PUTATIVE INHIBITORY ACTIONS OF SELECTED MEDICINAL PLANTS AGAINST EXONIC SPLICING ENHANCERS

ROSLINA BINTI RASHID

UNIVERSITI SAINS MALAYSIA

2020

PUTATIVE INHIBITORY ACTIONS OF SELECTED MEDICINAL PLANTS AGAINST EXONIC SPLICING ENHANCERS

by

ROSLINA BINTI RASHID

Thesis submitted in fulfilment of the requirements

for the degree of

Doctor of Philosophy

October 2020

ACKNOWLEDGEMENTS

I praise The Almighty Allah for giving me the strength and patience to complete this research. I would like to express my sincere appreciation and deepest gratitude to my supervisor Assoc. Prof. Dr. Teguh Haryo Sasongko. His guidance, advice, encouragement and endless support have helped me in completing this study. I would like to extend my deepest appreciation to my co-supervisors, Assoc. Prof. Dr.Muzaimi Mustapha, Prof. Dr. Zabidi Azhar Mohd Husin and Prof. Dr. Habibah Abdul Wahab for their never-ending ideas, motivations and advice in completing this thesis. I would like to acknowledge Research University Grant (RUI) for funding this research through Grant Scheme (1001/PPSP/813062) and MyBrain15 for providing a Postgraduate Scholarship to support this study. Thank you to all members of the Human Genome Centre, who have been very supportive in providing suggestions and advice. To all my friends Sathiya, Maziras, Fazreen, Ninie, Rani, Siti, Shima, Sha, Marin, Wati, Sarina and Shida for their friendship and encouragement. Lastly, I extend my special thanks to my family and my husband for their love, support and encouragement were given.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

result was also aligned with reference sequence using ClustalW multiple alignments

LIST OF ABBREVIATIONS

ANOVA	: analysis of variance
AONs	: antisense oligonucleotides
APOE	: apoliprotein E
APP	: amyloid precursor protein
Art	: artificial
ASMD	: alternative splicing mutation database
ASSP	: alternative splice site predictor
ATM	: AT mutated
bp	: base pair
Buffer AE	: elution buffer
Buffer BL	: lysis buffer
Buffer BW and TW	: wash buffer
Buffer EB	: elution buffer
Buffer PB	: purification buffer
Buffer PE	: wash buffer
cDNA	: complimentary deoxyribonucleic acid
Clk-4	: cyclic-dependent like kinase 4
cm^2	: centimetre square
CMD	: congenital muscular dystrophy
DAGs	: dystrophin-associated glycoproteins

DGE	: dystrophin-glycoprotein complex
DGGE	: denaturing gradient gel electrophoresis
dHPLC	: denaturing high performance chromatography
dH_2O	: distilled water
DMD	: Duchenne muscular dystrophy
DNA	: deoxyribonucleic acid
dNTPs	: dinucleotide triphosphates
DPC	: dystrophin-associated complex
dsDNA	: double strand deoxyribonucleic acid
dTTP	: deoxythymine triphosphate
DYRKs	: dual specificity tyrosine phosphorylation regulated kinases
E3	: exon-3 skipping
EGCg	: epigallocatechin gallate
eRF	: eukaryotic release factor
ESE	: exonic splicing enhancer
ESEfinder	: exonic splicing enhancer finder program
EST	: expressed sequence tag
FISH	: fluorescent in-situ hybridization
FM-CSCE	: fluorescent multiplex conformation sensitive capillary electrophoresis
FDA	: food and drug administration
g	: gram
Gen	: genuine

GTP : green tea polyphenol

GTE : green tea extract

GSK : GlaxoSmithKline

HATs : histone acetyltransferases

HD : Huntington's disease

HEK293 : human embryonic kidney 293

HIV : human immunodeficiency virus

hnRNP A1 : heterogenous nuclear ribonucleoprotein A1

HR-HPVs : high risk-human papillomaviruses

HR-MCA : high resolution melting curve analysis

HSF : human splicing finder

ISE : intronic splicing enhancer

ISS : intronic splicing silencer

kbp : kilo base pair

LGMD type I : limb-girdle muscular dystrophy type I

MAPH : multiplex amplifiable probe hybridization

MED : maximum entropy distribution

mg/ml : miligram per mililitre

$MgCI_2$: magnesium chloride

MIC_{50} : minimal inhibitory concentration 50

Ml : millilitre

mHH : mutated Huntington

mPCR	: multiplex polymerase chain reaction
mRNA	: messenger ribonucleic acid
NADI	: natural product discovery
NF1	: neurofibromatosis type 1
nNOS	: neuronal nitric oxide synthase
OE-PCR	: overlapping polymerase chain reaction
PCR	: polymerase chain reaction
PTB	: polypyrimidine tract binding protein
PKC	: protein kinase C
PWM	: position weight matric
PMO	: phosphorodiamidate morpholino oligomers
PS	: phosphorothioate
RESCUE	: relative enhancer and silencer classification method by unanimous enrichment
REST	: repressor element-1 silencing transcription factor exon-3 skipping
RNA	: ribonucleic acid
ROS	: reactive oxygen species
rpm	: revolutions per minute
RT-PCR	: reverse transcription PCR
rRNAs	: ribosomal RNAs
SDS-PAGE	: sodium dodecyl sulphate polyacrylamide gel electrophoresis
SE	: standard error
SR	: serine/arginine-rich

SP	: splicing potential
SMA	: spinal muscular atrophy
SMN2	: survival of motor neuron 2
SNP	: single nucleotide polymorphism
snRNPs	: small nuclear ribonucleo proteins
SPS	: san-lin-pai-tsu-san
SRPK1	: SR protein kinase 1
ssDNA	: single strand deoxyribonucleic acid
SSCP	: single strand conformation polymorphism analysis
TBE	: tris-borate-EDTA
TF	: tissue-factor
TR	: translational read-through
tRNAs	: transfer RNAs
Taq	: Thermophilus aquaticus
µg	: microgram
IUPAC	: International Union of Pure and Applied Chemistry
Ml	: microliter
µm	: micrometer
UMD	: universal mutation database
UTR	: untranslated region
UV	: ultra-violet

