Misregulation of alternative splicing as a novel target for drug intervention

TA Thanaraj[1], Olivier Cochet[2] and Stefan Stamm[3]

1. Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridgeshire, CB10 1SD, UK

2. Exonhit Therapeutics, 65 Boulevard Masséna 75013 Paris, France

3. University of Erlangen-Nurenberg, Fahrstraße 17, 91054 Erlangen, Germany. Tel: +49 9131 8524622, email: stefan@stamms-lab.net

Recent research has shown that alternative missplicing has been largely underestimated as a cause and consequence of disease. A better understanding of the mechanisms regulating splice site selections now allows for the correction of missplicing events in *in vivo* test systems. Several examples show that alternative missplicing events can be targeted by small molecules and present novel drug targets

The sequencing of the human genome has revealed that fewer genes than previously anticipated create the complex human proteome. Post-transcriptional mechanisms such as alternative pre-mRNA splicing contribute to the observed diversity of the transcriptome. About 90% of the pre-mRNA consists of intervening sequences (introns) located between sequences (exons) that are exported into the cytosol for translation[1]. In the majority of genes, parts of the pre-mRNA can be used alternatively – sometimes they are removed as introns; sometimes they are expressed as exons. EST comparisons and a detailed analysis of chromosomes 22 and 19 estimate the number of genes that undergo alternative splicing to be between 47 and 60%[1,2]. Often cells alter their usage of alternative exons in response to external stimuli, indicating that splicing modulation is an important physiological adaptation process[3].

Alternative exons introduce several features into the pre-mRNA. The best understood functions are stop codons or frameshifts that are introduced by 20–35% of alternative exons[3,4]. 74% of these exons fulfill the criteria for nonsense-mediated decay, indicating that at least 26% of transcripts could be switched off by the combination of alternative splicing and nonsense-mediated decay[4]. The function of the remaining alternative exons is mostly unclear. However, detailed analyses of several exons revealed that alternative exons can create soluble receptors, change intracellular localisation, alter ligand affinity, produce inactive protein variants and change the properties of ion channels[5].

SPLICE SITE SELECTION

The catalytic mechanism of pre-mRNA splicing has been determined in great detail[6]. In contrast, it is not clear how exons are accurately recognised on the pre-mRNA. Currently it is not possible to precisely predict exons from genomic RNA. Exons are hard to predict as their defining sequences, the 5', 3' splice sites and the branchpoint are only weakly conserved. Additional elements known as exonic/intronic enhancers are necessary for exon recognition[7]. As these elements are mostly in coding regions they are highly degenerate. This allows for the flexibility needed to encode proteins. Despite the intrinsic weakness of each signal, splice sites can be recognised *in vivo* with high fidelity. This is achieved by binding of proteins to these enhancer sequences. These proteins can be subdivided into two major classes: SR/SR-like proteins[8] and hnRNPs[9], which usually contain RNA binding and protein:protein interaction motifs. Although the interaction of each protein with a target sequence is weak, the high specificity needed for exon recognition is achieved through the interactions of multiple RNA elements with several RNA binding proteins that can interact with each other. As a result, exons are recognised *in vivo* with high specificity due to combinatorial control[10] (Figure 1). Since the concentrations of regulatory proteins differ between tissues, cell types or developmental stages, exons can be recognised alternatively. Several kinases have been shown to target splicing regulatory factors. As a result, signal transduction pathways that control these kinases can influence splice site selection[11].

MISSPLICING-ASSOCIATED DISEASE

The importance of alternative splicing is evidenced by the increasing numbers of diseases associated with missplicing events[12,13]. About 10–15% of human disease-causing mutations affect splicing [14,15]. They can be categorised into two types. Type I mutations affect invariant positions of splice sites, usually resulting in severe disease as recognition of the affected exon is destroyed. Type II mutations occur in variant positions of the splice sites and in enhancer or silencer regions. These mutations often lead to an altered ratio of alternative exon usage and the resulting phenotypes can be more subtle (Table 1). A change of alternative splicing without obvious mutations has been observed in a number of diseases. For example, in schizophrenia, altered isoform ratios of the GABA-A receptor[16], N-CAM[17] and NMDA R1 receptor[18] have been observed. This indicates that small defects in the pre-mRNA processing machinery could cause diseases that are manifested in altered alternative splicing (Table 2, overleaf). The effects of this altered splicing pattern can be small and manifest only over time or predispose an organism to a disease[12]. An interesting observation is that the same mutation-bearing allele shows different alternative

