Compensatory Relationship Between Exonic Splicing Enhancer, Splice Site and Protein Function

Thesis

Compensatory Relationship Between Exonic Splicing Enhancer, Splice Site and Protein Function

Alessia Falanga

A Thesis Submitted in Fulfillment of the Requirements for the Degree of Doctor of Philosophy in Life Sciences

International Centre for Genetic Engineering and Biotechnology,
ICGEB Trieste, Italy
The Open University, UK

Director of Studies: Prof. Francisco E. Baralle, M.D. Ph.D.
External Supervisor: Dr Colin Sharpe, MA DPhil.

October 2012

Date of Submission: 22 August 2012
Date of Award: 25 October 2012

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 – 1346

Table of Contents

LIST OF FIGURES AND TABLES

ABBREVIATIONS

The standard abbreviations used in this dissertation follow IUPAC rules. All the abbreviations are defined also in the text when they are introduced for the first time. The abbreviations mentioned only once are not included in this list.

aa	Amino acid
ALPL	Tissue-non specific alkaline Phosphatase gene
ALPP	Placental alkaline Phosphatase gene
Bp	Base pairs
cDNA	Complementary DNA
ddH20	Double-distilled water
DNA	Deoxyribonucleic acid
dNTPs	Deoxynucleoside triphosphate (A, C, G and T)
DTT	Dithiothreitol
EAWS	ESE Analyzer Web Server
EDTA	Ethylenediamine tetra-acetic acid
ESE	Exonic Splicing Enhancer
ESS	Exonic Splicing Silencer
hnRNP	Heterogenous ribonuclear protein
IPTG	Isopropyl-β-d-thiogalactopyranoside
ISE	Intronic Splicing Enhancer
ISS	Intronic Splicing Silencer
kb	Kilobase
Kcat	catalytic rate constant
kDa	Kilodalton
Km	Michaelis constant
mRNA	Messenger ribonucleic acid
N	Nucleotide (A or C or G or T)
nt	Nucleotides
PBS	Phosphate buffer saline
PLAP	Placental alkaline Phosphatase protein
pNPP	p-Nitrophenylphosphate
RNA	Ribonucleic acid
RT-PCR	Reverse transcriptase polymerase chain reaction

snRNP	Small nuclear ribonucleoprotein particles
SR	Arginine-serine rich protein
ss	Splice site
TBE	Tris-borate-EDTA (buffer)
TNAP	Tissue-non specific alkaline Phosphatase protein

ABSTRACT

The process of pre-mRNA splicing involves the removal of intronic sequences from the pre-mRNA and it is directed by intronic cis acting elements know as the 5' and 3' splice sites that mark the boundaries of the exons. Over the two decades, however, it has become clear that exons encode for auxiliary splicing signals that either enhance or perturb their inclusion in the final mRNA product. It is possible that the evolution of mRNA sequences could be conditioned by the presence of these exonic cis-acting splicing regulatory elements and not mainly by the selection of optimal protein function.

To explore this hypothesis, I have investigated how the need for ESE influences the gene evolution of a paralogous gene family, specifically the human Alkaline Phosphatases (ALPs). In this work, I have identified in correspondence to a weak 3'splice site, two ESE sequences in the placental ALP exon 4, and demonstrate that the ESE are necessary for the exon inclusion in the mRNA due to the weak 3'splice sites. Furthermore, I show that they are absent in the corresponding exon of the non-tissue specific ALP transcript, specifically exon 5 that carries a strong 3' splice site. Most importantly, the localization of the ESEs correspond to an area that in the paralogous non-tissue specific ALP gene differs in amino acid composition with respect, not only to the placental ALP where I mapped the ESEs but also to the other members of the family, where this area is well conserved. These amino acid changes may represent a possible evolutionary constraint on enzymatic activity, in keeping with this hypothesis, substituting the amino acids in the region of the ESE for those of the paralogous non-tissue specific ALP gene increases the enzymatic activity. Thus splicing-related constraints challenge the primacy of biochemical function in rates of protein evolution.

1. INTRODUCTION

1.1 Overview of splicing

The initial primary transcript, the precursor-messenger RNA (pre-mRNA), in eukaryotes is synthesized by RNA polymerase II (Pol II), and before leaving the nucleus undergoes several modifications that can occur co- and post- transcriptionally. These tightly regulated steps begin in the nucleus with the chromatin decompaction that facilitates the recruitment of the Pol II to the transcriptional start site. Soon after Pol II initiates transcription, the nascent RNA is modified by the addition of a 7-methyl guanosine cap structure at its 5' end that occurs co-transcriptionally, within only about 20-30 nucleotides (nt) of transcription. This cap serves initially to protect the new transcript from attack by nucleases and later serves as a binding site for proteins involved in export of the mature mRNA into the cytoplasm and its translation into protein (Lewis and Izaurralde, 1997). Coding sequences in the gene (exons) are interrupted by non-coding sequences (introns), which are removed by pre-mRNA splicing to generate mature transcripts (Gilbert, 1978). Over the past decade direct evidences have been shown that pre-mRNA can be spliced during its synthesis (Fong and Zhou, 2001; Kornblihtt et al., 2004). A multi-protein complex machine called the spliceosome carries out this reaction (Staley and Guthrie, 1998). Upon reaching the end of a gene, Pol II stops transcription and the newly synthesized RNA is cleaved and a poly(A) tail is added to the 3'end of the transcript ("polyadenylation") (Proudfoot et al., 2002). The processes by which information is transferred from DNA to RNA and from RNA to protein are physically separated in eukaryotes by the nuclear membrane. Therefore, processed mRNAs must be transported from the nucleus to the cytoplasm before translation can occur (Fig.1.1).

Thus, eukaryotic gene expression requires a careful orchestration of all the steps involved and some mechanisms are still unknown. Indeed, one of the fundamental issues in RNA splicing research is represented by understanding how the spliceosome can successfully define exons and introns in an ocean of similar sequences. Since its first description, researchers in this field have identified and characterized many fundamental elements and players capable of affecting the splicing process, both in a negative and positive manner. These elements range form the basic sequences that define the exon/intron boundaries to nucleosome positioning and epigenetic factors (de Almeida and Carmo-Fonseca, 2012).

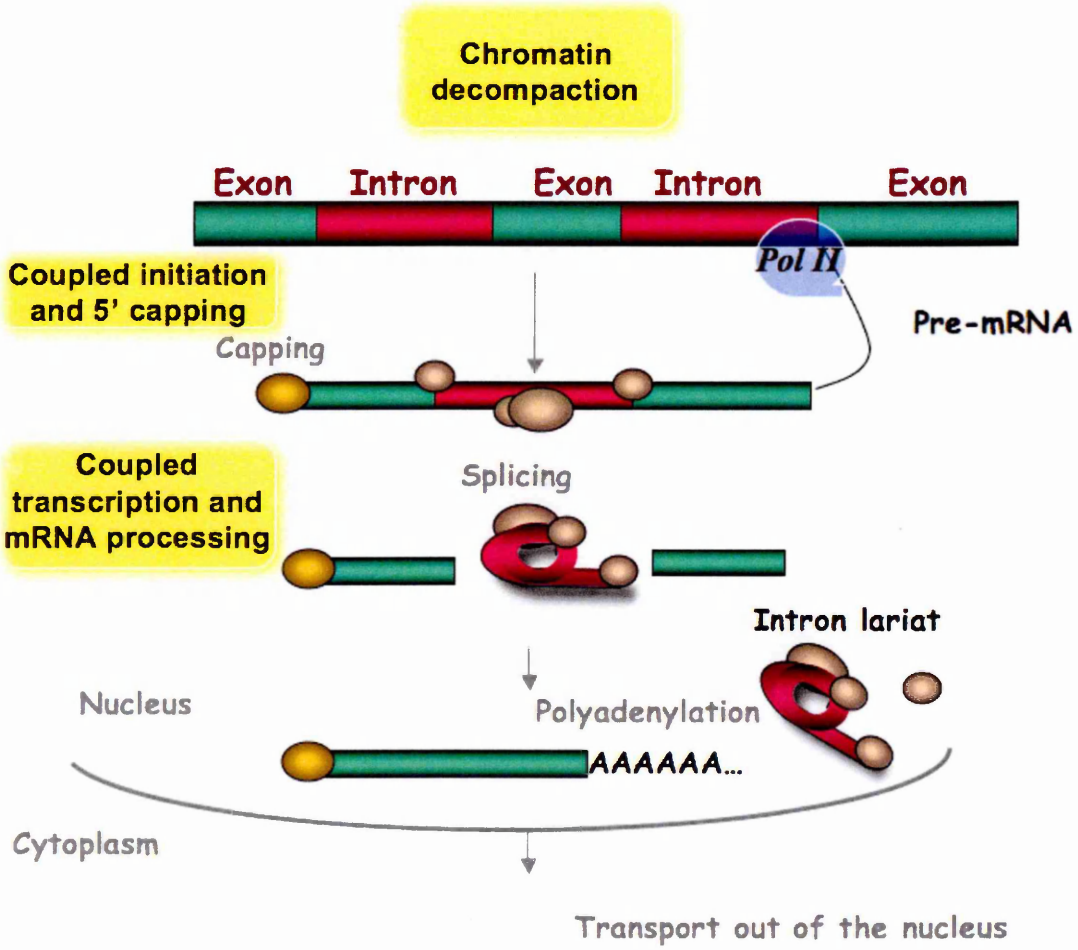


Figure 1.1 Schematic representation of the complex network during mRNA processing. The diagram illustrates as each step of gene expression is coupled one to another, from the chromatin decompaction that facilitate the recruitment of the Pol II enzymes to the transcriptional start site. Soon after Pol II initiates transcription, the nascent RNA is modified by the addition of a 7-methyl guanosine cap structure at its 5' end. Coding sequences in the gene (exons) are often interrupted by non-coding sequences (introns), which are removed by pre-mRNA splicing to generate mature transcripts. Upon reaching the end of a gene, Pol II stops transcription and the newly synthesized RNA is cleaved and a poly(A) tail is added to the 3'end of the transcript ("polyadenylation"). Once an mRNA has been fully processed, it must be transported to the site of protein translation in the cytoplasm. Figure modified from Orphanides et al. (2002).

1.2 Spliceosome complex formation

As mentioned the actual process of splicing is carried out by the spliceosome, one of the most complex macromolecular machines in the cell, characterized by highly dynamic associations and dissociations of several particles (Will and Luhrmann, 2011). The essential components have being characterised as five small ribonuclear proteins snRNPs U1, U2, U4, U5 and U6 that function in conjunction with over 200 non sn-RNP auxiliary proteins (de Almeida and Carmo-Fonseca, 2008; Hartmuth et al., 2002; Jurica et al., 2002). The assembly of the spliceosome has been characterized principally using in vitro systems where several distinct intermediates in an assembly pathway can be observed (Query et al., 1995) (Fig. 1.2).

The basic steps of spliceosomal components interaction with pre-mRNA starts with U1 snRNP that base-pairs with the 5' splice site (ss) and non-snRNP factors such as U2 auxiliary factor (U2AF) which is a heterodimer that includes a 65 kDa subunit (U2AF65), a protein containing tandem RNA recognition motifs that binds tightly to the polypyrimidine tract, and a 35 kDa subunit (U2AF35) which binds the actual 3' ss (Brow, 2002) and SF1/BBP (splicing factor 1/branch binding protein) that recognizes the branch point sequence. These factors, together with additional proteins, form a first discrete functional spliceosome complex called the E or commitment complex. The E complex bridges the intron and plays a crucial role in the initial recognition of the splice sites to be cleaved together. In a subsequent step, U2AF recruits the U2 snRNP, and an ATP dependent step allows the RNA portion of the U2 snRNP to base pair with a branch point and SF1/mBBP is displaced. This base pairing of the U2 snRNP with the branch point completes the A complex. After, ATP dependent recruitment of U4-U6-U5 tri-snRNPs to the 5'ss the B complex is formed. After the release of U1 and U4 and compositional rearrangments the complex is able to catalyze the first of the two transesterification steps of splicing described

below. After conformational rearrangements the C complex is formed, which catalyzes the second transesterification reactions on the splice sites (Konarska et al., 2006).

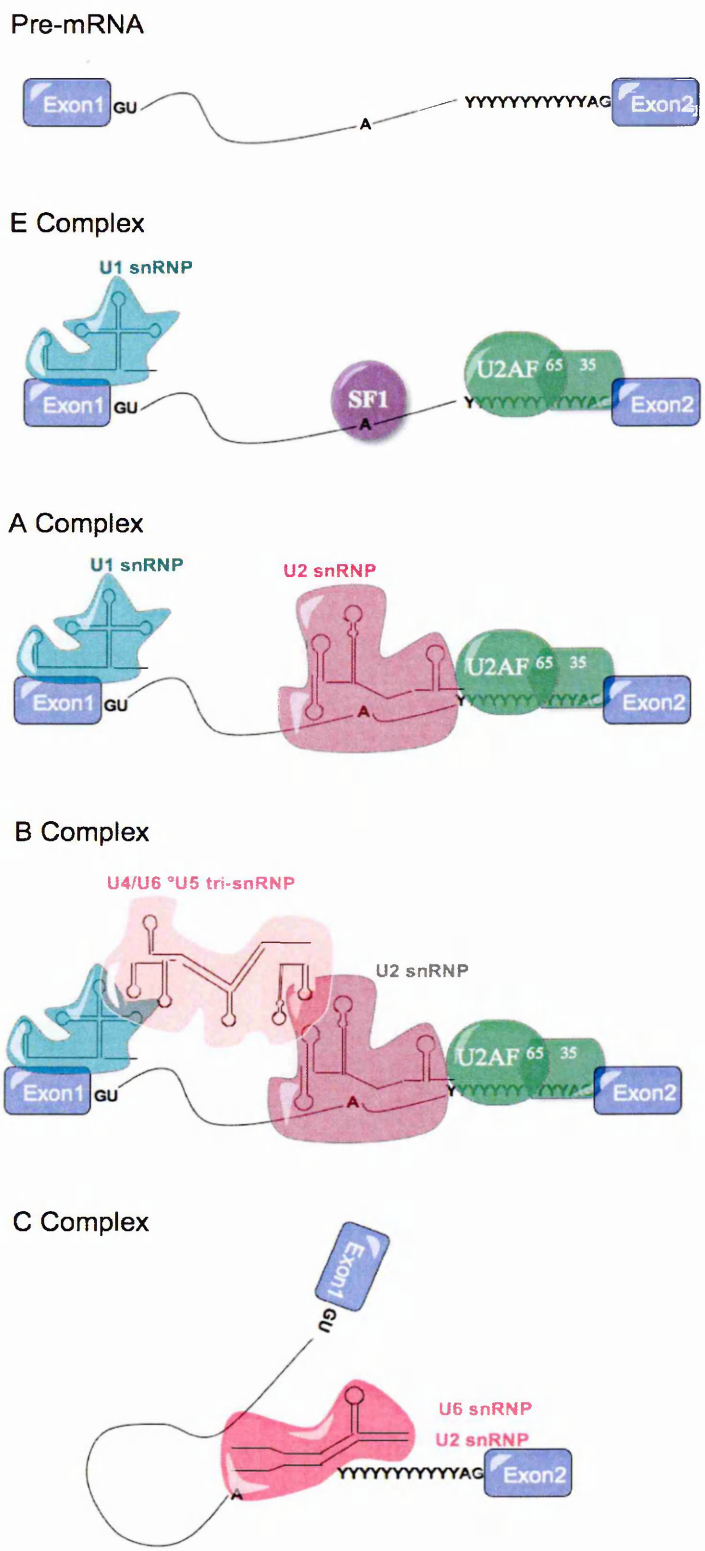


Figure 1.2. Spliceosome assembly. The spliceosome assembles onto the pre-mRNA in a stepwise manner. The E complex contains U1 snRNP bound to the 5' splice site, SF1 bound to the branch point, and U2AF65 and U2AF35 bound to the pyrimidine tract and 3' splice site AG, respectively. In the A complex, SF1 is replaced by U2 snRNP at the branch point. The U4/U6/U5 tri-snRNP then enters to form the B complex. Finally, a rearrangement occurs to form the catalytically active C complex, in which U2 and U6 interact, and U6 replaces U1 at the 5' splice site. Figure modified from Hertel et al. (2005)

During the actual transesterification reactions, in the first step the 2'-hydroxil group of the adenosine at the branch site carries out a nucleophilic attack on the 5'ss, breaking the phosphodiester bond and forming the 2'-5' linkage between the branch site and the 5'-terminal nucleotide of the intron. This reaction generates the splicing intermediates (free exon 1 and lariat-exon 2). In the second step the 3' hydroxyl group of the exon 1 attacks at the 3' splice site creating a new phosphodiester bond between exon 1 and exon 2 and a free intron released in form of "lariat" (Lamond, 1993; Moore and Sharp, 1993) (Fig. 1.3). After this second step, the mRNA is released, the snRNPs dissociate and than can take part in the next splicing reaction.

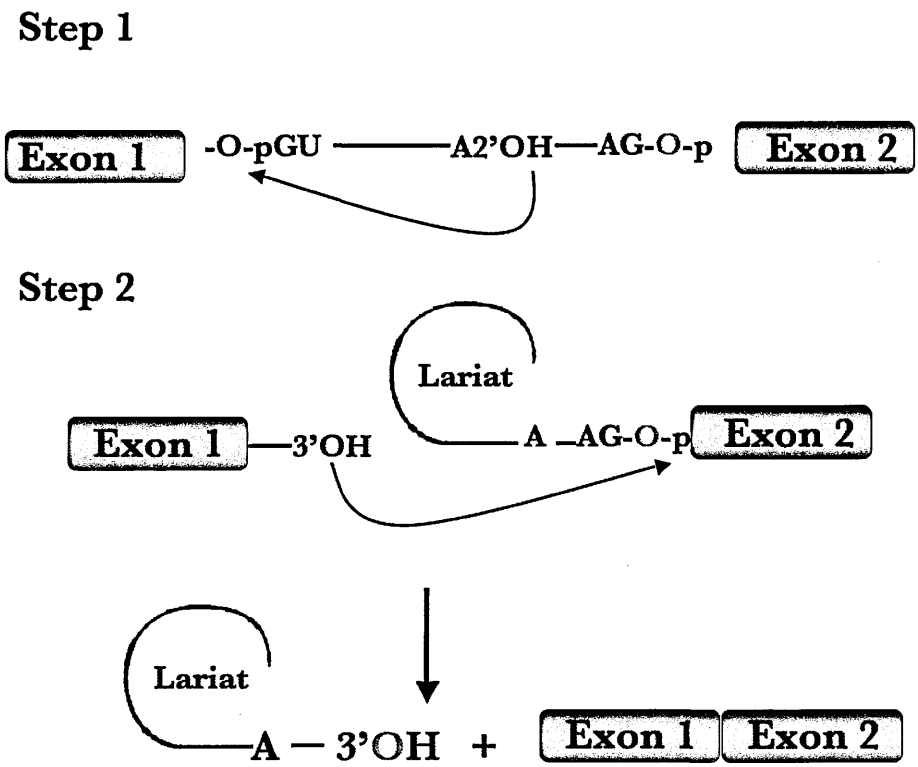


Figure 1.3. Schematic representation of sequential transesterification during splicing. In the first reaction, the ester bond between the 5' phosphorous of the intron and the 3' oxygen (red) of exon 1 is exchanged for an ester bond with the 2' oxygen (dark blue) of the branch-site A residue. In the second reaction, the ester bond between the 5' phosphorous of exon 2 and the 3' oxygen (light blue) of the intron is exchanged for an ester bond with the 3' oxygen of exon 1, releasing the intron as a lariat structure and joining the two exons. Arrows show where the activated hydroxyl oxygens react with phosphorous atoms.

1.3 Alternative splicing: where a common precursor pre- mRNA molecule can generate multiple mRNAs

As previously explained, the splicing reaction occurs in two transesterification steps within a large spliceosome complex. Splicing can be either *constitutive* (when the exon in question always forms part of the mRNA) or *alternative*. During alternative splicing, the spliceosome assembly is altered so that a splice site is optionally used depending on the cell type, developmental stage or sex, resulting in the inclusion or exclusion of alternative exon sequences in the mature mRNA. Alternative splicing is therefore a process by which the exons of the RNA produced by transcription are reconnected in multiple different ways increasing greatly protein diversity.

Transcripts from a gene can undergo many different patterns of alternative splicing. Typical possibilities of alternative splicing are inclusion or skipping of one or more exons (cassette exons), shortening or lengthening of an exon by alternative 5' and 3' splice site usages, mutual exclusion of two or more exons, and retained introns. Different promoters and different polyadenylation sites may specify alternative 5' and 3' terminal exons, respectively. Combinations of different basic types can form more complex alternative splicing patterns (Fig. 1.4).

Estimates of how commonly alternative splicing occurs in human protein coding genes have increased over the years, from an initial 5 % to more than 95 % (Calarco et al., 2011; Sharp, 1994). Recent studies have suggested that alternative splicing is nearly ubiquitous in human transcripts and is frequently controlled in a tissue specific manner (Pan et al., 2004; Wang et al., 2008). The consequences of alternative splicing include altered mRNA stability or subcellular localization and the addition or deletion of specific protein coding sequences. These mechanisms of alternative splicing are so tightly controlled that even subtle defects in alternative splicing factors or aberrant inclusion of alternative exon can result in genetic diseases (Faustino and Cooper, 2003; Kashima and Manley, 2003; Licatalosi and Darnell, 2006).

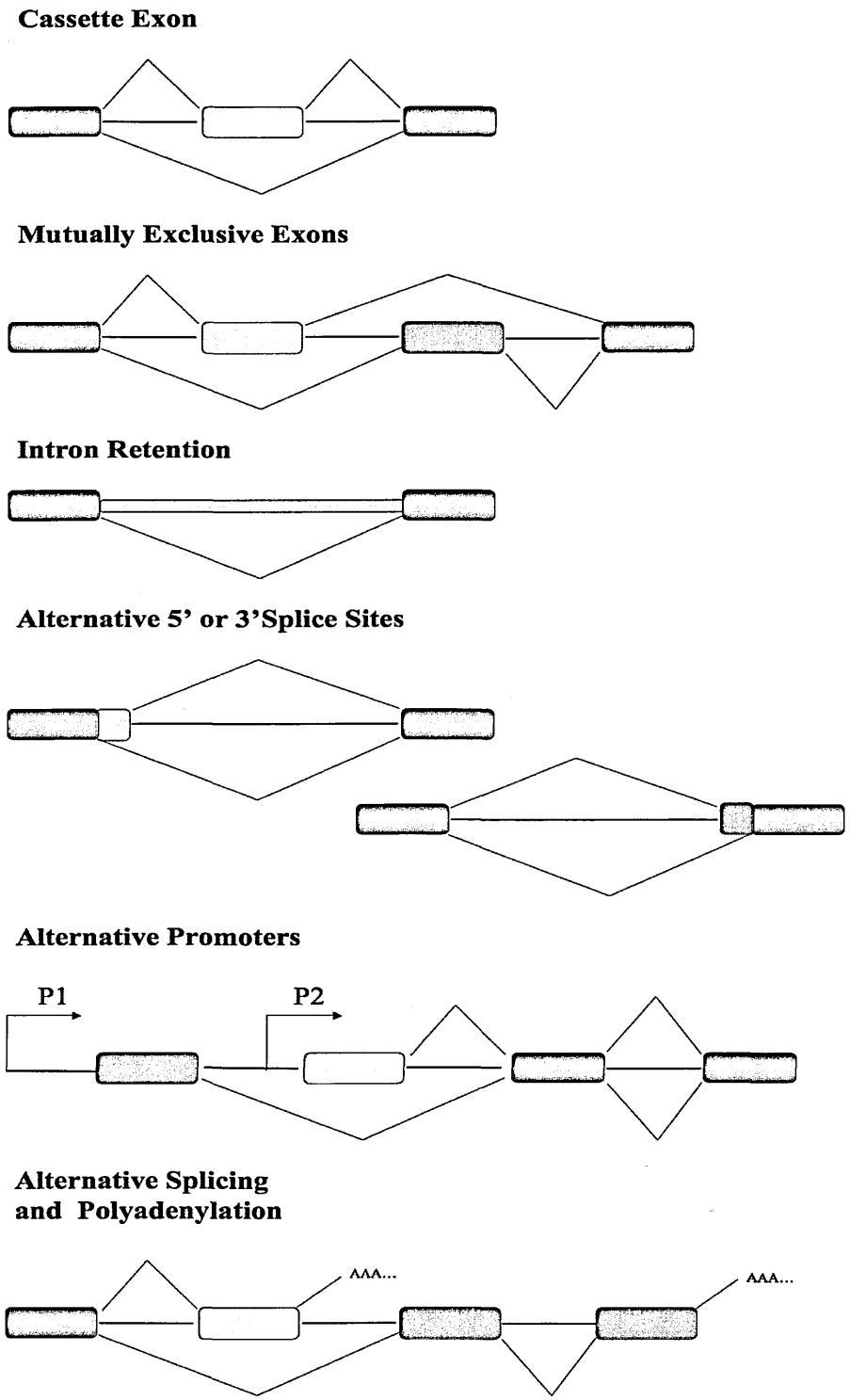


Figure 1.4. Modes of alternative splicing. Typical types of alternative splicing are inclusion or skipping of one or more exons (cassette exons), mutual exclusion of two or more exons, retained introns and shortening or lengthening of an exon by alternative 5' and 3' splice site. Different promoters (P) and different polyadenylation sites may specify alternative 5' and 3' terminal exons, respectively. In these graphics, exons are represented by boxes and introns by lines. Exon regions included in the messages by alternative splicing are colored while constitutive exons are shown in green. Promoters are indicated with arrows and polyadenylation sites with AAAA.

1.4 Donor and Acceptor splice site sequence

Discrimination between exon and intron sequences is a complex task for the splicing machinery. In general, it is directed by the presence of specific consensus sequences at the exon-intron junctions (Fig. 1.5).

The 5' ss motif (donor site) marks the exon/intron junction at the 5' end of the intron and in higher eukaryotes consists of nine partially conserved nucleotides, YAG/GURAGU (where Y is a pyrimidine, R is A or G and the slash the exon-intron boundary), at the exon-intron junction, spanning from position -3 to +6. This consensus sequence is however highly degenerate with only the underlined GU dinucleotide being universally conserved (Langford et al., 1984; Sun and Chasin, 2000) (Fig 1.5). Recognition of the 5'ss involves a nearly perfect Watson-Crick base pairing with the U1 snRNA that guides the early assembly of the spliceosome machinery upon the intron (Horowitz and Krainer, 1994). Indeed several studies have shown that the introduction of mutations that improve the match of weak splice sites to the consensus can lead to the constitutive recognition of alternatively skipped exons (Del Gatto et al., 1997; Huh and Hynes, 1993; Muro et al., 1998). The extent of sequence complementarity of the 5'ss to the U1 snRNA is used to calculate the strength of the ss with 5'ss that have a high complementarity with U1 snRNP shown to splice more efficiently than those with low complementarity (Roca et al., 2005). The possibility of attribute strength to the splice site, described by a consensus sequence, has being the focus of many studies principally based on position weight matrices that are calculated from collections of splice sites (Senapathy et al., 1990; Shapiro and Senapathy, 1987). Today, a number of bioinformatics resources, albeit with slightly different modes of calculation of 5' strength are available, the more utilized of which are listed in Table 1.1.

The 3'ss (acceptor site) is even more loosely defined than the 5'ss and is composed by three elements: the branch point sequence (BPS), polypyrimidine tract (PPT) and the actual 3'ss (intron/exon junction) (Fig. 1.5).

The branch point sequence (BPS) consists of the highly degenerated sequence YNYTRAY (where Y is C or T) containing the branch point adenosine that is involved in the first step of splicing reaction (A underlined above) (Query et al., 1994). Most branch points have been mapped within 18-40 nt of the 3'ss (Reed and Maniatis, 1988).

The polypyrimidine tract (PPT) is a run of pyrimidines ($Y_{(n)}$-eight bases in the average intron), located between the branch site and the terminal AG at the intron/exon junction.

The actual 3'ss defines the 3' border of the intron, just downstream to the PPT. This site is characterised by the short YAG/G sequence (Y denotes pyrimidines; the slash indicates the intron-exon boundary and the underlined nucleotides are conserved).

Also the 3'ss is described by consensus sequence. In general, the extent of the polypyrimidine tract defines the strength of the 3' splice sites: long polypyrimidine tracts insure high affinity binding sites for spliceosomal components and promote efficient exon recognition (Reed, 1989). Commonly used programs to predict the 3'ss strength are listed in Table 1.1.

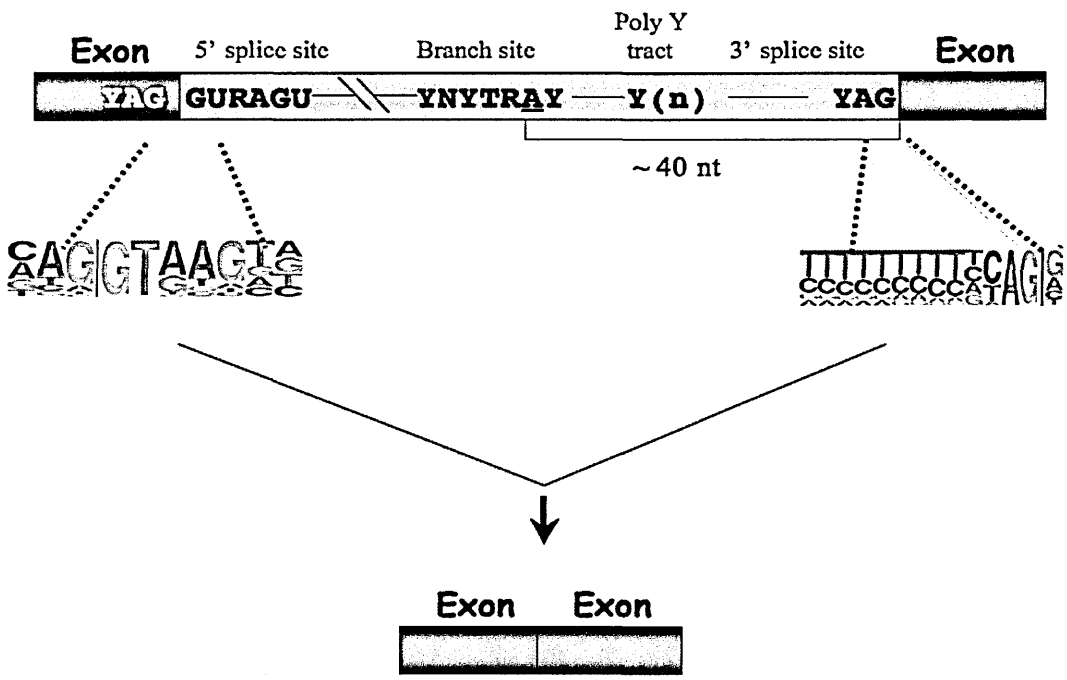


Figure 1.5. Schematic representation of mammalian exon-intron boundaries and consensus sequences for 5' and 3' splice sites and branch point. The two exons and the intron are indicated. Poly (Y) tract means region reach in pyrimidines. The universally conserved nucleotides are the dinucleotide cores of the 5' and 3' splice sites GU and AG respectively together with the branch point (A).

Some deviations from the canonical variants of splice signals are known to exist. As stated the vast majority of spliceosomal introns contain |GT at the donor splice site and AG| at the acceptor splice site. However, a distinct class of rare introns has been recognized on the basis of their unusual terminal dinucleotides: these introns contain |AT at the donor splice site and AC| at the acceptor splice site (Hall and Padgett, 1994; Jackson, 1991). Introns of this class are excised by a distinct, so-called minor or U12 spliceosome, which contains several specific, low-abundance snRNPs. It has been subsequently shown that some |GT-AG| introns are also removed by the U12 spliceosome (Dietrich et al., 1997).

Donor, acceptor site prediction	Web address	Role	Reference
a) MaxEntScan	http://genes.mit.edu/burgelab/maxent/Xmaxentscan_scoreseq.html	Scores 5' and 3'ss based on the maximum entropy principle (MEP).	(Yeo and Burge, 2004)
b) NNSplice	http://www.fruitfly.org/seq_tools/splice.html	Scores potential splice sites s based on a generalized Hidden Markov Model (GHMM)	(Reese et al., 1997)
c) NetGene2	http://www.cbs.dtu.dk/services/NetGene2/	Prediction of transition regions between introns and exons regulates a cutoff level for splice site assignment	(Brunak et al., 1991)
d) Cryp-Skip	http://cryp-skip.img.cas.cz/	Detection of cryptic splice sites	(Divina et al., 2009)
e) GeneSplicer	http://www.cbcb.umd.edu/software/genesplicer/	Prediction of splice sites based on maximal dependence decomposition (MDD) and Markov (MM) models	(Pertea et al., 2001)

Table 1.1. Programs used to analyze splice sites.

1.4.1 Relative strength and cooperation between signals

Several studies have disclosed the synergy existing either between the 5' and 3' splice sites or 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 (Nasim et al., 1990). On the other hand, sequences, upstream the 3' splice site of an exon can facilitate the use of a downstream 5' splice site (Tsukahara et al., 1994).

The polypyrimidine tract determines the location of the branch point sequence and indirectly the 3'ss (that is usually the first AG downstream the branch point). It has also been shown that a strong polypyrimidine tract can partially balance a weak branch point sequence. Likewise, a strong branch point site can partially balance a weak polypyrimidine tract (Buvoli et al., 1997). Finally, also the distance between the branch point sequence and the pyrimidine tract is critical for efficient lariat formation (Gattoni et al., 1988).

1.5 Exon/Intron Architecture

In addiction to splice site sequences, the exon/intron architecture in pre-mRNA is important for efficient splice site recognition (Deutsch and Long, 1999).

The architecture of exon and introns may be different across eukaryotic organisms. In general, vertebrate genes are characterized by short exons (average size of 170 nt.) separated by considerably longer introns (Sakharkar et al., 2005; Zhang, 1998). Interestingly, experiments aimed to investigate the influence of exon and intron sizes in pre-mRNA processing showed that spliceosome formation is strongly inhibited if exons lengths are expanded artificially to over 300 nt. However, some exceptionally long exons exist, and in these cases it has been observed that inclusion still occurs efficiently if the flanking introns are small (Sterner et al., 1996).

The variable size of exon-intron units suggested that two ways of recognizing intron and exons exist: intron and exon definition models (Fig. 1.6). Basically, when exons are small and introns are long the splicing machinery is more likely to identify and assemble across the small exon than across the large intron (Robberson et al., 1990). This model based on pairing between the splice sites across an exon that is called "exon definition model" (Fig. 1.6A) (Berget, 1995). In the alternative model, also known as the "intron definition" model, splice sites are initially paired across small introns rather than exons permitting splice site recognition within the same intronic splicing unit (Fig. 1.6B) (Berget, 1995; Lang and Spritz, 1983).

Recently, the intronic length above which exon definition becomes predominant has been estimated to be approximately 200-250 nucleotides (Fox-Walsh et al., 2005). On a practical level, however, there is no difference in spliceosomal complex assembly over exons or introns and both exon definition and intron definition may occur in different parts of the same pre-mRNA. At present, the transition from cross-exon to cross-intron interaction is still poorly understood and may often involve the action of additional cis-acting sequences to help the spliceosome to home in on the correct splice sites (Ram and Ast, 2007; Sharma et al., 2008).

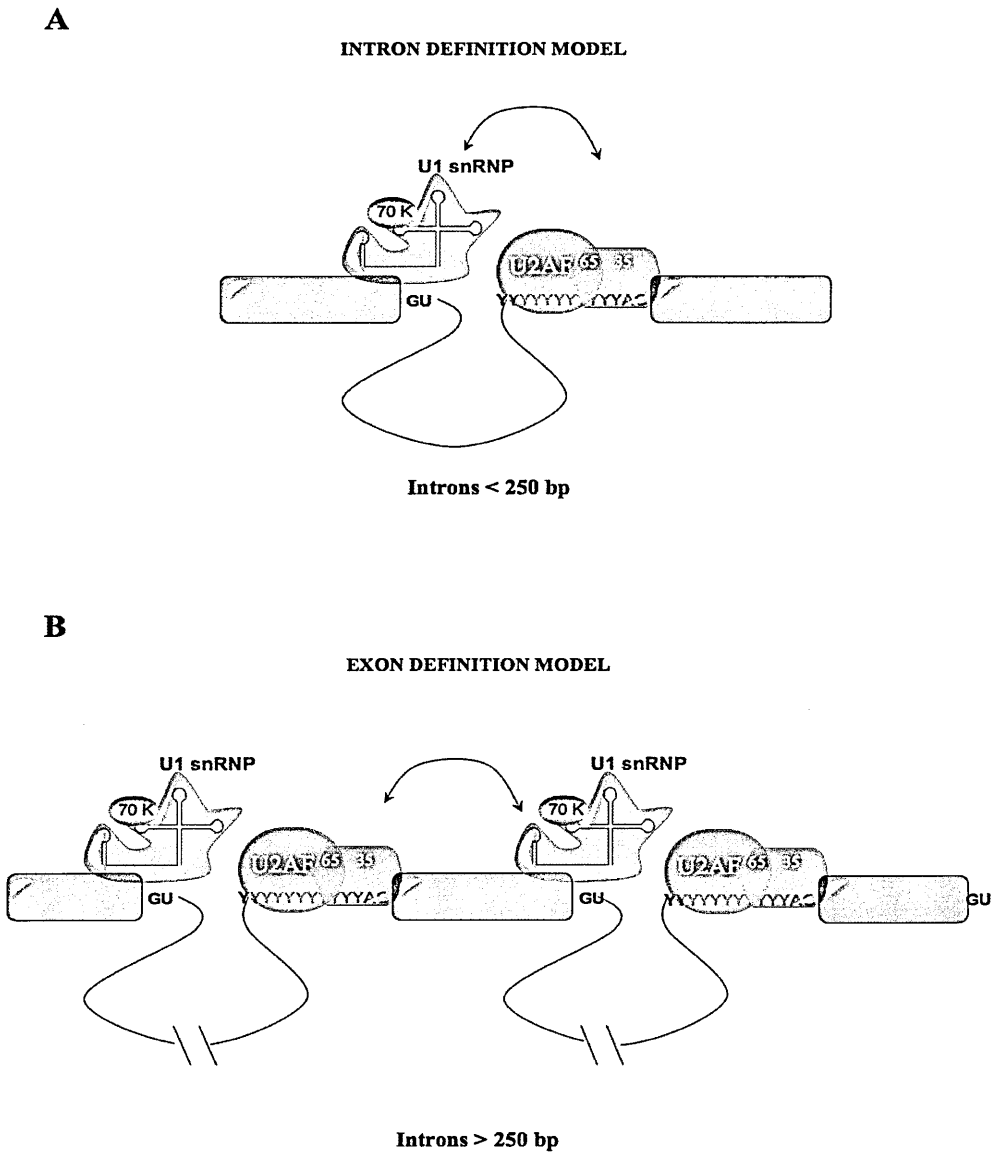


Figure 1.6. Exon and Intron definition models. (A) When the intron is small (less than about 250 nt) the spliceosome can recognize the splice sites that will be pared across the intron, referring to the "intron definition model" of splice site recognition. (B) When introns are large (greater than about 250 nt), splice site communication occurs across exons, referring to the "exon definition model" of splice site recognition.

1.6 Auxiliary splicing regulatory element

Although, the aforementioned information makes a strong case that the splice sites and exon/intron architecture are important for activating pre-mRNA splicing, these factors are not the only players.

It is well known that many potential splice sites are in fact present along all pre-mRNAs but the vast majority of these sequences, known as pseudosplice sites are never selected for splicing. For example, a computer search for potential splice sites in the 42kb human hprt (hypoxantine phospho-ribose transferase) gene, composed of nine exons and eight introns, found the eight real 5'ss but also found over 100 5' splice sites that were never recognized by the splicing machinery. The case was even more extreme for the 3'ss where 683 pseudo-sites were found with higher scores than the lowest scoring real site (Sun and Chasin, 2000).

This degeneration of the splicing code means that mammalian genes are full of apparently viable 5' ss and 3' ss consensus sequences but only a minority specifies for a "real" exon, thus the more arduous tasks of the spliceosome is to identify the *bona fide* splice sites from the numerous pseudo sites found in any pre-mRNA transcript. Certainly, the relative strength of the splice sites plays a major role in determining whether the fundamental snRNP factors will be able to bind or not (Roca et al., 2005). However, although necessary, the 5' and 3' splice sites are by no means sufficient to define the exon/intron junctions. Evidence for additional splicing regulatory elements (SRE) came about as early as 1987, with the observation that exonic sequences away from the 5' and 3' splice sites were involved in the definition of alternative splicing exons (Mardon et al., 1987).

Today far from being exceptional, the majority of pre-mRNA molecules are known to contain a myriad of auxiliary splicing *cis*-acting regulatory elements that either enhance or inhibit exon-intron recognition.

These SREs reinforce the limited information in the splice sites and depending on their

location and their function, these elements are referred as exonic splicing enhancers (ESE) or silencers (ESS) and intronic splicing enhancers (ISE) or silencers (ISS). Enhancers and silencers are involved in both constitutive and alternative splicing and in most cases they lack a well defined consensus sequence (Baralle and Baralle, 2005; Buratti et al., 2006; Cartegni et al., 2002).

These elements are not always unequivocally defined and their functions may overlap. In fact, in some systems it may be more appropriate to talk about composite exonic regulatory elements of splicing (CERES) as has been described for CFTR exons 9 and 12 (Pagani et al., 2003c; Pagani et al., 2000).

Ultimately, the combinatory effect of all the *cis*- acting elements located in proximity or within an exon affects the spliceosome assembly, promoting or inhibiting exon inclusion in the final mRNA (Matlin et al., 2005).

1.6.1 Enhancer Elements

Although originally discovered in regulated exons, ESEs are today also known to be components of constitutively spliced exons (Mayeda et al., 1999; Schaal and Maniatis, 1999). Regarding their mode of action, ESEs generally assist early spliceosomal complex formation by interacting with components of the splicing machinery that make up previously described the E complex (Reed, 1996).

