BIO F242
Introduction To Bioinformatics
Assignment 1

# Splicing Features

Group 10 :
Deshpande Tejas Girish : 2019B1A80947P
Gaurav Masand : 2019B1A40899P
Gosavi Aditya Sandeep : 2019B1A80963P
Sambhav Jain : 2019B1A10192P
Aman Bansal : 2019B1A41025P
Group 11 :
Manik Mehta : 2019B1A20612P
Neha Nallamalli : 2019B1A41038P
**Ravish Chand : 2019B1A31036P**
Suchismita Tripathy : 2019A7PS0554P
Akshat Jain : 2019B1A30987P
"All group members have contributed to this assignment"

7th February 2022

# Contents

# 1 Introduction

RNA splicing is the post-transcriptional process of formation of mature mRNA from precursor mRNA transcript discovered in the late 1970s. It can be done by the removal of introns (non-coding regions) thereby joining together exons (coding regions) and this process is called constitutive splicing. Another mode of splicing is alternative splicing wherein exons can be included or excluded in multiple combinations thereby creating a diverse array of mRNA transcripts from a single per mRNA, thus increasing the diversity of transcriptome. Alternate splicing occurs in more than 95% of mRNAs. Due to the complexity in mRNA splicing mechanism and its regulation, many diseases can be caused if any errors arise during the process like Familial dysautonomia, Spinal muscular atrophy, Myotonic dystrophy, etc. There are also a number of different approaches currently in development like trans-splicing,anti-sense oligonucleotides etc that hope to rectify splicing where it goes wrong, with the ultimate goal of therapeutic applications.

The aim of this review is to give an overview of the mechanisms involved in mRNA and protein splicing and also cover splicing mutations, the relation between splicing and diseased including cancer and the applications of protein splicing.

# 2 mRNA Splicing

Splicing is an important post-transcriptional modification in eukaryotic organisms to make different gene products using the same gene. We will go over the basic mechanism of gene splicing. Splicing in most organisms occurs in the nucleus itself after which the spliced mRNA is capped and tailed and the mature mRNA is formed which is ready to be translated. It is done by a complex of proteins that make up the spliceosome which helps to cut out the introns and then paste the exons together. Choices made during splicing by the machinery alters the final coding sequence and thus the protein product. Alternative selection of 3' and 5' splice sites can regulate gene expression as well as differential polyadenylation of mRNA can modify the 3' UTR or coding potential which affects the binding ability of microRNA and other regulatory RNAs to the mRNA. There are 3 main events in gene splicing:

1. Exon Skipping: Exon(s) are included or excluded from the final gene transcript, resulting in prolonged or truncated mRNA variations.

2. Intron Retention: When an intron is preserved in the final transcript, it is known as intron retention. Introns are shown to remain in 2-5% of human genes. The non-coding (junk) sections of the gene are retained by the gene splicing mechanism, resulting in a demornity in protein structure and functionality.

3. Alternative 3' Splice Site and 5' Splice Site: Joining of alternative splice sites which results in different mRNA sequences for alternative gene products which is the main step of splicing and the machinery is discussed below.

## 2.1 Chemistry of Splicing

Splicing is essentially a series of two transesterification reactions which take place after nucleophilic attack. The first step of splicing is to cut off the introns. Introns have specific sequences that are recognized by splicing machinery at their 5' end (GU) - 5'splice site and 3' end (AG) - 3' splice site as well as a characteristic internal sequence which starts with A and is followed by a sequence

of pyrimidines called the branch point. The first transesterification reaction happens between the branch point and the internal sequence by a nucleophilic attack by the lone pair of the branch point on the phosphodiester bond of the 5' splice site which results in a free -OH group on the 5' end which now becomes a 3' end having lost the splice site and they form a Lariat like structure by forming hydrogen bonds with the branch point site.


The exposed -OH group on the 5' splice site now attacks the 3' splice site and the second transesterification reaction occurs and the intron with the lariat structure is removed. After this step, ligation occurs which joins the two exons together.

