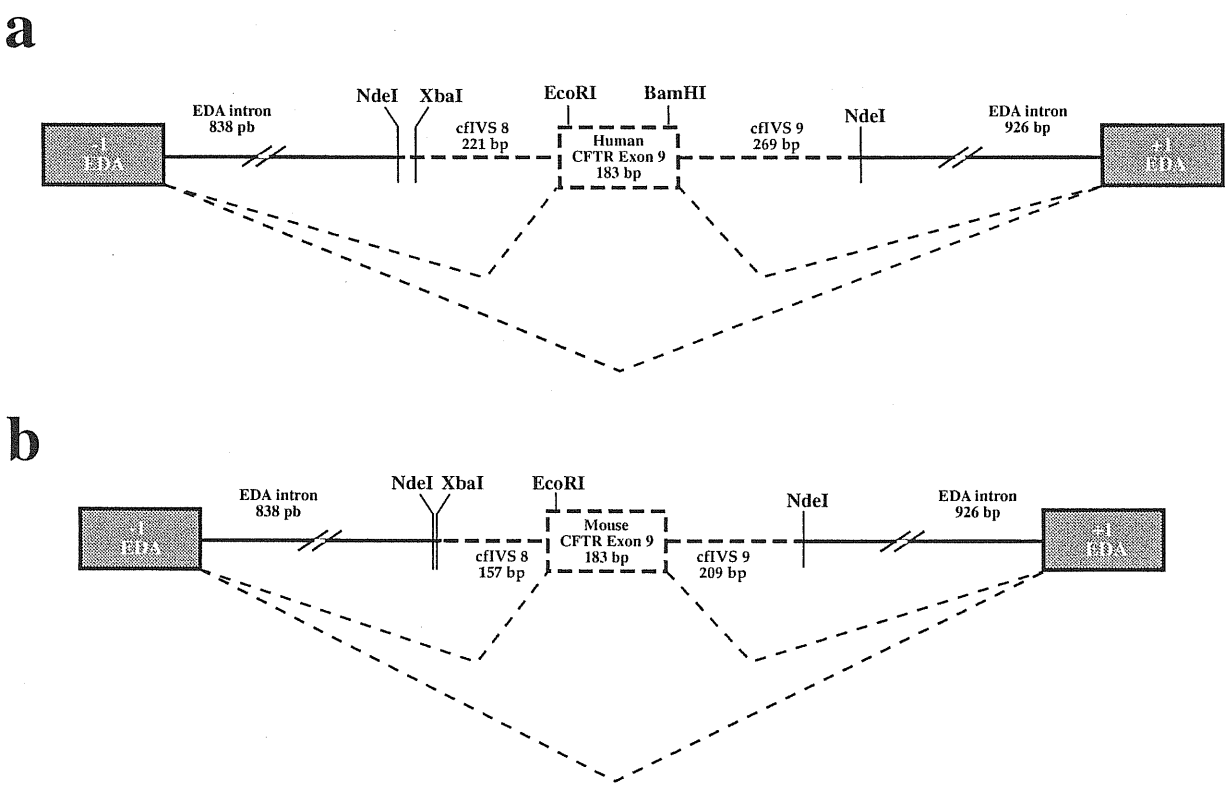


Figure 3.2. Schematic representation of human and mouse minigenes. a) Human exon 9 and its flanking intronic sequences cloned into NdeI site of pSV-mEDA plasmid. **b)**. Mouse exon 9 and its flanking intronic sequences cloned into NdeI site of pSV-mEDA plasmid. Sizes of CFTR exon 9, CFTR introns 8 and 9 and EDA introns are shown as well as position of relevant restriction enzyme target sites. Dot lines depict exon 9 alternative splicing. EDA+1, EDA-1, fibronectin EDA exons of the pSV-mEDA plasmid.

involved in regulation of EDA splicing [139] (Figure 3.2.a). Then, series of mutants were created by deletion or by substitution with related mouse sequences or non related sequences within the CFTR region of the original pSVCF vector. After transient transfection into suitable human and murine cell lines, RNA was extracted and retrotranscribed. Finally, the splicing pattern was determinated by quantitative analysis of the PCR products amplified using the oligonucleotides specific for the product of the expression vector. In this way, the effects on splicing regulation of the cis-acting elements within the cloned CFTR minigene were directly evidenced.

3.2.1. Tn-tract and alternative splicing of CFTR exon 9

Three constructs have been designed and used for the studies of the effects of the poly-thymidine tract length on alternative splicing. Each construct contained fixed number of (TG) dinucleotide repeats (TG11) whereas the number of poly-thymidine was variable and included five (pTG11/T5), seven (pTG11/T7) or nine (pTG11/T9) thymidines (Figure 3.3.a.).

Transient transfections into the NCI-H322 cell line followed by RT-PCR using primers specific for our minigene showed that the constructs carrying T9 or T7 genotypes generated mRNA with 100% exon 9 inclusion, whereas mRNA form lacking exon 9 was detectable only when with T5 genotype.

To characterise better the weakening of the exon 9 definition mediated by impairing the length of the polyU tract we have investigated whether it is due to the spacing-specific effect or instead to the sequence content within this polymorphic region. A set of mutants was generated: one in which the T-tract was completely deleted, leaving the (TG)11 repeats adjacent to the 3'-splice site of exon 9 (pΔT), and others in which T-tract was replaced by unrelated sequences, i.e. a non-polypyrimidine A-tract, creating two constructs carrying A5 (pTG11/A5) or A7 (pTG11/A7) (Figure 3.3.a.).

a

IVS 8 Exon 9

Tn

```
p(TG)11/T9   ...aactcatcttttatttttgatgtgtgtgtgtgtgtgtgtgtttttttttaacagGGATTTGGGG...
p(TG)11/T7   ...aactcatcttttatttttgatgtgtgtgtgtgtgtgtgtgtg--ttttttaacagGGATTTGGGG...
p(TG)11/T5   ...aactcatcttttatttttgatgtgtgtgtgtgtgtgtgtgtg----ttttaacagGGATTTGGGG...
p(TG)11/A7   ...aactcatcttttatttttgatgtgtgtgtgtgtgtgtgtgtg--aaaaaaaaacagGGATTTGGGG...
p(TG)11/A5   ...aactcatcttttatttttgatgtgtgtgtgtgtgtgtgtgtg----aaaaaaacagGGATTTGGGG...
p(TG)11/ΔT   ...aactcatcttttatttttgatgtgtgtgtgtgtgtgtgtgtg---------aacagGGATTTGGGG...
```

b

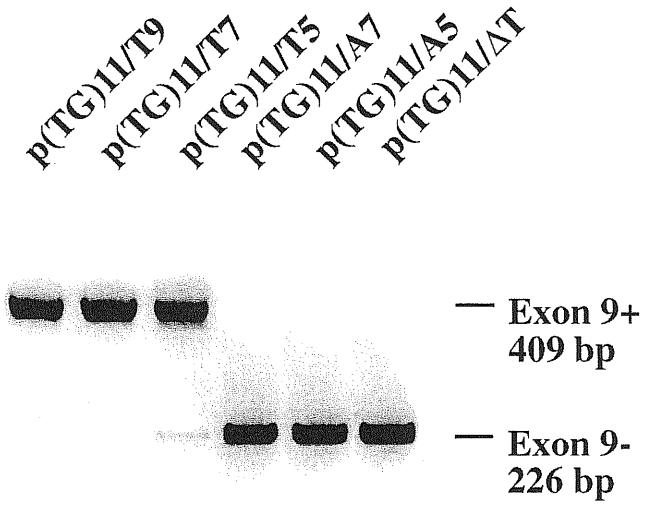


Figure 3.3. Effect of T-tract on alternative splicing of human exon 9. a) Sequence of intron 8 3′-end and exon 9 splice junction of constructs carrying 11-fixed (TG) repeats juxtaposed to 9 or 7 or 5 thymidines (p(TG)11/Tn). In p(TG)11/A7 and p(TG)11/A5 constructs 7 or 5 adenines replaced 7 or 5 thymidines, respectively. In p(TG)11/ΔT construct the T-tract was completely deleted. **b)** Analysis of pre-mRNA splicing of human constructs in NCI-H322 cells. The size of PCR products including (409bp) and excluding (226 bp) exon 9 are indicated. Amplicons were separated on 2.0 % (w/v) agarose gel. Proportion of exon 9 exclusion was determined by densitometry using the total densitometric units of 9+ and 9- amplicons as 100%. p(TG)11/T5, 6%.

Figure 3.3b shows how all these constructs transfected into NCI-H322 cell line lead to expression of mRNA totally excluding exon 9, thus demonstrating that the presence of a poly-thymidine tract in CFTR IVS 8, is essential for the utilization of the exon 9 3′ splice site. Secondarily, it is also shown that a stretch of polypyrimidines alterned with purines (within the TG tract) does not work as a functional continuous polypyrimidine tract.

3.2.2. T/(TG) repeats interplay and alternative splicing of CFTR exon 9

Previous observations in human subjects suggested that the (TG) dinucleotide repeats could affect the alternative splicing of exon 9 in a way independent of the allele found at the T locus [63]. We analysed how the length of the (TG) tract might influence the efficiency of exon 9 inclusion, using minigene system.
To this end, a new set of constructs was generated, where T5 allele was linked to 11, 12 or 13 (TG) repeats, respectively (Figure 3.4.a.).
Transient transfections of these constructs into NCI-H322 cell line have shown that, for the T5 allele, the proportion of exon 9 exclusion is further modulated by the number of (TG) repeats, since the longer is the (TG) tract the higher is the proportion of transcripts devoid of exon 9 (Figure 3.4.b.). These findings demonstrate that the exclusion rate of exon 9 is inversely proportional to the length of the T-tract and directly proportional to the number of (TG) repeats. This is not the case for the T7 allele: in fact, there is no weakening of exon 9 recognition occurs in constructs which carries T7 allele regardless of the number of TG repeats (Figure 3.4.b.). These results suggest that the (TG) repeats cannot control directly exon 9 skipping, i.e. independently from the T-tract, but they can only modulate alternative splicing determinated by T5 allele.

a

IVS 8 (TG)m Tn Exon 9

p(TG)11/T7 ...aactcatcttttatttttga----tgtgtgtgtgtgtgtgtgtgtttttttaacagGGATTTGGGG...

p(TG)13/T7 ...aactcatcttttatttttgatgtgtgtgtgtgtgtgtgtgtgtttttttaacagGGATTTGGGG...

p(TG)11/T5 ...aactcatcttttatttttga----tgtgtgtgtgtgtgtgtgtgtg--tttttaacagGGATTTGGGG...

p(TG)12/T5 ...aactcatcttttatttttga--tgtgtgtgtgtgtgtgtgtgtgtg--tttttaacagGGATTTGGGG...

p(TG)13/T5 ...aactcatcttttatttttgatgtgtgtgtgtgtgtgtgtgtgtg--tttttaacagGGATTTGGGG...

b

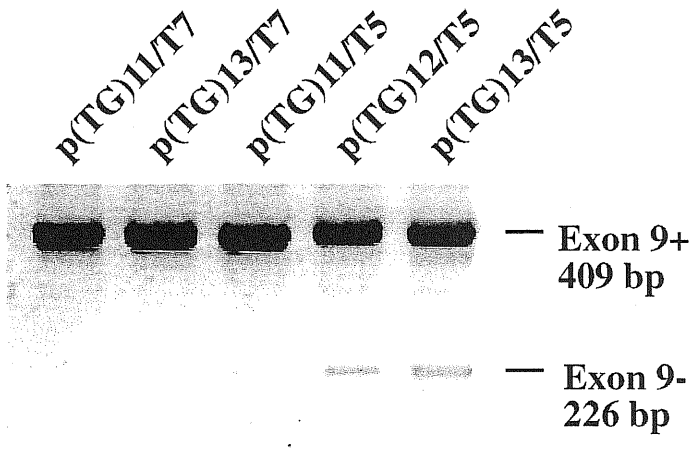


Figure 3.4. Effect of TG/T interplay on alternative splicing of human exon 9. a) Sequence of intron 8 3'-end and exon 9 splice junction of constructs carrying variable number of (TG) repeats juxtaposed to 7 or 5 thymidines (p(TG)11/Tn). b) Analysis of pre-mRNA splicing of human constructs in NCI-H322 cells. The size of PCR products including (409bp) and excluding (226 bp) exon 9 are indicated. Alternative splicing form is observed only when five thymidines are present. PCR products were separated on 2.0 % agarose gel. Proportion of exon 9 exclusion was determined by densitometry using the total densitometric units of 9+ and 9- amplicons as 100%. p(TG)11/T5, 6%; p(TG)12/T5, 10%; p(TG)13/T5, 18%.

3.2.3. 5′ splice site involvement in the recognition of the exon 9

The proper association of 5′ and 3′ splice sites is a critical step for pre-mRNA splicing.

To explore the structural reasons of this weak exon definition, a mutation was introduced in the donor splice site of exon 9. Sequencing of cloned PCR products has revealed that a cryptic 5′ splice site splice site, just 5 bases upstream of the default 5′ splice site revealing an exception to the "Chambon GT rule" (Figure 3.5.). Transfections with constructs carrying leaky exon 9 5′ splice site showed again that only T5 allele weakens exon 9 definition, and that the number of (TG) repeats acts in synergy with T5 allele to modulate the exclusion rate of exon 9 (Figure 3.5.). The concerning the results of transfections, experiments suggest that the 3′-splice site may affect exon 9 definition more significantly than the 5′-splice site.

3.3. Species-specific alternative splicing of CFTR exon 9

The CFTR coding region is highly conserved throughout the evolution, as evidenced by the high degree of nucleotide identity among the species. The comparison between human and mouse cDNA has disclosed a degree of identity of 83%, suggesting that the CFTR protein has an important role for the development of the higher eukaryotes [211, 212].

Afterwards, both in human and mouse, several species-specific CFTR splice mRNA variants have been found (Figure 3.6.a) and, in particular, the 9- mRNA form seemed to specific for humans [50].

We have tried to confirm this observation, amplifying by PCR the cDNA synthesised using the RNA extracted from different mouse tissues (lung, brain, intestine and testis). In each analysed tissue, only the 9+ mRNA form was detected (Figure 3.6.b).

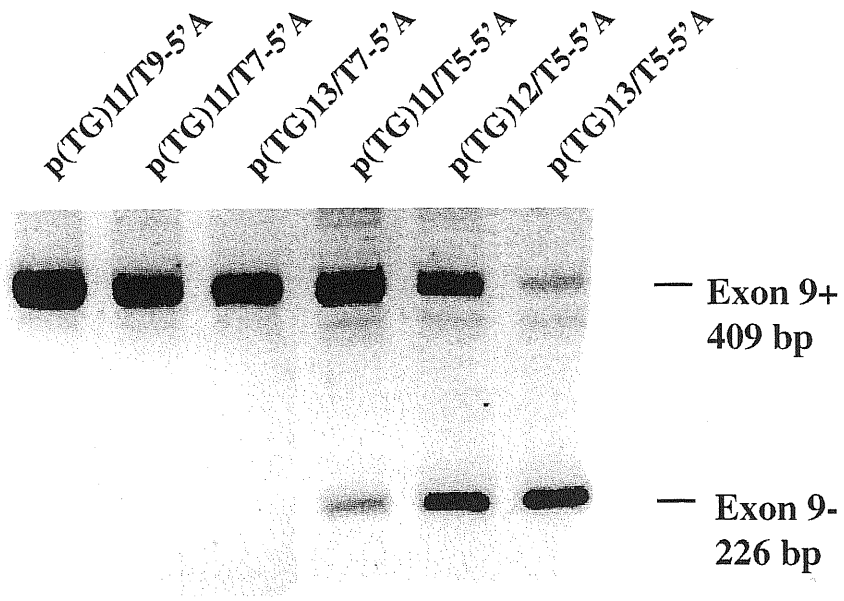


Figure 3.5. Effect of 5'-splice site mutation on alternative splicing of CFTR exon 9. The influence of exon definition on 3'-splice site recognition was investigated through mutagenesis of the 5'-splice site of exon 9 by replacing guanine in position +1 within intron 9 with an adenine (AAG/*g*tag -> AAG/*a*tag). Analysis of pre-mRNA splicing of human constructs mutagenised at 5'-splice site shows activation of exon 9 skipping only when five thymidines are present. PCR products were separated on 2.0 % agarose gel. Proportion of exon 9 exclusion was determined by densitometry using the total densitometric units of 9+ and 9- amplicons as 100%. p(TG)11/T5-5'A, 18%; p(TG)12/T5-5'A, 52%; p(TG)13/T5-5'A, 75%.

a

Splice variant	Mouse	Human
4-	<1	<2
5-	20-40	ND
4-/5-	1-8	ND
9-	ND	10-92
11b-89 bp	10-15	ND
11b-108 bp	<1	ND
12-	ND	5-30

Proportions of CFTR variants are expressed as a percentage of total CFTR transcript. ND, not detected.

Nature Genetics 1993, 4, 426

b

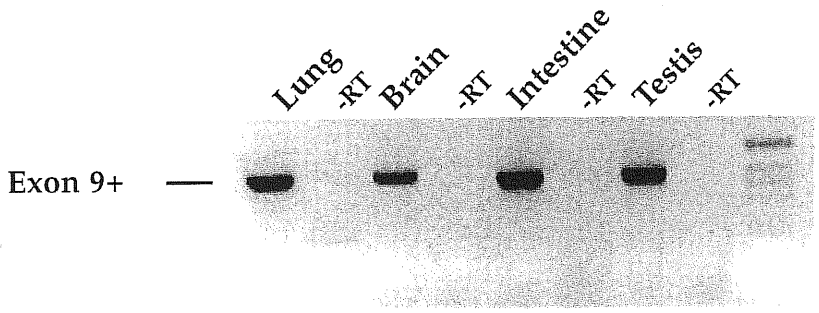


Figure 3.6. a) Different splice variants between human and mouse. 11b-89pb variant contains an additional exon containing 89bp of intron 11, 11b-108pb variant contains an additional 108 bp of intron 11. b) Mouse tissues do not express exon 9- of CFTR pre-mRNA. cDNA from different mouse (CD1 strain) tissues was amplified by PCR using the oligonucleotides spanning from exon 8 to exon 11.

These findings led us to investigate the molecular basis of the species specific alternative splicing of CFTR exon 9.

3.3.1. Difference in exon size between human and mouse introns

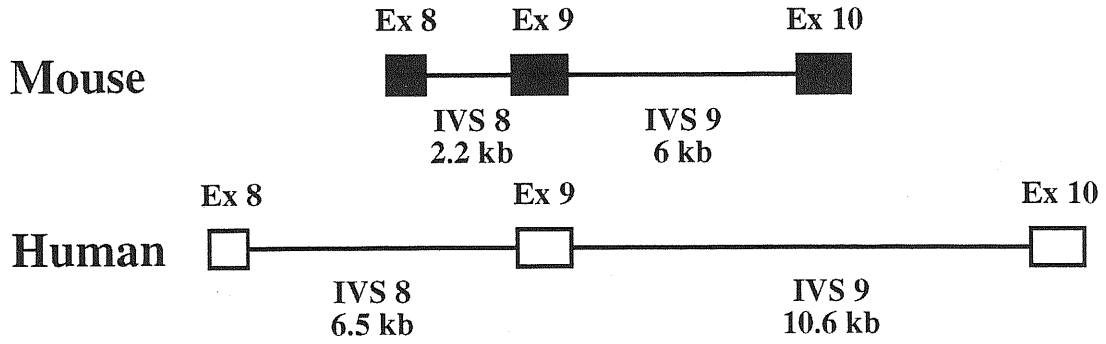
As a first step toward characterisation of the splice sites in the mouse gene, we amplified by long range PCR the genomic region of the murine (CD1 strain) CFTR gene spanning from the exon 8 to 9, and from exon 9 to 10. Surprisingly, a striking difference was observed in the size of both mouse introns compared to the size of human ones: mouse intron 8 is about 2.2 kb long , that is one third of the size of human intron 8; whereas mouse intron 9 was ~6 kb long, that is about half of the size of the human intron 9 (Figure 3.7.).

The differences in the size of introns between the human and the mouse CFTR exon 9 suggest the occurrence of rearrangement events within human CFTR gene during the evolution.

3.3.2. Comparison between human and mouse exon 9 and intron/exon junctions

To analyse the mouse exon 9 splice sites, a mouse genomic region containing CFTR exon 9 and about 300 bases of its flanking introns 8 and 9 was amplified by PCR and entirely sequenced. The comparison of this mouse sequence with the corresponding human counterpart has revealed the high degree of homology existing between exons and the poor homology existing between mouse and human introns (Figure 3.8.). The analysis of the primary sequences of the splicing sites of the mouse CFTR exon 9 shows that its 3′ splice site is poorly defined (but, however, mouse exon 9 is well recognised and constitutively spliced). A cytosine at the position -3 of the 3′ splice site in the human sequence is substituted by a less frequent thymidine in the mouse.

a



b

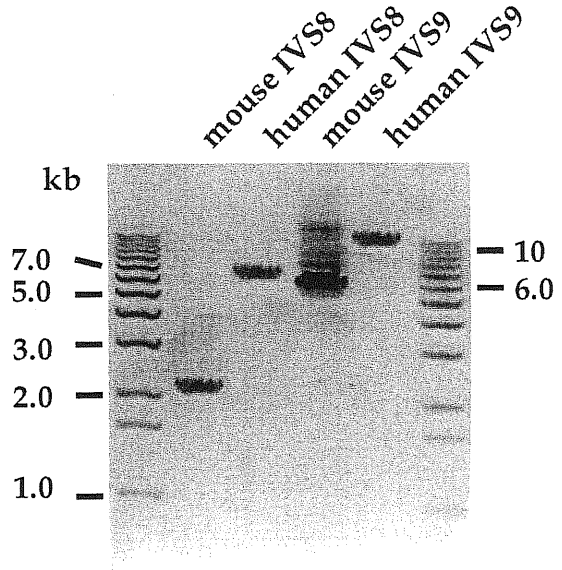


Figure 3.7. Human and mouse genomic regions spanning exon 8-10. a) Schematic representation of mouse and human exon 8-10.Length of the humanand mouse intron 8 and 9 were from PCR amplification.Length of human introns are confirmed by data found in GeneBank. Boxes correspond to exons, lines to introns. Sizes of exons and introns are shown. b) PCR amplification of the human and mouse genomic regions spanning exon 8 to 9 (mouse IVS8 and human IVS8) and exon 9 to 10 (mouse IVS9 and human IVS9).First and last lanes, 1 kb marker (Gibco).