LIST OF APPENDICES

TINDAKAN RENCATAN TERHADAP PENGGALAK PENYINGKIRAN EKSON MENGGUNAKAN TUMBUHAN HERBA TERPILIH

ABSTRAK

Kajian sebelum ini berjaya membuktikan penglibatan '*isodiospyrin'* sebagai perencat kepada faktor penyingkiran dan penggunaan molekul sebatian kecil sebagai pencetus kepada proses penyingkiran ekson. Dengan cara merencat SR protin menggunakan '*isodiospyrin*' dan homolognya, mampu untuk menyingkirkan exon dan secara tidak lansung mengembalikan jujukan protin. Pembinaan model novel minigen boleh digunakan bagi mengkaji kerumitan mekanisma penyingkiran ekson, yang mana berpotensi kepada identifikasi target terapi pada semua penyakit yang berkaitan seperti 'Duchenne Muscular Dystrophy' dimana 'dystrophin' gen terlibat. Objektif kajian ini adalah untuk mengenalpasti tindakan rencatan terhadap penggalak penyingkiran ekson menggunaka '*isodisospyrin'* dan homolognya daripada tumbuhan herba terpilih bagi menyingkirkan ekson di dalam minigen yang direka. Faktor pencetus pada proses penyingkiran ekson didalam dystrophin minigen di kenal pasti menggunakan perisian 'ESEfinder 3.0'. Dua jenis minigen yang direka bentuk iaitu minigen asli dan juga minigene buatan. Minigen asli termasuk 'Gen-Ex45', 'Gen-Ex 51' dan 'Gen-Ex53' manakala minigen buatan pula mempunyai 'ESE' yang spesifik iaitu 'Art-SF2/ASF', 'Art-SC35', 'Art-SRp40', 'Art-SRp5'5 dan 'Art-NO ESE'. Semua minigen yang direka di klon kan sebelum disaring menggunakan proses penjujukan dan di masukkan ke dalam sel 'HEK-293' bagi proses penyingkiran tersebut. Proses tersebut kemudian disahkan menggunakan 2 kaedah iaitu pengujian signal 'luciferase' dan juga kehadiran band sasaran menggunakan kaedah 'RT-PCR', akhirnya akan disaring melalui analisa

penjujukan ‘DNA’. Di samping itu juga, 6 jenis ekstrak dari 5 jenis spesis sebatian semulajadi yang menyerupai ‘*isodiospyrin’* di kenal pasti menggunakan perisian ‘NADI’. Berdasarkan keputusan proses penjujukan secara terus, tiada penyingkiran ekson di temui dalam semua minigen asli. Manakala berbeza dengan minigen buatan yang mana menunjukkan penyingkiran di dalam semua minigen melalui keputusan ‘RT-PCR’. Selepas analisa luciferase dilakukan, nilai penyingkiran tersebut masih jauh daripada nilai minigen rujukan yang mana secara tidak lansung menunjukkan kepekaan kaedah aktiviti luciferase berbanding ‘RT-PCR’. Lima jenis sebatian secara signifikan menunjukkan penyingkiran ekson berlaku selepas didedahkan terhadap minigen ‘Art-SRp55’, satu sebatian masing-masing terhadap ‘Art-SC35’ dan ‘Art-SRp40’. Manakala, tiada sebatian yang menunjukkan keputusan yang signifikan terhadap penyingkiran ekson didalam Art-SF2/ASF. Walaubagaimanapun, keputusan yang signifikan tersebut juga tidak menunjukkan nilai penyingkiran yang sepenuhnya seperti yang berlaku didalam minigen rujukan yang bertindak sebagai petunjuk piawaian penyingkiran. Menariknya, ‘*isodiospyrin’* dikenal pasti berpotensi sebagai sebatian penyingkir ‘ESE’ terhadap ‘Art-SRp40’ kerana ianya menunjukkan nilai signifikan yang sudah menghampiri kepada sempadan iaitu (p=0.049) walaupun dengan kehadiran motif perencat dan juga memiliki nilai ‘SP’ yang tinggi. Kesimpulannya, berdasarkan keputusan ‘luciferase’, Kesimpulannya, *isodiospyrin* dan sebatian yang menyerupainya mempunyai kebolehan sebagai pengaruh kepada penyingkiran ekson terhadap minigen buatan dengan atau tanpa kehadiran motif perencat dan juga ‘hnRNP A1’. Pendekatan ini sedikit sebanyak memberi gambaran kepada kajian yang berkaitan tentang penglibatan ‘ESE’ di dalam penyakit terutamanya yang berkaitan dengannya.

PUTATIVE INHIBITORY ACTIONS OF SELECTED MEDICINAL PLANTS AGAINST EXONIC SPLICING ENHANCERS

ABSTRACT

A previous study had demonstrated the successful use of *isodiospyrin* as an inhibitor of splicing factor and the use of a small-molecule compound as exon skipping inducer. Inhibition of serine and arginine-rich (SR) protein using *isodiospyrin* and their homolog results in exon skipping and indirectly restore the reading frame and protein product. Creating a novel minigene model can be used for studying the complexity of the splicing mechanism, potentially translatable into the identification of therapeutic targets in various related other conditions, such as Duchenne Muscular Dystrophy where the Dystrophin gene is affected. The objective of this study was to determine the inhibitory actions of *isodiospyrin* and *isodiospyrin* homolog of selected medicinal plant extracts for inducing skipping in the designed minigene against exonic splicing enhancers. Exonic splicing enhancers of dystrophin minigene was identified using ESEfinder 3.0 software. There are two subtypes of minigene which are genuine minigene and artificial minigene. Genuine minigene includes Gen-Ex45, Gen-Ex 51 and Gen-Ex53 while artificial minigenes with specific exonic splicing enhancers (ESE) are Art-SF2/ASF, Art-Sc35, Art-SRp40, Art-SRp55 and Art-NO ESE. All minigenes were constructed before being subjected to the cloning process and targeted minigenes were validated using sequencing before to transfection into the HEK-293 cell line for splicing assay. The assay was again validated using 2 methods which are luciferase assay by the fluorescent signal and another method by the presence of targeted size band after reverse transcriptase-polymerase chain reaction (RT-PCR) and

were then confirmed by sequencing analysis. Six extracts from 5 plants similar to *isodiospyrin* homolog were screened using NADI software. Direct sequencing further validated the absence of exon after compounds exposure to all minigene, results showed no skipping in all genuine minigene, different with artificial minigene which showed all skippings based on the RT-PCR results. After luciferase analyses, their skipping values were still far from mock minigene (standard skipping) which showed a higher threshold indicating that no skipping occurred and that luciferase assay was more sensitive than RT-PCR. Based on the result obtained, it was proven that fewer ESE sequences in the exon are unable to retain exon. Also, there was a higher potential of skipping to occur if there are few ESE in the sequence, the presence of a silencer motif as well as when that sequence consists of positive splicing potential value. Five of the compounds were shown significantly to induce skipping after exposure to the Art-SRp55 and one of each Art-SC35 and Art-SRp40, while no compounds showed significant skipping after exposure to the Art-SF2/ASF. However, it was shown that the skipping level was not as much as that which occurred in the mock minigene that acted as a skipping standard. Interestingly, *isodiospryin* showed to have a high tendency to become ESE skipper when exposed to the Art- SRp40 minigene, because it showed a significant skipping value (p= 0.049) although with the presence of silencer motif 1 and higher SP value. In a conclusion, *isodiopsyrin* and its homologs might have shown the capacity to induce skipping, although in an ESE-specific manner, even with or without the presence of silencer motif and hnRNP A1. This approach may provide a view to further study ESE on the disease-related conditions.

