

# 8 The Expanding Repertoire of RNA

The biochemical analyses of RNA in the 1960s detected many short RNAs in the nucleus and cytoplasm of eukaryotic cells<sup>a</sup> using new techniques of radioactive labeling, differential sedimentation and gel electrophoresis, and better procedures for isolating intact RNAs with detergents and chaotropic agents<sup>1</sup> to overcome degradation by RNases.

Initially there were concerns that the small RNAs are by-products of the biogenesis or degradation of larger RNAs. On the other hand, contrary to the generalization that nuclear RNAs are transient, destined for processing and export to the cytoplasm, the few groups studying these newly identified low molecular weight RNAs reported that they are highly expressed and account for ~20% of the nuclear RNA in mammalian cells.<sup>2</sup> They were also found to differ in size and sequence composition from tRNAs and rRNAs, and to be metabolically stable.<sup>3</sup>

At least 10 discrete RNA species were identified, some of which contained methylated nucleotides, localized in specific subnuclear fractions (the nucleoplasm, chromatin or the nucleolus), with others in the cytoplasm.<sup>4–9</sup> Many of those RNAs in the nucleus were uridine-rich, leading to their designation as ‘U RNAs’, numbered in the order of their discovery as U1, U2, U3 and so on.<sup>2,6</sup> Robert Weinberg and Sheldon Penman named them “small nuclear RNAs” (snRNAs).<sup>6,10</sup>

It was also found that RNA polymerase III is responsible for the transcription of many of these small RNAs, not RNA polymerase II, which transcribes mRNAs (and long non-protein-coding RNAs, Chapter 13), indicating that different RNA polymerases synthesize different classes of RNA.<sup>11–13</sup> RNA polymerase III products also include RNAs originating from repetitive sequences, only later characterized, such as those transcribed in human cells from Alu elements.<sup>13,14</sup>

The characterization of snRNAs in the following decades revealed that they had ‘housekeeping’ functions in the modification and maturation of rRNAs, tRNAs and mRNAs, as well as other functions in gene regulation and cellular processes such as protein

export. These years also saw RNAs encroach on the traditional domain of proteins, catalysis, which in turn led to a plausible explanation of the molecular origin of genetic information, with RNA at its core.

## SPLICEOSOMAL RNAs

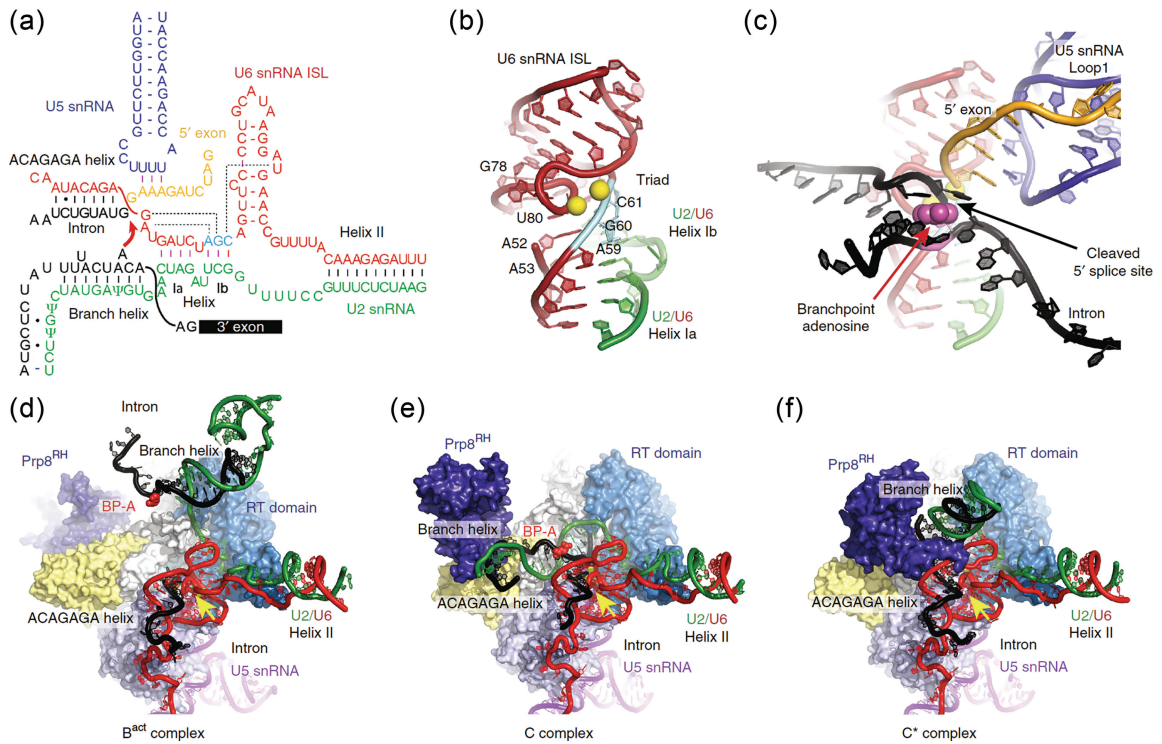
Although their roles were unknown, snRNAs were too small to function as mRNAs and, being mainly nuclear, did not seem to be directly involved in protein synthesis. On the other hand, some snRNAs contained sequences complementary to hnRNAs. This led Michael Lerner and Joan Steitz, and independently John Rogers and Randolph Wall, to propose in 1980 that these RNAs play a role in RNA splicing,<sup>15–17</sup> based on earlier work on sense-antisense interactions between mRNA and rRNAs sequences in translation initiation.<sup>18,19</sup> In particular, the complementarity of the U1 snRNA sequence to both the 5′ and 3′ splice site sequences of hnRNAs led to the hypothesis that splice sites are recognized and aligned through RNA–RNA interactions between the splice sites and U1 snRNA.<sup>15–17</sup>

The characterization of the functions of these RNAs in the 1980s was aided by the fortuitous discovery that antibodies in the serum of individuals suffering autoimmune disorders, such as lupus erythematosus, precipitated ribonucleoprotein complexes (RNPs) containing snRNAs.<sup>15,20–23</sup> Several of the snRNAs interacted with a common antigen, the Sm (Smith) antigen,<sup>b</sup> named after the first lupus patient in whom such antibodies were detected.<sup>20,23</sup> Other autoantigens were associated with other RNP complexes.<sup>25</sup>

These antibodies were used not only to purify the complexes but also to block the function of the corresponding snRNPs *in vitro*, which showed that splicing of pre-mRNA is inhibited by targeting the U1 RNP and therefore that snRNAs are required.<sup>26,27</sup> Chemical cross-linking confirmed that U1 and U2 RNAs do, in fact, base pair to hnRNAs in the nucleus,<sup>28,29</sup> and genetic complementation experiments confirmed that U1, U2, U4, U5 and U6

<sup>a</sup> Also in bacteria (Chapter 9).