ESEs have been the subject of many studies and most, but not all, are known to be recognised by members of the SR protein family of splicing factors (Blencowe, 2000; Coulter et al., 1997; Graveley, 2000; Manley and Tacke, 1996; Tacke and Manley, 1999). It should be noted that there are many exceptions for example, classical SR proteins have also been described to be involved in splicing repression in some cases (Kanopka et al., 1996; Pagani et al., 2000).

The SR proteins have a common domain structure of one or two RNA recognition motif (RRM) followed by an RS domain containing repeated arginine/serine dipeptides, which can be highly phosphorylated. This phosphorylation modulates protein-protein interaction, that serves as a bridge between the 5' and 3' splice sites across the introns and across the exons and/or between enhancers and adjacent splice site, within the spliceosome (Black, 2003; Caceres et al., 1998; Ram and Ast, 2007). The structural organization of SR proteins suggests a model for their function. The RRM mediates sequence-specific binding to the mRNA, whereas the RS domain seems to be involved mainly in protein-protein interactions (Cartegni et al., 2002; Graveley, 2000).

Two non-exclusive models have been proposed to explain the role of SR protein in pre-mRNA splicing (Fig. 1.7). One model is supported by the observation that ESE-dependent splicing, requires protein-protein interactions mediated by SR proteins (Hertel and Maniatis, 1999). Through this interaction, the splicing factor U2AF is thought to be recruited onto a suboptimal, upstream 3'splice site and stimulate spliceosome assembly. The recruitment of U2AF, seems to be important, especially in those cases in which recognition of weak pyrimidine tract is a rate-limiting step in splicing reaction (Graveley et al., 2001)(Fig. 1.7A).

The second model proposes that a SR protein, bound to an ESE, can antagonize the negative effect of a juxtaposed silencer element (Cartegni et al., 2002; Sanford et al., 2005). However, in some cases SR proteins can act in a negative fashion. The negative effect on splicing can be mediated by the binding to an intronic sequence (ISS) (Buratti et al., 2007; Ibrahim el et al., 2005) or by the inhibitory property of the protein itself, as reported for SRp38 and SRp86 (Barnard et al., 2002) (Fig. 1.7B).

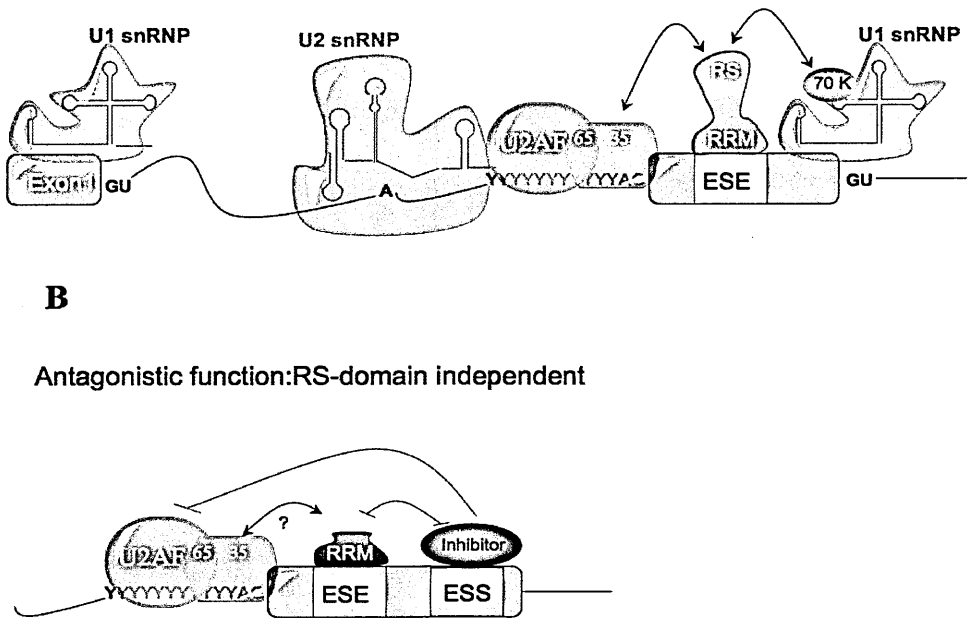


Figure 1.7. Models of SR protein action in exonic-splicing-enhancer-dependent splicing. (A) RS-domain-dependent mechanism. An SR protein binds to an exonic splicing enhancer (ESE) through its RNA-recognition motifs (RRM) and contacts the splicing factor U2AF35 and/or U1-70K at the adjacent splice sites through its RS domain. U2AF65 binds to the polypyrimidine (Y) tract, which here is interrupted by purines and is therefore part of a weak 3' splice site. U2AF65 also promotes binding of U2 snRNP to the branch site. The U1 snRNP particle binds to the upstream and downstream 5' splice sites through base paring of the U1 snRNA. The three sets of splicing-factor–pre-mRNA interactions (U2AF–3' splice site, U1 snRNP–5' splice site and SR protein–ESE) are strengthened by the protein–protein interactions (black arrows) that are mediated by the RS domain. (B) RS-domain-independent mechanism. Here, the main function of the SR protein that is bound to an ESE is to antagonize the negative effect on splicing of an inhibitory protein that is bound to a juxtaposed exonic splicing silencer (ESS). The SR protein is shown without its RS domain, although this domain is normally present and might still promote U2AF binding, or other domains might be involved in protein–protein interactions. Inhibitory interactions are shown (red), as is a putative stimulatory interaction (double-headed arrow). These models are not mutually exclusive, and the splicing of some introns might involve a combination of these mechanisms. Figure adapted from Cartegni et al. (2002).

Cis-acting enhancer regulatory elements in the intronic sequences have being studied to a lesser extent, however ISE do indeed exist and a number of examples have been studied. One well characterized ISE element is the triplet (GGG) which often occurs in clusters and can enhance recognition of adjacent 5' or 3' splice sites or the intronic (CA) repeats that in several cases can enhance splicing of upstream exons (Hui and Bindereif, 2005; McCullough and Berget, 1997).

1.6.2 Silencer Elements

In addition to sequences that promote exon inclusion, there are sequences that inhibit splicing, so called exonic or intronic splicing silencers. The silencers are less well characterized: they can be purine or pyrimidine-rich and bind a diverse array of proteins which have not been characterized to the same extent as the enhancers, however heterogeneous ribonuclear proteins (hnRNPs) have been generally implicated in interactions with these elements and in particular, ISS are usually recognized by the polypyrimidine track binding protein (PTB; also known as hnRNPI) (Fairbrother and Chasin, 2000; Garcia-Blanco et al., 1989). Several mechanisms have been proposed for ESS- or ISS-mediated splicing repression, indeed, hnRNP-bound splicing silencer have been shown to repress spliceosomal assembly through multimerization along exons (Zhu et al., 2001), blocking the recruitment of snRNPs (House and Lynch, 2006), or by looping out the exon (Martinez-Contreras et al., 2006).

1.6.3 Bioinformatic analysis of SRE

Although these additional elements are often conserved between species, highly degenerate sequence motifs characterize these sequences, making their identification difficult.

Due to the importance of SRE(s) a number of both experimental and computational approaches have also been developed in order to identify the auxiliary splicing regulatory motifs that regulate exon inclusion. The result of these studies led to the development of computational methods that predict splicing regulatory sequences and their binding partners. The experimental approaches have included classical SELEX experiments, functional SELEX, (Buratti and Baralle, 2005; Cartegni et al., 2002; Coulter et al., 1997) in vitro splicing assay (Smith et al., 2006) and in vivo screening for regulatory elements (Wang et al., 2004). In particular, the classical SELEX approach has been used for a variety of splicing factors, with a particular emphasis towards the splicing factors that up- or down regulate exonic inclusion. However, sequences belonging to naturally occurring enhancer or silencer elements do not always correspond to the optimal binding sites that have been identified using the selection techniques. Therefore, the functional SELEX approach has been introduced, combining the splicing functionality and protein-RNA binding efficiency. Exonic splicing enhancer (ESE) sequences can be functionally investigated by coupling the SELEX basic methodology with an in vitro splicing RNA template. To obtain selection specificity for a specific splicing factor, the splicing reaction can be performed in a depleted nuclear extract that can drive the splicing process only if complemented by a recombinant splicing factor of interest. These functional experiments have enabled the elaboration of binding matrices, which can then be used to search for potential SR-binding motifs in any RNA sequence of interest using a web-based application, ESE-Finder (http://rulai.cshl.edu/tools/ESE/) (Cartegni et al., 2003). This has resulted in the identification of high-affinity binding sites for four SR proteins: SF2/ASF, SC35, SRp40 and SRp55. These binding sites consist of purine-rich sequences known to function as a

splicing enhancer.

In addition to the ESE-finder application, another web-based application for predicting the occurrence of enhancer motifs (but based on the statistical analysis of exonic sequences rather than selection procedures) RESCUE (relative enhancer and silencer classification by unanimous enrichment)-ESE (http://genes.mit.edu/burgelab/rescue-ese), was tested (Fairbrother et al., 2004b). RESCUE-ESE looks for hexanucleotides that are significantly enriched in exons as opposed to introns and also more abundant in exons with weak (non-consensus) splice sites than in exons with strong splice sites. This approach identified 238 hexanucleotides, and a sample of them was successfully validated using minigene system for testing enhancer activity.

Although promising, this web-based approaches show a variable degree of reliability and high false positive rate, hence, at present there is no substitute for functional experiments to validate splicing predictions.

1.7 Combinatorial Mechanism

In higher eukaryotes, the gene complexity and the relatively low level of splice site conservation, make the precision of the splicing machinery in recognizing and pairing splice sites impressive.

Over the last few years, it has become increasingly clear that to reach this outcome, splice site selection has evolved to depend on multiple parameters such as splice site strength, the exon/intron architecture and RNA secondary structures. The sum of contributions from each of these parameters is necessary to control the generation of constitutive and alternatively spliced mRNA molecules (Hertel, 2008).

An example of combinatorial splicing regulation is the SRC tyrosine kinase gene in which a cassette exon N1 is included in neurons, but skipped in non-neuronal cells (Levy et al., 1987). The exon N1 (18 nt) is located between exon 3 and 4 and in non-neuronal cells, exon N1 is skipped to generate *c-src* transcript. PTB binds cooperatively to sequences on both

sites of the N1 exon. As a result, exon 4 splices to exon 3. In neuronal cells, N1 is included between exon 3 and 4 to form the *n-src* transcript. In this case PTB is replaced by less repressive neuronal PTB an enhancer complex assembles on the downstream control sequence resulting in the inclusion of N1 exon in the mRNA (Black, 2003; Modafferi and Black, 1997). Thus, in this as in most cases, the decision to splice is due to a balance of multiple positive and negative inputs.

1.8 Protein evolution: An Overview

Several important hypotheses have been proposed so far to determine the main players acting on protein evolution (Pal et al., 2006).

A landmark study from Zuckerkandl and Pauling (Zuckerkandl and Pauling, 1965) suggested that there is a uniform rate of amino acid changes, named molecular clock, among species. This observation was the basis for the development of the neutral theory of molecular evolution. The neutral theory (Kimura, 1983) suggested that advantageous mutations occur so infrequently that they do not significantly contribute to the rate of molecular evolution and consequently most amino acid substitutions are neutral and fixed by random genetic drift with a constant substitution rate.

In the 1970s, the sequencing of many proteins from diverse species provided opportunities to examine the nature of the protein clock. By 1971, it was shown that the rate of molecular evolution was constant only for a particular protein but it was vastly different among different proteins. It was also suggested that the surface residues of a specific protein could be constrained by the protein's interactions with other proteins, therefore some portions of the surface of certain proteins seemed to have higher functional constraint than others. For example, fibrinopeptides have little functional constraint and evolve very fast; instead cytochrome c has high functional constraints and evolves very slowly (Dickerson, 1971) In this case assuming that functionally important amino acids remain essentially unchanged in

the evolutionary process but others can change with neutral rate, the rate of amino acid substitution (r) was expressed by $r = fv$ where f is the proportion of neutral mutations and v is the mutation rate per site. This predicts that less important proteins should evolve at faster rates (higher r) because f should be greater for less important proteins.

In the following years Zuckerkandl also indicated that the rate of evolution is proportional to the density of protein functional domains. In the "functional density hypothesis" it was suggested that the selective constraint should be proportional to the number of residues involved in its function and that the domain architecture of the protein should be untouched to avoid selective pressure (Zuckerkandl, 1976). The functional density (F) can be expressed as $F = n_s / N$ where n_s is the number residues involved in specific function and cannot be easily substituted and N is the total number of sites in the protein. This formula should reflect the ratio of constrained to neutral amino acids for a given protein, hence, it should be proportional to the rate of sequence evolution. The measurement of functional density remained tricky because residues may contribute to protein function in unpredictable ways. Many recent surveys have analyzed other measures that may represent functional density, such as expression level and dispensability. Indeed, Wilson et al. proposed that two proteins subject to the same level of functional constraint, but differing in their dispensability, would evolve at different rates. Thus, essential genes (knockouts of which are lethal or infertile) should evolve slower than non-essential genes (knockouts of which are viable and fertile) (Wilson et al., 1977).

Most recently, a study has led to the "fitness density hypothesis" in which the rate of evolution was considered as a measure of the change in fitness of the mutant protein (relative to the wild type molecule). Consequently, it has been shown that non-essential protein have higher rate of evolution since they can accumulate mutations without resulting in a deleterious effect on the fitness. Conversely, essential proteins having stronger fitness effect are subjected to higher selective pressure, reducing the possibility to accumulate deleterious mutations (Hirsh and Fraser, 2001).

1.8.1 Codon based model of protein evolution

Many different mathematical models have been developed for identifying the way in which selection occurs, one of the most used is a codon substitution model. Mutation rate of protein coding sequences treat the codon as the unit of evolution and distinguish between synonymous and nonsynonymous rates of evolution. Synonymous mutations yield a different codon without changing the encoded amino acid and therefore do not affect the protein sequence. Nonsynonymous mutations, on the other hand, result in replacement of one amino acid with another. This distinction enables the calculation of two substitution rates: Ks, the number of synonymous substitutions per synonymous site and Ka, the number of nonsynonymous substitutions per nonsynonymous site (Goldman and Yang, 1994).

By distinguishing between Ks and Ka it is possible to draw inferences regarding the nature of the selection operating on the protein-coding sequence. In particular, the ratio of these rates (Ka/Ks) is commonly used to estimate ω (the amino acid selection pressure) corrected for π (the background nucleotide mutation rate). This follows from the fact that, because synonymous changes are silent at the protein level, synonymous sites are typically regarded as neutrally evolving. Therefore, the synonymous rate is dependent on the nucleotide mutation rate, π and not on amino acid selection pressure, ω. Nonsynonymous sites, on the other hand, evolve at a rate determined by both these processes.

In a neutrally evolving protein-coding sequence nonsynonymous mutations are as likely to be fixed as synonymous mutations (i.e., Ka/Ks is expected to equal one). This fact as led to the common use of the ratio Ka/Ks to estimate the nature and magnitude of different types of amino acid selection pressure. Values of $Ka/Ks < 1$ indicate the operation of purifying selection in causing a reduction in the fixation rate of amino acid changes that are deleterious relative to the silent synonymous rate. Positive selection for beneficial amino acid changes is frequently inferred when $Ka/Ks > 1$ (Hurst, 2002).

1.8.2 Splicing and the evolution of proteins in mammalian proteins

The higher variability of rates in multicellular organisms such as mammals could be generally due to different factors. First, the small size of population can influence the efficiency of selection against deleterious mutations. Second, the organization in tissue and organs in mammals is likely to be associated with selective constraints indeed, early studies of the impact of expression breadth on protein evolution in mammals have shown that the rate of amino acid substitution depends on the tissue in which proteins are expressed (Hughes, 1997; Kuma et al., 1995) and, also, that broadly expressed proteins are more conserved than tissue-specific one. Therefore, the slow evolution of a ubiquitously expressed protein could be due to an increase in the functional density of a sequence resulting from the dual requirement that the protein should function under a wide range of conditions where it encounters a wide range of molecular interaction partners.

Another important factor affecting the rate of evolution of mammalian proteins is the genome heterogeneity that produces the variation of synonymous substitution rate between mammalian genes (Wolfe et al., 1989). Although, Ka/Ks ratio has been widely used to measure the selective pressure on amino acids based on nucleotide mutations causing amino acid changes of the encoding genes (Yang and Nielsen, 2000) this ratio often produces ambiguous results due to the selection acting also on nucleotide mutation that do not cause amino acidic changes (Hurst and Pal, 2001). For example, alternative spliced exons show higher Ka/Ks ratio in sequence comparison indicating the importance of local segments like SREs of the gene for selection (Xing and Lee, 2005a) (see section1.8.3).

Several bioinformatics analyses have confirmed that the need to preserve information necessary for intron removal (e.g., exonic splicing enhancer) represent a strong force operating on protein evolution in multicellular organisms (Warnecke et al., 2008). In particular, the presence of splicing enhancers has been found to influence the usage of synonymous codon near the splice sites. Indeed it has been documented that codons preferred in ESE are recurrent near intron-exon boundaries (Berget, 1995). Willie and

Majewski have evidenced that although both synonymous triplets GAA and GAG encode for the glutamic acid, have different usage in humans, GAA triplet represents the most common codon in exonic splicing enhancers and is also most frequently found near the splice site because of its role in promoting splicing. (Willie and Majewski, 2004).

Conversely, a bioinformatic study showed that the AAG codon (lysine), even though it is abundant in ESE, exhibits a decrease frequency near the 5'end of the exon. This results have been explained with cryptic splice site avoidance theory in which Eskesen *et al.* have hypothesized that because the 3' ends of introns typically terminate AG, exons should avoid using this nucleotide at the 5' end to minimise the chance of deleterious aberrant splice forms. They also found some evidences also for GT avoidance at exonic 3'ends. (Chamary and Hurst, 2005; Eskesen et al., 2004).

More generally, analysis of the distribution of SNPs in human exons showed a decreased density near the 5' and 3' ends to preserve a correct recognition of the splice sites (Majewski and Ott, 2002). Additionally, RESCUE-ESE data set distribution showed an inverse trend compared with SNP density, with an increased concentration in vicinity of exon boundaries providing evidence that there is a purifying selection against mutations that affect ESE with lower rate of Ks than in non-ESE sequence.(Carlini and Genut, 2006; Fairbrother et al., 2004a) (Parmley et al., 2006). These bioinformatics studies are in accordance with in-depth analysis of genes that demonstrated that synonymous changes couldn't evolve freely, because they can compromise correct splicing (Baralle et al., 2006; Pagani et al., 2005; Raponi et al., 2007).

Since these studies have challenged the neutrality of Ks, Ke *et al.* have introduced the rate of intronic substitution (Ki) to estimate the neutral mutation rate. Using a genomic comparative approach between human, chimpanzee and macaque, they have observed: (1) lower Ks than Ki, supporting the role of RNA beyond protein coding; (2) the avoidance of mutations that disrupt predicted ESE or create ESS and that synonymous substitution in constitutive exon tend to create enhancer and eliminate silencer sequences; (3) when

splicing efficiency tend to decrease in one species due for example to the weakening of splice site consensus sequence, there is a tendency to compensate this negative event with predicted ESE gain or ESS losses. These results supported the hypothesis of a purifying selection that acts to preserve these splicing-promoting sequences and a positive selection for their formation. It has been suggested that this positive selection might be the result of splicing-positive events compensating for splicing-negative events as well as for mutations that weaken splice-site sequences (Ke et al., 2008). Thus, based on the splicing compensation model of exon evolution, the splicing elements seem coevolve in a way that preserves overall exon strength, allowing specific elements to substitute for loss or weakening of others (Xiao et al., 2007). In line with this proposed model, a recent bioinformatics survey by Parmley *et al.* suggested that the selection acts not only on the synonymous codon choice near intron-exon junction, but also the majority of amino acids show skewed usage because of their involvement in splicing. The authors have also reported that ESEs leave an imprint on the amino acid composition of proteins. In fact, some amino acids, such as lysine (K) and isoleucine (I), are strongly preferred near boundaries whereas others, such as proline (P) and alanine (A), are significantly avoided due to the effect on splicing definition at the nucleotide level (Parmley et al., 2007).

1.8.3 Alternative splicing and evolution

Evolutionary studies, which have revealed the formation of de novo alternative exons and the evolution of exon–intron architecture, highlight the importance of alternative splicing in the diversification of the transcriptome, especially in humans.

Alternative splicing may, therefore, be a mechanism that enables evolution to experiment with newly created exons. The obvious correspondence with gene duplication was made explicit by a study documenting differences in gene structure between human and mouse orthologs (Modrek and Lee, 2003). Modrek and Lee analyzed a set of 9,434 orthologous genes in human and mouse. This work introduces an additional distinction within

alternative exons: 'major form'' if the exon appears in at least 50% of the transcripts and ''minor form'' otherwise.

They also introduced the term "internal paralog" to describe a "minor-form transcript" whose alternative exons are generally not conserved between human and mouse. This minor form is free of selective constraint thus allowing it to accumulate changes more rapidly. The major-form of alternative exons are much more conserved between human and mouse. Modrek and Lee demonstrated that alternative splicing is likely to have facilitated many of the changes that have occurred in the structure of genes since the human/mouse divergence. This was illustrated by the finding that species-specific exons are 10 times more likely to be alternatively spliced than conserved exons. Notably, this facilitation is dependent on the low frequency incorporation of these species-specific exons into transcripts. Alternatively spliced exons specific to human or mouse are nearly eight times more likely to be spliced at low frequencies (i.e., as the minor form) than alternative exons conserved between these mammals (Modrek and Lee, 2003). It is this low-frequency expression, by alternative splicing, of species-specific "internal paralogs" that is likely to protect newly formed exons from selection while ensuring that the gene's ancestral function is not compromised.

Evidence is emerging for the existence of two contrasting selective pressures operating on alternatively spliced exons.

On the one hand, alternative splicing is associated with an apparent relaxation of negative selection. There is also evidence that purifying selection on amino acid changes (as measured by Ka /Ks) is up to seven-fold weaker in alternatively spliced exons (Xing and Lee, 2005b).

On the other hand, alternative splicing is also associated with an increased selective constraint. Alternatively spliced exons are observed to be under stronger selection to preserve reading frame (Resch et al., 2004) and to have fewer single nucleotide polymorphisms (Yeo et al., 2005). Strikingly, a more than six-fold reduction in

synonymous site divergence is seen among minor-form exons compared to constitutive exons (Xing and Lee, 2005a). The most likely explanation for the strong selective pressure on synonymous sites in alternative exons relates to the preservation of splicing regulatory motifs. Consistent with these findings the presumed neutrality of synonymous mutations has been challenged because of the presence of splicing regulatory elements overlapping with the amino acid code (Pagani and Baralle, 2004).

1.9 Impact of gene duplication on rates of molecular evolution

The duplication of single genes can occur by DNA-based tandem duplication and duplicative transposition. However, some duplication events do not generate perfect copies of their progenitor. The significance of gene duplication to genome evolution is estimated by the high birth rate of duplicated genes and frequently is a source for evolutionary innovation (Lynch and Conery, 2000). However, there is also a high frequency of gene duplicates lost from the genome.

1.9.1 Different mechanisms for paralogous genes preservation

Although the life of most gene duplicates is generally short there are several mechanisms that increase the survival chances of a newly formed gene duplicate.

The frequent loss of gene duplicates is consistent with Ohno's classical model under which gene duplication creates two paralogous genes that are functionally redundant (Ohno, 1970). This redundancy implies that one of the paralogs could evolve free from selective constraints. Under Ohno's model the ultimate fate of this unconstrained paralog is determined by the neutral accumulation of mutations that were previously forbidden by selection. Given the abundance of degenerative mutations, the most likely event is the fixation of a null allele that results in the nonfunctionalization and ultimate loss of the gene

duplicate. According to the classical model the only way to survive for a duplicate gene is the creation of a new function by the rare fixation of beneficial mutations. This process is referred to as neo-functionalization and can occur by the fixation of mutations in the duplicate gene's protein-coding or regulatory sequences. A characteristic of the neo-functionalization mechanism is that the survival of a newly formed gene duplicate is guaranteed by its gain of a novel function that differentiates it both from its duplicate and ancestral copy.

An alternative, more recently described, model proposes that gene duplicates can be retained without functional innovation and adaptation. The sub-functionalization model (Force et al., 1999) provides a neutral explanation for the retention of duplicate copies of a multifunctional gene by degenerative mutations that lead to the loss of different subfunctions in each duplicate. The complementary pattern of subfunction loss ensures that both duplicates are required to perform the ancestral set of subfunctions and therefore both must be retained in the genome. There is an alternative, adaptive, way for the preservation of duplicate copies of a multifunctional gene. Duplicating such a gene provides a chance to eliminate the negative pleiotropic constraints, and allows the refinement of each subfunction by positive selection (Hughes, 1994; Piatigorsky and Wistow, 1991).

It should be noted that gene duplicates retained in the genome provide considerable molecular substrate for the later development of evolutionary novelty. Therefore the preservation of gene duplicates either through increased protein dosage or as a result of sub-functionalization is compatible with later acquisition of novel functions (He and Zhang, 2005; Kondrashov et al., 2002).

Nevertheless, one member of a duplicate pair may directly diverge in function relative to the other. Therefore, when one paralog undergoes neofunctionalization, positive selection for this new function will result in a rate acceleration of the protein relative to its duplicated copy. An implicit assumption of the neofunctionalization model is that the second duplicate performs the ancestral gene's function and continues to evolve at the same rate as its parent.

This model shows that the preservation of gene duplicates by divergence in protein function can directly impact on rates of protein evolution.

However, the maintenance of duplicate genes by neofunctionalization or subfunctionalization can also proceed by divergence in gene expression profile. In this case, the immediate target of these preservational processes is the non-coding sequence responsible for gene expression regulation rather than the protein-coding sequence. The modularity of non-coding regulatory sequences makes them independently mutable and especially prone to the complementary degenerative mutations characteristic of subfunctionalization.

Expression divergence may also govern the sequence divergence of gene duplicates in multicellular eukaryotes. In mammals, population genetic considerations suggest that subfunctionalization is more likely to occur because of their small population sizes. As outlined above, the vulnerability of cis-regulatory sequences to degenerative mutations means that expression patterns are particularly susceptible to subfunctionalization. In fact, the partitioning of ancestral expression patterns can proceed both quantitatively (by a division of the ancestral expression level between duplicates so that their summed expression is required to fulfill ancestral function (Ferris and Whitt, 1979), spatially (by division of the constituent tissues of the ancestral expression domain among the duplicates (McClintock et al., 2002) or temporally (by division of expression at different developmental stages among the duplicates (Yan et al., 2005). In each case the divergence in expression is expected to result in an increase in evolutionary rate of the duplicates relative to the ancestral gene. In summary, it seems likely that asymmetry in expression between mammalian gene duplicates has the potential to explain some of their asymmetry in the rate of sequence divergence.

1.10 Alkaline phosphatase gene family

One protein family in which the need for an ESE has influenced gene evolution is the Alkaline Phosphatase (ALP) family. ALPs are a family of homodimeric enzymes that catalyze the hydrolysis of monoesters of phosphoric acid with release of inorganic phosphate (McComb and Bowers, 1972), schematically shown in the following general reaction:

$R\text{-}OP + H_2O \rightarrow R\text{-}OH + Pi$

Alkaline Phosphatases (ALPs) occur widely in nature and are found in many organisms from Escherichia coli to man. The catalytic site contains residues and cofactors that are preserved in different species and essential for enzymatic activity, i.e. the catalytic Ser and three metal ions (two Zn^{2+} and one Mg^{2+}) (Stec et al., 2000).

Although mammalian ALPs have different catalytic activity and affinity for the ligands this feature of the catalytic mechanism has been conserved during their evolution (Kim and Wyckoff, 1991). However, whereas the ALP in E. coli is situated in the periplasmic space, the enzymes, in mammals, are glycosylated ectoplasmic enzymes attached to the plasma membrane via a glycosylphosphatidylinositol (GPI) anchor.

In humans, ALPs are encoded by four distinct loci (Harris, 1990), traditionally named after the tissues in which they are expressed (Table 1.2). (1) The tissue non-specific ALP gene (ALPL) (Weiss et al., 1988), which encode the protein TNAP, has an ubiquitous expression, but with predominant presence in liver, bone and kidney, and in placenta during the first trimester of pregnancy (McComb et al., 1979). It has been shown to have an important role in mineralizing bone, where it is expressed on the plasma membrane of osteoblastic cells. (2) The intestinal ALP gene (ALPI) (Henthorn et al., 1988), which encode the protein IAP, has been found in the brush border of the intestinal cells. (3) The placental ALP gene (ALPP) (Knoll et al., 1988), which encode the protein PLAP, has been found expressed in the syncytiotrophoblast from the first trimester of pregnancy to term. (4)

The placental-like ALP gene (ALPPL2) (Millan and Manes, 1988), which encode the protein GCAP, has been found expressed in primordial germ cells and in small amount in testis and thymus.

The amino acid sequences of PLAP and GCAP are 98% similar. In the case of PLAP and IAP, there is 90% of similarity. The tissue-nonspecific (ALPL) is approx. 50% identical to the other three isozymes (Knoll et al., 1988; Le Du and Millan, 2002).

Gene name	Protein name	Definition	Chromosomal location	Function
ALPL	**TNAP**	Tissue-nonspecific alkaline phosphatase	chr1:p36.1-p34	Bone mineralization
ALPP	**PLAP**	Placental alkaline phosphatase	chr2:q34-q37	Unknown
ALPPL2	**GCAP**	Germ cell alkaline phosphatase	chr2:q34-q37	Unknown
ALPI	**IAP**	Intestinal alkaline phosphatase	chr2:q34-q37	Intestinal absorption?

Table 1.2 Summary of gene nomenclature, protein names, chromosomal location and function, if known (Table taken from Millan, 2006).

Furthermore the human ALPs have been distinguished from each other by different methods: thermostability, inhibition, and immunologic studies.

Thermostability: PLAP and GCAP are highly thermostable. In contrast the other enzymes are inactivated under high temperature conditions. However, IAP is more stable than TNAP when treated at 56 °C (Harris, 1980).

Inhibition studies: ALPs have been shown to respond differently when treated with inhibitors through an uncompetitive mechanism. PLAP, GCAP and IAP were found to be more sensitive to L-phenylalanine than TNAP. Conversely, the TNAP is inhibited by L-homoarginine that affects in much less extent the other enzymes (Lin and Fishman, 1972).

Instead L-leucine and an unrelated compound, Levamisole, are particularly strong inhibitor of GCAP and TNAP respectively.

Immunologic studies: TNAP is the only enzyme that does not cross-react with antisera raised in rabbits against placental ALP. Hence, IAP and GCAP are more closely related immunologically to PLAP than is TNAP (Lehmann, 1980).

In mammals only few compounds have been confirmed to be a natural substrate. Inorganic pyrophosphate (PP_i) has been well documented that the hydrolysis to P_i by TNAP, thus providing P_i needed for bone mineralization (McComb et al., 1979). A natural substrate of TNAP is also pyridoxal-5'-phosphate (PLP) (a phosphorylated form of vitamin B6). ALPs appear also to be involved in the metabolism of nucleotides and sugars (Say et al., 1991).

The gene structure of these 4 proteins also shares some similarity. The coding regions include a signal peptide (17-21 amino acids), which is cleaved off in the mature protein. The human tissue-specific genes, ALPP, ALPPL and ALPI, are clustered on human chromosome 2 (2q34-q37) and they are closely related to one another. Their structures are nearly identical, consisting of 11 exons at analogous positions and the similarity between all three genes suggests a divergent evolution. ALPL on the other hand is located on the short arm of chromosome 1 (1p36.12) with a length of 69 Kb, five times superior than each other three genes and has 12 exons compared with 11 in the tissue specific alkaline phosphatase genes (TSAP); the first exon is at the 5' end in the non coding region and the second exon contains also the start codon for translation. The difference in the length is due to larger introns in ALPL gene. Notwithstanding these differences the coding regions are interrupted however in similar positions in all genes and are comparable also in length (Fig 1.8)

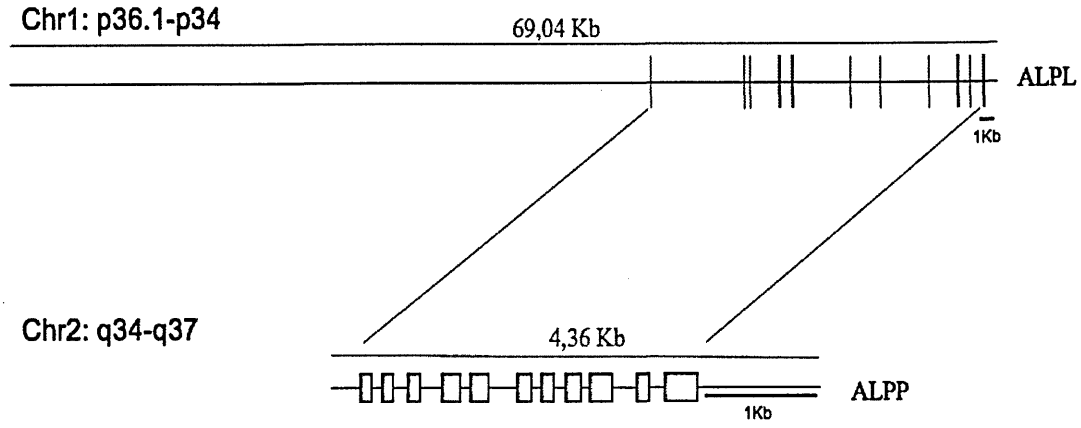


Figure 1.8. Diagram showing the genomic organization of ALPL and ALPP genes. In this graphic, exons are represented by boxes and introns by lines. Figure modified from (Harris, 1990).

2. AIM OF THE THESIS

It is well established that exonic sequences contain ESEs that overlap with coding capacity and that these may influence protein evolution. However, there has been a lack of experimental data that investigate the effect of how the need for ESE might influence the amino acid composition and protein activity of evolutionary correlated proteins.

For this reason, the aim of the PhD thesis was to investigate if the coexistence of exon splicing enhancer elements with amino acidic coding capacity might restrict the evolutionary selection of codon variants that could affect the protein function. Therefore, at least in a fraction of the exons that are present in the genome, suboptimal protein function might be tolerated to allow the persistence of sequences that are essential for exon inclusion (Pagani and Baralle, 2004; Pagani et al., 2005) (Fig.1.9). The principal line of investigation taken in my PhD is to see if there is a combinatorial effect between exonic splicing enhancers, splice sites strength and protein activity. In order to do so, ideal candidate gene families in which evolution at the protein level may be hindered by the presence of cis-acting regulatory elements need to be identified (bioinformatic collaboration). In particular, we searched for paralogous gene families in which differences in the amino acid sequence are associated with the presence or absence of hypothetical ESE, and the presence of these ESE could in turn be associated to weaker splice site strengths.

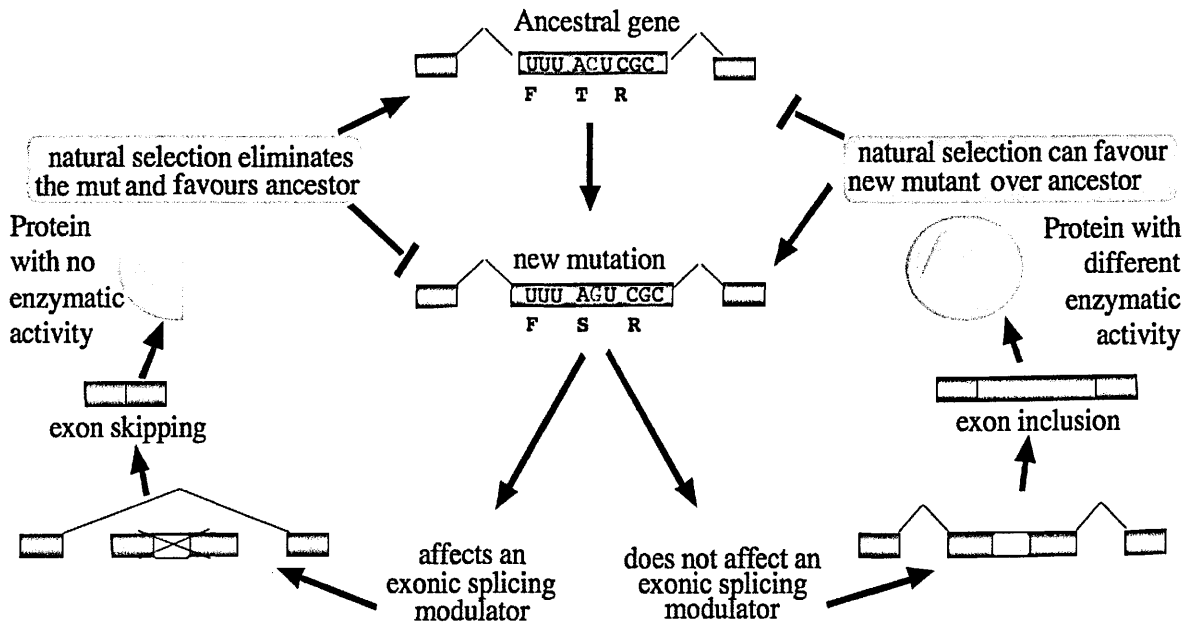


Figure 2.1. The primary selective pressure on exons is for their inclusion in mRNA. The selection of a new amino acid that leads to a better enzyme can occur only if the codon substitution caused by a genetic mutation, does not affect an exonic regulatory element. In this model it is assumed that the threonine to serine change at the catalytic site will produce a more active enzyme that has a selective advantage. However the C to G substitution has to be compatible with the splicing machinery that identifies the exon. If the inclusion is guaranteed, then the amino acid change is favoured. If not, exon skipping will result in an inactive enzyme and the ancestral threonine that produces a suboptimal protein will be kept to ensure the exon inclusion step. Figure modified from Pagani and Baralle (2004).

3. RESULTS

3.1 ESE Analyzer Web Server (EAWS) computational analysis

The nucleotides present in the exonic sequence not only code for the amino acids of the protein but can also play a role in splicing, coding for splicing cis-acting elements such as ESE. This has led to the hypothesis that the splicing regulatory elements may restrict the choice of amino acids in these areas and, therefore, possibly also influence protein activity (Pagani and Baralle, 2004; Pagani et al., 2005).

In order to identify possible candidate genes in which the evolution of exon sequences might be constrained by the presence of ESE, I looked at closely related human paralogous genes in which the protein function and exonic organization were maintained during evolution. The search criteria was to find a protein family which had a measurable enzymatic activity that would allow subsequent analysis of the constraint that may occur between optimal splicing and optimal protein functions with greater ease. This hypothetical family of paralogous proteins would have members with different amino acids within or close to the functional domain and these differences in residue composition would correspond to creation or disruption of an ESE. In order to aid the screen, as ESE are often present in exons with "weak" 3 or 5' ss (see section 1.5.1), I also searched for scenarios in which weak/strong splice sites are associated with the presence/absence of ESE within the same paralogous family. Therefore, ideal candidates were paralogous genes within which a member would have ESE constraining variation that would have effects on functional characteristics of the protein and that the presence or absence of these ESE sequences correlated respectively to weak and strong splice sites.

In order to perform this search at a level of a wide screen it was necessary to integrate information from several bioinformatic tools, a complex and time consuming task with visually hard data sets to follow. To overcome these problems, a bioinformatic platform was created to identify candidate paralogous genes: ESE Analyzer Web Server (EAWS). This program was developed in collaboration with Dr. Vlahovicek (University of Zagreb, Croatia) and provides information related to splicing regulatory sequences in a visually clear manner. Since as mentioned above, one of the difficulties in the screen was the visualization of the data sets, an important characteristic of EAWS is that it is built on user-friendly interface that is a crucial point in order to facilitate access to information for the user. Indeed, the user can obtain a list of candidate genes by either entering via the Interpro domain identification key (e.g. IPR001952), submitting the sequence in FASTA format, inserting Transcript/Exon ID or using a keyword (e.g. *protease, phosphatase*) (Fig. 3.1).

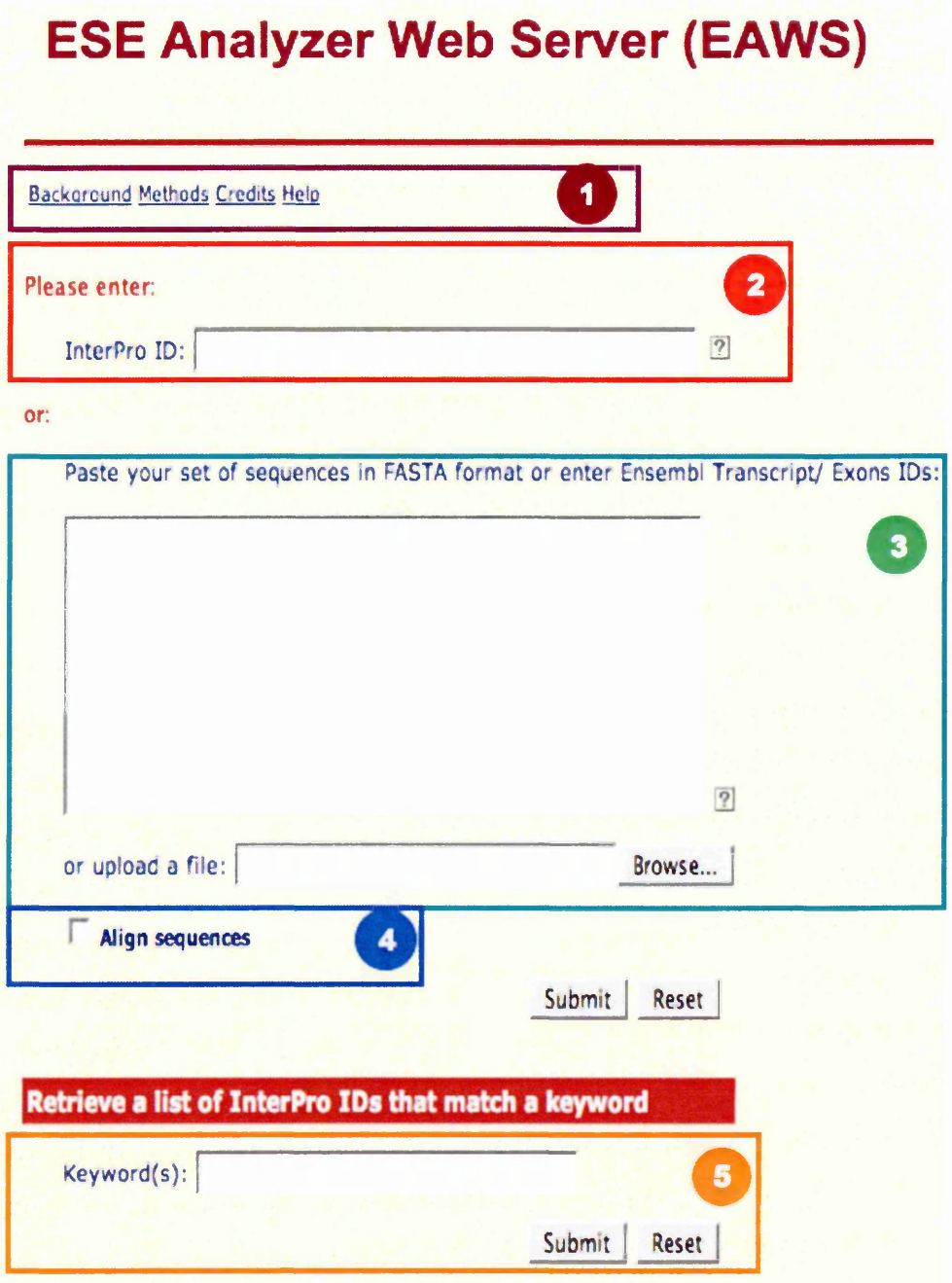


Figure 3.1. Start page of ESE Analyzer Web Server (EAWS). 1- Link to Introduction (background, methods, etc.,); 2- textbox for Interpro ID input; 3- textbox for user submitted sequence and/or Ensemble exon and transcript IDs; 4- choice to align submitted sequences; 5- choice to search domains by keyword.