CHAPTER 1

INTRODUCTION

1.1 Background of the study

1.1.1 DMD and splicing events

Isodiospyrin is a natural product of *Diospyros Morrisiana*, within the genus of Diospyros, consisting of an asymmetrical 1,2-binaphthoquinone chromophore (Hou et al., 2015; Ting et al., 2003). They are commonly known as ebony or persimmon tree and mostly grow in the tropics, including South Africa, Malaysia and, India. The role of *isodiospyrin* as a potential exon skipping inducer remains largely unexplored. *isodiospyrin* is commonly used as an antibacterial agent in tuberculosis and other bacterial diseases as well as an antiproliferative agent for tumor cells (Adeniyi et al., 2000; Das Sarma et al., 2007; Karkare et al., 2013). It was also known to induce exon skipping through kinetic inhibition against topoisomerase-I-related exon splicing enhancer (ESE). The assay was done within an isolated system of β-globin splicing without the presence of DNA, hence devoid of a real cellular assay system to reflect the dynamics of cellular components (Tazi et al., 2005; Ting et al., 2003). This is important as topoisomerase I was known to also bind to DNA (Karkare et al., 2013; Kurzwernhart et al., 2012; Mosad et al., 2012). It is unclear whether the presence of DNA in a real cellular setting interferes with the ability of *isodiospyrin* in exerting its inhibitory effects on ESE. Besides, considering the variability of splicing mechanisms such as splicing potential, the strength of splice site, and presence of silencer among different genes, although *isodiospyrin* has shown an effect on the isolated splicing system of β-globin, data pertaining to its effects towards the real cellular setting of

other genes commonly implicated in splicing-prone disorders such as Dystrophin gene remain elusive.

1.1.2 Splicing and Exonic Splicing Enhancer (ESE)

Splicing is the process where introns are removed and exons are ligated together which occurs in the nucleus. It can occur simultaneously during transcription or immediately after the transcription process. Splicing is important in eukaryote to create messenger ribonucleic acid (mRNA) and further translates into protein. Series of splicing events are catalyzed by spliceosomes which is a complex of small nuclear ribonucleoproteins (snRNPs) (Wahl et al., 2009). This is called constitutive splicing, and normally produce specific protein from single gene which differ from alternative splicing.

Alternative splicing occurs when gene-coding regions work to produce alternative processed mRNA sequences and result in multiple proteins spliced from a single gene. It results in a variety or unique protein compositions (Havens et al., 2013) due to intron retaining or exon skipping or exon extension in the final mRNA transcript (Gupta et al., 2004; Kan et al., 2002; Sharp, 2005). Transcriptome analysis suggested that up to 90% of human genes undergo alternative splicing (Pan et al., 2008; Wang et al., 2008) which directly contribute to the variety in species and organ specificities during the development and evolution process (Barbosa-Morais et al., 2012; Merkin et al., 2012; Nilsen and Graveley, 2010).

Alternative splicing is regulated by trans-acting proteins (repressor and activator) that bind to the cis-acting elements which are silencer and enhancers acting on a premature

transcript (pre-mRNA). Figure 1.1 shows the differences in the mRNA processes in constitutive and alternative splicing. Serine-Arginine rich (SR) proteins bind to exon splicing enhancer (ESE) to promote exon splicing but when they bind to heterogeneous ribonucleoprotein particles (hnRNPs), they block exon splicing in a region with exonic splicing silencers (ESS) Figure 1.2.

Several studies reported that mutations which disrupted normal splicing event comprised up to one-third of all disease-causing mutations. An estimated 15% of point mutations resulting in human genetic diseases arise from RNA splicing defects (Krawczak et al., 1992). Analysis of reconstructed mRNAs that are derived from chromosome 22 indicated that ~60% of genes are represented by two or more transcripts (Lander et al., 2001; Modrek and Lee, 2002). Though likely underestimated, such aberrant splicing is thought to account for over 50% of patients with mutations within NF1 (neurofibromatosis type 1) (Ars et al., 2000) and ATM (ataxia telangiectasia mutated) (Teraoka et al., 1999), 95% of Spinal Muscular Atrophy patients (Sasongko et al., 2009) and nearly 70% of Duchenne/Becker Muscular Dystrophy (Hoffman et al., 1987). Growing reports have correlated specific point mutations in coding regions with the skipping of the exon that harbors the mutation (Cieply and Carstens, 2015; Cooper et al., 2009).

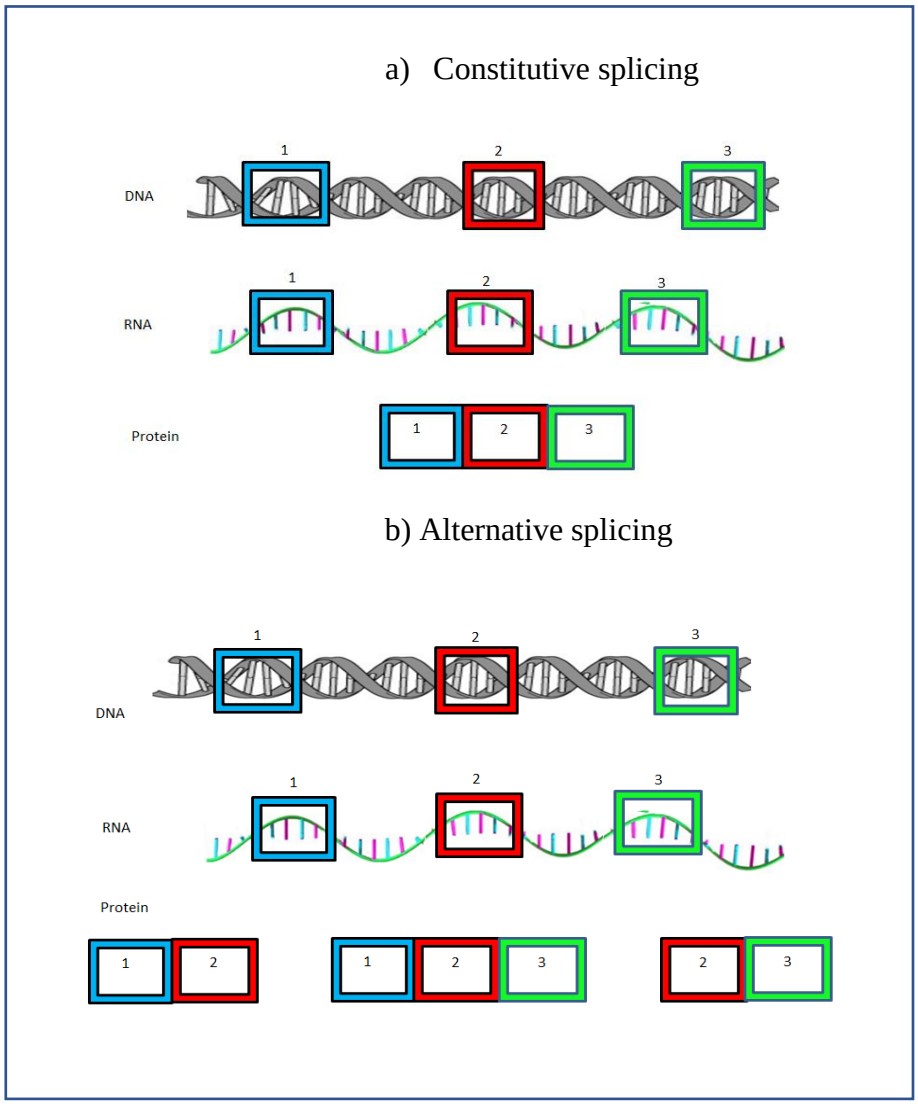

Figure 1.1: A diagram showing conceptual differences of spliced mRNA product in (a) constitutive and (b) alternative splicing when there are multiple protein isoforms are produced by alternative splicing as compared to constitutive splicing (Blencowe, 2006).