<sup>b</sup> Sm proteins participate in a wide variety of RNA transactions and predate the Archaea-Eukarya split.<sup>24</sup>



**FIGURE 8.1** The mechanism of splicing and the complexity of the RNA interactions and structures (a) The RNA interaction network before the first trans-esterification reaction. The dotted lines indicate the triplex interactions at the catalytic core. (b) Three-dimensional structure of the active site RNA in the C complex. Magnesium ions are represented by two yellow spheres located between the backbone of the catalytic triad and the highly twisted backbone at the bulge in the internal stem loop. (c) Structure of the 5' exon and branched intron bound to the active site (overlaid on the structure in b). (d–f) Interaction of the catalytic core of the spliceosome and movement of the branch helix, RNAs are color-coded as in a. The yellow arrows indicate the active site metals. (Reproduced from Fica and Nagai<sup>31</sup> with permission from Springer Nature.)

snRNAs function in splicing.<sup>25,30</sup> Characterization of the process revealed that large RNPs, which came to be known as ‘spliceosomes’, incorporate these snRNAs, wherein they interact with each other and with target pre-mRNAs to guide the process of splicing (Figure 8.1).<sup>25,30–33</sup>

Much later, it was found that U1 and U4 snRNAs also regulate transcriptional initiation, transcript structure and chromatin architecture,<sup>34–40</sup> and that U2 snRNA is required for RNA polymerase II pausing,<sup>41</sup> coupling transcription to splicing, which are intertwined processes (Chapters 14 and 16). It was also discovered that there is a minor class of spliceosome, which contains specific snRNAs (U11, U12, U4atac and U6atac) equivalent to but distinct from their counterparts in the major U2-type spliceosome (U1, U2, U4 and U6), and which recognizes a rare class of introns

initially referred to as AT–AC introns, now called U12-type introns.<sup>42–45</sup> The minor spliceosome is required for development and may have particular functions in the brain.<sup>45–48</sup>

### SMALL NUCLEOLAR RNAs

Other small RNAs had other functions. The highly conserved U3 RNA was found to associate with 28S rRNA<sup>8,49</sup> and to be localized in the nucleolus,<sup>11,50</sup> where ribosome biogenesis occurs, which led Jean-Pierre Bachelier to suggest in 1983 that U3 and other “small nucleolar RNAs” (snoRNAs)<sup>c</sup> participate in this process.<sup>50</sup> Subsequently it was also found

<sup>c</sup> Although they were originally named based on their nucleolar localization (in contrast to snRNAs), snoRNAs are also found beyond the nucleolus<sup>51</sup> and are secreted from cells.<sup>52</sup>

that autoimmune antibodies recognizing a nucleolar protein, fibrillarin,<sup>53</sup> would co-precipitate not only U3 but also the less abundant U8 and U13 RNAs.<sup>54</sup> It was later found also to bind U16.<sup>55</sup>

Similar to the observations that led to the elucidation of the roles of snRNAs in pre-mRNA splicing, snoRNAs were found to have short sequence motifs complementary to rRNA sequences, which indicated that snoRNAs were involved via base pairing in rRNA processing, modification or other aspects of ribosome biogenesis.<sup>50</sup>

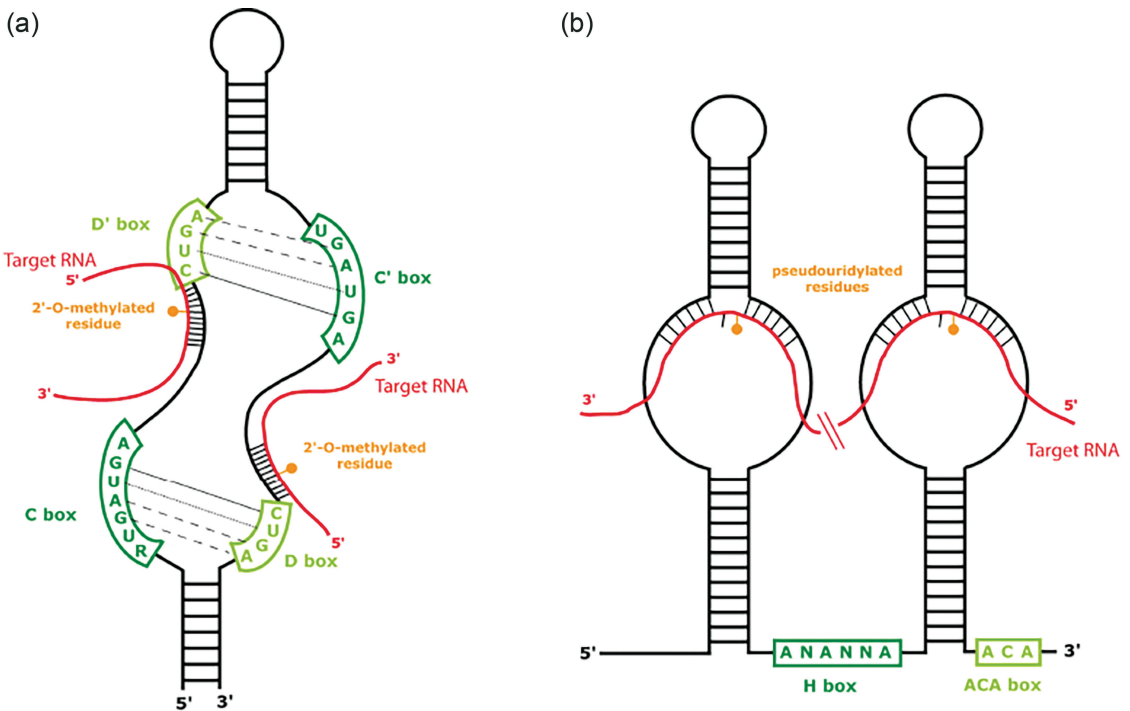
Nevertheless, it was only in 1990 that U3 was shown to be essential for rRNA processing,<sup>56</sup> and in 1996 that snoRNA U24 directs site-specific methylation of rRNAs<sup>57</sup> with subsequent studies showing that other snoRNAs perform similar functions via base pairing with target sequences adjacent to modification sites.<sup>51,58</sup> Thus, although first identified in the 1960s, it was three decades before snoRNAs were defined as a new “class of RNAs” with demonstrated functions.<sup>51,59</sup>