## 2.2 Spliceosome Machinery

Selection of exons occurs through interaction between cis and trans acting regulatory elements of genes in which generally cis-elements are enhancing and act additively. Under varied physiological or pathological situations, positive or negative splice-site recognition is regulated through numerous methods, such as the local concentration or activity of splicing regulatory factors. However, these findings only explain a portion of how these factors work together to accurately choose a controlled splice site. The complexes which make up spliceosome machinery were studied through systematic analyses of ESTs and microarray analysis. These snRNPs and protein complexes are recruited in a sequential manner and allow the splicing to occur in a step wise process and they help in holding the introns and exons in specific positions which enables the chemical reactions to occur. Splicing occurs in two major steps: assembly of spliceosome machinery and the actual splicing of premature mRNA. Many of the steps that take place during splicing are reversible in nature. The spliceosome is a complex of 150 proteins and 5 RNAs, consisting of mainly snRNPs (small nuclear ribonucleoproteins) which are catalytic in nature. The most common splicing pathway is the Lariat pathway which accounts for 99% of splicing which has been described The spliceosome is mainly composed of U1, U2 small nuclear ribonucleoproteins (snRNPs) and the U4/U6.U5 tri-snRNP which help in identification of the splicing signals in the pre-mRNA. Specific spliceosomal complexes (E, A, B and others) and eight evolutionarily conserved DExD/H-type RNA-dependent ATPases/helicases are assembled in a stepwise fashion which results in intron excision and exon ligation.

First, the U1 snRNP binds to the 5' splice site, U2AF1 and 2 (U2 associated factor) and BBP(Branch point Binding Protein) complexes bind near the 3' splice site an then BBP is released (ATP hydrolysed). U2 snRNP replaces SF1 by using another ATP and attaches to the branch point. There is a partial loop structure formed and U4/U6 and U5 complexes bind, U5 binds the 5' splice site and U6 binds to U2 after which the first transesterification reaction occurs. U1 snRNP dissociates. Now the free -OH end attacks the 3' splice site and second transesterification is underway. The lariat along with U4/U6, U5 and U2 is separated. The U5 exon to intron transition occurs, and the U6 binds to the 5' splice site. U6/U2 catalyzes transesterification, causing the 5' end of the intron to ligate to the A on the intron and create a lariat, U5 binds exon at the 3' splice site, and the 5' site is cleaved, forming the lariat. The 3' site is cleaved and exons are ligated utilizing ATP hydrolysis, whereas 2/U5/U6 remains attached to the lariat. The spliced RNA, the lariat, and the snRNPs are all released. The lariat is destroyed and the snRNPs are recycled for the next cycle of splicing. The

exons are now ligated giving the spliced mRNA. The basic mechanism for splicing different introns is the same but they have slight variations depending on which of the 3 groups (group I, II and III) the intron belongs to.

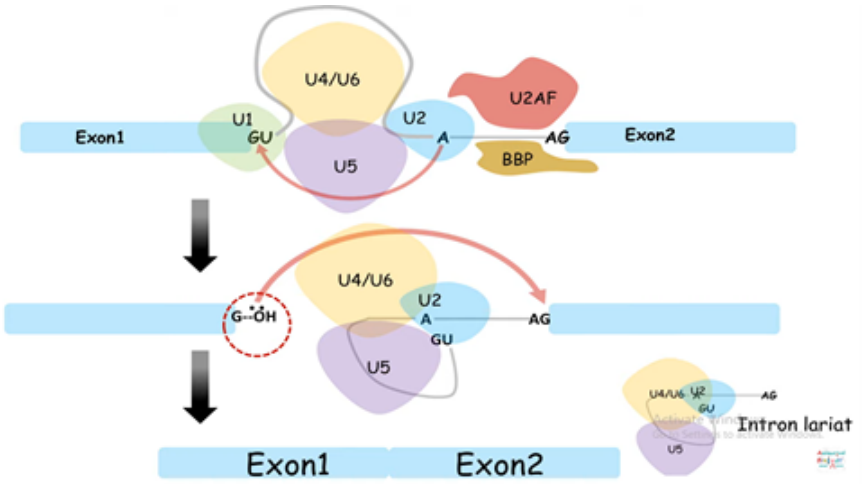


Figure 1: Spliceosome machinery
https://www.youtube.com/watch?v=IJQv1H-2IoI

## 2.3 Importance of RNA Splicing

Apart from being crucial to the formation of mature mRNA through the removal of introns in pre-mRNA, RNA splicing has other useful effects. By enabling the formation of different mRNA transcriots, splicing controls gene expression. It can cause different combinations of exons (alternative splicing) and hence different transcript variants to be formed, which finally change the protein produced through translation.

Mis-splicing or errors in the splicing process, caused by mutations or spliceosomal components, can result in non-functional protein products. These errors take place particularly when alternative splicing is involved. There are a number of diseases in humans associated with splicing errors, some of which are:

1. Myotonic Dystrophy Types 1 and 2 : Microsatellite expansions result in abnormal levels of some splicing factors which cause problems in alternative splicing
2. Spinal Muscular Atrophy : Skipping of an exon portion results in incorrect protein biproduct
3. Hutchinson-Gilford progeria syndrome : A point mutation of C -¿ T results in the altered sequence being recognised as a splice site, causing a truncated protein to be formed
4. Menkes Diseases : The mutations and their locations affect the severity of the disease and the extent to which normal splicing functions are affected

Splicing is also an important research area for the study of cancer as splicing variants can affect the proliferation of cells, cellular processes in cancer cells and their response to certain drugs and tumor suppressors also eg: CDKN2A is a gene that encodes for 2 suppressor proteins, but due to

slicing mutations, exon 2 for each protein gets skipped, resulting in incorrect protein products and an increased risk of melanoma. Splicing can also affect the radiosensitivity of cancer cells eg: it was found that chemo-radiation resistance of colon cancer cells is caused primarily by exon skipping due to splicing issues. Thus the study of splicing would be very important to understanding and treating/preventing cancer but there is still a lot left to be known in this field. Some of the details uncovered in the study of splicing and cancer are explained below.

