Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

CHAPTER

2

Virus Origins: Conjoined RNA Genomes as Precursors to DNA Genomes

Hugh D. Robertson and Olivia D. Neel

INTRODUCTION

The rapid and unexpected progress in RNA research during the past two decades has led to many theoretical and practical advances. RNA's unprecedented ability to act both as a template for information storage and as an enzymatic molecule has led to the proposal that primitive living systems were based on RNA, with protein synthesis and DNA templates for information storage added later. If this "RNA world" hypothesis is to be taken seriously, it is necessary to explain a number of developments during evolution at the RNA level, including not only coding and self-replication but also the ability of genetic information to rearrange, recombine and expand itself, creating ever more complex living systems. It is the purpose of this chapter to interpret certain RNA-level events that must have taken place in simple viral or pre-viral systems in this light.

In what follows, we will seek, in what is known today about simple self-replicating RNAs, enlightenment regarding how they acquired their singular nature. We will focus on a process that has been called "RNA conjunction" (Branch *et al.*, 1989) or "RNA capture" (Diener, 1989), in which two independent, functional RNA molecules become associated in such a way that each retains function and contributes properties to the resulting "conjoined RNA" or "RNA mosaic". As we shall see, RNA conjunction differs from both random RNA rearrangements and recombination between the RNAs of closely related RNA viruses, in that two independent activities, each embodied in a separately evolved RNA, are required to survive the conjunction or capture process that joins them together. This is not to suggest that the mechanisms that drive random or homologous RNA recombination are not used in RNA conjunction but rather that conjoined RNAs comprise a highly selected subset of successful multifunctional RNA mosaics.

Assuming that a prebiotic system of chemical evolution somehow produced RNA in the first place, there has been intense speculation about how early, small RNAs might have replicated. And, assuming that a genetic code leading from nucleic acids to proteins also evolved, there has been equally thorough scrutiny as to how the replicating and coding RNAs in such a hypothetical primitive time might have combined and expanded leading to viral and, ultimately, cellular RNAs. Once converted to the more stable DNA storage system, these molecules may have formed the basis for modern DNA viruses. Until recently, studies on the creation and properties of conjoined RNAs have been theoretical in nature, emphasizing computer modeling and mutational probabilities rather than experiments or molecular prototypes. Recent work to be reviewed below shows that there is one class of primitive life forms – the viroid-like

Origin and Evolution of Viruses
ISBN 0-12-220360-7

pathogens – whose properties today could help us to understand how primitive RNA-based self-replication may have been compatible with expansion to produce more complex RNAs. In summary, the causative agent for human hepatitis delta contains two specialized domains, one concerned with replication and the other encoding a single protein (Branch *et al.*, 1989; Purcell and Gerin, 1996; Taylor, 1996). From both theoretical and practical considerations, it now seems likely that the two RNA domains that embody these two functions arose separately and were subsequently joined together. The existence of a prototype conjoined viral RNA that is functional in the modern world provides a singular opportunity for testing some of the above ideas, and has already caused a redoubling of efforts to find other examples as well as to understand the one we have.

In this chapter, we will first review briefly current knowledge about RNA rearrangement and recombination, principally in viruses. We will cite evidence for various mechanisms catalyzing these events. We will also review some recent evidence that, at least in one system, RNA recombination can occur in what appears to be a spontaneous manner, as if it were an inherent property of the RNA. Such a potential, even at low frequency, would expand opportunities for RNA conjunction. Second, we will outline the significance of work on viroid-like pathogens, circular RNA replication and their potential relation to early RNA. We will then put delta agent RNA in context, discussing the relative significance of its dual RNA nature to the RNA recombination systems already cited. Finally, we will relate the early emergence of RNA mosaics to developments leading to today's DNA-based systems of viral gene expression.

RNA REARRANGEMENT: MECHANISMS OF VIRAL RNA RECOMBINATION

While most studies on genetic recombination have been carried out on DNA-based organisms, there are several examples from the field of RNA virology that clearly demonstrate that RNA recombination is a reality. Animal, plant and bacterial virus systems have all been identified in which RNA recombination takes place (reviewed in Lai 1992a,b, 1995, Nagy and Simon, 1997). In most of these studies, emphasis is placed on the types of RNA molecules that are joined (homologous versus non-homologous) and the mechanism by which two separate RNAs become recombined. The majority of RNA recombination appears to involve closely related, or homologous, sequences, in which mutant markers are reassorted in an orderly fashion. Most picornavirus and coronavirus RNA recombination takes place in this way (Lai, 1992b; Zhang and Lai, 1994; Pilipenko *et al.*, 1995; Duggal *et al.*, 1997), although there are exceptions. The same is true of most plant virus RNA recombination events (Gibbs and Cooper, 1995; Le Gall *et al.*, 1995; Figlerowicz *et al.*, 1997; Fraile *et al.*, 1997; Nagy and Bujarski, 1997, 1998), including those involving the well-studied brome mosaic virus (BMV) system. However, there are exceptions, as exemplified by the turnip crinkle virus (TCV)/satellite system (Carpenter *et al.*, 1995; Carpenter and Simon, 1996a,b; Nagy and Simon, 1997).