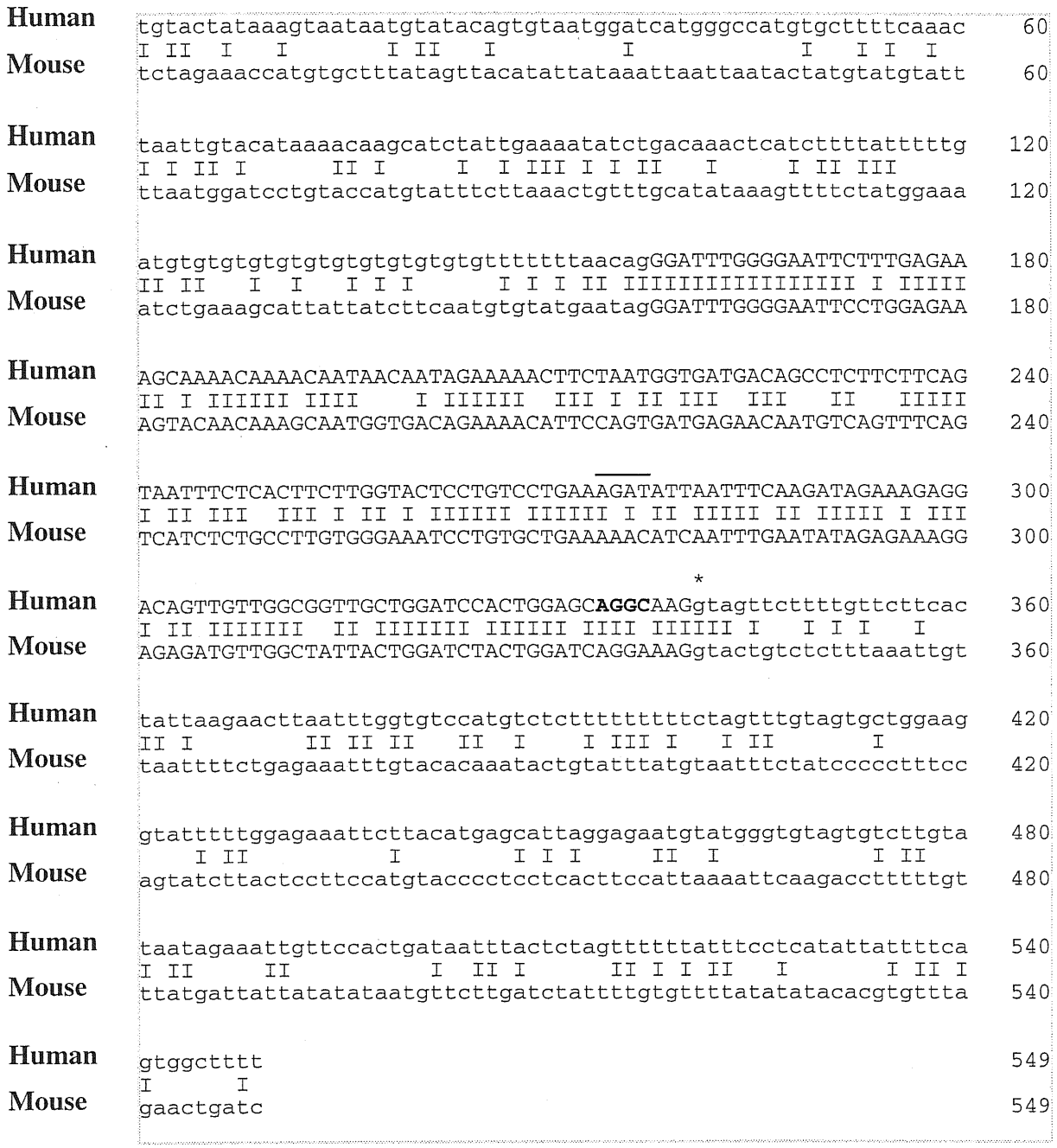


Figure 3.8. Alignment of mouse and human exon 9 (183 bp) with its flanking sequences of intron 8 (157 bp; Genebank submitted) and of intron 9 (209 bp; Genebank submitted). Human sequence, upper line; Mouse sequence, lower line. Exon 9, uppercase; Intron 8 and 9, lowercase. Asterisk (*) labels the guanine nucleotide in position +1 within human intron 9 replaced with an adenine nucleotide in constructs mutagenised at 5'-splice site. Bases in bold style indicate the human cryptic 5'-splice site activated after mutagenesis of wild-type 5' splice site (CCAG/GCAAG*a*tag). Superlined bases, 118 bp within the human exon (nt275, AG/AT), indicate the cryptic 3'-splice site activated in transfection of pm/H/m construct as detected by direct sequencing of PCR product.

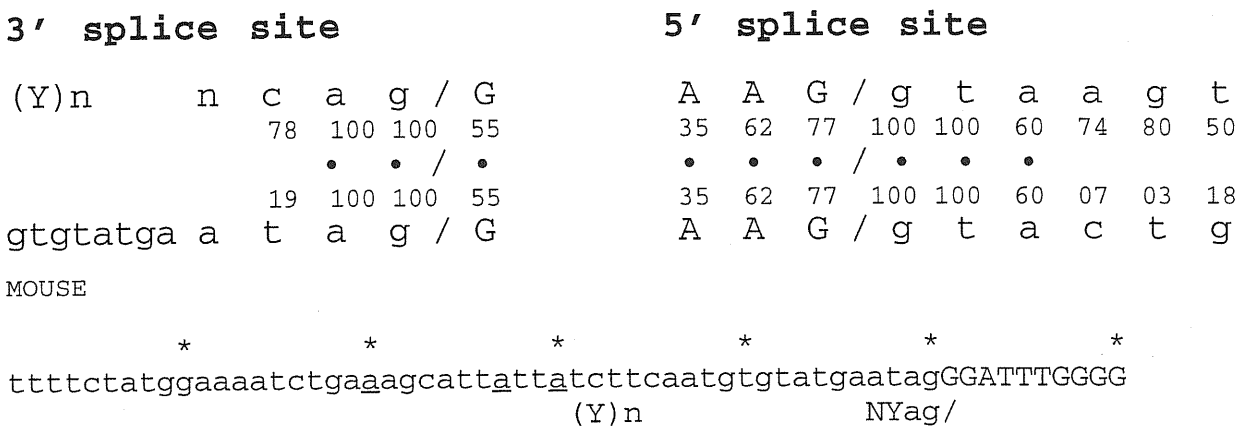


Figure 3.9. Comparison of splice sites from the mouse CFTR exon 9 with consensus sequences for 5′ and 3′ splice sites. a) 3′- and 5′-splice sites of mouse CFTR exon 9 (lower line) are compared with consensus sequences (upper line). Points indicate homology between the two sequences. Respective base frequencies are also indicated for both, consensus sequences and exon 9 splice site. N indicates any nucleotide, Y pyrimidine and R purine. Upper case, exon 9; lower case, introns. b) Organisation of the 3′ splice site of mouse CFTR exon 9 showing the three potential branch points and underlined a are the candidate branch sites), the polypyrimidine tract, (Y)n, and the 3′ splice junction, (consensus, NYAG/) .

Furthermore, none of two polymorphic loci, both TG and T repeats, are present in mouse. Moreover, the polypyrimidine tract in mouse is not well defined as the pyrimidines are interrupted with purines, leaving only five pyrimidines in a row just at the position -16 up to -20. By looking for a good branchpoint consensus sequence, using FindPattern program, and taking in consideration the distance from the 3′ splice site, we have selected three potential sequences (at position -21, -24 and -31) (Figure 3.9.). Then, at the 5′ splice site, bases at position +4 and +6 are different from the human sequence as well as from the consensus sequence.

This observations let us to conclude that mouse splice sites are weaker than the human ones and they do not explain the differences in the efficient recognition and splicing of exon 9 between human and mouse. In contrary, even if they look like weaker than human ones, yet the mouse exon is more efficiently spliced.

3.3.3. Effect of TG and T polymorphic is influenced by other cis-acting elements in human and/or mouse intron 9

Our results with the human minigene system show that the 600 bp of exon 9 and its flanking introns contain enough information to reproduce the splicing pattern of various alleles in a minigene context. Following this example, we have created the similar mouse minigene system. A 550 bp CFTR mouse gene fragment containing exon 9 and its flanking intronic sequences were amplified by PCR and cloned in the pSV-mEDAΔNdeI vector used for the human CFTR minigene (Figure 3.2b). The transient transfection of the mouse minigene confirmed the reliability of our expression system. As observed in vivo, the transcription of the mouse minigene resulted in the production of the mRNA exclusively containing mouse exon 9. This mouse system has opened us additional possibilities to test the involvement of TG and T repetitions in splicing regulation.

Considering that these two polymorphic loci in the close proximity of the 3′ splice site in human might be the only elements that determinate the alternative splicing, we hypothesised that the TG/T stretch placed in the heterologous environment of the constitutivley spliced mouse CFTR exon 9, should decrease recognition of this mouse exon at least in the same way observed in human.

In order to study the effect of the human 3′ splice site in the mouse context, human/mouse hybrids were designed, where the intron 8 of a mouse minigene was replaced with the human one containing 11TG/7T or 11TG/5T alleles (p(TG)11/T7/M/m and p(TG)11/T5/M/m) (Figure 3.10.a). The possibility that the results obtained by the transient transfection were species- specific was excluded by parallel expression of human as well as human/mouse hybrid constructs in two cell lines of the same tissue (liver) but different specie origin; NMuli murine, and Hep3B human liver cell line.

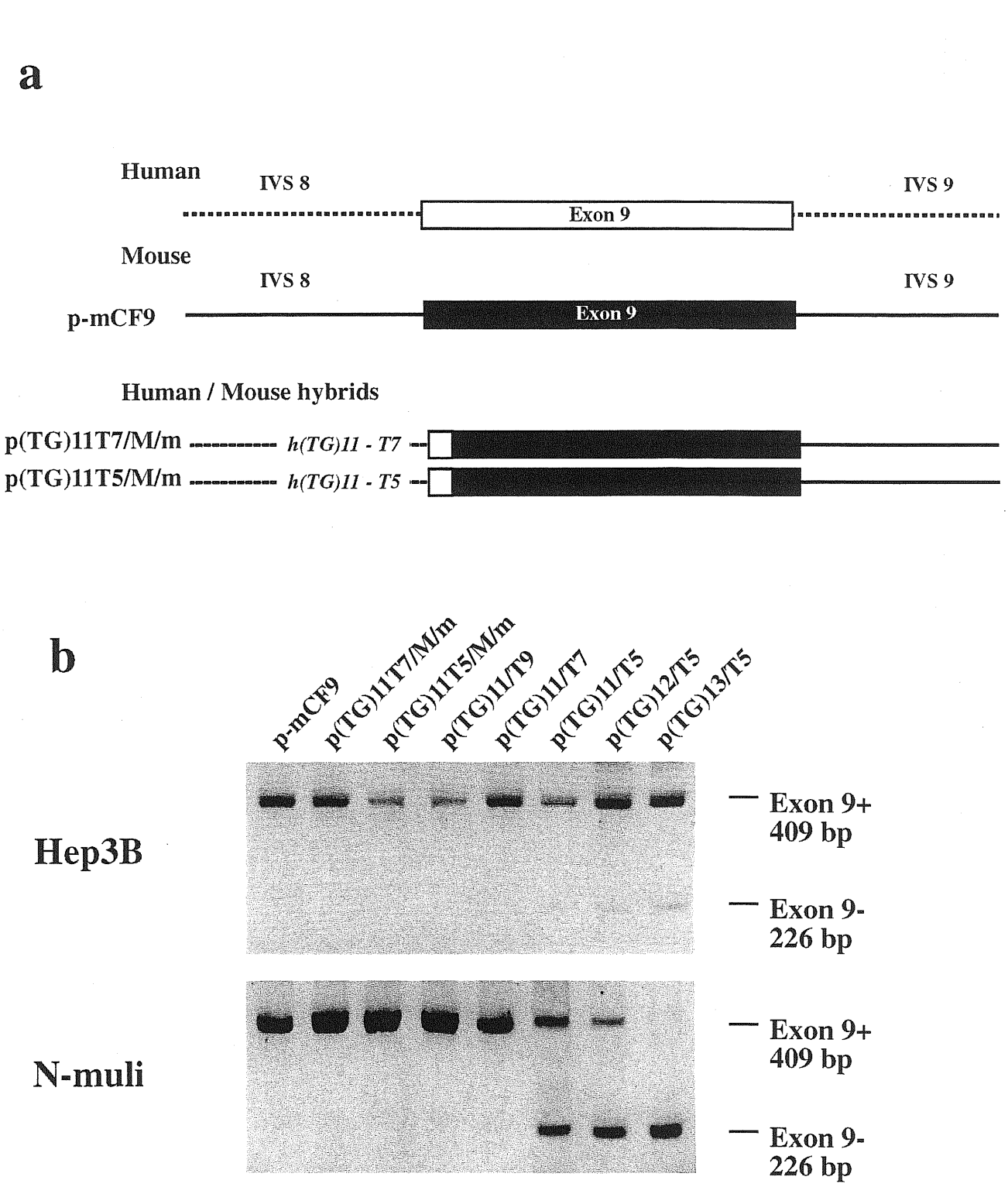


Figure 3.10. Analysis of pre-mRNA splicing from human/mouse hybrid constructs. a) Schematic representation of human/mouse hybrids. White box, human exon 9; black box, mouse exon 9. Dot lines, human intron 8 or 9. Continuous lines, mouse introns. h(TG)11-T7 and h(TG)11-T5, genotype of the polymorphic loci at 3' end of human intron 8. b) Analysis of pre-mRNA splicing of human/mouse hybrid constructs. PCR products were separated on 2.0 % agarose gel. Proportion of exon 9 exclusion was determined by densitometry using the total densitometric units of 9+ and 9- amplicons as 100%. Hep3B: p(TG)11/T5, 3%; p(TG)12/T5, 7%; p(TG)13/T5, 11%. N-Muli: p(TG)11/T5, 54%; p(TG)12/T5, 72%; p(TG)13/T5, 100%.

Firstly, in both cell lines, the transient transfection of human constructs resulted in expression of mRNA exclusively containing exon 9 in the case of p(TG)11/T7 and weakening of exon 9 recognition in p(TG)11/T5 (Figure 3.10.b.). The proportion of exon 9 skipping for the p(TG)11/T5 in NMuli was highly increased compared to Hep3B (Figure 3.10.b).

Secondarily, the transfections with the hybrid constructs p(TG)11/T7/M/m and p(TG)11/T5/M/m resulted in 100% inclusion of exon 9 both in human and mouse cell lines, regardless the number of Ts within the poly-thymidine tract (Figure 3.10.). The fact that human TG and T polymorphisms were able to shift the balance of exon skipping in human but not in the mouse sequence enviroment suggests that TG/T effect can be modified by some other cis-acting elements. Also, different behaviour of the TG/T sequence within the mouse and human context let us to conclude that these elements are not conserved between human and mouse.

In order to localise and identify these elements, possibly involved in exon 9 definition, the work was directed in two directions. Firstly, looking for the position of the elements (within human and/or mouse) that could modify the effect of the TG/T tract on splicing pattern a series of human/mouse hybrids were generated, where the murine sequences were progressively replaced with the human counterpart (Figure 3.11.a.) and effect of this changes in splicing pattern were analysed by transfection experiments. Second, by transfection experiments where the different minigene constructs, human and mouse, were cotransfected with the hnRNPA1 and ASF/SF2 (described in the section 3.5..) we tried to better characterise the role of these control elements.

The transfection of hybrid constructs p(TG)11/T7/H/m and p(TG)11/T5/H/m (Figure 3.11.a) in Hep3B as well as NMuli showed 100% inclusion of exon 9, so excluding the existance of different cis-acting elements within human and mouse exon 9. Splicing pattern was, also, not changed even when the mouse 5′ splice site was replaced with the human one (m/3′hT5). In this way the 5′ splice

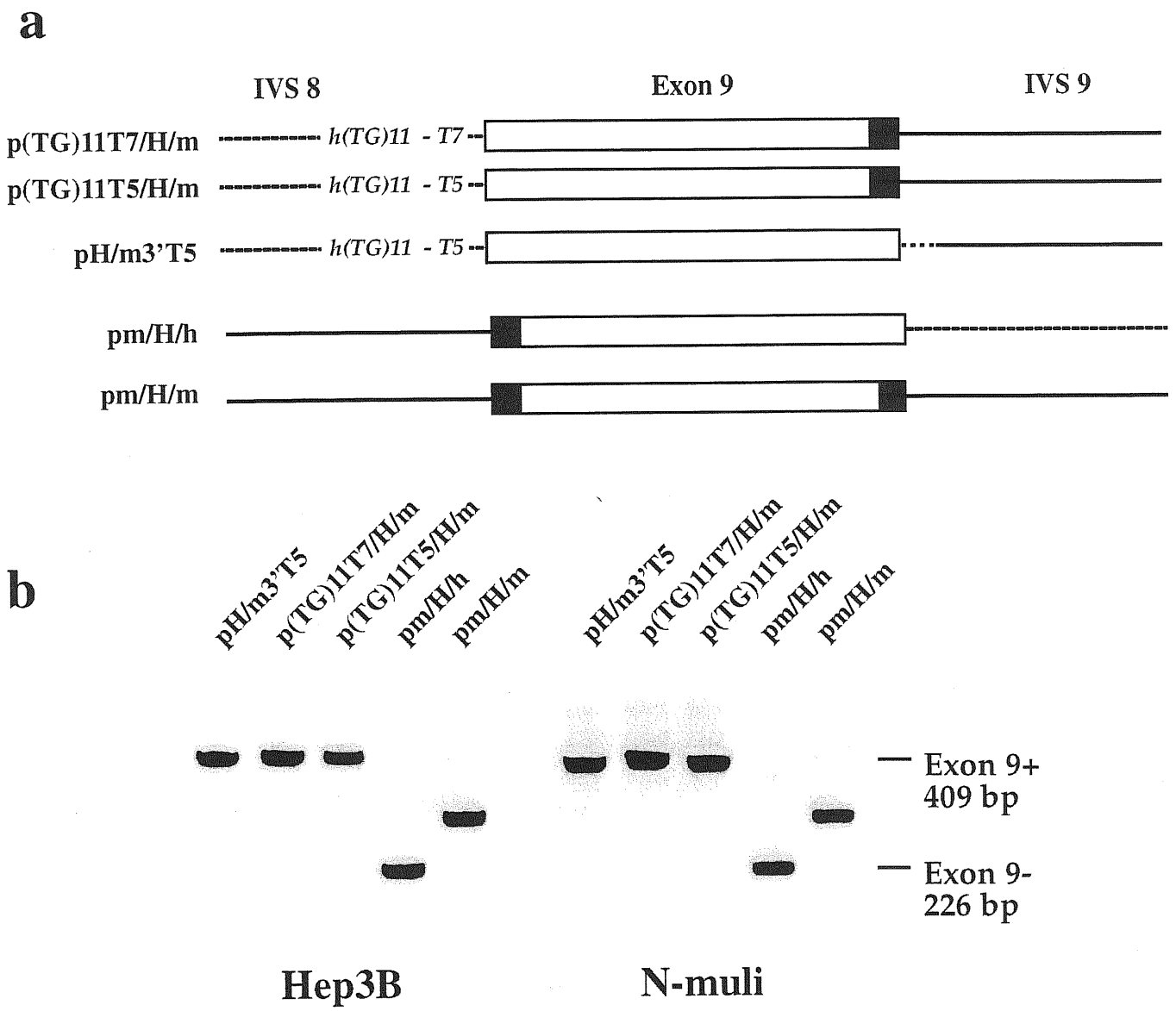


Figure 3.11. Analysis of pre-mRNA splicing from human/mouse hybrid constructs. a) Schematic representation of human/mouse hybrids. White box, human exon 9; black box, mouse exon 9. Dot lines, human intron 8 or 9. Continuous lines, mouse introns. h(TG)11/T7 and h(TG)11/T5, genotype of the polymorphic loci at 3'-end of human intron 8. b) Effect of transient transfection of human/mouse hybrid constructs into Hep3B and N-Muli cell lines. Transient transfections of the pH/m3'T5, the p(TG)11/T7/H/m and p(TG)11/T5/H/m constructs show 100% exon 9 inclusion. Transfection of pm/H/h construct shows 100% exon 9 exclusion. The transfection of pm/H/m construct shows activation of a cryptic splice site 118 bp within the human exon (nt275, Figure 3.8.). Sizes of PCR product including and excluding exon 9 are shown. PCR products were separated on 2.0 % agarose gel.

site junction of mouse and human were excluded to be responsable for the different splicing pattern of the mouse and human minigenes carrying the same 3′ splice site. Finally, we could conclude that sequence encompassing theregion from -10 up to - 200 of intron 9 (mouse or human) contains elements that contribute together with polyT and poly TG loci to the regulation of exon 9 splicing.