## 2.4 Splicing and Cancer

The notion that some alternatively spliced isoforms promote the oncogenic process and might be appealing therapeutic targets underscores the relevance of exploring linkages between alternative splicing and cancer. For example, SRSF1 (also known as ASF/SF2 or ASF), has been shown to be an oncogene9 and has been connected to the regulation of AS of genes that constitute multiple cancer hallmarks. SRSF1 is a protein that binds to RNA.

Most cancer hallmarks are affected by SRSF1-driven splicing isoforms. SRSF1 is one of the effectors of EGF-mediated global effects on AS (Alternative Splicing) during cell proliferation, according to recent research. This may happen through – (i) Growth suppressor evasion—Through an intron-4 retention mechanism, SRSF1 controls production of the oncogenic cyclin D1b splicing variant, which is highly expressed in prostate malignancies. (ii) Apoptosis escape—SRSF1 is implicated in promoting AS from pro-apoptotic to antiapoptotic variants in multiple genes, including BIN1, Bim16, and caspase 9; (iii) Angiogenesis—SRSF1 is one of the key regulators of pro- and antiangiogenic vascular endothelial growth factor (VEGF) isoforms. (iv) It controls over expression of Ron in numerous cancers.

As a result, it is possible that SRSF1 is a master regulator of splicing events in a variety of cancers, making it a promising therapeutic target.

1. **Replicative Immortality:**
   The ability of cancer cells to avoid the normal cellular senescence process caused by telomere shortening is one of their most important characteristics. Although little is documented about how AS controls this process, multiple splicing isoforms of hTERT, a key component of telomerase activity, have been discovered. Various malignancies, including as breast cancer21, gastric cancer22, and lymphoma progression, have been associated with aberrant splicing isoforms.

2. **Sustaining Proliferative Signaling:**
   The epidermal growth factor receptor (EGFR) is a receptor tyrosine kinase that is activated by EGF-ligand binding and downstream signalling mediators such as Akt, JAK/STAT, or ERK. It plays a key role in cellular proliferation and mobility. The characterization of a splice variation that lacks exon 4 (de4 EGFR), which is prevalent in numerous malignancies (gliomas, prostate, and ovarian cancer) but not in neighbouring normal tissues, is a good example of cancer cells adapting to AS manipulation. The missing exon is functionally translated into a constitutively active receptor that drives proliferation.

3. **Evading Growth Suppressors:**
   Tumour suppressors include proteins - p53 and retinoblastoma. DNp53, a splice variation of p53 lacking the first 40 amino acids, may still bind DNA and so fight with wild-type p53, affecting its normal function. Because SRSF1 binds to the RPL5/MDM2 ribosomal–protein complex in the cytoplasm to keep p53 stable, phosphorylation and nuclear localization of SRSF1 is expected to result in p53 breakdown and proliferation.

4. **Apoptosis:**
   Several apoptosis-related transcripts, including Bcl-x, appear to have a similar motif in which cancer cells control AS to change translation from pro- to antiapoptotic isoforms. RBM5, an RNA-binding peptide with splicing factor capabilities that is excessively synthesized in lung and breast cancers, influences Fas receptor exon 6 splicing, resulting in either a membrane-bound pro-apoptotic Fas receptor or an antiapoptotic soluble Fas receptor.

5. **Actiavting Invasion and Metastasis:**
   Most cancer related deaths happen due to metastasis, i.e. the invasion of the tumour to other tissues. EMT (Epithelial to Mesenchymal Transition) has a role to play in this. A new epithelium-specific splice factor, epithelial splicing regulatory protein (isoforms 1 and 2) (ESRP), has been demonstrated to be a master regulator of AS events triggered during EMT. Here is a selection of the more recent and well-researched events: (i) Exon 11 skipping in E-cadherin, a cell-to-cell adhesion molecule that is dysregulated in EMT, results in a splice variant that is overexpressed in several cancers and negatively associated with expression of the wildtype variant; (ii) CD44, a transmembrane glycoprotein involved in cell transformation, has several measured exons, resulting in a myriad of splice isoforms.

