Part One
Theory

Alternative pre-mRNA Splicing: Theory and Protocols, First Edition. Edited by Stefan Stamm, Chris Smith, and Reinhard Lührmann.

1
Splicing in the RNA World

Emanuele Buratti, Maurizio Romano, and Francisco E. Baralle

Key Concepts

- Alternative splicing is a key element of eukaryotic gene expression.
- Almost all polymerase II transcripts are alternatively spliced.
- RNA is chemically and structurally more flexible than DNA, and can act as a catalyst.
- RNA is an active player in mediating genetic information, not just a static messenger.
- Almost the entire human genome is transcribed into RNA and new classes of noncoding RNA molecules are emerging.
- The number of diseases known to be associated with alternative splicing is steadily increasing.

1.1
Introduction: The Fascination of Alternative Pre-mRNA Splicing

The genetic information is stored in DNA, which is transferred from one generation to the next. During the life of a cell, this DNA information is retrieved as RNA. Whereas DNA is chemically very stable and therefore well suited to archive the genetic information, RNA is chemically more reactive, and thus unstable. Therefore, with the exception of RNA viruses, RNA does not store the genetic information but rather acts as an intermediate between DNA and proteins.

However, RNA does not simply copy the genetic information, as the primary RNA transcript generated from DNA undergoes processing. Most human polymerase II transcripts contain exonic sequences that are finally exported into the cytoplasm (exons, for exported sequence), whereas the intervening sequences (introns) remain in the nucleus. The removal of the introns and the joining of the exons is known as pre-mRNA splicing [1–3]. Almost all human protein-coding genes undergo alternative splicing (AS; see Chapter 3 Hertel) [4]; this means that, depending on the cellular conditions, an alternative exon can be either included or removed from the final messenger RNA (mRNA). For example, the protein kinase CβII gene contains an alternative exon encoding a protein part that regulates the subcellular localization and substrate specificity of the kinase. In skeletal muscle, the inclusion of this exon is promoted by insulin, via a phosphatidylinositol 3-kinase-dependent pathway [5,6]. This example shows how the readout of the genetic information is regulated by AS in response to a daily activity, such as the eating of a meal. The carbohydrates in the food trigger an insulin response; the insulin binds to receptors on muscle cells that initiate a phosphorylation cascade which modulates the splicing machinery to use only certain parts of the genetic information, which in turn generates a regulatory protein with altered properties (see Chapter 48 Patel for signaling and splicing). This shows that the type of information transferred from the genome to the cell depends on inputs that the cell receives, which implies that the output of a gene is only defined in the context of the cellular state.

Alternative pre-mRNA Splicing: Theory and Protocols, First Edition. Edited by Stefan Stamm, Chris Smith, and Reinhard Lührmann.

RNA is therefore more than just a copy of the genetic information: RNA can "interpret" the genetic information depending on environmental cues that the cell receives. Alternative splicing is a central mechanism in this interpretation process, as it allows the expression of selected parts of genetic information.

Due to its role as a flexible "interpreter," AS strongly enhances the number of proteins that can be encoded by the genome. For example, by combining one exon out of four alternatively spliced regions that contain 12, 48, 33, and 2 alternative exons each, the *Drosophila Dscam* gene can generate 38 016 protein isoforms ($12 \times 48 \times 33 \times 2$) [7]. Deep sequencing results (see Chapters 50 Guigó, 51 Zhang for this method) indicate that the fly actually generates this large number of isoforms. Alternative splicing can thus generate from a single gene a number of protein isoforms that is larger than the total number of protein coding genes in *Drosophila*.

The ability to change the output of the genetic information depending on cellular states, and the ability to expand the information content of the genome, makes AS a central element in gene expression. About 30 years after the discovery of splicing [1,2], we are now beginning to understand on a molecular level how AS can be such a fascinating biological process (see Chapters 3 Hertel, 5 Lührmann, and 8 Smith).

1.2 RNA Can Adopt a Flexible Conformation

RNA molecules can be represented by a linear sequence of four classical bases: adenine and guanine (A/G, both purines); and cytosine and uracil (C/U, both pyrimidines). These bases can be subjected to more than 100 post-transcriptional modifications that are currently listed in the RNA modification database [8] http://library.med.utah.edu/RNAmods (see also Chapter 14 by Höbartner for synthetic available bases). In the RNA molecule, each of these bases (schematically represented in Figure 1.1a) is bound to the 1′ position of a ribose sugar that, through its 3′ position, utilizes a phosphate group to link with the 5′ position of the next ribose. The most important features that distinguish RNA (ribonucleic acid) from DNA (deoxyribonucleic acid) is the presence of a hydroxyl group (−OH) in the 2′ position of the ribose sugar (Figure 1.1b). The 2′ hydroxyl group is chemically reactive, which not only makes RNA more vulnerable to degradation but also it to participate in chemical reactions.

Although RNA molecules are described as a single-stranded sequence, most RNA molecules exhibit a high degree of double-helical character, as complementary segments of the RNA fold back on each other. The base-pairing of RNA is more flexible than that of DNA. In addition to the canonical Watson–Crick base pairs (cytosine with guanine, adenine with uracil), there are "noncanonical" base pairs, such as G–U pairing, and numerous other base pairings are possible [9]. Since the areas of complementarity in an RNA molecule are short, RNA molecules show local regions of base-pairing, which is referred to as "secondary structure." RNA secondary structures are locally confined, which is in contrast to the extended double-stranded DNA helix.