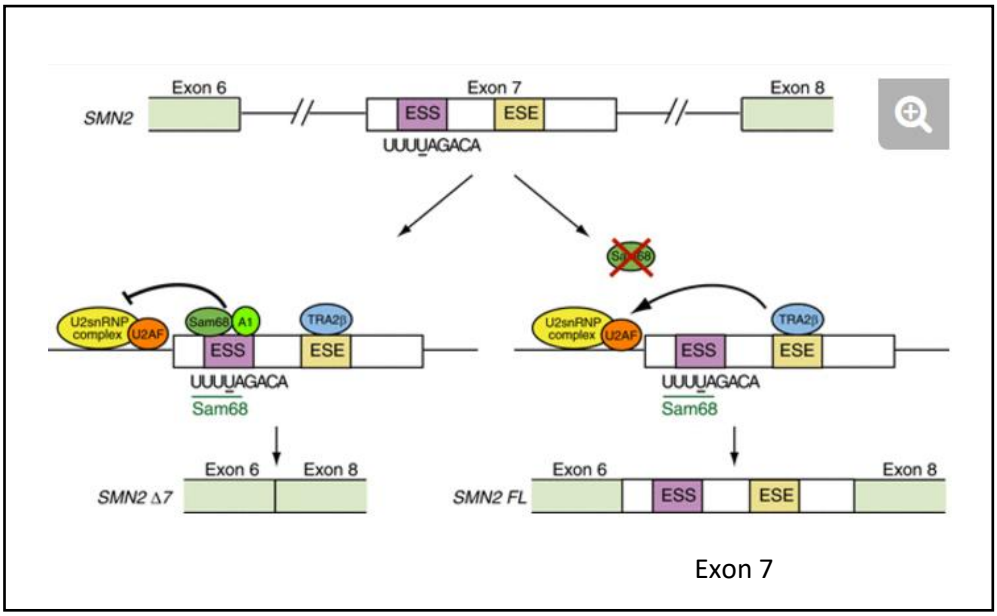

Figure 1.2: A diagram showing the splicing process in the alternative splicing of the SMN2 gene. It shows the model of Sam68 (encoded by KHDRBS1) in the recruitment of hnRNP A1 which binds to the ESS and induce exon skipping. If there Sam68 is absent, TRA2β would bind to the ESE to maintain exon-7 inclusion and result in full-length SMN2 mRNA (Pedrotti et al., 2010).

The mechanisms in which these mutations can disrupt splicing involve a direct inactivation or a splice site creation, either by activating a cryptic splice site or by interfering with splicing regulatory elements. Point mutations in the coding regions of genes had traditionally been assumed to exert their effects by altering single amino acids in the encoded proteins. However, some of these exonic mutations also affect pre-mRNA splicing. Nonsense, missense, and even translationally silent mutations can disrupt ESEs and cause the splicing machinery to skip the mutant exon, with the resulting dramatic effects on the structure of the gene product. ESEs are pre-mRNA cis-acting elements that are crucial for exon recognition during splicing (Wang et al., 2005). ESEs participate in both alternative and constitutive splicing and many of them act as binding sites for members of the serine-arginine-rich (SR) proteins family (Blencowe, 2000; Graveley, 2000). There are four SR proteins mainly focused as a target in splicing research: SF2/ASF, SRp55, SRp40 and SC35 (Cartegni et al., 2003).

1.1.3 Duchenne and Becker Muscular Dystrophy

Duchene and Becker Muscular Dystrophy (DMD) is an X-linked recessive neuromuscular disorder characterized by repeated necrosis and skeletal muscle regeneration leading to the fibrosis and muscle weakness (Emery, 1987; Koenig et al., 1987). The incidence of this Duchene and Becker is 1 in 3500-5000 newborns males (Helderman-van den Enden et al., 2013; Mendell et al., 2012; Moat et al., 2013; Muntoni et al., 2003). It is caused by a deficiency of Dystrophin due to various mutational events of the gene. Exonic deletions of the Dystrophin gene that disrupts the open reading frame, resulting in the absence of a functional Dystrophin protein at the sarcolemma of muscle fibers (Hoffman et al., 1987), account for most cases seen. This is seen due to the loss of exonic sequences of the Dystrophin gene that disrupts

the reading frame such as aberrant post-splicing products, thus total dysfunction of protein. It occurs when the deleted nucleotide is not divisible by three then the wrong amino acid will be translated into the protein. After that, it can lead to the premature stop codon and generally cause complete loss of protein function (Kashima and Manley, 2003; Liu et al., 2001; Moseley et al., 2002; Pagani et al., 2003). However. A mutation that retains some function of the gene usually causes Becker muscular dystrophy, while mutations which cause complete loss of protein cause Duchenne muscular dystrophy.

However, evidence showed that further skipping of adjacent exon(s) in the situation of exonic lost in DMD cases may restore the reading frame by partially restoring the splicing product, thus maintaining the function of the protein tend to change from severe DMD to the milder BMD (Pramono et al., 1996). In addition, a study by Rani and colleagues (2013) reported that skipping of exon 45 may reduce the phenotype severity in 24% of Malaysian DMD patients. Given its splicing pathophysiology, Dystrophin of DMD may serve as a gene model for studying the complexity of splicing mechanisms that potentially translatable into the identification of therapeutic targets in various related other conditions.

1.2 The rationale of the study

Exon skipping is known as one of the potential molecular approaches for the therapy of diseases where the splicing of exon plays a central pathology such as NF1, cystic fibrosis and DMD/BMD (Aartsma-Rus et al., 2009a; Aznarez et al., 2007; Tanganyika-de Winter et al., 2012; Zatkova et al., 2004). *Isodiospyrin* has been shown to have an exon-skipping property. A previous study demonstrated that *isodiospyrin*

induced exon skipping through kinetic inhibition against topoisomerase-I-related ESE. The assay was, however, done within an isolated system of β-globin splicing without the presence of DNA, hence devoid of a real cellular assay system to reflect the dynamics of cellular components (Ting et al., 2003; Tazi, 2005). This is important as topoisomerase I was known to also bind to DNA. It is unclear whether the presence of DNA in a real cellular setting interferes with the ability of *isodiospyrin* in exerting its inhibitory effects on ESE.