6. **Avoiding Immune Detection:**
   Tumour cells are detected as very aberrant phenotypes by the body, which triggers immune responses and eliminates them. Tumour cells have developed a vast array of defence mechanisms to escape immune cell detection and destruction. The production of atypical HLA molecules on the surface of tumour cells, such as HLA-G, which suppress immunocompetent cells, is a well-known method of immune escape in tumours. MHC-I splice variant lacking exon 7; this increases cytotoxic T lymphocyte activity and has been demonstrated to be more efficient than the conventional variant in inhibiting melanoma development in mice.

## 2.5 Splicing Mutations

Splicing is a complicated process that is essential for normal protein synthesis. Any changes to this process could result in a drop in the level of the specific messenger RNA and, as a result, a protein deficiency, which could lead to abnormal cellular metabolism and function. The abnormal splicing of pre-mRNA caused by point mutations, such as nucleotide substitutions, that change the splicing regulatory sequences in a gene might even result in certain hereditary monogenic disorders.

**Mutations in the canonical acceptor and donor sites (type I and IV mutations):**
Strongly conserved regions that determine exon-intron borders are affected by mutations in the canonical acceptor and donor locations. The components of the spliceosome recognise the 5 splice site (CAG/GUAAGU sequence) and the 3 splice site (NYAG/G sequence). As a result, any changes to these canonical sequences may affect the interaction of pre-mRNA with proteins involved in intron removal. The most common mutations impact residues + 1 and + 2 at the 5 donor splice site, as well as residues 1 and 2 at the 3 acceptor splice site. Single exon skipping is frequently the result of conventional splice sequences. If the splice site is present in the intron or exon, this can result in the inclusion of the intron or the removal of an exon segment if the splice site is weak. Type I Autosomal dominant congenital cataract is a disease caused by donor splice site mutation, c.606+1G¿A in the MIP gene.

**Deep intronic variants as a cause of cryptic exon inclusion (type II mutations):**
Deep intronic mutations are substitutions that occur in large introns and result in the presence of

an intron fragment (also known as a cryptic exon or pseudoexon) in the final transcript. These mutations create new acceptor or donor sites that are identified by the splicing complex and combined with the intronic cryptic splice sites. Deep intronic mutations may also result in the development of new regulatory elements (e.g., splicing enhancers) and recognition of specific intronic sequences as exonic sequences. These mutations are not common but have significant impact on transcript splicing and futher protein synthesis. Neurofibromatosis type 1 (NF1) is a disease caused by type II mutation, c.888+651T¿A, in the NF1 gene.

**Exonic mutations affecting splicing (type III and V) mutations:**
The pattern of pre-mRNA splicing may be affected by alterations in exonic sequences. Exonic mutations could have a two-fold effect. They can add a new 5 or 3 splice site or activate a cryptic splice site that is stronger than the original, causing changes in pre-mRNA processing and exon fragment loss or type III splicing mutation. hey can add a new 5 or 3 splice site or activate a cryptic splice site that is stronger than the original, causing changes in pre-mRNA processing and exon fragment loss or type III splicing mutation.

They can add a new 5 or 3 splice site or activate a cryptic splice site that is stronger than the original, causing changes in pre-mRNA processing and exon fragment loss or type III splicing mutation. The existence of exonic mutations that impair exonic splicing enhancers may result in complete exon skipping or type V splicing mutation.

Ehlers-Danlos syndrome is caused because of c.655-2A¿G mutation, which leads to major product exon 5 and 6 skipping and partial deletion of exon 4.

**Detection and Confirmation of Splicing Mutation:**
Genomic sequencing is used to identify the mutations in splicing features of the sequence. For predicting their effects, bioinformatic algorithms are a valuable resource employed for examining the potential effects of sequence mutations. ESE Finder, based on functional SELEX technique or RESCUE-ESE, RESCUE-based MODEL HEXplorer score, and ESRsearch, all three based on relative categorization of enhancer and silencer by unanimous enrichment approach employing frequencies of hexameric sequences, can be used to examine splicing mutations. These models are based on the results of functional analyses of sequence data for enhancer or silencer qualities using minigene assays or the direct interaction between certain splicing factors and their RNA target motifs. Functional analysis of the mRNA transcripts is also done in-vitro to confirm the predictions.

## 2.6 Recent Advances in Splicing Features

2 recent advances in the splicing features are discussed here, on alternative splicing in neurons and on the involvement of snRNPs in splicing.