FIGURE 1. EXON

Exons are indicated as thin lines. S elements (enhancer shown as gray box thin boxes in intro site (CAGguaagu) $(y)_{10}$ncagG, as well (ynyyray), are indica g, c or u). Upper-c nucleotides that re mRNA. Two major protein:RNA intera snRNP close to th (shown in red). Th multi-protein:RNA frequent in pre-m sites, and is antag yellow: hnRNPs;

splicing patterns de background, which tive splicing is a po

Since it is hard spliced exons *in s* native exons depe ison between RN EST or cDNA d sequence. In ad databases are a experimentally ve leaf). Datasets d verified exons or but accurate an biological proper relying on ESTs but are larger a 5' and 3' regio these problem containing bo computer gene oped by the AS base) consor Currently, a d delineated a using detailed sequences wi a database of collected from are available.

STRATEGIE

Several expe ciple that dis of missplicin major appro

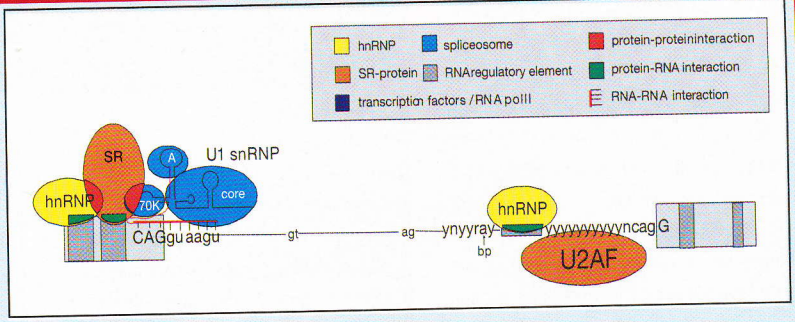

FIGURE 1. EXON RECOGNITION

Exons are indicated as boxes, introns as thin lines. Splicing regulator elements (enhancers or silencers) are shown as gray boxes in exons or as thin boxes in introns. The 5' splice-site (CAGguaagu) and 3' splice-site $(y)_{10}$ncagG, as well as the branch point (ynyyray), are indicated (y=c or u, n=a, g, c or u). Upper-case letters refer to nucleotides that remain in the mature mRNA. Two major groups of proteins, hnRNPs (yellow) and SR or SR related proteins (orange), bind to splicing regulator elements; the protein:RNA interaction is shown in green. This protein complex assembling around an exon enhancer stabilises binding of the U1 snRNP close to the 5' splice-site, for example due to protein:protein interaction between an SR protein and the RS domain of U170K (shown in red). This allows hybridization (thick red line with stripes) of the U1 snRNA (red) with the 5' splice-site. The formation of the multi-protein:RNA complex allows discrimination between proper splice-sites (bold letters) and cryptic splice-sites (small gt ag) that are frequent in pre-mRNA sequences. Factors at the 3' splice-site include U2AF, which recognises pyrimidine rich regions of the 3' splice-sites, and is antagonised by binding of several hnRNPs (eg, hnRNP I) to elements of the 3' splice-site. Orange: SR and SR related proteins; yellow: hnRNPs; green: protein:RNA interaction; red: protein:protein interaction; thick red line with stripes: RNA:RNA interaction.

splicing patterns depending on the genetic background, which suggests that alternative splicing is a potential genetic modifier[19].