The Dystrophin gene is highly conserved; homologs have been identified not only in vertebrates (mammals, birds and fish) but also in the popular invertebrate laboratory models *Caenorhabditis elegans* and *Drosophila melanogaster*. The dystrophin gene is chosen as the study model because it is high conserved in vertebrates (birds, fish and mammals) and popular invertebrate models such as Caenorhabditis elegans and *Drosophila melanogaster* (Collins and Morgan, 2003). Then, it gives a high possibility of identifying successful treatment through the integration of studies in multiple different species to addressing particular research questions. Besides that, the Dystrophin gene with largely splicing pathology in DMD/BMD contains a unique frame rule. The splicing pathology in DMD/BMD can be corrected through exon skipping. Thus, these characteristics of the Dystrophin gene provide the researcher with a useful model for testing out compounds with potential skipping properties.

A previous study showed that skipping of exon 45 may rescue at least 24% of the DMD phenotypes (Rani et al., 2013). Exons 45, 51 and 53 of the Dystrophin gene were selected as skipping models in this study because these exons represent the major hotspot mutant region of the gene. Theoretically, up to 63% of the patients could be treated by skipping the entire exons 45-55 region (Beroud et al., 2007; Koenig et al.,

1987; Rani et al., 2013). Based on the previous studies, deletion of exons 45-55 were normally associated with mild clinical phenotypes and sometimes almost asymptomatic (Nakamura et al., 2008).

This study opens a new approach where natural compounds may be exploited for inhibiting ESE to induce exon skipping as a potential therapeutic approach in splicing-prone disorders.

1.3 Research Questions

1. Is *isodiospyrin* able to inhibit ESE in genuine minigene or artificial specific minigene or both minigene and cause exon skipping?
2. Can *isodiospyrin* homolog compound act as an ESE inhibitor similar to *isodiospyrin* and cause exon skipping?

1.4 Objectives of the research

1.4.1 General Objective

To determine of the inhibitory actions of *isodiospyrin* by inducing skipping in the designed minigene against exonic splicing enhancers.

1.4.2 Specific objectives

1. To design artificial minigene with ESE specific splicing activity (Art-ESE specific minigene).

2. To determine the inhibitory effects of *isodiospyrin* against splicing of the Art-ESE specific minigenes.

3. To characterize the exonic splicing enhancer motifs contributing to splicing of Gen-Ex45, 51 and 53 of the Dystrophin gene.

4. To determine the inhibitory effects of *isodiospyrin* against splicing of Gen-Ex45, 51 and 53 of the Dystrophin gene.

5. To validate the inhibitory effects of *isodiospyrin* using different natural substances of similar structure.

CHAPTER 2

LITERATURE REVIEW

2.1 Duchenne muscular dystrophy (DMD)

DMD is caused by a mutation in the *Dystrophin* gene which is complicated due to the larger size of the gene in the human genome with 79 exons. Mutation normally was observed at the major hot spot region which is the distal hot spot. Dystrophin is important to maintain muscle fiber (Koenig et al., 1987). Muscle loss normally affect s the pelvis and thigh followed by the arms, leads to difficulty in standing up. About 95% of the DMD can be detected by genetic testing during pregnancy (Helderman et al., 2013). Even though there is no cure for this disease, however braces, physical therapy and some corrective surgery may help in certain symptoms. Besides that, ongoing medical treatment normally aimed to control the onset of symptoms and give better lifestyle example like AON, natural substances, chemical compound and splicing modified compounds to reduce the severity of the DMD to the BMD (Jarver, P and Gait, 2014; Flagilan et al., 2011; Lunn and Wang, 2008; Aartsma-Rus et al., 2006a).

2.2 Splicing Mechanism

In molecular biology, mature messenger ribonucleic acid (mRNA) is transcribed from precursor (Pre) mRNA by removing the introns resulting in ligation of the exons. The word “Intron” is derived from the term “intragenic region” which is the intervening segment of DNA between two exons (Will and Luhrmann, 2001). Introns are found in most organisms including many viruses and can generate ribosomal RNAs (rRNAs),

proteins and transfer RNAs (tRNAs) (Roy and Gilbert, 2006). Splicing normally takes place inside the nucleus either co-transcriptionally or right after the transcription process (Tilgner et al., 2012). Splicing is important to create mRNAs that can be translated into proteins via the translation process. Most of the intron splicing processes in eukaryotes are carried out by the catalysis of spliceosomes.

2.2.1 Spliceosome

The spliceosome is a large and complex molecule consisting of small nuclear ribonucleoproteins (snRNPs) found in the nucleus of eukaryotic cells. The spliceosome removes the introns from the transcribed pre-mRNA to convert into mature mRNAs – known as splicing (Will and Luhrmann, 2011). Spliceosomal introns normally reside within the eukaryotic protein-coding genes. Three sites are important for a splicing event to take place namely - 1) donor site (5' end of intron), 2) branch site (near the 3' end of the intron) and 3) acceptor site (3' end of the intron). An invariant GU sequence is observed at the 5' end, whereas, AG sequence at the 3' end of the introns. 5' upstream of the invariant AG sequence, there is a region with high pyrimidine tract (C and U nucleotides) known as branchpoint which includes an adenine nucleotide-associated lariat formation (Black, 2003).

As regards to the consensus sequence from the UIPAC nucleic acid notations, reported G-G-[cut]-G-U-R-A-G-U (donor site) then intron sequence followed by Y-U-R-A-C (branch sequence about 20-50 nucleotides upstream from the acceptor site), then Y-rich-N-C-A-G-[cut]-G (acceptor site). It is noted that specific sequence within the intronic element and the number of nucleotides between branch point nearest to 3' acceptor site can affect the splice site selection (Figure 2.1) (Taggart et al., 2012).

Besides, point mutations also may cause an error during transcription by activating the cryptic splice sites in part of a transcript and resulting not spliced out with missing parts of an exon. Each of the spliceosomes is normally composed of five small nuclear RNAs (snRNA) and when combined with protein factor it is formed an RNA-protein complex known as snRNPs.

There are two types of spliceosomes 1) major and 2) minor which contains different snRNPs. Major spliceosomes splice those introns which consist of AT the 3' end and GU at the 5' splice site while U1, U2, U4, U5 and U6 snRNPs active in the nucleus (Qian and Liu, 2014). This type of splicing is called canonical splicing or lariat pathway which occurs more than 99% of the splicing events. However, if intronic flanking sequences do not follow the GU-AG rule, then noncanonical splicing occurs (Ng et al., 2004). At least 70% of the human genes undergo alternative splicing that can create evolution in the composition of the same mRNA either by extending, skipping or retaining intronic regions (Johnson et al., 2003; Cartegni and Krainer, 2002). Based on the study comparison between expressed sequence tag (EST) with genomic or complementary DNA (cDNA), about 35-60% of the human gene produces multiple transcripts which explained the choice of splice site (Brett et al., 2002; Lander et al., 2001; Modrek et al., 2001; Croft et al., 2000).