### 2.6.1 Alternative Splicing in Neurons

The nervous system is a complex and a veritably technical network of numerous neurons which are nerve cells. Neuronal isolation involves complex reprogramming of gene expression. Indispensable splicing of precursor mRNAs increases the complexity of transcriptomes and diversifies protein functions at the post-transcriptional position. Indeed, indispensable splicing plays an important part in neuronal isolation, axon guidance, synaptogenesis, synaptic transmission, and malleability. Because the delicate structure and function of neurons make them particularly susceptible to dysregulation

of splicing, aberrant expression or function of splicing factors may beget neuronal diseases. Thus, it's important to ameliorate our understanding of the mechanisms and physiological functions of indispensable splicing regulation in neurons. Regulation of indispensable splicing primarily involves the list of nonsupervisory factors to specific cis- rudiments of precursor mRNAs, and interplay between splicing factors may lead to fine tuning of splicing regulation, thereby diversifying the skeleton of mature products. In addition, recap rate and the vacuity of the rudimentary splicing ministry may also impact indispensable splicing. Lately, our understanding of the mechanisms underpinning indispensable splicing have been advanced from studies of several neuronal splicing factors; these studies have employed inheritable knockout or complaint models as well as genome-wide analysis of mRNA isoforms. Our understanding of splicing regulation mechanisms and splicing nonsupervisory networks has been advanced mainly by recent studies using gene inactivation ways and genome-wide experimental and computational examination of indispensable splicing events. In the past decade, CLIP in confluence with colorful types of mRNA identification systems has been used considerably for in vitro study of splicing factors and their regulation mechanisms. Ablation of splicing factors in dressed cells by RNA hindrance has also been extensively used for mechanistic studies of indispensable splicing of endogenous or journalist minigene reiterations. Nonetheless, we're still at the morning of our understanding of the mechanistic and, in particular, physiological aspects of indispensable splicing regulation.

Our understanding of the physiological consequences of indispensable splicing still largely relies on inheritable approaches. For illustration, knockout of splicing factors in creatures in combination of mRNA isoform comparison can grease the identification of their in vivo targets and natural functions. Study of complaint-affiliated splicing factors can in particular give perceptivity into pathogenesis of aberrant splicing. Also, knock-in or knockout of specific mRNA isoforms can help to unveil their functional consequences, which is inadequately understood, and may indeed allow delineation of unproductive goods (Moroy  Heyd, 2007). Still, progress has been fairly slow owing to limitations of inheritable ways in mammalian systems. At present, effective recombination technologies are being developed to grease high-out- outturn gene knockout in embryonic stem cells (Valenzuela, etal., 2003) which may allow large-scale analysis of natural functions of splicing factors as well as mRNA isoforms. Besides more effective/ accessible inheritable tools, high-outurn whole-transcriptome sequencing and expansive bioinformatics tools have proved their advantage. With these ways, we will begin to establish a more accurate paradigm for mRNA splicing nonsupervisory networks with physiological significance.

### 2.6.2 Involvement of snRNPs in Splicing

Discrete, stable small RNA motes are planted in the capitals of cells from a wide variety of eukaryotic organisms2. Numerous of these small nuclear RNA (snRNA) species, which range in size from about 90 to 220 nucleotides, have been well-characterized biochemically 3 – 6, and some sequence 7,. Still, their function has remained obscure. The most abundant snRNA species live as a nearly related set of RNA – protein complexes called small nuclear ribonucleoproteins (snRNPs) 9. snRNPs are the antigens recognised by antibodies from some cases with lupus erythematosus (LE), an autoimmune rheumatic disease10,. Anti-RNP antibodies from lupus sera widely precipitate snRNP species containing Ula7 and Ulb9 RNAs from mouse Ehrlich ascites cell capitals, whereas anti-Sm antibodies bind these snRNPs and four others containing U2 (ref. 8), U4, US and U6 (ref. 9) RNAs. Both antibody systems precipitate the same seven prominent nuclear proteins. All notes of the snRNAs U1, U2, U4, U5 and U6 appear to live in the form of antigenic snRNPs9. The patches deposition at about 10S and each presumably contains a single snRNA patch. Circular immunofluorescence stud-

ies (refs 12, 13, and unpublished compliances) using anti-RNP and anti-Sm sera confirm the nuclear (butnon-nucleolar) position of the antigenic snRNPs. Then we present several lines of substantiation that suggest a direct involvement of snRNPs in the splicing of hnRNA. Utmost interesting is the observation that the nucleotide sequence at the 5  end of U1 RNA exhibits expansive complementarity to those across splice junctions in hnRNA motes.

# 3 Protein Splicing

Protein splicing is an intramolecular event in which an internal protein segment (known as an intein) is excised from a precursor protein on both sides by ligation of C-terminal and N-terminal external proteins (known as exteins). The precursor protein's splicing junction contains mostly cysteine or serine, which are amino acids with a nucleophilic side chain. Exogenous cofactors or energy sources like adenosine triphosphate (ATP) or guanosine triphosphate (GTP) are not required for protein splicing reactions that are known now. Splicing is normally solely linked with pre-mRNA splicing. This precursor protein is made up of three segments: an N-extein, an intein, and a C-extein. The resultant protein has the N-extein coupled to the C-extein after splicing, this splicing product is also known as an extein.