As RNAs do not form a long-range double-stranded structure, the short RNA helices themselves can interact with each other to form what is known as the "tertiary structure" [10]. The rules that exactly define the final outcome of these folding processes, and the various factors that influence them, remain the subject of many active studies. In contrast to proteins, it is currently not possible to predict *in vivo* RNA tertiary structures accurately [10] (see Chapter 54 Bujnicki for structure prediction of splicing proteins). X-ray crystallography experiments have clearly shown defined tertiary structures for metabolically stable RNAs, such as transfer RNAs (tRNAs) or ribosomal RNAs (rRNAs) [11]. In contrast, structures in pre-mRNAs that form the substrate of the spliceosome can currently be predicted only indirectly by mutagen-

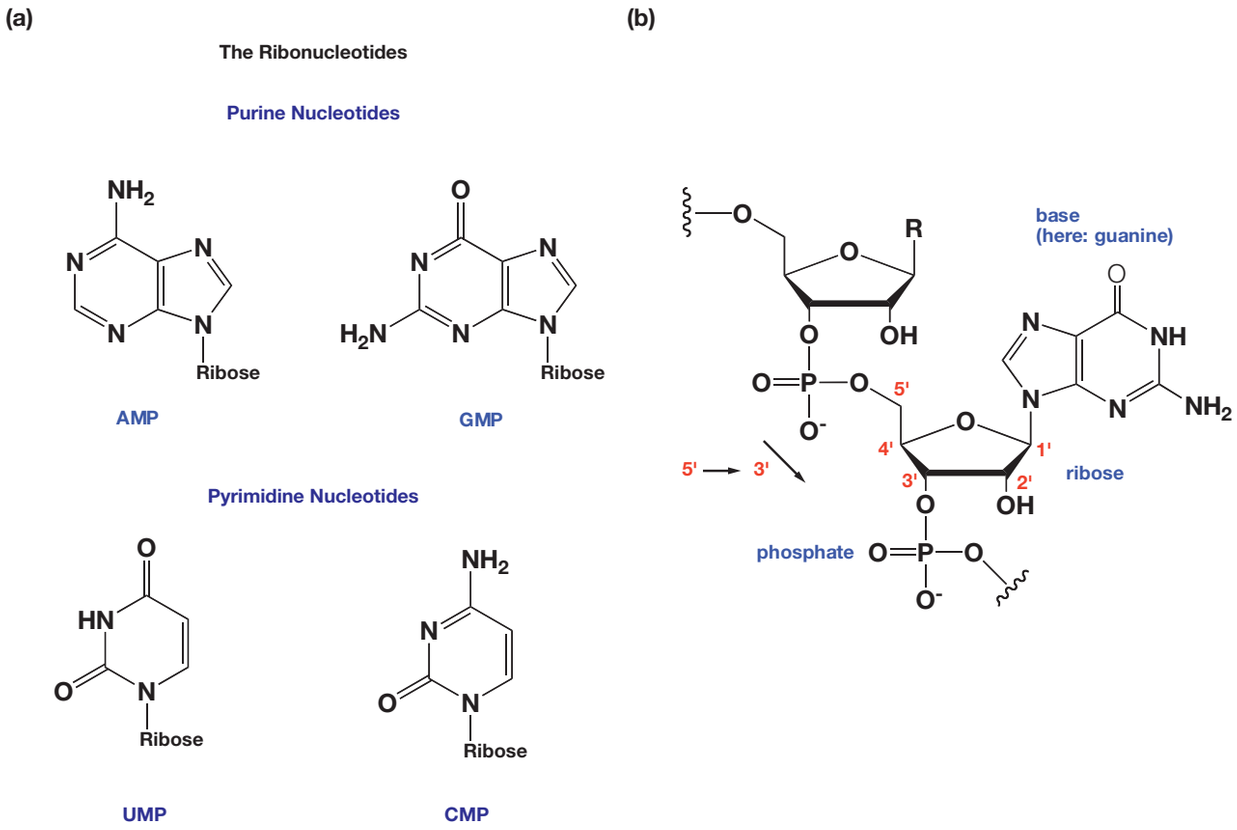

Fig. 1.1 The structure of RNA. (a) Schematic representation of the four major bases of the ribonucleotides: adenosine (AMP) and guanine (GMP) that are both purines, and uracil (UMP) and cytosine (CMP) that belong to the pyrimidines; (b) The structure of RNA.

esis or bioinformatic analyses, as structures on pre-mRNA will most likely be formed only transiently (see Chapter 53, Hiller).

1.3 Enzymatic RNAs and the RNA World

The lack of a complementary strand, and the presence of the 2′ hydroxyl (OH) group, which confers chemical reactivity, combined with the ability to fold into complex tertiary structures, allow RNA to perform catalytic reactions [12]. The first example of an RNA with catalytic activity was the self-splicing pre-ribosomal RNA from the ciliate *Tetrahymena* [13], followed by the discovery of RNase P, a ribonucleoprotein complex that cleaves tRNAs [14]. The general catalytic mechanism of these RNA enzymes is an activation of the 2′-OH by a base, followed by a nucleophilic attack of the activated 2′-O^- oxygen on the cleavable 3′-O-phosphobond. The outcome of this attack differs between the RNA classes: self-cleaving RNA, such as the hammerhead ribozymes, forms 2′,3′ cyclic phosphates, whereas group I, II and spliceosomal introns undergo splicing via a transesterification (see Chapter 5 Lührmann, see Figure 5.1 for the mechanism). In most cases, the base that activates the −OH group is a metal ion. Since these RNA molecules act as enzymes, they were named “ribozymes”; similarly to their protein relatives, ribozymes form specific three-dimensional (3-D) conformations that form solvent-protected active sites and undergo sterical changes during the reaction.