Moreover, transfections of two other hybrid minigenes, pm/H/h and of pm/H/m, showed 100% of exon 9 exclusion and activation of a cryptic 3′ splice site 118 base pairs downstream of the default one, respectively (Figure 3.11.b.). On one hand, these results confirm the observation of mouse sequence analysis that mouse intron 8 3′ splice site is functionally weaker than human counterpart. On the other hand, they suggest that the polymorphic loci at the 3′ end of the intron 8 might be influenced by some other elements that could be either presence of the enhancer in mouse, the silencer in the human region of intron 9, or both.

3.4. Trans acting elements involved in regulation of exon 9 splicing in human CFTR

The splicing factors regulate alternative splicing of many pre-mRNA. They recognise and bind specific sequences on the pre-mRNA, and thus they mediate the accessibility of other sequence elements to be recognised by the splicing machinery [90, 213, 214].

Considering the limited variations in sequence composition associated to changes in the polyU stretch length, we suppose that it could be due to the changes of its affinity for the U2AF^{65}. On the other hand, the observations related to the UG tract led us to hypothetise that this sequence could be the target of some factors, whose binding in proximity to the 3′ splice site might have an inhibitory effect on the splicing efficiency of exon 9.

3.4.1. A protein recognise and bind exclusively to poly-UG sequence

To address the possibility that the UG polymorphic tract might interact with some trans-acting factors so affecting the 3′ splice site exon 9 recognition, both, functional and UV crosslinking assays were performed.

The set of construct used in this study included a wild type sequence (p(TG)11/T7), one deleted in the polyT stretch (p(TG)11/ΔT), another deleted in the poly (TG) stretch, (pΔTG/T7) and finally one lacking the whole polymorphic region pΔ(TG)/ΔT (Figure 3.12.a).

Firstly, a functional role of polymorphic sequences was analysed by transfecting the minigene with different deletions in NCI-H322. The transfection of p(TG)11/ΔT construct resulted in expression of RNA excluding exon 9, confirming the positive role of the U stretch in the splicing regulation. On the other hand, the transfection of two other constructs (pΔ(TG)/ΔT and pΔ(TG)/T7) resulted in 100% exon 9 inclusion (Figure 3.12.) probably because, as a consequence of the deletions, a stretch of continuous thymidines was brought in proximity of the 3′ splice site, replacing the deleted polypyrimidine tract.

These mutants carrying different deletions were placed in pBS SK vector under the T7 promoter and were transcribed in presence of the α-32P-UTP. After incubation of transcribed mRNAs in HeLa nuclear extract, UV crosslinking experiments were performed and the results were evidenced through SDS-PAGE. Our results showed no difference in the UV crosslinked protein pattern obtained with the wild type ((UG)11/U7) and the (UG)11/ΔU mRNA. Instead, the Δ(UG)/ΔU and Δ(UG)/U7 mRNA templates did not UV crosslinked a cellular protein of apparent molecular weight of 52kDa (note that the correct molecular size of that protein have might be influenced by interaction of the protein to the labelled RNA) that is present in ((UG)11/U7) and the (UG)11/ΔU (Figure 3.12.c).

Additionally, UV crosslinking/competition assays were performed to confirm these results. The Δ(UG)/ΔU, Δ(UG)/U7 and (UG)11/ΔU mRNAs were used as

a

```
p(TG)11/T7 ...aactcatcttttatttttga----tgtgtgtgtgtgtgtgtgtgtttttttaacagGGATTTGGGG...
p(TG)11/TΔ ...aactcatcttttatttttgatgtgtgtgtgtgtgtgtgtgtgtg-------aacagGGATTTGGGG...
p(TG)Δ/T7  ...aactcatcttttatttttga-------------------------ttttttttaacagGGATTTGGGG...
p(TG)Δ/TΔ  ...aactcatcttttatttttga-------------------------------aacagGGATTTGGGG...
```

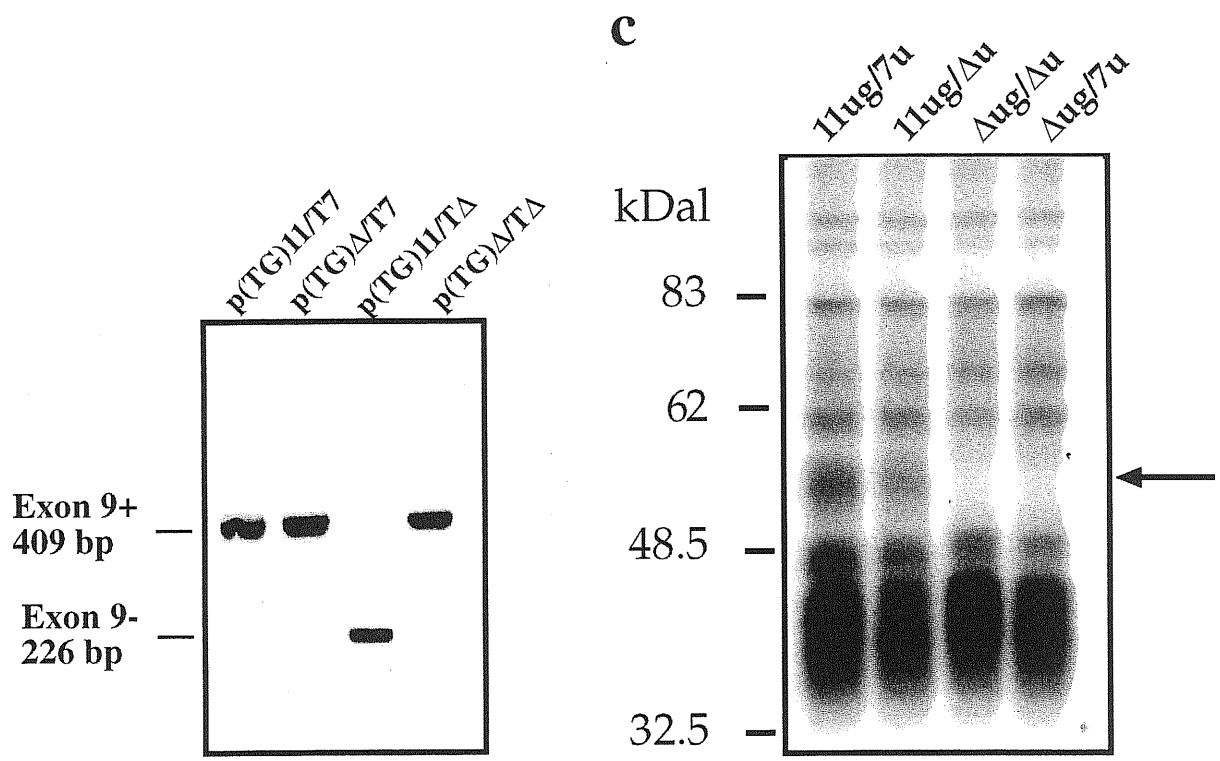


Figure 3.12. Effect of the TG and/or T tract deletions on RNA splicing and RNA/protein interaction. a) Sequence of intron 8 3′ splice site of constructs carrying different TG and/or T deletions. b) Analysis of pre-mRNA splicing of the constructs in the NCI-H322 cells. The size of the products including (409bp) and excluding (226bp) exon 9 are indicated. c) UV crosslinking reactions of RNA transcripts synthesised from the constructs by the T7 RNA polymerase and cutted by BamHI (restricted site at the end of human exon 9). Transcripts, named11ug/7u, Δug/7u, Δυ/11ug and Δug/Δu, respectively were incubated with HeLa nuclear extracts under standard conditions, UV irradiated, digested with RNAse, resoved on SDS-PAGE and autoradiographed. Positions of molecular weight markers are indicated.

cold competitors of the p(UG)11/U7. The (UG)11/ΔU successfully competed, whereas the others Δ(UG)/ΔU, and Δ(UG)/U7, failed in competing with the wild type mRNA for binding 52 kDA cellular protein (Figure 3.13.).

Altogether, these results demonstrate that the UG stretch negatively affects exon 9 splicing efficiency and that a specific protein of 52 kDa recognises and binds the UG dinucleotide repeats of human intron 8.

3.4.2. Binding specificity of the UG binding protein

In order to confirm that only the UG tract was involved in the binding of this protein, we have performed a competition analysis using very short cold RNAs. This RNAs were obtained by transcripion of the pBS SK vector containing sequences of interest, (TG)12 or (TCTT)3. Figure 3.14. shows that a cold RNA consisting only of 12 UG repeats was able to compete for the binding of this protein by the UG11/U7 mRNA. Considering that the apparent molecular weight of this protein was similar to that of PTB (poly-pyrimidine binding protein) and that the UG has been demonstrated to act as a poly-pyrimidine tract in certain contexts, as the first effort to identify this protein, we used a short RNA containing three repeats of a known consensus binding sequence of PTB, $(UCUU)_3$ as evidenced by SELEX analysis [184]. Our results show that this sequence is not able to compete for the 52 kDa protein bound by human wild type transcript so it was possible to exclude PTB as the UG binding protein (Figure 3.14.).

In addition, we have found that the binding specificity of this protein is not confined to RNA but also to DNA since a TG stretch can be used as an effective competitor (Figure 3.15.).

In fact, to further investigate the binding specificity of the studied protein we performed UVcrosslinking/competition assay of human transcript using as competitors the single stranded and double stranded DNA. Figure 3.15. shows

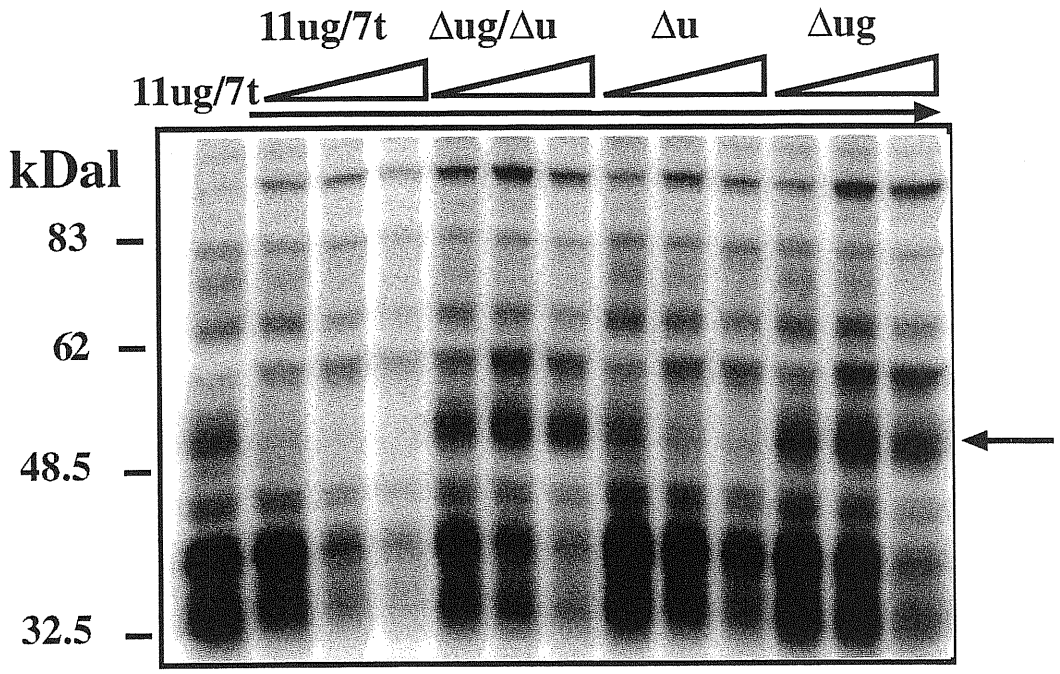


Figure 3.13. UV crosslinking assay of transcribed linearised p(TG)11/T7 construct using HeLa nuclear extract. Increasing amount of transcribed p(TG)11/T7, pΔ(TG)/ΔT, p(TG)11/ΔT, pΔ(TG)/T7 constructs were used as competitors Competitor transcripts were named 11ug/7u, Δug/7u, Δυ/11ug and Δug/Δu, respectively. The molar ratio of competitorRNA to labelled RNA used for three data points were 2, 5, and 10. First line, no competitors added. Position of the protein competed by the Δu RNA was indicated by an arrow.

a

RNA single-stranded (5'-3')	
$(ug)_{12}$	ugugugugugugugugugugugug
$(ucuu)_3$	uccuccuccuucuucuucuucagg

b

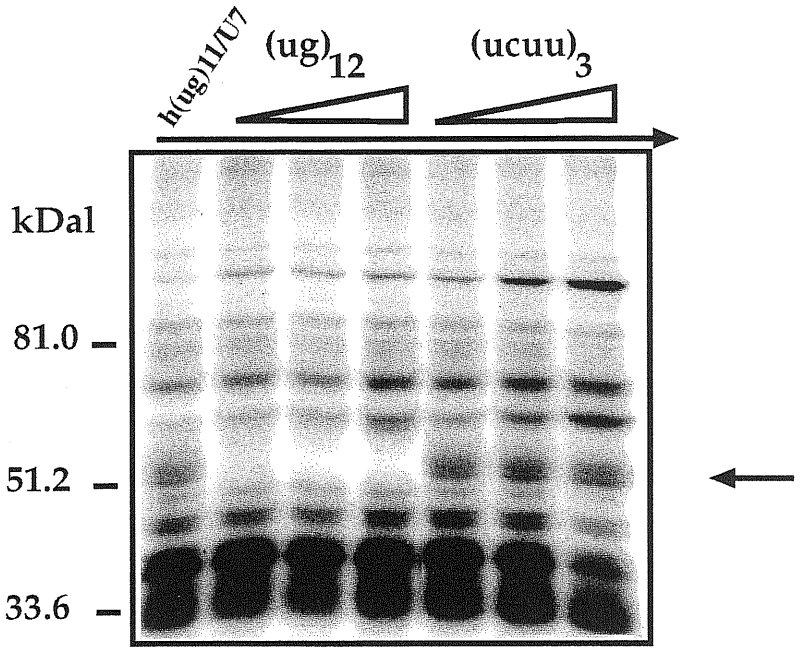


Figure 3.14. Binding specificity of the protein that interact with UG tract of intron 8 in human CFTR for different RNAs. a) Table of RNA oligonucleotide sequences used as a competitors in UV crosslinking/competition assay. b) UV crosslinking competition assay of RNA transcribed from TG11/T7 variant. The molar ratio of competitor RNA to labelled RNA used for the three data points were approximatelly 3, 8, 17.Competitors were incubated with HeLa nuclear extract 5 min. before addituion of labelled RNA transcript. Transcripts were incubated with HeLa nuclear extracts under standard conditions, UV irradiated, digested with RNAse, resolved on SDS-PAGE and autoradiographed. Arrow indicates the position of the protein competed by the $(UG)_{12}$. First line, no competitors added. Positions of molecular weight markers are indicated.

a

DNA single-stranded (5'-3')	
$(tg)_{12}$	tgtgtgtgtgtgtgtgtgtgtg
$(tctt)_3$	tcctcctccttcttcttcttcagg
$(agaa)_3$	aggaggaggaagaagaagaagtcc
(mex9as)	gaaaattaacaatttaaa
DNA double-stranded (5'-3')	
(tgds)	tgtgtgtgtgtgtgtgtgtgtgtg acacacacacacacacacacacac
(tcttds)	tcctcctccttcttcttcttcagg aggaggaggaagaagaagaagtcc

b

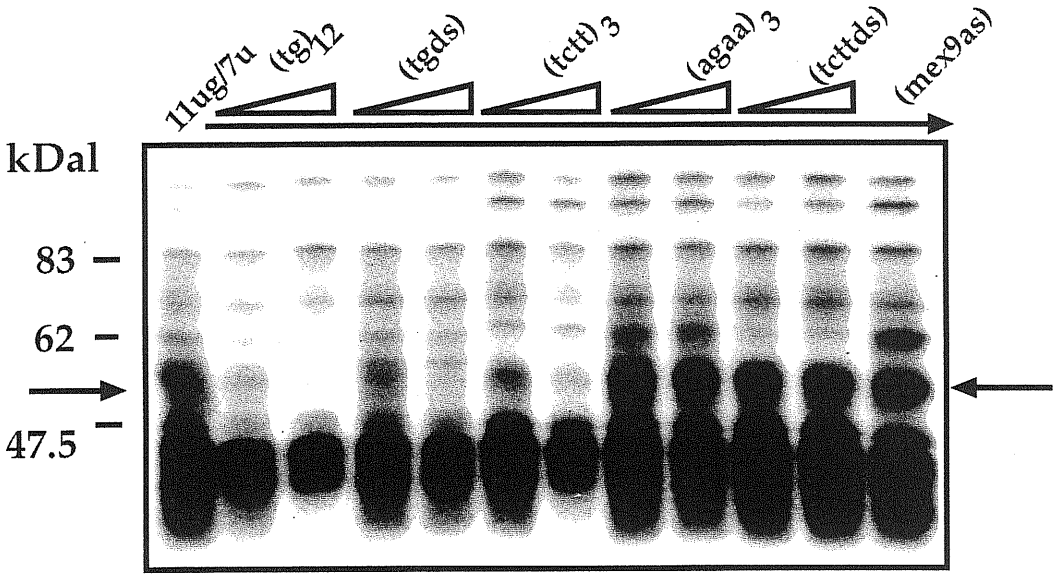


Figure 3.15. Binding specificity for the protein that interact with UG tract of intron 8 in human CFTR. a) Table of single stranded and double stranded oligonucleotide sequences used as competitors in UV crosslinking/competition assay. b) UV crosslinking competition assay of RNA transcribed from p(TG)11/T7 variant. Competitors RNA's were incubated with HeLa nuclear extract 5 min. before addition of labelled RNA transcript. Transcripts were incubated with HeLa nuclear extracts under standard conditions, UV irradiated, digested with RNAse, resoved on SDS-PAGE and autoradiographed. The molar ratio of competitor DNA to labelled RNA used for the two data points performed with each oligo was 5 and 10. First line, no competitors added; last line, mouse exon 9 RNA added as a competitor. Positions of molecular weight markers are indicated. Arrow indicates a position of the protein competed by (tg)12, (tctt)3 and tgds.

that the investigated protein has a broad spectrum of binding affinity for different single- and double- stranded DNA sequences (ssTG12, ss(tctt) and tgds).

3.4.3. Identification of the UG binding cellular protein

In order to identify the UG binding cellular protein we have set up an affinity procedure that involves the crosslinking of synthetic RNAs to adipic acid dehydrazide agarose beads following sodium-m-periodate treatment. As crosslinking target sequence we used (UG)12 RNA oligonucleotides and we incubated them with the HeLa nuclear extract. After the incubation and repeated washings, the proteins bound to the RNA on the beads were eluted, separated on the SDS-PAGE gel and stained with Coomassie blue. As a negative control we used a UCUU derivatized beads which should have affinity for the PTB protein. As shown in Figure 3.16., contrarily to the 57 kD doublet pulled down by the UCUU beads and consistent with the MW of PTB, the UG derivatized beads bound a 43 kDa protein that is consistent with previous studied protein of 52 kDa molecular size, considering that the increase in molecular weight could be due to the crosslinked RNA. The band containing this protein was excised from the gel and sent for an internal sequence analysis to the Eurosequence service. Internal sequence analysis results yielded a 17-mer peptide with the following aminoacid sequence: EPNQAFGSGNNSYSGSN. This sequence is a 100% identical to a residues 362 to 378 of a TDP-43 protein. This protein is a cellular factor originally found to bind Human Imunodeficiency Virus I TAR DNA element where it seems to be involved in repression of its transcription [215].

3.5. SR protein and its involvement in modulation of splicing

We have already shown that effect of TG and T polymorphisms on exon 9 recognition can be modified by some other elements placed in human or/and

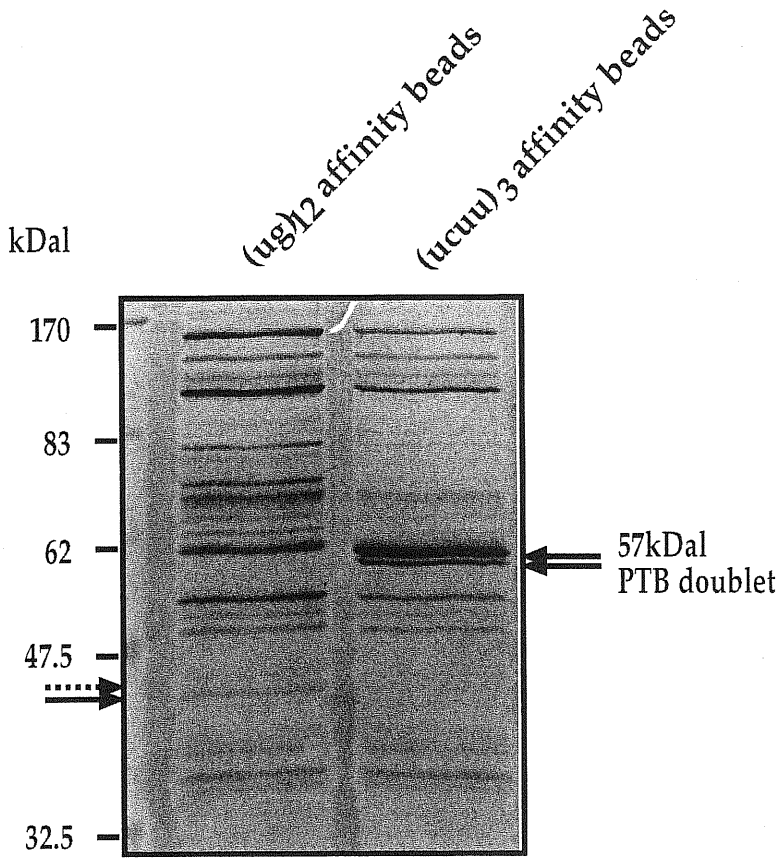


Figure 3.16. Results of pull down assay. Adipic acid dehydrazide beads were (ug)12 and (ucuu)3 derivatized. Bound proteins were stained with Comassie Blue. Left arrows (lane (ug)12) indicate the main and faint protein bands that are absent in the lane (ucuu)3. Right arrows (lane (ucuu)3) indicate 57kDa doublet corresponding to PTB.

mouse. Considering that the effect of cis elements is mostly due to their posibility to interact with some protein factors we have investigated if splicing of human and mouse CFTR exon 9 can be modulated by the ASF/SF2 and/or hRNPA1 overexpression, considering their possible interactions either with the 3′ splice site and/or the 5′ splice site.