## 3.1 Mechanism

First, the side chain of the first residue (a serine, threonine, or cysteine) of the intein portion of the precursor protein nucleophilically attacks the peptide bond of the residue immediately upstream to it. This leads to the formation of a linear ester (or thioester) intermediate. Now, the side chain of the first C-extein residue attacks the newly formed ester. After this process, the N-terminal end of the intein is released. This results in a branched intermediate to which the N-extein and C-extein are attached, though not via a peptide bond. The last residue of the intein is always an aspargine. The amide nitrogen atom of the asparagine side chain cleaves the peptide bond between the intein and the C-extein which results in a free intein segment with a terminal cyclic imide. Finally, the free amino group of the C-extein attacks the (thio)ester that connects the N- and C-exteins. A peptide bond and the functional, ligated protein are created via an O-N or S-N shift.

The diagram below depicts the splicing mechanism in detail:

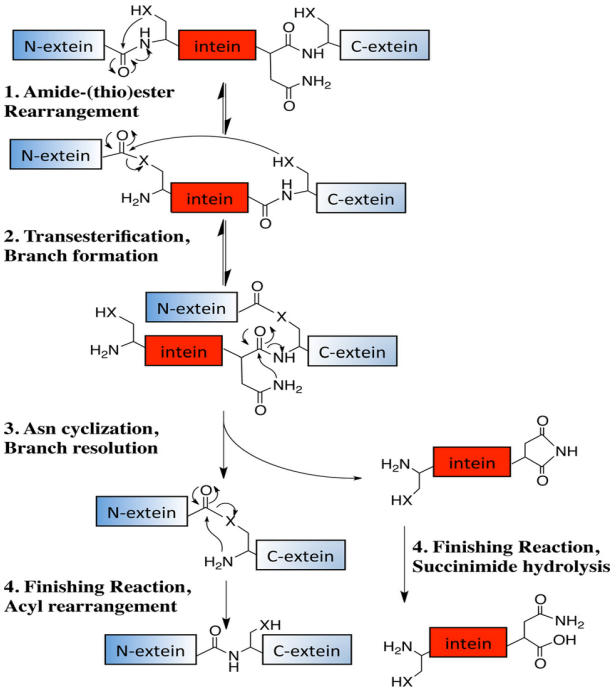


Figure 2: Protein splicing: how inteins escape from precursor proteins
https://www.researchgate.net/publication

## 3.2 Conditional Protein Splicing

To regulate protein activity in vivo, splicing takes place conditionally either in trans or in cis. The activation of CPS can be triggered by temperature, a small molecule, light, pH or change in redox state. CPS remain to be shown that inteins are sensitive to signals in the context of their native exteins, expressed in a host organism. Inteins are seen as selfish genetic elements, which are difficult to remove as they interfere with key proteins such as DNA pol and recombinase. But some inteins may be beneficial for the host, due to which they should be retained.