The favored recombination mechanism for viral RNA molecules is one involving template switching (analogous to the copy/choice mechanism of DNA recombination), in which the RNA-dependent RNA polymerase of the virus ceases the copying of a particular strand in midreaction, moves to a second strand with nascent RNA still attached and resumes synthesis at a point in the second strand near the place where copying ended in the first one (Lai, 1992a, b). Other possibilities include a cleavage/ligation reaction resembling trans RNA splicing (Maroney *et al.*, 1996) and a recently identified transesterification process (Chetverin *et al.*, 1997). The majority of well-studied examples of both plant and animal viruses have been assigned to the template switching category of RNA recombination, and much effort has gone into the identification of regions of sequence or secondary structure which would promote the template switching event.

An orthodox view of viral RNA recombination would thus include the involvement of known components – the viral RNA-dependent

RNA polymerase and viral RNA strands for template switching; previously known ribozymes or conventional RNA processing enzymes for the break/rejoin reactions analogous to trans RNA splicing. The ability to harness such mechanisms to reassort viral RNA genomes and promote new, and perhaps more fit, combinations is viewed as a significant contributing factor to the evolution of RNA viruses.

Included in the catalog of RNA recombination studies are a few cases in which viruses have picked up host-cell RNA sequences. These examples will be important when we consider the nature of the reactions that produced delta agent RNA. In plant viruses, Mayo and Jolley (1991) have shown the occasional uptake of RNA sequences encoded in host chloroplast DNA. Because the acquired sequence is part of an open reading frame, and the recombination site is within 7 bases of an exon–intron boundary, the authors speculate that the recombination occurred by a trans splicing (or break/rejoin) mechanism. That plant viruses can recombine with cellular mRNAs to promote new sequence combinations was proved unmistakably when Greene and Allison (1994) demonstrated that, in transgenic plants containing a viral RNA sequence now expressed in the cell as a conventional mRNA, recombination with exogenous virus could take place as its RNA replicated in the cells of the transgenic plant. Presumably such events can occur with more conventional cellular mRNAs as well, although none were reported in this system.

In animal virus systems, there are several examples in which host cell RNAs are incorporated into viral RNA. In influenza viral RNA, for example (Khatchkian *et al.*, 1989) a segment of host 28S rRNA is incorporated into the hemagglutinin gene by a mechanism involving nonhomologous recombination. It is reported that the viral species containing the host RNAs have increased viral pathogenicity, although it is not known whether this trait conferred a selective advantage upon the recombinant influenza virus population. In Sindbis virus (Monroe and Schlesinger, 1983), tRNA sequences are sometimes incorporated into the 5′ termini of defective RNAs; while in the TCV satellite system, the incorporation of nonviral sequences has been reported (Carpenter *et al.*, 1995; Nagy and Simon, 1997).

Perhaps the most striking example of RNA recombination between a virus and host gene sequences is the bovine viral diarrhea virus, BVDV, a pestivirus with a single-stranded RNA genome 12.5 kb in length that encodes a single polyprotein (Collett *et al.*, 1989; Meyers *et al.*, 1991). In the course of an investigation of changes in cytopathogenicity among different BVDV strains, many were found to have acquired cellular RNA sequences into a domain encoding a non-structural protein. The most frequently observed inserts consisted of sequences from the host ubiquitin gene (Meyers *et al.*, 1991). It is not known whether expression of the acquired sequences took place, but the virus was clearly able to survive this acquisition into its polyprotein, and in some cases to acquire a selective advantage. This phenomenon could be reproduced, and the investigators concluded that some unknown features of BVDV RNA not only facilitated the recombination process but also conferred some selective advantage to the recombinants.

Thus today's viral RNA recombination mechanisms can occasionally lead to the acquisition of host sequences apparently unrelated to the virus. Before concluding that such events must always be mediated by one of only two mechanisms – template switching by the viral RNA polymerase or specific cleavage and ligation of RNA resembling trans splicing – it is as well to consider some recent findings from a phage system, which suggest that RNA recombination may take place by a more general, chemical mechanism. In the Qβ phage system, Chetverin *et al.* (1997) have discovered non-homologous RNA recombination, which takes place in a cell-free system at a variety of sequence locations. The non-homologous recombinations observed are entirely dependent on the 3′ hydroxyl group of the 5′ fragment in the joining reaction. Chetverin *et al.* (1997) believe that the mechanism by which these recombinants are generated is "entirely different from copy choice".

Nagy and Simon (1997), in reviewing the above work from a perspective favoring template switching, concede that the data of Chetverin *et al.* can all be explained by an RNA-

mediated transesterification mechanism, but that a template-switching mechanism is not excluded. While further controls need to be done, uncoupling the recombination events from the Qβ replicase-dependent amplification needed to detect the results, it seems probable that an RNA-mediated breakage and ligation accounts for at least a fraction of Qβ RNA recombinants. And, while Nagy and Simon (1997) correctly point out that "it is difficult to estimate how widespread [such a system] might be in natural virus systems", the prospect that RNA molecules have a certain probability for spontaneous rearrangement provides additional scope for the evolution of viral RNAs.