Hep3B cell line was transfected with the minigene constructs, carrying 11TG/T5, 11TG/T7 and 11TG/T9 at their polymorphic loci together with increasing amount of plasmid carrying the cDNA encoding for the ASF/SF2 protein. The effects of ASF/SF2 on alternative splicing of CFTR exon 9 were analysed by RT-PCR amplification. Our results have shown that overexpressed ASF/SF2 increased exon 9 skipping in all variants (T5, T7 and T9) (Figure 3.17.). Furthermore, to test if effect of ASF/SF2 on splicing efficiency of CFTR exon 9 is common for human and mouse, mouse minigene construct (pmCF9) was cotransfected together with the ASF/SF2. The overexpression of ASF/SF2 splicing factor did not result in any significant effect on the splicing of mouse CFTR exon 9 (Figure 3.17.b).

To locate sequences that possibly interact with the ASF/SF2 in human but not mouse context, a contransfection experiments were performed, using series of mouse/human hybrid plasmids and contrasfecting them with ASF/SF2.

The splicing of the hybrids p(TG)11T5/M/m and was not modified by the ASF/SF2 showing 100% inclusion. This finding suggests that the 3′ end of human intron 8 is not the target site of splicing inhibition of mediated by ASF/SF2, or that mouse elements able to counteract the effect negative effect of ASF/SF2 in splicing efficiency. Also, splicing of p(TG)11T5/H/m was not modified by ASF/SF excluding exon 9 as a target for its binding.

Altogether, these results show that ASF/SF2 splicing factor, usually involved in the positive regulation of exon splicing, have a negative effect on the splicing of the human CFTR exon 9. The possible localisation of cis-acting elements which might be recognised by ASF/SF2 splicing factor is within the range -10 - -269

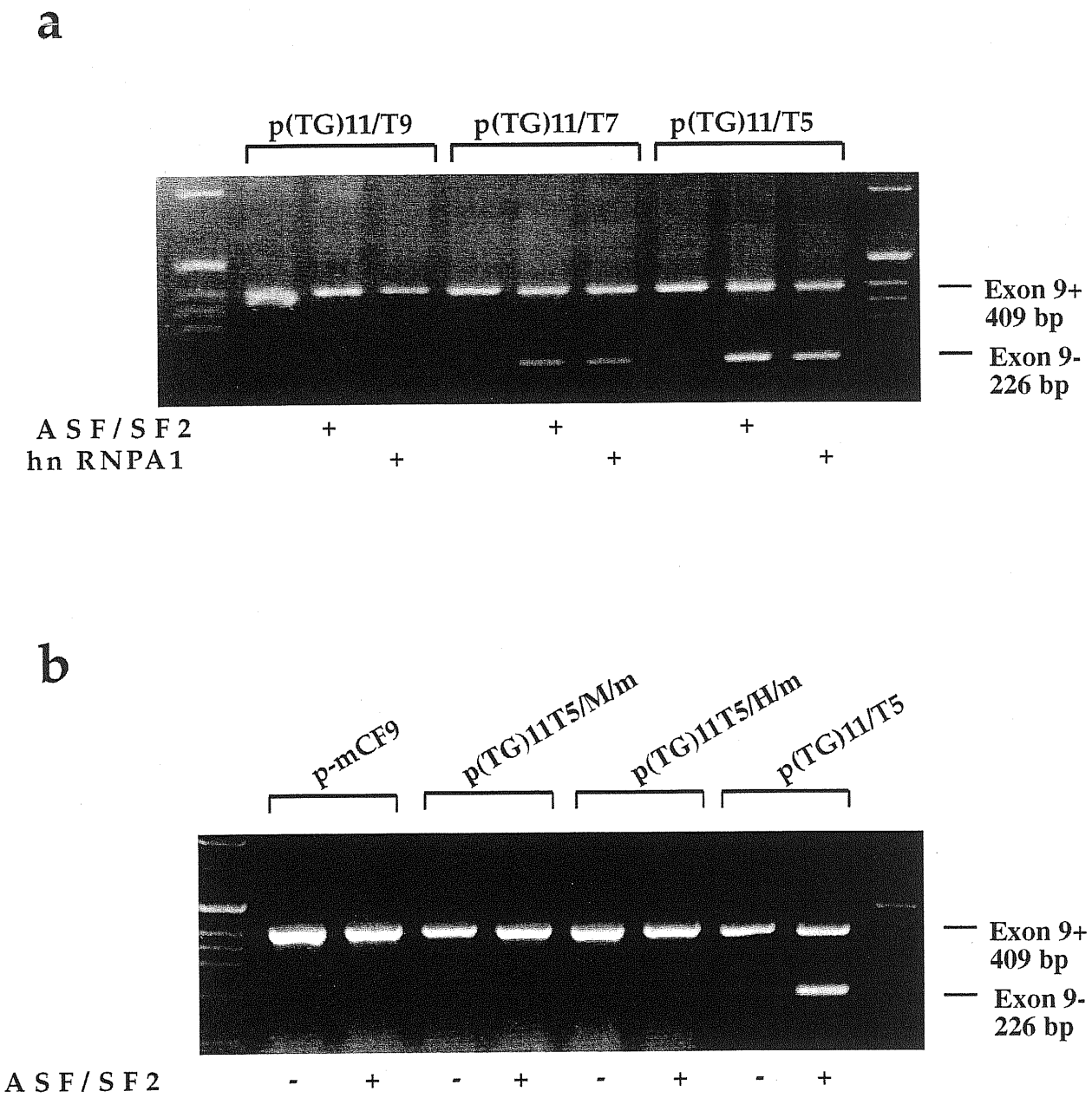


Figure 3.17. Effect of overexpression of ASF/SF2 and hnRNPA1 on splicing efficiency of the human CFTR exon 9. a) Hep3B cells were transfected with constructs carrying a different number of tymidines at the 3′ splice site (T5, T7, T9) and an 500ng of plasmid coding for ASF2/SF protein or hnRNPA1. b) Expression of the human and a mouse CFTR exon 9 constructs in the presence of ASF/SF2 . Mouse CFTR exon 9 is not negatively regulated by SR proteins. The Hep3B were cotransfected with the 500 ng of the indicated splicing factor. RNA splicing variants were detected by RT-PCR and analysed on the 2.0 %agarose gel. The size of PCR products including (409bp) and excluding (226 bp) exon 9 are indicated.

nucleotide of human intron 9. This evidence can further explain previous results of transfection with pm/H/h and pm/H/m hybrids since the negative effect of human intron 9 for the exon 9 recognition was eliminated by replacing it with the murine intron 9 in pm/H/m and support an idea that a negative regulatory element might exist in human intron 9.

To test the role of hnRNPA1 on splicing of human CFTR exon 9 Hep3B cells were transfected with minigene constructs carrying T9, T7 or T5 along with plasmid containing cDNA encoding for hnRNPA1. HnRNPA1 overexpression induced exon skipping in a way similar to that observed with ASF/SF2 (Figure 3.17.).

3.5.1. Localisation of cis-acting elements within the human intron 9

In order to establish a functional correlation between the human intron 9 and its species specific splicing inhibition by trans-acting factors, we have analysed the human and mouse CFTR intron 9 transcripts for their ability to interact with nuclear proteins from HeLa nuclear extract.

For this reason UV crosslinking assays were performed using mouse or human intron 9 as a templates. Our results have revealed that, beside a set of proteins shared by both introns, there were three additional bands in a range ~35 to 44 kDa and a band of ~75 kDa binding specifically human intron 9. This observation was confirmed by competition/UV crosslinking assay where we tested both human and mouse intron 9 for their ability to compete for proteins that bind human intron 9. In fact, we have found that mouse intron 9 was not able to compete for the proteins that crosslink specifically to the human intron 9 RNA (Figure 3.18.b, left panel).

In order to localise more precisely sequences that recognise specifically human intron 9, a new series of competitors was synthesised. These competitors contained short regions of the human intron 9 sequences, with 77 bp or 117 bp or

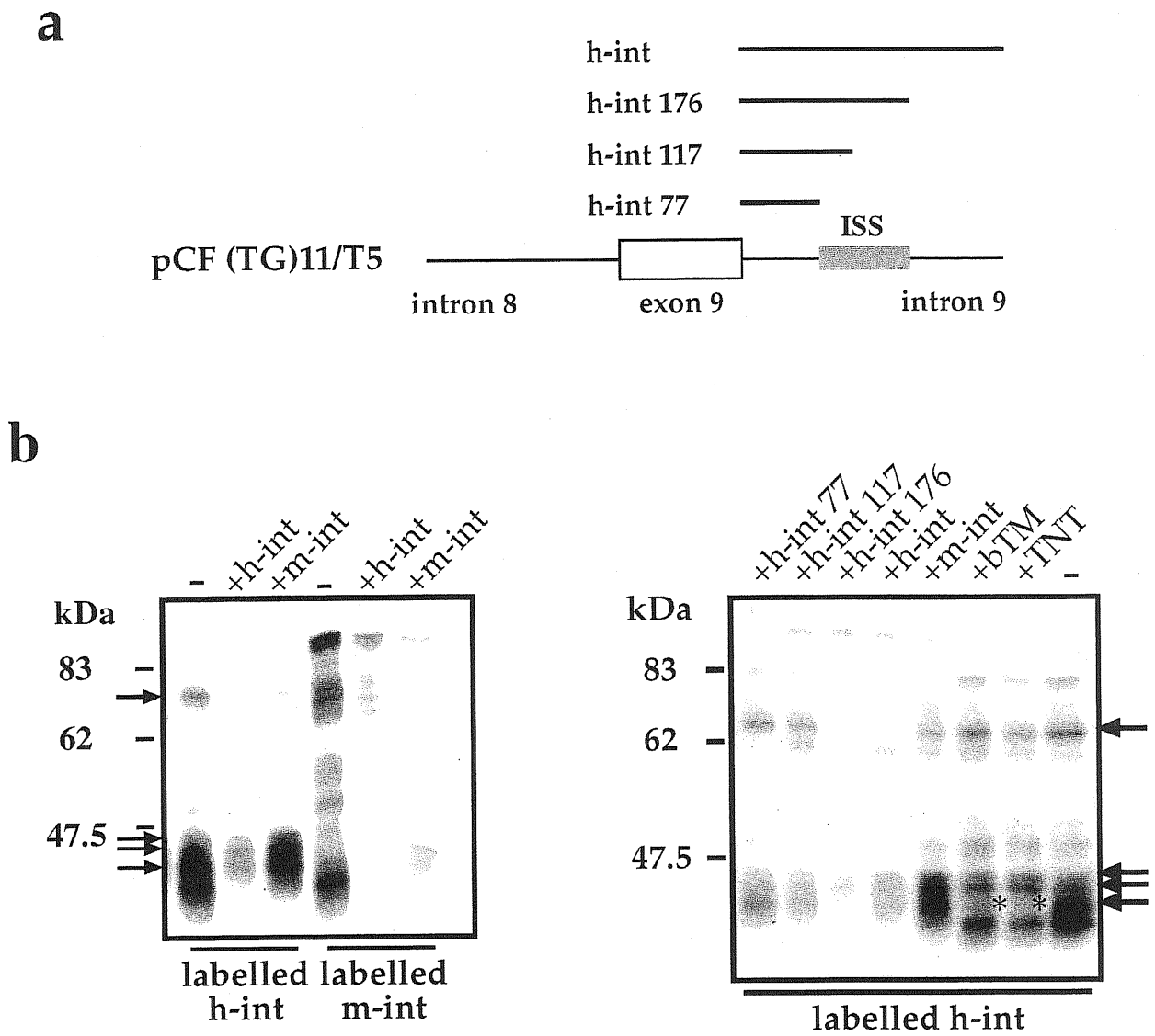


Figure 3.18. The CFTR exon 9 ISS element in human intron 9 binds SR proteins.
a) Shematic representation of the human sequences used in UV crosslinking/competition assay: m-int sequence is a mouse conterpart of indicated h-int.
b) UV crosslinking of labelled human (h-int) and mouse (m-int) introns challenged with the indicated competitor RNAs. Labelled RNAs were were incubated with HeLa nuclear extracts under the standard splicing conditions, UV irradiated, digested with RNAse, resolved on polyacrilamide gels and autoradiographed. TNT sequence; 5′-aagaggaagaauggcuugaggaagacgacg-3′, β-TM, 5′-agggaaagacagggagggagagagaaagagaaagg-3′. Position of molecular weight markers are indicated.

176 bp of the 5′ end of human intron 9 (Figure 3.18a. and b. right panel). A competition effect, for human specific proteins, was observed only with competitors containing 100 bp region located between 77 and 176 bases downstream of the 5′ splice site, indicating that this region is the target for the proteins.

Furthermore, in order to better characterise these proteins, we have generated the sequences that have well defined binding specificity to SR proteins and used them as a competitors. These sequences derive from β-tropomyosin alternative exon 6A and the cardiac troponin T alternative exon, where it is demonstrated their binding affinity for ASF/SF2, SC35 and SRp75 [195, 216]. The results of UV crosslinking/competition assays have shown that at least one of the UV crosslinked bands in the 35-44 kDa range completely competed by β-TM and TNT competitors (Figure 3.18.b, βTM and TNT). The molecular weights of the competed protein are consistent with those of ASF/SF2 and/or SC35.

Therefore, these results suggest that some SR proteins (possibly ASF/SF2 and/or SC35) might be the trans-acting factors interacting with some sequence within the region between 117 and 176 bp of human intron 9 and through this interaction they could influence exon 9 definition.

3.6. Cell type specific CFTR splicing regulatory mechanism

The SR proteins influence the selection of alternative splice sites in concentration dependant manner [89, 90, 155]. The fact that level of different SR proteins vary between different tissues [194]. indicate its possible involvement in cell-type specific regulation of splicing. Our findings that SR ASF/SF2 is able to modulate the alternative splicing of human CFTR exon 9 suggest that it could be one of the factors involved in tissue- dependant regulation of splicing of this exon.

3.6.1. The CFTR splicing regulatory mechanism is different in different cell types

We have tested whether splicing of the CFTR exon 9 is modulated in a cell type dependant manner. Hanamura et al. have analysed the level of expression of ASF/SF2 in different rat tissues and have shown that its concentration vary between these tissue in the relation testis>lung>kidney>liver. Taking into consideration these observation we have used the human bronchial carcinoma NCI-H322, human embrional kidney carcinoma, 293, and human liver carcinoma Hep3B as cell lines representing tissues analysed by Hanamura. Each of this cell line express reasonable level of CFTR mRNA since it was possible to amplify by PCR the cDNA obtained by reverse transcription of nontrasfected cells (data not shown).

CFTR minigene constructs carrying TG11/T9, TG11/T7, TG11/T5 genotypes were used for transient transfection in each of these cell line.

Alternative splicing was observed in each cell line only when transfected with p(TG)11/T5 allele and that the efficiency of splicing varied slightly among the cell lines. The higher proportion of exon 9 exclusion was observed in NCI-H322, and then in Hep3B, while in 293 was only fairly visible (Figure 3.19.).

Furthermore, we transfected the same set of plasmids in cell lines of different species origin as a mouse hepatocarcinoma (NMuli), and monkey transformed kidney (COS7). Also in these cell lines the exon 9 definition was affected by the polymorphic T tract locus in the same manner as in the human cell lines. However, interestingly, a higher proportion of 9- form was detected in the NMuli cell line.

These findings show the existence of a tissue variability for the splicing efficiency of human CFTR exon 9, probably due to the different concentration of splicing factors present in each cell line.

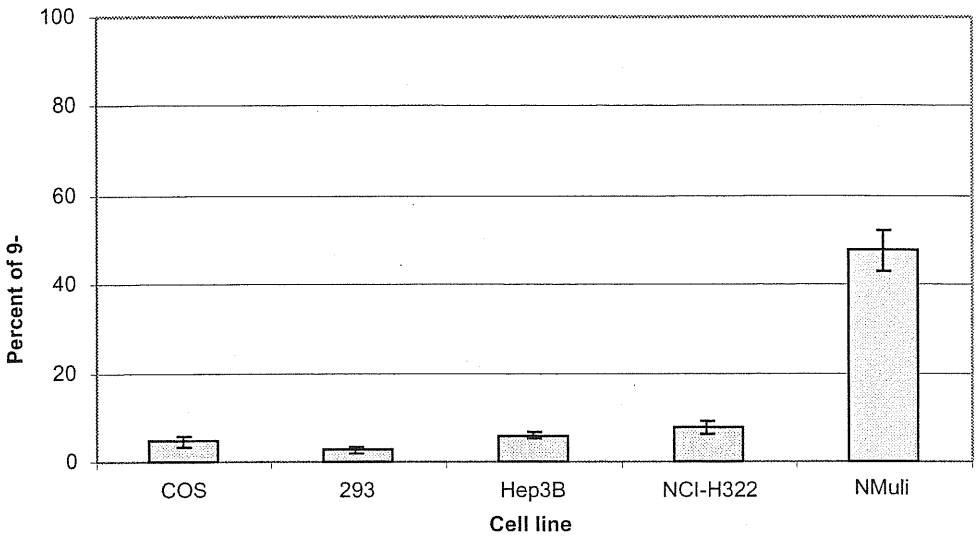


Figure 3.19. Levels of 9- form in different cell types. Different cell types were transfected with the minigene carrying pCF(TG)11/T5. Chart shows the proportion of exon 9 exclusion determined by densitometry using the total densitometric units of 9+ and 9- amplicon as 100%.

CHAPTER 4

DISCUSSION

The splicing of nuclear pre-mRNA is a highly regulated process in which introns are accurately and efficiently removed to produce mature mRNA. The selection of the splicing sites in the alternative and constitutive splicing is determined by the protein-RNA and protein-protein interactions. A considerable progress in understanding of splicing regulation has been made in recent years identifying the factors involved in splicing of different genes.

Our study have focused on the understanding the molecular basis of the alternative splicing of exon 9 in the human CFTR gene.

4.1. The 3′ splice site connection

Human CFTR exon 9 and its flanking introns share most of the canonical features of splicing sites. In fact, 3′ and 5′ splice junctions are well defined and resemble the consensus sequences [201]. Among the possible candidates, two putative branchpoints were selected, placed at about -60 and -40 bases upstream of the 3′ splice site (the exact positions depend on the TG/T length). Taking into consideration that most of the brancpoints identified are placed in the range of 18-40 bp upstream of the 3′ splice site we favour the hypothesis that the sequence at the position of -40 is a functional branchpoint. However, a functional mapping has to be performed to confirm our hypothesis.

In the human CFTR gene, the polypyrimidine tract at the 3′-splice site of exon 9 is composed of two polymorphic sequences. Firstly, it is composed exclusively by thymidines, polymorphic for its length and immediately downstream it is placed another polymorphic locus made of a variable number of (TG) repeats. It has been previously shown that thymidines (T) work efficiently as a polypyrimidine

tract when they are at least 6 in a row [142, 143]. Moreover, alterned stretch of T and G are also able to work as a functional polypyrimidine tract in constitutive splicing of at least one natural gene. In fact, the stretch of 16 (TG) repeats in intron 2 of the human Apo AII gene form a fully functional polypyrimidine tract [217]. Coolidge et al. have demonstrated that a TG tract is able to work as a polypyrimidine tract also in other contexts [143]. However, a polypyrimidine tract made of such a combination of TG and T polymorphic repeats is a feature unique to the human CFTR exon 9.

In this study, we have demonstrated that this peculiar architecture of intron 8 3'-splice site modulates the alternative splicing of human CFTR exon 9. In our minigene system, we have found that a stretch of five thymidines but not seven or nine thymidines weakens the exon 9 recognition. This is consistent with the relevance of polypyrimidine tract length observed in other genes [218]. The transfection experiments indicate that decreasing the number of thymidines within the polypyrimidine tract up to 5 is critical and provides the basis for a lower exon definition. Therefore, the function of the polymorphic polypyrimidine tract might be to rescue a poorly defined exon, at least in its 3'-splice site where five thymidines seem to be the threshold for a correct exon 9 recognition.

In addition, we have demonstrated that the effect of the (TG) tract is different from those reported in other contexts where it is able to form a fully functional polypyrimidine tract [143, 217]. In fact, transient transfection experiments with the construct lacking the T tract resulted exclusively in expression of mRNA lacking exon 9.