SnoRNAs are ~60–300nt in length and are classified into two families (based on typical sequence

motifs and structural features) that guide enzyme complexes to perform 2'-O-ribose methylation ('C/D box' RNAs) or pseudouridylation ('H/ACA box' RNAs) respectively of target nucleotides (Figure 8.2), not only in rRNAs and tRNAs but also in snRNAs.<sup>60</sup> Homologs of C/D box and H/ACA box snoRNAs occur in archaea, where they also guide modifications of tRNAs, indicating that they first evolved over 3 billion years ago.<sup>61–63</sup>

There are also many snoRNAs that show tissue-specific expression and whose targets are unknown, described as “orphan” snoRNAs.<sup>60</sup> Some were later found also to be involved in RNA processing, including the C/D box snoRNAs U8, U14, and U22, as well as H/ACA box snoRNAs snR10, snR30, E2 and E3, which direct site-specific cleavage of pre-rRNAs.<sup>65</sup> Yet, other snoRNAs were shown to regulate alternative splicing by base pair recognition<sup>66</sup> and to guide other modifications such as acetylation of specific cytosine residues in 18S rRNA.<sup>67</sup>

More was to come. In 2001, Beáta Jády and Tamás Kiss identified a sno-like RNA, U85, containing



**FIGURE 8.2** Schematic representation of C/D (a) and H/ACA (b) box snoRNAs. Consensus box sequences are highlighted in green. The conserved structures within these snoRNAs guide effector protein complexes that catalyze 2'-O-methylation or pseudouridylation respectively. (Reproduced from Abel and Rederstorff<sup>64</sup> with permission from Elsevier.)

both H/ACA and C/D box motifs, which is localized in ‘Cajal bodies’ (a subnuclear domain associated with nucleoli, discovered by Santiago Ramón y Cajal in 1903<sup>68</sup>) and guides 2'-O-ribose methylation and pseudouridylation of the U5 spliceosomal snRNA.<sup>69</sup> The snRNA U7 is also localized in Cajal bodies and participates in histone pre-mRNA 3' end formation.<sup>70,71</sup> Other small Cajal body-specific ('sca') RNAs have since been identified, some with a composite structure similar to U85, while others have only H/ACA box or C/D box domains, which guide modifications of spliceosomal snRNAs<sup>72</sup> and tRNAs, the latter as part of a stress response.<sup>73,74</sup> The RNA component of the human telomerase complex also contains a characteristic H/ACA box scaRNA-like structure and also localizes in Cajal bodies.<sup>75–77</sup> Primate-specific H/ACA snoRNA-like RNAs, called AluACA RNAs, are derived from intronic Alu-repeat RNAs<sup>78</sup> and new functions of snoRNAs continue to be discovered, such as the maintenance of chromatin accessibility.<sup>79</sup>

There are over 700 known snoRNA and snoRNA-like RNAs encoded in the human genome.<sup>80</sup> Most snoRNAs are produced by processing of intronic RNAs excised from host transcripts,<sup>81</sup> commonly those of genes encoding proteins involved in translation or ribosome biogenesis, including ribosomal proteins, translation factors and nucleolar proteins such as fibrillarin,<sup>60,82</sup> the first evidence of parallel genetic output. Many other snoRNAs are derived from the introns of transcripts that do not encode proteins,<sup>83–88</sup> some involving species-specific alternative splicing,<sup>89</sup> and whose primary function in many cases is uncertain, although others, like *Gas5*,<sup>90</sup> have demonstrated roles as long regulatory RNAs (Chapters 9 and 13). Some snoRNAs are expressed exclusively in the brain.<sup>51,91–93</sup>

The complexity of the relationships between small RNAs is illustrated by the later discovery that snoRNAs, from yeast to humans, are processed to produce three subspecies, one of which functions as a microRNA in the RNA interference pathway<sup>64,94–96</sup> (Chapter 12). The complexity of the networks, and an indication of how much is yet to be understood about them, is highlighted, for instance, by the observation that a human-specific snoRNA is attached to the end of a longer non-coding RNA that regulates rRNA biogenesis and nucleolar structure (via phase separation, Chapter 16).<sup>97–99</sup> In addition, many snoRNAs accumulate in the form of stable lariats instead of fully processed snoRNP particles, with as yet unknown functions.<sup>100</sup>

Aberrant expression of snoRNAs has been linked to human disease. There is a large cluster of C/D box snoRNAs in a parentally imprinted gene that is normally expressed in the brain from the maternally derived allele, perturbations of which are associated with Angelman and Prader-Willi syndromes.<sup>88,91,101,102</sup> One of the snoRNAs in this region, HBII-52, contains an 18 nucleotide sequence that is complementary to an exon in the serotonin receptor 5-HT(2C) mRNA and mediates its alternative splicing.<sup>103</sup>