VIROID-LIKE AGENTS, CIRCULAR RNA REPLICATION AND EARLY RNA GENOMES

Early reports of viroid-like RNA pathogens centered on plant viroids and their relatives (Gross *et al.*, 1978; Diener, 1979; Semancik, 1987; Branch *et al.*, 1990). More recently, the causative agent for delta hepatitis in humans was confirmed to be a circular viroid-like RNA (Kos *et al.*, 1986; Wang *et al.*, 1986; Makino *et al.*, 1987; Taylor *et al.*, 1987). Delta RNA is about four times the size of plant viroids. The principal effort which led to the working out of the replication cycle for these agents took place between 1981 and 1987, at a time when the role of RNA in the evolution of primitive, self-replicating systems was just coming into focus. For example, two proposals based on both the template and enzymatic qualities of RNA (Sharp, 1985; Gilbert, 1986) appeared during that time. The potential for RNA circles to simplify the tasks required for replication in a primitive environment is considerable, and includes at least four elements. First, as with circular DNA genomes (Reanney and Ralph, 1968), there are advantages involving the ability to tolerate gene duplication and subsequent variation while preserving the initial sequence; second, as pointed out previously (Robertson, 1992), the synthesis of multimeric copies on a circular complementary template leads automatically to the unwinding of each copy from duplex structure with its template as it is displaced by the next copy; third, as also pointed out by Diener (1989), the need for a specific initiation point at one end of a linear genome is eliminated; and fourth, circular RNAs with no free ends – especially if they also contain extensive secondary structure as do the RNAs of viroid-like pathogens – are less susceptible to breakdown by ribonucleases than normal RNA molecules.

Many advocates of the "RNA world" hypothesis (Gesteland and Atkins, 1993) have proposed a set of common assumptions. One is that RNA molecules evolved self-replication first, then the property of protein coding and finally an information storage system using DNA copies. This idea leads to the prediction that genetic systems of today will contain features reflecting such a history. In the context of viroid-like RNAs, one way to test these assumptions is to consider the way today's viroid-like RNAs, including that of the delta agent, are thought to replicate. We proposed the rolling circle pathway as a general mechanism for viroid-like RNA replication (Branch *et al.*, 1981; Branch and Robertson, 1984), in which multimeric copies of RNA strands are synthesized and then processed to yield monomeric progeny molecules (Figure 2.1). This pathway has been demonstrated for a number of viroid-like RNAs, including the delta agent (Chen *et al.*, 1986). Host enzymes are required for the RNA synthetic steps of this pathway, and are the only proteins absolutely required for replication (since examples of RNA-catalyzed cleavage of multimers and ligation to form circles have been documented in several systems).

Studies by Cech and others (reviewed in Cech, 1989) have begun to demonstrate how RNA may have first begun to copy itself. These proposals reveal several potential problems, e.g. how to copy accurately (and protect from exconuclease cleavage) the ends of such molecules; how to unwind the newly synthesized RNA strand from a stable duplex with its template so that subsequent rounds of copying can proceed; and how to initiate synthesis without a pre-existing set of initiation factors in a way that guarantees accurate inheritance of every base by the progeny RNA. As mentioned above, a circu-

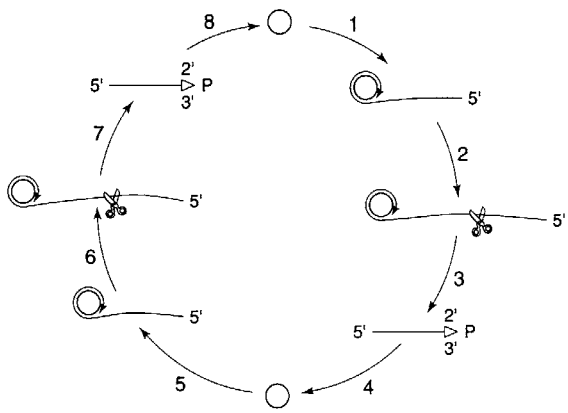

FIGURE 2.1 The rolling circle replication pathway for viroid-like RNAs. In this depiction, the circular genomic ("+") strands are copied into multimeric antigenomic ("–") strands (steps 1 and 2), cleaved to unit length (step 3) and ligated to give minus strand circles (step 4). These antigenomic monomeric circles serve as templates for multimeric genomic strands (steps 5 and 6) which are cleaved to unit length (step 7) and circularized to produce progeny genomic RNAs (step 8).

lar template simplifies all of these difficulties. Since it has no ends, its structure is stabilized in a fashion impossible for linear molecules. In addition, a circular template undergoing copying by an RNA polymerase (whether a primitive one composed of RNA or a modern host protein) will lead to production of greater than unit length multimeric copies, so that the first copy will be displaced from its duplex association with the template as the second copy is synthesized, overcoming the unwinding dilemma. Furthermore, initiation at any point on a circular template leads to a complete copy with no risk of losing ends or other domains. We conclude that the case for primitive circular self-replicating RNAs is a persuasive one.

Another line of thought concerning viroid-like RNAs in evolution emerged shortly after the discovery of eukaryotic RNA splicing, in which a number of investigators speculated on the potential relationship between viroids and intervening sequences, or introns. Shortly after split genes and the need for mRNA splicing were first announced, Roberts (1978) speculated about a connection between viroid-like RNAs and introns, guessing (correctly) that RNA splicing mechanisms might turn out to be reciprocal, not only joining exons but producing circular introns as well. Crick (1979) observed that introns might be excised as circles, while both Diener (1981) and Dickson (1981) pointed out sequence homologies between the plant viroid PSTV and the small nuclear RNAs involved in mRNA splicing. Gilbert (1987) focused on the possibility that early ribozyme-containing introns in the "RNA world" might somehow have served as insertion sequences.