The greatest innovation of this platform is that EAWS is able to integrate information from different sources available to date regarding splicing from the existing web servers: Gene Splicer that predicts the 3' and 5' ss strength (Pertea et al., 2001); FAS-ESS that predicts exonic splicing silencer (Wang et al., 2004); ESE RESCUE (Fairbrother et al., 2002); ESEfinder (Cartegni et al., 2003). Regarding the latter, the sensitivity threshold can be adjusted. Furthermore, EAWS contains all Interpro (Hunter et al., 2009) protein domains that have annotated transcripts in Ensembl (Flicek et al., 2010). All genes that have conserved protein domains were fetched from Ensembl and transcripts analyzed for the presence of splicing regulatory sequences. It also allows, at the same time, to align transcripts and protein sequences of similar functional domains (Fig. 3.2). This multiple sequences alignment is done through the use of MUSCLE program. Protein sequence of only one domain instance per gene was selected for an alignment, from a reference transcript of that gene. If there are multiple domain repeats within a gene, only the first repeat was selected. In this way, each gene is represented with a part of its sequences that codes for distinct protein domain. Therefore, the alignment of functional domains from human paralogs would provide ideal candidates in which evolution at the protein level may be hindered by the presence of cis-acting regulatory elements.

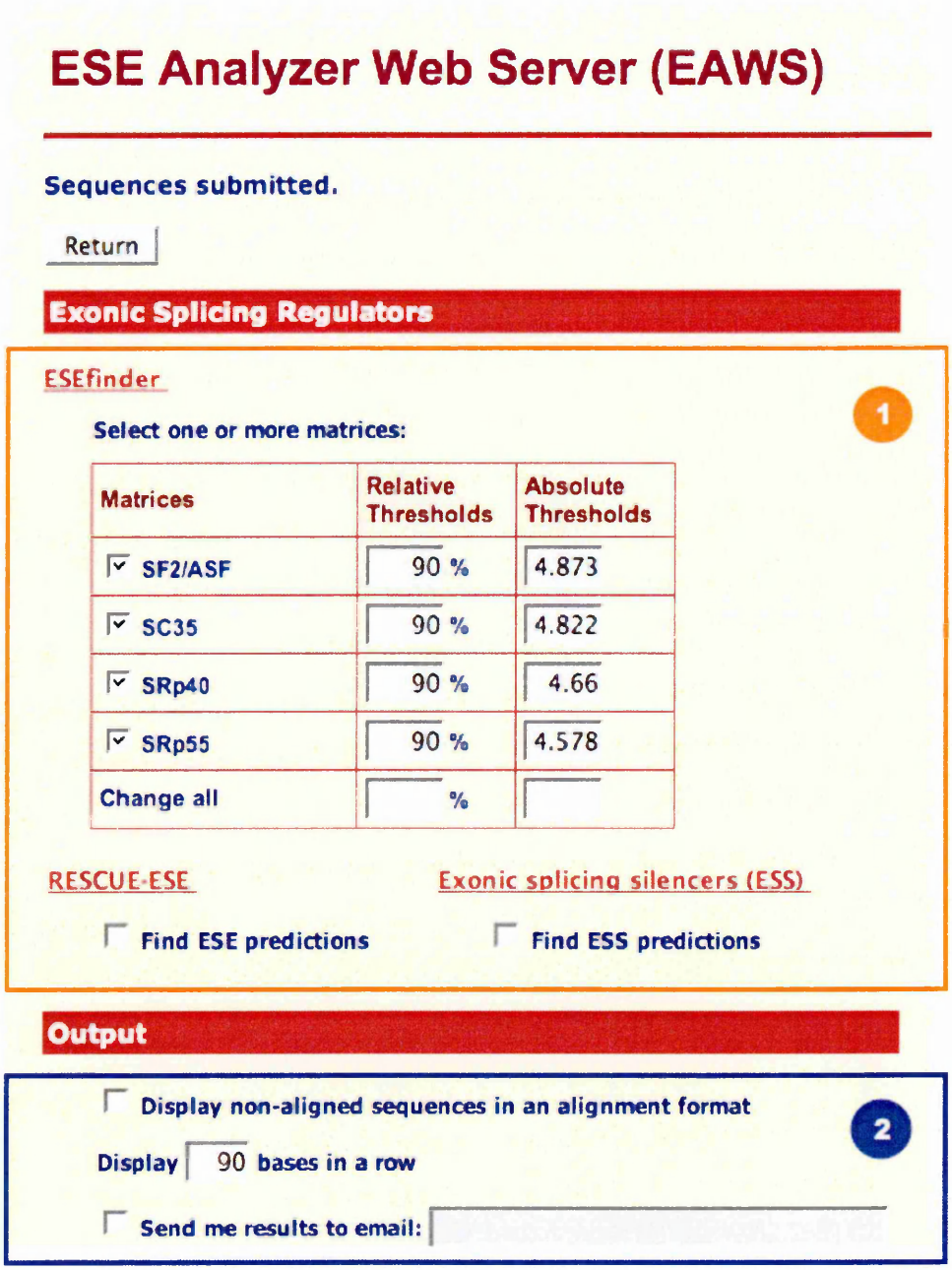


Figure 3.2. Page for selecting options in EAWS. 1- different bioinformatics servers for the analysis of exonic splicing regulators, 2- output options as well as thresholds.

An example of an analysis is shown in figure 3.3. Entering via the Interpro domain identification key IPR023412 (Ribonuclease A domain), EAWS fetch in Ensembl all the transcripts that have RNase A domain. The selected transcripts are analyzed for ESE Finder, ESE-RESCUE and ESS motifs and aligned.

The first line, in green, represents the consensus sequence with above 50% of amino acid conservation. The amino acids and corresponding nucleotide triplets are aligned below.

The numbers at the beginning of each block line indicate the amino acid/nucleotide positions in the corresponding proteins/transcripts. Nucleotides in red are predicted

ESEfinder motifs, underlined are predicted ESE-RESCUE hexanucleotides. In the transcripts, exon borders (the last nucleotide of the preceding exon and the first nucleotide of the next exon) are highlighted in different colours depending of the splice site strength (darker is stronger donor/acceptor); and in protein sequence highlighted in orange is the amino acid principally involved in the active site.

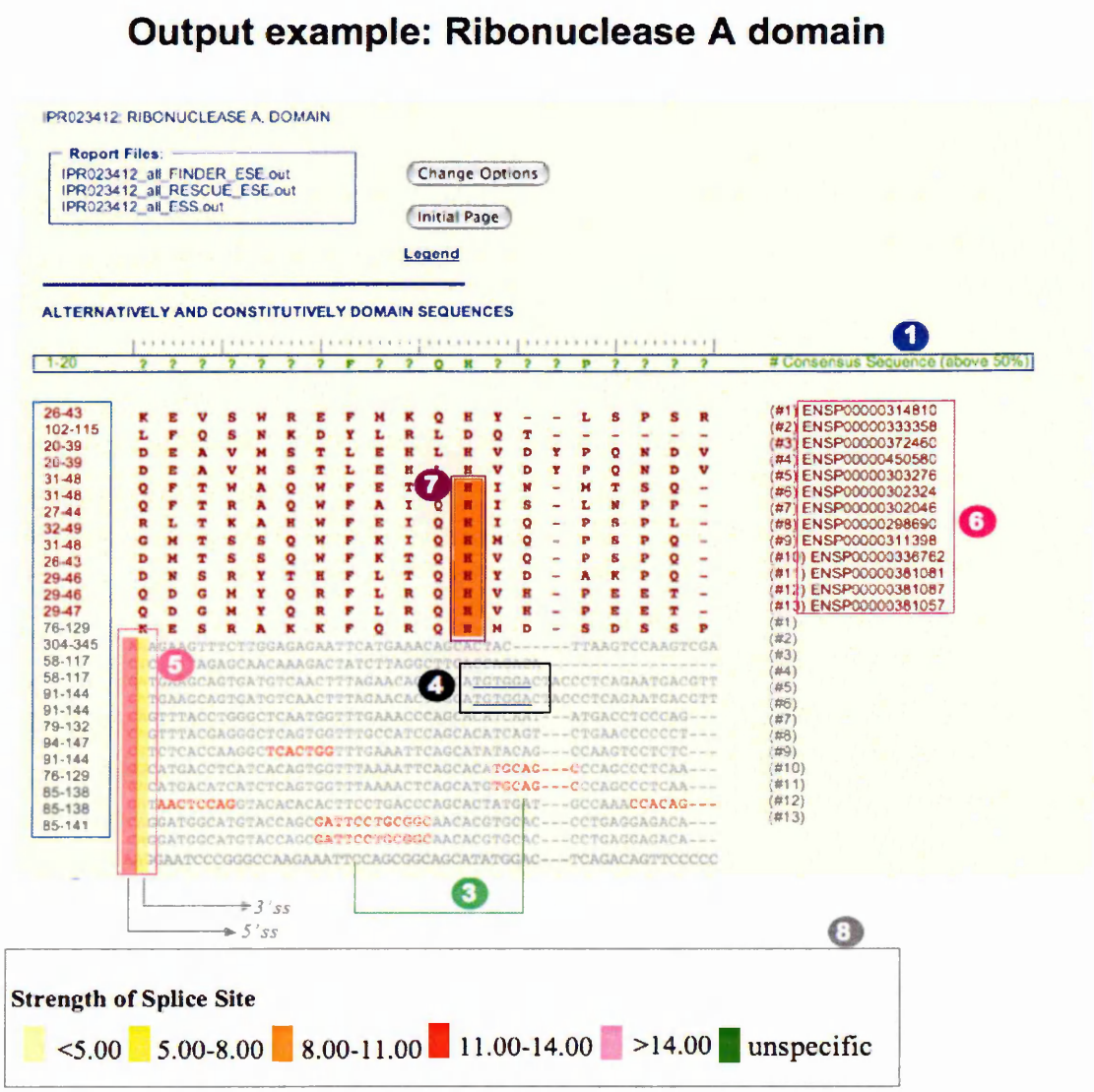


Figure 3.3. A typical output page of EAWS for Ribonuclease A domain. An example on how to read the EAWS output. 1 – consensus sequence above 50% of amino acid conservation; 2 – range of residues in the line for each sequence (The numbers at the beginning of each block line indicate the amino acid/nucleotide positions in the corresponding proteins/transcripts); 3 – predicted ESEfinder motifs; 4 – predicted RESCUE-ESE (underlined); 5 – exon borders (the last nucleotide of the preceding exon and the first nucleotide of the next exon) highlighted in different colors depending of the splice site strength (darker is stronger donor/acceptor); 6 – Protein ID; 7 - the amino acid mainly involved in the active site; 8- Color coding of the output of the splice site strength.

3.1.1 Candidate protein families selected via EAWS

Using EAWS platform several possible candidate genes were found to comply with our search criteria. Figure 3.4 shows three examples that I regarded as the most promising candidates: (A) Vacuolar-type proton pump (H^+-ATPase) G subunit (ATP6V1G), (B) Protein Phosphatase type 2A B′ subunit (PPP2AR) and (C) Alkaline Phosphatase (ALP). Briefly, in each case amino acid differences were observed in family members corresponding to nucleotide changes that in turn created or disrupted ESE and the presence of these ESE could be associated with weak 5' or 3' ss.

Specifically, the multi-subunit *H^+-ATPase* comprises two functional domains, V_1 and V_0. The peripheral V_1 domain binds and hydrolyzes ATP, providing the energy for H^+ translocation across the integral membrane V_0 domain and each domain contain several subunits. In particular, ATP6V1G is referred to H^+-ATPase G subunits localized in the catalytic domain V_1 (Beyenbach and Wieczorek, 2006). In this case, in the exon three of the transcripts ATP6V1G3 (line #3) and ATP6V1G1 (line #5), associated with a weak 3'ss, I observed a correlated increase in the number of predicted ESEs. In ATP6V1G1 there was also an amino acid variation in the area of a predicted ESE encoding for an Arg with respect to the other members where the Gln is highly conserved or as in the case of ATP6V1G3 a predicted ESE encoded for an Ile where in all the other members a Val residue was present (Fig. 3.4A).

The core enzyme of the *Protein Phosphatase type 2A (PP2A)* comprises a dimer (PP2AD), consisting of a catalytic subunit (PP2AC) and a regulatory subunit termed A subunit. A third regulatory B subunit can be associated with this core structure. At present, four different families of B subunits have been identified, termed the B, B′, B′′ and B′′′ families. In particular PPP2AR is referred to the PP2A regulatory B′ subunit (Lechward et al., 2001). In this case, the EAWS analysis showed the absence of a predicted ESEfinder motif in the 3' end of the PPP2R5A exon two (line #2), resulting in Ile amino acid variation where Leu is instead present in all the other members (area 1) and the increase in acceptor

strength in the exon five of the transcript PPP2R5B (line #3) is compensated by the decrease of an ESE (area 2) that was not associated with any amino acid variation (Fig. 3.4B).

However, the most promising candidate protein family were those of the human *Alkaline Phosphatase* paralogous genes (ALPs) (Fig. 3.4C).

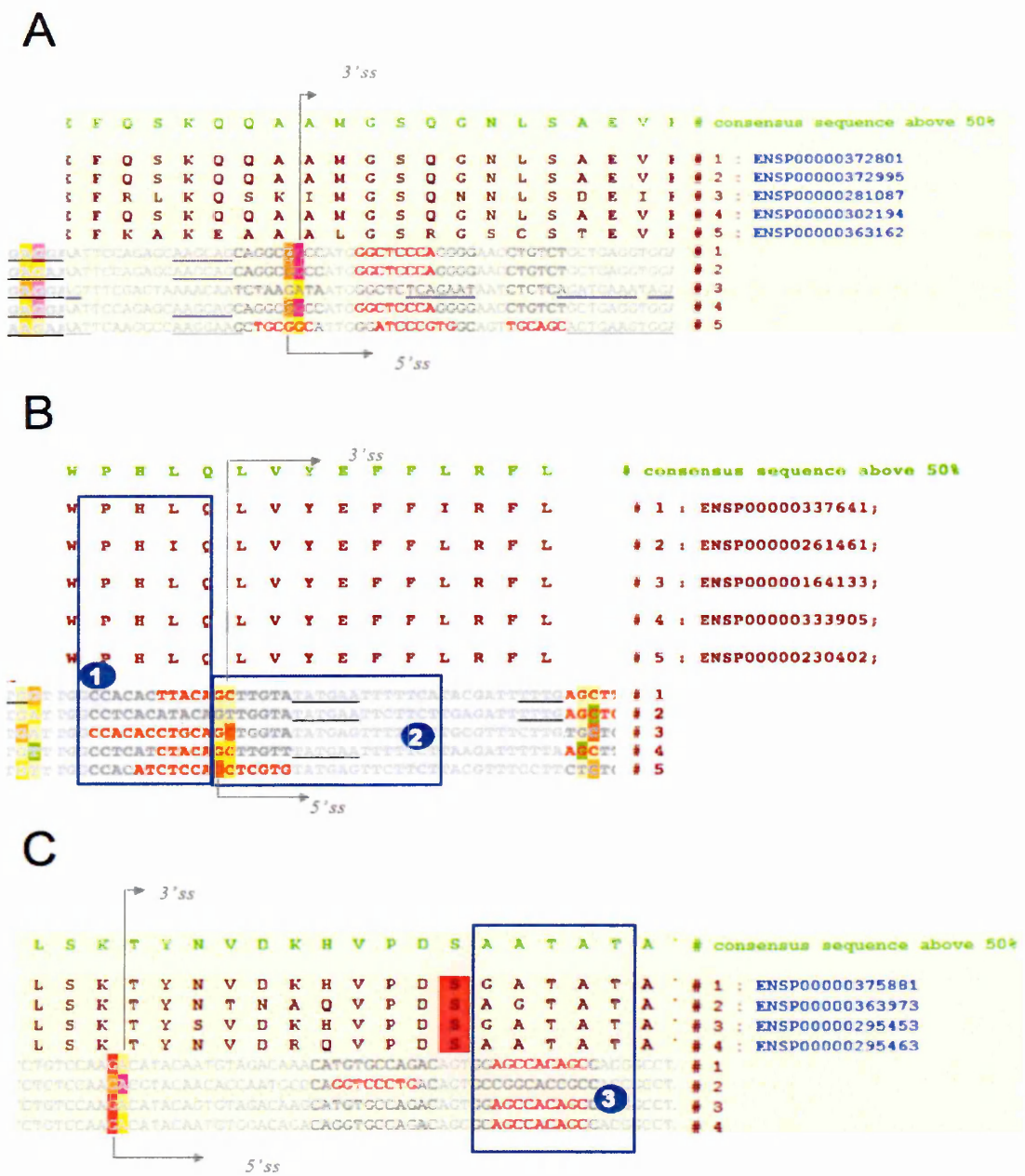


Figure 3.4. Partial alignment of candidate gene families. (A) Partial alignment of Vacuolar (H^+)ATPase G subunit (ATP6V1G) (Interpro domain IPR005124) localized in the catalytic domain V_1. In this case, in the exon three of the transcripts ATP6V1G3 (line #3) and ATP6V1G1 (line #5), associated with a weak 3'ss, there is an increase in the number of predicted ESEs. In ATP6V1G1 there was also an amino acid variation in the area of a predicted ESE encoding for an Arg and in ATP6V1G3 a predicted ESE encoded for an Ile with respect to the other members of the family where the these amino acids are highly conserved. (B) Partial alignment of Protein Phosphatase 2A, regulatory B´ subunit (PP2AR) (Interpro domain IPR002554). In this case, the EAWS analysis showed the absence of a

predicted ESEfinder motif in the 3' end of the PPP2R5A exon two (line #2), resulting in Ile amino acid variation where Leu is instead present in all the other members (area 1) and the increase in acceptor strength in the exon five of the transcript PPP2R5B (line #3) is compensated by the decrease of an ESE (area 2) that was not associated with any amino acid variation. (C) Partial alignment of Alkaline Phosphatase gene family (Interpro domain IPR001952) in which compensatory changes in acceptor strength follow changes in ESE and amino acid variations (area 3) close to the active Serine, in orange (see below).

In humans, ALPs are encoded by four distinct loci, traditionally named after the tissues where they are predominantly expressed (Table 1.2). The placental ALP (ALPP), the germ cell ALP (ALPPL) and the intestinal ALP (ALPI) isozymes are tissue-specific and 90-98% homologous, while the tissue-nonspecific (ALPL) is approx. 50% identical to the other three isozymes (see introduction). From the analysis the ALPs human transcripts, using the ESE Analyzer Web Server, the exon 4 of the tissue- specific genes and the exon 5 of the non-tissue specific isoform, that enclose the active site of the enzymes, match at the amino acid level with a high degree of homology and fulfilled many of my search criteria (Fig. 3.5). Indeed this analysis showed: a weak 3' ss of exon 4 in the tissue specific ALP genes and a strong 3' ss in exon 5 of ALPL gene. The investigation of the cis-acting splicing regulatory elements through ESE RESCUE and ESE Finder, highlighted several elements. As the ESE RESCUE motifs identified were not associated with changes in amino acid composition correlated to the ALP's with the "weak" 3' splice site or where universally present in all the ALPs I decided to concentrate the initial studies on the splicing cis acting regulatory elements identified by ESE finder. In this case ESEfinder identified in correspondence to the ALPP transcript exon 4, that has a weak 3'ss as calculated by Gene Splicer (see Tab. 1.1), two hypothetical ESEfinder motifs, which are absent in the corresponding sequences of the ALPL gene, in particular exon 5, that carries a strong 3'ss (Fig. 3.5, areas 1 and 2). The nucleotide changes in the corresponding region of ALPL exon

5 that resulted in the absence of these hypothetical ESE also resulted in amino acid variations in this region that, in the other members of the family is well conserved. A third putative ESEfinder motif is also present, partially overlapping in all 4 ALPs. There was also an amino acid variation in the area of predicted ESEs of ALPL exon 5 with respect to the other ALPs where the Ala is highly conserved. However, in this scenario there is no connection between presence or absence of ESE with 3' ss strength (Fig. 3.5, area 3). Intriguingly, the amino acids of interest that differ between PLAP and TNAP, in the predicted ESEfinder motifs, may play a role in the catalytic protein function, for the proximity to the active Serine highlighted in orange.

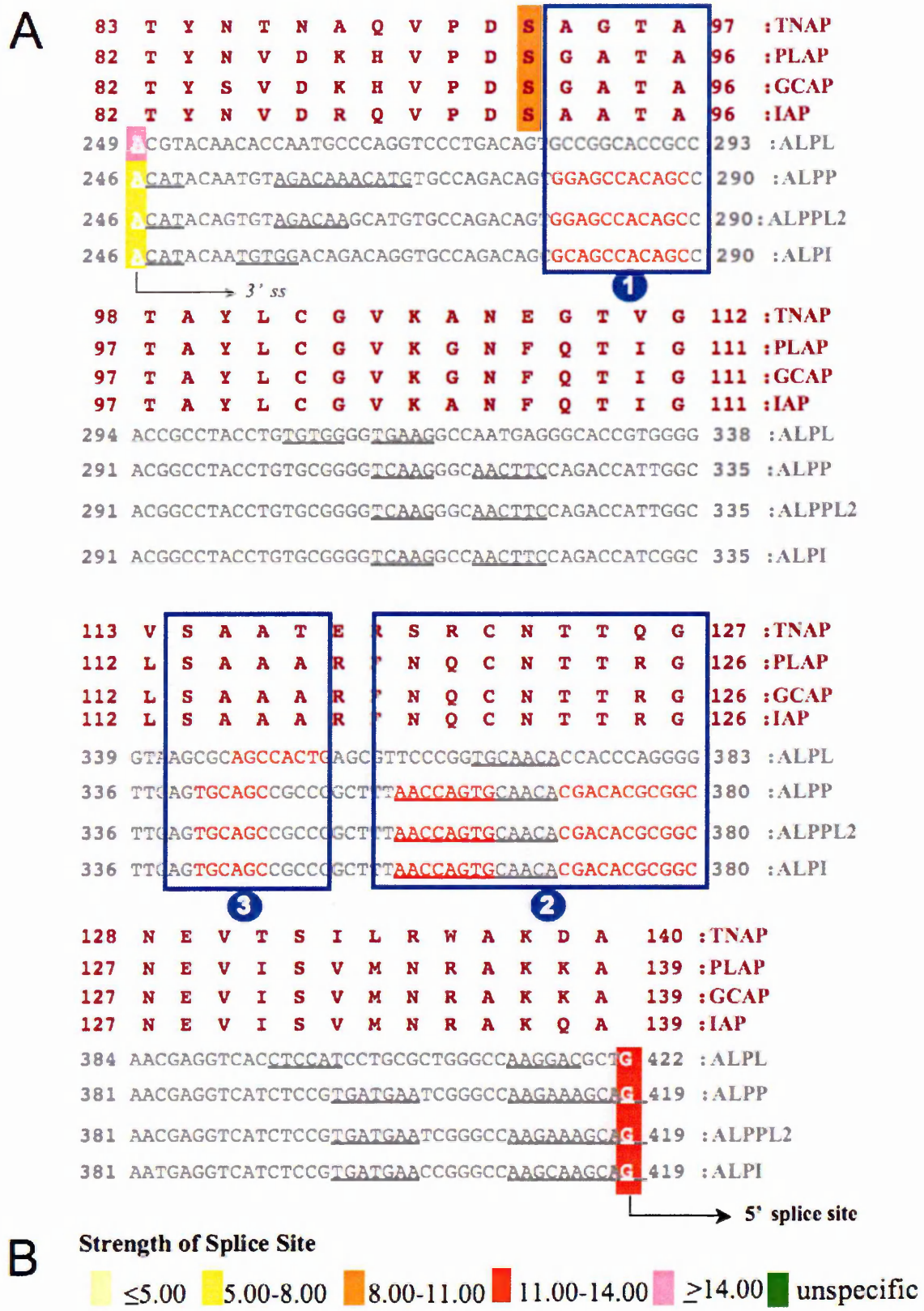


Figure 3.5. Output of EAWS showing the comparative analysis of the exon encoding for the active site of human Alkaline Phosphatase (ALP) family. (A) Alignment of exonic sequences and corresponding amino acidic sequences of human placental (ALPP), germ

cell (ALPPL2), intestinal (ALPI) and tissue non-specific (ALPL). The numbers at the beginning of each block line indicate the amino acid/nucleotide positions in the corresponding proteins/transcripts. Nucleotides in red are predicted ESEfinder motifs and sequences underlined are ESE-RESCUE motifs. The ESE areas and its homologous sequence in ALPL are boxed (1-3). The exon borders are highlighted in different colors depending of the splice site strength (darker is stronger donor/acceptor); (B) Color coding of the output of the splice site strength.

3.2 Analysis of ESE bioinformatics predictions

Since the current understanding of the properties of human ALPs comes from studies using ALPP and ALPL as paradigms, (Le Du and Millan, 2002) I decided to use these in the following study. In order to confirm the bioinformatic predictions I initially set up a minigene assay to test if this was a viable methodology with which to map the presence or lack of ESE in ALPP and ALPL respectively (see material and methods for a detailed description). As it has often being shown that the adjacent flanking sequences are also important for the definition of the exon I decided to create three exon-two intron minigene, thus creating (Buratti et al., 2006) as natural as possible environment for the specific exon under study. I therefore generated the minigenes ALPP and ALPL that spanned exons 3 to 5 and exons 4 to 6 respectively, with the exon under study always being the central one (Fig. 3.6A).

Transfection in HeLa cells followed by RT-PCR analysis showed that in the case of ALPL minigene a unique PCR product of 467 bp is observed. This was identified, through sequencing, as normal inclusion of exon 5 in the mature transcript (Fig. 3.6B, Lane 1). The RT-PCR of the ALPP minigene carrying the exon 4 with its flanking intronic and exonic sequence, showed a major amplicon of 464 bp, corresponding to the normal inclusion of the exon 4 and a very minor amplicon of higher molecular weight (Fig. 3.6B, Lane 2).

Sequencing of the cDNA fragment showed that the minor fragment was composed of a complete inclusion of the intron 4.

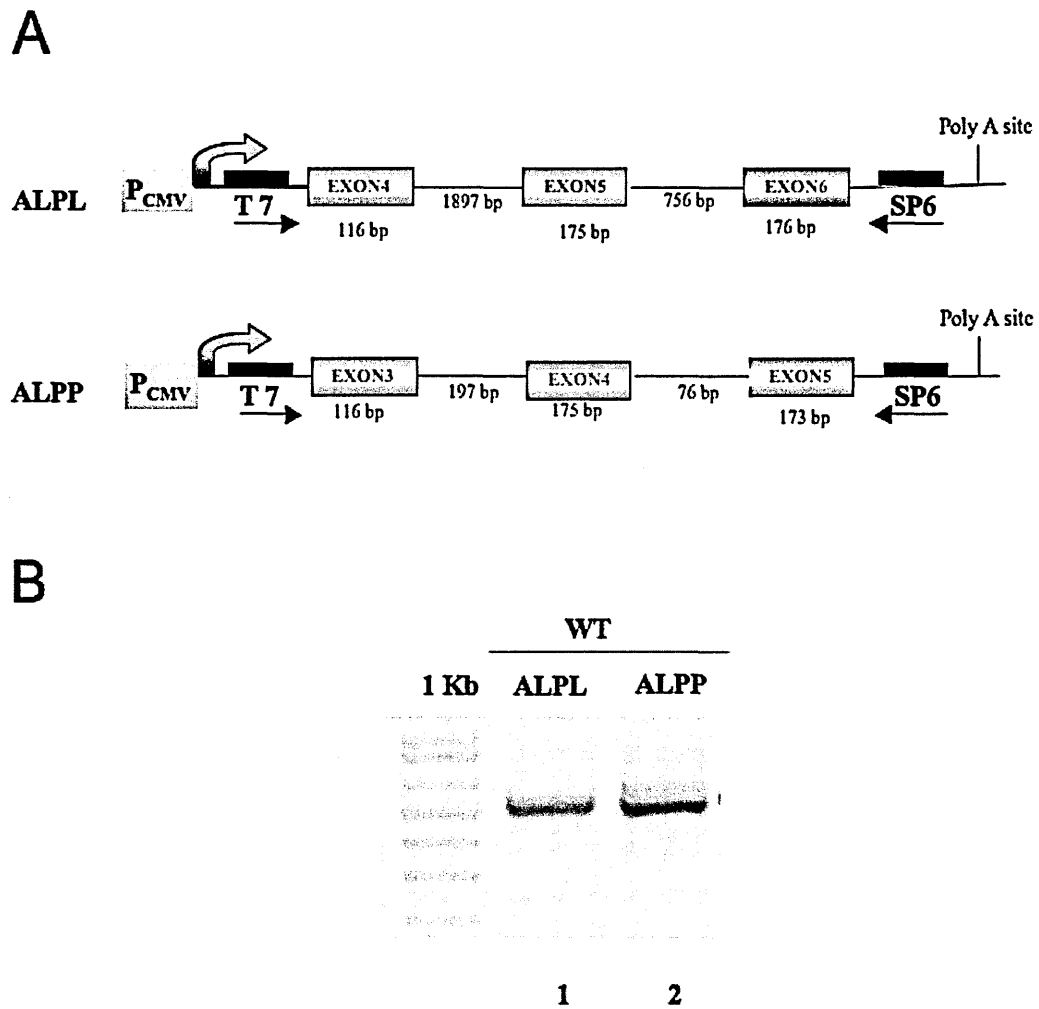


Figure 3.6. Wild type splicing patterns of ALPP and ALPL minigenes. (A) Schematic representation of the wild type hybrid minigenes used in transfection experiments. The exons of interest are shown in green. (B) The amplified RT–PCR products stained with ethidium bromide after transfection in HeLa cells. RT-PCR analysis of ALPL minigene showed a unique PCR product of 467 bp corresponding to normal inclusion of exon 5 in the mature transcript (Lane 1). The RT-PCR of the ALPP minigene carrying the exon 4 with its flanking intronic and exonic sequence, showed a major amplicon of 464 bp, corresponding to the normal inclusion of the exon 4 and a very minor amplicon of higher molecular weight produced by the retention of the intron 4. A faint band at about 700 bp is most probably due to contamination of plasmid DNA (Lane 2).

Although minigenes represent a useful tool with which to analyze splicing mechanisms, they can result in artificial outcomes due to the reduced amount of gene sequence utilized. Regarding this issue, previous studies have clearly shown that apparent discrepancies in comparisons between minigenes and endogenous situations can sometimes be due to long distance interactions and the balance of these controls the correct splicing outcome. In particular, it has been shown that in the NF1 (neurofibromatosis 1) gene the genomic context plays a very important role in explaining apparent discrepancies between minigenes and endogenous situations (Baralle et al., 2006). Moreover, using hybrid minigene experiments, it has been shown that changes in promoter, in this case using CMV promoter of the pcDNA3 minigene instead of the endogenous, could strongly affect splice site selection (Kornblihtt, 2005; Pagani et al., 2003b).

In order to establish whether the retention of intron 4 in ALPP minigene occurs also endogenously in human placental tissue or if it represented an artefact of the minigene system, I analyzed the endogenous splicing pattern of the exon 4 from placental tissue. As can be seen in figure 3.7 the RT-PCR did not reveal any retention of the intron 4 *in vivo,* but a unique PCR product of 464 bp, corresponding to the normal processing. Based on the assumption that this intron retention was an artefact of the minigene system, the corresponding splicing product was not considered in further studies.

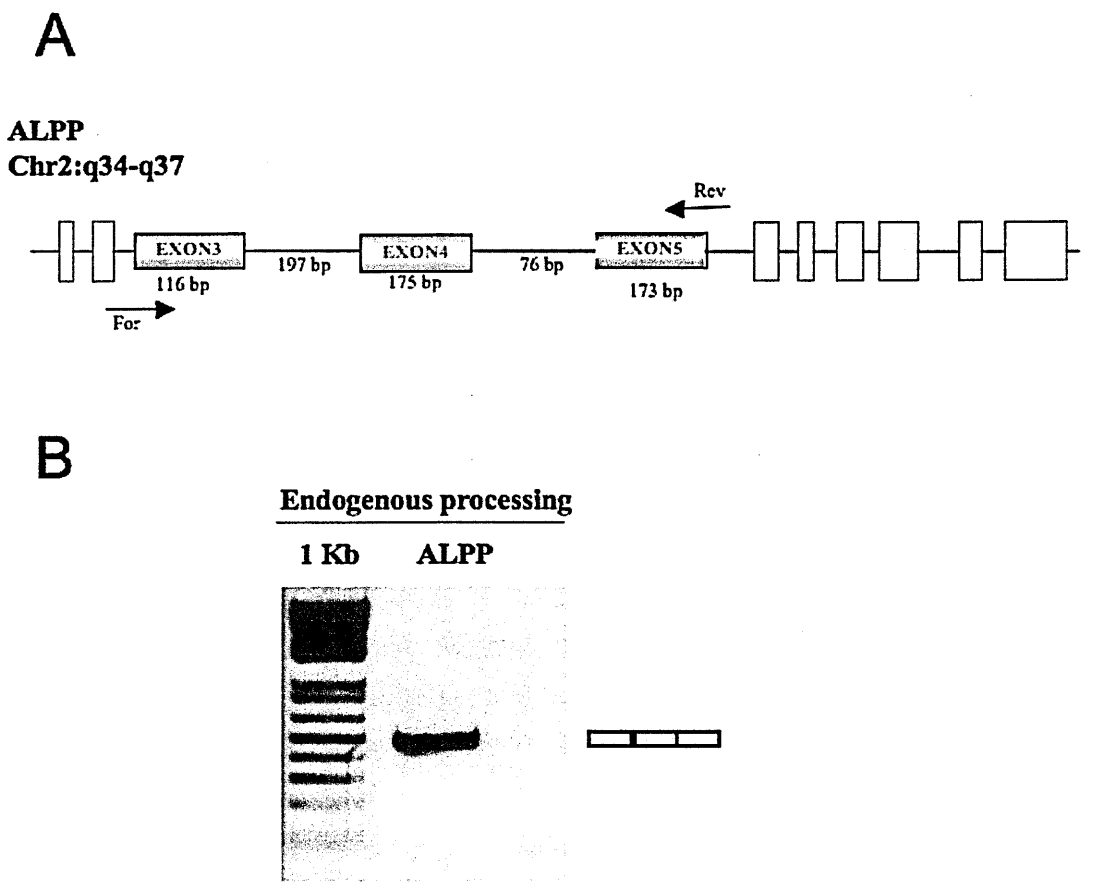


Figure 3.7. Wild type splicing patterns of endogenous ALPP exon 4 processing. (A) Schematic representation of the amplified portion from the endogenous transcript. (B) The amplified RT–PCR products from human placental RNA stained with ethidium bromide. The RT-PCR did not reveal the retention of the intron 4 *in vivo,* but a unique PCR product of 464 bp, corresponding to the normal processing.

3.2.1 Analysis of ALPP exon 4 processing after mutations in two regions where the presence of ESE is predicted and associated with weak 3' splice sites

From EAWS bioinformatics analysis of ALPP and ALPL transcripts, the ALPP exon 4 showed predicted ESE motifs that were absent in corresponding sequences of ALPL exon 5, these differences were also associated with amino acidic variations close to the active site of the enzymes (Fig. 3.5, areas 1 and 2). I started the investigation verifying the predictions of the bioinformatics analysis. A schematic representation of the central part of the wild type (WT) minigenes, the exon 4 and its flanking intronic region in the case of ALPP minigene and exon 5 and its surrounding intronic sequences for ALPL construct is shown in figure 3.8A. The analysis of ALPP exon 4 for putative ESE with ESEfinder highlighted two regions containing a total of 3 putative ESE associated with a weak 3' ss as well as with amino acids changes that were not detected in the analogous region of ALPL (1^{st} seq and 2^{nd} seq) exon 5 that carried a strong 3' splice site. The 2^{nd} ESE region contained two putative enhancers, however, given their proximity this area was initially analyzed as one. To test the bioinformatic predictions I decided to generate minigenes where the regions corresponding to the ALPP 1^{st} and 2^{nd} putative ESE were exchanged individually with the corresponding ALPL sequences (1^{st} seq and 2^{nd} seq) (Fig. 3.8A, MUT1-2), which were predicted not to have ESEs. Minigenes were then transfected in HeLa cells. RNA was extracted and than subjected to RT-PCR analysis. The rationale behind this experiment was that if it was indeed the case that the ESEs in ALPP were needed for correct splicing of this exon, and that if this was not the case in the corresponding region of ALPL, an aberrant splicing of ALPP exon 4 would be observed.

As can be seen in Figure 3.8B the replacement of the 1^{st} and 2^{nd} predicted ESEs with the corresponding area in ALPL resulted in a decrease in exon definition with a partial exon 4 skipping in both cases. These results confirmed the bioinformatics prediction, that these two regions of the exon 4 contain enhancer elements responsible for promoting the exon

inclusion in HeLa cells and it is for such reason that I will refer them to as the 1st and 2nd ESE element regions.

To see if the ESEs were acting in union I decided to create a minigene in which I exchanged the 1st and 2nd ESE element regions for the corresponding region in ALPL (1st seq and 2nd seq) simultaneously, creating the minigene MUT3 (Fig. 3.8A). Transfection into HeLa cells, followed by RT-RCR analysis and agarose gel electrophoresis, showed that exon 4 skipping increases drastically when both ESE element regions were substituted for the corresponding region of ALPL, hence demonstrating a cumulative effect of ESE sequences present in these two areas (Fig. 3.8C).

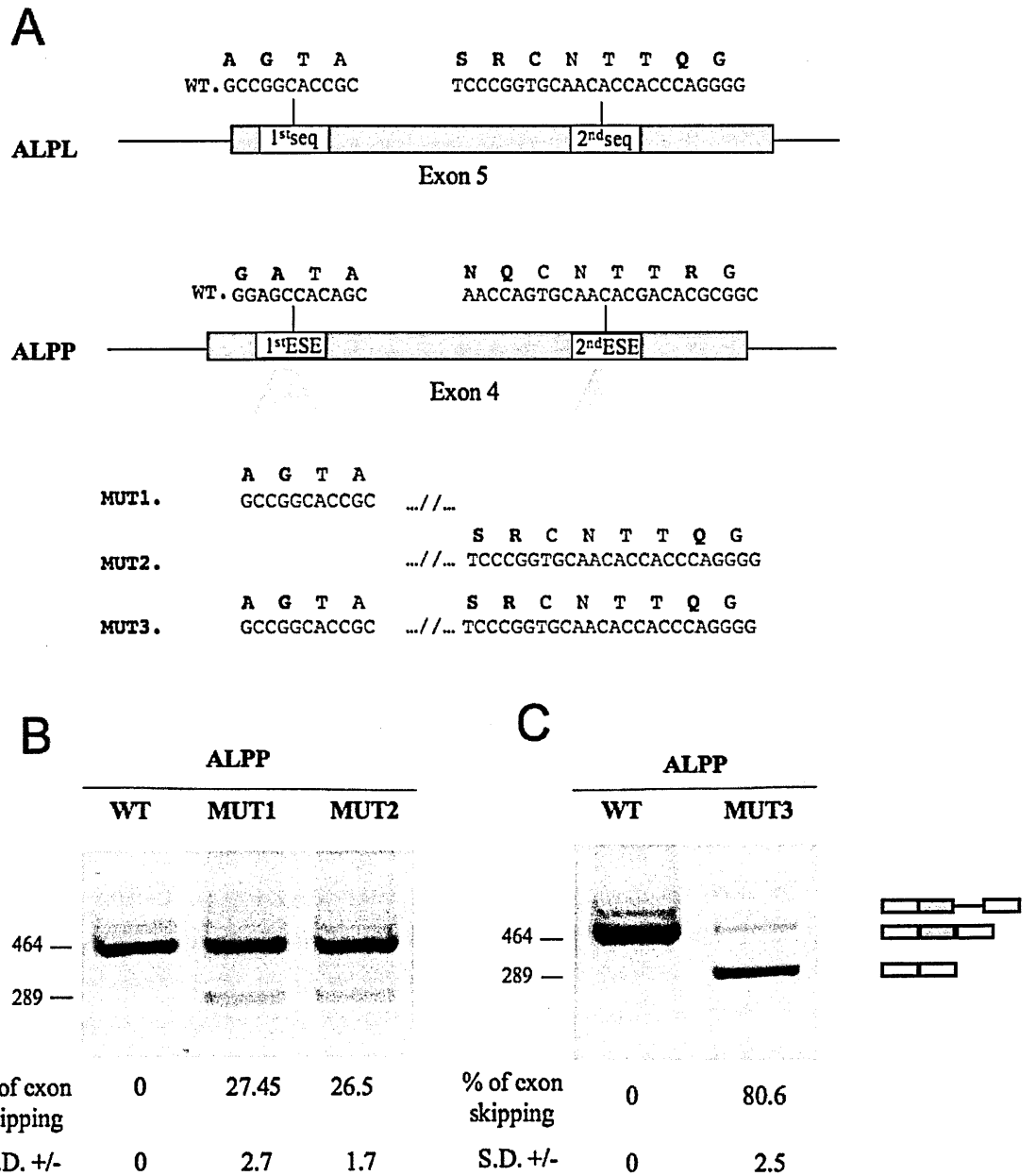


Figure 3.8. Analysis of ALPP exon 4 splicing after swapping the regions 1st and 2nd ESE with that of ALPL. (A) A schematic representation of the central part of the wt ALPP and ALPL constructs, the exon of interest is shown in green and its flanking intronic region as black line and the investigated wt sequences are represented above the exon (1st and 2nd seq in ALPL exon 5; 1st and 2nd ESE in ALPP exon 4). The sequences reported nearby the mutants (MUT1-3) show the substitutions performed on the minigenes. (B) The amplified RT–PCR products stained with ethidium bromide after transfection in HeLa cells. The

inversion of the 1st and 2nd predicted ESEs with the corresponding area in ALPL resulted in a decrease in exon definition with a partial exon 4 skipping. (C) Transfection into HeLa cells, followed by RT-RCR analysis and agarose gel electrophoresis, showed that exon 4 skipping increases when both ESE sequences were substituted for the corresponding region of ALPL. On the right hand site of the gels a schematic representation of the splicing product obtained can be observed. The percentage of skipping and standard deviation (SD) are indicated below each lane and represent the mean of three experiments.