## OTHER SMALL GUIDE, SCAFFOLDING AND REGULATORY RNAs

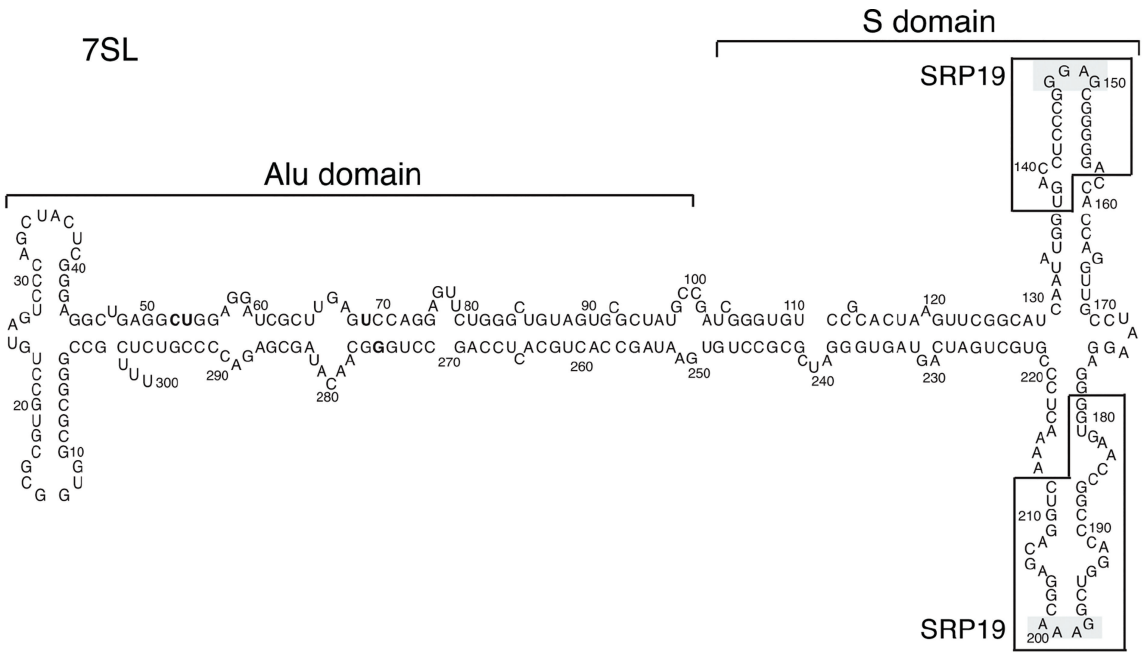
7SL and 7SK RNA species, their names reflecting their sedimentation coefficient, were identified by Penman in 1976.<sup>104</sup> These RNAs were initially thought to have a viral origin<sup>3</sup> but were shown to be present in uninfected cells.<sup>104,105</sup>

7SL RNA is ~300nt in length. It is ubiquitous in eukaryotic cells and was found, accidentally,<sup>d</sup> to be an essential component of the protein export ‘signal recognition particle’ (SRP),<sup>106</sup> characterized in the 1970s and 1980s by Günter Blobel and colleagues.<sup>107</sup> The SRP associates with the ribosome and targets nascent proteins to the endoplasmic reticulum via an N-terminal ‘signal’ or ‘leader’ sequence for membrane insertion or secretion into the extracellular milieu,<sup>106,108,109</sup> with 7SL RNA acting as a scaffold upon which the six proteins of the SRP assemble.<sup>109,110</sup> Similar RNAs were later shown to be involved in protein export in bacteria and archaea.<sup>111,112</sup>

7SL RNA was subsequently found to be required for the selective packaging of the RNA/DNA modifying enzymes APOBEC3G and 3F into retroviral particles<sup>113–117</sup> and to repress the translation of the tumor suppressor TP53.<sup>118</sup> It is also a precursor of the Alu elements in the human genome (Figure 8.3).<sup>119–121</sup>

7SK RNA is highly expressed in vertebrates and, like most RNAs, has a complex structure.<sup>122</sup> Early evidence indicated that 7SK regulates transcription and transcription termination in a tissue- and species-specific manner,<sup>123,124</sup> but mechanistic insights would not emerge until the early 2000s.

<sup>d</sup> Blobel and colleagues were expecting only proteins to be components of the SRP. The presence of RNA was revealed by a strong UV absorbance signal at a wavelength typical of nucleic acids in the SRP preparation, detected as a result of an incorrect setting on the detector.<sup>106</sup>



**FIGURE 8.3** The structure of 7SL RNA, showing the part coopted into Alu transposable elements in primates. SRP19 binding sites are shown. (Reproduced with minor modifications from Itano et al.<sup>117</sup> with permission from John Wiley and Sons.)

These studies showed that 7SK RNA acts as a trans-acting negative regulator, uncovered serendipitously in biochemical assays to identify factors that regulate RNA polymerase II, similar to the “general transcription factor” role identified for the small 6S RNA in bacteria<sup>125</sup> (Chapter 9), but having additional roles in gene expression regulation in animals<sup>126–134</sup> (Chapter 13). 7SK is also present in invertebrates, and, despite the fact that it is little primary sequence similarity, its identification was possible due to its conserved secondary structural motifs and domains.<sup>135,136</sup>

In one of the earliest demonstrated regulatory RNA roles, in 1980, Hugh Pelham, Robert Roeder and colleagues discovered that the transcription factor TFIIIA<sup>e</sup> not only binds the 5S rRNA gene<sup>f</sup>

but also its transcript, which results in a feedback loop that titrates the transcription factor away from the gene, inhibiting further transcription and stabilizing the transcript until required for ribosome assembly.<sup>140,141</sup>

Y RNAs were identified in 1981 as components, like snRNAs, of autoantigens in systemic lupus patients.<sup>21,142</sup> There are four distinct and highly conserved Y RNAs (in humans ranging from ~80 to ~110nt),<sup>143</sup> which are structural components of the Ro autoantigen.<sup>144–146</sup> The Ro protein, lack of which causes a lupus-like syndrome in mice, appears to prevent autoimmunity by recognizing misfolded RNAs, with Y RNAs regulating the process.<sup>147–149</sup> Y RNAs also occur in bacteria where they associate with orthologs of mammalian Ro and are involved in rRNA maturation and stress responses,<sup>150–152</sup> with a modular structure that includes a domain that mimics tRNAs,<sup>153</sup> indicating an ancient function in cellular RNA biology.