The idea that emerges, then, is that introns originally arose as circular self-replicating RNAs, with a replication pattern that presaged both RNA capture and modern mRNA splicing. The earlier speculations cited above (Roberts, 1978; Crick, 1979; Diener, 1981; Dickson, 1981; Gilbert, 1987) did not focus on the way in which the rolling circle mode of replication used by viroid-like RNAs (Branch and Robertson, 1984) combines stable RNA circles with ribozymes that cleave and ligate RNA (although Crick (1979) does point out that, if introns were excised as circles, "There is little difficulty in thinking of interesting functions which such a single-stranded circular RNA might perform," and gives a reference to viroids). Subsequent publications (Robertson and Branch, 1987; Diener, 1989; Branch *et al.*, 1989; Robertson, 1992) recognized these and other advantages of circular RNA, leading to the idea that viroid-like RNA circles could have developed into introns over evolutionary time. Indeed, if the rolling circle pathway was in fact employed in the RNA world (Gesteland and Atkins, 1993), events taking place during each cycle of RNA synthesis – in which linear monomers built into multimers by repeated copying of a circular RNA template are cut apart and then circularized by ribozyme action – could foreshadow the development of RNA introns. The existence of self-replicating RNA circles equipped with the ribozyme machinery to cleave newly synthesized chains and then join the newly formed ends would lead naturally to events in which cleavage could be followed by the joining of two different molecules – perhaps rarely at first, and then more often. Alternatively, template switching during rolling circle replication could also lead to the joining of a viroid-like RNA and coding

segment. In either case, as the population of circular self-replicating RNAs increased, the likelihood of their acquiring newly arisen coding sequences would go up in parallel.

DELTA AGENT RNA

The genome of the hepatitis delta agent or virus is a circular, highly structured single-stranded RNA. As shown in Figure 2.2, delta RNA has two domains, each with features of two quite different RNA types. Shown on the left is the viroid-like region, comprising about one-quarter of the bases. This domain is named after the 350–400 base viroid RNAs that infect plants, and contains most of the sites important for the rolling circle replication pathway, including the ribozyme activities, whose cleavage sites are indicated. The three-quarters of the delta RNA map depicted on the right of Figure 2.2 is very different: as shown, it encodes mRNA for the delta antigen, a minor structural protein of delta particles. The protein-coding region is much more divergent in sequence than the viroid-like region. The differences between these two domains led to the postulate that the delta RNA genome arose from two domains and the suggestion that this RNA in its present form is the product of RNA conjunction (Branch *et al.*, 1989); and the broadening of this hypothesis (Robertson, 1992).

In common with plant viroids, many of which have a core "central conserved region" involving about 50 bases, which are nearly invariant (Keese and Symons, 1985), part of the delta viroid-like region comprising bases 613–771 and 842–957 shows similar evidence of conservation (Krushkal and Li, 1995). In contrast, the protein coding region can have up to 20% variability between two given strains whose viroid-like sequences are completely conserved (Imazeki *et al.*, 1990). Thus the viroid-like and protein-coding regions have many properties to suggest they originated separately and were joined by a process which – if better understood – could help us to understand not only RNA evolution but also its present-day function. These conclusions would remain valid regardless of whether delta RNA itself turned out to be a primitive molecule – a "missing link" in molecular evolution – or whether delta represents a recent recurrence of processes that have a constant potential to occur among informational RNA molecules which are active in present-day cells.

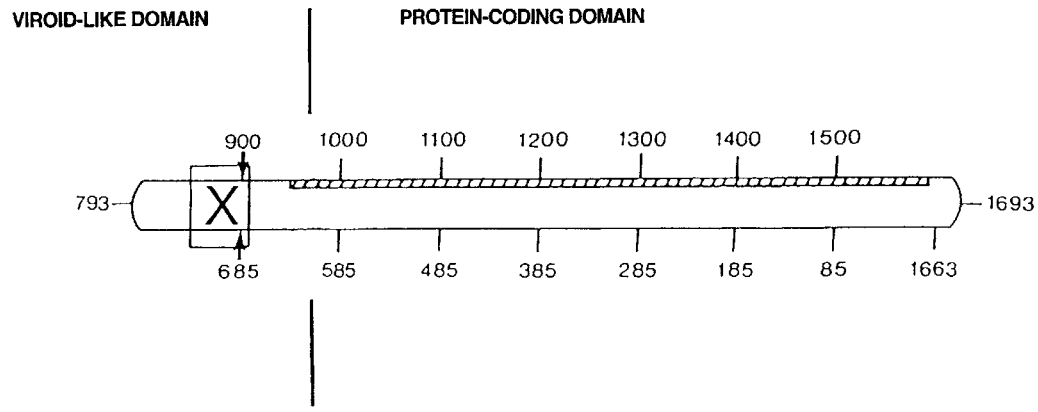

FIGURE 2.2 The genomic RNA of the delta hepatitis agent. As shown, the circular single-stranded RNA molecule comprises two separate domains, the viroid-like region (left) and the protein-coding region (right). Residues 613 and 980 form the boundary between the two regions. The hatched area corresponds to the residues that encode the mRNA for the delta antigen protein, whose coding sequence is complementary to the region shown here. Residues 685 and 900 (arrows) depict the sites of ribozyme cleavage in the genomic and antigenomic strands respectively. The box to their left marked with an "X" is a region of local tertiary structure thought to be involved in interaction with host proteins.