It has been demonstrated that RNA catalyzes peptide bond formation in the ribosome [15], and it has been proposed that RNA is responsible for catalysis in the human spliceosome [16], which in turn raises the question of the function of proteins

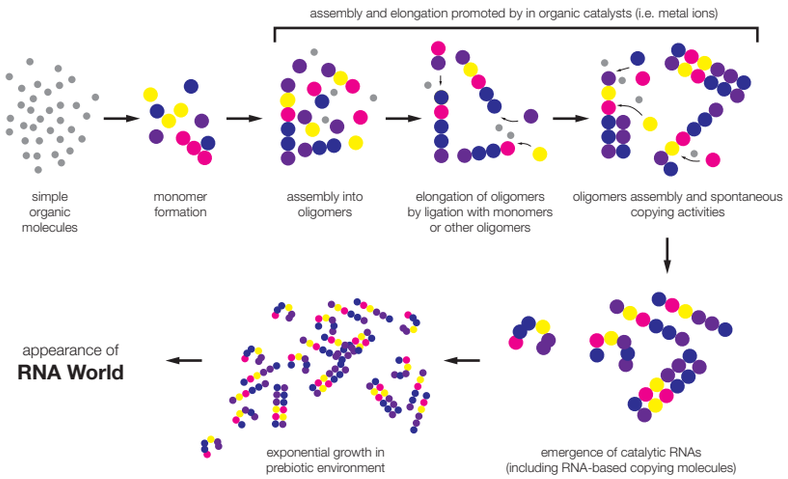

Fig. 1.2 The RNA World hypothesis. The genesis of a hypothetical RNA World includes a series of events, summarized schematically in this figure, that include the abiotic synthesis of RNA monomers from common inorganic/organic molecules and their assembly into oligomers (in the likely presence of metal catalysts) that would have had to serve as templates for their own spontaneous copying or replication. At this stage, a set of catalytic RNAs may have evolved (among them, RNA-based RNA copying molecules), and this would have sustained their exponential growth in the prebiotic environment.

in ribonucleoproteins (RNPs). The study of RNase P showed the importance of protein components associating with ribozymes. RNase P is an RNA–protein complex that cleaves tRNA precursors. In bacteria, the catalytic activity resides within the RNA [14], but in human mitochondria RNase P catalyzes the reaction without RNA, demonstrating that proteins can substitute for RNA functions [17]. The question then is, since similar biological functions can be performed by either RNA or protein complexes, why did evolution select RNPs such as ribosomes and spliceosomes for protein synthesis and pre-mRNA processing?

One possibility is that proteins facilitate the conformational changes of RNA that are necessary for catalysis. The spliceosome catalyzes the reaction between two structurally different substrates, which necessitates large spatial rearrangements during the reaction (see Chapter 5, Lührmann), which could be stabilized by additional RNA–protein interactions in the spliceosome. In fact, the spliceosome is an excellent example of an RNP machine, where the degree of interdependence between RNA and protein for catalytic function is such that it is justified to consider it a veritable RNP enzyme.

The discovery of the enzymatic activity of RNAs led to the concept of a primitive "RNA World" which could have existed before the appearance of modern proteins and DNA [18]. A schematic depiction of the RNA World hypothesis is shown in Figure 1.2. It is impossible to prove the existence of a pure RNA World in the prehistoric Earth, and RNA could have been coexisting with small peptides. However, the RNA World concept has been useful in analyzing the mechanism of RNA-based machines, such as the ribosome or spliceosome, as it points to RNA as the catalytic moiety [19]. One of the predictions of the RNA World hypothesis is that the core catalytic activity should be conserved in spliceosomes from different phyla. The comparison of spliceosomes between human (Chapter 5 Lührman), yeast (Chapter 6, Rymond), and plants (Chapter 7 Barta) shows which is the case, and suggests that they derived from a common precursor. Another echo of the RNA World is the fact that most of the human genome is transcribed into noncoding RNAs [20], which further suggests a larger regulatory role of RNA.

1.4 Common Classes of Eukaryotic RNA

Only about 1.2% of the human genome encodes proteins. The ENCODE project, which carefully analyzed gene expression in 1% of the human genome, showed that at least 93% of the human genome in this region is transcribed [20]. It is not clear however, whether other genomic regions are transcribed in a similarly active manner, as recent RNAseq data have suggested that most transcription is associated with