Afterward, I decided to investigate also if 1st and 2nd seq element in ALPL could differentially affect the efficiency of the exon 5 inclusion when exchanged for the corresponding sequence of ALPP. Therefore, I generated two minigines in which 1st and 2nd seq element were separately replaced with the 1st and 2nd ESE in the ALPL minigene and a minigene in which I exchanged the 1st and 2nd seq element regions for the corresponding region in ALPP (1st and 2nd ESE) simultaneously (Fig. 3.9A). The substitution of the 1st and 2nd seq, in the exon 5, with the 1st and 2nd ESE elements, as expected, did not compromise the correct inclusion of the exon 5 in the mRNA (Fig. 3.9B).

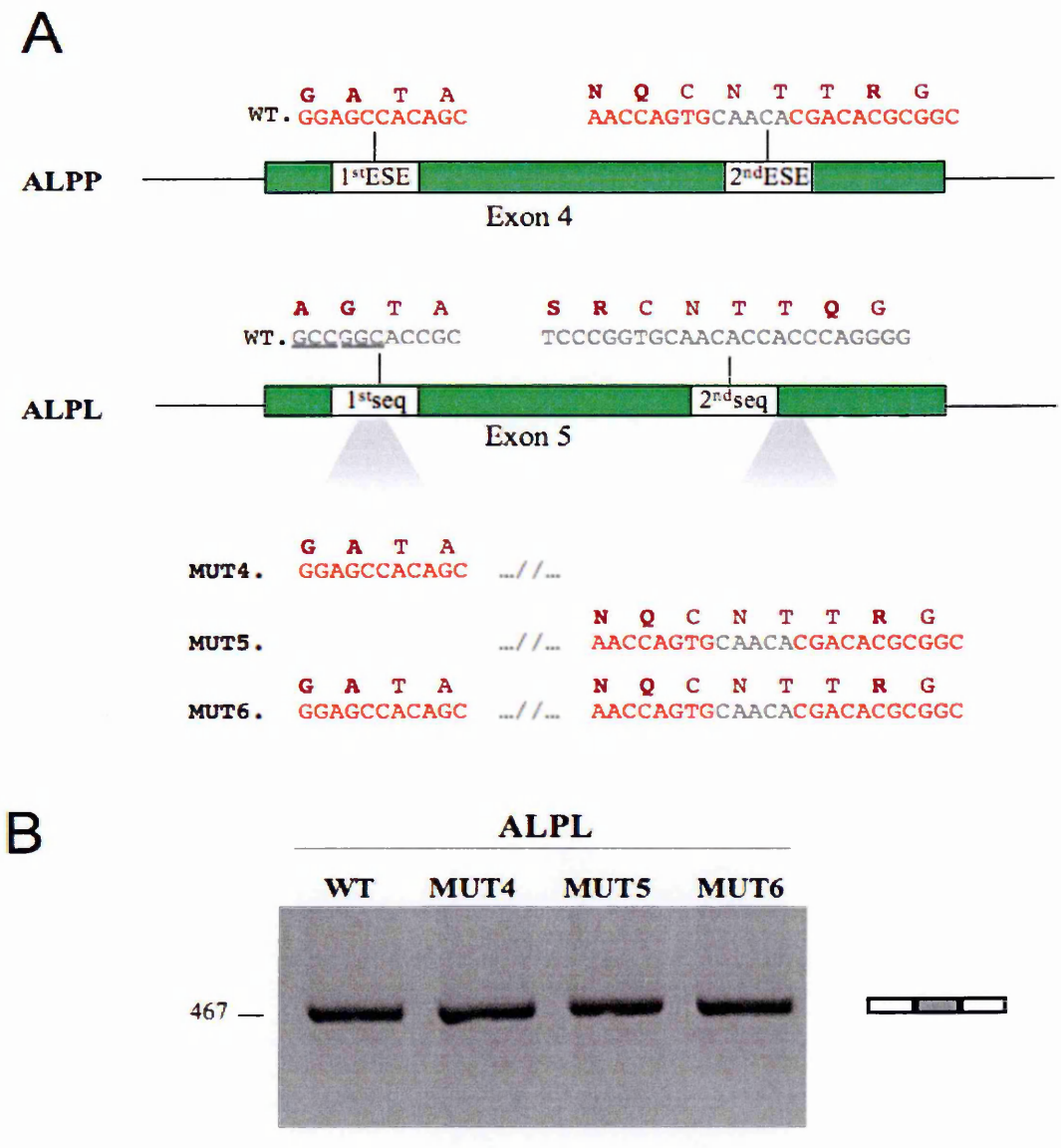


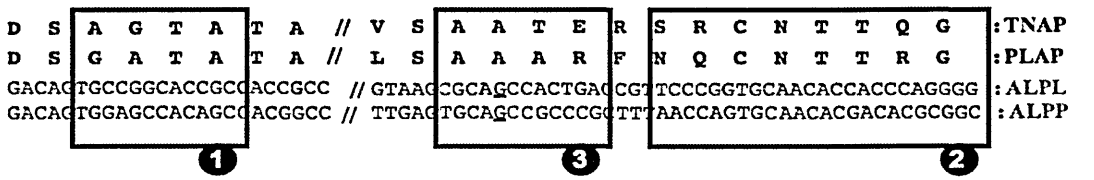
Figure 3.9. Analysis of ALPL exon 5 splicing after swapping the regions 1st and 2nd seq with that of ALPP. (A) A schematic representation of the central part of the wt ALPP and ALPL constructs, the exon of interest is shown in green and its flanking intronic region as black line and the investigated wt sequences are represented above the exon (1st and 2nd seq in ALPL exon 5; 1st and 2nd ESE in ALPP exon 4). The sequences reported nearby the mutants (MUT4-6) show the substitutions performed on the minigenes. The constructs are analyzed with pcDNA3 minigene system in HeLa cells. (B) RT-PCR performed from RNAs isolated from HeLa cells previously transfected with the mutated minigene. In

MUT4-6 the inversion of the 1st and 2nd seq with the corresponding area in ALPP did not compromise the correct inclusion of the exon 5 in the mRNA (n=3). On the right hand site of the gels a schematic representation of the splicing product obtained can be observed.

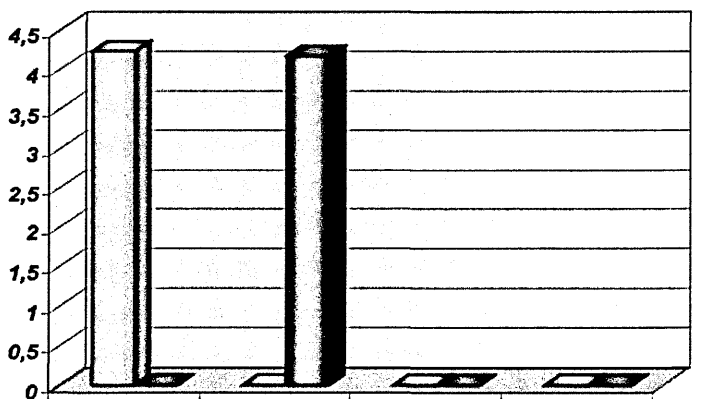
3.2.2 Experimental validation of ESE present in both ALPP and ALPL exon 4 and 5 respectively

As shown in figures 3.5 and 3.10A the bioinformatics analysis also identified a putative ESE sequence through ESEfinder in ALPP exon 4 in addition to those present in the 1st and 2nd ESE region analyzed above. Although the region covered by the ESE was also associated with an amino acid difference between the tissue specific and non tissue specific ALP in the latter, ALPL, the putative ESE was found to partially overlap with a predicted ESEfinder motif but in this case the predicted enhancer makes no distinction between presence or absence of ESE with 3' ss strength. Even though the search criteria, in this case, were not completely fulfilled I investigated the functional role of these putative ESE in the inclusion of ALPP exon 4, disrupting this putative ESE motif and analyzing whether this mutation affects splicing. In this case simple convers ion of the sequence for that of ALPL would not provide the answer we were looking for as an ESE also exists in this region. Using ESEfinder prediction I found that a single base deletion could cause the loss of ESEfinder motif in both ALPL and ALPP. Indeed, it showed that after G nucleotide deletion the sequences are no more recognized as enhancer (Fig. 3.10B).

A



B



	ALPL 1st putative ESE	ALPP 3rd putative ESE	ALPL 1st putative ESE ΔG	ALPP 3rd putative ESE ΔG
SRp55	4,23	0	0	0
SRp35	0	4,17	0	0

Figure 3.10. Putative ESE present in both ALPP and ALPL exon 4 and 5 respectively. (A) Alignment of exonic sequences and corresponding amino acid sequences of human placental (ALPP) and tissue non-specific (ALPL). Overlapping predicted ESE in paralogous genes are boxed. Nucleotides in red are predicted ESEfinder motifs. (B) Results from analysis with ESEfinder web-resource (http://rulai.cshl.edu/tools/ESE/). The y-axis is the numerical-scale for the ESE score; x-axis shows the four different sequences analyzed. The ESEfinder prediction shows that after nucleotide deletion (ΔG; above underlined) the sequences are no more recognized as enhancer sequence.

Therefore, I created a series of constructs with deletion of the overlapping G nucleotide (TGCAGC, underlined nucleotide) in the ALPP minigene (Fig.13.11A, MUT7-10). In particular, MUT7 carried only ΔG mutation in ALPP wt minigene context and was made to see whether disrupting this putative enhancer could have an effect in the exon 4 pre-mRNA splicing. The MUT8, 9 and 10 were made on the base of the previous mutated minigenes MUT1, 2 and 3, respectively, with the addition of G nucleotide deletion in the 3rd putative ESE (Fig. 3.11A). After RT-PCR analysis, I observed that the MUT7, 8, 9, and 10 had no consequences on the inclusion of ALPP exon 4 in the mRNA (Fig. 3.11B).

As this deletion, did not have any effect on the inclusion of exon 4, I therefore decided to not consider this putative ESE sequence as important in the exon 4 inclusion and was not further investigated.

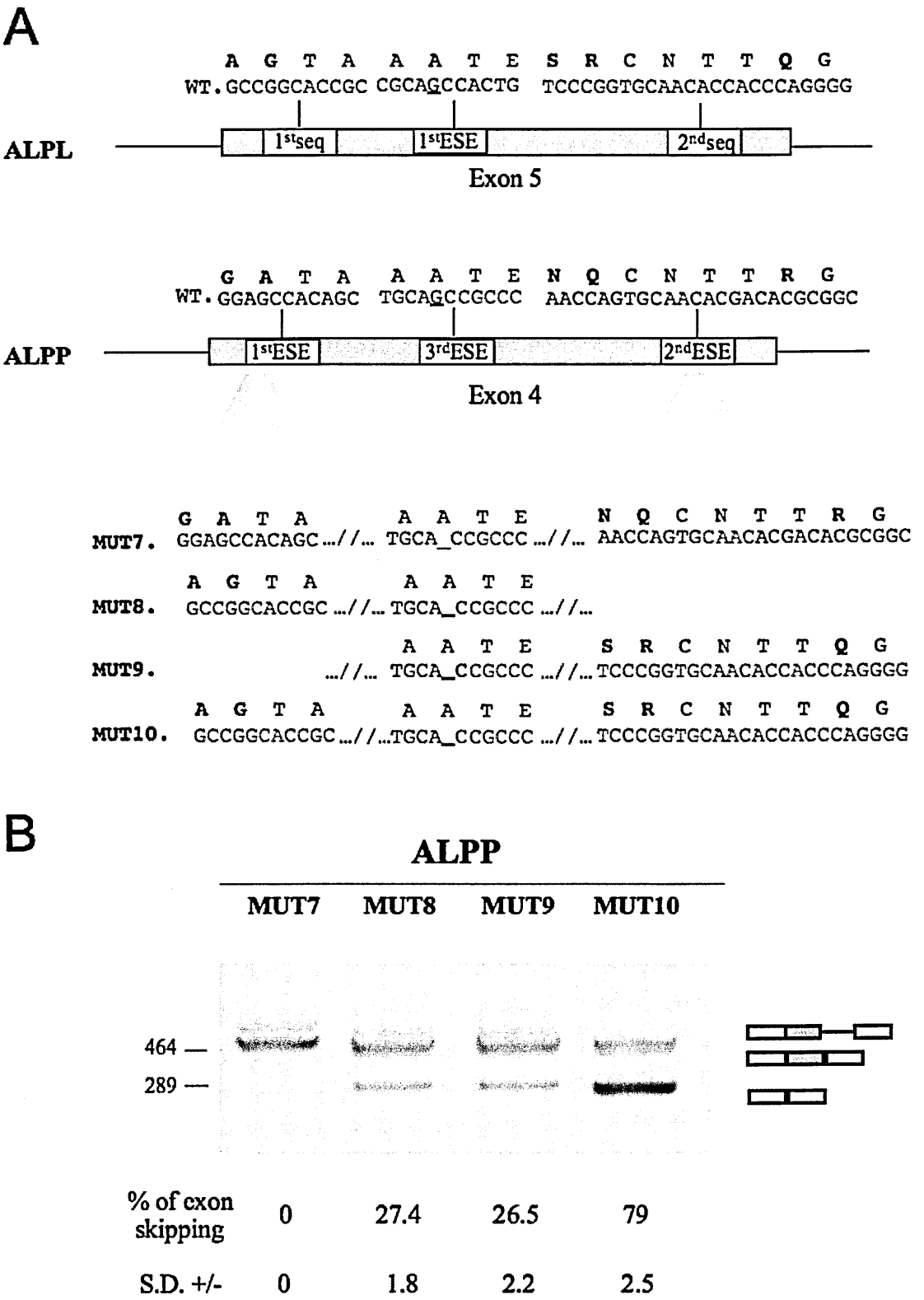


Figure 3.11. Analysis of ALPP exon 4 splicing after ΔG mutation in the 3rd putative ESE. (A) A schematic representation of the central part of the wt ALPP and ALPL constructs, the exon of interest is shown in green and its flanking intronic region as black line and the

investigated wt sequences are represented above the exon (1st and 2nd seq in ALPL exon 5; 1st and 2nd ESE in ALPP exon 4). The sequences reported nearby the mutants (MUT7-10) show the substitutions performed on the minigenes. In particular, the nucleotide deletion in the 3rd putative ESE sequence is underscored in the sequences overhead. (B) The splicing pattern observed upon transfection of these constructs in HeLa cells. The ΔG mutation did not have any effect on the inclusion of exon 4. On the right hand site of the gel a schematic representation of the splicing product obtained can be observed. The percentage of skipping and standard deviation (SD) are indicated below each lane and represent the mean of three experiments.

3.2.3 Fine mapping of the ESE

Having established the presence of the ESEs motifs in the ALPP exon 4 and their absence in the corresponding sequence of ALPL I decided to further define these regions, to know which amino acid changes were directly linked to ESE activity as the first mapping was performed in a broad fashion as a first approach and contained 5 amino acid differences (Fig. 3.5). The 1st and 2nd ESE regions were therefore further analyzed by exchanging each nucleotide triplet with the corresponding nucleotides encoding for the amino acidic observed in ALPL. Initially, given that the 1st ESE region presents two different amino acidic changes in ALPL, Gly93 and Ala94, when compared to the corresponding region in ALPL (1st seq) I created a series of minigenes in which I mutated each triplet encoding for glycine 93 or alanine 94 into their counterparts in ALPL, alanine and glycine codons (underlined in figure 3.12A), as well as a third minigene in which both triplets were mutated (Fig. 3.12A, MUT11-13). Transfection and RT-PCR of these constructs showed in all cases a partial exon 4 skipping (Fig. 3.12B), indicating that both triplets in the 1st ESE region are necessary for the correct processing of the exon 4. Considering the fact that these triplets are adjacent it is very likely that they are part of the same ESE element.

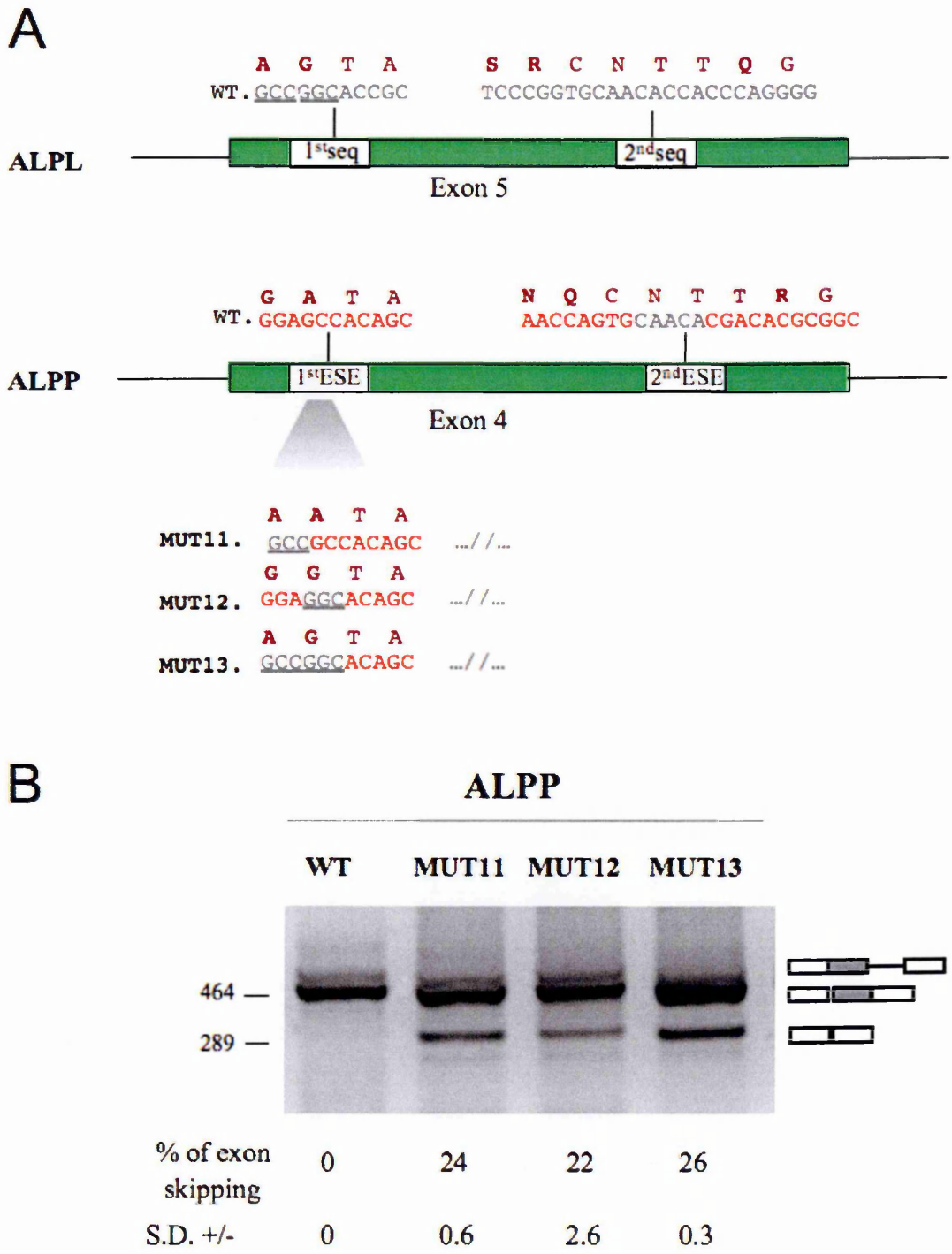


Figure 3.12. Analysis of ALPP exon 4 splicing after mutations in the 1st ESE codons that lead to amino acid difference in ALPL. (A) A schematic representation of the central part of the wt ALPP and ALPL constructs, the exon of interest is shown in green and its flanking intronic region as black line and the investigated wt sequences are represented above the exon (1st and 2nd seq in ALPL exon 5; 1st and 2nd ESE in ALPP exon 4). The sequences reported nearby the mutants (MUT11-13) show the substitutions performed on the

minigenes. In particular, the nucleotide changes in the 1st ESE sequence are underlined in the sequences overhead. (B) The splicing pattern observed upon transfection of these constructs in HeLa cells. MUT11-13 contribute to alter the splicing pattern of the exon 4. On the right hand site of the gel a schematic representation of the splicing product obtained can be observed. The percentage of skipping and standard deviation (SD) are indicated below each lane and represent the mean of three experiments.

The ALPP 2nd ESE region is, instead, associated with three amino acidic changes in the triplets encoding for Asn119, Gln120 and Arg125 when compared to the equivalent region in ALPL (2nd seq). Therefore, as I did for the 1st ESE region, I investigated the effect of the nucleotide triplet differences also for the 2nd ESE region. I converted the single codons encoding for asparagine 119, glutamine 120 and arginine 125, into serine, arginine and glutamine (underlined in figure 3.13A), respectively, by site directed mutagenesis, giving rise to the new minigene constructs (Fig. 3.13A, MUT14-16). Interestingly, transfection followed by RT-PCR analysis showed that, among these three constructs, only the triplet corresponding to the arginine 125 affects the splicing in this region (Fig. 3.13B). Thus only this triplet was considered to be associated with ESE function.

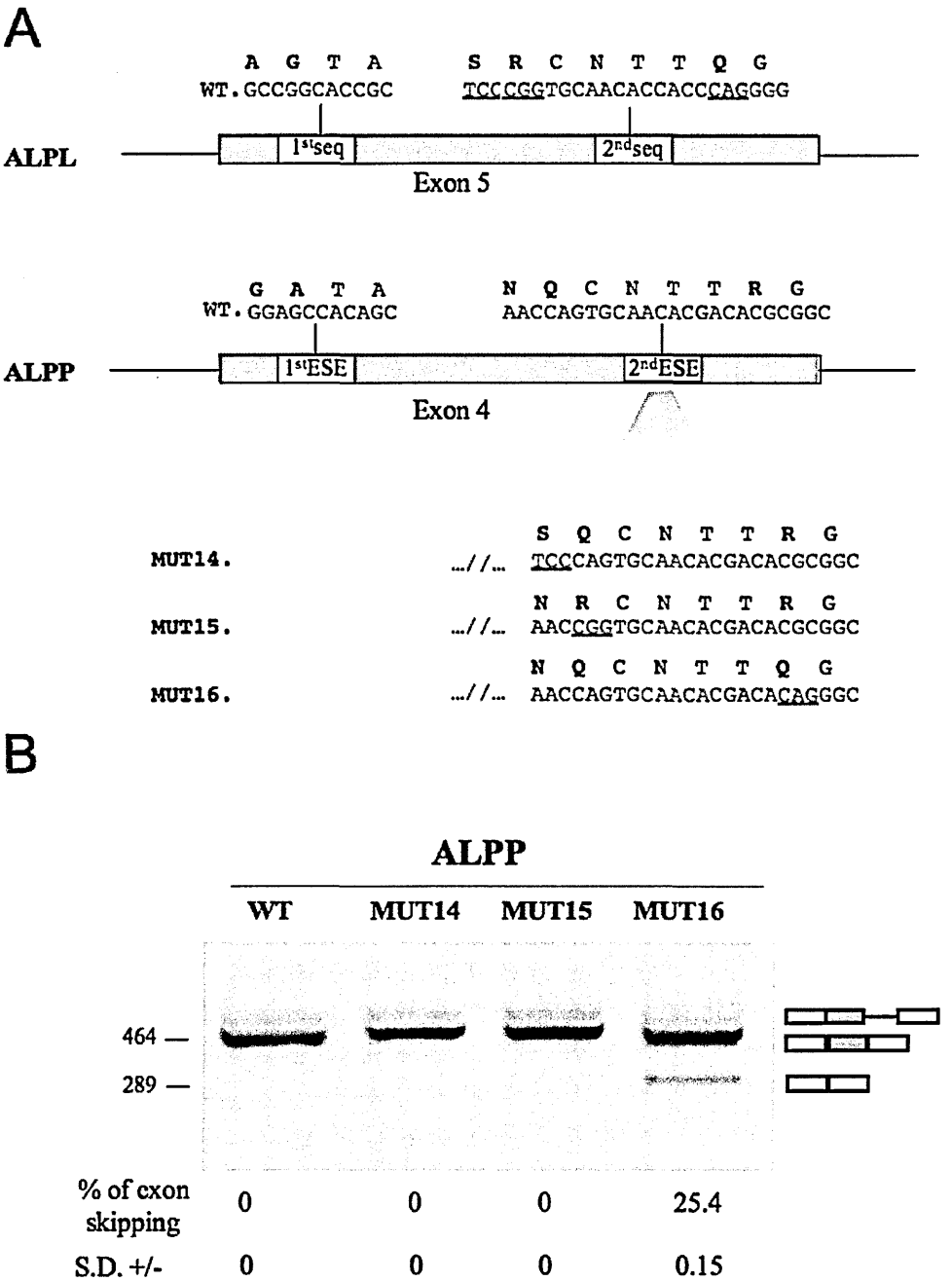
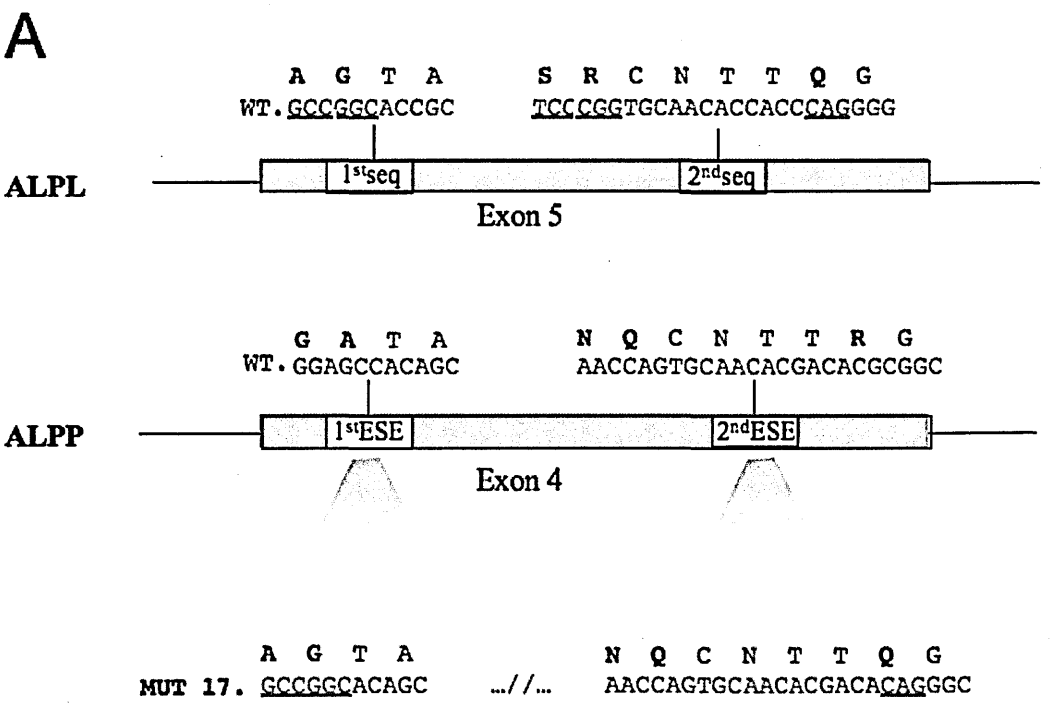


Figure 3.13. Analysis of ALPP exon 4 splicing after mutations in the 2nd ESE codons that lead to amino acid difference in ALPL. (A) A schematic representation of the central part of the wt ALPP and ALPL constructs, the exon of interest is shown in green and its flanking intronic region as black line and the investigated wt sequences are represented above the exon (1st and 2nd seq in ALPL exon 5; 1st and 2nd ESE in ALPP exon 4). The sequences reported nearby the mutants (MUT14-16) show the substitutions performed on the minigenes. In particular, the nucleotide changes in the 2nd ESE sequence are underlined in the sequences overhead. (B) The splicing pattern observed upon transfection of these constructs in HeLa cells. RNA splicing variants corresponding to ALPP exon 4 inclusion

and exclusion are shown. Only MUT16 affects the splicing in this region. On the right hand site of the gel a schematic representation of the splicing product obtained can be observed. The percentage of skipping and standard deviation (SD) are indicated below each lane and represent the mean of three experiments.

To address the question of whether the synergic effect of the two ESE elements still occurs as was observed with ESE region 1 and 2, see figure 3.8 MUT3, I examined the effect that these finer mapped enhancer sequences have in combination on the splicing efficiency of the ALPP exon 4. The minigene used to do this had the two triplets encoding for Gly93, Ala94 and the triplet encoding for the Arg125 exchanged for the corresponding triplets in ALPL (Fig. 3.14A, MUT17). As shown in Fig. 3.14B, the mRNA processing showed that the MUT17 now behaves as the MUT3 minigene, displaying an exon skipping of about 80%. Also in this case the cooperative effect of enhancer sites is maintained for the specific recognition of the exon, leading to a correct pre-mRNA splicing.



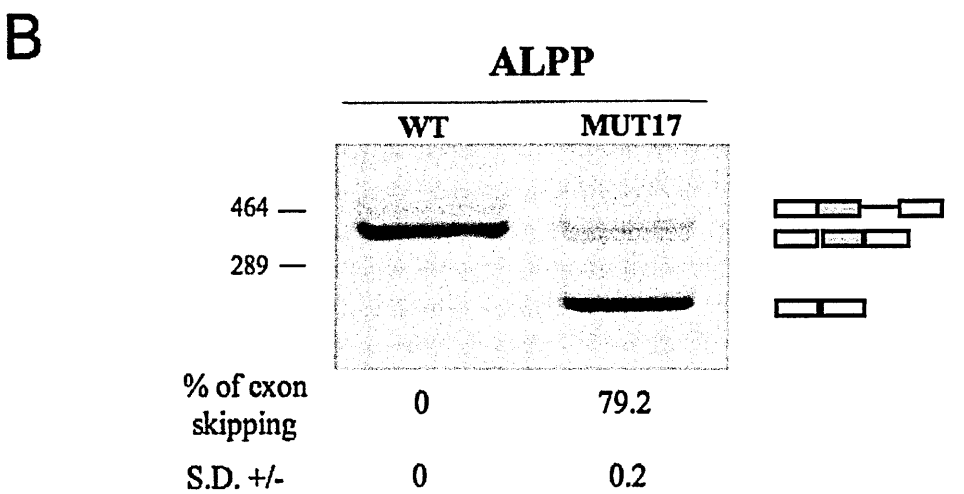


Figure 3.14. Analysis of ALPP exon 4 splicing after swapping of redefined regions 1st and 2nd ESE with that of ALPL associated to amino acidic differences. (A) A schematic representation of the central part of the wt ALPP and ALPL constructs, the exon of interest is shown in green and its flanking intronic region as black line and the investigated wt sequences are represented above the exon (1st and 2nd seq in ALPL exon 5; 1st and 2nd ESE in ALPP exon 4). The sequences reported nearby the mutant (MUT17) show the substitution performed on the minigene. The nucleotide changes in the 1st and 2nd ESE

sequence are underlined in the sequences overhead. (B) RT-PCR products after the transient transfection of the minigenes constructs. RNA splicing variants corresponding to ALPP exon 4 inclusion and exclusion are shown. MUT17 behaves as the MUT3 minigene, displaying an exon skipping of about 80%. On the right hand site of the gel a schematic representation of the splicing product obtained can be observed. The percentage of skipping and standard deviation (SD) are indicated below each lane and represent the mean of three experiments.

3.2.4 The ESEs in ALPP are necessary due to a non consensus 3' ss

It has been assumed that one function of ESEs is the recruitment of spliceosomal components to a weak splice site of an exon (Lam and Hertel, 2002). Indeed improving the strength of the splice site has already been demonstrated to counteract the enhancer requirements (Tian and Maniatis, 1994).

As previously mentioned the EAWS analysis included the evaluation of the splice sites strength. The 5'ss of the ALPP exon 4 and ALPL exon 5 was in both cases considered strong. On the other hand, the 3'ss strength of the ALPP exon 4 was suboptimal with the ALPL 3' ss being much stronger. This observation suggested that the weakness of the 3'ss strength in ALPP might represent the cause of the need of additional splicing regulatory elements to allow an efficient exon processing. To test this hypothesis and to investigate the influence of a weak 3' ss strength on ALPP exon 4 splicing, I exchanged the weak 3'ss of ALPP with the stronger one of ALPL, creating the minigene MUT18 (Fig. 3.15A). To see if the substitution would have any effect on the inclusion of exon 4 the minigene was analyzed and after transfection and RT-PCR analysis (Fig. 3.15B), I observed that it functioned in an analogous way as to the ALPP wt minigene (Fig. 3.6B, Lane 2).

Using this construct (ALPP with the the stronger splice site of ALPL) as a backbone, I then created a series of minigenes, initially with the exchanged ESE region 1 and 2 (Fig. 15A, MUT18-21), and subsequently in the context of finely mapped ESEs (Fig.3.16A, MUT22-

24).

Transient transfection experiments in HeLa cells and subsequent RT-PCR analysis showed that the improvement of the 3' splice site strength eliminates the need for ESE. Indeed, inclusion of ALPP exon 4 occurs, where before were observed different degrees of exon 4 skipping, as shown comparing ALPP minigenes with endogenous splice site (weak), (Fig. 8, 12, 13 and 14) and ALPP carrying the stronger splice site (Fig.15B and 16B).

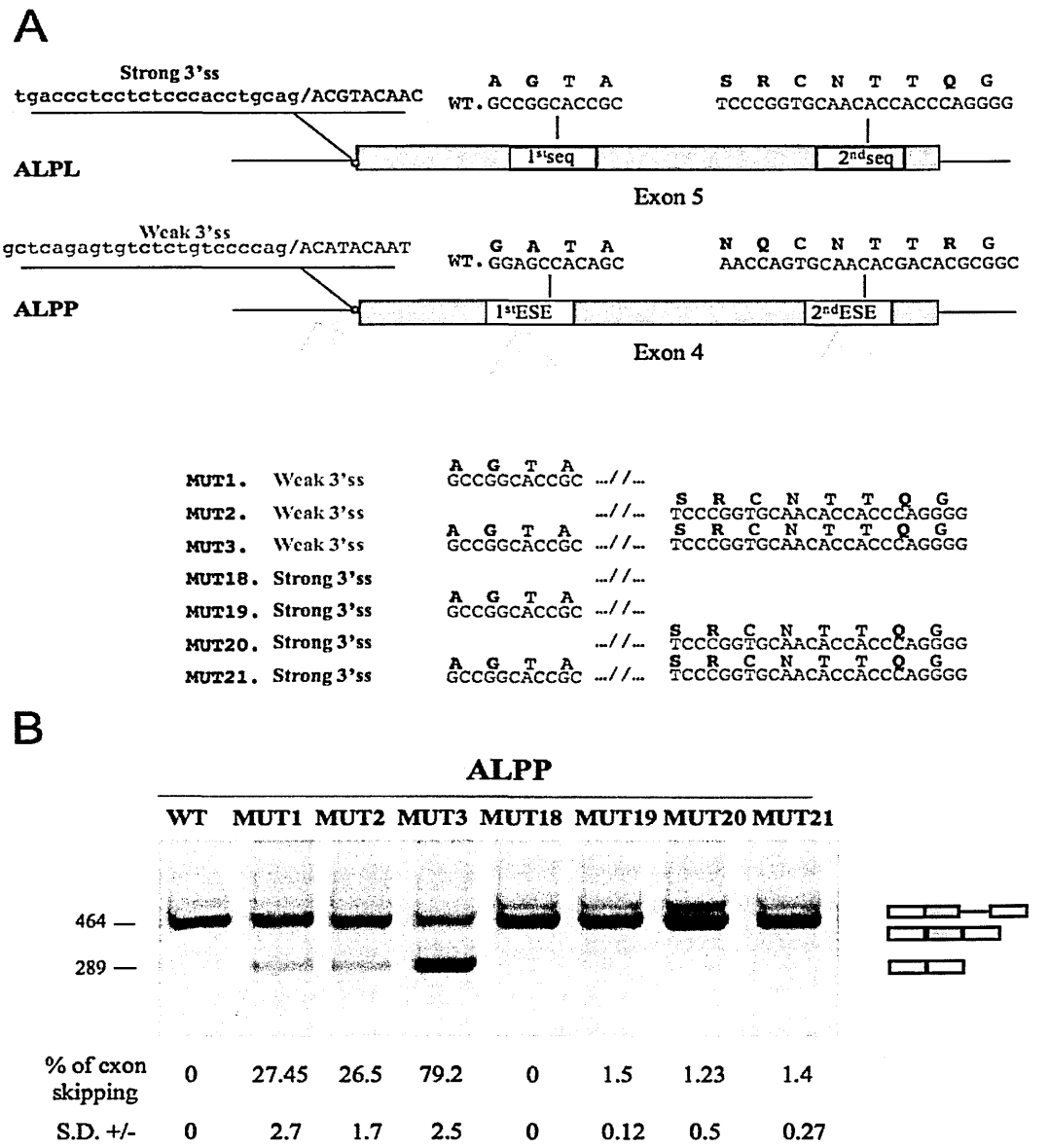


Figure 3.15. Analysis of ALPP exon 4 splicing after swapping of weak 3'ss with that strong of ALPL exon 5 in minigenes that lack of ESE motifs. (A) A schematic representation of the central part of the wt ALPP and ALPL constructs, the exon of interest is shown in green and its flanking intronic region as black line and the investigated wt sequences are represented above the exon (1st and 2nd seq in ALPL exon 5; 1st and 2nd ESE in ALPP exon 4). The sequences reported nearby the mutants (MUT1-3; MUT18-21) show the substitutions performed on the minigenes. The exact exchanged sequences of both ALPP 3'ss (pink) and ALPL 3'ss (blue) are reported in full. The constructs are analyzed with pcDNA3 minigene system in HeLa cells. (B) RT-PCR performed from RNAs isolated from HeLa cells previously transfected with the mutated minigene. MUT19-21 the increase of the 3'ss strength resulted in the recovery of the exon. On the right hand site of the gel a schematic representation of the splicing product obtained can be observed. The percentage of skipping and standard deviation (SD) are indicated below each lane and represent the mean of three experiments.