It has been proposed a possible interplay between TG/T tract to modulate the extent of exon 9 alternative splicing in a way that cooperatively determines the distance between the branch point and 3' splice junction [63]. We have shown that two genotypes,TG11/T7 and TG12/T5, with the same total base number ,-29-, do not show the same splicing pattern, since the TG12/T5 but not the TG11/T7

genotype weakened exon 9 recognition. However, we have found that the hypothesis proposed by Cuppens is valid for the TG number when juxtaposed to the T5. In fact, the TG stretch have a detrimental effect for exon 9 inclusion when associated with the T5 allele and the most simple explanation might be that the length of TG tract starts to be critical, affecting the distance between the branch point and the 3′ splice junction, only when the polypyrimidine tract is weak.
Previous studies on the effect of 5′ splice site mutations on splicing have shown either exon skipping or recognition of a closest 5′ cryptic splice site [111, 114]. In the CFTR exon 9 context, the mutation (AG/gt->AG/at) activated AG/gc splice site, that provides an exception to the GT rule. However, similar exceptions have been previously reported [219, 220] for the rabbit b-globin gene where the same cryptic splice site was activated following mutation of the default 5′ splice site. On one hand, we can speculate about the fact that the cryptic splice site was located close to 5′ splice site, as an evidence for the exon definition model in the CFTR exon 9 context. On the other hand, the mutation of the 5′ splice site did not changed the splicing pattern ruled by TG/T stretches. Alternative splicing was observed only with the constructs with T5 allele, where proportion of exon exclusion was higher than that of the corresponding wild-type 5′-splice site constructs and it was directly proportional to the length of the (TG) tract. Therefore, considering that the cryptic 5′ splice site is weaker than the wild type and that splicing pattern was not subverted completely, it is possible to conclude that the contribution of 5′-splice site to exon 9 recognition is less critical than that of the 3′-splice site.

4.2. 3′ splice site and protein interaction

The weakening of exon 9 recognition by imparing the length of tymidines suggests that some changes in the binding ability of $U2AF^{65}$ to the polypyrimidine tract occurs, [142, 218] lowering the efficiency of 3′-splice site

selection. On the other hand, it is possible that changes in the length of the polypyrimidine tract might change the binding specificity or increase the binding affinity of some other splicing factor/s that negatively regulate intron 8 - exon 9 junction recognition. Recent studies demonstrated that hnRNPA1 can exert an effect at the 3'-splice site, promoting exon skipping [178], so we can hypothesise that the decrease in length of the T tract might reduce the affinity for a splicing factor recognising the polypyrimidine tract ($U2AF^{65}$) and this, in turn, might allow a novel or stronger binding of hnRNP A1 protein to the 3'-splice site of exon 9, resulting in alternative splicing. This hypothesis is supported by the observation that the 3'-splice junction of exon 9 has a high homology with the hnRNP A1 consensus target sequence (TAGGGCTTAGGGT) (Figure 4.1.) [221, 222]. Following this reasoning, it is possible that variations in the accessibility of 3'-splice site to trans-acting factors, such as hnRNPA1, might affect its splicing efficiency.

```
hnRNPA1                                  UAGGGCUUAGGGU
          *          *          *          *          *          *
mouse     ttattatcttcaatgtgtatgaatagGGATTTGGGGAATTACTGGAGAAAGTACA
          I   I I I      I I I II IIIIIIIIIIIIIIII IIIIIIII I I
human     tgtgtgtgtgtgttttttaacagGGATTTGGGGAATTATTTGAGAAAGCAAA
```

Figure 4.1. The match between the hnRNPA1 consensus and the 3' splice site of exon 9 in human and mouse

However, the weakening of exon 9 recognition, when the length of poly T tract is critical does not explain further modulation of the splicing efficiency by the TG tract. It is questionable if TG is just a spacer between branch point and 3' splice junction or its interaction with trans-acting factors influence the spliceosome assembly. Interestingly, we have found, by UV crosslinking experiments, that a 43 kDa protein specifically binds to a UG tract. Affinity purification of this protein has allowed its identification as the TDP-43 cellular protein. Originally, TDP-43 was isolated following its binding to the pyrimidine rich motifs of Human

Immunodeficiency Viruses type I TAR DNA sequence where is likely to be critical in the control of gene expression [215]. In fact, it is a potent inhibitor of HIV-1 gene expression, during the early phase the HIV-1 cycle. The structural organisation of this protein is similar to other RNA binding proteins: it contains three RRM (RNA recognition motifs) and a glycine rich domain. However, its cellular role is presently unknown [215]. The group of cellular proteins to which TDP-43 belongs binds to a variety of different RNAs [67, 223]. Interestingly, Ou et al. have demonstrated binding ability of this protein for the double stranded DNA, exclusively for polypyrimidine sequences containing at least 8 pyrimidines in the row [215]. In contrast with this findings, our studies have shown that short RNA sequence consisting of alternated pyrimidine/purine, (UG)12, are successfully recognised and bound by TDP-43. In fact, UG but not U repeats within the intron 8 of human CFTR mRNA binds TDP-43. Morover, we have found that TDP-43 is able to bind different DNA and RNA sequences.

Up to now, it is not clear which is the function, if any, of this cellular factor in the splicing regulation. It is possible that the increase in length of the UG tract would result in an increase of the binding affinity for this protein. Alternatively, the UG tract might bind more molecules of TDP-43, where number of the bound molecules could increase with the increasing number of UG repeats. In fact, the glycine rich domain of the TDP-43 suggests its ability to form protein-protein interaction.

Through its binding to UG sequence, TDP-43 might influence splicing assembling in two ways: its binding might indirectly mask poly U stretch modulating the binding efficiency of $U2AF^{65}$ to the polypyrimidine tract or, alternatively, TDP-43 could directly be involved in the formation of new nonfavored protein-protein interaction with other proteins of the splicing machinery thus affecting exon 9 recognition. In any case, it is conceivable that binding of TDP-43 to the UG tract might lower exon 9 definition by acting as a "disturbing" element placed in proximity of its 3′ splice site junction.

4.3. Cis-acting and trans-acting elements regulating mouse and human exon 9 definition

The peculiarity of the 3′ splice site of human CFTR exon 9 is further highlighted by the comparison with the 3′ splice site of its mouse counterpart. This comparison outlined the high degree of identity between the exons 9 and low homology between human and mouse introns. Interestingly, mouse does not present the polymorphic loci and does not show alternative splicing.

These findings led us to think that TG and T tract at the 3′ splice site of human exon 9 could be the only element involved in weakening the exon 9 definition and modulation of splicing efficiency in humans. To test if the alternative splicing of human depends only on the particular configuration of the 3′ splice site, we have interchanged human and mouse 3′ splice sites. When the human TG/T-tract was juxtaposed to mouse exon 9, no alternative spliced form was observed, even when T5 allele was present. This data pointed that the effect of the TG an T tract might be influenced by some other elements. The different results of transfections p(TG)11/T5 and p(TG)11/T5/M/m, (alternative splicing was detected only in p(TG)11/T5) indicate that these elements are different in human and mouse context. Murine exon 9 and intron 9 enviroment is functionally stronger than human and this could be due to the presence of an enhancer in mouse, or a silencer in human, or both. In order to localise the elements (in human or mouse) which are responsible for the different involvement of TG/T tract in exon 9 recognition we progressively replaced human with mouse sequences.

The first information obtained from transfection with pm/H/h construct concerns the weakness of mouse 3′ splice site compared to the human one.

Then, the transfections with p(TG)11/T5/M/m and p(TG)11/T5/H/m and h/m3′T5 constructs (which showed 100% exon 9 inclusion) permitted to exclude

exon 9 as possible location of other regulatory elements and pointed out to murine intron 9 as a stronger element compared to the human one.

Futhermore, these results along with those concerning the constructs mutated at the 5′ splice site (which continued to show weakening of exon 9 definition only when T5 allele was present) suggest that the 5′ splice junctions are not responsible for the differences in splicing pattern between human and mouse when the same TG/T genotype is present at 3′ splice sites. In addition, the comparison of the mouse and human 5′ splice site by computer analysis has shown that human exon 9 5′ junction is closer to the consensus than the murine one.

Finally, the results of transfection with pm/H/h and pm/H/m hybrid constructs confirmed that mouse intron 9 is more positive than the human one.

Considering all these data, the reason of these observations might be due to the presence of regulatory elements, that might be either positive in mouse (i.e. an enhancer) or negative in human (i.e. a silencer), in the region of intron 9 spanning between base -77 and -176 and even both possibilities might coexist. However, indication of the functional role of the intronic human sequences arises from the experiments of trans-acting factors overexpression.

Surprisingly, the following characterization of the trans-acting factors involved in the regulation of the human exon 9 alternative splicing has revealed that both ASF/SF2 and hnRNPAI have a negative effect in splicing of human exon 9. Previous studies have shown that ASF/SF2 can, in certain circumstances, inhibit splicing. In fact, it has been shown to bind to a splicing repressor sequence in the Rous sarcoma virus RNA [224] and to interact with the purine rich sequence in the IIIa intron of the adenovirus L1 transcript repressing IIIa splicing in vitro, probably by blocking U2 snRNP binding to the branch site [156]. It is thought that negative regulatory function of ASF/SF2 reported for the viruses might be important to maintain the balance between spliced and unspliced RNA, which is important for the their life cycle [225].

We have shown that human but not mouse intronic sequences located between 77 and 176 bases downstream of the 5′ splice site show binding specificity for some SR proteins, possibly, ASF/SF2 and/or SC35. Therefore, it is plausible that the negative role of this SR proteins in splicing of human exon 9 might be due to its binding within the localised intron sequence.

It has to be mentioned that our data, which indicate that the negative effect in exon recognition in the human exon 9 is due to the elements within the intron 9 (position -77- -176) do not exclude the existance of the elements in mouse that have an positive effect for mouse exon 9 recognition. The elements that are involved in splicing of mouse exon 9 still has to be characterised.

Finally, concerning that the negative effect of the hnRNPA1 on the human exon 9 reconition, we suggest that it might be due to its interaction with the 3′ splice site as was described in the paragraph at the section 4.2.).

4.4. An evolutionary point of view

The different regulation of the human and mouse CFTR exon 9 might be interpreted from the evolutionary point of view.

In the mammalian genomes there are abundant traces of recombination events via transposons or retroviruses that resulted in substantial changes in specific regions of the genomes [226, 227]. It has been proposed that the human CFTR gene underwent a retrotransposition event and then evolved and diverged from the cognate CFTR gene early during the evolution [228]. This hypothesis is supported by the findings of Rozmahel et al., who have found that fragments of CFTR gene spanning exon 9 and its flanking introns together with its polymorphic loci are present in multiple copies in the human genome with an identity degree of 92-96%. Also, the poly A tract and the 9 bp sequence (AAACAGACA) detected at the junctions of these multiple CFTR sequence copies suggest retro-transposition type of insertion event, possibly mediated by L1

elements [227, 228]. Since the exon 9 sequence is present in only one copy in Old World monkeys CFTR, retrotransposition and amplification events in human can be dated between 7 and 10 million years ago [228].

Our findings have now shown that mouse intronic sequences flanking exon 9 do not show the homology with human counterpart and that human and mouse introns 8 and 9 differ strikingly in their size. Therefore, it is conceivable that insertion- and rearrangement- events occurred in intervening sequences surrounding human CFTR exon 9, placing the new cis-acting elements in proximity of its splice junctions which lowered exon definition and provided a basis for exon 9 skipping in CFTR mRNA.

4.5. Models of human CFTR exon 9 alternative splicing

On the base of our data, at least two models can be hypothesised to explain the alternative splicing of the human CFTR exon 9.

The first is based on the decrease of the exon 9 definition through two independent mechanisms, one, determined by the polymorphic loci at 3′ end of intron 8 and the other, controlled by the ISS placed within the 5′ region of human intron 9. In this case, the intracellular concentration of trans-acting factors might be critical for the regulation of splicing. At the 3′ splice site, the binding of TDP-43 to UG tract might directly compete with the binding of polypyrimidine binding proteins ($U2AF^{65}$ or other proteins of splicing machinery) to the polyT tract. This competition might become more evident when the polypyrimidine tract is short (T5). If we consider that the binding of TDP-43 to the UG tract is constitutive, this could result in a sterical interference for the binding of $U2AF^{65}$ to the poly-T. This interaction could be significantly affected when the number of thymidines progressively decreases to five. It is possible that the displacement of the polypyrimidine binding proteins would

open the access for the binding of hnRNPAI to the 3′ splice junction, further decreasing the recognition efficiency of the 3′ splice site (Figure 4.2.).

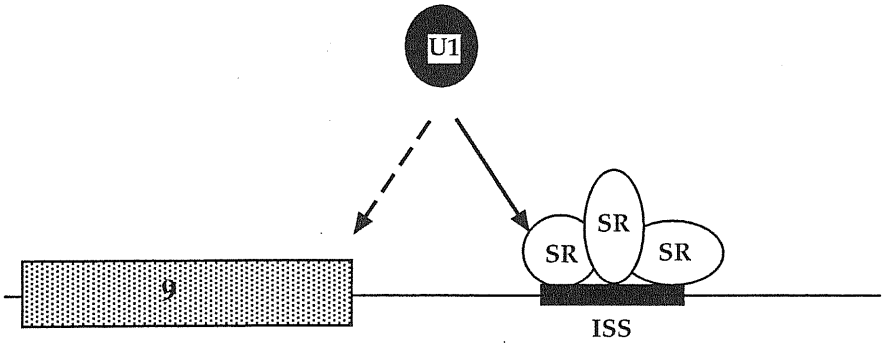


Figure 4.3. Model for 5′ splice site independent inhibition.

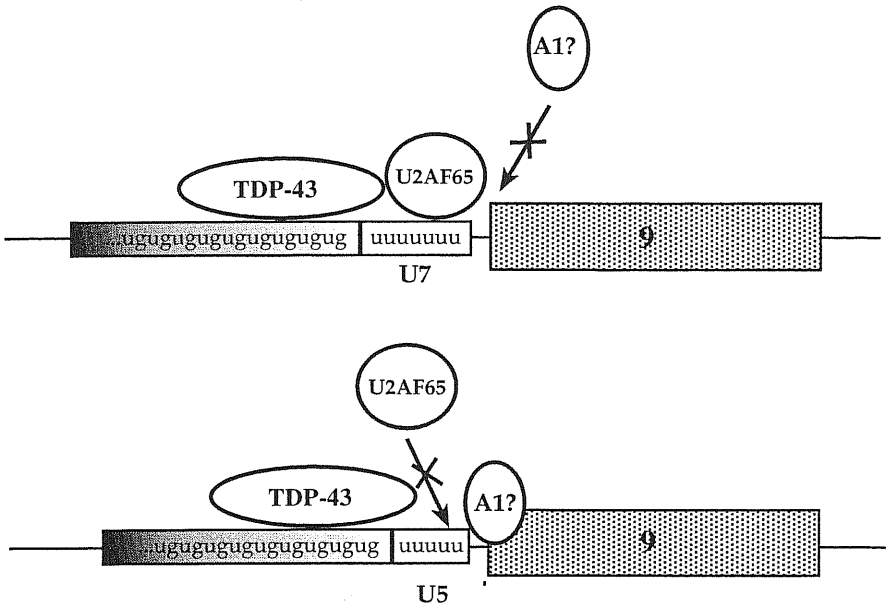


Figure 4.2. Model for 3′ splice site independent inhibition.

At the 5′ splice site, the binding of SR proteins within the 5′ end of intron 9 might exert an inhibitory effect on the recognition of the 5′ splice site recruiting U1snRNP far from the 5′ splice junction (Figure 4.3.).

The second model is based on that recently proposed by Blanchette and Chabot, where skipping of alternative exon 7B from hnRNPAI pre-mRNA is regulated by combined action of different intron elements at 3′ and 5′ site of the exon. Binding

of trans-acting factors to the intronic sequences causes change in the pre-mRNA conformation and increases commitment between 3′ and 5′ splice sites of neighbouring exons [229]. A similar model can be hypothesised to explain the splicing regulation of exon 9.

A bridge between proteins bound at the 3′ end of intron 8 (TDP-43) and at 5′ end of intron 9 (ASF/SF2 or SC35) might increases the frequency of exon 9 skipping. However, up to now there is no evidence for TDP-43-SR interactions (Figure 4.4.).

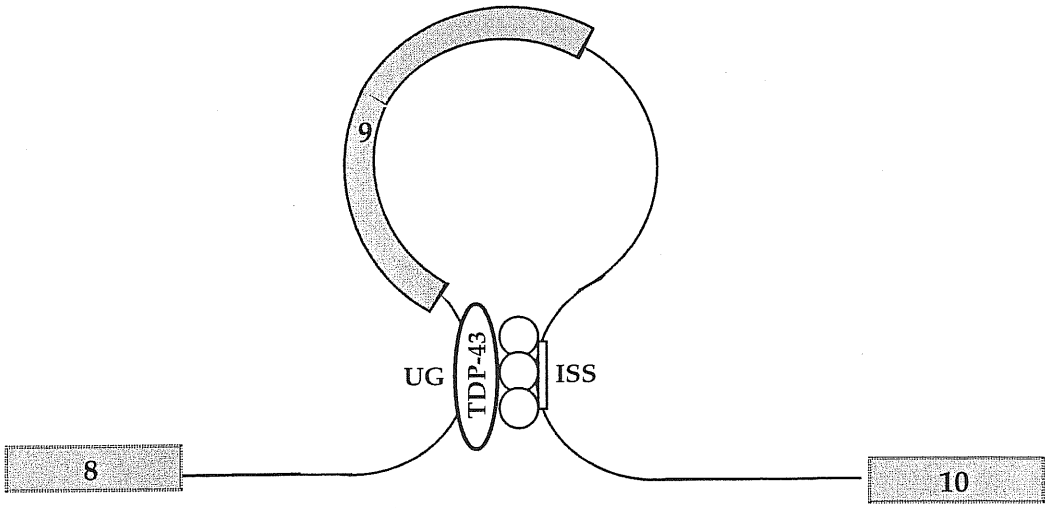


Figure 4.4. **Model of exon 9 exclusion through bridging across the introns.**

4.6. Tissue specific expression of the splicing factors and regulation of the splicing in CFTR exon 9

Several reports have shown that proportion of exon 9- form of CFTR mRNA differ in different tissues [230, 231]. The transfections of the cell lines of different tissue- and specie- origin with constructs carrying different TG/T number confirm previous observations and strongly suggest that the splicing environment for the CFTR transcript can vary among the different cell types. In fact, a slight but reproducible difference in the exclusion rate of exon 9 was

observed in the NCI-H322 human cell line compared with the Hep3B and 293 human cell lines.

The results of transfections in different cell lines are consistent with the observations by Hanamura et al. who have detected the high level of expression of ASF/SF2 and hnRNPA1 in lung and pancreas and lower levels in kidney and liver [154]. Therefore, our studies suggest that the concentration of the ASF/SF2, hnRNPA1 and possibly of TDP-43, whose expression is also tissue- regulated [215] might be additional elements that mediate the tissue- specific regulation of the alternative splicing of exon 9.

However, in our system, the proportion of exon 9 skipping in NCI-H322 does not reach the reported levels of exon 9 exclusion observed in the individuals with the T5 allele [35, 50, 231, 232] although has the same qualitative trend. The quantitative differences might be due to the variations in cell- and/or tissue-specific trans-acting factors involved in the regulation of the CFTR exon 9 splicing. It should be noticed that in our construct, exon 9 and its intronic region are flanked by unrelated genomic sequences lacking CFTR distant exons and, possibly, some other crucial cis-acting element. Hence we cannot rule out that the minigene context might influence the level of exon 9 exclusion. In any case, the TG/T constructs show a very clear qualitative difference due only to the polymorphic loci as they differ only in the length of the TG and/or T tract making them a suitable system for studying the alternative splicing of CFTR exon 9.

It is, however, surprising that Hep3B and Nmuli give such a stricking difference in efficiency of CFTR exon 9 alternative splicing, considering that both cells have the same tissue of origin. It is unlikely that we are seeing a species-specific interactions of the murine and human splicing factors because they are highly homologous [194]. It is possible that the cells represent different stages of hepatocyte differentiation and may have qualitative and quantitative differences in the set of trans-acting factors.

In conclusion, it is conceivable that CBAVD is the consequence of the development- and tissue-specific levels of expression of the several splicing factors that, alone or in cooperation, reduce the recognition of exon 9 within CFTR.

4.7. FUTURE DIRECTIONS

The work described in this thesis presents a first effort in understanding mechanisms of splice site selection of the alternatively spliced exon 9 of the human CFTR gene. We demonstrated: 1. at least three cis acting elements involved in splice site selection, 2. TDP-43 is a protein that recognise and binds UG sequence; 3. overexpression of the ASF/SF2 has an negative effect on the exon definition; 4; exon 9 splicing pattern changes in different cell types.

However, despite the advances concerning exon 9 splicing mechanism, a lot of questions still have to be answered.

The identification of the TDP-43 protein gave us possibility to speculate about the functional involvement of this protein in 3′ splice site selection but this hypothesis still have to be proved. The experiments where the expression vector carrying TDP-43 cDNA would be cotransfected with the p(TG)11/T5 and p(TG)11/T7, might indicate its direct role in splicing. Another approach could be the set up of an in vitro splicing system where it might be possible to investigate the effects of TDP-43 depletion from nuclear extract (using antibodies) on CFTR exon 9 splicing pattern. In such a way, it will be demonstrated if and how complementation of the TDP-43 deficient nuclear extracts with the recombinant protein may restore the original splicing pattern. Also, taking into consideration that it contains protein binding motif, it would be interesting to see if and which are the proteins that interact with TDP-43.