The possibility that some of the above models might be subject to experimental test arose when Brazas and Ganem (1996) detected an mRNA molecule in uninfected liver cells encoding a protein that can strongly influence the replication of hepatitis delta RNA, and which appears to be related in amino-acid sequence and composition to the delta antigen protein, HDAg, for which it has a binding affinity. This host protein, named Dip-A (for Delta-interacting protein A), has been isolated and its gene cloned. The 202-amino-acid Dip-A protein is similar in size to HDAg and 56% of its amino acids are identical or very similar to the corresponding delta antigen residues. The question of whether this sequence similarity is statistically significant has been the subject of two additional publications (Brazas and Ganem, 1997; Long *et al.*, 1997). The predominant conclusion, reached using the latest protein sequence comparison programs devised by Dr S. Altschul of NIH (cited in Brazas and Ganem, 1997), is that the similarity between Dip-A and HDAg remains striking enough for the possibility that they have diverged from a common ancestor to be vigorously pursued.

If such events have taken place, we need to propose testable mechanisms for the creation of viable RNA mosaics. The rolling circle mechanism illustrated in Figure 2.1 and the map of delta RNA in Figure 2.2 allow us to visualize one process of RNA conjunction that might accomplish this end. If a primitive viroid-like RNA 368 bases long occupies residues 613–980 (reading 5′ to 3′ (clockwise) in Figure 2.2) of the delta map, then there must have been a way either during or after RNA synthesis to join residue 980 to the 5′ end of a copy of the cellular RNA to be captured. Brazas and Ganem (1996) have suggested a copy-choice mechanism, in which the enzyme system copying the viroid negative strand template switches, presumably at low frequency, to copying a cellular RNA such as Dip-A mRNA. As outlined in the section on RNA rearrangent above, a number of RNA viruses that undergo RNA rearrangement are thought to use such a template-switching mechanism, in which the virally encoded RNA polymerase moves from one RNA to another while continuing synthesis (Kierkegaard and Baltimore, 1986; Lai, 1992a,b; Carpenter *et al.*, 1995; Nagy and Bujarski, 1996). In contrast, delta agent RNA encodes no RNA polymerase, using instead a host activity.

An alternate mechanism for RNA capture by delta would involve the sites of RNA cleavage and ligation shown in Figure 2.2 (whereby the multimeric RNA precursors produced by rolling circle replication yield monomers that are then circularized): the termini of linear monomers could occasionally be misaligned, joining a linear negative viroid RNA strand to a cellular mRNA species. Such a trans splicing mechanism was proposed by Mayo and Jolly (1991) to account for the occasional acquisition by potato leafroll virus RNA of unexpressed host RNA fragments. The formation of RNA mosaics by either mechanism would be expected to preserve the ability of the delta viroid-like domain to replicate, and the resulting RNA would encode an advantageous protein.

While other cases in which RNA viruses acquire altered RNAs from each other or even from host sources by RNA recombination have been cited above, the delta agent is the only known case where a coding sequence thought to be cellular in origin has entered a functional association with a replicating virus-like element at the level of RNA only, with no DNA involvement, so that the translation-level expression of the protein has survived *and* the replicating element remains functional. If this is indeed the result of RNA capture, leading to a molecule that fits the definition of a conjoined RNA given in the introduction to this chapter, we would assume that such successful events would be rare, and when they happened it is likely that they would usually prove fatal to the capturing organism. If an RNA mosaic were to survive, we would assume both that its ability to replicate has been preserved and that the encoded protein must contribute somehow to the welfare of the mosaic. For example, Brazas and Ganem (1996) show that Dip-A has sequence elements in common with nucleic acid binding proteins and transcription factors. RNA encoding a protein like this, which might help a highly structured RNA molecule like delta to simplify replication, might confer increased survival value to the replicating RNA.

More recent work from the Ganem laboratory on the nature and function of the Dip-A protein and the origin of its mRNA has focused on two issues: the likelihood that delta RNA replication indeed involves host RNA polymerase II and the nature and cellular function of the Dip-A protein itself. The former consideration is important for any speculation about delta RNA recombination mediated by a template-switching mechanism. If host RNA polymerase II can copy delta RNA, then its occasional switch – during delta genomic strand synthesis – to a host nuclear mRNA template would not be entirely far-fetched. The resulting conjoined RNA would contain the delta genomic strand attached to the template for the host mRNA.

Additional mapping experiments carried out by Brazas and Ganem (personal communication) using the mouse Dip-A locus have strengthened the idea that this gene may encode a protein related to transcription. The mouse Dip-A locus (which turns out to be 98% homologous at the amino-acid level to the human Dip-A previously reported (Brazas and Ganem, 1997) maps only 600 bases away from a mouse homolog of the Fra (fos-related antigen) transcription factor family. Since previous work (Brazas and Ganem, 1996) had establishd that Dip-A protein is homologous to Fra-1 over a limited range, this mapping result supports the notion that Dip-A is a member of the Fra family. It is important to recall that – in addition to the finding that Dip-A is a host cellular protein with extensive homology to the delta antigen protein – Dip-A also has an affinity for the delta antigen which may play a role in delta replication. To test this, knockout mice are being developed by Ganem and co-workers, and it should soon be possible to evaluate whether the host Dip-A protein plays a part in the delta RNA synthetic cycle (R. Brazas and D. Ganem, personal communication).