Since it is hard to predict alternatively spliced exons *in silico*, databases of alternative exons depend on sequence comparison between RNA sequences present in EST or cDNA databases with genomic sequence. In addition, manually curated databases are available that depend on experimentally verified exons (Table 3, overleaf). Datasets derived from experimentally verified exons or cDNA are inevitably small but accurate and contain annotations of biological properties. In contrast, databases relying on ESTs lack biological annotations, but are larger and biased to events on the 5' and 3' regions of genes. To overcome these problems, an integrated database containing both human curated and computer generated data is being developed by the ASD (alternative splicing database) consortium (www.ebi.ac.uk/asd). Currently, a database of computationally delineated alternative splice events using detailed alignments of EST/cDNA sequences with genome sequences, and a database of alternatively spliced exons as collected from peer-reviewed journal articles are available.

STRATEGIES FOR INTERVENTION

Several experiments have proven the principle that diseases caused by an alteration of missplicing can be reversed *in vivo*. Two major approaches are emerging. Defective regulatory sequences can be masked with oligonucleotides or the concentration of regulatory factors can be changed. Oligonucleotides have been used to alter missplicing events[20] and were shown to revert aberrant splicing in beta-thalassemias[21], FTDP-17[22], spinal muscular atrophy[23], duchenne muscular dystrophy[24], and myasthenia gravis associated with missplicing of acetylcholine esterase[25].

Since alternative splice site recognition depends on a combinatorial control of various splicing factors, changing the relative ratios of these factors can correct missplicing *in vivo*. For example, overexpression of the splicing factor tra2-beta1 stimulates exon inclusion of the SMN2 gene which could then substitute for the SMN1 gene product missing in spinal muscular atrophy[26]. Mutations in exonic and intronic enhancer elements of tau exon 10 change the ratio tau isoforms which lead to FTDP-17[27]. Hyperphosphorylation of splicing factors by transfecting the appropriate kinases can revert this missplicing event *in vivo*[28]. Despite the obvious problems associated with the delivery of these agents, these studies prove the principle that alternative missplicing can be reversed *in vivo*.

TABLE 1. SOME ALTERNATIVE MISSPLICING EVENTS ASSOCIATED WITH HUMAN DISEASE CAUSED BY MUTATIONS IN EXONIC ENHANCERS

Disease	Gene	Reference
FTDP-17	Tau	[35]
Spinal muscular atrophy	SMN2	[36]
Sandhoff disease	Beta-hexaminidase	[37]
Acute intermittent porphyria	Porphobilinogen deaminase	[38]
Tyrosinemia, Type I	Fumarylacetoacetat hydrolase	[39]
Leigh's encephalomyelopathy	Pyruvat dehydrogenase E1	[40]
Menkes disease	MNK	[41]
Immunodeficiency	Adenosine deaminase	[42]
Metachromic leukodystrophy	Arylsulfatase A	[43]
Marfan Syndrom	Fibrillin-1	[44]
Cerebrotenidinous xanthomatosis	CYP 27	[45]
Beta-Thalassemia	Beta-globin	[46]
Breast and ovarian cancer	BRCA1	[47]
Neurofibromatosis type I	NF-1	[48]
Myotonic dystrophy	DMPK	[49]
Occipital horn syndrome	ATP7A	[50]
Cystic fibrosis	CFTR	[51]

EXAMPLES OF DRUG INTERVENTION

The use of RNA binding proteins, such as gentamicin, chloramphenicol and tetracycline clearly demonstrates that drugs can be targeted against RNA and/or RNA binding proteins. Several examples demonstrate that splice site selection can be influenced by small drugs. For example, aclarubicin can reverse the wrong splicing pattern of the SMN2 gene allowing it to substitute for the loss of SMN1 in spinal muscular atrophy[29]. The same exon is influenced by sodium butyrate treatment that most likely changes the expression of regulatory splicing factors[30].