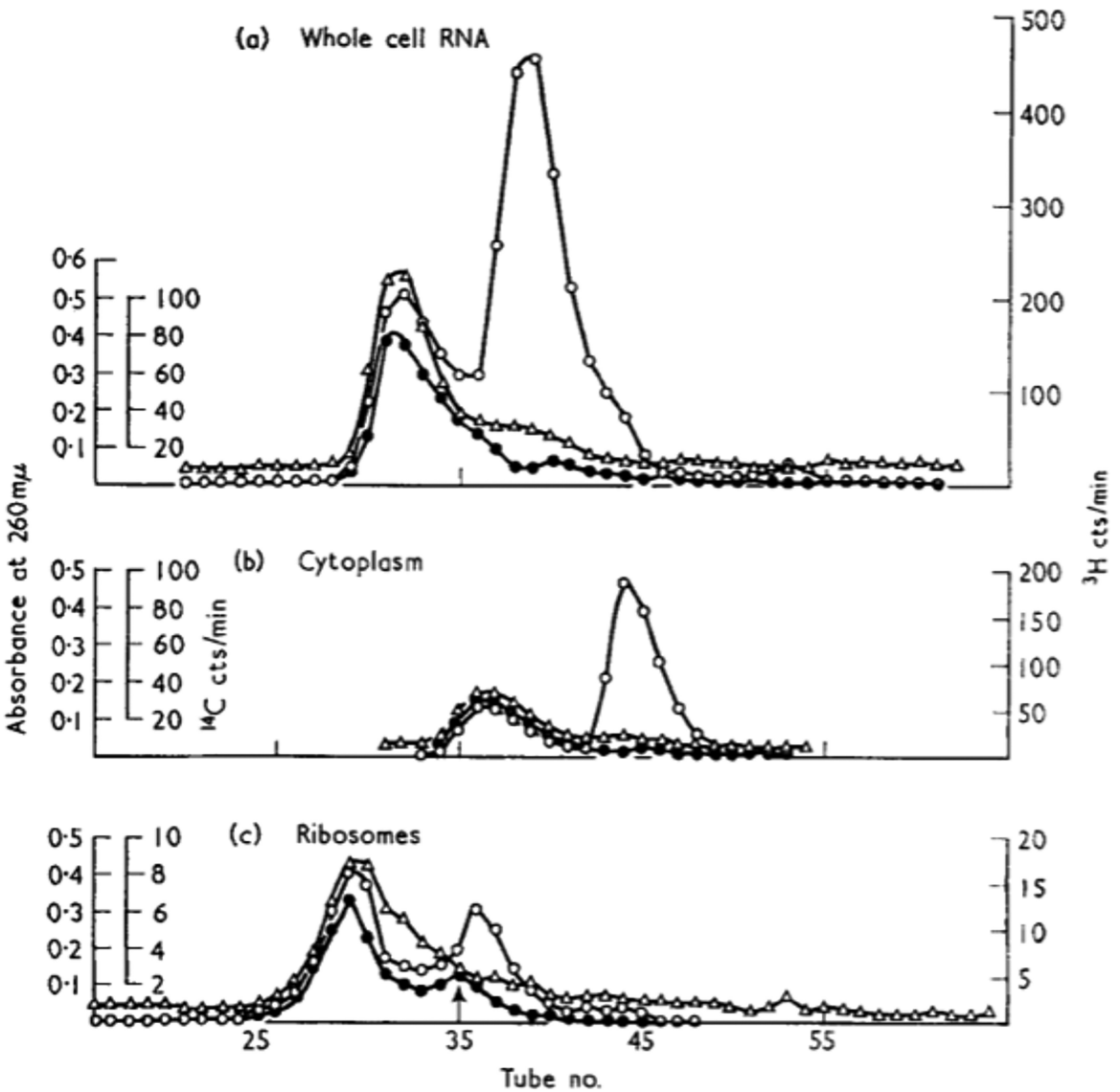
Vault RNAs, short 80–150nt RNAs transcribed by RNA polymerase III, so named because of their presence in large ovoid ribonucleoprotein particles in the cytoplasm of eukaryotic cells that resemble the arches of cathedral vaults, were discovered in

<sup>e</sup> TFIIIA was shown by Aaron Klug and colleagues to interact with RNA and DNA via repetitive domains stabilized by zinc, called ‘zinc fingers’.<sup>137</sup> Zinc finger transcription factors were later found to be the largest class of transcription factors in plants and animals, comprising 3% of the genes in the human genome<sup>138</sup> (Chapter 16).

<sup>f</sup> A non-coding RNA, termed 5S-OT, is transcribed from 5S rDNA loci in eukaryotes and has been shown to regulate transcription of 5S rRNA in mammals. An antisense Alu element has inserted at the 5S-OT locus in monkeys, apes and humans and regulates alternative splicing of other genes via Alu/anti-Alu pairing.<sup>139</sup>

1986 by Nancy Kedersha and Leonard Rome.<sup>154,155</sup> Vault RNAs are considered essential for eukaryotic cell biology because of their high conservation and near ubiquitous presence.<sup>156</sup> Their function is not well understood, but recent evidence indicates that they play a role in regulating autophagy,<sup>157</sup> i.e., the degradation and recycling of cellular components in lysosomes,<sup>158</sup> as well as apoptosis<sup>159,160</sup> and signaling pathways involved in neuronal synapse formation and plasticity.<sup>161</sup>

Other RNAs discovered in these years were the highly abundant viral RNAs in infected cells, some of which interact with autoantigens. The ~160nt VA RNAs (virus-associated RNAs) present in cells infected with adenovirus<sup>163</sup> (Figure 8.4) were in fact the first non-coding RNAs described after the 'canonical' RNAs (rRNA, tRNA and mRNA) and the first non-coding RNAs shown to be expressed from mammalian viruses, reported in the same year as the identification of the *lac* repressor.



**FIGURE 8.4** The discovery of low molecular weight virus-associated RNA (VA RNA): size chromatography of newly synthesized (radiolabeled) RNA isolated from infected (open circles) and uninfected (closed circles) cells infected with adenovirus (UV absorbance shown by triangles). (Reproduced from Reich et al.<sup>163</sup> with permission from Elsevier.)