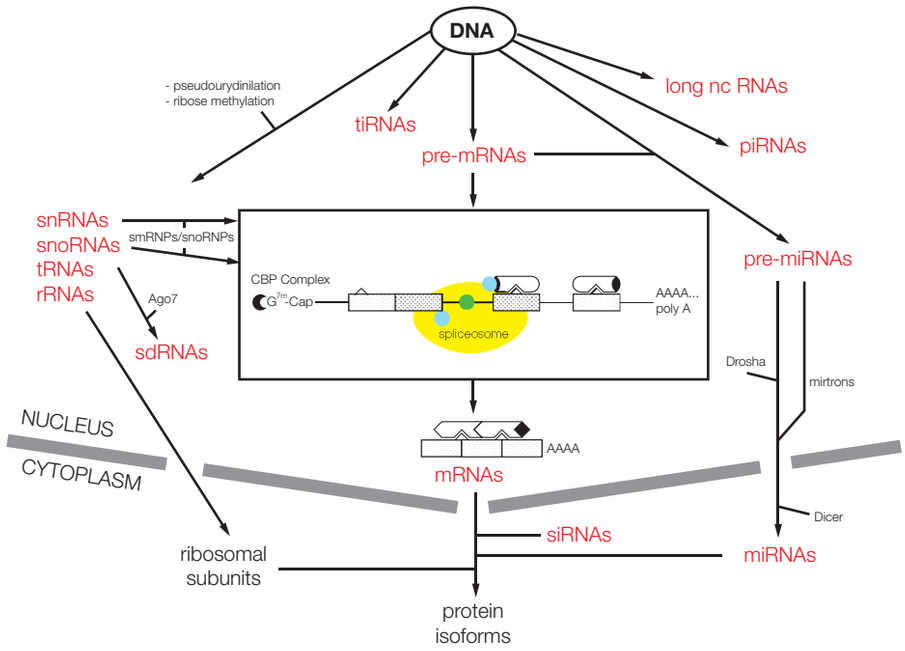

Fig. 1.3 Expression of RNA. A schematic view of the most common noncoding and coding RNA families currently described in the eukaryotic cell.

known genes [21]. Most of the RNA expression consists of short, nuclear, non-protein-coding RNAs (ncRNAs) (Figure 1.3), although it is not clear whether these RNAs simply represent noise or have functions. As a large fraction of the ncRNAs show cell type-specific expression and derive from evolutionary highly conserved promoter regions, it is likely that they represent a pool of sequences that can be recruited by evolution to regulate gene expression [22], possibly by yet unknown mechanisms.

The most abundant cellular RNAs are the rRNAs, which are the core of the ribosomes – the ribonucleoprotein particles in charge of translating the information encoded in mRNAs into proteins. The ribosomes in eukaryotes are formed by two subunits – the 60S and the 40S – named according to their sedimentation coefficients (see Chapter 13 Hartmuth for sedimentation analysis). These subunits contain the 28S/5S rRNA and 18S rRNA tightly associated with proteins. The amino acids are brought to the ribosome by tRNAs, whereby each tRNA is associated with an amino acid and recognizes the mRNA through a three-nucleotide sequence known as a "codon." Formation of the peptide bonds which connect the amino acids is performed by the RNA part of the ribosome, which acts as a ribozyme [15].

An additional class of relatively abundant small RNAs is formed by the small nucleolar RNAs (snoRNAs) [23]. As indicated by their name, these localize to the nucleolus and are mainly involved in rRNA maturation, although they also play important functions in protein translation, mRNA splicing, and genome stability. There are two classes of snoRNA (C/D and H/ACA box) that function as ribonucleoprotein (RNP) complexes to guide the enzymatic modification of target RNAs. Generally, the C/D box snoRNAs guide the methylation of target RNAs, while the H/ACA box snoRNAs guide pseudouridylation [24]. It has also been recently discovered that snoRNAs can be additionally processed to yield smaller molecules, termed sno-derived RNAs (sdRNAs), that are associated with Ago2 and may thus be linked to gene silencing and transcriptional repression processes [25]. One of the shorter sdRNAs has been demonstrated to regulate AS in neurons [26,27].

Another well-characterized family of RNAs are the small nuclear RNAs (snRNAs). Based on sequence homology and common protein factors, the snRNAs can be divided in two classes – the Sm and Lsm (like Sm) classes. The name Sm is derived from Stephanie Smith, a patient with lupus erythematosus whose blood contained antibodies against snRNA-associated proteins [28,29], and which were used for the purification of such proteins. The sequence of the U1 snRNA shows a complementarity to the 5′ splice site, which at an early stage led to the correct hypothesis that these

RNAs function in splicing [30]. The Sm class is composed of U1, U2, U4, $U4_{atac}$, U5, U7, U11, and U12, whereas U6 and $U6_{atac}$ are associated with the Lsm class of proteins. While U1, U2, U4, U5, and U6 are components of what is termed the "major spliceosome" (which splices introns with GU at the 5′ splice site and AG at the 3′ splice site), the U11, U12, $U4_{atac}$, and $U6_{atac}$ RNAs are components of the so-called "minor spliceosome" (which splices introns that have AT–AC at their 5′ and 3′ ends) [31]. After assembly with small nuclear ribonucleoproteins (snRNPs), all of the resulting snRNP particles form the core of the spliceosome (major or minor), and catalyze the removal of introns from pre-mRNA (see Chapter 5 Lührmann). The only exception to this is represented by U7 snRNP, which functions in histone pre-mRNA 3′-end processing (see Chapter 45 Schümperli).

Other classes of ncRNAs are micro RNAs (miRNAs) and short interference RNAs (siRNAs). The miRNAs are 21- to 23-nucleotide (nt) RNAs that regulate gene expression through binding to mRNAs via an imperfect complementarity. The siRNAs recognize perfect complementary RNAs, and induce their cleavage and subsequent degradation. Both of these RNA classes are discussed in Chapter 2 Meister and summarized in Table 2.1), while the application of siRNAs to knockdown genes is described in Chapter 15 Gabellini.

Piwi-interacting RNAs (piRNAs) are another class of ncRNAs, which are expressed only in the germline of flies, fish, and mammals; here, the name Piwi (P-element induced wimpy testis in *Drosophila*) refers to a class of RNA-binding proteins in *Drosophila*. These proteins were observed to interact with a novel class of longer-than-average miRNAs (26–31 nt), termed piRNAs. The expression of both piRNAs and Piwi proteins is restricted to the male germline.