TABLE 2. SOME HUMAN DISEASES ASSOCIATED WITH CHANGES IN THE RATIOS OF ALTERNATIVE SPLICED PRODUCTS (more examples at the ASD site: www.ebi.ac.uk/asd)

Disease	Gene	Reference
Wilms tumour	WT1	[52]
Breast cancer	MDM2	[53]
Melanoma	Bin1	[54]
Prostate cancer	FGFR-2	[55]
Sporadic amyotrophic lateral sclerosis	EAAT2	[56]
Sporadic amyotrophic lateral sclerosis	NOS	[57]
Schizophrenia	N-CAM	[17]
Schizophrenia	NMDA R1	[18]
Schizophrenia	GABA-A	[16]

TABLE 3. EXISTING DATABASES THAT COMPILE HUMAN ALTERNATIVE EXONS, REGULATORY PROTEINS AND DISEASES

Database	Entry mode	Species covered	Reference	URL
ASD-AEDB	Manual	All animals	[58]	www.ebi.ac.uk/asd/
ASD- AltExtron	Computer generated	Human and eight other species	[59, 60]	www.ebi.ac.uk/asd/
ASD-AltSplice	Computer generated	Human	[61]	Sapiens.wustl.edu/~zkan/TAP/ALTSEQ.htm
AltRefSeq	Computer generated	All animals	[62, 63]	Cbcg.nersc.gov/asdb
ASDB	Computer generated	Human and six other species	[64]	www.bioinf.mdc-berlin.de/asforms
Asforms	Computer generated	Mammalian	[65]	http://166.111.30.65/ASMAMDB.html
AsMAMDB	Computer generated	Human	[66]	www.ebi.ac.uk/~thanaraj/gcag
GC-AG introns	Computer generated	Human	[2]	www.bioinformatics.ucla.edu/HASDB
HASDB	Computer generated	All species	[67]	lsis.bit.uq.edu.au/front.html
ISIS	Manual	Animals, neurons	[68]	ftp://phage.cshl.org/pub/science/alt_exon
Neuron-specific exons	Computer generated	Metazoan splicing factors	[69]	www.sanger.ac.uk/cgi-bin/Pfam/getacc?PF00076
RRM containing proteins	Manual	Human	[70]	www.cookie.imcb.osaka-u.ac.jp/nakai/asdb.html
Splice site mutations	Manual	Human	[15]	www.hgmd.org
Mutations				

An important function of alternative spliced genes is the introduction of new stop codons and their misregulation can cause disease. For example, an estimated 5% of all mutant alleles causing cystic fibrosis carry a premature stop codon. Aminoglycoside antibiotics such as gentamycin or tobramycin can suppress premature stop codons *in vivo* and their administration was shown to be effective in treating the G542X cystic fibrosis mutation in transgenic mice[31] and shows promising effects in humans[32]. A number of alternative splicing events determine ligand binding specificity. It is therefore not surprising that in return a drug can act in an isoform specific fashion. The cyclooxygenase-1 gene generates several isoforms via alternative splicing, but only the COX-3 isoform seems to be the target of nonsteroidal anti-inflamatory drugs, such as acetaminophen[33]. Interestingly, the length of the alternatively spliced intron could be polymorphic, which would explain differences in drug action between individuals[33,34].

A member of the ASD consortium, ExonHit Therapeutics is a pioneer in the analysis of splicing events and has developed DATAS (differential analysis of transcripts with alternative splicing). This technology detects differences between the various mRNAs generated during DNA transcription. Using DATAS, targets for amyotrophic lateral sclerosis (ALS) treatment were identified. Tissue samples of hSOD-1 G93A transgenic mice (a well-characterised mouse model for ALS) and non-transgenic siblings were compared. A number of mRNAs were identified, which are differentially spliced between the transgenic mice and controls. Of these, 13 have already been implicated in the pathogenesis of ALS or in neurodegenerative disease. For one of the pertinent proteins, a chemical modulator (EHT201) was known, and appeared appropriate for testing in humans. A phase II clinical trial with 400 ALS patients from 12 European centres is currently ongoing to evaluate the safety and the efficacy of EHT 0201.

Since it is now obvious that alternative splicing plays an important role in physiological and pathological situations, this mechanism is a large novel target for drug intervention. ●

REFERENCES

Due to the number of citations in this article, references will be sent by email on request. Email the editor on steve.handley@hhc.co.uk with the words 'References for pv/splicing' in the subject line.