VA RNAs were later shown (among other roles) to inhibit protein kinase R (PKR)<sup>§</sup> to curb the innate immune response and enhance the translation of viral RNAs.<sup>163–169</sup>

RNAs that are not highly conserved across large evolutionary distances but have specific expression patterns were also identified during the 1980s. Examples include neuronal RNAs that are transported into dendrites, such as B2, a ~180nt RNA transcribed by RNA polymerase III from repeated sequences in mice, which shows higher expression in some tumor cells and heat-shocked cells,<sup>170–172</sup> and BC1 (brain cytoplasmic RNA 1), a ~150nt RNA transcribed in rats from repeated sequences derived from a tRNA,<sup>173–175</sup> with a human equivalent, BC200.<sup>176,177</sup> Transgenic mice lacking BC1 have no obvious developmental deficiency, but display reduced exploration activity and increased anxiety, a phenotype that is invisible in the cage but likely lethal in the wild.<sup>178</sup>

Small RNAs are also required as primers for DNA replication<sup>179,180</sup> and for the maintenance of telomeres,<sup>h</sup> shown by Elizabeth Blackburn, Carol Greider and Jack Szostak to be accomplished by retrotransposon-derived RNAs operating on repeat sequences to replicate chromosome ends.<sup>182–189</sup>

Small common RNAs such as snRNAs and snoRNAs were relatively easy to find. Others were identified later in genomic datasets by characteristic sequence motifs and secondary structures, using a growing suite of bioinformatic tools<sup>190–193</sup> and databases such as *Rfam*<sup>194,195</sup> and *RNAcentral*.<sup>196</sup> On the other hand, lower copy number transcripts that are less conserved and expressed only in particular circumstances or cells were difficult to detect, a problem compounded by the lack of anticipation of the existence of cell-specific transcripts beyond mRNAs and their nuclear precursors (Chapters 12 and 13).

## CATALYTIC RNAs AND THE ANCIENT RNA WORLD HYPOTHESIS

The participation of RNAs in a variety of cellular processes and the existence of many RNPs with different functions signaled that RNAs are versatile. As put by Crick in 1966, referring to the ability of RNA

to form complex secondary structures: “tRNA looks like Nature’s attempt to make RNA do the job of a protein.”<sup>197</sup>

In 1957, an influential symposium in Moscow on the origin of life speculated that RNA most likely preceded proteins in the origin of life, a view supported there by Brachet, Mirsky, Oparin<sup>i</sup> and others.<sup>199</sup> In 1962, Alex Rich proposed that RNAs had a central role in the origin of life.<sup>200</sup> In the late 1960s, Orgel, Woese and Crick also hypothesized that RNAs might have preceded proteins in a pre-cellular world, predicting that RNAs possessed the required enzymatic activities.<sup>197,201,202</sup>

Nonetheless, the existence of catalytic RNAs in extant organisms was completely unexpected.<sup>203</sup> In 1982, Tom Cech and colleagues discovered that RNA can perform autocatalytic ‘self-splicing’ rearrangements, removing the intervening sequence of rRNA precursors by excision and cyclization in the ciliate protozoan *Tetrahymena*. They named these catalytic RNAs “ribozymes”,<sup>204</sup> later categorized as ‘self-splicing group I introns’.

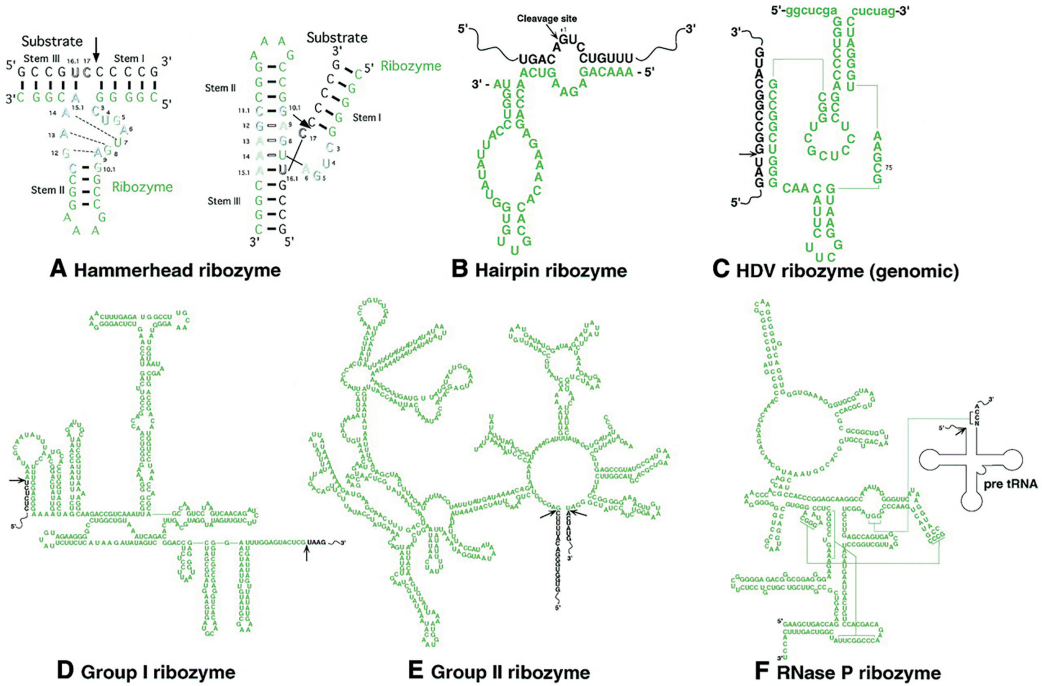
Group II self-splicing introns were recognized in 1983 in organelle genomes by François Michel and Bernard Dujon.<sup>205</sup> They consist of a catalytically active intron RNA and an intron-encoded reverse transcriptase, enabling intron proliferation within genomes. Group II intron RNA catalyzes its own splicing via transesterification reactions that are the same as those of spliceosomal introns, yielding spliced exons and an excised intron lariat RNA.<sup>206–208</sup> Thus, group II introns appear to be the ancestors of modern spliceosomal introns<sup>209–213</sup> and likely entered the eukaryotic lineage through the bacterial ancestor of the mitochondrion.<sup>214,215</sup>