Concerning the region that we named ISS, it needs a more fine mapping, for example, by mutagenesis. Then, a possible way to verify its functional role is to place it in a heterologous context.

It will be also important to verify the identity of the factor/s that bind to the supposed ISS and this might be carried out using monoclonal antibodies specific for ASF/SF2 and SC35.

Also our demonstration that exon recognition varies in different cell types is interesting and deserve to be studied more profoundly. It is intriguing that this tissue and developmentally regulated splicing seems to be responsible for the pathology of the monosymptomatic disease, as CBAVD. Since we have demonstrated the negative effect of the overexpression of the ASF/SF2 on the splicing efficiency it is logical to conclude that the endogenous concentration of the splicing factors can shift the splicing pattern. The endogenous concentration of the SR proteins among different cell lines (for example Hep3B and Nmuli) can be analysed by Western blot and correlated with the proportion of the exon 9- or alternatively the phosphorylation status of ASF/SF2 can also be determined. Another approach, would be to check the SR status within the same cell type during its differentiation. It is already shown that exon 9 splicing pattern changes 8 fold within the nondifferentiated and differentiated HT-29 cell line and this cell line could be a good model for an analysis of the changes of the splicing pattern due to the changes in the concentration of the SR proteins.

CHAPTER 5

MATERIALS AND METHODS

5.1. MATERIALS

5.1.1. Chemical reagents

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

5.1.2. Enzymes

Restriction enzymes were from Parmacia 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 Boehringer Manheim 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 insterile water and boiled for 10 minutes to destroy trace amounts of DNAase activity. All enzymes were used following manufacturer instructions.

5.1.3. Synthetic oligonucleotides

Synhetic DNA and RNA oligonucleotydes were purchhased from from Roche Diagnostic.

5.1.4. Radioactive isotopes

Radioactive α-^{32}P dCTP, γ-^{32}P dATP, ^{35}S, α-^{32}P-UTP were supplied by Amersham U.K. Ltd.

5.1.5. Bacterial culture

The K12, *E. coli* family strain DH5α 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. Where appropriate antibiotics, ampicillin and tetracycline, were added to media to a final concentration of 200 µg/ml (ampicillin) and 7 µg/ml (tetracycline).

5.2. Cell culture

The following cell lines were used:

1. **NMuli**; mouse, liver, epithelial,
2. **Hep3B**; human, liver tumor, fibroblast-like,
3. **293**; human, embryonal kidney, epithelial,
4. **cos-7**; African green monkey, kidney, fibroblast,
5. **NCI-H322**; (ECACC 95111734); human, bronchial carcinoma.

5.3. METHODS

5.3.1 Nucleic Acids Preparations

5.3.1.2. 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. (1989) [233]. The final pellet was resuspended in 50 µl of dH_2O and 5 µl of such preparation were routinely taken for analysis by restriction enzyme digests.

5.3.1.3. 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 hugh amount of plasmid, we were dealing with 300- 500 ml of overnight inoculum using LB or TB medium.

5.3.1.4. 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 two times. Supernatant was then precipitated with absolute EtOH. The final pellet was then resuspended in 40 µl of ddH_2O and frozen at -80°C. The RNA quality was checked by electrophoresis on 1% agarose gels.

5.3.1.5. Preparation of RNA from tissues

To prepare RNA from tissues the basic protocol of Chomczynski [234] was followed. Frozen tissue (0.5-1 g) was homogenised in a correspondent volume (0.5-1 ml) of D solution (guanidine thiocyante, 4 mM, Na-citrate, 25 mM, β-mercaptoethanol, 100 mM, lauroyl- sarcosine, 0.5%, antifoam A, 0.1%) using a UltraTurrax homogenizer. A further phenol-chloroform step was added to the basic protocol to improve the sample purification from proteins and fatty acids. The final pellet was then resuspended in 100 µl of ddH_2O and frozen at -80°C.

5.4. Estimation Of Nucleic Acid Concentration

5.4.1. Spectrophotometric

An optical density of 1.0 at 260 nm is roughly equivalent to a concentration of 50 µg/µl for double stranded DNA, 33 µg/µl for single stranded DNA and 40 µg/µl 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.

5.4.2 Enzymatic Modification of DNA

5.4.2.1. 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 µl containing 5 U of the appropriate restriction enzyme per µg 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 µg 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.

5.4.2.2. 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.

5.4.2.3. 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.

5.4.2.4. 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 16°C for 6-12 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.

5.4.2.5. 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 µl of gel buffer. Aliquots were pooled and DNA recovered by ethanol precipitation. To remove small gel fragments remaining the aqueous solution was centrifuged for 10 minutes at 4°C and the supernatant was transferred to a fresh tube.

5.4.2.6. PCR reaction for a cloning purposes

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 µl with 1x Taq buffer, dNTP mix 200 µM 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.

5.4.2.7. Expand Long Range Polymerase Chain Reaction

When a fragments over 3 kb were amplified a Expand Long Template PCR kit (Boehringer Mannheim) was used according to the manufacturer instruction.
PCR were performed in a final volume of 50 µl adding 100-200ng of DNA.

5.4.2.8. Sequence analysis

Sequence analysis of plasmid DNA were performed when was necessary by the dideoxinucleotyde chain termination method using a T7-based DNA sequencing kit (Amersham-Pharmacia). Approximatelly, 1 µg 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 dryed and exposed to kodak x-omat films for 12 hours.

5.5. 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′-ttttcatatggggccgctctaggacttgataatgggcaaatatctta-3′; hcfIVS9 rev, 5′-cccctcgaccatatgctcgccatgtgcaagatacag-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.

To permit subsequent clonings, Eco RI target site was introduced through a two-step PCR overlap extension method [235] 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)m alleles, a XbaI-EcoRI cassette was generated by PCR using a common sense primer (5′-catatggggccgctctagga-3′) and antisense primer (5′-aaagaattccccaaatccctgttaaaaaaacacacacacacacacacacacatcaaaaataaaagatgagtt-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 EDA-intron of the pSV-mEDA NdeI-digested vector, previously described [139]

In order to study the influence of exon definition on 3′-splice site recognition, we mutagenised the 5′-splice site of exon 9 by replacing guanine in position +1 within intron 9 with an adenine (AAG/*g*tag -> AAG/*a*tag) using a two-step PCR overlap extension method. After mutagenesis, the new 5′-splice site selected for exon 9 splicing was placed 5 bases upstream the wild type one (CCAG/GCAAG*a*tag) (Fig.6b).

Mouse exon 8-9 and exon 9-10 were amplified by long-range PCR of CD1 strain as template and the following oligos: 8dir 5′-agtataacttaatgaccacaggcataatc-3′; 9rev, 5′-tccagtagatccagtaatagccaac atctc-3′; 9dir, 5′-gagatgttggctattactggatctactgga-3′; 10rev, 5′-atctgtactcatcataggaaacac caaaga-3′.

The 2.2 kb exon 8-9 and the 6 kb exon 9-10 fragments were blunt-end cloned into pBluescript and exon and their junctions were sequenced. The sequences were submitted in a genebank; intron 8 (157 bp; Genbank n. AF176095) and of intron 9 (209 bp; Genbank n. AF176096).

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′-ttttcatatgtctagaaaccatgtgctttatagt-3′, that include Nde I and Xba I target sites; mCF9i rev, 5′-aaaacatatgataggttatccaatcttaagtgatcagtt ctaaacacgtgta-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′-ccaggaattccccaaatccctattcata-3′and 5′-atagggatttggggaattcctggaga-3′.

Mouse minigene was NdeI excised from pBluescript and transferred into the EDA-intron of the pSV-mEDA 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/m3′dir, 5′-tgttggcggttgctggatctactggatcaggaaaggtact-3′, h/m3′rev 5′-agtacctttcctgatccagtagatccagcaaccaccaaca-3′. Before expression, identity of all construct was checked by sequencing.

5.6. Maintenance and analysis of cells in culture

Cells were propagated in recomanded media and were maintained in 100x20 mm Falcon tissue culture plates, incubated at 37°C and with 5% carbon dioxide. NCI-H322 cell line (ECACC 95111734) was maintained in culture in RPMI supplemented with 10% fetal calf serum, 50 μg/ml gentamicin and 4 mM glutamine. Hep3B, N-Muli, COS-7 and 293 cell lines were growen 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.

5.6.1. 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 in two ways; NCI-H322 and COS7 by DOTAP Liposomal Transfection Reagent (Boehringer Mannheim); 3μg of expression plasmid was mixed with 20μg of lipofectin and the mixture was incubated at room temperature for 15min. 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.

Hep3b, Nmuli, 293 were transfected by calcium-phosphate technique folowing instructions of Maniatis (chapter 16.33.).

In, cotransfection experiments, cells were transfected with 3μg of the minigene variants and 500ng of the plasmid carrying cDNA coding for ASF/SF2 or hnRNPA1. These plasmids were kindly provided by dr. J. Caceres.

An equal amount of plasmid DNA pβ5'SVBglII [236] was cotransfected as a source of SV40 T antigen required for stimulation of replication of the test plasmids which all contained an SV40 origin of replication.

Also, 200 ng of vector carrying the gene encoding the human growth hormone (hGH) was contrasfected in each transfection experiment, whose expression was quantitatively measured by hGH ELISA and used to normalise the transfection efficiency.

Each transfection experiment was repeated at least 3 times.

5.6.2. 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 Boehringer Mannheim. The 200μl of diluted medium was 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 200μl of working dilution of anti-hGH-DIG was added in each well following incubation of 1h at 37°. After five washings, 200μl 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.

5.7. mRNA Analysis by Polymerase Chain Reaction

5.7.1. cDNA synthesis

In order to synthetise cDNA, the 3 μg of total RNA from cells or 5 μg of total RNA from tissues were mixed with 100ng specific oligonucleotides and a sterile water in final volume of 20 μl. After denaturation at 95°C speific buffer (10mM Tris-HCL (pH8.4), 50mM KCl, 2.5mM, $MgCl_2$) 10mM DTT, 1mM of each dNTPs, RNAsin 0.1U/μl 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.

5.7.2. PCR analysis

PCRs were carried out for 35 cycles (45 sec at 94°C, 30 sec at 60°C, 30 sec at 72°C) in 50 μl reaction volumes. For radioactive amplifications, a 0.1μl of α-^{32}P-dCTP (1 μCi) was added to PCR reaction. Oligonucleotides used were those specific for the minigene system, placed at the a-globinEDA-1 and a-globin EDA+1 junctions (mFN-αglob dir 5′-cactgcctgctggtgacgtac-3′; mFN-αglob rev 5′-tgggcggccagggtcacggc-3′).

5.7.3. Analysis of the PCR-Amplified Products

5.7.3.1. Not radiolabelled PCR products

10 μl of each amplification reaction were then analysed by electrophoresis on a 2.0% (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.

5.7.3.2. 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% 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.

5.8.1. Radioactive labeling of the RNA

Plasmids used for the UV crosslinking were linearised by digestion with appropriate restriction enzyme. Transcription with T7 RNA polymerase (Stratagene) was performed using 1-2 µg of linearised plasmid, in the presence of α-^{32}P-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 $4x10^6$ c.p.m./µg of RNA.

5.8.2. UV crosslinking

HeLa nuclear extract were prepared from HeLa cells grown in suspension with 10% calf serum as described [237]. The total protein concentration was then measured by using Bio-Rad protein assay. The UV crosslinking assays were performed by adding α-^{32}P-UTP labeled RNA probes ($1x10^6$ 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 $MgCl^2$, 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 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 and 6U of RNAseT1 (1:250) by the incubation at 37°C for 30min in a water bath. Samples were then analised by 8% sodium-dodecil-sulfate polyacrylamide gel electrophoresis (SDS-PAGE) followed by autoradiography.

5.8.3. Preparations of competitors

Different CFTR genes containing different deletions at the polymorphic Tm and Tn were cloned under the T7 promotor of pBS. Transcription with T7 RNA polymerase (Stratagene) was performed using 1-2 µg of linearised plasmid (BamHI), without α-^{32}P-UTP added.

Short RNAs used as a competitors transcribed from a plasmids p(TCTT)3, p(TG)12, pTM-β and pTNT. These plasmids were obtained by the annealling of following oligonucleotides 5′-tgtgtgtgtgtgtgtgtgtgtg-3′ and 5′-acacacacacaca cacacacacac-3′ (p(TG)12); 5′-tcctcctccttcttcttcttcagg-3′ and '5-cctgaagaagaaggaggagga-3′(p(tctt)3); 5′-agggaaagacagggagggagagagaaagagaaagg-3 and 5′-cctttctctttctctctccctccctgtctttccct-3′(pTMβ); 5′-aagaggaagaatggcttgaggaagacgacg-3′ and 5′-cgtcgtcttcctcaagccattcttcctctt-3′ (pTNT) and cloning in pBS SK vector, cut with ClaI and blunt ended. Plasmids were linerised with Hind III and transcribed with T7 polymerase as described above.

Oligonucleotides were used as a single stranded DNA competitors, and their sequences were outlined in the tables above the appropriate figures.

DNA double stranded competitors were prepared by cutting of the plasmids p(TCTT)3 and p(TG)12 by HindIII and SalI and purifying the fragment of interest by elution from a 12% polyacrylamide gel.

5.8.3. Cross-linking of RNA to adipic dehydrazide agarose beads for affinity purification of cellular factors

500 pmoles of RNA (approx. 15 µg of a 100mer RNA) were placed in a 400 µl reaction mixture containing 100 mM NaOAC pH 5.0 and 5 mM sodium m-periodate (Sigma). The raction mixtures were incubated for 1 hour in the dark at room temperature. The RNA was then ethanol precipitated and resuspended in 500µl of 0.1 M NaOAC, pH 5.0. Then, 400µl 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, 300µl of 0,1 M NaOAC pH 5.0 were added to the beads. The slurry was then mixed with the periodate-treated RNA and incubate 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 650µl 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 $MgCl_2$. After the final centrifugation the proteins bound to the RNA on the beads were eluted by addition of 60 µl 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.

REFERENCES

1. Boat, T.F., Welsh, M.J., Beauder, A.L. (1989) *The metabolic basis of inherited disease*. McGraw-Hill, New York.
2. Welsh, M.J., Tsui, L., Boat, F.T. (1995) Cystic fibrosis. In *The Metabolic and Molecular Bases of Inherited Disease*, Volume 3McGraw-Hill, New York 3799-3786.
3. Knowles, M., Gatzy, J., Boucher, R. (1981) Increased bioelectric potential difference across respiratory epithelia in cystic fibrosis. *N Engl J Med* **305**, 1489-95.
4. Quinton, P.M. (1983) Chloride impermeability in cystic fibrosis. *Nature* **301**, 421-2.
5. Kopelman, H., Durie, P., Gaskin, K., Weizman, Z., Forstner, G. (1985) Pancreatic fluid secretion and protein hyperconcentration in cystic fibrosis. *N Engl J Med* **312**, 329-34.
6. Rommens, J.M., Iannuzzi, M.C., Kerem, B., Drumm, M.L., Melmer, G., Dean, M., Rozmahel, R., Cole, J.L., Kennedy, D., Hidaka, N., et al. (1989) Identification of the cystic fibrosis gene: chromosome walking and jumping. *Science* **245**, 1059-65.
7. Anderson, M.P., Gregory, R.J., Thompson, S., Souza, D.W., Paul, S., Mulligan, R.C., Smith, A.E., Welsh, M.J. (1991) Demonstration that CFTR is a chloride channel by alteration of its anion selectivity. *Science* **253**, 202-5.
8. Gregory, R.J., Cheng, S.H., Rich, D.P., Marshall, J., Paul, S., Hehir, K., Ostedgaard, L., Klinger, K.W., Welsh, M.J., Smith, A.E. (1990) Expression and characterization of the cystic fibrosis transmembrane conductance regulator [see comments]. *Nature* **347**, 382-6.
9. Cheng, S.H., Gregory, R.J., Marshall, J., Paul, S., Souza, D.W., White, G.A., O'Riordan, C.R., Smith, A.E. (1990) Defective intracellular transport and processing of CFTR is the molecular basis of most cystic fibrosis. *Cell* **63**, 827-834.
10. Hyde, S.C., Emsley, P., Hartshorn, M.J., Mimmack, M.M., Gileadi, U., Pearce, S.R., Gallagher, M.P., Gill, D.R., Hubbard, R.E., Higgins, C.F. (1990) Structural model of ATP-binding proteins associated with cystic fibrosis, multidrug resistance and bacterial transport [see comments]. *Nature* **346**, 362-5.
11. Riordan, J.R., Rommens, J.M., Kerem, B., Alon, N., Rozmahel, R., Grzelczak, Z., Zielenski, J., Lok, S., Plavsic, N., Chou, J.L., et al. (1989)

Identification of the cystic fibrosis gene: cloning and characterization of complementary DNA. *Science* **245**, 1066-73.

12. Chang, X., Hou, Y., Jensen, T.J., Riordan, J.R. (1994) Mapping of cystic fibrosis transmembrane conductance regulator membrane topology by glycosylation site insertion. *J Biol Chem* **269**, 18572-75.
13. Tabcharani, J., Rommens, J., Hou, Y., Chang, X., Tsui, L., al., e. (1993) Multi-ion pore behaviour in the CFTR chloride channel. *Nature* **366**, 79-82.
14. Akabas, M.H., Kaufmann, C., Cook, T.A., Archdeacon, P. (1994) Amino acid residues lining the chloride channel of the cystic fibrosis transmembrane conductance regulator. *J Biol Chem* **269**, 14865-8.
15. McDonough, S., Davidson, N., Lester, H.A., McCarty, N.A. (1994) Novel pore-lining residues in CFTR that govern permeation and open- channel block. *Neuron* **13**, 623-34.
16. Sheppard, D.N., Rich, D.P., Ostedgaard, L.S., Gregory, R.J., Smith, A.E., Welsh, M.J. (1993) Mutations in CFTR associated with mild-disease-form Cl- channels with altered pore properties [see comments]. *Nature* **362**, 160-4.
17. Travis, S., Carson, M., Ries, D., Welsh, M. (1993) Interaction of nucleotides with membrane-associated cystic fibrosis transmembrane conductance regulator. *J Biol Chem* **268**, 15336-39.
18. Carson, M., Travis, S., Welsh, M. (1995) The two nucleotide-binding domains of the cystic fibrosis transmembrane conductance regulator (CFTR) have distinct functions in controlling channel activity. *J Biol Chem* **270**, 1711-17.
19. Winter, M.C., Welsh, M.J. (1997) Stimulation of CFTR activity by its phosphorylated R domain. *Nature* **389**, 294-6.
20. Dulhanty, A., Riordan, J. (1994) Phosphorylation by cAMP-dependent protein kinase causes a conformational change in the R domain of the cystic fibrosis transmembrane conductance regulator. *Biochemistry* **33**, 4072-79.
21. Gadsby, D.C., Nairn, A.C. (1994) Regulation of CFTR channel gating. *Trends Biochem Sci* **19**, 513-8.
22. Trezise, A.E., Buchwald, M. (1991) In vivo cell-specific expression of the cystic fibrosis transmembrane conductance regulator. *Nature* **353**, 434-7.
23. Yoshimura, K., Nakamura, H., Trapnell, B.C., Chu, C.S., Dalemans, W., Pavirani, A., Lecocq, J.P., Crystal, R.G. (1991) Expression of the cystic fibrosis transmembrane conductance regulator gene in cells of non-epithelial origin. *Nucleic Acids Res* **19**, 5417-23.
24. Yoshimura, K., Nakamura, H., Trapnell, B.C., Dalemans, W., Pavirani, A., Lecocq, J.P., Crystal, R.G. (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.

25. Denning, G.M., Ostedgaard, L.S., Cheng, S.H., Smith, A.E., Welsh, M.J. (1992) Localization of cystic fibrosis transmembrane conductance regulator in chloride secretory epithelia. *J Clin Invest* **89**, 339-49.
26. Lukacs, G.L., Chang, X.B., Kartner, N., Rotstein, O.D., Riordan, J.R., Grinstein, S. (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.
27. Barasch, J., Kiss, B., Prince, A., Saiman, L., Gruenert, D., al-Awqati, Q. (1991) Defective acidification of intracellular organelles in cystic fibrosis [see comments]. *Nature* **352**, 70-3.
28. Bradbury, N.A., Jilling, T., Berta, G., Sorscher, E.J., Bridges, R.J., Kirk, K.L. (1992) Regulation of plasma membrane recycling by CFTR [see comments]. *Science* **256**, 530-2.
29. Ismailov, II, Awayda, M.S., Jovov, B., Berdiev, B.K., Fuller, C.M., Dedman, J.R., Kaetzel, M., Benos, D.J. (1996) Regulation of epithelial sodium channels by the cystic fibrosis transmembrane conductance regulator. *J Biol Chem* **271**, 4725-32.
30. Tsui, L.C. (1992) The spectrum of cystic fibrosis mutations. *Trends Genet* **8**, 392-8.
31. Hamosh, A., Trapnell, B.C., Zeitlin, P.L., Montrose-Rafizadeh, C., Rosenstein, B.J., Crystal, R.G., Cutting, G.R. (1991) Severe deficiency of cystic fibrosis transmembrane conductance regulator messenger RNA carrying nonsense mutations R553X and W1316X in respiratory epithelial cells of patients with cystic fibrosis. *J Clin Invest* **88**, 1880-1885.
32. Anderson, M.P., Welsh, M.J. (1992) Regulation by ATP and ADP of CFTR chloride channels that contain mutant nucleotide-binding domains [published erratum appears in Science 1992 Dec 11;258(5089):1719]. *Science* **257**, 1701-4.
33. Romey, M.C., Guittard, C., Chazalette, J.P., Frossard, P., Dawson, K.P., Patton, M.A., Casals, T., Bazarbachi, T., Girodon, E., Rault, G., Bozon, D., Seguret, F., Demaille, J., Claustres, M. (1999) Complex allele [-102T>A+S549R(T>G)] is associated with milder forms of cystic fibrosis than allele S549R(T>G) alone. *Hum Genet* **105**, 145-50.
34. Chiba-Falek, O., Kerem, E., Shoshani, T., Aviram, M., Augarten, A., Bentur, L., Tal, A., Tullis, E., Rahat, A., Kerem, B. (1998) The molecular basis of disease variability among cystic fibrosis patients carrying the 3849+10 kb C-->T mutation. *Genomics* **53**, 276-83.