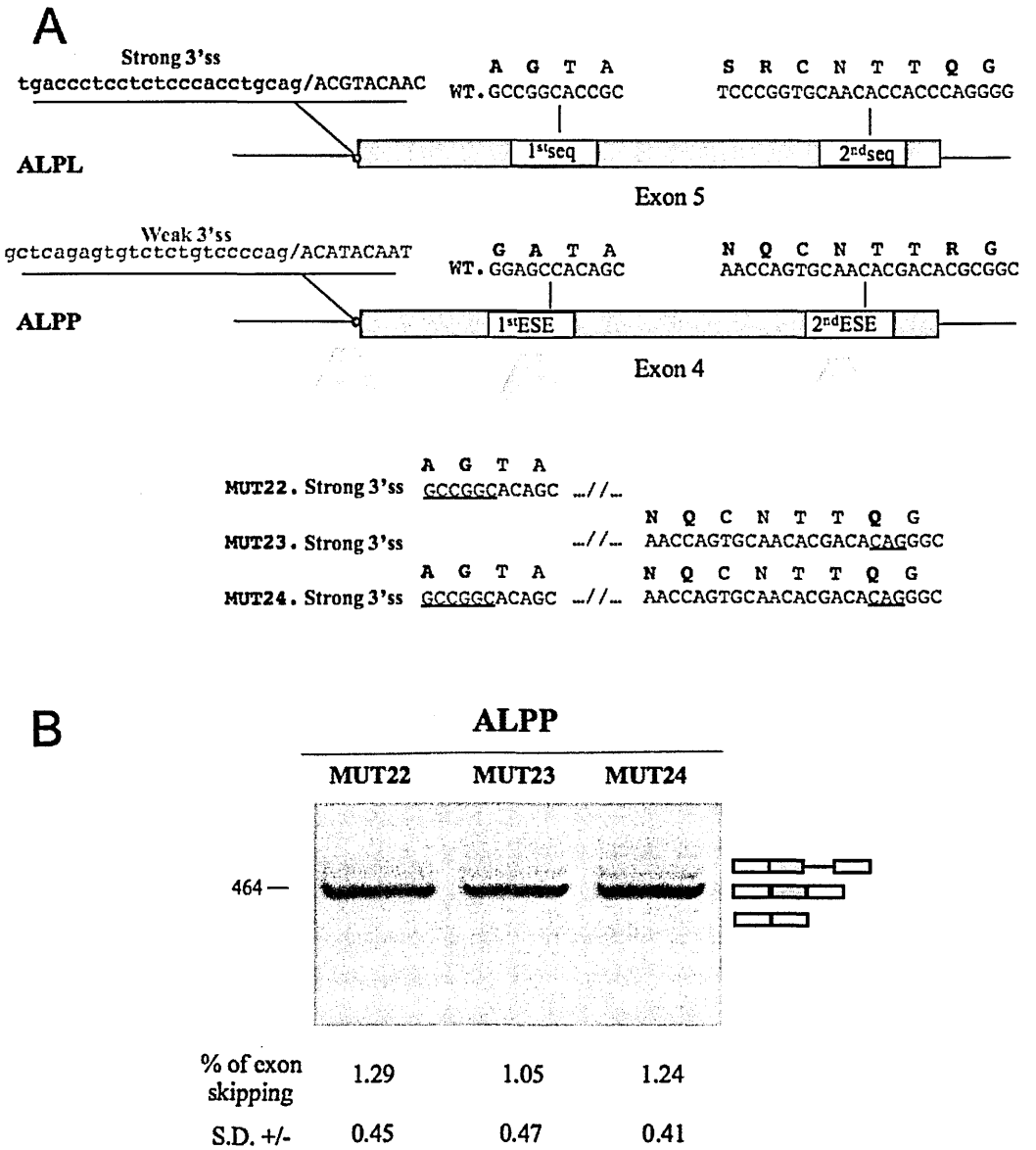
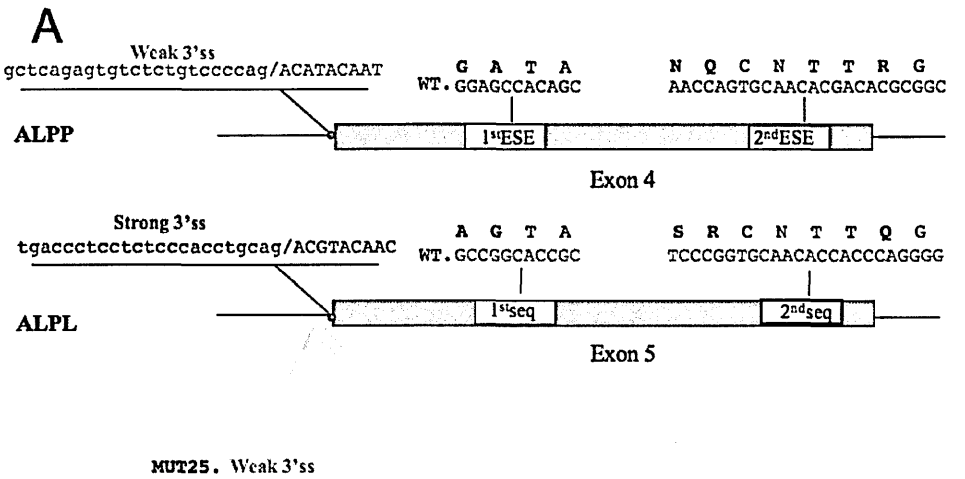


Figure 3.16. Analysis of ALPP exon 4 splicing after swapping of weak 3'ss with the stronger one of ALPL exon 5 in minigenes that lack of refined ESE motifs. (A) A schematic representation of the central part of the wt ALPP and ALPL constructs, the exon of interest is shown in green and its flanking intronic region as black line and the investigated wt sequences are represented above the exon (1st and 2nd seq in ALPL exon 5; 1st and 2nd ESE in ALPP exon 4). The sequences reported nearby the mutants (MUT22-24) show the substitutions performed on the minigenes. The exact exchanged sequences of both ALPP 3'ss (pink) and ALPL 3'ss (blue) are reported in full. The nucleotide changes in the 1st and 2nd ESE sequence are underlined in the sequences overhead. The constructs are analyzed with pcDNA3 minigene system in HeLa cells. (B) RT-PCR performed from RNAs isolated from HeLa cells previously transfected with the mutated minigene. MUT22-24 the increase of the 3'ss strength resulted in the recovery of the exon (n=3). On the right hand site of the gel a schematic representation of the splicing product obtained can be

observed. The percentage of skipping and standard deviation (SD) are indicated below each lane and represent the mean of three experiments.

Conversely, the bioinformatic analysis of ALPL exon 5 for putative ESEfinder elements, identified no predicted ESE motifs in the region corresponding to the 1st and 2nd ESE in the exon 4. This finding could be due to the fact that the 3' splice site of the ALPL is well defined, as shown by the bioinformatic prediction that reported indeed an optimal consensus score.

To investigate the influence of a strong 3' ss strength on ALPL exon 5 splicing, I generated a minigene where I exchanged the 3'ss of ALPL with the weaker one of ALPP (Fig. 3.17A). Transfection and RT-PCR of this minigene, confirmed that the absence of ESE sequences in ALPL exon 5 as a partial exon 5 skipping when the 3'ss is weakened, can be observed (Fig. 3.17B). These results indicate that well defined 3'ss in ALPL exon 5 might not require the presence of additional *cis*-acting elements, such as ESEs, to be efficiently recognized by the splicing machinery, whereas reducing the 3'ss strength of ALPL, the exon 5 is no more able to correctly include the exon in the mature mRNA.



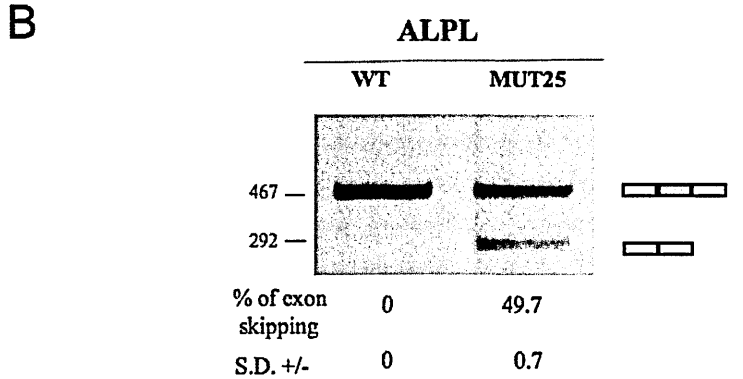


Figure 3.17. Analysis of ALPL exon 5 splicing after swapping of strong 3'ss with the weaker one of ALPP exon 4. (A) A schematic representation of the central part of the wt ALPP and ALPL constructs, the exon of interest is shown in green and its flanking intronic region as black line and the investigated wt sequences are represented above the exon (1^{st} and 2^{nd} seq in ALPL exon 5; 1^{st} and 2^{nd} ESE in ALPP exon 4). The sequences reported nearby the mutant (MUT25) show the substitution performed on the minigene. The exact exchanged sequences of both ALPP 3'ss (pink) and ALPL 3'ss (blue) is reported in full. The constructs are analyzed with pcDNA3 minigene system in HeLa cells. (B) RT-PCR performed from RNAs isolated from HeLa cells previously transfected with the mutated minigene. RNA splicing variants corresponding to ALPL exon 5 inclusion and exclusion are shown. In MUT25 the weakness of the 3'ss strength resulted in a partial exon 5 skipping. The percentage of skipping and standard deviation (SD) are indicated below each lane and represent the mean of three experiments.

3.3 Testing the biochemical effect of the amino acid differences within the enhancer elements of PLAP and corresponding region of TNAP

I have thus far mapped the nucleotides associated with two ESEs in ALPP and the lack of it in ALPL linked also to amino acid differences between PLAP and TNAP (at protein level) in a region that is conserved at the amino acid level among the alkaline phosphates asides TNAP. This fine mapping of the ESEs, was crucial in order to investigate if the need of ESE could dictate the amino acid code in such a way as to affect protein function.

In order to test this hypothesis I decided to investigate if exchanging the amino acids encoded by the nucleotides that were shown to be part of ESE motifs, into those of TNAP protein would have biochemical consequences.

In the work described above I demonstrated that in the 1st ESE region the two triplets GCC and GGC and in the 2nd ESE region the CAG triplet, (Fig. 3.12-3.13, nucleotide differences in ALPL are underlined) encoding for Gly93, Ala94 and Arg125, respectively, were associated with ESE function as well as causing amino acid differences between PLAP and TNAP protein.

Intriguingly, the analysis of 3D structure (made by Prof. JL Millan's Lab) in figure 3.18, modulated on the basis of human PLAP crystal structure (Le Du et al., 2001) highlighting the PLAP active site, shows that the amino acids of interest that differ between PLAP and TNAP, in the predicted ESEfinder motifs, may play a role in the catalytic protein function, for the proximity to the active site.

In particular, the figure 3.18A highlights the amino acid differences encoded by the 1st ESE, the G93 and A94 that are depicted as sticks in light blue and blue, respectively. The tridimentional structure showed that Gly93 and Ala94, besides being near the active site, are close to the conserved Ser92 residue, which is phosphorylated in the course of the catalysis. Figure 3.18B highlights the amino acid difference encoded by the nucleotides

encoding within the 2nd ESE, R125 depicted as stick in purple. Also in this case the proximity of the amino acid to the active site, suggests a possible involvement in the protein activity.

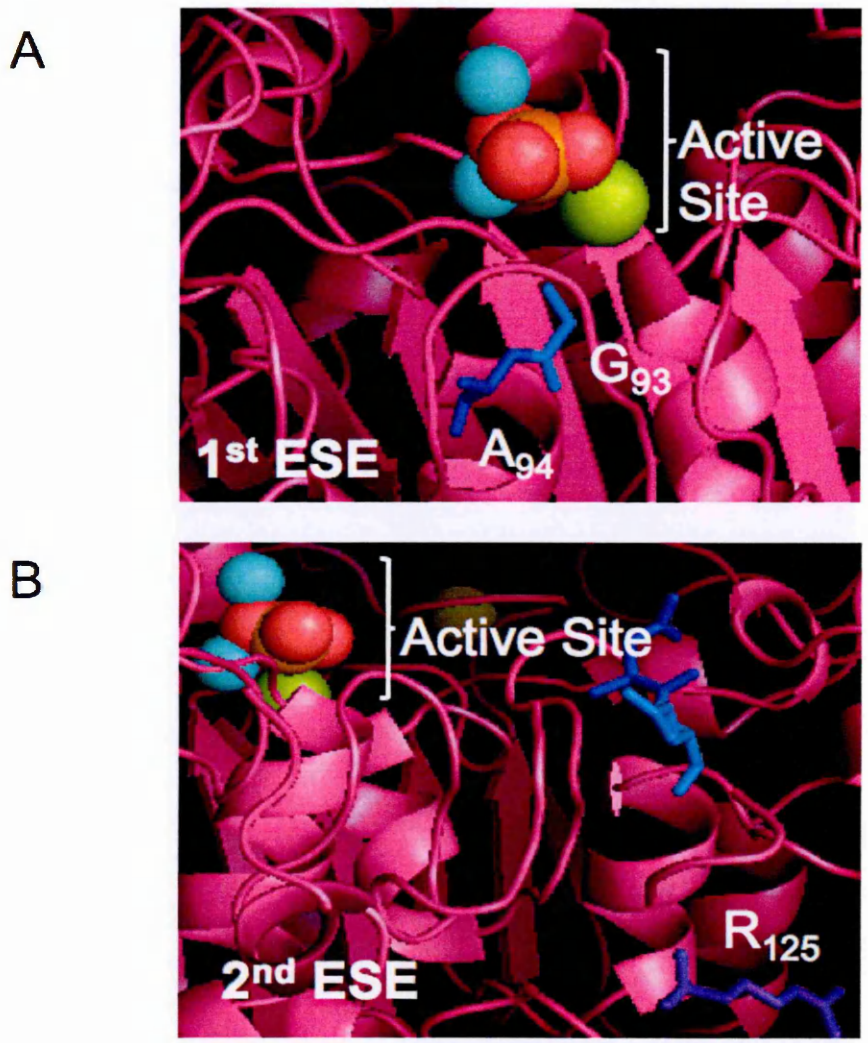


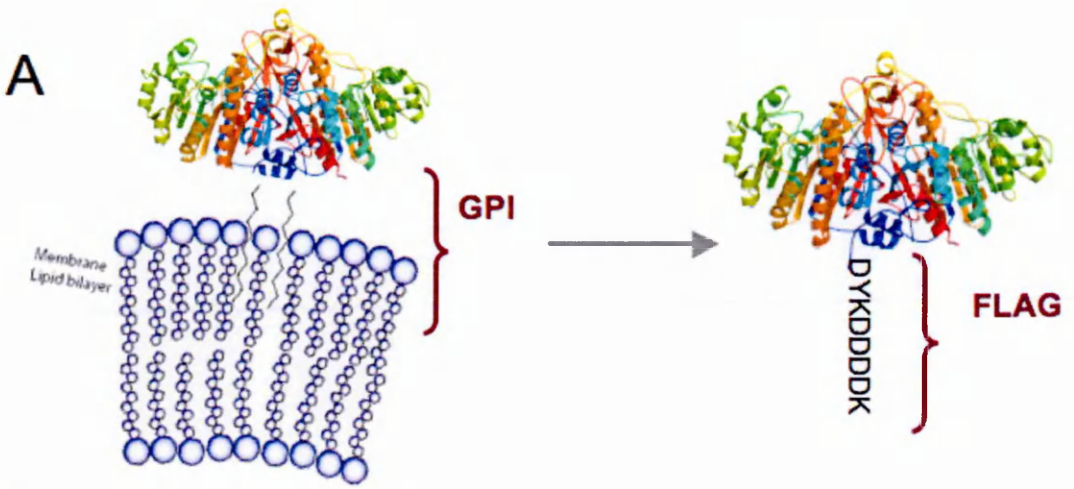
Figure 3.18. Active site region of human PLAP. The highlighted residues are those that differ between TNAP and PLAP, within the 1st ESE, in light blue G93 and in blue A94 (A) and 2nd ESE, in violet R125 (B). PLAP is shown as a ribbon and investigated residues as sticks. The zinc atoms are shown in turquoise, the magnesium is in green, and the phosphate moiety is shown in orange and red. The figures were produced made by JL Millan's Lab.

3.4 Setup of methodology for recombinant Human Alkaline Phosphatase protein expression

In order to test the biochemical effect of the amino acid differences I initially set up a methodology for the production of recombinant ALPs. I decided to use a mammalian expression system in order to ensure the post-translational modifications on the protein were taking place correctly. To simplify the recovery and the purification of the recombinant enzymes, the GPI anchoring signal peptide sequence of ALPs was replaced by the FLAG octapeptide (DYKDDDDK) (Fig 3.19A). The recombinant proteins were thus expressed as secreted, epitope-tagged enzymes. Previous studies have demonstrated that the addition of FLAG tag does not interfere with the kinetic properties of the enzymes (Di Mauro et al., 2002). It should be stated that we opted for FLAG sequence and not for the more widely used His-tag system out of concern that the high affinity of this extraneous stretch of His residues would interfere with the binding of Zn1 and Zn2 in the active site pocket of the enzyme. Instead, the usage of secreted FLAG enzyme was widely adopted for a large number of ALP studies.

The expression vector pcDNA3.1 carrying the wild type PLAP and TNAP cDNA was kindly donated by JL Millan (Sanford-Burnham Medical Research Institute, USA). The FLAG epitope was introduced after Leu489 and Thr483 in TNAP and PLAP, respectively, followed by a termination codon to eliminate the glycosylphosphatidylinositol-anchoring signal (Fig. 3.19, red lettering). Since the original wt constructs showed several polymorphisms that altered the coding sequence, we decided to mutate these sites into wt (non-polymorphic) sequences to resemble a more physiological context.

The constructs were transfected into COS-1 cells for transient expression as described in Materials and Methods (session 6.17). Transfected cells were cultured in OPTI-MEM serum free medium, and conditioned media, containing secreted enzyme, was collected 48 hours after transfection.



B

TNAP WT

LVPEKEKDPKYWRDQAQETLKYALELQKLNTNVAKNVIMFLGDGMGVSTVTAARILKGQLHHN
PGEETRLEMDKFPFVALSKTYNTNAQVPDSAGTATAYLCGVKANEGTVGVSAATERSRCNTTQG
NEVTSILRWAKDAGKSVGIVTTTRVNHATPSAAYAHSADRDWYSDNEMPPEALSQGCKDIAYQL
MHNIRDIDVIMGGGRKYMYPKNKTDVEYESDEKARGTRLDGLDLVDTWKSFKPRYKHSHFIWN
RTELLTLDPHNVDYLLGLFEPGDMQYELNRNNVTDPSLSEMVVVAIQILRKNPKGFFLLVEGGRI
DHGHHEGKAKQALHEAVEMDRAIGQAGSLTSSEDTLTVVTADHSHVFTFGGYTPRGNSIFGLAP
MLSDTDKKPFTAILYGNGPGYKVVGGERENVSMVDYAHNNYQAQSAVPLRHETHGGEDVAVFS
KGPMAHLLHGVHEQNYVPHVMAYAACIGANLGHCAPASSAGSL$_{489}$
AAGPLLLALALYPLSVLF

PLAP WT

IIPVEEENPDFWNREAAEALGAAKKLQPAQTAAKNLIIFLGDGMGVSTVTAARILKGQKKDKLG
PEIPLAMDRFPYVALSKTYNVDKHVPDSGATATAYLCGVKGNFQTIGLSAAARFNQCNTTRGNE
VISVMNRAKKAGKSVGVVTTTRVQHASPAGTYAHTVNRNWYSDADVPASARQEGCQDIATQL
ISNMDIDVILGGGRKYMFRMGTPDPEYPDDYSQGGTRLDGKNLVQEWLAKRQGARYVWNRTE
LMQASLDPSVTHLMGLFEPGDMKYEIHRDSTLDPSLMEMTEAALRLLSRNPRGFFLFVEGGRID
HGHHESRAYRALTETIMFDDAIERAGQLTSEEDTLSLVTADHSHVFSFGGYPLRGSSIFGLAPGK
ARDRKAYTVLLYGNGPGYVLKDGARPDVTESESGSPEYRQQSAVPLDEETHAGEDVAVFARGP
QAHLVHGVQEQTFIAHVMAFAACLEPYTACDLAPPAGTTD$_{483}$
AAHPGRSVVPALLPLLAGTLLLLETATAP

Figure 3.19. Recombinant Human Alkaline Phosphatase protein expression. (A) Schematic representation of Alkaline Phosphatase, indicating the glycosylphosphatidylinositol (GPI) anchor and the soluble form after GPI substitution with FLAG epitope. (B) Complete amino acid sequence of human tissue non-specific (TNAP), placental (PLAP), highlighting in purple the residues that are removed from the C-terminus and substituted with FLAG epitope (DYKDDDDK).

3.4.1 Purification and quantification of wt FLAG secreted protein

The wt PLAP-FLAG construct was used to set up a system for the purification and quantification of the recombinant proteins. After transfection, affinity purification of the soluble protein was carried out with anti-FLAG M2 monoclonal antibody affinity gel (SIGMA) as described in the Materials and Methods. Bound protein was eluted with five fractions of 1 ml glycine (100 mM, pH 3.5) (Fig. 3.20A). Only the first elution fraction of wt PLAP-FLAG sample exhibited a single band at approximately 64 kDa (Fig. 3.20A, Lane 1), corresponding to the subunit molecular weight of Placental Alkaline Phosphatase (Greene and Sussman, 1973).

Subsequently this was also done for TNAP and in order to assess the efficiency and purity of this procedure, 50µl of eluted wt PLAP-FLAG and TNAP-FLAG proteins were analyzed by Coomassie staining and Western blot procedure (Fig. 20B and C). As shown in Figure 19B, only a 64 kDa band, corresponding to the recombinant wt PLAP-FLAG protein was detected when the elutes were analyzed by Coomassie staining (Fig. 3.20B, line 1). Western blot analysis with an antibody anti-FLAG however showed the TNAP protein to be present albeit in a reduced amount (Fig. 3.20C).

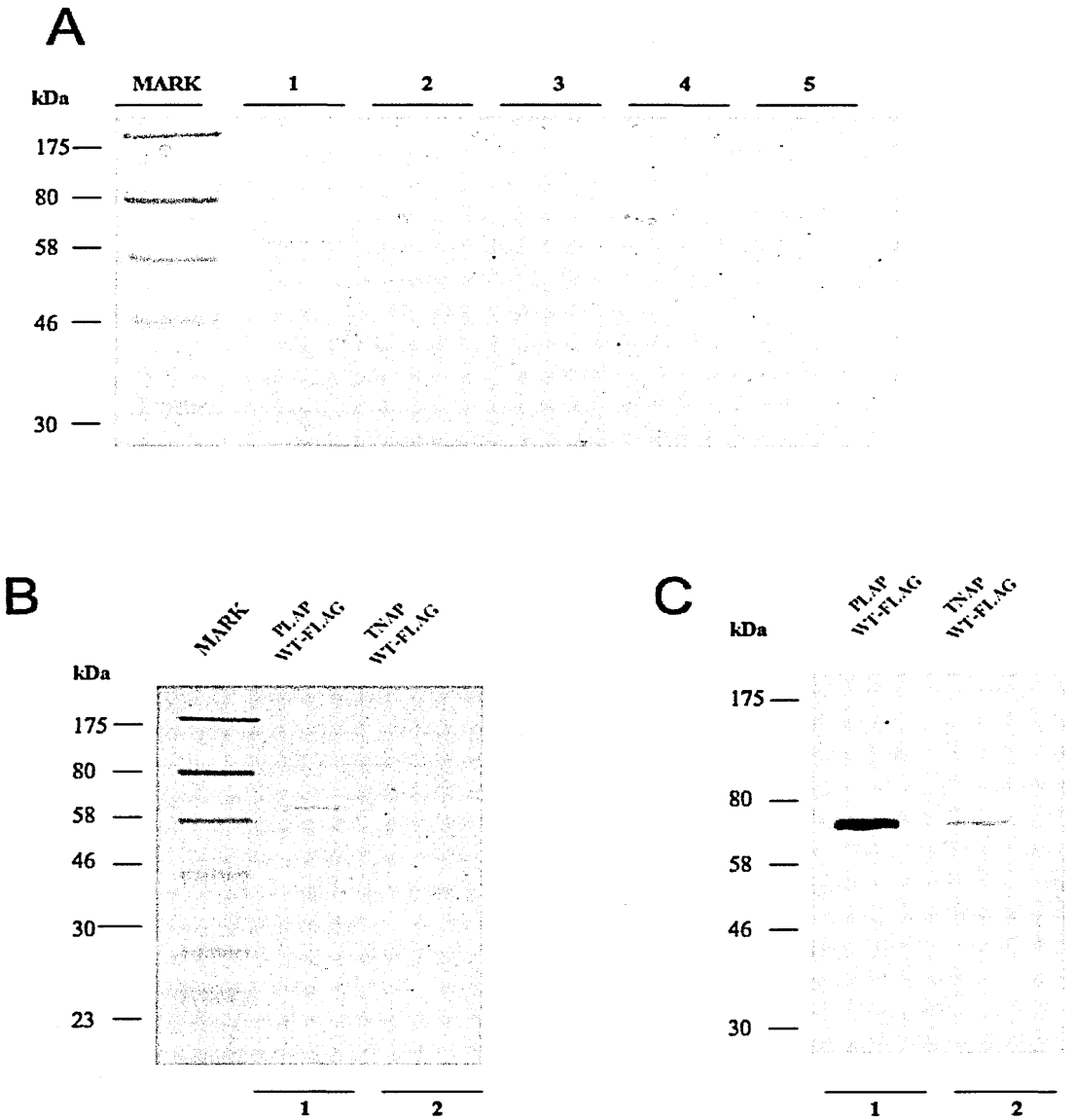


Figure 3.20. Purification of the FLAG secreted enzymes. After 48 hours from transfection, the culture medium is collected and the flag tagged protein is purified by anti-FLAG M2 resin. (A) 5% of each elution fraction of wt PLAP-FLAG loaded on 10% SDS PAGE gel stained with Coomassie Blue. (B) 5% of the first elution fraction of wt PLAP and TNAP loaded on 10% SDS-PAGE gel after purification and stained with Coomassie Blue. (C) Western Blot of the wt PLAP and TNAP eluted proteins, using the same protein amount as in B experiment.

3.4.2 Evaluating purification yield

Since this severe decrease in TNAP protein production could generate difficulties in quantification, essential for the measurement of enzymatic activity, the first problem to solve was to calculate the protein concentration.

Several approaches were investigated simultaneously to quantify the amount of the purified proteins. As can be seen in figure 3.21 where I approximately calculated the PLAP-FLAG concentration from BSA scale (~ 5ng/µl) whereas the purity of the proteins is extremely high the expression level is quite low making quantification by Bradford protein assay unfeasible. For this reason I tried to calculate the protein concentration with Micro Bicinchoninic Acid (BCA) Protein Assay (Pierce) for determining the protein concentration of dilute samples (0.5 - 20µg/mL). The amount of protein present in a solution can be quantified by measuring the absorption spectra and the standard curve is obtained by plotting the absorbance (Abs) versus known concentration of standard protein. Unfortunately, also micro BCA assay failed to evaluate protein yield, as the elution buffer that I used for purification was not compatible with this system.

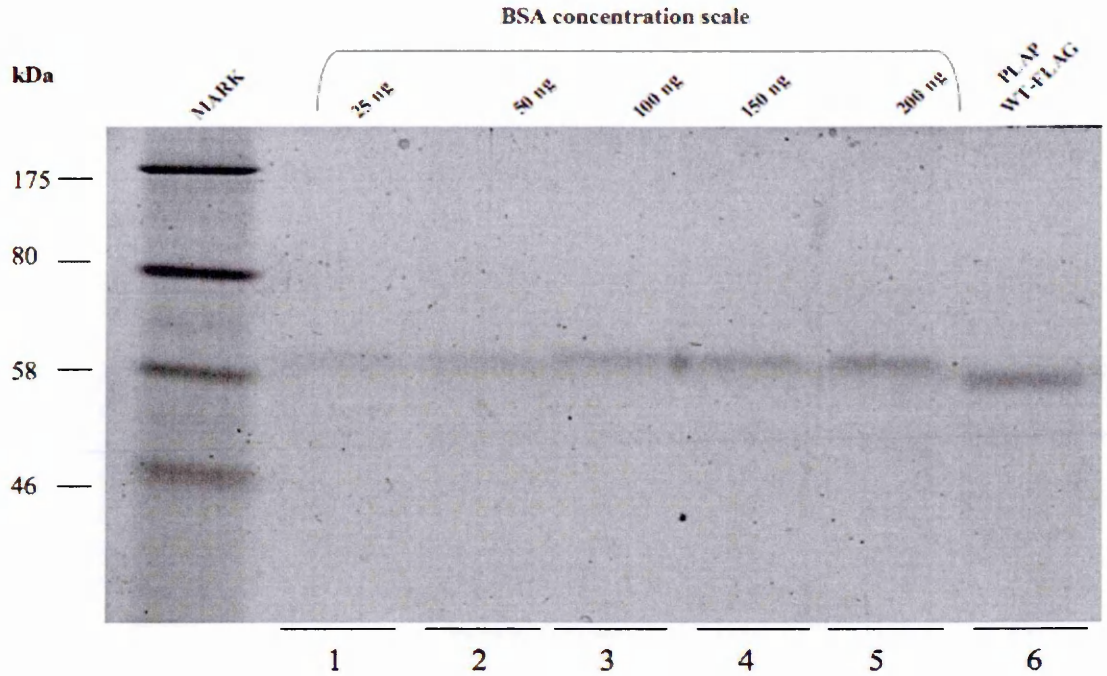
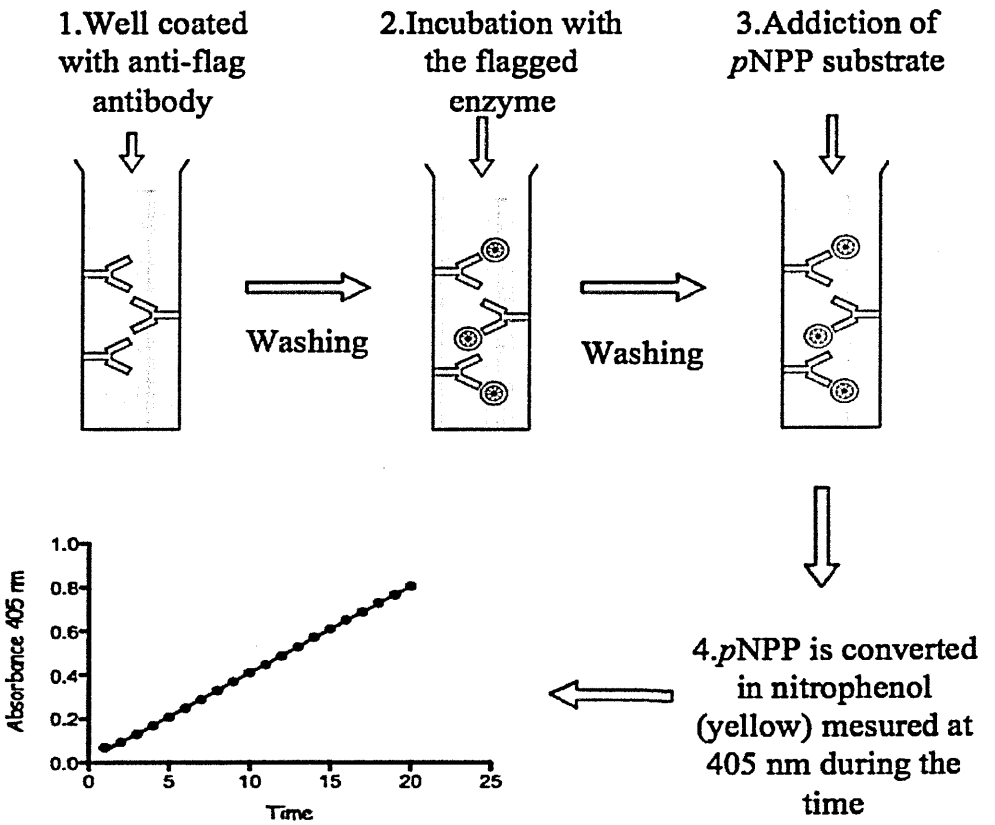


Figure 3.21. BSA concentration scale for the approximate estimation of PLAP-FLAG yield. SDS PAGE gel at 10% stained with Coomassie Blue in which the first five lines indicate the BSA concentration scale (from 25 to 200 ng) and the lane 6 shows the wt PLAP-FLAG in which was loaded the 5% of total protein obtained from the purification experiment.

I also attempted to set up an immunoenzymatic assay in order to measure the ALP catalytic activities. This system bypasses the calculation of the protein concentration itself. The experiments were done during a 3 weeks stage at Prof. Millan's Laboratory, at the Burnham Institute in San Diego. The principle of the assay, represented schematically in figure 3.22, was based on the measurement of relative specific enzymatic activities using microtiter plates coated with 0.2 μg/ml M2 anti-Flag. The wells were than incubated with the recombinant enzyme carrying the FLAG epitope, recognized by anti-FLAG antibody. When p-nitrophenyl phosphate (pNPP, 0.01-20 mM final concentration.) was added as

substrate it is possible to measure directly the activity at 405 nm during the time (Fig. 3.22A). For the calculation of catalytic rate constants (k_{cat}), the wt PLAP-FLAG construct with a known k_{cat} is used as a reference for each microtiter plate. Although, this system has been extensively used in Prof. Millan's Lab for the measurement of enzymatic activity when I analyzed our wt PLAP and TNAP-FLAG enzymes, changing the missense-polymorphic sites into wt (non-polymorphic) sequences, with this procedure it unfortunately did not yield results (Fig. 3.22B). One explanation was that the changes in the original wt cDNA construct caused a reduction in production levels compared to the original one used in the laboratory (section 2.4), making impossible the measurement of the activity with this system. A more sensitive substrate was also tested to detect the enzymatic reaction: CDP-Star (phenylphosphate substituted 1,2 dioxetane) chemiluminescent substrate. It enables extremely sensitive and fast detection of enzymatic dephosphorylation by producing visible light at 466 nm. However even using chemiluminescent substrate did not improve the detection of catalytic activities.

A **Immunoenzymatic assay technique:**



B

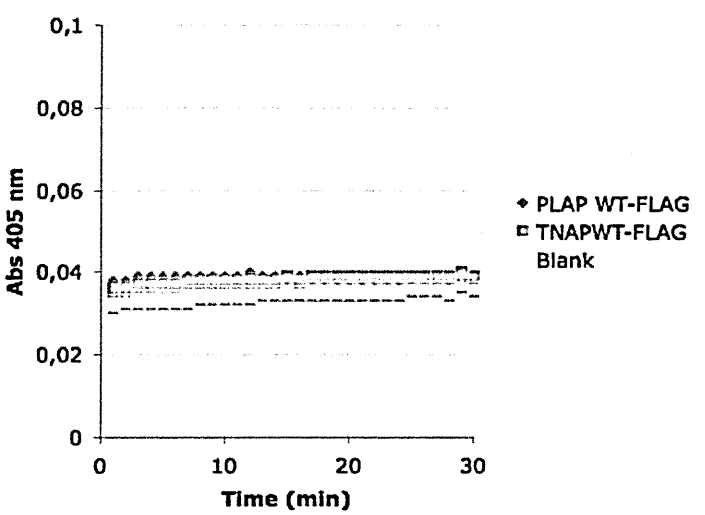


Figure 3.22. ALPs Immunoenzymatic assay (IEA). (A) Schematic representation of IEA used in JL Millan's lab. 1.Mictrotiter plates were coated with anti-FLAG antibody. 2.Recombinant ALP-FLAG proteins were incubated with the plates for 2 h, after which each well was washed with PBS. 3-4. The activity of bound enzymes was than measured at 405 nm as a function of time, after addiction of the substrate (p-NPP). (B) Wt PLAP-FLAG and TNAP-FLAG activity measured at 405 nm during the time.

Finally, a quantitative slot blot assay was developed to calculate the protein concentration, which together with scanning densitometry allows protein detection with high sensitivity (Zhu et al., 2005). Amino-terminal FLAG-Bacterial Alkaline Phosphatase (BAP-FLAG) fusion protein was used as a standard. Figure 2.23 (A and B) shows a typical analysis of FLAG-BAP standard with recombinant ALP proteins detected with slot blot system. Seven two-fold serial dilutions of FLAG-BAP standard protein and purified enzymes were prepared with PBS. The BAP-FLAG standard dilutions ranged from 3.5 to 225 ng (Fig. 3.23A, lanes 1 and 2). The scanned image of the film and the intensity of each band on the slot was measured with ImageJ64 Software. Thus each peak in the graph represented a slot containing detected protein (Fig. 3.23B). Standard curves were obtained by plotting peak area versus known concentration of BAP-FLAG and fitting the data points with linear regression equation (Fig. 3.23C). Notably, the regression curves obtained always had good correlation coefficients (R^2) of 0,9907 (n=12). The peak areas of the serially diluted purified protein samples that fell within the detection range of each respective standard curve were used to calculate the protein concentration. The mean value of 3 peaks from each protein dilution series was used per assay with a coefficient of variation (CV) that ranged from 3% to 12%. PBS control samples gave no measurable signals on the membranes (data not shown).

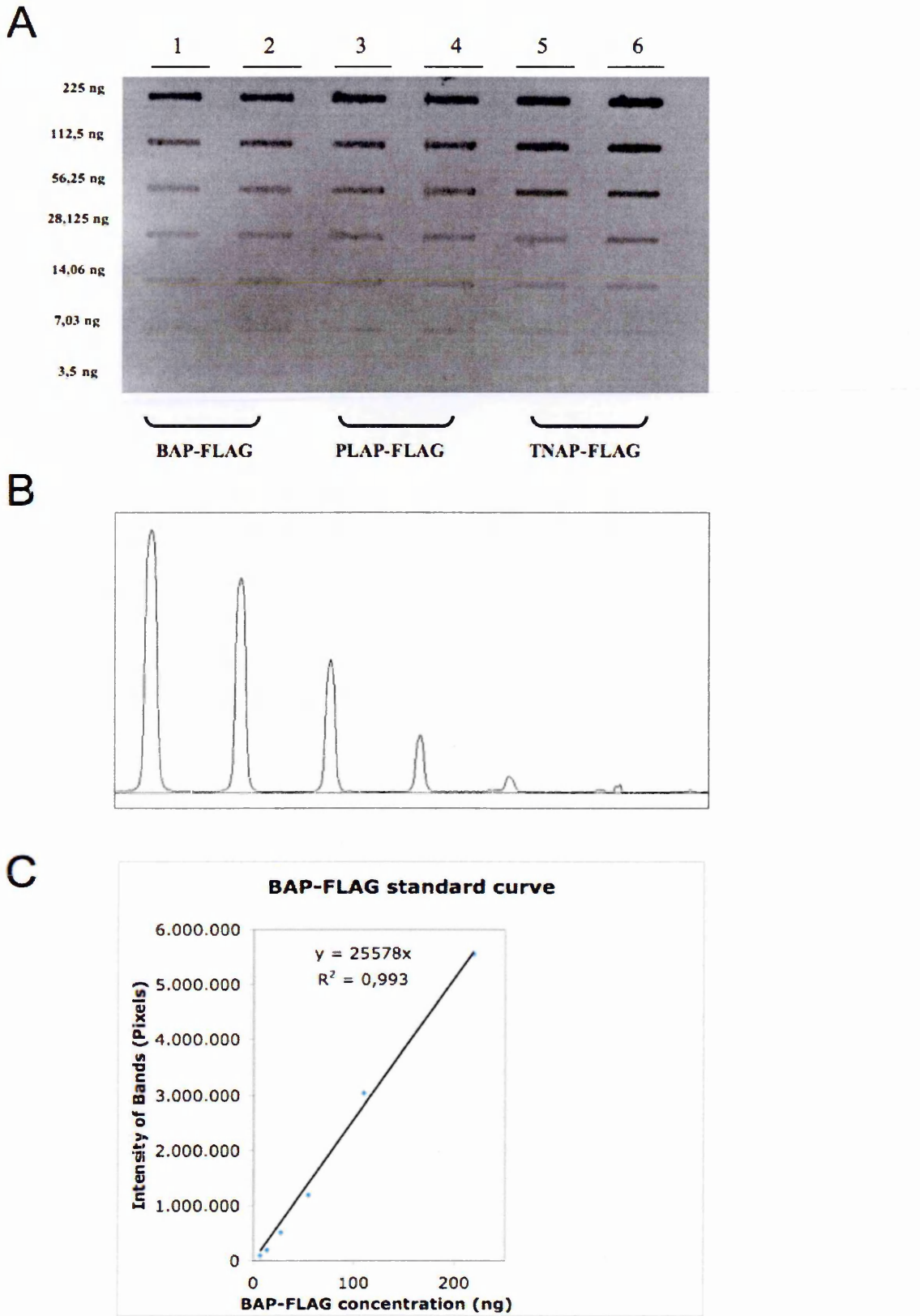


Figure 3.23 Slot blot detection of ALPs using anti-FLAG antibody. (A) Lanes 1 and 2 contain BAP-FLAG standard, with the first (top left) slot loaded with 225 ng and followed by 2-fold serial dilutions. Lanes 3-4 contain purified recombinant wt PLAP-FLAG protein. Lanes 5-6 contain purified recombinant wt TNAP-FLAG protein. (B) A typical analysis of FLAG-BAP standard and recombinant ALP proteins in which the ImageJ64 Software was

used to determine the density of the slot blot signals from a scanned image of the film. Thus each peak in the graph represented a slot containing detected protein. (C) Data are plotted as peak area versus known concentration of BAP-FLAG and fitting the data points with linear regression equation.

3.5 Construction of protein expression vectors by site–directed mutagenesis of PLAP cDNA

To establish if the amino acid differences between PLAP and TNAP associated with ESE, in the former, may play a role in PLAP function I decided to test whether they have a biochemical effect by mutating them for the corresponding amino acids in TNAP.

The 1st ESE region in PLAP contains the amino acids Gly93 and Ala94 (Fig. 3.24A). I mutagenized these two amino acids to the corresponding residues in TNAP, i.e., Ala and Gly, respectively, in the context of wt PLAP-FLAG cDNA, creating a new construct: PLAP-FLAG 1st ESE [G93A; A94G] (Fig. 3.24C). In the 2nd ESE region the amino acid difference between PLAP and TNAP found to be associated with the ESE was Arg125 (Fig. 3.13). This was therefore mutated to the corresponding residue in TNAP, i.e., Gln, in the context of wt PLAP-FLAG cDNA giving rise to the new expression construct: PLAP-FLAG 2nd ESE [R125Q] (Fig. 3.24C). The results previously observed in the figure 3.17 demonstrated that the amount of ALPP exon 4 skipping increases drastically when both ESE sequences were substituted for the corresponding region of ALPL, suggesting that they can act in a synergistic manner. Based on this finding, I also created a mutant in which residues encoded by the 1st and 2nd ESE region in PLAP, i.e. Gly93, Ala94 and Arg125 were mutagenized into the corresponding amino acids of TNAP, Ala, Gly and Gln, respectively generating the mutant PLAP-FLAG 1st/2nd ESE [G93A; A94G; R125Q] (Fig. 3.24C).

A

```
83  T Y N T N A Q V P D S A G T A 97  :TNAP
82  T Y N V D K H V P D S G A T A 96  :PLAP

249 ACGTACAACACCAATGCCCAGGTCCCTGACAGTGCCGGCACCGCC 293 :ALPL
246  CATACAATGTAGACAAACATGTGCCAGACAGTGGAGCCACAGCC 290 :ALPP
```

B

```
113 V S A A T E R S R C N T T Q G 127 :TNAP
112 L S A A A R F N Q C N T T R G 126 :PLAP

339 GTAAGCGCAGCCACTGAGCGTTCCCGGTGCAACACCACCCAGGGG 383 :ALPL
336 TTGAGTGCAGCCGCCCGCTTTAACCAGTGCAACACGACACGGGC 380 :ALPP
```

C

PLAP- mutants:

PLAP-FLAG 1st ESE [G93A;A94G]

...TYNVDKHVPDSAGTATAYLCGVKGNFQTIGLSAAARFNQCNTTRGEVISVMNRAKKA...