Also in 1983, Sidney Altman, Norman Pace and colleagues showed that RNA is the catalytic component of the bacterial RNase P complex, which produces mature tRNAs by cleaving a 5' end sequence in a process analogous to splicing.<sup>216</sup> The RNA in RNaseP is one of the only two ribozymes found in all domains of life:<sup>217</sup> a closely related eukaryotic RNA was also described in the early 1980s<sup>218</sup> and later shown to be the catalytic center of a complex involved in sequence-specific processing of mitochondrial and other RNAs,<sup>219</sup> hence called RMRP

<sup>§</sup> PKR is in fact variously inhibited or activated by trans-acting RNAs.<sup>162</sup>

<sup>h</sup> The existence of telomeres to protect chromosome ends was inferred by McClintock and Muller in the 1930s. Muller coined the term ‘telomere’ from the Greek ‘telos’ (end) and ‘meros’ (part).<sup>181</sup>

<sup>i</sup> Oparin previously had a long running debate with Hermann Muller, with Oparin maintaining that life was the outcome of a step-wise process of pre-cellular evolution of membrane-bound polymolecular systems, whereas Muller argued that life started with the appearance of the first nucleic acid molecule.<sup>198</sup>



**FIGURE 8.5** Secondary structures of six types of ribozymes. (Reproduced from Takagi et al.<sup>222</sup> with permission of Oxford University Press. The ribozyme or intron portion is printed in green. The substrate or exon portion is printed in black.)

(RNase MRP), although it is mainly located in the nucleolus and has other important regulatory functions (Chapter 13) (Figure 8.5).<sup>220,221</sup>

The demonstration that RNA molecules are able to cleave and join themselves or other RNAs, and capable of the phosphodiester bond transfers needed for RNA synthesis, prompted Walter Gilbert to formalize the “RNA World” hypothesis.<sup>223</sup> In this view, RNA molecules, not proteins, were the precursors of existing life, having performed the catalytic and information storage functions<sup>j</sup> in the pre-cellular world.<sup>k</sup>

Since then self-cleaving ribozymes have been found in bacteria, protists, fungi, plants, nematodes,

arthropods, insects and vertebrates, including human,<sup>222,227–234</sup> one of which has been shown to methylate other RNAs.<sup>235</sup> More surprisingly for its implications, B2 and Alu elements – ‘repetitive’ sequences that occur in ~350,000 copies and over 1 million copies in the mouse and human genomes respectively – have been shown to harbor self-cleaving ribozyme activity that is induced upon stress and T-cell activation by binding to the Polycomb histone methyltransferase protein EZH2<sup>233</sup> (Chapter 16).

It has been reasonably postulated that modified nucleotides may have enhanced the early catalytic capacities of RNAs and facilitated their path to self-replication.<sup>236</sup> Ribosomal RNAs, small nucleolar RNAs and spliceosomal RNAs are all heavily modified, and RNA modification has been widely deployed as a mechanism to introduce plasticity into RNA regulatory circuits (Chapter 17).

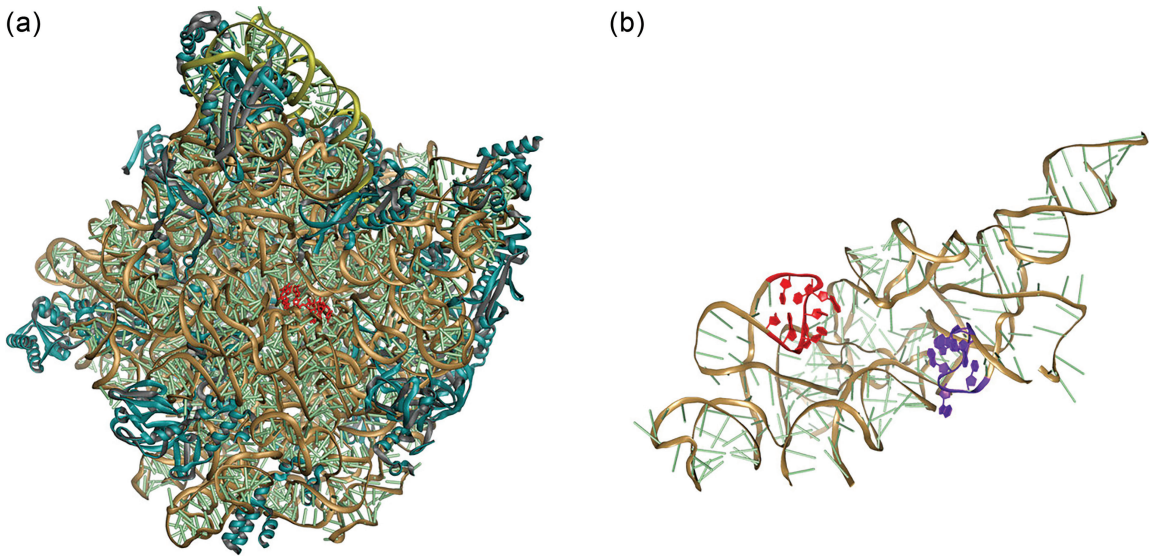
### THE CATALYTIC HEART OF SPLICING AND TRANSLATION

In 1992, Harry Noller and colleagues showed that rRNA is not just a structural scaffold for the

<sup>j</sup> This view is supported by many observations, including that an RNA polymerase ribozyme obtained by *in vitro* evolution can copy complex RNA templates, including itself, albeit at low fidelity.<sup>224</sup> There are also plausible scenarios for the prebiotic synthesis of the pyrimidine and purine building blocks of RNA.<sup>225</sup>

<sup>k</sup> RNA can also nucleate ‘liquid crystal’ phase-separated domains, explored early on by Oparin (Chapter 2), who worked extensively on the role of RNA as a polyanion in the formation of ‘coacervates’,<sup>226</sup> a property that would come to the fore as central to both modern cell biology and prebiotic evolution as RNAs would have been able to sequester organic molecules in a proto-cell (Chapter 16).