35. Chu, C.S., Trapnell, B.C., Curristin, S., Cutting, G.R., Crystal, R.G. (1993) Genetic basis of variable exon 9 skipping in cystic fibrosis transmembrane conductance regulator mRNA. *Nat Genet* **3**, 151-6.
36. Fulmer, S.B., Schwiebert, E.M., Morales, M.M., Guggino, W.B., Cutting, G.R. (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.
37. Highsmith, E.W., Burch, L.H., Ahou, Z., Olsen, J.C., Boat, T.E., al., e. (1994) A novel mutation in the cystic fibrosis gene in patients with pulmonary disease but normal sweat chloride concentrations. *N Engl J Med* **331**, 974-980.
38. Zielenski, J., Markiewicz, D., Lin, S.P., Huang, F.Y., Yang-Feng, T.L., Tsui, L.C. (1995) Skipping of exon 12 as a consequence of a point mutation (1898 + 5G-- >T) in the cystic fibrosis transmembrane conductance regulator gene found in a consanguineous Chinese family. *Clin Genet* **47**, 125-32.
39. Kristidis, P., Bozon, D., Corey, M., Markiewicz, D., Rommens, J., Tsui, L.C., Durie, P. (1992) Genetic determination of exocrine pancreatic function in cystic fibrosis. *Am J Hum Genet* **50**, 1178-84.
40. Anguiano, A., Oates, R.D., Amos, J.A., Dean, M., Gerrard, B., Stewart, C., Maher, T.A., White, M.B., Milunsky, A. (1992) Congenital bilateral absence of the vas deferens. A primarily genital form of cystic fibrosis. *Jama* **267**, 1794-7.
41. Strong, T.V., Smit, L.S., Turpin, S.V., Cole, J.L., Hon, C.T., Markiewicz, D., Petty, T.L., Craig, M.W., Rosenow, E.C.d., Tsui, L.C., et al. (1991) Cystic fibrosis gene mutation in two sisters with mild disease and normal sweat electrolyte levels. *N Engl J Med* **325**, 1630-4.
42. Dork, T., Wulbrand, U., Richter, T., Neumann, T., Wolfes, H., Wulf, B., Maass, G., Tummler, B. (1991) Cystic fibrosis with three mutations in the cystic fibrosis transmembrane conductance regulator gene. *Hum Genet* **87**, 441-6.
43. Kiesewetter, S., Macek, M., Jr., Davis, C., Curristin, S.M., Chu, C.S., Graham, C., Shrimpton, A.E., Cashman, S.M., Tsui, L.C., Mickle, J., et al. (1993) A mutation in CFTR produces different phenotypes depending on chromosomal background. *Nat Genet* **5**, 274-8.
44. Chu, C.S., Trapnell, B.C., Curristin, S.M., Cutting, G.R., Crystal, R.G. (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.

45. Bremer, S., Hoof, T., Wilke, M., Busche, R., Scholte, B., Riordan, J.R., Maass, G., Tummler, B. (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.
46. Hull, J., Shackleton, S., Harris, A. (1994) Analysis of mutations and alternative splicing patterns in the CFTR gene using mRNA derived from nasal epithelial cells. *Hum Mol Genet* **3**, 1141-6.
47. Melo, C.A., Serra, C., Stoyanova, V., Aguzzoli, C., Faraguna, D., Tamanini, A., Berton, G., Cabrini, G., Baralle, F.E. (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.
48. Will, K., Stuhrmann, M., Dean, M., Schmidtke, J. (1993) Alternative splicing in the first nucleotide binding fold of CFTR. *Hum Mol Genet* **2**, 231-5.
49. Yoshimura, K., Chu, C.S., Crystal, R.G. (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.
50. Delaney, S.J., Rich, D.P., Thomson, S.A., Hargrave, M.R., Lovelock, P.K., Welsh, M.J., Wainwright, B.J. (1993) Cystic fibrosis transmembrane conductance regulator splice variants are not conserved and fail to produce chloride channels. *Nat Genet* **4**, 426-31.
51. Delaney, S.J., Koopman, P., Lovelock, P.K., Wainwright, B.J. (1994) Alternative splicing of the first nucleotide binding fold of CFTR in mouse testes is associated with specific stages of spermatogenesis. *Genomics* **20**, 517-8.
52. Trezise, A.E., Buchwald, M., Higgins, C.F. (1993) Testis-specific, alternative splicing of rodent CFTR mRNA. *Hum Mol Genet* **2**, 801-2.
53. Strong, T.V., Wilkinson, D.J., Mansoura, M.K., Devor, D.C., Henze, K., Yang, Y., Wilson, J.M., Cohn, J.A., Dawson, D.C., Frizzell, R.A., 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.
54. Pignatti, P.F., Bombieri, C., Benetazzo, M., Casartelli, A., Trabetti, E., Gile, L.S., Martinati, L.C., Boner, A.L., Luisetti, M. (1996) CFTR gene variant IVS8-5T in disseminated bronchiectasis [letter]. *Am J Hum Genet* **58**, 889-892.
55. Zielenski, J., Patrizio, P., Corey, M., Handelin, B., Markiewicz, D., Asch, R., Tsui, L.C. (1995) CFTR gene variant for patients with congenital absence of vas deferens [letter]. *Am J Hum Genet* **57**, 958-60.

56. Burger, J., Macek, M., Jr., Stuhrmann, M., Reis, A., Krawczak, M., Schmidtke, J. (1991) Genetic influences in the formation of nasal polyps [letter]. *Lancet* **337**, 974.
57. Culard, J.F., Desgeorges, M., Costa, P., Laussel, M., Razakatzara, G., Navratil, H., Demaille, J., Claustres, M. (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.
58. Mercier, B., Verlingue, C., Lissens, W., Silber, S.J., Novelli, G., Bonduelle, M., Audrezet, M.P., Ferec, C. (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.
59. Chillon, M., Casals, T., Mercier, B., Bassas, L., Lissens, W., Silber, S., Romey, M.C., Ruiz-Romero, J., Verlingue, C., Claustres, M., 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.
60. Costes, B., Girodon, E., Ghanem, N., Flori, E., Jardin, A., Soufir, J.C., Goossens, M. (1995) Frequent occurrence of the CFTR intron 8 (TG)n 5T allele in men with congenital bilateral absence of the vas deferens. *Eur J Hum Genet* **3**, 285-93.
61. Hamosh, A., Oates, R., Amos, J., Cutting, G.R. (1994) Correlation of the level of full-length CFTR transcript with pulmonary phenotype in patients carrying R117H and 1342-1,-2delAG mutations in 8th annual North American Cystic Fibrosis conference., Orlando, Florida,.
62. Osborne, L.R., Lynch, M., Middleton, P.G., Alton, E.W., Geddes, D.M., Pryor, J.P., Hodson, M.E., Santis, G.K. (1993) Nasal epithelial ion transport and genetic analysis of infertile men with congenital bilateral absence of the vas deferens [published erratum appears in Hum Mol Genet 1993 Nov;2(11):1990]. *Hum Mol Genet* **2**, 1605-9.
63. Cuppens, H., Lin, W., Jaspers, M., Costes, B., Teng, H., Vankeerberghen, A., Jorissen, M., Droogmans, G., Reynaert, I., Goossens, M., Nilius, B., Cassiman, J.J. (1998) Polyvariant mutant cystic fibrosis transmembrane conductance regulator genes. The polymorphic (Tg)m locus explains the partial penetrance of the T5 polymorphism as a disease mutation. *J Clin Invest* **101**, 487-96.
64. Berget, S.M., Moore, C., Sharp, P.A. (1977) Spliced segments at the 5′ terminus of adenovirus 2 late mRNA. *Proc Natl Acad Sci U S A* **74**, 3171-5.
65. Chow, L.T., Gelinas, R.E., Broker, T.R., Roberts, R.J. (1977) An amazing sequence arrangement at the 5′ ends of adenovirus 2 messenger RNA. *Cell* **12**, 1-8.

66. Jeffreys, A.J., Flavell, R.A. (1977) The rabbit beta-globin gene contains a large large insert in the coding sequence. *Cell* **12**, 1097-108.
67. Dreyfuss, G., Matunis, M.J., Pinol-Roma, S., Burd, C.G. (1993) hnRNP proteins and the biogenesis of mRNA. *Annu Rev Biochem* **62**, 289-321.
68. Visa, N., Alzhanova-Ericsson, A.T., Sun, X., Kiseleva, E., Bjorkroth, B., Wurtz, T., Daneholt, B. (1996) A pre-mRNA-binding protein accompanies the RNA from the gene through the nuclear pores and into polysomes. *Cell* **84**, 253-64.
69. Zeng, C., Kim, E., Warren, S.L., Berget, S.M. (1997) Dynamic relocation of transcription and splicing factors dependent upon transcriptional activity. *Embo J* **16**, 1401-12.
70. Sleeman, J.E., Lamond, A.I. (1999) Nuclear organization of pre-mRNA splicing factors. *Curr Opin Cell Biol* **11**, 372-7.
71. Corden, J.L., Patturajan, M. (1997) A CTD function linking transcription to splicing. *Trends Biochem Sci* **22**, 413-6.
72. Bentley, D. (1999) Coupling RNA polymerase II transcription with pre-mRNA processing. *Curr Opin Cell Biol* **11**, 347-51.
73. Lamond, A.I. (1993) The spliceosome. *Bioessays* **15**, 595-603.
74. Shapiro, M.B., Senapathy, P. (1987) RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. *Nucleic Acids Res* **15**, 7155-74.
75. Reed, R. (1989) The organization of 3′ splice-site sequences in mammalian introns. *Genes Dev* **3**, 2113-23.
76. Nelson, K.K., Green, M.R. (1989) Mammalian U2 snRNP has a sequence-specific RNA-binding activity. *Genes Dev* **3**, 1562-71.
77. Kambach, C., Walke, S., Nagai, K. (1999) Structure and assembly of the spliceosomal small nuclear ribonucleoprotein particles. *Curr Opin Struct Biol* **9**, 222-30.
78. Konarska, M.M., Sharp, P.A. (1987) Interactions between small nuclear ribonucleoprotein particles in formation of spliceosomes. *Cell* **49**, 763-74.
79. Flickinger, T.W., Salz, H.K. (1994) The Drosophila sex determination gene snf encodes a nuclear protein with sequence and functional similarity to the mammalian U1A snRNP protein. *Genes Dev* **8**, 914-25.
80. Gunderson, S.I., Beyer, K., Martin, G., Keller, W., Boelens, W.C., Mattaj, L.W. (1994) The human U1A snRNP protein regulates polyadenylation via a direct interaction with poly(A) polymerase. *Cell* **76**, 531-41.
81. Scherly, D., Boelens, W., Dathan, N.A., van Venrooij, W.J., Mattaj, I.W. (1990) Major determinants of the specificity of interaction between small

nuclear ribonucleoproteins U1A and U2B'' and their cognate RNAs. *Nature* **345**, 502-6.

82. Behrens, S.E., Tyc, K., Kastner, B., Reichelt, J., Luhrmann, R. (1993) Small nuclear ribonucleoprotein (RNP) U2 contains numerous additional proteins and has a bipartite RNP structure under splicing conditions. *Mol Cell Biol* **13**, 307-19.
83. Behrens, S.E., Luhrmann, R. (1991) Immunoaffinity purification of a [U4/U6.U5] tri-snRNP from human cells. *Genes Dev* **5**, 1439-52.
84. Will, C.L., Luhrmann, R. (1994) *RNA Processing.*

snRNP structure and function. Oxford University Press, Oxford, U.K.

85. Ruskin, B., Zamore, P.D., Green, M.R. (1988) A factor, U2AF, is required for U2 snRNP binding and splicing complex assembly. *Cell* **52**, 207-19.
86. Zamore, P.D., Green, M.R. (1991) Biochemical characterization of U2 snRNP auxiliary factor: an essential pre-mRNA splicing factor with a novel intranuclear distribution. *Embo J* **10**, 207-14.
87. Zuo, P., Maniatis, T. (1996) The splicing factor U2AF35 mediates critical protein-protein interactions in constitutive and enhancer-dependent splicing. *Genes Dev* **10**, 1356-1368.
88. Krainer, A.R., Conway, G.C., Kozak, D. (1990) Purification and characterization of pre-mRNA splicing factor SF2 from HeLa cells. *Genes Dev* **4**, 1158-71.
89. Manley, J.L., Tacke, R. (1996) SR proteins and splicing control. *Genes Dev* **10**, 1569-79.
90. Valcarcel, J., Green, M.R. (1996) The SR protein family: pleiotropic functions in pre-mRNA splicing. *Trends Biochem Sci* **21**, 296-301.
91. Zahler, A.M., Lane, W.S., Stolk, J.A., Roth, M.B. (1992) SR proteins: a conserved family of pre-mRNA splicing factors. *Genes Dev* **6**, 837-47.
92. Fu, X.D. (1993) Specific commitment of different pre-mRNAs to splicing by single SR proteins [see comments]. *Nature* **365**, 82-5.
93. Labourier, E., Rossi, F., Gallouzi, I., Allemand, E., Divita, G., Tazi, J. (1998) Interaction between the N-terminal domain of human DNA topoisomerase I and the arginine-serine domain of its substrate determines phosphorylation of SF2/ASF splicing factor. *Nucleic Acids Res* **26**, 2955-62.
94. Xiao, S.H., Manley, J.L. (1997) Phosphorylation of the ASF/SF2 RS domain affects both protein-protein and protein-RNA interactions and is necessary for splicing. *Genes Dev* **11**, 334-44.
95. Moore, M.J., Query, C.C., Sharp, P.A. (1993) Splicing of Precursors to mRNA by the Spliceosome. In *RNA World* (R. F. Gesteland. and J. F. Atkins., eds), CSH Press, New York 303_357.

96. Sharp, P.A. (1994) Split genes and RNA splicing. *Cell* **77**, 805-15.
97. Michaud, S., Reed, R. (1993) A functional association between the 5′ and 3′ splice site is established in the earliest prespliceosome complex (E) in mammals. *Genes Dev* **7**, 1008-20.
98. Jamison, S.F., Garcia-Blanco, M.A. (1992) An ATP-independent U2 small nuclear ribonucleoprotein particle/precursor mRNA complex requires both splice sites and the polypyrimidine tract. *Proc Natl Acad Sci U S A* **89**, 5482-6.
99. Zamore, P.D., Patton, J.G., Green, M.R. (1992) Cloning and domain structure of the mammalian splicing factor U2AF. *Nature* **355**, 609-14.
100. Valcarcel, J., Gaur, R.K., Singh, R., Green, M.R. (1996) Interaction of U2AF65 RS region with pre-mRNA branch point and promotion of base pairing with U2 snRNA [corrected] [published erratum appears in Science 1996 Oct 4;274(5284):21]. *Science* **273**, 1706-9.
101. Wu, J., Manley, J.L. (1989) Mammalian pre-mRNA branch site selection by U2 snRNP involves base pairing. *Genes Dev* **3**, 1553-61.
102. Sawa, H., Ohno, M., Sakamoto, H., Shimura, Y. (1988) Requirement of ATP in the second step of the pre-mRNA splicing reaction. *Nucleic Acids Res* **16**, 3157-64.
103. Brosi, R., Groning, K., Behrens, S.E., Luhrmann, R., Kramer, A. (1993) Interaction of mammalian splicing factor SF3a with U2 snRNP and relation of its 60-kD subunit to yeast PRP9. *Science* **262**, 102-5.
104. Madhani, H.D., Guthrie, C. (1992) A novel base-pairing interaction between U2 and U6 snRNAs suggests a mechanism for the catalytic activation of the spliceosome. *Cell* **71**, 803-17.
105. Lang, K.M., Spritz, R.A. (1983) RNA splice site selection: evidence for a 5′ leads to 3′ scanning model. *Science* **220**, 1351-5.
106. Goguel, V., Rosbash, M. (1993) Splice site choice and splicing efficiency are positively influenced by pre-mRNA intramolecular base pairing in yeast. *Cell* **72**, 893-901.
107. Chen, I.T., Chasin, L.A. (1994) Large exon size does not limit splicing in vivo. *Mol Cell Biol* **14**, 2140-6.
108. Hawkins, J.D. (1988) A survey of intron and exon length. *Nucleid Acid Research* **16**, 9893-9908.
109. Talerico, M., Berget, S.M. (1994) Intron definition in splicing of small Drosophila introns. *Mol Cell Biol* **14**, 3434-45.
110. Berget, S.M. (1995) Exon recognition in vertebrate splicing. *J Biol Chem* **270**, 2411-2414.
111. Robberson, B.L., Cote, G.J., Berget, S.M. (1990) Exon definition may facilitate splice site selection in RNAs with multiple exons. *Mol Cell Biol* **10**, 84-94.

112. Dominski, Z., Kole, R. (1991) Selection of splice sites in pre-mRNAs with short internal exons. *Mol Cell Biol* **11**, 6075-83.
113. 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.
114. Talerico, M., Berget, S.M. (1990) Effect of 5′ splice site mutations on splicing of the preceding intron. *Mol Cell Biol* **10**, 6299-305.
115. Kuo, H.C., Nasim, F.H., Grabowski, P.J. (1991) Control of alternative splicing by the differential binding of U1 small nuclear ribonucleoprotein particle. *Science* **251**, 1045-50.
116. Xu, R., Teng, J., Cooper, T.A. (1993) The cardiac troponin T alternative exon contains a novel purine-rich positive splicing element. *Mol Cell Biol* **13**, 3660-74.
117. Dominski, Z., Kole, R. (1994) Identification and characterization by antisense oligonucleotides of exon and intron sequences required for splicing. *Mol Cell Biol* **14**, 7445-54.
118. Carothers, A.M., Urlaub, G., Grunberger, D., Chasin, L.A. (1993) Splicing mutants and their second-site suppressors at the dihydrofolate reductase locus in Chinese hamster ovary cells. *Mol Cell Biol* **13**, 5085-98.
119. Hoffman, B.E., Grabowski, P.J. (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.
120. Izaurralde, E., Lewis, J., McGuigan, C., Jankowska, M., Darzynkiewicz, E., Mattaj, I.W. (1994) A nuclear cap binding protein complex involved in pre-mRNA splicing. *Cell* **78**, 657-68.
121. Zuo, P., Manley, J.L. (1994) The human splicing factor ASF/SF2 can specifically recognize pre-mRNA 5′ splice sites. *Proc Natl Acad Sci U S A* **91**, 3363-7.
122. Eperon, I.C., Ireland, D.C., Smith, R.A., Mayeda, A., Krainer, A.R. (1993) Pathways for selection of 5′ splice sites by U1 snRNPs and SF2/ASF. *Embo J* **12**, 3607-17.
123. Kohtz, J.D., Jamison, S.F., Will, C.L., Zuo, P., Luhrmann, R., Garcia-Blanco, M.A., Manley, J.L. (1994) Protein-protein interactions and 5′-splice-site recognition in mammalian mRNA precursors. *Nature* **368**, 119-24.
124. Staknis, D., Reed, R. (1994) Direct interactions between pre-mRNA and six U2 small nuclear ribonucleoproteins during spliceosome assembly. *Mol. Cell. Biol.* **14**, 2994-3005.

125. Wu, J.Y., Maniatis, T. (1993) Specific interactions between proteins implicated in splice site selection and regulated alternative splicing. *Cell* **75**, 1061-70.
126. Tronchere, H., Wang, J., Fu, X.D. (1997) A protein related to splicing factor U2AF35 that interacts with U2AF65 and SR proteins in splicing of pre-mRNA. *Nature* **388**, 397-400.
127. Chabot, B. (1996) Directing alternative splicing: cast and scenarios. *Trends Genet* **12**, 472-8.
128. Kennedy, C.F., Kramer, A., Berget, S.M. (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.
129. Blencowe, B.J., Issner, R., Nickerson, J.A., Sharp, P.A. (1998) A coactivator of pre-mRNA splicing. *Genes Dev* **12**, 996-1009.
130. Abovich, N., Rosbash, M. (1997) Cross-intron bridging interactions in the yeast commitment complex are conserved in mammals. *Cell* **89**, 403-12.
131. Boggs, R.T., Gregor, P., Idriss, S., Belote, J.M., McKeown, M. (1987) Regulation of sexual differentiation in D. melanogaster via alternative splicing of RNA from the transformer gene. *Cell* **50**, 739-47.
132. Cline, T.W., Meyer, B.J. (1996) Vive la difference: males vs females in flies vs worms. *Annu Rev Genet* **30**, 637-702.
133. Bell, L.R., Maine, E.M., Schedl, P., Cline, T.W. (1988) Sex-lethal, a Drosophila sex determination switch gene, exhibits sex- specific RNA splicing and sequence similarity to RNA binding proteins. *Cell* **55**, 1037-46.
134. Nakai, K., Sakamoto, H. (1994) Construction of a novel database containing aberrant splicing mutations of mammalian genes. *Gene* **141**, 171-7.
135. Sterner, D.A., Carlo, T., Berget, S.M. (1996) Architectural limits on split genes. *Proc Natl Acad Sci U S A* **93**, 15081-5.
136. Langford, C.J., Klinz, F.J., Donath, C., Gallwitz, D. (1984) Point mutations identify the conserved, intron-contained TACTAAC box as an essential splicing signal sequence in yeast. *Cell* **36**, 645-53.
137. Huh, G.S., Hynes, R.O. (1993) Elements regulating an alternatively spliced exon of the rat fibronectin gene. *Mol Cell Biol* **13**, 5301-5314.
138. Del Gatto, F., Breathnach, R. (1995) A Crouzon syndrome synonymous mutation activates a 5′ splice site within the IIIc exon of the FGFR2 gene. *Genomics* **27**, 558-9.
139. Muro, A.F., Iaconcig, A., Baralle, F.E. (1998) Regulation of the fibronectin EDA exon alternative splicing. Cooperative role of the exonic enhancer element and the 5′ splicing site. *FEBS Lett* **437**, 137-41.