Despite the rapid emergence of new ncRNA classes, their characterization remains in its infancy, and the majority of the ncRNAs do not fall into defined classes and have no function attributed [22]. It is likely, therefore, that new classes of ncRNAs with specialized functions will be discovered.

1.5 Alternative Pre-mRNA Splicing as a Central Element of Gene Expression

Messenger RNAs are the only RNAs that encode the information to make proteins, and have therefore been extremely well studied. The mRNAs are generated by processing from their pre-mRNAs precursor molecules. Pre-mRNAs are the first product of gene transcription performed by RNA polymerase II, and are processed while their synthesis is still ongoing (see Chapter 9 Neugebauer). In this case, there are sequences that are removed from the pre-mRNA, which are referred to as "introns" (for intragenic or intervening sequence), while the sequences that are joined and exported to the cytoplasm are referred to as "exons" (for expressed or exported sequence) [32]. Since the processing of the pre-mRNA begins during its synthesis, it is unlikely that an RNA corresponding to the whole DNA gene sequence exists. The splicing reaction, which forms the central step in the production of mRNAs, involves the recognition of exon boundaries by the spliceosomal machinery, the excision of the introns, and a concomitant ligation of the exons. Splicing can be either *constitutive* (when the exon in question always forms part of the mRNA) or *alternative* (when the specific exon can be excised from a proportion of the mRNAs) (see Chapters 5 Lührmann, 3 Hertel, and 6 Rymond). The mRNA molecule is further modified by 5′-end capping, in addition to cleavage and polyadenylation at the 3′ end.

Mature mammalian mRNAs can be divided into three regions: the protein-coding sequence; the 5′ untranslated region (5′UTR); and the 3′ untranslated region (3′UTR). The coding sequence stretches from the translation initiation codon (AUG) to the stop codon (UAA, UAG, or UGA).

Pre-mRNA splicing occurs in the nucleus, and is linked to other events in gene expression, which is important for its regulation. Pre-mRNA splicing occurs during

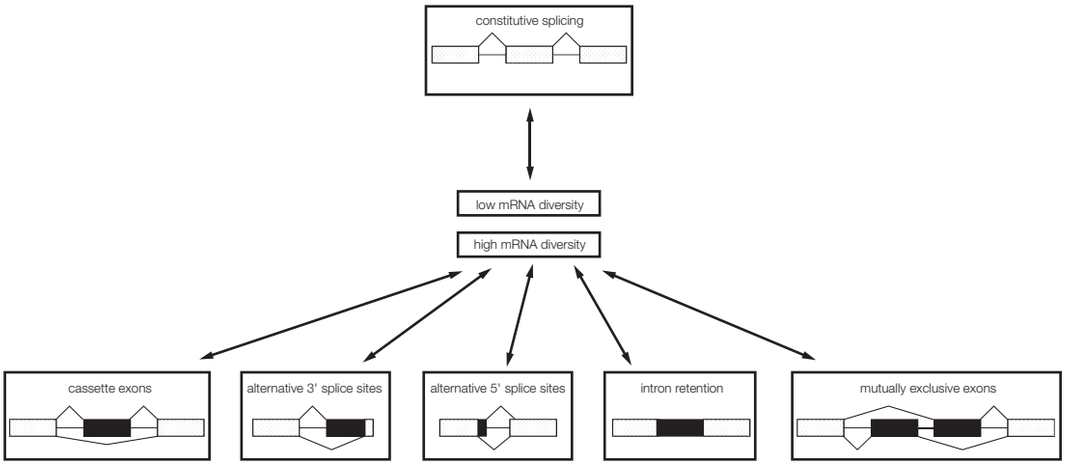

Fig. 1.4 Products of alternative splicing. A schematic representation of alternative splicing models. Constitutive exons are shown as gray boxes, introns as horizontal lines, and alternative exons as blue–red-shaded boxes.

transcription, and therefore the speed and protein composition of the polymerase II complex will impact on alternative exon usage (see Chapter 9 Neugebauer and Chapter 37 Kornblihtt). It is likely that the speed of the polymerase activity is influenced by the chromatin structure, which indicates that the location of an exon in a certain chromatin region influences its regulation (Chapter 37 Kornblihtt) [33] An early event in transcription is formation of the 5′ cap, which promotes spliceosome assembly [34]; similarly, the splicing of the last intron of a pre-mRNA is coupled to the polyadenylation [34]. Finally, the splicing reaction deposits "exon junction complexes" just upstream of the newly formed exon–exon junction; these are necessary for an efficient RNA export, and they also enhance translation [35]. In addition, splicing regulatory proteins – such as SF2/ASF – can stimulate translation [36] (some of these possibilities are shown schematically in Figure 1.4).

These examples, which are further discussed in Chapters 9 Neugebauer and 37 Kornblihtt, show that pre-mRNA splicing takes a central role in the production of pre-mRNA. Numerous crosstalks occur between the various stages of pre-mRNA processing [37] that allow the splicing reaction to integrate different cellular signals, in order to decide which part of the genetic information should be expressed.

1.6 Increasing Numbers of Human Diseases are Associated with "Wrong" Splice Site Selection

Given the importance of pre-mRNA splicing in gene expression, it is not surprising that defects in this process lead to human diseases [38–40]. With a few exceptions, such as PRPF31/U4-61k [41,42], PRP8 [43], PRPF3/U4/U6-90K [44,45] and PRPF4/U4/U6-60K [46], no mutations in components of the core spliceosome are known to cause a disease. Whilst this may suggest that defects in the general splicing machinery are generally incompatible with life, changes in AS can often be tolerated by an organism.