RELATION TO DNA SYSTEMS

It is evident that if RNA conjunction leading to delta-like RNA mosaics with both replicating and functionally translatable protein-coding domains has taken place, we need to consider the consequences both for the evolution of primitive RNA systems, yielding today's DNA-based cellular information system, and for present-day RNA-level events. We suggest that what have now become modern mRNA introns arose because their ancestors carried activities essential for RNA replication in pre-DNA times. The rolling circle yields progeny multimers to be cleaved out and ligated to give progeny circles. This primitive RNA replication pathway could have led first to sequence capture and the formation of RNA mosaics and then to RNA splicing, both using the same pre-existing ribozyme activities. RNA mosaics would build up in this way, interspersing viroid-like replicating elements with newly evolved coding segments (some of which could produce larger proteins by developing the ability to be periodically spliced together using ribozyme action). It is clear that template switching by the primitive RNA-synthesizing machinery could also contribute to this process.

In this context, the ongoing possibility of RNA variation would have been accompanied by a constant tendency to combine, re-assort and test new RNA combinations. The resulting RNA mosaics would eventually be copied into DNA with the arrival of reverse transcriptase activities. The RNAs transcribed from such DNAs, since they in turn would need to be spliced for proper expression of their encoded proteins, would be expected to preserve the built-in splicing capability inherited from their RNA-coded ancestors. Such reactions would be further enhanced by the presence of RNA segments encoding RNA-binding proteins, some of which would evolve into transcription factors for the DNA-directed enzymes now in place. However, the role in RNA replication previously associated with the viroid-like RNA domains would be lost.

What we see, then, is a system in which DNA encoding RNA mosaics emerges with a ready-made RNA splicing system to liberate the coding segments for mRNA function. This early emergence of RNA splicing may explain why it has been retained so far by eukaryotic genomes despite its many problems. Furthermore, the potential for intervening sequences to retain at least those functions required for cleavage and

ligation steps in RNA splicing suggests that they may be less inert than previously suspected. The idea considered here – that intervening sequences began as viroid-like RNAs – carries with it the possibility that self-replicating RNAs may re-emerge from transcripts of DNA, combine with other RNA segments and give rise to new combinations of genetic material. In this regard, it is possible that directed versions of such a process could lead to a new class of RNA-level vectors capable of changing the genetic expression of the cells containing them. It is also possible that naturally occurring RNA mosaics with pathogenic potential continue to arise with some regularity *in vivo*. Whether the delta agent RNA arose in this way, or whether it instead represents an ancestral intermediate in molecular evolution, remains to be determined.

Whatever their role may have been in the evolution of RNA viruses, it is clear that the mechanisms and processes for RNA-level genetic assortment and reassortment are alive and well today. Whether they play a role in the continued development of DNA-based viral or retroviral systems needs further evaluation. Now that intron-containing RNAs of both viral and cellular origin are being transcribed from DNA templates, and their rates of sequence variation studied, it is well to remember that such rates would be profoundly influenced if it should turn out that what was thought to be a DNA gene had spent even a short time in RNA form. Such an "escaped gene" would temporarily be free of the usual DNA-based constraints on mutation and recombination, and could thus rapidly turn into something quite different. This "RNA uncertainty principle" applies, for example, in the case of the Dip-A versus delta antigen mRNA sequence comparison, where the divergence of at least one of this pair of sequences descended from a hypothetical common ancestor has taken place at the RNA, not the DNA, level. Further studies on the Dip-A gene sequences and their RNA-based counterparts should allow the extent of "RNA uncertainty" to be discerned and then used to enhance our understanding of DNA *and* RNA virus evolution.

ACKNOWLEDGEMENT

We thank Drs D. Ganem and R. Brazas for helpful discussions and sharing of results prior to publication. Work in our laboratory has been supported by the US National Institutes of Health and by the New York State Science and Technology Foundation.

REFERENCES

Branch, A.D. and Robertson, H.D. (1984) A replication cycle for viroids and other small infectious RNAs. *Science*, **223**, 450–454.

Branch, A.D., Robertson, H.D. and Dickson, E. (1981) Longer-than-unit-length viroid minus strands are present in RNA from infected plants. *Proc. Natl Acad. Sci. USA*, **78**, 6381–6385.

Branch, A.D., Benenfeld, B.J., Baroudy, B.M., Wells, F.V., Gerin, J.L. and Robertson, H.D. (1989) An ultraviolet-sensitive RNA structural element in a viroid-like domain of the hepatitis delta virus. *Science*, **243**, 659–652.

Branch, A.D., Levine, B.J. and Robertson, H.D. (1990) The brotherhood of circular RNA pathogens: viroids, circular satellites, and the delta agent. *Sem. Virol.*, **1**, 143–152.