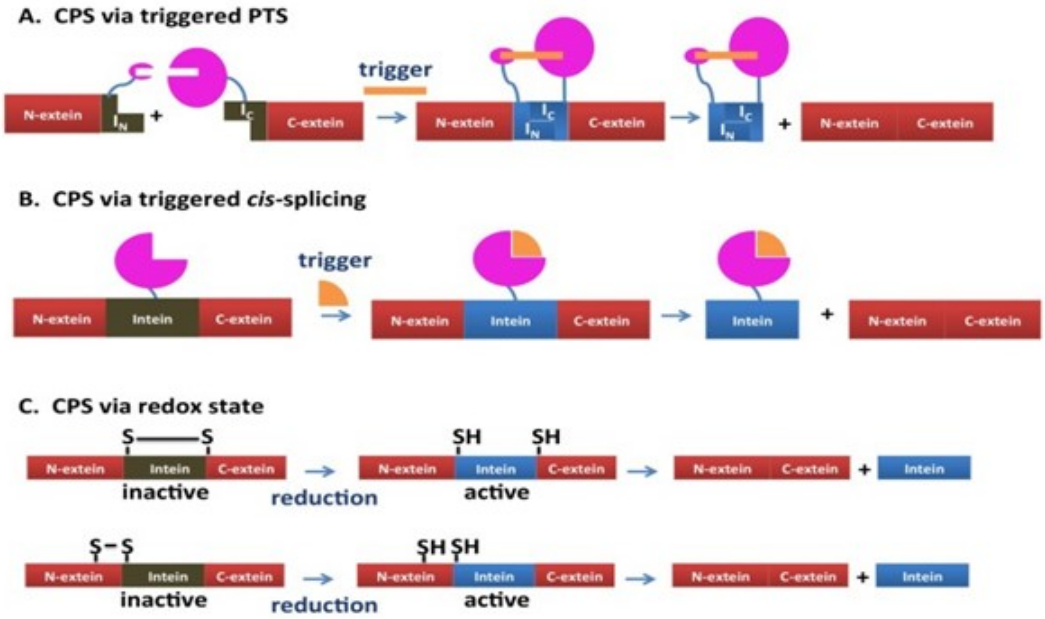


Figure 3: Schematic representation of conditional protein splicing (CPS) : Brown boxes indicate an inactive intein, whereas blue boxes indicate an active intein
https://media.springernature.com/lw685/springer-static/image/art%3A10.1186%2F1759-8753-5-5/MediaObjects/13100_2013_Article_91_Fig5_HTML.jpg?as=webp

To control cis-protein splicing some engineered inteins have also been created. Earlier Mtu RecA intein was interrupted by the human estrogen receptor to its ligand-binding domain rather than its endonuclease domain and modified by directed evolution to splice only with the addition of 4-hydroxytamoxifen in S. cerevisiae. Later it was extended to the mammalian system as well to facilitate splicing of reporter proteins and transcription factors which involves in the hedgehog pathway, and to create a CPS-activated histone H2A variant in S. cerevisiae.

Other than this Mtu RecA intein was also interrupted by the human thyroid hormone receptor and have been shown to regulate splicing of -lactamase and -galactosidase in E. coli in response to thyroid hormone.Therefore, it is possible to design inteins that can regulate (ON/OFF) with help of small molecules. Such small molecules can be used to control the active, non-engineered inteins.

Recently, cisplatin has been found to inhibit protein splicing both in vitro and in E. coli and M. tuberculosis and divalent cations can prevent protein splicing in vitro.Recently, cisplatin has been found to inhibit protein splicing both in vitro and in E. coli and in M. tuberculosis and divalent cations can prevent protein splicing in vitro.

## 3.3 Applications

Protein splicing is an autocatalytic process in which an intervening protein domain (intein) excises itself from the polypeptide in which it is embedded, concomitantly creating a new peptide bond be-

tween its two flanking regions (exteins). Several engineered inteins have been developed that allow access to recombinant protein $\alpha$-thioester derivatives by thiolysis of the corresponding C-terminal intein fusions. Inteins have also been engineered to allow the introduction of an N-terminal cysteine moiety into recombinant proteins. Simple access to reactive proteins without any size restriction through molecular biology techniques suddenly enabled the application of NCL to the modification of a much larger fraction of the proteome.

The simplest application of EPL or PTS is the modification of the N- or C-terminal regions of a protein because this can be achieved in a single ligation step involving a synthetic peptide fragment, containing the desired chemical probe(s) and a recombinant protein fragment. Central regions of the protein of interest can also be labeled, but a three-piece ligation strategy is then required, which is more technically challenging. It should be noted that EPL and PTS can be used to link a recombinant protein to a nonpeptidic moiety, provided it has the necessary reactive handles for ligation. For example: The attachment of proteins to surfaces, polymers, and nucleic acids.

Protein trans-splicing (PTS) mediated by naturally-occurring or artificially designed split-inteins has been also widely used for the modification of proteins both in vitro and in cell.

Phosphorylation is one of the most common and extensively studied PTMs. It should not be surprising then that EPL has been heavily utilized for the preparation of proteins containing this modification. Glycosylation is arguably the winner among the PTMs in terms of shear chemical complexity. The attached sugars can be composed of several different monosaccharide building blocks linked together in elaborate branched structures whose tailoring can differ from molecule to molecule. PTS has also been used in vivo to add synthetic probes, such as noncoded fluorescent probes.

## 3.4 Intein based Biosensors

Both the research of human illnesses and the validation of new therapeutic molecules require the development of diagnostic tools for the investigation of complicated biological challenges. Many fascinating novel assays and biosensors based on proteins have been developed, allowing for efficient and accurate investigation of key processes. In vivo diagnostics of protein–protein interactions, protein translocation to the nucleus or mitochondria, and protein release from mitochondria are a few examples.