Mutations that cause a change in AS can be viewed as "experiments" that have been selected by evolution. Alu elements (short repetitive elements found only in primates) represent a good example of how the splicing machinery interacts with an evolving genome. The Alu elements represent about 10% of the total human genome sequence (for a review, see Ref. [47]), they contain potential splice sites, and they can evolve into exons [48]. It has been estimated that up to 5% of human alternative exons may be derived from Alu sequences [49], thus demonstrating how the splicing machinery can use new genetic information.

By far the largest number of currently known splicing diseases are caused by mutations in the pre-mRNA, which leads to aberrant exon usage (Chapter 10, Baralle). The analysis of these mutations has been highly informative for the mechanism of splice site regulation, and has provided insights into disease mechanisms that allowed rational therapies first to be devised. For example, the sequencing of disease-associated mRNAs led to the detection of numerous synonymous mutations. Initially, it was difficult to understand how these mutations could lead to a disease, as synonymous mutations did not alter the protein encoded by the mRNA. However, the disease mechanism became clear when it was realized that exons contain sequences that regulate their AS [50]. Synonymous mutations can, therefore, act by changing alternative exon usage. Based on this insight into the disease mechanism, it is possible to test therapeutic approaches. Currently, the most common genetic cause of death in children is spinal muscular atrophy (SMA). This deadly disease is caused by deletion of the *SMN1* gene which, unfortunately, cannot be substituted by the almost identical *SMN2* gene that is present in all patients. The difference between the two genes is a synonymous mutation that causes exon skipping in *SMN2*. Yet, the disease could be treated if the inclusion of this particular exon into the *SMN2* mRNA could be promoted. Towards this aim, the regulation of the exon has been investigated in great detail (see Chapters 18 and 19 Singh), and this has led not only to the development of therapeutic approaches currently undergoing clinical trials [51], but also to new experimental approaches to alter splice site selection (see Chapter 45 Schümperli).

The majority of currently known mutations that have a clear effect on alternative exon usage have become apparent in only a few patients. One such example – the Hutchinson–Gilford progeria syndrome (HGPS) – highlights the benefits of studying these rare diseases. HGPS is genetic disorder that is characterized phenotypically by many features of premature aging, with patients typically dying at the age of 13 years. Mutations causing HGPS have been identified in the nuclear lamin A/C (*LMNA*) gene, and three out of 14 mutations affecting lamin A/C have been reported to specifically alter lamin A splicing (see Chapter 36, Tazi). This particular splicing event causes an aberrant farnesylation of the resulting protein, which causes the disease. A screening effort has identified a previously tested farnesyltransferase inhibitor that could be used to treat the disease [52]. HGPS, which has an incidence of 1 per 4–8 million live births, may prove to be highly informative for the normal aging process, as the mutant splice variant accumulates in the skin of aging individuals [53]. This example shows that screening for substances that change alternative splice site selection (as shown in Chapter 46 Stoilov) may, in time, have a huge impact on human health.

The development of new experimental techniques was a strong driving force in research on AS. The current experimental protocols collected in this book have been built on more than 30 years of experimental experience. For example, the identification of antisera against splicing components [28] allowed their purification such that, today, is it possible to purify different stages of spliceosomes (see Chapters 13 and 31 Lührmann) and to generate specific antisera (as shown in Chapter 43 Fishman). The identification of alternative exons is based on comparing cDNA and genomic sequences. The completion of the human genome showed, for the first time, the unexpected high usage of AS on an organismal level [54]. The sequencing techniques have been improved (as shown in Chapters 25 Guigo and 51 Zhang), and it is now possible to rapidly sequence the genome of an individual [55]. The knowledge of individual genome sequences, and their analysis by genome-wide DNA arrays (see Chapter 24 de la Grange), marks the beginning of personalized medicine. This will allow the analysis of potential changes in the AS of individuals by PCR (Chapter 21 Smith), cell-based (Chapters 35–37, Stamm, Tazi, Kornblihtt), and *in vitro* assays (Chapter 30 Krainer). Knowledge of disease-causing mechanisms may lead to a better genetic counseling (see Chapter 11 Baralle), and might also pave the way to the development of therapies (as discussed in Chapters 45–48 Schümperli, Stoilov, Annemieke, Patel).

Acknowledgments

This work was supported by Telethon, and by the EC grant EURASNET.