Brazas R. and Ganem, D. (1996) A cellular homolog of hepatitis delta antigen: implications for viral replication and evolution. *Science*, **274**, 90–94.

Brazas R. and Ganem, D. (1997) Response to [Delta-interacting protein A and the origin of hepatitis delta antigen]. *Science*, **276**, 825.

Carpenter, C.D. and Simon, A.E. (1996a) Changes in locations of crossover sites over time in de novo generated RNA recombinants. *Virology*, **223**, 165–173.

Carpenter, C.D. and Simon, A.E. (1996b) In vivo restoration of biologically active 3′ ends of virus-associated RNAs by nonhomologous RNA recombination and replacement of a terminal motif. *J. Virol.*, **70**, 478–486.

Carpenter, C.D., Oh, J. W., Zhang, C. and Simon, A.E. (1995) Involvement of a stem-loop

structure in the location of junction sites in viral RNA recombination. *J. Mol. Biol.*, **245**, 608–622.

Cech, T.R. (1989). RNA enzymes. *Adv. Enyzmol*, **62**, 1–36.

Chen, P.J., Kalpana, G., Goldberg, J. *et al.* (1986) Structure and replication of the genome of the hepatitis delta virus. *Proc. Natl Acad. Sci. USA*, **83**, 8774–8778.

Chetverin, A.B., Chetverina, H.V., Demidenko, A.A. and Ugarov, V.I. (1997) Nonhomologous RNA recombination in a cell-free system: evidence for a transesterification mechanism guided by secondary structure. *Cell*, **88**, 503–513.

Collett, M.S., Moening, V. and Horzinek, M.C. (1989) Recent advances in pestivirus research. *J. Gen. Virol.*, **70**, 253–266.

Crick, F. (1979) Split genes and RNA splicing. *Science*, **204**, 264–271.

Dickson, E. (1981) A model for the involvement of viroids in RNA splicing. *Virology*, **115**, 216–221.

Diener, T.O. (1979) *Viroids and Viroid Diseases*, John Wiley, New York.

Diener, T.O. (1981) Are viroids escaped introns? *Proc. Natl Acad. Sci. USA*, **78**, 5104–5105.

Diener, T.O. (1989) Circular RNAs: relics of precellular evolution? *Proc. Natl Acad. Sci. USA*, **86**, 9370–9374.

Duggal, R., Cuconati, A., Gromeier, M. and Wimmer, E. (1997) Genetic recombination of poliovirus in a cell-free system. *Proc. Natl Acad. Sci. USA*, **94**, 13786–13791.

Figlerowicz, M., Nagy, P.D. and Bujarksi, J.J. (1997) A mutation in the putative RNA polymerase gene inhibits nonhomologous, but not homologous, genetic recombination in an RNA virus. *Proc. Natl Acad. Sci. USA*, **94**, 2073–2078.

Fraile, A., Alonso-Prados, J.L., Aranda, M.A., Bernal, J.J., Malpica, J.M. and Garcia-Arenal, F. (1997) Genetic exchange by recombination or reassortment is infrequent in natural populations of a tripartite RNA plant virus. *J. Virol.*, **71**, 934–940.

Gesteland, R.F. and Atkins, J.F. (1993) *The RNA World*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Gibbs, M.J. and Cooper, J.I. (1995) A recombinational event in the history of luteoviruses probably induced by base-pairing between the genomes of two distinct viruses. *Virology*, **206**, 1129–1132.

Gilbert, W. (1986) The RNA world. *Nature*, **319**, 618.

Gilbert, W. (1987) The exon theory of genes. *Cold Spring Harbor Symp. Quantit. Biol.*, **52**, 901–905.

Golding, B.G., Tsao, N. and Pearlman, R.E. (1994) Evidence for intron capture: an unusual path for the evolution of proteins. *Proc. Natl Acad. Sci. USA*, **91**, 7506–7509.

Greene, A.E. and Allison, R.F. (1994) Recombination between viral RNA and transgenic plant transcripts. *Science*, **263**, 1423–1425.

Gross, H.J., Domdey, H., Lossow, C. *et al.* (1978) Nucleotide sequence and secondary structure of potato spindle tuber viroid. *Nature*, **273**, 203–211.

Imazeki, F., Omata, M. and Ohto, M. (1990) Heterogeneity and evolution rates of delta virus RNA sequences. *J.Virol.*, **64**, 5594–5599.

Keese, P. and Symons, R.H. (1985) Domains in viroids: evidence of intermolecular RNA rearrangements and their contribution to viroid evolution. *Proc. Natl Acad. Sci. USA*, **82**, 4582–4586.

Khatchkian, D., Orlich, M. and Rott, R. (1989) Increased viral pathogenicity after insertion of a 28S ribosomal RNA sequence into the hemagglutinin gene of an influenza virus. *Nature*, **340**, 156–157.

Kierkegaard, K. and Baltimore, D. (1986) The mechanism of RNA recombination in poliovirus. *Cell*, **47**, 433–443.

Kos, A., Kijema, R., Arnberg, A. C., van der Meide, P.H. and Schellekens, H. (1986) The hepatitis delta virus possesses a circular RNA. *Nature*, **323**, 558–560.

Krushkal, J. and Li, W.H. (1995) Substitution rates in hepatitis delta virus. *J. Mol. Evol.*, **41**, 721–726.