A biosensor for cytosolic caspase activity has been developed thanks to an unique use of SICLOPPS technology. Caspases are involved in apoptosis and have been linked to a variety of disorders, including ischemia, cancer, Alzheimer's disease, and autoimmune diseases. Understanding caspase activity in vivo is critical for figuring out how these illnesses work and generating effective therapies. Kanno et al. used the inverted Ssp DnaE split-intein to cyclize firefly luciferase in order to establish a specific in vivo/in vitro test for caspase activity. This protein was modified so that an internal luciferase site was fused to the intein fragments, and a solvent accessible caspase-3 cleavage site was employed to connect the enzyme's native N- and C-termini. The ability of the N- and C-terminal domains of luciferase to close when substrate is bound is critical to the enzyme's catalytic activity, and joining the termini was thought to prevent this. Both the non-spliced precursor and the cyclized luciferase including the bridging protease site show no luminous activity in control trials. The bridge region is cut in the presence of caspase-3, releasing the two domains and restoring luminescence. Kanno et al. subcutaneously inserted HeLa cells into the backs of living mice to demonstrate the system's power,

and then used staurosporine to trigger caspase-3. Caspase-3 activity in vivo was seen in real time by measuring luminescence and demonstrated dosage dependent on caspase-3 production. Because this assay technique can be simply adapted to test for any protease, it will be of immediate use in the development and validation of protease inhibitor medications and biological processes that trigger proteases. Skretas et al. created a bacterial system capable of detecting human nuclear hormone receptor ligands in another recent work. By fusing a non-splicing Mtu RecA intein to the ligand binding domain of the human oestrogen receptor and T4 bacteriophage thymidylate synthase, then expressing this fusion in an E. coli strain, a highly sensitive bacterial sensor for endocrine modulators was developed. The thyA strain of E. coli. Following that, the system was further refined, and it was shown that this bacterial sensor can appropriately detect agonists and antagonists for the human oestrogen receptor. This is the first artificial system to discriminate between the two types of effectors, and it offers promise for quick screening of new nuclear hormone receptor modulators utilising a bacterial growth phenotype. Intein splicing is a structural component that stabilises the oestrogen receptor fold and allosterically conveys the receptor's binding status to the thymidylate synthase subunit.

The intein scaffold's stable yet flexible structure appears to allow it to convey tiny structural changes in the ligand binding domain to the dimerization interface of thymidylate synthase, resulting in the system's high sensitivity.

# 4 References


1. https://www.youtube.com/watch?v=IJQv1H-2IoI
2. http://www.premierbiosoft.com/tech_notes/gene-splicing.html#:~:text=Gene%20splicing%20is%20a%20post,important%20source%20of%20protein%20diversity
3. https://www.spandidos-publications.com/10.3892/br.2014.407
4. Nature volume 283, pages 220–224 (1980).
5. Douglas AG, Wood MJ. RNA splicing: disease and therapy. Brief Funct Genomics. 2011 May;10(3):151-64. doi: 10.1093/bfgp/elr020. PMID: 21628314.
6. Di C, Syafrizayanti, Zhang Q, Chen Y, Wang Y, Zhang X, Liu Y, Sun C, Zhang H, Hoheisel JD. Function, clinical application, and strategies of Pre-mRNA splicing in cancer. Cell Death Differ. 2019 Jul;26(7):1181-1194. doi: 10.1038/s41418-018-0231-3. Epub 2018 Nov 21. PMID: 30464224; PMCID: PMC6748147.
7. Scotti MM, Swanson MS. RNA mis-splicing in disease. Nat Rev Genet. 2016 Jan;17(1):19-32. doi: 10.1038/nrg.2015.3. Epub 2015 Nov 23. PMID: 26593421; PMCID: PMC5993438.
8. Oltean S, Bates DO. Hallmarks of alternative splicing in cancer. Oncogene. 2014 Nov 13;33(46):5311-8. doi: 10.1038/onc.2013.533. Epub 2013 Dec 16. PMID: 24336324.
9. Anna A, Monika G. Splicing mutations in human genetic disorders: examples, detection, and confirmation. J Appl Genet. 2018 Aug;59(3):253-268. doi: 10.1007/s13353-018-0444-7. Epub 2018 Apr 21. Erratum in: J Appl Genet. 2019 May;60(2):231. PMID: 29680930; PMCID: PMC6060985.

10. Arnold U, Hinderaker MP, Nilsson BL, Huck BR, Gellman SH, Raines RT. Protein prosthesis: a semisynthetic enzyme with a beta-peptide reverse turn. J Am Chem Soc. 2002 Jul 24;124(29):8522-3. doi: 10.1021/ja026114n. PMID: 12121081.

11. Cheriyan M, Perler FB. Protein splicing: A versatile tool for drug discovery. Adv Drug Deliv Rev. 2009 Sep 30;61(11):899-907. doi: 10.1016/j.addr.2009.04.021. Epub 2009 May 13. PMID: 19442693.

12. Topilina NI, Mills KV. Recent advances in in vivo applications of intein-mediated protein splicing. Mob DNA. 2014 Feb 4;5(1):5. doi: 10.1186/1759-8753-5-5. PMID: 24490831; PMCID: PMC3922620.