References

1 Berget, S.M., Moore, C., and Sharp, P.A. (1977) Spliced segments at the 5′ terminus of adenovirus 2 late mRNA. *Proc. Natl Acad. Sci. USA*, **74**, 3171–3175.
2 Chow, L.T., Roberts, J.M., Lewis, J.B., and Broker, T.R. (1977) A map of cytoplasmic RNA transcripts from lytic adenovirus type 2, determined by electron microscopy of RNA:DNA hybrids. *Cell*, **11**, 819–836.
3 Sharp, P.A. (1994) Split genes and RNA splicing. *Cell*, **77**, 805–815.
4 Pan, Q., Shai, O., Lee, L.J., Frey, B.J., and Blencowe, B.J. (2008) Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat. Genet.*, **40**, 1413–1415.
5 Patel, N.A., Chalfant, C.E., Watson, J.E., Wyatt, J.R., Dean, N.M., Eichler, D.C., and Cooper, D.R. (2001) Insulin regulates alternative splicing of protein kinase C beta II through a phosphatidylinositol 3-kinase-dependent pathway involving the nuclear serine/arginine-rich splicing factor, SRp40, in skeletal muscle cells. *J. Biol. Chem.*, **276**, 22648–22654.
6 Weg-Remers, S., Ponta, H., Herrlich, P., and Konig, H. (2001) Regulation of alternative pre-mRNA splicing by the ERK MAP-kinase pathway. *EMBO J.*, **20**, 4194–4203.
7 Celotto, A.M. and Graveley, B.R. (2001) Alternative splicing of the *Drosophila* Dscam pre-mRNA is both temporally and spatially regulated. *Genetics*, **159**, 599–608.
8 Rozenski, J., Crain, P.F., and McCloskey, J.A. (1999) The RNA Modification Database: 1999 update. *Nucleic Acids Res.*, **27**, 196–197.
9 Leontis, N.B. and Westhof, E. (2001) Geometric nomenclature and classification of RNA base pairs. *RNA*, **7**, 499–512.
10 Brion, P. and Westhof, E. (1997) Hierarchy and dynamics of RNA folding. *Annu. Rev. Biophys. Biomol. Struct.*, **26**, 113–137.
11 Noller, H.F. (2005) RNA structure: reading the ribosome. *Science*, **309**, 1508–1514.
12 Doudna, J.A. and Cech, T.R. (2002) The chemical repertoire of natural ribozymes. *Nature*, **418**, 222–228.
13 Kruger, K., Grabowski, P.J., Zaug, A.J., Sands, J., Gottschling, D.E., and Cech, T.R. (1982) Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of *Tetrahymena*. *Cell*, **31**, 147–157.
14 Guerrier-Takada, C., Gardiner, K., Marsh, T., Pace, N., and Altman, S. (1983) The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. *Cell*, **35**, 849–857.
15 Steitz, T.A. (2008) A structural understanding of the dynamic ribosome machine. *Nat. Rev. Mol. Cell Biol.*, **9**, 242–253.
16 Valadkhan, S., Mohammadi, A., Wachtel, C., and Manley, J.L. (2007) Protein-free spliceosomal snRNAs catalyze a reaction that resembles the first step of splicing. *RNA*, **13**, 2300–2311.
17 Holzmann, J., Frank, P., Loffler, E., Bennett, K.L., Gerner, C., and Rossmanith, W. (2008) RNase P without RNA: identification and functional reconstitution of the human mitochondrial tRNA processing enzyme. *Cell*, **135**, 462–474.
18 Orgel, L.E. (1998) The origin of life – a review of facts and speculations. *Trends Biochem. Sci.*, **23**, 491–495.
19 Cech, T.R. (2009) Crawling out of the RNA world. *Cell*, **136**, 599–602.
20 Birney, E., Stamatoyannopoulos, J.A., Dutta, A., Guigo, R., Gingeras, T.R., Margulies, E.H., Weng, Z., Snyder, M., Dermitzakis, E.T., Thurman, R.E. *et al.* (2007) Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature*, **447**, 799–816.
21 van Bakel, H., Nislow, C., Blencowe, B.J., and Hughes, T.R. (2010) Most 'dark matter' transcripts are associated with known genes. *PLoS Biol.*, **8** (5), e1000371.
22 Amaral, P.P., Dinger, M.E., Mercer, T.R., and Mattick, J.S. (2008) The eukaryotic genome as an RNA machine. *Science*, **319**, 1787–1789.
23 Dieci, G., Preti, M., and Montanini, B. (2009) Eukaryotic snoRNAs: a paradigm for gene expression flexibility. *Genomics*, **94**, 83–88.
24 Matera, A.G., Terns, R.M., and Terns, M.P. (2007) Non-coding RNAs: lessons from the small nuclear and small nucleolar RNAs. *Nat. Rev. Mol. Cell Biol.*, **8**, 209–220.
25 Taft, R.J., Glazov, E.A., Lassmann, T., Hayashizaki, Y., Carninci, P., and Mattick, J.S. (2009) Small RNAs derived from snoRNAs. *RNA*, **15**, 1233–1240.
26 Kishore, S., Khanna, A., Zhang, Z., Hui, J., Balwierz, P., Stefan, M., Beach, C., Nicholls, R.D., Zavolan, M., and Stamm, S. (2010) The snoRNA MBII-52 (SNORD 115) is processed into smaller RNAs and regulates alternative splicing. *Hum. Mol. Genet.*, **19**, 1153–1164.
27 Kishore, S. and Stamm, S. (2006) The snoRNA HBII-52 regulates alternative splicing of the serotonin receptor 2C. *Science*, **311**, 230–232.
28 Lerner, M.R. and Steitz, J.A. (1979) Antibodies to small nuclear RNAs complexed with proteins are produced by patients with systemic lupus erythematosus. *Proc. Natl Acad. Sci. USA*, **76**, 5495–5499.
29 Reeves, W.H., Narain, S., and Satoh, M. (2003) Henry Kunkel, Stephanie Smith, clinical immunology, and split genes. *Lupus*, **12**, 213–217.
30 Lerner, M.R., Boyle, J.A., Mount, S.M., Wolin, S.L., and Steitz, J.A. (1980) Are snRNPs involved in splicing? *Nature*, **283**, 220–224.
31 Will, C.L. and Lührmann, R. (2005) Splicing of a rare class of introns by the U12-dependent spliceosome. *Biol. Chem.*, **386**, 713–724.
32 Gilbert, W. (1978) Why genes in pieces? *Nature*, **271**, 501.
33 Kornblihtt, A.R. (2007) Coupling transcription and alternative splicing. *Adv. Exp. Med. Biol.*, **623**, 175–189.
34 Schwer, B. and Shuman, S. (1996) Conditional inactivation of mRNA capping enzyme affects yeast pre-mRNA splicing *in vivo*. *RNA*, **2**, 574–583.
35 Luo, M.L., Zhou, Z., Magni, K., Christoforides, C., Rappsilber, J., Mann, M., and Reed, R. (2001) Pre-mRNA splicing and mRNA export linked by direct interactions between UAP56 and Aly. *Nature*, **413**, 644–647.
36 Sanford, J.R., Gray, N.K., Beckmann, K., and Caceres, J.F. (2004) A novel role for shuttling SR proteins in mRNA translation. *Genes Dev.*, **18**, 755–768.
37 Moore, M.J. (2005) From birth to death: the complex lives of eukaryotic mRNAs. *Science*, **309**, 1514–1518.
38 Cooper, T.A., Wan, L., and Dreyfuss, G. (2009) RNA and disease. *Cell*, **136**, 777–793.
39 Jeanteur, P. (2006) *Alternative Splicing and Disease*, Springer.
40 Tazi, J., Bakkour, N., and Stamm, S. (2009) Alternative splicing and disease. *Biochim. Biophys. Acta*, **1792**, 14–26.
41 Vithana, E.N., Abu-Safieh, L., Allen, M.J., Carey, A., Papaioannou, M., Chakarova, C., Al-Maghtheh, M., Ebenezer, N.D., Willis, C., Moore, A.T. *et al.* (2001) A human homolog of yeast pre-mRNA splicing gene, PRP31, underlies autosomal dominant retinitis pigmentosa on chromosome 19q13.4 (RP11). *Mol. Cell*, **8**, 375–381.
42 Wilkie, S.E., Vaclavik, V., Wu, H., Bujakowska, K., Chakarova, C.F., Bhattacharya, S.S., Warren, M.J., and Hunt, D.M. (2008) Disease mechanism for