**FIGURE 8.6** The active site of the ribosome, the peptidyl transferase center, is located on the large ribosomal subunit within a highly conserved region of the ribosomal RNA (red). (a) Model of RNA structure. (b) detail. (Image courtesy of Marina Rodnina and Wolfgang Wintermeyer (Max Planck Institute for Biophysical Chemistry, Göttingen).)

ribosome, as had been widely assumed, but harbors the central peptidyl transferase activity for protein chain extension in translation, making the ribosome a complex and conserved ribozyme<sup>237,238</sup> (Figure 8.6). RNAs must therefore have pre-existed proteins.

As noted above, RNA splicing is also an RNA catalyzed reaction,<sup>1</sup> with sequence and mechanistic similarities to group II self-splicing introns in bacteria, capable of inserting into new locations by reversal of the splicing reaction.<sup>209,239</sup>

Many small molecules, including antibiotics, target ribosomal and other RNAs: these include tetracyclins, aminoglycosides such as streptomycin and gentamycin, chloramphenicol, carbomycin A, blastidicin S, puromycin and hygromycin B.<sup>240–242</sup> They also include poisons such as ricin, an N-glycosidase RNA-modifying enzyme that depurinates a conserved loop of 28S rRNA and leads to irreversible arrest of protein synthesis.<sup>243</sup> Indeed, therapeutic targeting of diverse RNA types by small molecules is an area of rapidly growing interest.

## THE DIGITAL AND ANALOG FACES OF RNA

This period also saw the expansion of RNA structural biology, led by Noller, Eric Westhof, Tom Steitz, Robin Gutell, Jennifer Doudna and others, who showed that RNAs have extraordinarily complex structures, capable of binding proteins, as a consequence of being able to form hydrogen bonds on all three faces – the Watson-Crick face, the Hoogsteen face (within the double-stranded groove) and the ribose, because of its 2' hydroxyl, which is lacking in DNA.<sup>244–246</sup> They also have exposed sequences that can base pair with other RNAs and DNA, through RNA:RNA duplexes, R-loops (RNA:DNA duplexes with displaced single stranded DNA) and RNA:DNA:DNA triplexes, which are common in eukaryotic chromatin<sup>247,248</sup> (Chapter 16).

The structural versatility of RNA has been explored and exploited *in vitro*. In the 1990s, the groups of Larry Gold and Jack Szostak developed SELEX ('Systematic Evolution of Ligands by Exponential Enrichment') to evolve artificial RNAs that bind specific ligands ('aptamers') or have other activities,<sup>249–252</sup> speculating that the same will have occurred *in vivo*.<sup>253</sup>

In addition, "free" low molecular weight circular RNA molecules lacking protein-coding capacity but

<sup>1</sup> U2 and U6 snRNAs interact to form the conserved structure of the catalytic triplex, coordinating two magnesium ions to form the active site of the spliceosome.<sup>212</sup>

having “peculiar” secondary structures and ribozyme activity were discovered in the 1970s to infect and autonomously replicate in plant cells (baptized as ‘viroids’),<sup>254,255</sup> and postulated to represent living fossils<sup>256</sup> subject to Darwinian evolution in a prebiotic world.<sup>257</sup> Other curious virally associated RNAs were also described in the late 1970s and 1980s.<sup>258–260</sup>

## CANDLES IN THE DARK

By 1985, it was becoming apparent that RNAs are multifaceted molecules that have specific subcellular locations, form complex structures, interact with (many) proteins and perform a vast array of functions beyond protein synthesis, from gene regulation to acting as components of cellular complexes, catalytic molecules and antisense guides. These observations did little to challenge the assumption that the destination of genetic information is (nearly always) the production of a protein and were regarded as interesting but idiosyncratic additions to the tapestry of molecular biology, rather than the first indications of a wider role for RNA in cell and developmental biology.

In his 1986 article describing the RNA World hypothesis, Gilbert proposed that, after the emergence of DNA as the carrier of genetic information, RNA was then “relegated to the intermediate role that it has today – no longer the centre of the stage, displaced by DNA and the more effective protein enzymes”.<sup>223</sup> Others concurred.<sup>261</sup>

Penman lamented in 1991:

If genes just make proteins and our proteins are the same, then why are we so different?  
... we have the bizarre proposal dominating biology that the incredibly complex living systems are described entirely by component

proteins and their coding sequences. Where is the genetic information that executes the design of an organism? We do not have to look far for a candidate. There is plenty of information in the more than 95% of the genome that is devoid of open reading frames. These sequences, heavily transcribed in all cells, appear to have little, if anything, to do with making proteins.<sup>262</sup>

## FURTHER READING

- Darnell J.E. (2011) *RNA: Life's Indispensable Molecule* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY).
- Gesteland R.F., Cech T. and Atkins J.F. (2006) *The RNA World: The Nature of Modern RNA Suggests a Prebiotic RNA World* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY).
- Joyce G.F. (1991) The rise and fall of the RNA world. *The New Biologist* 3: 399–407.
- Lazcano A. (2012) The biochemical roots of the RNA world: From zymonucleic acid to ribozymes. *History & Philosophy of the Life Sciences* 34: 407–23.
- Mount S.M. and Wolin S.L. (2015) Recognizing the 35th anniversary of the proposal that snRNPs are involved in splicing. *Molecular Biology of the Cell* 26: 3557–60.
- Müller, F., et al. (2022) A prebiotically plausible scenario of an RNA–peptide world. *Nature* 605: 279–284.
- Robertson M.P. and Joyce G.F. (2012) The origins of the RNA world. *Cold Spring Harbor Perspectives in Biology* 4: a003608.
- Vicens Q. and Kieft J.S. (2022) Thoughts on how to think (and talk) about RNA structure. *Proceedings of the National Academy of Sciences USA* 119: e2112677119.
- Wilkinson M.E., Charenton C. and Nagai K. (2020) RNA splicing by the spliceosome. *Annual Review of Biochemistry* 89: 359–88.