Lai, M.M.C. (1992a) RNA recombination in animal and plant viruses. *Microbiol. Rev.*, **56**, 61–79.

Lai, M.M.C. (1992b) Genetic recombination in RNA viruses. *Curr. Topics Microbiol. Immunol.*, **176**, 21–32.

Lai, M.M.C. (1995) Recombination and its evolutionary effect on viruses with RNA genomes. In: *Molecular Basis of Virus Evolution* (eds Gibbs, A.J., Calisher, C.H., and Garcia-Arenal, F.), pp.119–132. Cambridge University Press, Cambridge

Le Gall, O., Lanneau, M., Candresse, T. and Dumez, J. (1995) The nucleotide sequence of the RNA-2 of an isolate of the English serotype of tomato black ring virus: RNA recombination in the history of nepoviruses. *J. Gen. Virol.*, **76**, 1279–1283.

Long, M., deSouza, S.J. and Gilbert, W. (1997) Delta-interacting protein A and the origin of hepatitis delta antigen. *Science*, **276**, 824–825.

Makino, S., Chang, M.F., Sheih C.K. *et al.* (1987) Molecular cloning and sequencing of a human hepatitis delta virus RNA. *Nature*, **329**, 343–346.

Maroney, P.A., Yu, Y.T., Jankowska, M. and Nilsen, T.W. (1996) Direct analysis of nematode *cis*- and *trans*-spliceosomes: a functional role for U5 snRNA in *spliced* leader addition *trans-splicing* and the identification of novel Sm snRNPs. *RNA*, **2(8)**, 735–745.

Mayo, M.A. and Jolly, C.A. (1991) The 5′-terminal sequence of potato leafroll virus RNA: evidence of recombination between virus and host RNA. *J. Gen. Virol.*, **72**, 2591–2595.

Meyers, G., Tautz, N., Dubovi, E.J. and Thiel., H.-J. (1991) Viral cytopathogenicity correlated with integration of ubiquitin-coding sequences. *Virology*, **180**, 602–616.

Monroe, S.S. and Schlesinger, S. (1983) RNAs from two independently isolated defective interfering particles of Sindbis virus contain a cellular tRNA sequence at their 5′-ends. *Proc. Natl Acad. Sci. USA*, **80**, 3279–3283.

Nagy, P.D. and Bujarksi, J.J. (1996) Homologous RNA recombination in brome mosaic virus: AU-rich sequences decrease the accuracy of crossovers. *J. Virol.*, **70**, 415–426.

Nagy, P.D. and Bujarksi, J.J. (1997) Engineering of homologous recombination hotspots with AU-rich sequences in Brome mosaic virus. *J. Virol.*, **71**, 3799–3810.

Nagy, P.D. and Bujarksi, J.J. (1998) Silencing homologous RNA recombination hot spots with GC-rich sequences in Brome mosaic virus. *J. Virol.*, **72**, 1122–1130.

Nagy, P.D. and Simon, A.E. (1997) New insights into the mechanisms of RNA recombination. *Virology*, **234**, 1–9.

Pilipenko, E.V., Gmyl, A.P. and Agol, V.I. (1995) A model for rearrangements in RNA genomes. *Nucl. Acids Res.*, **23**, 1870–1875.

Purcell, R.H. and Gerin, J.L. (1996) Hepatitis delta virus. In: *Fields Virology*, vol 2 (eds Fields, B. N., Knipe, D.M. and Howley, P.M.), pp. 2819–2829. Lippincott-Raven, Philadelphia, PA.

Reanney, D.C. and Ralph, R.K. (1968) Genetic circularity and evolution. *J. Theoret. Biol.*, **21**, 217–228.

Roberts, R.J. (1978) Intervening sequences excised *in vitro*. *Nature*, **274**, 530.

Robertson, H.D. (1992) Replication and evolution of viroid-like pathogens. *Curr. Topics Microbiol. Immunol.*, **176**, 214–219.

Robertson, H.D. and Branch, A.D. (1987) The viroid replication process. In: *Viroids and Viroid-like Pathogens* (ed. Semancik, J. S.), pp. 49–70. CRC Press, Boca Raton, FL.

Semancik, J.S. (ed.) (1987) Viroids and Viroid-like Pathogens. CRC Press, Boca Raton, FL.

Sharp, P. A. (1985) On the origin of RNA splicing and introns. *Cell*, **42**, 397–400.

Taylor, J.M. (1996) Hepatitis delta virus and its replication. In: *Fields Virology*, vol. 2 (eds Fields, B.N., Knipe, D.M. and Howley, P.M.), pp. 2809–2818. Lippincott-Raven, Philadelphia, PA.

Taylor, J., Mason, W., Summers, J. *et al.* (1987) Replication of human hepatitis delta virus in primary cultures of woodchuck hepatocytes. *J. Virol.*, **61**, 2891–2895.

Wang, K.S., Choo, Q.L., Weiner, A. J. *et al.* (1986) Structure, sequence, and expression of the hepatitis delta viral genome. *Nature*, **323**, 508–514.

Zhang, X. and Lai, M.M.C. (1994) Unusual heterogeneity of leader-mRNA fusion in a murine coronavirus: implications for the mechanism of RNA transcription and recombination. *J. Virol.*, **68**, 6626–6633.