retinitis pigmentosa (RP11) caused by missense mutations in the splicing factor gene PRPF31. *Mol. Vis.*, **14**, 683–690.

43 Boon, K.L., Grainger, R.J., Ehsani, P., Barrass, J.D., Auchynnikava, T., Inglehearn, C.F., and Beggs, J.D. (2007) prp8 mutations that cause human retinitis pigmentosa lead to a U5 snRNP maturation defect in yeast. *Nat. Struct. Mol. Biol.*, **14**, 1077–1083.

44 Chakarova, C.F., Hims, M.M., Bolz, H., Abu-Safieh, L., Patel, R.J., Papaioannou, M.G., Inglehearn, C.F., Keen, T.J., Willis, C., Moore, A.T. *et al.* (2002) Mutations in HPRP3, a third member of pre-mRNA splicing factor genes, implicated in autosomal dominant retinitis pigmentosa. *Hum. Mol. Genet.*, **11**, 87–92.

45 Gonzalez-Santos, J.M., Cao, H., Duan, R.C., and Hu, J. (2008) Mutation in the splicing factor Hprp3p linked to retinitis pigmentosa impairs interactions within the U4/U6 snRNP complex. *Hum. Mol. Genet.*, **17**, 225–239.

46 Schmidt-Kastner, R., Yamamoto, H., Hamasaki, D., Parel, J.M., Schmitz, C., Dorey, C.K., Blanks, J.C., and Preising, M.N. (2008) Hypoxia-regulated components of the U4/U6. U5 tri-small nuclear riboprotein complex: possible role in autosomal dominant retinitis pigmentosa. *Mol. Vis.*, **14**, 125–135.

47 Hasler, J., Samuelsson, T., and Strub, K. (2007) Useful 'junk': Alu RNAs in the human transcriptome. *Cell Mol. Life Sci.*, **64**, 1793–1800.

48 Lev-Maor, G., Sorek, R., Shomron, N., and Ast, G. (2003) The birth of an alternatively spliced exon: 3′ splice-site selection in Alu exons. *Science*, **300**, 1288–1291.

49 Sorek, R., Ast, G., and Graur, D. (2002) Alu-containing exons are alternatively spliced. *Genome Res.*, **12**, 1060–1067.

50 Cooper, T.A. and Mattox, W. (1997) The regulation of splice-site selection, and its role in human disease. *Am. J. Hum. Genet.*, **61**, 259–266.

51 Darras, B.T. and Kang, P.B. (2007) Clinical trials in spinal muscular atrophy. *Curr. Opin. Pediatr.*, **19**, 675–679.

52 Capell, B.C., Olive, M., Erdos, M.R., Cao, K., Faddah, D.A., Tavarez, U.L., Conneely, K.N., Qu, X., San, H., Ganesh, S.K. *et al.* (2008) A farnesyltransferase inhibitor prevents both the onset and late progression of cardiovascular disease in a progeria mouse model. *Proc. Natl Acad. Sci. USA*, **105**, 15902–15907.

53 McClintock, D., Ratner, D., Lokuge, M., Owens, D.M., Gordon, L.B., Collins, F.S., and Djabali, K. (2007) The mutant form of lamin A that causes Hutchinson-Gilford progeria is a biomarker of cellular aging in human skin. *PLoS One*, **2**, e1269.

54 Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W. *et al.* (2001) Initial sequencing and analysis of the human genome. *Nature*, **409**, 860–921.

55 Wheeler, D.A., Srinivasan, M., Egholm, M., Shen, Y., Chen, L., McGuire, A., He, W., Chen, Y.J., Makhijani, V., Roth, G.T. *et al.* (2008) The complete genome of an individual by massively parallel DNA sequencing. *Nature*, **452**, 872–876.