140. Zhuang, Y.A., Goldstein, A.M., Weiner, A.M. (1989) UACUAAC is the preferred branch site for mammalian mRNA splicing. *Proc Natl Acad Sci U S A* **86**, 2752-6.
141. Krawczak, M., Reiss, J., Cooper, D.N. (1992) The mutational spectrum of single base-pair substitutions in mRNA splice junctions of human genes: causes and consequences. *Hum Genet* **90**, 41-54.
142. Roscigno, R.F., Weiner, M., Garcia-Blanco, M.A. (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.
143. Coolidge, C.J., Seely, R.J., Patton, J.G. (1997) Functional analysis of the polypyrimidine tract in pre-mRNA splicing. *Nucleic Acids Res* **25**, 888-96.
144. Shelley, C.S., Sharpe, C.R., Baralle, F.E., Shoulders, C.C. (1985) Comparison of the human apolipoprotein genes. Apo AII presents a unique functional intron-exon junction. *J Mol Biol* **186**, 43-51.
145. Reed, R., Maniatis, T. (1988) The role of the mammalian branchpoint sequence in pre-mRNA splicing. *Genes Dev* **2**, 1268-76.
146. Bouck, J., Litwin, S., Skalka, A.M., Katz, R.A. (1998) In vivo selection for intronic splicing signals from a randomized pool. *Nucleic Acids Res* **26**, 4516-23.
147. Smith, C.W., Nadal-Ginard, B. (1989) Mutually exclusive splicing of alpha-tropomyosin exons enforced by an unusual lariat branch point location: implications for constitutive splicing. *Cell* **56**, 749-58.
148. Nasim, F.H., Spears, P.A., Hoffmann, H.M., Kuo, H.C., Grabowski, P.J. (1990) A Sequential splicing mechanism promotes selection of an optimal exon by repositioning a downstream 5′ splice site in preprotachykinin pre- mRNA. *Genes Dev* **4**, 1172-84.
149. Tsukahara, T., Casciato, C., Helfman, D.M. (1994) Alternative splicing of beta-tropomyosin pre-mRNA: multiple cis- elements can contribute to the use of the 5′- and 3′-splice sites of the nonmuscle/smooth muscle exon 6. *Nucleic Acids Res* **22**, 2318-25.
150. Buvoli, M., Mayer, S.A., Patton, J.G. (1997) Functional crosstalk between exon enhancers, polypyrimidine tracts and branchpoint sequences. *Embo J* **16**, 7174-83.
151. Gattoni, R., Schmitt, P., Stevenin, J. (1988) In vitro splicing of adenovirus E1A transcripts: characterization of novel reactions and of multiple branch points abnormally far from the 3′ splice site. *Nucleic Acids Res* **16**, 2389-409.
152. Clouet d'Orval, B., d'Aubenton Carafa, Y., Sirand-Pugnet, P., Gallego, M., Brody, E., Marie, J. (1991) RNA secondary structure repression of a muscle-specific exon in HeLa cell nuclear extracts. *Science* **252**, 1823-8.

153. Hertel, K.J., Lynch, K.W., Hsiao, E.C., Liu, E.H., Maniatis, T. (1996) Structural and functional conservation of the Drosophila doublesex splicing enhancer repeat elements. *Rna* **2**, 969-81.
154. Muro, A.F., Caputi, M., Pariyarath, R., Pagani, F., Buratti, E., Baralle, F.E. (1999) Regulation of fibronectin EDA exon alternative splicing: possible role of RNA secondary structure for enhancer display. *Mol Cell Biol* **19**, 2657-71.
155. Tian, M., Maniatis, T. (1994) A splicing enhancer exhibits both constitutive and regulated activities. *Genes Dev* **8**, 1703-1712.
156. Kanopka, A., Muhlemann, O., Akusjarvi, G. (1996) Inhibition by SR proteins of splicing of a regulated adenovirus pre- mRNA. *Nature* **381**, 535-538.
157. Liu, H.X., Zhang, M., Krainer, A.R. (1998) Identification of functional exonic splicing enhancer motifs recognized by individual SR proteins. *Genes Dev* **12**, 1998-2012.
158. Lavigueur, A., La Branche, H., Kornblihtt, A.R., Chabot, B. (1993) A splicing enhancer in the human fibronectin alternate ED1 exon interacts with SR proteins and stimulates U2 snRNP binding. *Genes Dev* **7**, 2405-17.
159. Watakabe, A., Tanaka, K., Shimura, Y. (1993) The role of exon sequences in splice site selection. *Genes Dev* **7**, 407-418.
160. Dominski, Z., Kole, R. (1994) Identification of exon sequences involved in splice site selection. *J Biol Chem* **269**, 23590-6.
161. Coulter, L.R., Landree, M.A., Cooper, T.A. (1997) Identification of a new class of exonic splicing enhancers by in vivo selection [published erratum appears in Mol Cell Biol 1997 Jun;17(6):3468]. *Mol Cell Biol* **17**, 2143-50.
162. Hertel, K.J., Lynch, K.W., Maniatis, T. (1997) Common themes in the function of transcription and splicing enhancers. *Curr Opin Cell Biol* **9**, 350-7.
163. Zhang, L., Ashiya, M., Sherman, T.G., Grabowski, P.J. (1996) Essential nucleotides direct neuron-specific splicing of gamma 2 pre- mRNA. *Rna* **2**, 682-98.
164. Ryan, K.J., Cooper, T.A. (1996) Muscle-specific splicing enhancers regulate inclusion of the cardiac troponin T alternative exon in embryonic skeletal muscle. *Mol Cell Biol* **16**, 4014-23.
165. Cook, C.R., McNally, M.T. (1999) Interaction between the negative regulator of splicing element and a 3′ splice site: requirement for U1 small nuclear ribonucleoprotein and the 3′ splice site branch point/pyrimidine tract. *J Virol* **73**, 2394-400.

166. McNally, M.T., Gontarek, R.R., Beemon, K. (1991) Characterization of Rous sarcoma virus intronic sequences that negatively regulate splicing. *Virology* **185**, 99-108.
167. Caputi, M., Casari, G., Guenzi, S., Tagliabue, R., Sidoli, A., Melo, C.A., Baralle, F.E. (1994) A novel bipartite splicing enhancer modulates the differential processing of the human fibronectin EDA exon. *Nucleic Acids Res* **22**, 1018-1022.
168. Staffa, A., Acheson, N.H., Cochrane, A. (1997) Novel exonic elements that modulate splicing of the human fibronectin EDA exon. *J Biol Chem* **272**, 33394-401.
169. Gooding, C., Roberts, G.C., Smith, C.W. (1998) Role of an inhibitory pyrimidine element and polypyrimidine tract binding protein in repression of a regulated alpha-tropomyosin exon. *Rna* **4**, 85-100.
170. Chabot, B., Blanchette, M., Lapierre, I., La Branche, H. (1997) An intron element modulating 5′ splice site selection in the hnRNP A1 pre-mRNA interacts with hnRNP A1. *Mol Cell Biol* **17**, 1776-86.
171. Krainer, A.R., Conway, G.C., Kozak, D. (1990) The essential pre-mRNA splicing factor SF2 influences 5′ splice site selection by activating proximal sites. *Cell* **62**, 35-42.
172. Reed, R. (1996) Initial splice-site recognition and pairing during pre-mRNA splicing. *Curr Opin Genet Dev* **6**, 215-220.
173. Weighardt, F., Biamonti, G., Riva, S. (1996) The roles of heterogeneous nuclear ribonucleoproteins (hnRNP) in RNA metabolism. *Bioessays* **18**, 747-56.
174. Lamm, G.M., Lamond, A.I. (1993) Non-snRNP protein splicing factors. *Biochim Biophys Acta* **1173**, 247-65.
175. Krecic, A.M., Swanson, M.S. (1999) hnRNP complexes: composition, structure, and function. *Curr Opin Cell Biol* **11**, 363-71.
176. Mayeda, A., Krainer, A.R. (1992) Regulation of alternative pre-mRNA splicing by hnRNP A1 and splicing factor SF2. *Cell* **68**, 365-75.
177. Fu, X.D., Mayeda, A., Maniatis, T., Krainer, A.R. (1992) General splicing factors SF2 and SC35 have equivalent activities in vitro, and both affect alternative 5′ and 3′ splice site selection. *Proc Natl Acad Sci U S A* **89**, 11224-8.
178. Bai, Y., Lee, D., Yu, T., Chasin, L.A. (1999) Control of 3′ splice site choice in vivo by ASF/SF2 and hnRNP A1. *Nucleic Acids Res* **27**, 1126-34.
179. Swanson, M.S., Dreyfuss, G. (1988) RNA binding specificity of hnRNP proteins: a subset bind to the 3′ end of introns. *Embo J* **7**, 3519-29.

180. Garcia-Blanco, M.A., Jamison, S.F., Sharp, P.A. (1989) Identification and purification of a 62,000-dalton protein that binds specifically to the polypyrimidine tract of introns. *Genes Dev* **3**, 1874-86.
181. Gil, A., Sharp, P.A., Jamison, S.F., Garcia-Blanco, M.A. (1991) Characterization of cDNAs encoding the polypyrimidine tract-binding protein. *Genes Dev* **5**, 1224-36.
182. Mullen, M.P., Smith, C.W., Patton, J.G., Nadal-Ginard, B. (1991) Alpha-tropomyosin mutually exclusive exon selection: competition between branchpoint/polypyrimidine tracts determines default exon choice. *Genes Dev* **5**, 642-55.
183. Mulligan, G.J., Guo, W., Wormsley, S., Helfman, D.M. (1992) Polypyrimidine tract binding protein interacts with sequences involved in alternative splicing of beta-tropomyosin pre-mRNA. *J Biol Chem* **267**, 25480-7.
184. Tuerk, C., Gold, L. (1990) Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science* **249**, 505-10.
185. Ashiya, M., Grabowski, P.J. (1997) A neuron-specific splicing switch mediated by an array of pre-mRNA repressor sites: evidence of a regulatory role for the polypyrimidine tract binding protein and a brain-specific PTB counterpart. *Rna* **3**, 996-1015.
186. Lou, H., Helfman, D.M., Gagel, R.F., Berget, S.M. (1999) Polypyrimidine tract-binding protein positively regulates inclusion of an alternative 3'-terminal exon. *Mol Cell Biol* **19**, 78-85.
187. Chen, C.D., Kobayashi, R., Helfman, D.M. (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.
188. Min, H., Chan, R.C., Black, D.L. (1995) The generally expressed hnRNP F is involved in a neural-specific pre- mRNA splicing event. *Genes Dev* **9**, 2659-71.
189. Chou, M.Y., Rooke, N., Turck, C.W., Black, D.L. (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.
190. Swanson, M.S., Dreyfuss, G. (1988) Classification and purification of proteins of heterogeneous nuclear ribonucleoprotein particles by RNA-binding specificities. *Mol Cell Biol* **8**, 2237-41.
191. Wilusz, J., Feig, D.I., Shenk, T. (1988) The C proteins of heterogeneous nuclear ribonucleoprotein complexes interact with RNA sequences downstream of polyadenylation cleavage sites. *Mol Cell Biol* **8**, 4477-83.

192. Wilusz, J., Shenk, T. (1990) A uridylate tract mediates efficient heterogeneous nuclear ribonucleoprotein C protein-RNA cross-linking and functionally substitutes for the downstream element of the polyadenylation signal. *Mol Cell Biol* **10**, 6397-407.
193. Wang, J., Manley, J.L. (1997) Regulation of pre-mRNA splicing in metazoa. *Curr Opin Genet Dev* **7**, 205-11.
194. Hanamura, A., Caceres, J.F., Mayeda, A., Franza, B.R., Jr., Krainer, A.R. (1998) Regulated tissue-specific expression of antagonistic pre-mRNA splicing factors. *Rna* **4**, 430-44.
195. Gallego, M.E., Gattoni, R., Stevenin, J., Marie, J., Expert-Bezancon, A. (1997) The SR splicing factors ASF/SF2 and SC35 have antagonistic effects on intronic enhancer-dependent splicing of the beta-tropomyosin alternative exon 6A. *Embo J* **16**, 1772-84.
196. Gallego, M.E., Sirand-Pugnet, P., Durosay, P., Clouet d'Orval, B., d'Aubenton-Carafa, Y., Brody, E., Expert-Bezancon, A., Marie, J. (1996) Tissue-specific splicing of two mutually exclusive exons of the chicken beta-tropomyosin pre-mRNA: positive and negative regulations. *Biochimie* **78**, 457-465.
197. Biamonti, G., Bassi, M.T., Cartegni, L., Mechta, F., Buvoli, M., Cobianchi, F., Riva, S. (1993) Human hnRNP protein A1 gene expression. Structural and functional characterization of the promoter. *J Mol Biol* **230**, 77-89.
198. Caceres, J.F., Misteli, T., Screaton, G.R., Spector, D.L., Krainer, A.R. (1997) Role of the modular domains of SR proteins in subnuclear localization and alternative splicing specificity. *J Cell Biol* **138**, 225-38.
199. Mayeda, A., Munroe, S.H., Caceres, J.F., Krainer, A.R. (1994) Function of conserved domains of hnRNP A1 and other hnRNP A/B proteins. *Embo J* **13**, 5483-95.
200. Xiao, S.H., Manley, J.L. (1998) Phosphorylation-dephosphorylation differentially affects activities of splicing factor ASF/SF2. *Embo J* **17**, 6359-67.
201. Lopez, A.J. (1998) Alternative splicing of pre-mRNA: developmental consequences and mechanisms of regulation. *Annu Rev Genet* **32**, 279-305.
202. Valcarcel, J., Singh, R., Zamore, P.D., Green, M.R. (1993) The protein Sex-lethal antagonizes the splicing factor U2AF to regulate alternative splicing of transformer pre-mRNA. *Nature* **362**, 171-5.
203. Burtis, K.C., Baker, B.S. (1989) Drosophila doublesex gene controls somatic sexual differentiation by producing alternatively spliced mRNAs encoding related sex-specific polypeptides. *Cell* **56**, 997-1010.

204. Hedley, M.L., Maniatis, T. (1991) Sex-specific splicing and polyadenylation of dsx pre-mRNA requires a sequence that binds specifically to tra-2 protein in vitro. *Cell* **65**, 579-86.
205. Tian, M., Maniatis, T. (1993) A splicing enhancer complex controls alternative splicing of doublesex pre-mRNA. *Cell* **74**, 105-114.
206. Lynch, K.W., Maniatis, T. (1995) Synergistic interactions between two distinct elements of a regulated splicing enhancer. *Genes Dev* **9**, 284-293.
207. Tian, M., Maniatis, T. (1992) Positive control of pre-mRNA splicing in vitro. *Science* **256**, 237-40.
208. Amrein, H., Hedley, M.L., Maniatis, T. (1994) The role of specific protein-RNA and protein-protein interactions in positive and negative control of pre-mRNA splicing by Transformer 2. *Cell* **76**, 735-46.
209. Burtis, K.C., Coschigano, K.T., Baker, B.S., Wensink, P.C. (1991) The doublesex proteins of Drosophila melanogaster bind directly to a sex-specific yolk protein gene enhancer. *Embo J* **10**, 2577-82.
210. Kelley, R.L., Wang, J., Bell, L., Kuroda, M.I. (1997) Sex lethal controls dosage compensation in Drosophila by a non-splicing mechanism. *Nature* **387**, 195-9.
211. Wine, J.J., Glavac, D., Hurlock, G., Robinson, C., Lee, M., Potocnik, U., Ravnik-Glavac, M., Dean, M. (1998) Genomic DNA sequence of Rhesus (M. mulatta) cystic fibrosis (CFTR) gene. *Mamm Genome* **9**, 301-5.
212. Tebbutt, S.J., Wardle, C.J., Hill, D.F., Harris, A. (1995) Molecular analysis of the ovine cystic fibrosis transmembrane conductance regulator gene. *Proc Natl Acad Sci U S A* **92**, 2293-7.
213. Fu, X.D. (1995) The superfamily of arginine/serine-rich splicing factors. *Rna* **1**, 663-80.
214. Will, C.L., Luhrmann, R. (1997) Protein functions in pre-mRNA splicing. *Curr Opin Cell Biol* **9**, 320-8.
215. Ou, S.H., Wu, F., Harrich, D., Garcia-Martinez, L.F., Gaynor, R.B. (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.
216. Ramchatesingh, J., Zahler, A.M., Neugebauer, K.M., Roth, M.B., Cooper, T.A. (1995) A subset of SR proteins activates splicing of the cardiac troponin T alternative exon by direct interactions with an exonic enhancer. *Mol Cell Biol* **15**, 4898-907.
217. Shelley, C.S., Baralle, F.E. (1987) Deletion analysis of a unique 3′ splice site indicates that alternating guanine and thymine residues represent an efficient splicing signal. *Nucleic Acids Res* **15**, 3787-99.

218. Norton, P.A. (1994) Polypyrimidine tract sequences direct selection of alternative branch sites and influence protein binding. *Nucleic Acids Res* **22**, 3854-60.
219. Wieringa, B., Meyer, F., Reiser, J., Weissmann, C. (1983) Unusual splice sites revealed by mutagenic inactivation of an authentic splice site of the rabbit beta-globin gene. *Nature* **301**, 38-43.
220. Wieringa, B., Hofer, E., Weissmann, C. (1984) A minimal intron length but no specific internal sequence is required for splicing the large rabbit beta-globin intron. *Cell* **37**, 915-25.
221. Burd, C.G., Dreyfuss, G. (1994) RNA binding specificity of hnRNP A1: significance of hnRNP A1 high- affinity binding sites in pre-mRNA splicing. *Embo J* **13**, 1197-204.
222. Varani, G., Nagai, K. (1998) RNA recognition by RNP proteins during RNA processing. *Annu Rev Biophys Biomol Struct* **27**, 407-45.
223. Kenan, D.J., Query, C.C., Keene, J.D. (1991) RNA recognition: towards identifying determinants of specificity. *Trends Biochem Sci* **16**, 214-20.
224. McNally, L.M., McNally, M.T. (1996) SR protein splicing factors interact with the Rous sarcoma virus negative regulator of splicing element. *J Virol* **70**, 1163-72.
225. Wang, J., Xiao, S.H., Manley, J.L. (1998) Genetic analysis of the SR protein ASF/SF2: interchangeability of RS domains and negative control of splicing. *Genes Dev* **12**, 2222-33.
226. Kazazian, H.H., Jr. (1998) Mobile elements and disease. *Curr Opin Genet Dev* **8**, 343-50.
227. Kazazian, H.H., Jr., Moran, J.V. (1998) The impact of L1 retrotransposons on the human genome. *Nat Genet* **19**, 19-24.
228. Rozmahel, R., Heng, H.H., Duncan, A.M., Shi, X.M., Rommens, J.M., Tsui, L.C. (1997) Amplification of CFTR exon 9 sequences to multiple locations in the human genome. *Genomics* **45**, 554-61.
229. Blanchette, M., Chabot, B. (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.
230. Mak, V., Jarvi, K.A., Zielenski, J., Durie, P., Tsui, L.C. (1997) Higher proportion of intact exon 9 CFTR mRNA in nasal epithelium compared with vas deferens. *Hum Mol Genet* **6**, 2099-107.
231. Larriba, S., Bassas, L., Gimenez, J., Ramos, M.D., Segura, A., Nunes, V., Estivill, X., Casals, T. (1998) Testicular CFTR splice variants in patients with congenital absence of the vas deferens. *Hum Mol Genet* **7**, 1739-43.

232. Chu, C.S., Trapnell, B.C., Murtagh, J.J., Jr., Moss, J., Dalemans, W., Jallat, S., Mercenier, A., Pavirani, A., Lecocq, J.P., Cutting, G.R., 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.
233. Sambrook, J., Fritsch, E.F., Maniatis, T. (1989) *Molecular Cloning - A Laboratory Manual*. Cold Spring Harbor Laboratory Press.
234. Chomczynski, P., Sacchi, N. (1987) Single-step method of RNA isolation by acid guanidinium thiocyanate- phenol-chloroform extraction. *Anal Biochem* **162**, 156-9.
235. McPherson, M.J., Quirke, P., Taylor, G.R. (1991) *PCR. A practical approach*, New York.
236. Grosveld, G.C., de Boer, E., Shewmaker, C.K., Flavell, R.A. (1982) DNA sequences necessary for transcription of the rabbit beta-globin gene in vivo. *Nature* **295**, 120-6.
237. Zerivitz, K., Akusjarvi, G. (1989) An improved nuclear extract preparation method. *Gene Anal Tech* **6**, 101-9.