PLAP-FLAG 2nd ESE [R125Q]

...TYNVDKHVPDSGATATAYLCGVKGNFQTIGLSAAARFNQCNTTQGEVISVMNRAKKA...

PLAP-FLAG 1st ESE/ 2nd ESE [G93A;A94G;R125Q]

...TYNVDKHVPDSAGTATAYLCGVKGNFQTIGLSAAARFNQCNTTRGEVISVMNRAKKA...

Figure 3.24. Comparison between PLAP and TNAP amino acid sequence in a region that includes the ESE sequences. (A) The region enclosed by the grey box highlights the amino acids that differ between PLAP and TNAP within the 1st ESE and corresponding nucleotide variations. (B) The region enclosed by the grey box highlights the amino acid that differs between PLAP and TNAP within the 2nd ESE and corresponding nucleotide variations. (C) Partial amino acid sequence of PLAP mutants. Amino acid changes are highlighted in turquoise.

Cos1 cells were transfected with the mutated construct. The conditioned media, containing secreted mutant enzymes, was collected 48 hours after transfection and then purified using the anti-FLAG M2 gel. The mutated purified proteins was subjected to 10% SDS-PAGE stained with Coomassie blue. This analysis revealed that the production of PLAP-FLAG 1st ESE [G93A; A94G] is approximately the same obtained for the recombinant wt PLAP-FLAG protein whereas that of PLAP-FLAG 2nd ESE [R125Q] and PLAP-FLAG 1st /2nd ESE [G93A; A94G; R125Q] was only approximately 50% of that was obtained for the wt enzyme (Fig. 3.25).

Not withstanding these low levels of expression I was able, using the quantitative slot blot assay, to calculate the protein concentration, loading much more quantity of PLAP-FLAG 2nd ESE and PLAP-FLAG 1st /2nd ESE through the slot to obtain a good signal on the membrane.

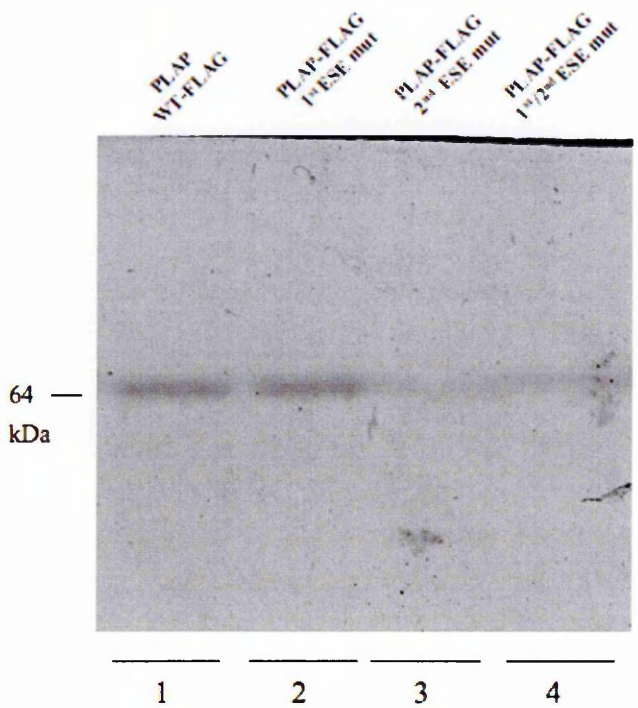
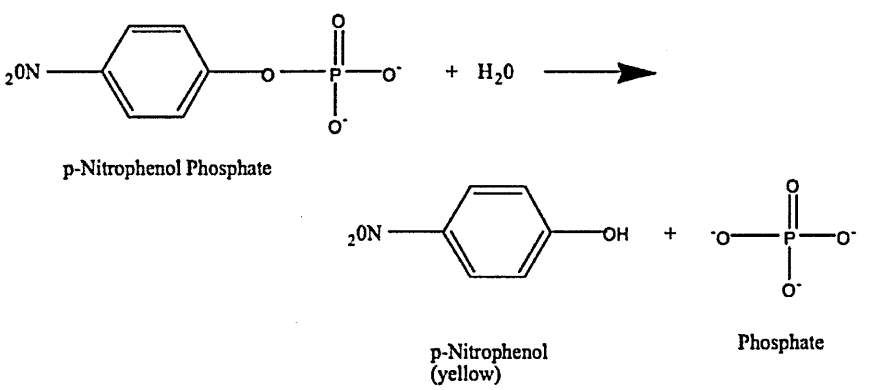


Figure 3.25. An example of SDS PAGE gel at 10% stained with Coomassie Blue. Controlling the purity production of recombinant proteins. In each case the 5% percent of the first elution was loaded.

3.6 Kinetic studies of ALPs

To study if the differences in the amino acid sequence in the 1st and 2nd ESE of PLAP, compared to the corresponding region in TNAP, which I hypothesized to be related to the need of splicing regulatory elements in these areas, could affect the kinetics of the protein, I tested if these changes affect PLAP through an enzymatic assay to measure the kinetics of this protein based on the following reaction:



The enzymatic reaction consisted in the removal of a phosphate from the p-NPP, used as substrate, to generate p-nitrophenol. While the substrate, pNPP, is colorless, the product p-nitrophenol is yellow, so it was possible to follow the reaction progress by measuring the generation of yellow color. Hence, the rate of the enzyme-catalyzed reaction can be traced by taking absorption readings of the amount of product accumulated at 405nm as a function of time.

Initially to set up the system I tested wt PLAP and TNAP-FLAG constructs. Enzyme kinetic determinations of wt ALP proteins were performed in microtiter plates by adding the enzyme to pNPP in 1.0 M diethanolamine (DEA) buffer, pH 9.8 (see materials and methods). The substrate concentration was varied between 0.01 and 20 mM with eleven different enzymatic reactions for each wt enzymes. The assay mixture also contained $MgCl_2$ and $ZnCl_2$ because each catalytic site contains three metal ions, i.e., two Zn and one

Mg, necessary for enzymatic activity.

The ALP reaction proceeded following Michaelis-Menten kinetics. In order to evaluate the catalytic parameters I calculated initial velocity (v_o) for each different substrate concentration. An important assumption is that the concentration of enzyme-substrate complex ([ES]) should remain unchanged during the initial velocity measurement, hence when the absorbance increases at a linear rate. The terminology describing this phase is "*steady-state*", because the concentration of ES is steady during the time interval used for enzyme kinetic work.

The initial rate of reaction was evaluated simply as a change in absorbance per unit of time: for pNPP formation this would be Δ_{405}/min, related to an increase in the concentration of the product, p-nitrophenol, per minute (Δc/min). The corresponding Δc/min was determined using Beer-Lambert's law, converting the absorbance value to the actual p-nitrophenol formed per minute. The reaction was followed for 1h but the v_o was calculated only from the linear part of the time-curve, generally 15 min (Fig. 3.26).

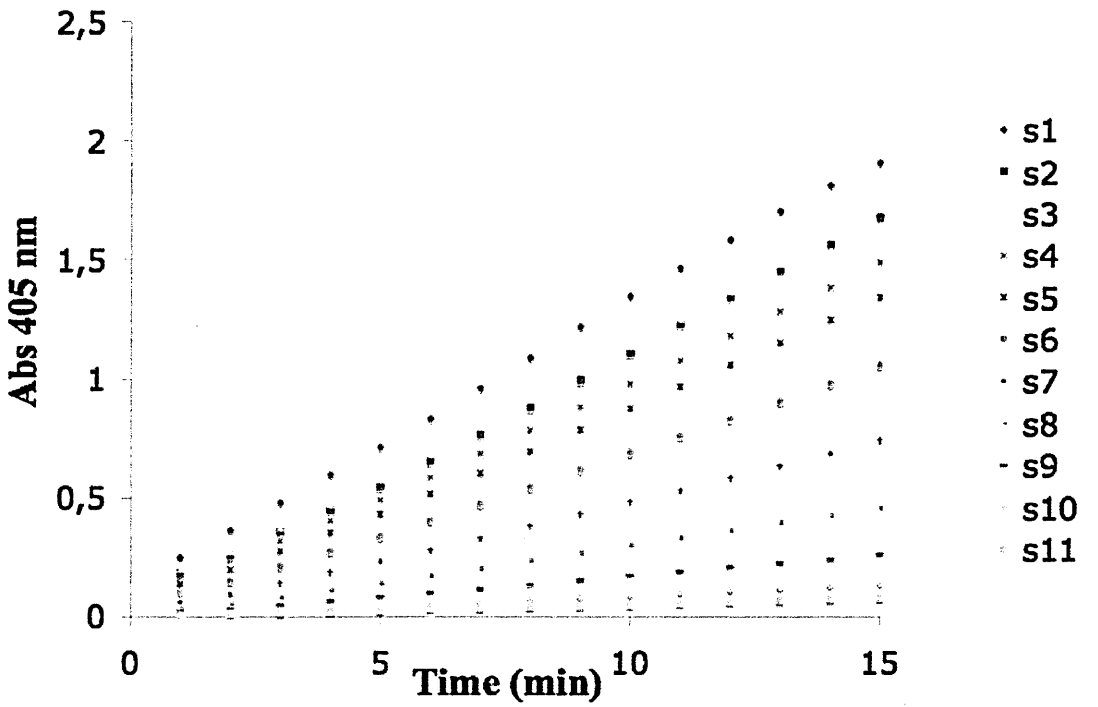


Figure 3.26. An example of p-nitrophenol formation measured as an increase in absorbance at 405 nm (Abs 405) during the time Δ_{405}/min (15 min). Eight ng of enzyme preparation were added to 200 ul of DEA buffer containing different concentration of p-NPP (s1-11), i.e., s1 contained 20mM and followed by 2-fold serial dilutions.

Since the best way for the measurement of catalytic activities is to fit a hyperbola directly to the substrate-velocity data, for this reason I used Prism program that automatically computed the binding parameters. Hence, the substrate concentration ([S]) on the *x* axis was plotted against the initial velocity (v_o) on the *y* axis, and the data were fitted in Michaelis-Menten equation. Four parameters were obtained from the Michaelis-Menten equation: (1) V_{max} is the maximum rate that can be achieved at a given concentration of enzyme; (2) K_m, the Michaelis constant, which is defined as the substrate concentration that produces an

initial velocity equal to one-half of V_{max} and is a measure of the affinity for the substrate; (3) K_{cat}, the enzymatic turnover number, which gives the frequency with which an enzyme, operating at saturation, can convert substrate to product; (4) K_{cat}/K_m that considers how well the substrate is bound to the enzyme and how rapidly it is converted from that point on to product.

Analysis of catalytic values for wt PLAP and TNAP are presented in Table 3.1. From our biochemical study, and in accordance to what is already known in literature, the V_{max} of TNAP was found to be double that of PLAP due to a more rapid movement of substrate into and out the active site, instead PLAP has a greater affinity for the substrate with a consequent lower K_m value (Fig. 3.27). These differences can be better observed visually in the Lineweaver-Burk plot where I transformed the data to double reciprocal plot (1/[S] vs. $1/v_0$) that was linear over the range of substrate concentration used, as shown in figure 3.27B. The determined kinetic parameters are listed in the Table 3.1.

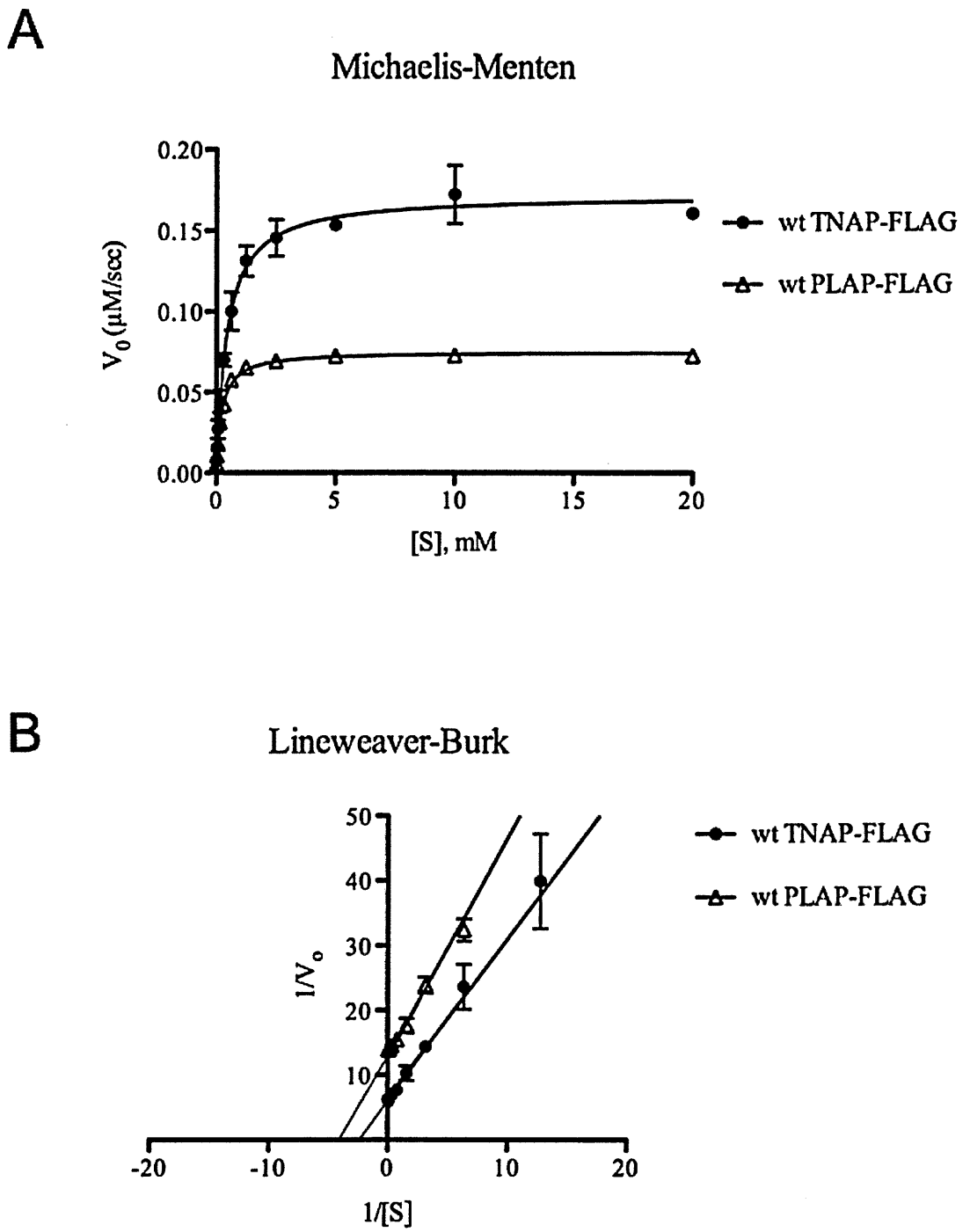


Figure 3.27. The Michaelis-Menten and Lineweaver-Burk plots for the hydrolysis of p-NPP by wt PLAP-FLAG (△) and wt TNAP-FLAG (●). Eight ng of enzyme preparation were added to 200 ul of DEA buffer containing different concentration of p-NPP (from 0.01 to 20 mM). (A) The initial velocities (v_o) were related to the concentration of substrate, [S], and the data were fitted in Michaelis-Menten equation. (B) In the Lineweaver-Burk plot the data were transformed to double reciprocal plot (1/[S] vs. $1/v_0$) that was linear over the range of substrate concentration used.

3.6.1 Effect of Gly93>Ala and Ala94>Gly amino acid substitutions on the kinetic activity of PLAP.

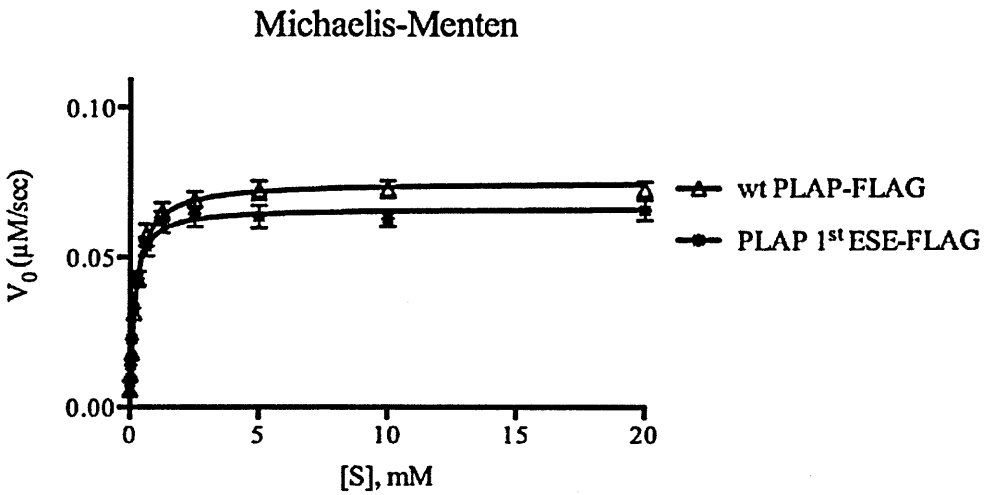
After setting up the kinetic assay and the purification of the proteins, I tested whether substituting the amino acids encoded within the 1st ESE for the residues found in TNAP, (Gly93 and Ala94 for Ala and Gly) would have any functional effects on catalysis. To verify this possibility, I compared wt PLAP-FLAG and PLAP-FLAG 1st ESE mutated [G93A; A94G] enzymatic kinetic parameters.

The best-fit values, from Michaelis-Menten equation, for the kinetic constants revealed a significant difference, between the wt PLAP-FLAG and 1st ESE mutated recombinant enzymes, in their affinity for the substrate (Fig. 3.28A). Indeed, K_m values were 0.2239 ± 0.017 mM and 0.1519 ± 0.015 mM, respectively (p <0.01) (Fig. 3.30). This difference can better be visually appreciated in the Lineweaver-Burk plot where I transformed the data to double reciprocal plot (1/[S] vs. $1/v_0$) that was linear over the range of substrate concentration used, as shown in Fig. 3.28B. The decrease in K_m was not accompanied by any significant variation in V_{max} and consequently in K_{cat} between the two enzymes. Data from wt and 1st ESE enzymes showed normal distribution. The statistical significance was evaluated using Students t-test. The kinetic parameters are listed in the Table 3.1.

Taken together these findings suggest that when specific residues in the 1st ESE, i.e., G93 and A94 are mutated for that of TNAP an increase in substrate-binding affinity to the catalytic site of the enzyme occurs, consistent with selection towards a more favourable amino acid content after release from splicing-related constrain. However, the observed difference in the Km or Vmax of mutant enzymes in which the amino acids in the region of the ESE where substituted for those of the TNAP could take a different direction respect to the original TNAP protein activity. This was observed with the substitution of Ala93Gly and Gly94Ala in PLAP 1st ESE mutated enzyme that showed an increase in the affinity for the substrate (K_m= 0.1519 ± 0.015) compared with wt TNAP (K_m= 0.4365 ± 0.05). Several

reasons can explain this behaviour, firstly it is important to keep in mind that the identity between the two enzymes is approx. 50%, hence, mutating the amino acids of PLAP, corresponding to the ESE region for that of TNAP, does not necessarily mean that the Km and Vmax of PLAP will become more TNAP. Secondly although the residues involved in the active site and those coordinating the two zinc atoms and magnesium ion are largely conserved among the family, any amino acid change in the active site could change the conformation of the binding-pocket altering the catalytic properties in an unpredictable way. The important conclusion from these experiments is that introducing residues present in TNAP, where the selection at the splicing level did not occur because of the presence of strong 3'ss, in the context of PLAP it was possible to release just a portion of this protein from splicing-related constrain, allowing a better performance.

A



B

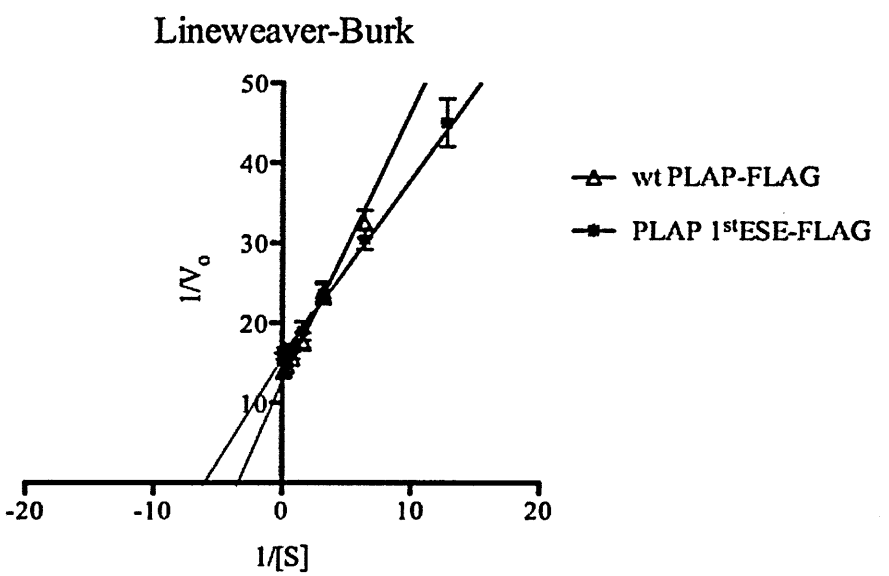


Figure 3.28. The Michaelis-Menten and Lineweaver-Burk plots for the hydrolysis of p-NPP by wt PLAP-FLAG (△) and PLAP-FLAG 1st ESE mut [G93A;A93G] (✳). Eight ng of enzyme preparation were added to 200 μl of DEA buffer containing different concentration of p-NPP (from 0.01 to 20 mM). (A) The initial velocities (v_0) were related to the concentration of substrate, [S], and the data were fitted in Michaelis-Menten equation. (B) In the Lineweaver-Burk plot the data were transformed to double reciprocal plot (1/[S] vs. $1/v_0$) the Y-axis and X-axis interceptions represent $1/V_{max}$ and $-1/K_m$ respectively.

3.6.2 Effect of Arg125>Gln amino acid substitution on the kinetic activity of PLAP

The PLAP-FLAG 2nd ESE [R125Q] enzyme was also tested for its activity. As in the previous experiment the initial velocities were plotted against the different concentrations of pNPP and values for the kinetic parameters Vmax and Km were derived by fitting data directly to the Michaelis-Menten equation (Fig. 3.29A)(GraphPad Prism 2.0). To analyze a possible effect of this mutation on enzymatic activity, I compared the V_{max}, K_m, and K_{cat} of the mutant enzyme to the wt. The best-fit values for the kinetic parameters revealed a significant difference, between the wt and 2nd ESE mutant recombinant enzymes, in their maximal rate. Indeed, V_{max} were estimated to be 0,07512 ± 0,001181 $s^{-1}\cdot\mu M$ and 0,1030 ± 0,001 $s^{-1}\cdot\mu M$, respectively (p <0.001) (Fig. 3.31). Accordingly, the mutant enzyme showed higher K_{cat} (329.6 ± 0.0009 s^{-1}) than PLAP wt (240.4 ± 0.001 s^{-1}) and consequently higher K_{cat}/K_m value (1.543 $s^{-1}\cdot\mu M^{-1}$). The increase in V_{max}, and consequently in K_{cat}, was not accompanied by any significant variation in Km (Table 3.1). Even data from 2nd ESE mut [R125Q] enzyme showed normal distribution. The statistical significance was evaluated using Students t-test.

Also for PLAP-FLAG 2nd ESE [R125Q] enzyme, I transformed the data to Lineweaver-Burk plot (1/[S] vs. $1/v_0$), as shown in Fig. 3.29B. In Lineweaver-Burk plot, the V_{max} values were represented by y-intercept of the graph ($1/V_{max}$) but the differences were graphically less pronounced than in the Michaelis-Menten plot.

These findings indicate that when the amino acid R125, associated with ESE function in the 2nd ESE region, is mutated for that of TNAP the consequence in protein function is a more rapid movement of substrate into and out the active site increasing the maximal rate and the catalytic efficiency of the enzyme.

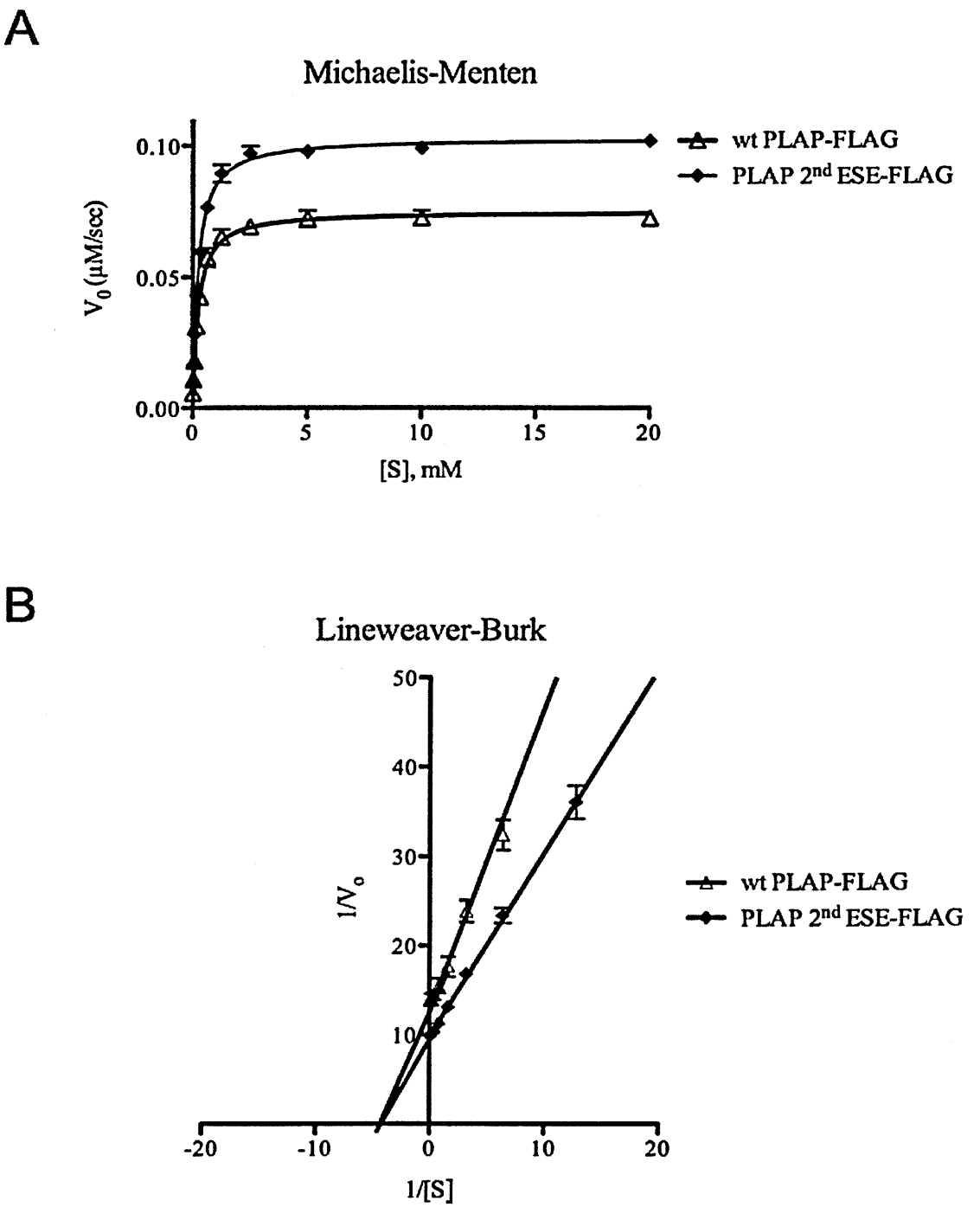


Figure 3.29. The Michaelis-Menten and Lineweaver-Burk plots for the hydrolysis of p-NPP by wt PLAP-FLAG (△) and PLAP-FLAG 2nd ESE mut [R125Q] (◆). Eight ng of enzyme preparation were added to 200 µl of DEA buffer containing different concentration of p-NPP (from 0.01 to 20 mM). (A) The initial velocities (v_o) were related to the concentration of substrate, [S], and the data were fitted in Michaelis-Menten equation. (B) In the Lineweaver-Burk plot the data were transformed to double reciprocal plot (1/[S] vs. $1/v_0$) the Y-axis and X-axis interceptions represent $1/V_{max}$ and $-1/K_m$ respectively.

3.6.3 Effect of Gly93>Ala, Ala94>Gly and Arg125>Gln amino acid substitutions on the kinetic activity of PLAP

To further analyze the possible synergic role played by A93G/G94A and R125Q, in the phosphatase activity of PLAP, I expressed an engineered double mutant enzyme the PLAP-FLAG 1^{st} /2^{nd} ESE mut [A93G; G94A; R125Q]. When the recombinant enzyme was tested for its activity mutation of these three amino acids in PLAP clearly affected alkaline phosphatase catalysis in PLAP-FLAG 1^{st} /2^{nd} ESE mut [A93G; G94A; R125Q], abrogating the activity altogether.

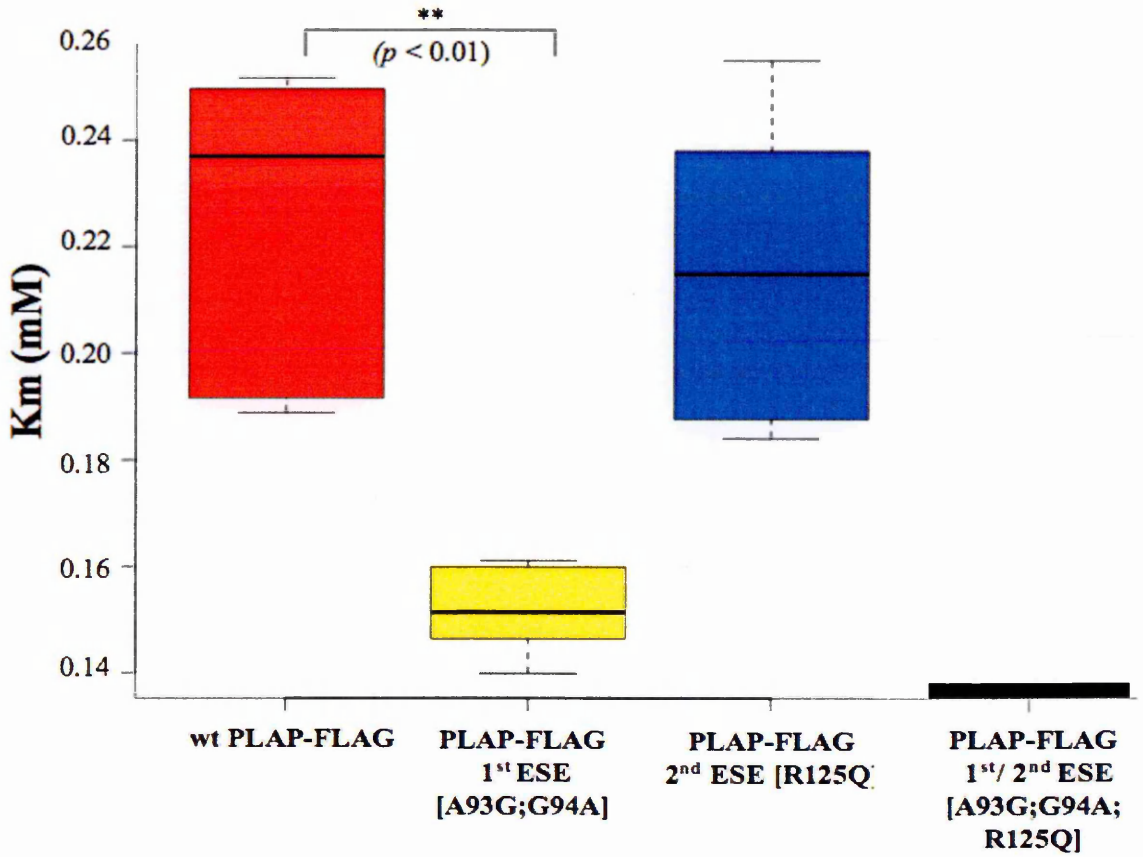


Figure 3.30. Box-plot showing the Michaelis-Menten constant (Km) in wt PLAP-FLAG, PLAP-FLAG 1st ESE mut [A93G; G94A], PLAP-FLAG 2nd ESE mut [R125Q] and PLAP-FLAG 1st /2nd ESE mut [A93G; G94A; R125Q]. Middle horizontal line inside box indicates median. Bottom and top of the box are 25th and 75th percentiles, respectively (n=6).

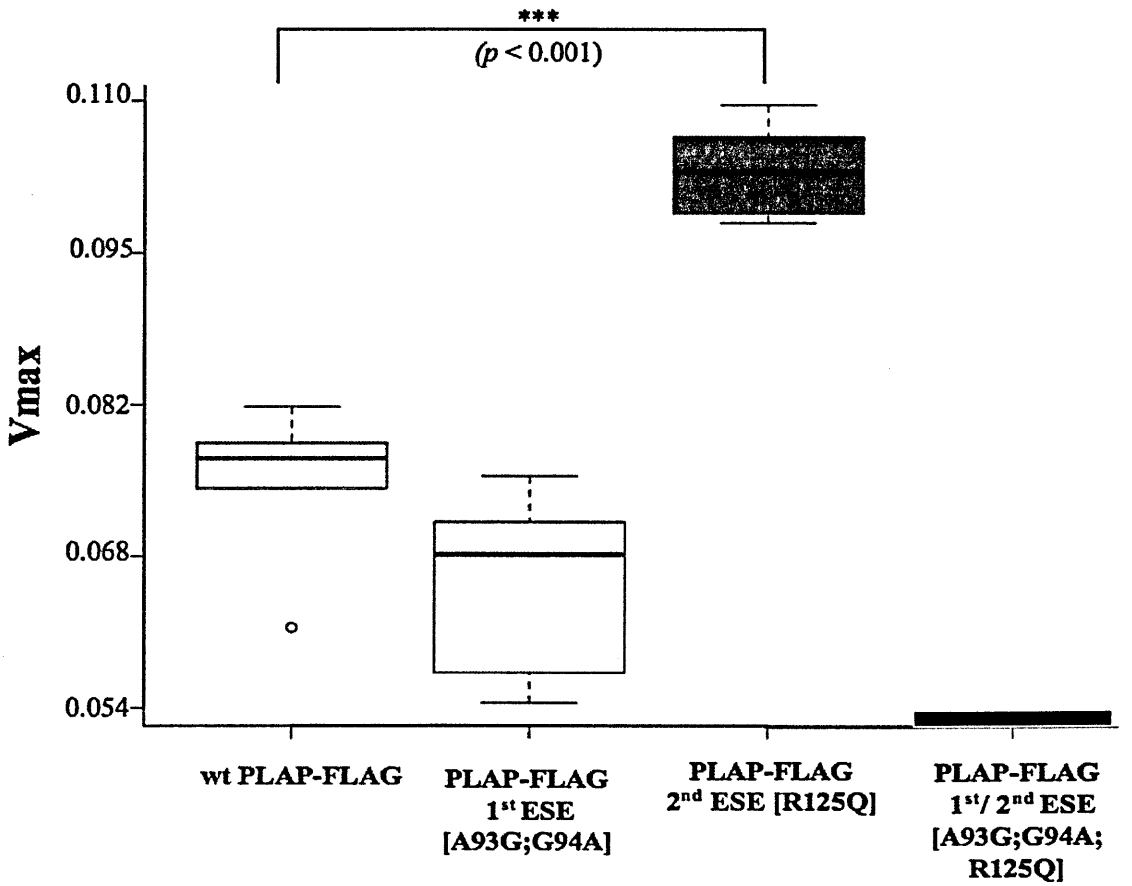


Figure 3.31. Box-plot showing the maximal rate of reaction (Vmax) in wt PLAP-FLAG, PLAP-FLAG 1st ESE mut [A93G; G94A], PLAP-FLAG 2nd ESE mut [R125Q] and PLAP-FLAG 1st /2nd ESE mut [A93G; G94A; R125Q]. Middle horizontal line inside box indicates median. Bottom and top of the box are 25th and 75th percentiles, respectively. The outliers are represented as circles and are the points beyond the lower and upper extreme values of the box plot (n=6).

TABLE

Kinetic Parameters of ALP and Mutants

Enzyme *Alkaline Phosphatase (ALP)*	*pNPP*			
	K_m *(mM)*	V_{max} *(μMs^{-1})*	K_{cat} *(s^{-1})*	K_{cat}/K_m *($s^{-1}\ \mu M^{-1}$)*
wt PLAP-FLAG	0.2239 ± 0.017	0,07512± 0,0011	240.4 ± 0.001	1.073
wt TNAP-FLAG	0.4365 ± 0.05	0,1719 ± 0,005	573 ± 0.0013	1.312
PLAP -FLAG 1st ESE [G93A;A94G]	0.1519 ± 0.015	0,06637 ± 0,0013	212.4 ± 0.0013	1.398
PLAP-FLAG 2nd ESE [R125Q]	0.2135 ± 0.009	0,1030 ± 0,001	329.6 ± 0.0009	1.543
PLAP-FLAG 1st ESE [G93A;A94G] /2nd ESE [R125Q]	ND	ND	ND	ND

Measurements were done at pH 9.8 using pNPP as substrate in the presence of 1 mM of $MgCl_2$ and 20μM of $ZnCl_2$.

Table 3.1. Kinetic parameters of ALP and mutants. The difference in the Km or Vmax of mutant enzymes in which the amino acids in the region of the ESE where substituted for those of the TNAP could take a different direction respect to the original TNAP protein activity, as observed with PLAP 1st ESE mutated enzyme that showed an increase in the affinity for the substrate (K_m) compared with wt TNAP.

4. DISCUSSION

The existence of exonic splicing enhancers (ESE) is well established and confirmed by many reports through over two decades of research. In fact, it has been demonstrated that ESEs participate in exon recognition, controlling both alternative and constitutive splicing processes (Wang and Burge, 2008) and that their presence is widely distributed among eukaryotic organisms, having a very early evolutionary origin (Webb et al., 2005). The fact that ESE-mediated splicing was an early phenomena in eukaryotic organisms makes for fascinating implications with regards to the potential relationships between coding and splicing regulatory regions during the course of evolution.

Although it is plain that these ESE overlap with coding capacity, the effect of the conflict between ensuring splicing efficiency and preserving the coding capacity has not been specifically analyzed. The impact that the double purpose of the coding sequence might have on gene evolution is that proteins may not be as optimised as they could be, due to the fact that their sequence has to comply with possible conflicting functions: the pressure to maintain the exon included in the mature mRNA and the selection of sequence variants that encode amino-acids that enhance the particular protein function. The former selective pressure might be more important than the latter, because exon inclusion in the final mRNA is a pre-condition for its translation (Pagani and Baralle, 2004). Consequently, this suggests that any amino acidic substitution should be selected also at splicing level in order to occur and confer particular advantages to protein performance.

In keeping with this concept, as already mentioned in the introduction, several studies have uncovered the presence of extensive purifying selection against substitutions in ESE as determined by a strong bias in codon choice at the intron-exon junction, with codons favoured in known ESEs found in the boundaries (Parmley et al., 2007).

Selection operating on splice control elements is further supported by the rarity of SNPs (single-nucleotide polymorphisms) density in these regions, this owing to the need to preserve their function (Carlini and Genut, 2006; Fairbrother et al., 2004a). Analogously, analysis of retrogenes in which the residues located in the original parental copy were at the intron-exon junctions showed that they have a higher rate of evolution. This suggest that constraints exist near intron-exon borders that have been released in the retrogenes and the higher rate of evolution in these sites is coherent with selection towards a more favourable amino acid content after release from splicing-related selection (Parmley et al., 2007). However it should be kept in mind that in these studies the presence of the ESE were only hypothetical and caution should be employed when making these kind of comparisons on the basis of bioinformatic studies (Irimia et al., 2009). For this reason, it is better to support any eventual conclusion with functional experiments that might either sustain or not the bioinformatics considerations. At the experimental level, the observation that splicing is subject to intense selective pressure during the course of evolution has been shown to occur in several exons such as CFTR exon 9 and 12 and fibronectin EDA exon (Haque et al., 2010; Pagani et al., 2005; Pagani et al., 2003a; Zago et al., 2011). The effect on the splicing efficiency indicates that exonic sequences are significantly constrained by splicing requirements that in turn restrict evolutionary productive genome variability. Therefore missense mutation can segregate in human population during the evolution and can be responsible of deleterious effects on protein function and the fate of this change may be the elimination by purifying selection (Hellmann et al., 2003) or the maintenance to preserve a proper exon inclusion.

Since several studies have been performed on the possible evolutionary connections between coding capacity and splicing regulation in different species it was also interesting to analyze how splicing regulatory elements evolve in the human genome, and how they behave in gene duplication events. Duplicate genes arise frequently in eukaryotic genomes and generally undergo accumulation of mutations. Because deleterious mutations occur

much more frequently than beneficial ones there is a higher probability that these mutations lead to loss of function of one gene copy, becoming pseudogenes. Nevertheless, it is also possible that duplicate genes are preserved for long time periods, even if the mechanism is still unclear. In addition to the classical model for the evolution of gene duplicates, in which the only mechanism by which a duplicate gene can be preserved is the acquisition of new beneficial function (Ohno, 1970) an alternative model was proposed by Force et al. (Force et al., 1999). In this model gene duplicates are frequently preserved by subfunctionalization in which complementary degenerative mutations in different regulatory elements of duplicated genes can facilitate the preservation of both duplicated genes and after these mutations each daughter gene adopts part of the functions of their parental gene. Indeed, several observations indicate that a common fate of the members of duplicate-gene pairs is the partitioning of tissue-specific patterns of expression of the ancestral gene that might reduce the pleiotropic constraints (Lynch and Force, 2000).