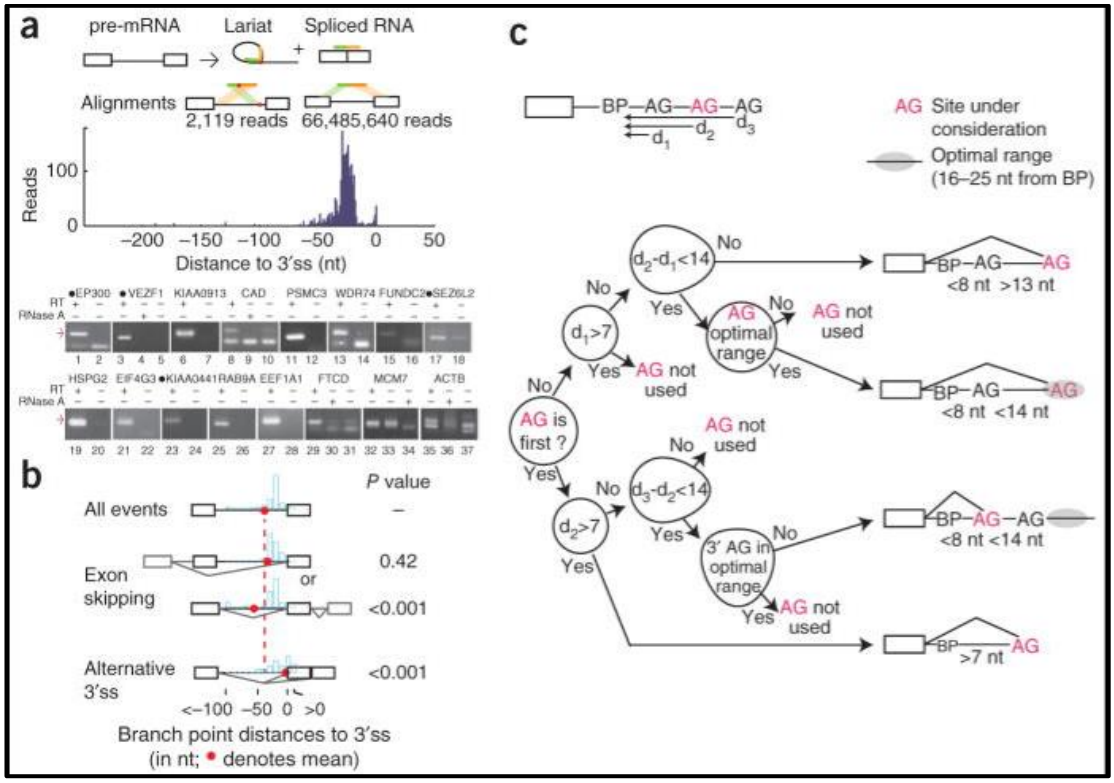

Figure 2.1: Diagram showing pre-mRNA lariats in human transcripts (Taggart et al., 2012)

2.2.2 Exonic splicing enhancer and silencer

Splicing is regulated by trans-acting protein as activators/repressors and corresponds with the cis-acting regulatory site such enhancer/silencer which present on the pre-mRNA sequence (Qian and Liu, 2014; Cartegni et al., 2002). Trans-acting components include spliceosome which is made up of 5 small nuclear ribonucleoproteins (snRNPs), heterogeneous nuclear ribonucleoproteins (hnRNP) and also regulatory complex as Figure 2.2 (Jian et al., 2014). There are several types of cis-acting element components including 5' splice site (5' ss) which is the junction between an exon and intron while 3' ss junction between intron and exon. Splicing silencer plays a role as splicing repressor protein which binds and reduces the chance of a nearby site to be used as a splice junction while splicing enhancers work oppositely. Splicing enhancers present in both intronic and exonic parts; splicing enhancers in intronic as known as intronic splicing enhancer (ISE), whereas the exon splicing enhancer (ESE) are in the exons. Moreover, for the splicing repressors, they are called as intronic splicing silencer (ISS), exonic splicing silencer (ESS). Another component is the branch point which is 10 of the nucleotide sequences located upstream of 3 ss (Jian et al., 2014).

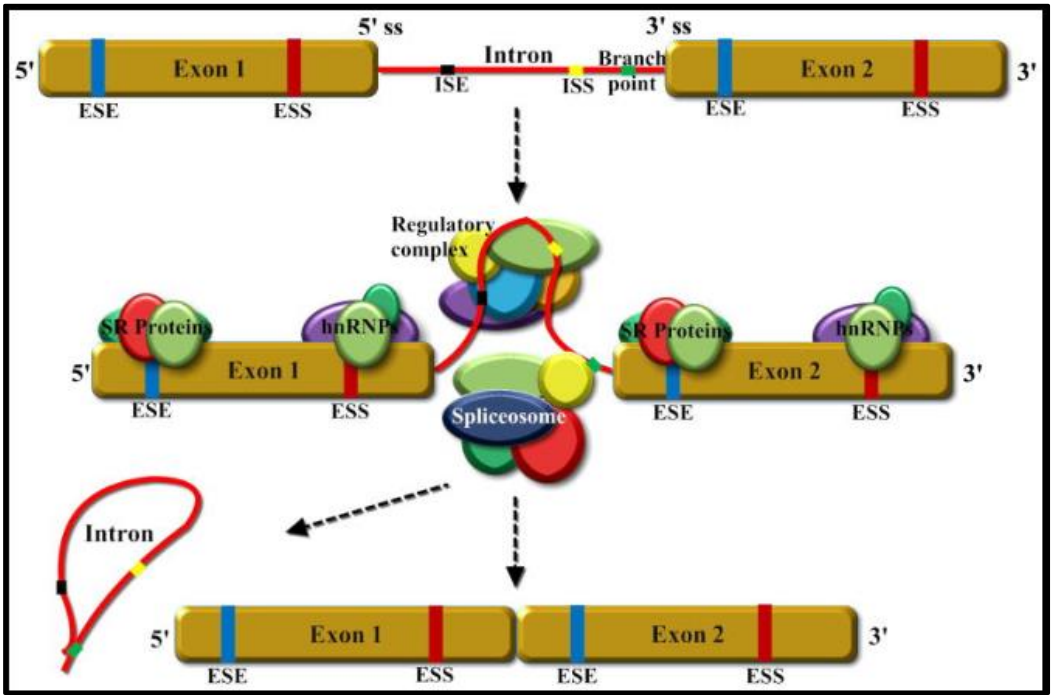

Figure 2.2: Diagram showing pre-mRNA splicing. 5′ ss and 3′ ss are recognized by the spliceosome and the intron is excised and exons are spliced. The overall process is regulated by *trans*-acting elements such as hnRNPs, SR proteins and the regulatory complex (Jian et la., 2014)**.**