In this thesis I have attempted to address the question whether the presence of ESEs conditioned the coding capacity in human genes by investigating the evolutionary constraints of ESE sequence maintenance on the enzymatic activity of the paralogous ALP protein family.

4.1 ESE Analyzer Web Server (EAWS) computational analysis.

Identifying a candidate paralogous protein family

In the first part of this study a bioinformatic platform ESE Analyzer Web Server (EAWS) was developed in collaboration with Dr. Vlahovicek (University of Zagreb, Croatia), and used for the identification of candidate protein families in which the evolution of mRNA sequences was conditioned by the presence of exonic cis-acting splicing regulatory elements. These elements encode also for amino acid that, even if sub-optimal for protein function, were maintained in the course of evolution most likely only to ensure a

correct splicing. In particular EAWS is able to integrate information from different sources available to date regarding splicing from the existing web servers (i.e., Gene Splicer, FAS-ESS, ESE RESCUE, ESEfinder). It also allows, at the same time, to align transcripts and protein sequences of similar functional domains and in particular in this study I looked at human paralogous gene families to find out examples in which the need for ESE may influence the protein activity of evolutionary correlated proteins (paralogs).

By using this platform, the most promising candidate protein family were those of the human *Alkaline Phosphatase* paralogous genes (ALPs; Fig. 3.5) in which the need for ESE has influenced the codon choice, protein activity and most likely the evolution of this protein family. Specifically, the exon encoding for the active site of the enzymes showed compensatory changes in acceptor strength followed by changes in ESE and amino acid variations close to the active Serine among the paralogs.

Although not the aim of the study, this type of tool might in the future enhance the positive hit rate of splicing regulatory elements (SRE) predictions. This would be a useful tool in clinical diagnostics, especially if we consider that in study of neurofibromatosis type1 (NF1)(Ars et al., 1999) and ataxia teleangectasia (ATM)(Teraoka et al., 1999) up to 50 % new mutations were found to be splicing mutations and their identification as such is of paramount importance. Indeed, a key issue raised in molecular diagnosis is the correct interpretation of the biological consequences of a gene sequence variant in which the real disease-causing mutation has not yet been identified; e.g., their putative impact on splicing. Accordingly, as mentioned above several lines of evidence suggest that many unclassified genetic variants might turn out to result in splicing abnormalities (Baralle and Baralle, 2005; Baralle et al., 2009).

The reason that the role of splicing mutations has been realized relatively late is because it has been difficult to show a clear correlation between the suspected mutation and disease, in contrast with other types of genetic mutation, in which this connection is relatively straightforward. From this exigency the usage of more reliable software to predict

whether a given variation would interfere with mRNA splicing would be useful for genetic counselling in the clinic.

The EWAS platform will align paralogous genes and analyze the transcripts for the presence of splicing regulatory elements, this ability to integrate information from different sources and to fetch directly from Ensembl transcripts belonging to the same family would increase the ease with which to analyze such mutations. For example, if a mutation occurs in an area very conserved among the family members and corresponds also to a hypothetical ESE throughout the family most likely such mutation will be deleterious for the protein. EWAS also calculates splice site strength and the user could easily check in the user friendly output, the splice site strength of the exon within which the mutation is found. Since it is difficult to experimentally analyze each sequence variant that could affect gene function in a routine, high-throughput manner, EAWS represents an easy source for clinicians and researchers, and can help to save a lot of work and money, identifying the most likely potential disease-causing splicing mutation.

4.2 Validation of bioinformatics ESE predictions

I have used a minigene system and mutagenesis analysis to confirm the presence of ESEs. Our data shows that exon 4 of ALPP contains two enhancer elements (1^{st} and 2^{nd} ESE) necessary for promoting exon inclusion due to the weak upstream 3'ss (section 3.2). The substitution of these sequences with corresponding one of ALPL, whose nucleotide content is not associated with ESE function, resulted in exon 4 skipping (Fig. 3.8; 3.12-3.14) while improving the splice site strength exchanging the ALPP 3'ss with that of ALPL made ESEs superfluous (Fig. 3.15- 3.16).

On the other hand, the results of our analysis have also demonstrated that in the exon 5 of ALPL (corresponding to exon 4 of ALPP), the lack of ESE in the regions 1^{st} seq and 2^{nd} seq (Fig. 3.8) most probably reflect the presence of strong 3'ss, indeed in this case the exon is

correctly recognized even in absence of splicing enhancers, conversely reducing the 3'ss strength, the exon 5 is no more able to correctly include the exon in the mature mRNA (Fig. 3.17).

Taken together, these results suggest that there is a tightly regulated network between the strength of the splice sites and the other splicing cis-acting regulatory elements in the exons under study controlling the outcome of splicing event.

4.3 Testing the biochemical effect of the amino acid differences within the enhancer elements of PLAP and corresponding region of TNAP, the protein products of ALPP and ALPL respectively

As mentioned before, the realisation that ESEs overlap with coding capacity is well established, however the effect of ensuring splicing efficiency and preserving the coding capacity has not been specifically analyzed to date. In this thesis I mapped two ESEs in ALPP not present in ALPL. These ESEs correspond to an area that in TNAP differs in amino acid sequence to PLAP representing a possible splicing constraint on the enzymatic activity of the protein. Indeed, substituting the amino acids in the region of the ESE for those of the TNAP, where the ESEs are absent, highlighted a significant constraint of the ESE on protein function. It was then of interest to analyse the effects these substitutions had on the enzymatic activity of the protein. That this was indeed the case in this scenario was quite high as these amino acid differences were close to the Ser92, the phosphate-binding residue involved in the catalytic reaction as determined from inspection of the crystal structure of the PLAP active site. Indeed substitution of Ala93Gly and Gly94Ala in PLAP produced a significant increase in the affinity for the substrate compared with wt PLAP (Fig. 3.28; Table 3.1). The single amino acid substitution Arg125Gln in PLAP that corresponded to the amino acid difference in the second ESE mapped in ALPP increased the protein performance improving the maximal rate with a faster processing of the

substrate in the active site pocket of the enzyme (Fig. 3.29; Table 3.1). The exchange of the amino acids corresponding to both the ESEs resulted in the complete abolition of the function (Table 3.1).

These observations confirm the hypothesis that sequences that perform both coding and splicing regulatory functions introduce splicing–related constraints on amino acid sequences that may affect protein performance.

Interestingly, although this is not the aim of the study, this kind of research can also be used to develop proteins for gene therapy or biotechnological application. Indeed, the usage of modified enzyme with increased catalytic constants will allow a good employment in the industries where proteins with better performance respect to the native condition could find several applications.

4.4 Alkaline phosphatase and evolution

Current views of the evolutionary history of ALP propose that the ancestral gene expresses tissue-unspecific alkaline phosphatase, as for example occurs with the E. coli enzyme, and that the actual tissue unspecific isoenzyme that populate liver, bone, kidney and the first-trimester placenta (ALPL) maintains the ancestral gene's characteristics while the other paralogs later acquired the tissue- specificity together with other features. Indeed, through a process of gene duplication and additional mutations, the intermediate intestinal gene arises from the tissue-unspecific precursor gene (ALPL like) and subsequently becomes the progenitor of intestinal, germ cell and placental genes (Fishman, 1990) (Fig. 4.1). In particular, ALPP gene represents a late evolutionary event since enzymes with the properties of PLAP (protein product of ALPP gene) were found expressed only in placenta of chimpanzee, orangutan and human (Doellgast and Benirschke, 1979) while PLAP and

TNAP (protein product of ALPL gene) are equivalent in several other mammalian species (Harris, 1980). The human Alkaline Phosphatase paralogous gene family has already been proposed as an example of subfunctionalization since these genes are expressed in different tissues still preserving the same general function (Rump et al., 2001). In particular, the tissue specificity can be explained by different alterations in the promoter regions of the genes, consistent with the findings from alkaline phosphatase promoter studies (Kiledjian and Kadesch, 1990; Millan, 1987).

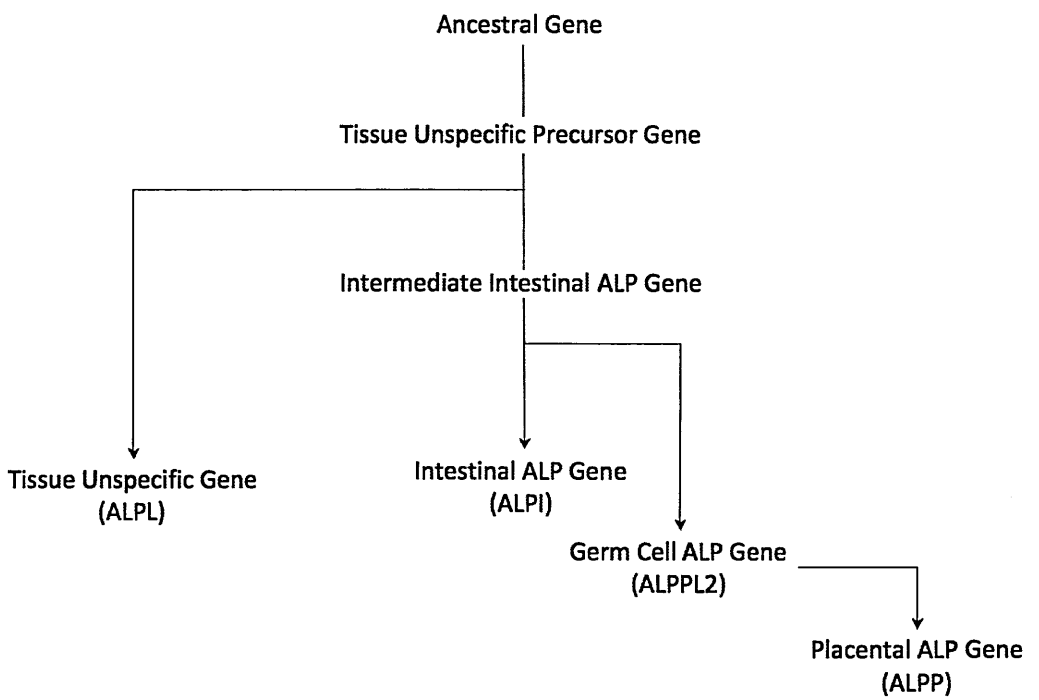


Figure 4.1. Possible evolutionary path of the ALP isoenzymes. Figure taken from Fishman (1990).

The ESEs I have mapped in ALPP, that are absent in ALPL, are necessary due to a weak 3'ss in the former. Since, as mentioned at the beginning of this section, ALPL has the same characteristics as the ancestral copy, these changes in splicing regulatory elements, within the catalytic exon of ALPP, from ALPL (ancestral like) to ALPP (later evolutionary event)

have followed the evolutionary history of this family as summarized in the model depicted in figure 4.2. Specifically, after the last common ancestral gene, which accounts for ALPL gene, at least two hypotheses are possible to explain the temporal succession of these changes in splicing motifs, but of course many others are feasible. The possible pathways imply always gene duplication from the unique initial non-tissue specific gene (ALPL like) and additional mutations.

In particular, the first hypothesis (Fig. 4.2; left panel), represents a scenario where the first random mutations in the exon led to creation of the ESE and consequently variations in the amino acid sequence; these mutations would not have any consequence on the splicing but only possibly on the protein function since a strong 3'ss is still present. Thus any amino acid change that may have occurred would produce a functional protein with the characteristics of ALPP, hence different from the parental copy. Subsequently, the second nucleotide changes could occur either in the exonic region generating another ESE or within the 3' ss. If within the exonic region, it would have no effect on splicing but the outcome would only be on protein activity (for the same reasons explained above). If, on the other hand, it were to occur in the 3' ss reducing the strength I have previously demonstrated that with just one ESE present associated with weak 3' splice site aberrant splicing would occur. However, this is only around 25 percent therefore the majority of the mRNA molecules can still be correctly synthesized. After that a second ESE was created by chance allowing the complete inclusion of the exon.

The second hypothesis (Fig. 4.2; right panel), is based on the splicing compensation model (Ke et al., 2008), mutation in the splice site occurred first, weakening the 3' ss strength, even if we don't know to which extent (maybe still providing partial exon inclusion), after that a compensation in the form of a gain of two ESEs, concurrent or subsequent, preserved the complete inclusion of the exon, hence, a correct splicing. However the latter alternative (Ke et al 2008) does not explain how the exon was maintained without drifting until the ESE was formed unless all the mutations, i.e. ESE and 3'ss, were simultaneous. However,

compensatory model paints a picture of exon evolution as a dynamic interplay between helpful and harmful mutations, continuously at work.

Since ALPP is 90% homologous to the other two tissue specific isoenzymes most likely the same results obtained in ALPP would occur also in ALPPL2 and ALPI, making possible to extend these theory to a more general evolutionary theory of splicing motifs from the tissue unspecific ancestral gene to the less distant tissue specific isoforms.

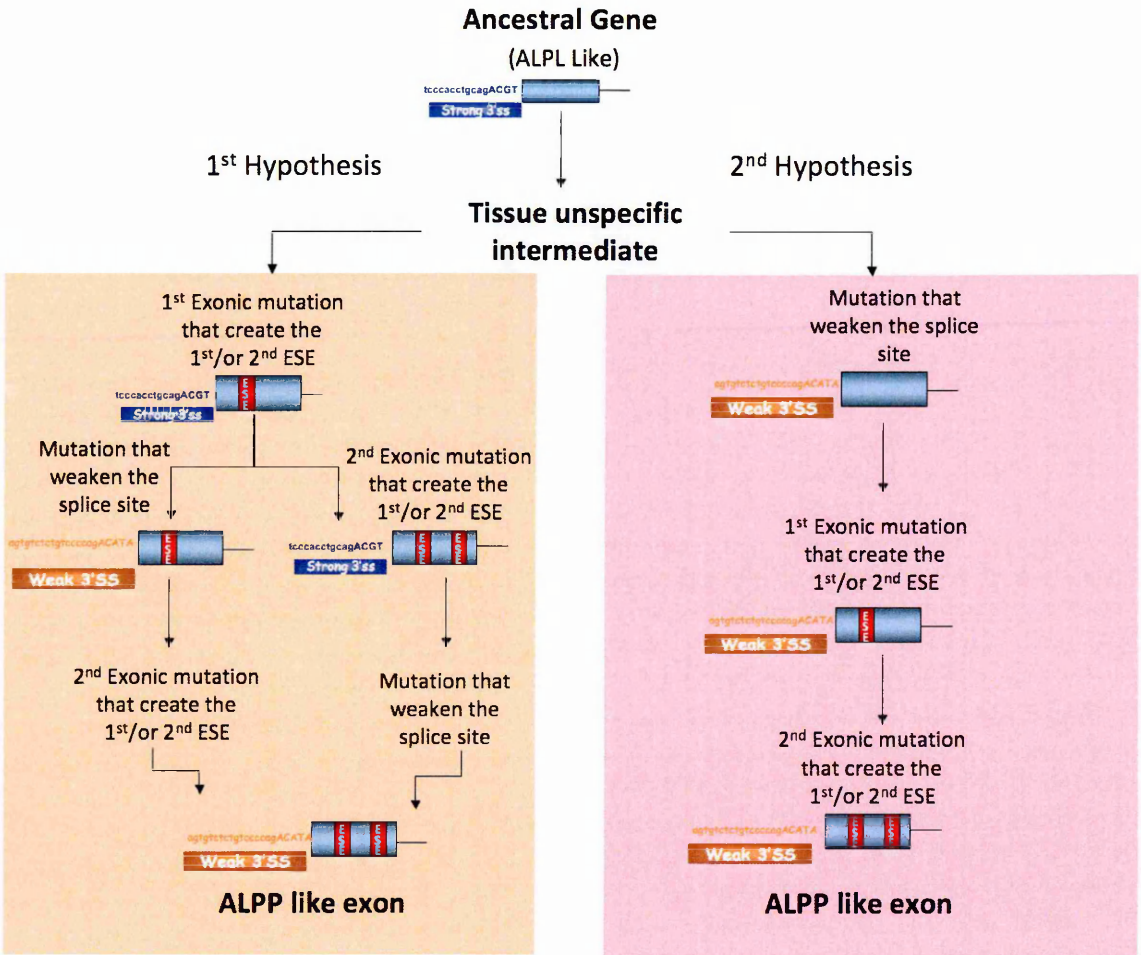


Figure 4.2. Two possible hypotheses to explain the temporal succession of changes in splicing motifs in the ALP tissue-specific exon 4. The hypothesis on the left represent a scenario where the first casual mutations generated occurs in the region of the ESE this casual positive event made possible second nucleotide changes that could occur either within the 3' ss or in the exonic region generating another ESE. In the first case "degenerative" mutations that weakened the 3'ss strength, in presence of only one ESE will cause partial exon skipping, later on a second ESE was created by chance or, in the second case, firstly occurred a casual positive formation of the second ESE that allowed the mutations that weaken the splice site to take place, as the correct splicing was already fully guaranteed. Second hypothesis on the right, is based on the splicing compensation model

(Ke et al., 2008), mutation in the splice site occurred first, weakening the 3' ss strength, even if we don't know to which extent. Subsequently compensatory changes, such as the creation of new ESEs, act to ameliorate the effects of the deleterious mutations and are positively selected.

5. CONCLUSIONS

In conclusion, in this study I have identified a family of paralogous genes where there is interplay between splice site strength and presence or absence of ESE sequences. These ESE sequence also dictate in part the catalytic characteristics of the enzyme. The splicing constraints thus generated a suboptimal paralogous protein that was tolerated to allow the persistence of sequences that are essential for exon inclusion.

Also, this study has established experimentally that the relationship between the constraints of ESE sequence on amino acid variability are enough to strongly condition the evolution of the protein. If a certain function would be desirable for a certain physiological role of the protein this may not be possible unless there are other changes that can ensure a correct splicing of the exon.

Hence, the changes A93G, G94A and R125Q, that may improve the substrate binding affinity and/or enzyme kinetics in PLAP, are not compatible with enzyme function as long as other splicing determinants (weak splice site) are in place.

Finally it is important to note, as mentioned above, that this kind of studies are not only important for a general understanding of protein evolution but can also help to develop proteins for biotechnological application or gene therapy, where it is crucial to optimize the protein performance.

6. MATERIALS AND METHODS

6.1 Chemical reagents

General chemicals were purchased from Sigma Chemical Co., Merck, Gibco BRL, Boehringer Mannheim, Carlo Erba and Serva.

6.1.1 Standard solutions

All the solutions are identified in the text when used apart from the following:

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

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

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

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

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

f) TBS: 50 mM Tris-HCl, 150 mM NaCl, 2 mM CaCl2, pH 7.5

6.2 Enzymes

Restriction enzymes were from New England Biolabs, Inc. DNA modifying enzymes such as Taq Polymerase, DNaseI RNase free, and T4 DNA ligase were obtained from Roche Diagnostic. Klenow fragment of *E. coli* DNA polymerase I and T4 polynucleotide Kinase were from New England Biolabs, Inc. RNase A was purchased from Sigma Chemicals Ltd. A 10 mg/ml solution of RNase A was prepared in sterile water and boiled for 10 minutes to

destroy trace amounts of DNase activity. All enzymes were used following manufacturer's instructions.

6.3 Synthetic oligonucleotides

Synthetic DNA oligonucleotydes were purchased from Sigma-Genosy and IDT (Integrated DNA Technologies).

6.4 Bacterial culture

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

6.5 Cell culture

The cell line used for transfection and cotransfection experiments were:

a) HeLa cells, an immortal cell line derived from cervical cancer cells.

b) COS-1 cell, an immortal cell line is derived from CV-1 simian cells.

6.6 DNA preparation

6.6.1 Small scale preparation of plasmid DNA from bacterial cultures

The alkaline lysis of recombinant bacteria was performed by resuspending the bacterial pellet in 200 µl of ddH2O; 150 µl of solution II (0.2 M NaOH, 1 % w/v SDS) were then added and the contents mixed by inversion. 250µl of solution III (3 M potassium acetate pH 5.2) were then added and the contents mixed by inversion. The bacterial lysate was then centrifuged in an Eppendorf microcentrifuge at maximum speed and the supernatant transferred to a new tube. An equal volume of 1:1 v/v phenol:chloroform solution was added to the supernatant. The tube was then vortexed and centrifuged as above. The aqueous phase containing the DNA was transferred to a new tube. An equal volume of chloroform was added to the supernatant. The tube was then vortexed and centrifuged as above. The aqueous phase containing the DNA was then recovered and the DNA pelleted by ethanol precipitation. The final pellet was resuspended in 50 µl of ddH2O and 5 µl of such preparation were routinely taken for analysis by restriction enzyme digests.

6.6.2 Large scale preparations of plasmid DNA from bacterial cultures

For large-scale preparations of plasmid DNA that was necessary for the transfection experiments, JetStar purification kit (Genomed) was used according to the manufacturer's instructions. In order to get a good amount of plasmid, an inoculation in 50 ml of LB medium is grown overnight at 37°C.

6.7 RNA preparation from cultured cells

Cultured cells were washed with PBS and RNA extracted using RNA Trizol, (Invitrogen inc) according to the manufactures instructions. Briefly, 750ul of Trizol was added/p6 well and allowed to incubate for 5 minutes. The solution was subsequently moved to a 1.5ml eppendoff tube and 200ul of chloroform was added. After centrifugation at 10000 rpm the supernatant was collected and the RNA precipitated with 0.75 vol. of isopropanol. The pellet was resuspended in 100 µl of ddH2O and digested with 1U of DNase RNase-free by incubation at 37 °C for 30 minutes, and then the RNA was purified by acid phenol extraction. The final pellet was resuspended in 35 µl of ddH2O and frozen at –80 °C. The RNA quality was checked by electrophoresis on 1% agarose gels.

6.8 Estimation of nucleic acid concentration

The RNA quantity was detected using the Nanodrop spectrophotometer instrument (Thermo Scientific) and equal amounts were used for cDNAs synthesis. The ratio of values for optical densities measured at 260 nm and 280 nm is considered as 1.8 for pure sample of DNA and 2 for RNA and these are reduced by protein contaminants (Sambrook et al., 1989). Therefore, these values were used to determinate not only the concentration but also the purity of the samples.

6.9 Enzymatic modification of DNA

6.9.1 Restriction enzymes

Restriction endonucleases were used in the construction and analysis of recombinant plasmids. Each restriction enzyme functions optimally in a buffer of specific ionic strength. All buffers were supplied by the same company that supplied the enzymes and were used according to the manufacturer's instructions.

For analytical digests 100-500 ng of DNA were digested in a volume of 20 µl containing 5 U of the appropriate restriction enzyme. The reaction was incubated for 2-3 hours at 37 °C. Preparative digestion was made of 5-10 µg DNA using the above conditions and 5 U of restriction enzyme for µg of DNA in 200 µl reaction volume.

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

These enzymes were used to treat PCR products for blunt-end ligation during construction of recombinant plasmids. The large fragment of DNA Polymerase I (Klenow) is a proteolytic product of *E. coli* DNA Polymerase I. It retains polymerization and 3'->5' exonuclease activity, but has lost 5'->3' exonuclease activity. This was useful for digesting specific residues added by Taq DNA polymerase at the 3' terminus to create compatible ends for ligation. T4 Polynucleotide Kinase catalyses the transfer of phosphate from ATP to the 5' hydroxyl terminus of DNA. It was used for example in the addition of 5'-phosphate to PCR products to allow subsequent ligation. Klenow fragment (2.5 U) was added to 23 µl of PCR product in 5 mM $MgCl_2$ buffer. The mixture was incubated at room temperature for 10 minutes. EDTA to a final concentration of 0.2 mM, ATP to a final concentration of 1 mM, 10 U of T4 Polynucleotide Kinase and the proper quantity of Kinase buffer were

added to the above mixture and incubated at 37 °C for 30 min. The enzymes were inactivated by incubation at 80 °C for 20 min.

6.9.3 T4 DNA ligase

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

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

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

6.10 Agarose gel electrophoresis of DNA

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

6.11 Elution and purification of DNA fragments from agarose gels

This protocol was used to purify small amounts (less than 1 μg) of DNA for sub-cloning. The DNA samples were electrophoresed onto an agarose gel as described previously. The DNA was visualized with UV light and the required DNA fragment band was excised from the gel. This lab was cut into pieces, and the JETquick Spin Column Technique (Genomed) was used according to the manufacturer's instructions. Briefly, 600 μl of gel solubilisation solution L1 ($NaClO_4$, Na acetate and TBE) were added for each 100 mg of the gel slice pieces and incubated at 55 °C for 15 min vortexing every 5 min. The mixture was loaded into a prepared JETquick column and it was centrifuged at maximum speed for 1 min. The flowthrough was discarded. 700 μl of washing and reconstituted solution L2 (ethanol, NaCl, EDTA and Tris-HCl) were added into the spin column and after 5 min, the column was centrifuged in the same conditions twice. The flowthrough was again discarded both times. To elute the bound DNA, 30-50 μl of pre-warmed sterile water were added onto the centre of the silica matrix of the spin column and the system was centrifuged for 2 min. The amount of DNA recovered was approximately calculated by UV fluorescence of intercalated ethidium bromide in an agarose gel electrophoresis .

6.12 Preparation of bacterial competent cells

Bacterial competent cells, *E. Coli* strains, were grown overnight in 3 ml of LB at 37°C. The following day, 300 ml of fresh LB were added and the cells were grown at room temperature for 4-5 h until the OD_{600} was 0.3-0.4. The cells were then put in ice and centrifuged at 4 °C and 1000g for 15 min. The pellet was resuspended in 30 ml of cold TSS solution (10% w/v PEG, 5% v/v DMSO, 35mM Mg Cl2, pH 6.5 in LB medium). The cells were aliquoted, rapidly frozen in liquid nitrogen and stored at –80°C. Competence was determined by transformation with 0.1 ng of pUC19 and was deemed satisfactory if this procedure resulted in more than 100 colonies.

6.13 Transformation of bacteria

Transformations of ligation reactions were performed using 1/2 of the reaction volume. Transformation of clones was carried out using 20 ng of the plasmid DNA. The DNA was incubated with 60 µl of competent cells for 20 min on ice and at 42°C for 1.5 minutes. At this point 60 µl of LB were added and the bacteria allowed to recover for 10 min at 37 °C. The cells were then spread onto agar plates containing the appropriate antibiotic. The plates were then incubated for 12-15 hours. When DNA inserts were cloned into β-galactosidase-based virgin plasmids, 25 µl of IPTG 100 mM and 25 µl of X-Gal (4 % w/v in dimethylformamide) were spread onto the surface of the agarose before plating to facilitate screening of positive clones (white colonies) through identification of β-galactosidase activity (blue colonies).

6.14 Amplification of selected DNA fragments

The polymerase chain reaction was performed on genomic or plasmid DNA following the basic protocols of the Roche Diagnostic Taq DNA Polymerases. The volume of the reaction was 50 µl. The reaction buffer was: 1X Taq buffer, dNTP mix 200 µM each, oligonucleotide primers 1 nM each, Taq DNA Polymerase 2.5 U. As DNA template, 0.1 ng of plasmid or 100-500 ng of genomic DNA were used for amplification. When a DNA fragment longer than 2000bp was amplified, DMSO 3% was also added to the mixture. The amplification conditions are described for each particular PCR. The amplifications were performed on a Cetus DNA Thermal Cycler (Perkin Elmer) or on a Gene Amp PCR System (Applied Biosystems).

6.15 Sequence analysis for cloning purpose

Sequence analysis of plasmid DNA was performed using the CEQ 2000 sequencer (Beckman Coulter) or sequencing service (BMR Genomics). The plasmid DNA of interest (approximately 100 ng) was purified through a MicroSpin S-400 HR Column (Amersham Pharmacia Biotech). The DNA was then amplified using fluorescent labeled dideoxy nucleotide terminators according to the manufacturer's instructions. The samples were analized by loading them into the automatic sequencer.

6.16 Generation of minigenes

6.16.1 PCR-directed mutagenesis

In this thesis several minigenes were generated by PCR-directed mutagenesis and this has been done through PCR amplification of the exon under study with its intronic and exonic flanking regions. Two steps PCR technique has been used for generating the hybrid mutants using different minigenes as a template. A first set of PCR reactions was performed using a combination of primers: forward external primer and reverse internal primer and forward internal and reverse external primer. The DNA fragments were gel extracted and used as templates for a second PCR reaction using both external primers. The DNA insert obtained was then cloned into pUC19 vector (Fermentas) using the enzyme restriction site Sma I and through sequencing using oligonucleotides Universal For and Rev the absence of any other nucleotide variation in the entire amplified fragments was checked. Finally, the obtained DNA inserts was digested with the KpnI and XbaI enzymes (New England Biolabs), gel extracted (EuroClone) and ligated with T4 DNA ligase (New England Biolabs) into the pcDNA3 plasmid, using the KpnI and XbaI restriction sites, which carries the CMV promoter suitable for cell culture experiments. Constructs were transfected in HeLa cell lines and the minigene splicing products were evaluated by reverse transcriptase PCR (RT-PCR) analysis using the primers T7 and SP6 (Fig 6.1).

Universal forward

5'GTAAAACGACGGCCAGT3'

Universal reverse

5'GAAACAGCTATGACCAT3'

T7

5'TAATACGACTCACTATAGGG3'

SP6

5'ATTTAGGTGACACTATAGAATA3'

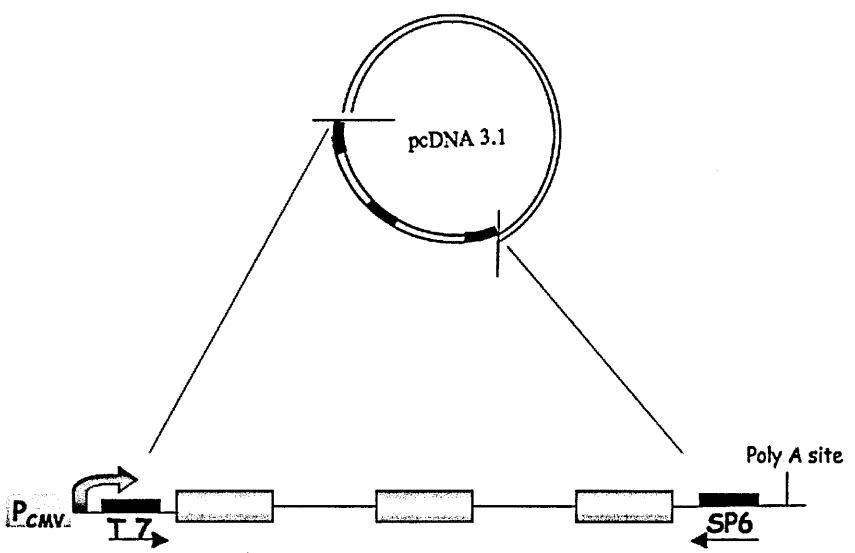


Figure 6.1. Schematic representation of pcDNA 3.1 minigene. Gray boxes represents the three exons amplified from genomic DNA. The minigene contains a functionall polyadenylation site at the 3' end and at the 5' end a CMV promoter. Black arrows before the exons indicate the position of primers,T7 and SP6, used in RT-PCR reaction when analyzing the mRNA produced by the minigene.

6.16.2 Quick Change Mutagenesis PCR method

Mutagenesis was also at times performed using the QuikChange site-directed methodology. Briefly, oligonucleotide primers, each complementary to opposite strands of the area of interest and carrying the mutant nucleotides are designed and utilized during temperature cycling by *PfuTurbo* DNA polymerase. Incorporation of the oligonucleotide primers generates a mutated plasmid containing staggered nicks. Following temperature cycling, the product is treated with *Dpn* I. The *Dpn* I endonuclease (target sequence: 5′-Gm6ATC-3′) is specific for methylated and hemimethylated DNA and is used to digest the parental DNA template and to select for mutation-containing synthesized DNA. DNA isolated from almost all *E. coli* strains is dam methylated and therefore susceptible to *Dpn* I digestion. The nicked vector DNA containing the desired mutations is then transformed into competent cells. The small amount of starting DNA template required to perform this method, the high fidelity of the *PfuTurbo* DNA polymerase, and the low number of thermal cycles all contribute to the high mutation efficiency and decreased potential for generating random mutations during the reaction (for the primers that were used see section 4.16.3). Briefly, two separate primer extension reactions were set up (one for each forward and reverse primer) containing:

5 microliters 10X Pfu Buffer (supplied with enzyme)

1 microliter 10 micromolar primer (0.13 microgram 45-mer)

0.1 – 0.2 microgram plasmid template

1 microliter 10 mM dNTP mix

H_2O to a final volume of 50 microliters

1 microliter Pfu turbo polymerase (Stratagene)

PCR reaction was done using standard conditions:

1. 94 deg, 30 sec

2. 95 deg, 30 sec

3. 55 deg, 1 min

4. 68 deg, 2 min/kb up to 10 KB plasmid

Digestion of amplification product was done by adding into a PCR reaction 10 units of Dpn I enzyme, mixed well and incubated at 37 deg for at least one hour. 1 microliter of the reaction was used to transform DH5α competent cells. 60 microliters of cells on separate plates were plated.

6.16.3 A complete list of the primers used in the section 3 in this thesis

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

External primers contained restriction sites: the two forward primers introduce a *KpnI* cut site, and the reverse primers introduces an *XbaI* cut site.

The ALPP wild type (wt) minigene was constructed by amplifying the genomic human DNA from the exon 3 up to the exon 5 using two oligos ALPP FOR WT and ALPP WT REV; while the ALPL wt minigene enclosed the exons 4, 5 and 6 using the following primers: ALPP WT FOR and ALPP WT REV (see primer list). These wt minigenes were then used as backbone for the mutant hybrid minigenes.

The oligonucleotides used for all the constructs are summarized:

External oligos:

ALPL FOR WT

5'CGGGGTACCCCATGGGTGTCTCCACAGTGACGGCTGCCCG3'

ALPL REV WT

5'TGCTCTAGAGCACTAGTCAATGTCCCTGATGTTATGCATGAGCTGG3'

ALPP WT FOR

5'GGGGTACCCCATGGGGGTGTCTACGGTGAC3'

ALPP WT REV

5'TGCTCTAGAGCACTAGTCAATGTCCATGTTGGA3'

Internal oligos:

ALPP MUT FOR

5'GTGCCGGCACCGCCACGGCCTACCTGTGCG3'

ALPP MUT REV

5'GCGGTGCCGGCACTGTCTGGCACATG3'

ALPP MUT2 FOR

5'TGCAACACCACCCAGGGGAACGAGGTCATCTCCGTGATG3'

ALPP MUT2 REV

5'GGTGGTGTTGCACCGGGAAAAGCGGGCGGTGCACTCAA3'

ALPP ΔG FOR

5'TTGAGTGCACCGCCCGCTTTAACCAG3'

ALPP ΔG REV

5'GGGCCGGTGCACTCAAGCCAA3'

ALPP 1°ESE G>A FOR

5'CATGTGCCAGACAGTgccGCCACAGCCACGGCC3'

ALPP 1°ESE G>A REV

5'GGCCGTGGCTGTGGCggcACTGTCTGGCACATG3'

ALPP 1° ESE A>G FOR

5'GTGCCAGACAGTGGAGGCACAGCCACGGCCTAC3'

ALPP 1° ESE A>G REV

5'GTAGGCCGTGGCTGTGCCTCCACTGTCTGGCAC3'

ALPP FOR N>S

5'TGCAGCCGCCCGCTTTTCCCAGTGCAACACGAC3'

ALPP REV N>S

5'GTCGTGTTGCACTGGGAAAAGCGGGCGGCTGCA3'

ALPP FOR Q>R

5'GCCGCCCGCTTTAACCGGTGCAACACGACACGC3'

ALPP REV Q>R

5'GCGTGTCGTGTTGCACCGGTTAAAGCGGGCGGC3'

ALPP FOR R>Q

5'CCAGTGCAACACGACACAGGGCAACGAGGTCATC3'

ALPP REV R>Q

5'GATGACCTCGTTGCCCTGTGTCGTGTTGCACTGG3'

ALPL FOR 1ESE

CAGGTCCCTGACAGTGgaGcCACaGCCACCGCCTACCT

ALPL REV 1ESE

AGGTAGGCGGTGGCtGTGgCtcCACTGTCAGGGACCTG

ALPL FOR 2ESE

CAGCCACTGAGCGTaaCcaGtgcAACACgACaCgcGGcAACGAGGTCACC

ALPL REV 2ESE

GGTGACCTCGTTCCCgcGGGTGGTGTTgcaCtgGttACGCTCAGTGGCTG

ALPL FOR 3'SS MUT

5'TGTCCCCAGACATACAATGTAGACAAACAGGTCCCTGACAGTGCCGGCAC3'

ALPL REV 3'SS MUT

5'CTGGGGACAGAGACACTCTGACAGGAGATGGCCAGGCCTTC3'

ALPP FOR 3'SS MUT

5'CCCACCTGCAGACGTACAACACCAATGCCCATGTGCCAGACAGTGCCGGCAC

3'

ALPP REV 3'SS

5'CTGCAGGTGGGAGAGGAGGGTGCTCTTCTCTGGGGCAGACAC3'

ALPP 1° ESE G>A/A>G FOR

5'CATGTGCCAGACAGTGCCGGCACAGCCA3'

ALPP 1° ESE G>A/A>G REV

5'TGGCTGTGCCGGCACTGTCTGGCACATG3'

Table 6.1 Oligonucleotides List. List of the primers used for PCR reactions.

The primers and the templates used to yield each mutated minigene are schematically represented in the following table (Table 6.1):

Name of the hybrid	Primers	Template
MUT1	ALPP MUT FOR ALPP MUT REV	ALPP WT
MUT2	ALPP MUT2 FOR ALPP MUT2 REV	ALPP WT
MUT3	ALPP MUT2 FOR ALPP MUT2 REV	MUT2
MUT4	ALPL FOR 1ESE ALPL REV 1ESE	ALPL WT
MUT5	ALPL FOR 2ESE ALPL REV 2ESE	ALPL WT
MUT6	ALPL FOR 2ESE ALPL REV 2ESE	MUT4
MUT7	ALPP ΔG FOR ALPP ΔG REV	ALPP WT
MUT8	ALPP ΔG FOR ALPP ΔG REV	MUT1
MUT9	ALPP ΔG FOR	MUT2

	ALPP ΔG REV	
MUT10	ALPP ΔG FOR ALPP ΔG REV	MUT3
MUT11	ALPP 1°ESE G>A FOR ALPP 1°ESE G>A REV	ALPP WT
MUT12	ALPP 1° ESE A>G FOR ALPP 1° ESE A>G REV	ALPP WT
MUT13	ALPP 1° ESE G>A/A>G FOR ALPP 1° ESE G>A/A>G REV	ALPP WT
MUT14	ALPP FOR N>S ALPP REV N>S	ALPP WT
MUT15	ALPP FOR Q>R ALPP REV Q>R	ALPP WT
MUT16	ALPP FOR R>Q ALPP REV R>Q	ALPP WT
MUT17	ALPP FOR R>Q ALPP REV R>Q	MUT13
MUT18	ALPP FOR 3'SS ALPP REV 3'SS	ALPP WT
MUT19	ALPP FOR 3'SS ALPP REV 3'SS	MUT1
MUT20	ALPP FOR 3'SS ALPP REV 3'SS	MUT2
MUT21	ALPP FOR 3'SS ALPP REV 3'SS	MUT3
MUT22	ALPP FOR 3'SS	MUT13

	ALPP REV 3'SS	
MUT23	ALPP FOR 3'SS ALPP REV 3'SS	MUT16
MUT23	ALPP FOR 3'SS ALPP REV 3'SS	MUT16
MUT24	ALPP FOR 3'SS ALPP REV 3'SS	MUT17
MUT25	ALPL FOR 3'SS MUT ALPL REV 3'SS MUT	ALPL WT

Table 6.2. Primers and templates used for the creation of each mutated minigene

6.17 Maintenance and analysis of cells in culture

HeLa and COS-1cell lines was grown in Dulbecco's Mem with Glutamax I (Gibco) (Dulbecco's modified Eagle's medium with glutamine, sodium pyruvate, pyridoxine and 4.5 g/l glucose) supplemented with 10% fetal calf serum (Euro Clone) and Antibiotic Antimycotic (Sigma) according to the manufacturer's instructions. Plates containing a confluent monolayer of cells were treated with 0.1% w/v trypsin as follows. Cells washed with PBS solution, were incubated at 37°C with 1-2 ml of PBS/EDTA/trypsin solution (PBS containing 0.04% w/v EDTA and 0.1% w/v trypsin) for 2 minutes or until cells were dislodged. After adding 10 ml of media, cells were pelleted by centrifugation and resuspended in 5 ml pre-warmed medium. 1-2 ml of this cell suspension was added to 10 ml medium in a fresh plate and was gently mixed before incubation

6.18 Transfection of minigene plasmids

The DNA used for transfections was prepared with JetStar purification kit (Genomed) as previously described. Liposome-mediated transfections of 3x10(5) HeLa cells were performed using Effectene reagent (Qiagen). 0.5 μg of construct DNA was mixed with 4 μl of Enhancer for each transfection and the mixture was incubated at room temperature for 5 minutes to allow the condensation of the DNA. Then, 5 μl of Effectene were added to the mixture and an incubation of 10 minutes has been performed. After the addition of 500 μl of complete culture medium the mixture was added to the cells in 3 ml of the same medium and incubated at 37°C. After 6 h the medium was replaced with fresh medium and 12 h later, the cells were harvested. RNA isolation followed as described.

6.19 mRNA analysis by Polymerase Chain Reaction

6.19.1 cDNA synthesis

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

6.19.2 cDNA analysis

PCR analysis of cDNA was carried out for 35 cycles (94 °C 30 sec, 58 °C 30 sec, 72 °C 1 min) in 50 µl reaction volumes using the following oligonucleotides which recognize specific regions of the pcDNA3 expression vector. Finally the PCR products representing the splicing products were separated on 2% agarose gels and ethidium bromide-stained. Quantification of band intensities in the acquired pictures was performed using the ImageJ64 software (available at http://rsb.info.nih.gov/ij).

Oligonucleotide name pcDNA3 vector	**Sequence of the oligonucleotide (5'-3')**
T7	AATACGACTCACTATAG
SP6	ATTTAGGTGACACTATAGAATA

6.20 Construction of the expression plasmids.

The PLAP and TNAP cDNA vectors were provided by JL Millan (Sanford-Burnham Medical Research Institute, USA). To facilitate isolation of the recombinant enzymes, a FLAG epitope was introduced after the Leu489 and Thr483 in TNAP and PLAP, respectively, followed by a termination codon to eliminate the glycosylphosphatidylinositol anchoring signal. The PLAP-FLAG mutants were generated by Quick change PCR amplification using PfuI polymerase followed by DpnI digestion (Promega). Constructs were sequenced to verify the correct introduction of mutations into the sequence.

Name of the expression plasmids	Primers	Template
PLAP-FLAG 1st ESE [A93G;G94A]	ALPP 1° ESE G>A/A>G FOR ALPP 1° ESE G>A/A>G REV	PLAP-FLAG WT
PLAP-FLAG 2nd ESE [R125Q]	ALPP FOR R>Q ALPP REV R>Q	PLAP-FLAG WT
PLAP-FLAG 1st/2nd ESE [A93G;G94A; R125Q]	ALPP FOR R>Q ALPP REV R>Q	PLAP-FLAG 1st ESE [A93G;G94A]

6.20.1 Transfection of expression plasmids

The expression constructs were transfected into COS-1 cells for transient expression by the Effectene transfection reagent (Quiagen). Cells were transfected with 5 μg of FLAG-tagged expression consctructs. Transfected cells were cultured in OPTI-MEM serum free medium, and conditioned media (25ml/150 mm plate), containing secreted wt and mutant enzymes, were collected 48 hours after transfection.

6.21 Purification of FLAG-tagged enzymes

Each secreted FLAG-tagged mutant enzyme was purified using an anti-FLAG M2 monoclonal antibody affinity column (Sigma, St. Louis, MO, USA).

The anti-flag M2 monoclonal antibodies agarose was first packed into chromatographic column. The anti-flag column was equilibrated with TBS for at least 10 bed-volumes. The conditioned media (25ml/150 mm plate), containing secreted enzyme was then loaded onto the column at the flow rate of 1 ml/min. Contaminating proteins were washed out with 10–20 column volumes TBS buffer. Then, the protein fused with the C-terminal Flag tag was eluted with 1ml of elution buffer (100 mM Glycine, pH 3.5). Each 1 ml of eluted fractions was collected in a 1.5 ml tube containing 20 µl Tris-HCl (1 M, pH 8.8) for neutralization. The column was then extensively washed with 10–20 column volumes TBS buffer and stored in storing buffer (TBS buffer, 50% glycerol, pH 7.5) at 4°C for the long-term storage.

6.22 Denaturing polyacrylamide gel electrophoresis (SDS-PAGE)

Proteins from each sample were mixed with 10X protein loading buffer (20% w/v SDS, 1M DTT, 0.63M Tris-HCl pH 7, 0.2% w/v bromophenol blue, 20% v/v glycerol, 10mM EDTA pH 7). Conventional slab gel SDS PAGE (Laemmli, 1970) was performed in vertical gels with the required percentage of polyacrylamide (37,5:1 acrylamide:bis-acrylamide, ProtoGel, National Diagnostics), depending on each case. The gels were run at 40 mA in 1X SDS-PAGE running buffer (50 mM Tris, 0.38 M glycine, 0.1 % w/v SDS). After running, gels were either stained with coomassie Blue R250 in methanol-water-acetic acid 45:45:10 (v/v/v) or Western blot analysis was performed.

6.23 Western blots and antibodies

SDS-PAGE gels were blotted on standard nitrocellulose membrane and incubated in blocking solution (PBS1X, 0.2% Tween, 2% milk) over night. Primary antibodies were incubated with the membrane in the same buffer for 2 hours. Afterwards, the membrane was then washed three times in washing solution (PBS1X, 2%Tween), incubated with the proper HRP-secondary antibodies (anti-mouse/rabbit, Dako) for 1 hour, washed twice and stained with ECL reagent (Thermo Scientific). Finally, an autoradiography was taken on Kodak Biomax XAR films.

The primary antibody used in this study: Anti-flag epitope: (1:2000) (Sigma F1804)

6.24 Micro Bicinchoninic Acid (BCA) Protein Assay

Micro Bicinchoninic Acid (BCA) Protein Assay (Pierce) for determining the protein concentration of dilute samples (0.5 - 20μg/mL).

The Micro BCA Protein Assay is a highly sensitive colorimetric assay primarily relies on two reactions. Firstly, the peptide bonds in the protein sample reduce Cu^{2+} ions, in a temperature dependent reaction, from the copper solution to Cu^{+}. The amount of Cu^{2+} reduced is proportional to the amount of protein present in the solution. Next, two molecules of BCA chelate with each Cu^{+} ion form a purple-colored product that strongly absorbs light at a wavelength of 562 nm. The amount of protein present in a solution can be quantified by measuring the absorption spectra and the standard curve is obtained by plotting the absorbance (Abs) versus known concentration of standard protein.

6.25 Immune Enzymatic Assay PLAP/TNAP

The principle of the assay was based on the measurement of relative specific enzymatic activities using microtiter plates coated with 0,2 µg/ml M2 anti-Flag antibody in TBS, $MgCl_2$ 1mM and ZnAc 20 µM. Each well was blocked with 200 µl 1% BSA. The plate was washed five times with washing buffer (working buffer + 0,002 % Tween 20). Saturating concentration of sample was added to each well for 3 hours at RT. After washing, increasing concentration of *p*-nitrophenylphosphate (pNPP, Sigma) (0.01 - 20 mM final concn.) in 1.0 M diethanolamine (DEA) buffer, pH 9.8, containing 20 µM ZnCl2 and 1 mM MgCl2 was added as substrate and the activity was measured directly the at 405 nm during the time. For the calculation of catalytic rate constants, the wt PLAP-FLAG construct with a known k_{cat} is used as a reference for each microtiter plate.

6.26 Slot blot protein determination

The enzyme concentration of each purified sample was determined by the Bio-Dot SF Apparatus (Bio-Rad) using Amino-terminal FLAG-Bacterial Alkaline Phosphatase (BAP) Fusion protein (Sigma) as a standard (Zhu et al., 2005). 96-well plates (Nunc MaxiSorp®) were blocked with 400 µl/well blocking buffer (3% skim milk in PBS) at room temperature for 30 min. The plates were washed five times with PBS/ 0.01% Tween 20 (PBST, Sigma). Eight two-fold serial dilutions of FLAG-BAP standard protein and purified enzymes were prepared with PBS in the blocked plates. The FLAG-BAP standard dilutions ranged from 3.5 to 225 ng. The proteins were applied to the slot blot apparatus and drown by vacuum through a pre–wetted nitrocellulose membrane (0.45 µM, Amersham Biosciences). The membrane slot was washed by drawing two times 300 µl of PBS through each well, and was incubated for 1 hour at room temperature with blocking solution (5% skim milk in PBS). The membrane was than incubated 2h at room temperature or, alternatively overnight

at 4°C 1:1000 dilution of M2 anti-FLAG antibody (Sigma). The membrane was washed three times with PBST, and was incubated for 1 hour at room temperature with 1:2000 dilution with the secondary anti-mouse antibody conjugated with horseradish peroxidase for 1 hour (Dako Cytomation). The membrane was then washed again three times with PBST and developed using the Pierce Chemi-Luminescen (ECL) western blotting substrate. The membrane was scanned and the intensity of each band on the slot was measured with ImageJ64 Software. Thus each peak in the graph represented a slot containing detected protein. Standard curves were obtained (using excel) by plotting peak area versus known concentration of FLAG-BAP and fitting the data points with linear regression equation. The peak areas of the serially diluted purified protein samples that fell within the detection range of each respective standard curve were used to calculate the protein concentration. Generally the mean value of 4 peaks from each protein dilution series is used per assay. The standard curve was included in every slot.

6.27 ALP assay and Kinetic measurements

ALP kinetic determinations were performed as described previously (Hoylaerts et al., 1992) in presence of increasing concentration of *p*-nitrophenylphosphate (pNPP, Sigma) (0.01 - 20 mM final concn.) as substrate in 1.0 M diethanolamine (DEA) buffer, pH 9.8, containing 20 µM ZnCl2 and 1 mM MgCl2. After the addition of recombinant ALPs (8 ng), the activity was measured as the change in absorbance at 405 nm over time at 37 °C. The kinetic analysis was done from those parts of the curve where the absorbance versus time was linear (15 min). To determine the initial rate, the ΔAbs_{405}/min was calculated taking into the account an extinction coefficient for reaction product p-nitrophenol of 12.2×10^3 M^{-1} cm^{-1}. The ΔAbs_{405}/min corresponds to an increase in the concentration of the product, *p*-nitrophenol, per minute (Δc/ minute) and an equal decrease in the concentration of the substrate. The corresponding Δc/minute can be determined by using the extinction

coefficient absorption coefficient (ε) for *p*-nitrophenol using Beer's law (A = ε x l x c, where l = path length (usually 1 cm), c = concentration of *p*-nitrophenol, and ε = extinction coefficient, in this case 12.2 $\times 10^3$ M^{-1} cm^{-1}). This Δc/min is also known as v_o, the initial velocity. Eleven different substrate concentration were used to measure the K_m and V_{max} and the resulting v_o were fit by nonlinear regression to the Michaelis-Menten equation using GraphPad Prism version 3.02 (GraphPad Software, San Diego, CA). The K_{cat} was calculated dividing the V_{max} for the concentration of enzyme used for each reaction.

6.28 Statistical analysis

The Km and Vmax values were calculated from 6 independent experiments. The statistical software R version 2.13.1 was used for statistical analysis. All values were checked for normal distribution by the Shapiro-Wilk normality test. Data from WT and mutants enzymes showed normal distribution and were expressed as mean values (± SD). Differences between mean values were analyzed using the Student's t-test.

The asterisks indicate the range of the different P values as shown in the scheme below:

P-value	Symbol
$P<0.05$	*
$P<0.01$	**
$P<0.001$	***

7. REFERENCES

Ars E., Kruyer H., Gaona A., Serra E., Lazaro C., Estivill X. (1999) Prenatal diagnosis of sporadic neurofibromatosis type 1 (NF1) by RNA and DNA analysis of a splicing mutation. Prenat Diagn 19:739-42.

Baralle D., Baralle M. (2005) Splicing in action: assessing disease causing sequence changes. J Med Genet 42:737-48.

Baralle D., Lucassen A., Buratti E. (2009) Missed threads. The impact of pre-mRNA splicing defects on clinical practice. EMBO Rep 10:810-6.

Baralle M., Skoko N., Knezevich A., De Conti L., Motti D., Bhuvanagiri M., Baralle D., Buratti E., Baralle F.E. (2006) NF1 mRNA biogenesis: effect of the genomic milieu in splicing regulation of the NF1 exon 37 region. FEBS Lett 580:4449-56.

Barnard D.C., Li J., Peng R., Patton J.G. (2002) Regulation of alternative splicing by SRrp86 through coactivation and repression of specific SR proteins. RNA 8:526-33.

Berget S.M. (1995) Exon recognition in vertebrate splicing. J Biol Chem 270:2411-4.

Beyenbach K.W., Wieczorek H. (2006) The V-type H+ ATPase: molecular structure and function, physiological roles and regulation. J Exp Biol 209:577-89.

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

Blencowe B.J. (2000) Exonic splicing enhancers: mechanism of action, diversity and role in human genetic diseases. Trends Biochem Sci 25:106-10.

Brow D.A. (2002) Allosteric cascade of spliceosome activation. Annu Rev Genet 36:333-60.

Brunak S., Engelbrecht J., Knudsen S. (1991) Prediction of human mRNA donor and acceptor sites from the DNA sequence. J Mol Biol 220:49-65.

Buratti E., Baralle F.E. (2005) Another step forward for SELEXive splicing. Trends Mol Med 11:5-9.

Buratti E., Baralle M., Baralle F.E. (2006) Defective splicing, disease and therapy: searching for master checkpoints in exon definition. Nucleic Acids Res 34:3494-510.

Buratti E., Stuani C., De Prato G., Baralle F.E. (2007) SR protein-mediated inhibition of CFTR exon 9 inclusion: molecular characterization of the intronic splicing silencer. Nucleic Acids Res 35:4359-68.

Buvoli M., Mayer S.A., Patton J.G. (1997) Functional crosstalk between exon enhancers, polypyrimidine tracts and branchpoint sequences. Embo J 16:7174-83.

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

Calarco J.A., Zhen M., Blencowe B.J. (2011) Networking in a global world: establishing functional connections between neural splicing regulators and their target transcripts. RNA 17:775-91.

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

Cartegni L., Chew S.L., Krainer A.R. (2002) Listening to silence and understanding nonsense: exonic mutations that affect splicing. Nat Rev Genet 3:285-98.

Cartegni L., Wang J., Zhu Z., Zhang M.Q., Krainer A.R. (2003) ESEfinder: A web resource to identify exonic splicing enhancers. Nucleic Acids Res 31:3568-71.

Chamary J.V., Hurst L.D. (2005) Biased codon usage near intron-exon junctions: selection on splicing enhancers, splice-site recognition or something else? Trends Genet 21:256-9.

Coulter L.R., Landree M.A., Cooper T.A. (1997) Identification of a new class of exonic splicing enhancers by in vivo selection. Mol Cell Biol 17:2143-50.

de Almeida S.F., Carmo-Fonseca M. (2008) The CTD role in cotranscriptional RNA processing and surveillance. FEBS Lett 582:1971-6.

de Almeida S.F., Carmo-Fonseca M. (2012) Design principles of interconnections between chromatin and pre-mRNA splicing. Trends Biochem Sci.

Del Gatto F., Plet A., Gesnel M.C., Fort C., Breathnach R. (1997) Multiple interdependent sequence elements control splicing of a fibroblast growth factor receptor 2 alternative exon. Mol Cell Biol 17:5106-16.

Deutsch M., Long M. (1999) Intron-exon structures of eukaryotic model organisms. Nucleic Acids Res 27:3219-28.

Di Mauro S., Manes T., Hessle L., Kozlenkov A., Pizauro J.M., Hoylaerts M.F., Millan J.L. (2002) Kinetic characterization of hypophosphatasia mutations with physiological substrates. J Bone Miner Res 17:1383-91.

Dickerson R.E. (1971) The structures of cytochrome c and the rates of molecular evolution. J Mol Evol 1:26-45.

Dietrich R.C., Incorvaia R., Padgett R.A. (1997) Terminal intron dinucleotide sequences do not distinguish between U2- and U12-dependent introns. Mol Cell 1:151-60.

Divina P., Kvitkovicova A., Buratti E., Vorechovsky I. (2009) Ab initio prediction of mutation-induced cryptic splice-site activation and exon skipping. Eur J Hum Genet 17:759-65.

Doellgast G.J., Benirschke K. (1979) Placental alkaline phosphatase in Hominidae. Nature 280:601-2.

Eskesen S.T., Eskesen F.N., Ruvinsky A. (2004) Natural selection affects frequencies of AG and GT dinucleotides at the 5' and 3' ends of exons. Genetics 167:543-50.

Fairbrother W.G., Chasin L.A. (2000) Human genomic sequences that inhibit splicing. Mol Cell Biol 20:6816-25.

Fairbrother W.G., Yeh R.F., Sharp P.A., Burge C.B. (2002) Predictive identification of exonic splicing enhancers in human genes. Science 297:1007-13.

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

Fairbrother W.G., Yeo G.W., Yeh R., Goldstein P., Mawson M., Sharp P.A., Burge C.B. (2004b) RESCUE-ESE identifies candidate exonic splicing enhancers in vertebrate exons. Nucleic Acids Res 32:W187-90.

Faustino N.A., Cooper T.A. (2003) Pre-mRNA splicing and human disease. Genes Dev 17:419-37.

Ferris S.D., Whitt G.S. (1979) Evolution of the differential regulation of duplicate genes after polyploidization. J Mol Evol 12:267-317.

Fishman W.H. (1990) Alkaline phosphatase isozymes: recent progress. Clin Biochem 23:99-104.

Flicek P., Aken B.L., Ballester B., Beal K., Bragin E., Brent S., Chen Y., Clapham P., Coates G., Fairley S., Fitzgerald S., Fernandez-Banet J., Gordon L., Graf S., Haider S., Hammond M., Howe K., Jenkinson A., Johnson N., Kahari A., Keefe D., Keenan S., Kinsella R., Kokocinski F., Koscielny G., Kulesha E., Lawson D., Longden I., Massingham T., McLaren W., Megy K., Overduin B., Pritchard B., Rios D., Ruffier M., Schuster M., Slater G., Smedley D., Spudich G., Tang Y.A., Trevanion S., Vilella A., Vogel J., White S., Wilder S.P., Zadissa A., Birney E., Cunningham F., Dunham I., Durbin R., Fernandez-Suarez X.M., Herrero J., Hubbard T.J., Parker A., Proctor G., Smith J., Searle S.M. (2010) Ensembl's 10th year. Nucleic Acids Res 38:D557-62.

Fong Y.W., Zhou Q. (2001) Stimulatory effect of splicing factors on transcriptional elongation. Nature 414:929-33.
Force A., Lynch M., Pickett F.B., Amores A., Yan Y.L., Postlethwait J. (1999) Preservation of duplicate genes by complementary, degenerative mutations. Genetics 151:1531-45.
Fox-Walsh K.L., Dou Y., Lam B.J., Hung S.P., Baldi P.F., Hertel K.J. (2005) The architecture of pre-mRNAs affects mechanisms of splice-site pairing. Proc Natl Acad Sci U S A 102:16176-81.
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.
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.
Gilbert W. (1978) Why genes in pieces? Nature 271:501.
Goldman N., Yang Z. (1994) A codon-based model of nucleotide substitution for protein-coding DNA sequences. Mol Biol Evol 11:725-36.
Graveley B.R. (2000) Sorting out the complexity of SR protein functions. Rna 6:1197-211.
Graveley B.R., Hertel K.J., Maniatis T. (2001) The role of U2AF35 and U2AF65 in enhancer-dependent splicing. RNA 7:806-18.
Greene P.J., Sussman H.H. (1973) Structual comparison of ectopic and normal placental alkaline phosphatase. Proc Natl Acad Sci U S A 70:2936-40.
Hall S.L., Padgett R.A. (1994) Conserved sequences in a class of rare eukaryotic nuclear introns with non-consensus splice sites. J Mol Biol 239:357-65.
Haque A., Buratti E., Baralle F.E. (2010) Functional properties and evolutionary splicing constraints on a composite exonic regulatory element of splicing in CFTR exon 12. Nucleic Acids Res 38:647-59.
Harris H. (1980) Multilocus enzyme systems and the evolution of gene expression: the alkaline phosphatases as a model example. Harvey Lect 76:95-124.
Harris H. (1990) The human alkaline phosphatases: what we know and what we don't know. Clin Chim Acta 186:133-50.
Hartmuth K., Urlaub H., Vornlocher H.P., Will C.L., Gentzel M., Wilm M., Luhrmann R. (2002) Protein composition of human prespliceosomes isolated by a tobramycin affinity-selection method. Proc Natl Acad Sci U S A 99:16719-24.
He X., Zhang J. (2005) Gene complexity and gene duplicability. Curr Biol 15:1016-21.
Hellmann I., Zollner S., Enard W., Ebersberger I., Nickel B., Paabo S. (2003) Selection on human genes as revealed by comparisons to chimpanzee cDNA. Genome Res 13:831-7.
Henthorn P.S., Raducha M., Kadesch T., Weiss M.J., Harris H. (1988) Sequence and characterization of the human intestinal alkaline phosphatase gene. J Biol Chem 263:12011-9.
Hertel K.J. (2008) Combinatorial control of exon recognition. J Biol Chem 283:1211-5.
Hertel K.J., Maniatis T. (1999) Serine-arginine (SR)-rich splicing factors have an exon-independent function in pre-mRNA splicing. Proc Natl Acad Sci U S A 96:2651-5.
Hertel K.J., Graveley B.R. (2005) RS domains contact the pre-mRNA throughout spliceosome assembly. Trends Biochem Sci 30:115-8.
Hirsh A.E., Fraser H.B. (2001) Protein dispensability and rate of evolution. Nature 411:1046-9.
Horowitz D.S., Krainer A.R. (1994) Mechanisms for selecting 5' splice sites in mammalian pre-mRNA splicing. Trends Genet 10:100-6.
House A.E., Lynch K.W. (2006) An exonic splicing silencer represses spliceosome assembly after ATP-dependent exon recognition. Nat Struct Mol Biol 13:937-44.

Hoylaerts M.F., Manes T., Millan J.L. (1992) Molecular mechanism of uncompetitive inhibition of human placental and germ-cell alkaline phosphatase. Biochem J 286 (Pt 1):23-30.
Hughes A.L. (1994) The evolution of functionally novel proteins after gene duplication. Proc Biol Sci 256:119-24.
Hughes A.L. (1997) Rapid evolution of immunoglobulin superfamily C2 domains expressed in immune system cells. Mol Biol Evol 14:1-5.
Huh G.S., Hynes R.O. (1993) Elements regulating an alternatively spliced exon of the rat fibronectin gene. Mol Cell Biol 13:5301-14.
Hui J., Bindereif A. (2005) Alternative pre-mRNA splicing in the human system: unexpected role of repetitive sequences as regulatory elements. Biol Chem 386:1265-71.
Hunter S., Apweiler R., Attwood T.K., Bairoch A., Bateman A., Binns D., Bork P., Das U., Daugherty L., Duquenne L., Finn R.D., Gough J., Haft D., Hulo N., Kahn D., Kelly E., Laugraud A., Letunic I., Lonsdale D., Lopez R., Madera M., Maslen J., McAnulla C., McDowall J., Mistry J., Mitchell A., Mulder N., Natale D., Orengo C., Quinn A.F., Selengut J.D., Sigrist C.J., Thimma M., Thomas P.D., Valentin F., Wilson D., Wu C.H., Yeats C. (2009) InterPro: the integrative protein signature database. Nucleic Acids Res 37:D211-5.
Hurst L.D. (2002) The Ka/Ks ratio: diagnosing the form of sequence evolution. Trends Genet 18:486.
Hurst L.D., Pal C. (2001) Evidence for purifying selection acting on silent sites in BRCA1. Trends Genet 17:62-5.
Ibrahim el C., Schaal T.D., Hertel K.J., Reed R., Maniatis T. (2005) Serine/arginine-rich protein-dependent suppression of exon skipping by exonic splicing enhancers. Proc Natl Acad Sci U S A 102:5002-7.
Irimia M., Rukov J.L., Roy S.W. (2009) Evolution of alternative splicing regulation: changes in predicted exonic splicing regulators are not associated with changes in alternative splicing levels in primates. PLoS One 4:e5800.
Jackson I.J. (1991) A reappraisal of non-consensus mRNA splice sites. Nucleic Acids Res 19:3795-8.
Jurica M.S., Licklider L.J., Gygi S.R., Grigorieff N., Moore M.J. (2002) Purification and characterization of native spliceosomes suitable for three-dimensional structural analysis. RNA 8:426-39.
Kanopka A., Muhlemann O., Akusjarvi G. (1996) Inhibition by SR proteins of splicing of a regulated adenovirus pre-mRNA. Nature 381:535-8. DOI: 10.1038/381535a0.
Kashima T., Manley J.L. (2003) A negative element in SMN2 exon 7 inhibits splicing in spinal muscular atrophy. Nat Genet 34:460-3.
Ke S., Zhang X.H., Chasin L.A. (2008) Positive selection acting on splicing motifs reflects compensatory evolution. Genome Res 18:533-43.
Kiledjian M., Kadesch T. (1990) Analysis of the human liver/bone/kidney alkaline phosphatase promoter in vivo and in vitro. Nucleic Acids Res 18:957-61.
Kim E.E., Wyckoff H.W. (1991) Reaction mechanism of alkaline phosphatase based on crystal structures. Two-metal ion catalysis. J Mol Biol 218:449-64.
Kimura M. (1983) Rare variant alleles in the light of the neutral theory. Mol Biol Evol 1:84-93.
Knoll B.J., Rothblum K.N., Longley M. (1988) Nucleotide sequence of the human placental alkaline phosphatase gene. Evolution of the 5' flanking region by deletion/substitution. J Biol Chem 263:12020-7.
Konarska M.M., Vilardell J., Query C.C. (2006) Repositioning of the reaction intermediate within the catalytic center of the spliceosome. Mol Cell 21:543-53.
Kondrashov F.A., Rogozin I.B., Wolf Y.I., Koonin E.V. (2002) Selection in the evolution of gene duplications. Genome Biol 3:RESEARCH0008.

Kornblihtt A.R. (2005) Promoter usage and alternative splicing. Curr Opin Cell Biol 17:262-8.
Kornblihtt A.R., de la Mata M., Fededa J.P., Munoz M.J., Nogues G. (2004) Multiple links between transcription and splicing. RNA 10:1489-98.
Kuma K., Iwabe N., Miyata T. (1995) Functional constraints against variations on molecules from the tissue level: slowly evolving brain-specific genes demonstrated by protein kinase and immunoglobulin supergene families. Mol Biol Evol 12:123-30.
Laemmli U.K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 227:680-5.
Lam B.J., Hertel K.J. (2002) A general role for splicing enhancers in exon definition. RNA 8:1233-41.
Lamond A.I. (1993) The spliceosome. Bioessays 15:595-603.
Lang K.M., Spritz R.A. (1983) RNA splice site selection: evidence for a 5' leads to 3' scanning model. Science 220:1351-5.
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.
Le Du M.H., Millan J.L. (2002) Structural evidence of functional divergence in human alkaline phosphatases. J Biol Chem 277:49808-14.
Le Du M.H., Stigbrand T., Taussig M.J., Menez A., Stura E.A. (2001) Crystal structure of alkaline phosphatase from human placenta at 1.8 A resolution. Implication for a substrate specificity. J Biol Chem 276:9158-65.
Lechward K., Awotunde O.S., Swiatek W., Muszynska G. (2001) Protein phosphatase 2A: variety of forms and diversity of functions. Acta Biochim Pol 48:921-33.
Lehmann F.G. (1980) Human alkaline phosphatases. Evidence of three isoenzymes (placental, intestinal and liver-bone-kidney-type) by lectin-binding affinity and immunological specificity. Biochim Biophys Acta 616:41-59.
Levy J.B., Dorai T., Wang L.H., Brugge J.S. (1987) The structurally distinct form of pp60c-src detected in neuronal cells is encoded by a unique c-src mRNA. Mol Cell Biol 7:4142-5.
Lewis J.D., Izaurralde E. (1997) The role of the cap structure in RNA processing and nuclear export. Eur J Biochem 247:461-9.
Licatalosi D.D., Darnell R.B. (2006) Splicing regulation in neurologic disease. Neuron 52:93-101.
Lin C.W., Fishman W.H. (1972) L-Homoarginine. An organ-specific, uncompetitive inhibitor of human liver and bone alkaline phosphohydrolases. J Biol Chem 247:3082-7.
Lynch M., Conery J.S. (2000) The evolutionary fate and consequences of duplicate genes. Science 290:1151-5.
Lynch M., Force A. (2000) The probability of duplicate gene preservation by subfunctionalization. Genetics 154:459-73.
Majewski J., Ott J. (2002) Distribution and characterization of regulatory elements in the human genome. Genome Res 12:1827-36.
Manley J.L., Tacke R. (1996) SR proteins and splicing control. Genes Dev 10:1569-79.
Mardon H.J., Sebastio G., Baralle F.E. (1987) A role for exon sequences in alternative splicing of the human fibronectin gene. Nucleic Acids Res 15:7725-33.
Martinez-Contreras R., Fisette J.F., Nasim F.U., Madden R., Cordeau M., Chabot B. (2006) Intronic binding sites for hnRNP A/B and hnRNP F/H proteins stimulate pre-mRNA splicing. PLoS Biol 4:e21.
Matlin A.J., Clark F., Smith C.W. (2005) Understanding alternative splicing: towards a cellular code. Nat Rev Mol Cell Biol 6:386-98.

Mayeda A., Screaton G.R., Chandler S.D., Fu X.D., Krainer A.R. (1999) Substrate specificities of SR proteins in constitutive splicing are determined by their RNA recognition motifs and composite pre-mRNA exonic elements. Mol Cell Biol 19:1853-63.

McClintock J.M., Kheirbek M.A., Prince V.E. (2002) Knockdown of duplicated zebrafish hoxb1 genes reveals distinct roles in hindbrain patterning and a novel mechanism of duplicate gene retention. Development 129:2339-54.

McComb R.B., Bowers G.N., Jr. (1972) Study of optimum buffer conditions for measuring alkaline phosphatase activity in human serum. Clin Chem 18:97-104.

McComb R.B., Bowers G.N., Posen S. (1979) Alkaline phosphatase Plenum Press, New York.

McCullough A.J., Berget S.M. (1997) G triplets located throughout a class of small vertebrate introns enforce intron borders and regulate splice site selection. Mol Cell Biol 17:4562-71.

Millan J.L. (1987) Promoter structure of the human intestinal alkaline phosphatase gene. Nucleic Acids Res 15:10599.

Millan J.L. (2006) Alkaline Phosphatases : Structure, substrate specificity and functional relatedness to other members of a large superfamily of enzymes. Purinergic Signal 2:335-41.

Millan J.L., Manes T. (1988) Seminoma-derived Nagao isozyme is encoded by a germ-cell alkaline phosphatase gene. Proc Natl Acad Sci U S A 85:3024-8.

Modafferi E.F., Black D.L. (1997) A complex intronic splicing enhancer from the c-src pre-mRNA activates inclusion of a heterologous exon. Mol Cell Biol 17:6537-45.

Modrek B., Lee C.J. (2003) Alternative splicing in the human, mouse and rat genomes is associated with an increased frequency of exon creation and/or loss. Nat Genet 34:177-80.

Moore M.J., Sharp P.A. (1993) Evidence for two active sites in the spliceosome provided by stereochemistry of pre-mRNA splicing. Nature 365:364-8.

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.

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.

Ohno S. (1970) Evolution by gene duplication Springer-Verlag, Berlin, New York,.

Orphanides G., Reinberg D. (2002) A unified theory of gene expression. Cell 108:439-51.
Pagani F., Baralle F.E. (2004) Genomic variants in exons and introns: identifying the splicing spoilers. Nat Rev Genet 5:389-96.

Pagani F., Raponi M., Baralle F.E. (2005) Synonymous mutations in CFTR exon 12 affect splicing and are not neutral in evolution. Proc Natl Acad Sci U S A 102:6368-72.

Pagani F., Buratti E., Stuani C., Baralle F.E. (2003a) Missense, nonsense, and neutral mutations define juxtaposed regulatory elements of splicing in cystic fibrosis transmembrane regulator exon 9. J Biol Chem 278:26580-8.

Pagani F., Stuani C., Zuccato E., Kornblihtt A.R., Baralle F.E. (2003b) Promoter architecture modulates CFTR exon 9 skipping. J Biol Chem 278:1511-7.

Pagani F., Stuani C., Tzetis M., Kanavakis E., Efthymiadou A., Doudounakis S., Casals T., Baralle F.E. (2003c) New type of disease causing mutations: the example of the composite exonic regulatory elements of splicing in CFTR exon 12. Hum Mol Genet 12:1111-20.

Pagani F., Buratti E., Stuani C., Romano M., Zuccato E., Niksic M., Giglio L., Faraguna D., Baralle F.E. (2000) Splicing factors induce cystic fibrosis transmembrane regulator exon 9 skipping through a nonevolutionary conserved intronic element. J Biol Chem 275:21041-7.

Pal C., Papp B., Lercher M.J. (2006) An integrated view of protein evolution. Nat Rev Genet 7:337-48.

Pan Q., Shai O., Misquitta C., Zhang W., Saltzman A.L., Mohammad N., Babak T., Siu H., Hughes T.R., Morris Q.D., Frey B.J., Blencowe B.J. (2004) Revealing global regulatory features of mammalian alternative splicing using a quantitative microarray platform. Mol Cell 16:929-41.

Parmley J.L., Chamary J.V., Hurst L.D. (2006) Evidence for purifying selection against synonymous mutations in mammalian exonic splicing enhancers. Mol Biol Evol 23:301-9.

Parmley J.L., Urrutia A.O., Potrzebowski L., Kaessmann H., Hurst L.D. (2007) Splicing and the evolution of proteins in mammals. PLoS Biol 5:e14.

Pertea M., Lin X., Salzberg S.L. (2001) GeneSplicer: a new computational method for splice site prediction. Nucleic Acids Res 29:1185-90.

Piatigorsky J., Wistow G. (1991) The recruitment of crystallins: new functions precede gene duplication. Science 252:1078-9.

Proudfoot N.J., Furger A., Dye M.J. (2002) Integrating mRNA processing with transcription. Cell 108:501-12.

Query C.C., Moore M.J., Sharp P.A. (1994) Branch nucleophile selection in pre-mRNA splicing: evidence for the bulged duplex model. Genes Dev 8:587-97.

Query C.C., Strobel S.A., Sharp P.A. (1995) The branch site adenosine is recognized differently for the two steps of pre-mRNA splicing. Nucleic Acids Symp Ser:224-5.

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

Raponi M., Baralle F.E., Pagani F. (2007) Reduced splicing efficiency induced by synonymous substitutions may generate a substrate for natural selection of new splicing isoforms: the case of CFTR exon 12. Nucleic Acids Res 35:606-13.

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

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

Reed R., Maniatis T. (1988) The role of the mammalian branchpoint sequence in pre-mRNA splicing. Genes Dev 2:1268-76.

Reese M.G., Eeckman F.H., Kulp D., Haussler D. (1997) Improved splice site detection in Genie. J Comput Biol 4:311-23.

Resch A., Xing Y., Alekseyenko A., Modrek B., Lee C. (2004) Evidence for a subpopulation of conserved alternative splicing events under selection pressure for protein reading frame preservation. Nucleic Acids Res 32:1261-9.

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.

Roca X., Sachidanandam R., Krainer A.R. (2005) Determinants of the inherent strength of human 5' splice sites. RNA 11:683-98.

Rump A., Kasper G., Hayes C., Wen G., Starke H., Liehr T., Lehmann R., Lagemann D., Rosenthal A. (2001) Complex arrangement of genes within a 220-kb region of double-duplicated DNA on human 2q37.1. Genomics 73:50-5.

Sakharkar M.K., Perumal B.S., Sakharkar K.R., Kangueane P. (2005) An analysis on gene architecture in human and mouse genomes. In Silico Biol 5:347-65.

Sanford J.R., Ellis J., Caceres J.F. (2005) Multiple roles of arginine/serine-rich splicing factors in RNA processing. Biochem Soc Trans 33:443-6.

Say J.C., Ciuffi K., Furriel R.P., Ciancaglini P., Leone F.A. (1991) Alkaline phosphatase from rat osseous plates: purification and biochemical characterization of a soluble form. Biochim Biophys Acta 1074:256-62.

Schaal T.D., Maniatis T. (1999) Multiple distinct splicing enhancers in the protein-coding sequences of a constitutively spliced pre-mRNA. Mol Cell Biol 19:261-73.

Senapathy P., Shapiro M.B., Harris N.L. (1990) Splice junctions, branch point sites, and exons: sequence statistics, identification, and applications to genome project. Methods Enzymol 183:252-78.

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.

Sharma S., Kohlstaedt L.A., Damianov A., Rio D.C., Black D.L. (2008) Polypyrimidine tract binding protein controls the transition from exon definition to an intron defined spliceosome. Nat Struct Mol Biol 15:183-91.

Sharp P.A. (1994) Split genes and RNA splicing. Cell 77:805-15.

Smith P.J., Zhang C., Wang J., Chew S.L., Zhang M.Q., Krainer A.R. (2006) An increased specificity score matrix for the prediction of SF2/ASF-specific exonic splicing enhancers. Hum Mol Genet 15:2490-508.

Staley J.P., Guthrie C. (1998) Mechanical devices of the spliceosome: motors, clocks, springs, and things. Cell 92:315-26.

Stec B., Holtz K.M., Kantrowitz E.R. (2000) A revised mechanism for the alkaline phosphatase reaction involving three metal ions. J Mol Biol 299:1303-11.

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

Sun H., Chasin L.A. (2000) Multiple splicing defects in an intronic false exon. Mol Cell Biol 20:6414-25.

Tacke R., Manley J.L. (1999) Determinants of SR protein specificity. Curr Opin Cell Biol 11:358-62.

Teraoka S.N., Telatar M., Becker-Catania S., Liang T., Onengut S., Tolun A., Chessa L., Sanal O., Bernatowska E., Gatti R.A., Concannon P. (1999) Splicing defects in the ataxia-telangiectasia gene, ATM: underlying mutations and consequences. Am J Hum Genet 64:1617-31.

Tian M., Maniatis T. (1994) A splicing enhancer exhibits both constitutive and regulated activities. Genes Dev 8:1703-12.

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.

Wang E.T., Sandberg R., Luo S., Khrebtukova I., Zhang L., Mayr C., Kingsmore S.F., Schroth G.P., Burge C.B. (2008) Alternative isoform regulation in human tissue transcriptomes. Nature 456:470-6.

Wang Z., Burge C.B. (2008) Splicing regulation: from a parts list of regulatory elements to an integrated splicing code. RNA 14:802-13.

Wang Z., Rolish M.E., Yeo G., Tung V., Mawson M., Burge C.B. (2004) Systematic identification and analysis of exonic splicing silencers. Cell 119:831-45.

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

Webb C.J., Romfo C.M., van Heeckeren W.J., Wise J.A. (2005) Exonic splicing enhancers in fission yeast: functional conservation demonstrates an early evolutionary origin. Genes Dev 19:242-54.

Weiss M.J., Ray K., Henthorn P.S., Lamb B., Kadesch T., Harris H. (1988) Structure of the human liver/bone/kidney alkaline phosphatase gene. J Biol Chem 263:12002-10.

Will C.L., Luhrmann R. (2011) Spliceosome structure and function. Cold Spring Harb Perspect Biol 3.

Willie E., Majewski J. (2004) Evidence for codon bias selection at the pre-mRNA level in eukaryotes. Trends Genet 20:534-8.

Wilson A.C., Carlson S.S., White T.J. (1977) Biochemical evolution. Annu Rev Biochem 46:573-639.

Wolfe K.H., Sharp P.M., Li W.H. (1989) Mutation rates differ among regions of the mammalian genome. Nature 337:283-5.
Xiao X., Wang Z., Jang M., Burge C.B. (2007) Coevolutionary networks of splicing cis-regulatory elements. Proc Natl Acad Sci U S A 104:18583-8.
Xing Y., Lee C. (2005a) Evidence of functional selection pressure for alternative splicing events that accelerate evolution of protein subsequences. Proc Natl Acad Sci U S A 102:13526-31.
Xing Y., Lee C. (2005b) Assessing the application of Ka/Ks ratio test to alternatively spliced exons. Bioinformatics 21:3701-3.
Yan Y.L., Willoughby J., Liu D., Crump J.G., Wilson C., Miller C.T., Singer A., Kimmel C., Westerfield M., Postlethwait J.H. (2005) A pair of Sox: distinct and overlapping functions of zebrafish sox9 co-orthologs in craniofacial and pectoral fin development. Development 132:1069-83.
Yang Z., Nielsen R. (2000) Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. Mol Biol Evol 17:32-43.
Yeo G., Burge C.B. (2004) Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. J Comput Biol 11:377-94.
Yeo G.W., Van Nostrand E., Holste D., Poggio T., Burge C.B. (2005) Identification and analysis of alternative splicing events conserved in human and mouse. Proc Natl Acad Sci U S A 102:2850-5.
Zago P., Buratti E., Stuani C., Baralle F.E. (2011) Evolutionary connections between coding and splicing regulatory regions in the fibronectin EDA exon. J Mol Biol 411:1-15.
Zhang M.Q. (1998) Statistical features of human exons and their flanking regions. Hum Mol Genet 7:919-32.
Zhu D., Saul A.J., Miles A.P. (2005) A quantitative slot blot assay for host cell protein impurities in recombinant proteins expressed in E. coli. J Immunol Methods 306:40-50.
Zhu J., Mayeda A., Krainer A.R. (2001) Exon identity established through differential antagonism between exonic splicing silencer-bound hnRNP A1 and enhancer-bound SR proteins. Mol Cell 8:1351-61.
Zuckerkandl E. (1976) Evolutionary processes and evolutionary noise at the molecular level. II. A selectionist model for random fixations in proteins. J Mol Evol 7:269-311.
Zuckerkandl E., Pauling L. (1965) Molecules as documents of evolutionary history. J Theor Biol 8:357-66.

ACKNOWLEDGEMENT

Though only one name appears on the cover of this dissertation, many people have contributed to its production. I owe my gratitude to all those people who have made this thesis possible. First and foremost, I offer my sincerest gratitude to my Ph.D. supervisor, Prof. Francisco E. Baralle, who has supported me throughout my thesis with his patience, motivation, enthusiasm and knowledge that I will never forget. He has been a wealthy source of advice and guidance and importantly he formed a laboratory environment that was both friendly and productive.

My sincere thanks also go to my external supervisor Dr. Colin Sharpe, for his encouragement, insightful comments, and discussions.

I am also very grateful to Marco Baralle Ph.D, who has been always there to listen, encourage and give practical advice during these four years. Thanks for reading my reports, commenting on my views, helping me in enriching my ideas and especially for the corrections and comments on countless revisions of this thesis.

I am also thankful to the members of my laboratory, Molecular Pathology Group, who have provided me with invaluable assistance, helping me to hurdle all the obstacles in the completion of this research work. In particular, a special thanks to my best friends at the ICGEB for being like a family to me. My gratitude also goes to Kristian Vlahovicek, our collaboration has been invaluable to the completion of this thesis. Also thanks to Prof. Josè Luis Millan, whose collaboration has greatly contributed to this thesis and his laboratory for the hospitality during my time in San Diego and for all the efforts to arrange my visit. I would like to express my heart-felt gratitude to Massimo Piccotto for providing helpful comments on the manuscript. Last but not least, none of this would have been possible without the love and patience of my family. My parents, Francesco Falanga and Rosa Cirillo, and my sister which has been a constant source of love, concern, support and strength along all these years and to whom I dedicated this thesis.