2.2.3 SR protein

Splicing is also involved in the interference of intricate pathways, in initiating signals and stimuli to the cells to integrate serine and arginine-rich proteins (SR protein) (Bourgeois et al., 2004). SR protein is one of the splicing factors which play a pivotal role in the alternative splicing process to fine-tune the gene expression at the post-transcriptional level (Yin et al., 2012). SR proteins intervene very early during spliceosome assembly and stabilize to form complex A by UI snRNP particles bind to 5' GU splice donor and U2 with the assistant of the U2 snRNP auxiliary factor (U2AF) binds with branch point A within the 3' branch site (Sanford et al., 2003; Graveley, 2000). Indeed, the formation of the complex A is a key step to determine the end of the exon to be retained while ends of the intron to be spliced out (Matlin et al., 2005). Other spliceosome proteins including U4, U5 and U6 complex binds and U6 and replaces U1 position. Then, U1 and U2 cleaves and another complex initiate two transesterification reactions - a) at 5' end, the intron is cleaved from an upstream exon and bind with the branch site A via 2',5'-phosphodiester linkage and b) at 3' end, the intron is cleaved from the downstream exon. Finally, two exons are joined together through a phosphodiester bond and the intron is released as lariat form and degraded (Black, 2003).

Most of the splicing repressors are from hnRNPs including hnRNPA1 and polypyrimidine tract binding protein (PTB). Most of the activator protein binds to ISE/ESE are from SR protein members (Wang and Burge, 2008). However, predicting of ESE is difficult because of two reasons, firstly SR protein may degenerate RNA-binding specificities (Tacke et al., 1997; Tacke and Manley, 1995) and secondly, SR protein and others positive acting factor normally recognize exonic elements

cooperative. Therefore, studies of SR protein function separately might fail to predict the real function of specific exon *in-vivo* (Li et al., 2000).

ESE acts as a specific recognition of the splice sites in both constitutive and alternative splicing processes (Fairbrother et al., 2002; Liu et al., 1998; Tian and Kole, 1995). Important of ESE in splice-site recognition was first reported in an artificial 0–globin RNA study in which the selection of two competing splice sites is dependent on the adjacent exonic sequences (Reed and Maniatis, 1986). A study by Cooper and Mattox (1997) recognized the best-defined ESEs and alteration or deletion of that sequence might lead to the change of the specific splice site recognition. ESEs have been extensively studied in many of the contexts including alternative splicing (Black, 2003) as well as constitutive splicing (Cartegni et al., 2002; Mayeda et al., 1999; Schaal and Maniatis, 1999). ESEs are normally located in the splicing site vicinity (Bailey et al., 2002) and unlike transcriptional enhancers, it functions as a strong position-dependent manner because it enhances splicing when present downstream 3 ss and/or upstream 5 ss (Edelmann et al., 1999), but represses when present at the intronic regions (Lupski, 1998; Mazzarella and Schlessinger, 1998).

Thus, it is suggested that strong ESEs are crucial for constitutively spliced exons in avoiding the presence of intronic sequence near splicing sites (Fairbrother et al., 2002). There are computational methods for identifying ESE including 'Relative Enhancer and Silencer Classification by Unanimous Enrichment' (RESCUE) which can predict splicing phenotypes on the mutant exons (within ESE activity sequence) by statistical analysis based on the exon, intron and splice site composition (Fairbrother et al., 2002). Researchers have predicted 10 ESE motifs and confirmed by reporter experiment that those 10 sequences display enhancer activity in vivo and mutant sequence exhibit

reduced activity. Furthermore, Fairbrother and colleagues (2004) identified 238 ESE candidates in the human genome which is served as n ESE indicator for the ESE quality which might cause variant to biological phenotype. The selection of those ESE candidates is based on the two criteria which are found significantly in human exon rather than intron. The second criterion is significantly found in exon with weak (non-consensus) splice site than exon with a strong (consensus) splicing site. Those 238 ESE from out of 4096 possible hexamers were then clustered into 10 groups, 2 groups of ESEs that present in 5' splice site, 5 groups in 3' splice site and another 3 present in both the groups. *In vivo* test on 10 non-redundant ESE motifs were chosen of splicing enhancer confirmatory step. Results were then confirmed for that 10 cases and when predicted ESE is disrupted by selected point mutation, reduce exon inclusion by more than 2 times in 9 out of 10 cases (Fairbrother et al., 2002)

2.2.4 Post-translational modifications

Thus, implied that those variations cause exons skipping and synonymous mutation presumed with no effect to the human need to be reconfirmed, because it may also change splicing phenotypes. The key step in understanding the splicing enhancer function is ESE being recognized by SR binding protein groups. Three types of posttranslational modifications occur in SR protein including methylation, acetylation and phosphorylation respectively (Zhou and Fu, 2013; Wang and Cooper, 2007).

Many of the arginine methylations are observed in RNA binding sites especially hnRNP protein (Liu and Dreyfuss (1995). A study on arginine methylation was first reported on Np13p in the budding of yeast (Siebel and Guthrie, 1996). It is located

near the end of the protein adjacent to the phosphorylation site. Np13p plays an important role in transcription and also splicing within the nucleus (Moehle et al., 2012; Kress et al., 2008). However, stable of arginine methylation blocks phosphorylation that requires Np13p to interact with nuclear import Mtr10p receptor and results in hypermethylation of Np13p and causes defect on the phenotype of nuclear import (Yun and Fu, 2000). A study by Sinha et al. (2010) reported that three methylated arginines were identified in the mammalian SR protein as well as SRSF1 which play a positive role in promoting the nuclear import of SR protein. Defective SRSF1 enhanced translational SR protein in the cytoplasm is observed and reduced in the nucleus such as regulation of alternative splicing and coupling with nonsense-mediated mRNA decay. Choudhary and coworkers (2009) revealed extensive lysine acetylation in SR protein and SR protein-specific kinases represent an important class of non-histone substrate that is recognized by various histone acetyltransferases (HATs) within mammalian cells. HAT Tip60 was reported to modify SRSF2 and counter-regulation by deacetylase HDAC6 (Edmond et al., 2011). This finding indicates the possibility of different SR proteins and regulators may be controlled by the different combinations of HATs and HDACs resulting in a complex regulatory network in modulating splice site selection in mammalian cells.

Phosphorylation of SR proteins is catalyzed by multiple kinases that belong to the CMGC family (Kannan and Neuwald, 2004). Discovery of SR protein phosphorylation was a coincident during the discovery of the SR protein family because of antibodies against endogenous SR proteins known as phosphoepitopes in their RS domain which automatically retard migration of the SR protein in the sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) (Zahler et al., 1992;

Roth et al., 1991). Only SR protein kinases 1 (SRPK1) family in yeast known as Sky1p shown to phosphorylate several SR proteins such as Np13p and Hrb1p (Porat et al., 2006; Siebel and Guthrie, 1996). In the case of Np13p, Sky1p-mediated phosphorylation appears to regulate and facilitate protein-protein and protein-RNA interactions which is highly related to those mediated by SRPKs in mammalian cells (Gilbert et al., 2001; Yeakley et al., 1999). Interestingly, SRPK1 is ubiquitously expressed including in the nervous system while SRPK3 largely in muscle cells (Nakagawa et al., 2005; Wang et al., 1998a) indicating each of the SRPK family members has a unique function in different types of cell development. Besides SRPKs, others kinases also were shown to be able to transfer a phosphate group to the SR protein *in vitro* including protein kinase C (PKC) and cAMP-dependent protein kinase A (Colwill et al., 1996b), Akt (Blaustein et al., 2005; Patel et al., 2005), topoisomerase I (Rossi et al., 1996), dual-specificity tyrosine phosphorylation regulated kinases (DYRKs) (Aranda et al., 2011; de Graaf et al., 2004) and cyclin-dependent like kinases (Clk1–4) (Colwill et al., 1996a; Duncan et al., 1998). However, only SRPKs and Clks were shown to be directly responsible toward SR protein phosphorylation in-vivo either by genetic ablation or chemical inhibition (Zhong et al., 2009; Yomoda et al., 2008; Fukuhara et al., 2006; Hayes et al., 2006). Cytoplasmic SRPKs may responsible for initiating phosphorylation to facilitate SR protein in nuclear import, whereas Clks may fine-tune the phosphorylation in the nucleus. Therefore, these 2 families of protein kinases may work synergistically and/or complementary in regulating splicing, RNA export as well as other RNA metabolism process in mammalian cells (Zhou et al., 2013).

There are several studies reported on the manipulation of the SR protein able to improve or treat some diseases including human immunodeficiency virus (HIV)-1 by inhibition SRp55 (Tranell et al., 2011; Tranell et al., 2010) and by inhibiting the phosphorylation of SC35 may treat high risk-human papillomaviruses (HR-HPVs) (McFarlane et al., 2015), upregulation of Bim and SRp55 may treat patients with melanoma cells with selective BRAF inhibitors (Lai et al., 2012) and regulation of SRp55 in the eye development especially in eye organogenesis and size (Fic et al., 2007). Upregulation of SF2/ASF in adult and fetal cells may increase the expression of a human neurotropic virus, JC virus (JCV) and control reactivation of JCV in the brain (Sariyer and Khalili, 2011). Moreover, manipulation of SR protein also may involve in combination forms such as SRp30 with SRp55 responsible for misregulating Tau exon 2 and 10 to form neurodegenerative diseases such as Alzheimer's, Parkinson and amnesia (Wang et al., 2005b). Then, SF2/ASF and SRp55 participate in modulating the splicing of tissue-factor (TF) in the biosynthesis of human monocytic cells (Tardos et al., 2008).

Pre-mRNA is an immature single-stranded RNA that is involved in gene expression to remove or splice out the introns. When the pre-mRNA completes their process, then it is termed as mature messenger RNA or simply messenger RNA which consists only exons attached through the splicing process (McManus and Graveley, 2011). Pre-mRNA splicing is an important step for gene expression in higher eukaryotes group which is controlled by multiples splicing factors such as cis-acting elements, basal splicing machinery (spliceosome) as well as auxiliary regulatory factor element (Wang and Burge, 2008). There are different elements in recognizing alternative and constitutive splicing processes for exon usage determination.

Normally alternative splicing is regulated by cell-specific patterns and their regulation also based on the specific sets of cis-acting and trans-acting factors (McManus and Graveley, 2011). If removal of the intronic part within the same pre-mRNA molecule is referred to RNA cis-splicing, while in two different molecules it is called RNA trans-splicing. Other cis-acting elements including exonic splicing silencers, exonic splicing enhancers, intronic splicing silencers and intronic splicing enhancers also play important roles either direct or indirect for alternative RNA splicing (De Conti et al., 2013; Garcia-Blanco et al., 2004). However, RNA-trans splicing is rarely found in mammal although separate forms of trans-splicing have been reported in others species including rodents and human (Dorn and Krauss, 2003; Labrador and Corces, 2003; Finta and Zaphiropoulos, 2002; Flouriot et al., 2002; Caudevilla et al., 1998; Bruzik and Maniatis, 1992; Rajkovic et al., 1990; Murphy et al., 1986; Sutton and Boothroyd, 1986).

2.3 Alternative Splicing and Disease

The majority of the protein-coding genes are from alternatively spliced products which increase the capacity of the human genome. Human protein-coding genes have an average mean of 8.8 exons with a mean size of 145 nucleotides. The mean of introns is 3365 nucleotides and 5’ and 3’ UTR is 770 and 300 nucleotides for each and result in gene spans about 27 kilobase pairs (kbp). After the pre-mRNA process, the average mRNA exported into the cytosol is 1340 nucleotide coding sequence, 1070 nucleotide UTR and also a poly-A tail (Lander et al., 2001). Thus, shows that more than 90% of the total of pre-mRNA is removed as intron during the transcription process and only 10% of average pre-mRNA is joined as exonic sequences. Exon usage in often alternative because the cell decides whether to remove it as part of the intron or remain

as mature mRNA as an alternative exon (Ben-Dov et al., 2008). Alternative splicing normally creates protein isoforms which differ biological properties normally in the aspect of protein-protein interaction, catalytic ability as well as subcellular localization (Stamm et al., 2005). More than a quarter of alternative exons results in a premature stop codon in mRNAs and causes truncated proteins or mRNA degradation in nonsense-mediated decay. Then it is not surprising that disturbance in alternative splicing is frequently associated with human disease (Tazi et al., 2009). Various diseases have been reported to be associated with changes in alternative splicing (Love et al., 2015; Fan and Tang, 2013; Lara-Pezzi et al., 2013; Buratti et al., 2006; Faustino and Cooper, 2003). Another example is non-syndromic deafness in humans after G to T substitution in the DFNA1 gene located at position 1 in intron 25 disrupted splice donor sequence and result in premature truncated 32 amino acids of protein (Lynch et al., 1997).

Moreover, several studies illustrated by the survival of motor neuron (SMN) 2 gene models that the defect of the SMN2 splicing process can develop different diseases (Cartegni et al., 2006; Cartegni and Krainer, 2002). Spinal muscular atrophy (SMA) is an autosomal recessive neuromuscular disorder caused by a mutation in SMN 1 (Wirth et al., 2006; Lefebvre et al., 1995). SMN plays important role in mRNA biogenesis and important for spliceosomal snRNP assembly in cytoplasm especially in the recycling of snRNPs and other splicing factors (Pellizzoni et al., 1998). Interestingly, there is a similarity between SMN1 and SMN2 genes, where, only 1 nucleotide is different in the exon 7 and 8. Besides, more than 32 (kilobase) kb regions is > 99% homolog between these two genes. However, the full transcript of cDNA is produced