3

The Genomic Tag Hypothesis: What Molecular Fossils Tell Us about the Evolution of tRNA

Nancy Maizels and Alan M. Weiner
Departments of Molecular Biophysics and Biochemistry, and of Genetics
Yale University School of Medicine
New Haven, Connecticut 06520-8024

INTRODUCTION

Holley's realization that tRNA could be folded into a two-dimensional cloverleaf posed more questions than it answered (Dudock et al. 1969). One of the most perplexing was whether the three-dimensional structure of tRNA would turn out to be an "integral fold" in which all parts were essential for the correct structure, or whether tRNA could be decomposed into smaller, structurally independent units. The crystal structure of tRNA immediately revealed that tRNA is composed of two perpendicular coaxial stacks (Quigley and Rich 1976): a stack of the acceptor stem on the dihydrouridine stem/loop (the "top half") and a stack of the TψC stem/loop on the anticodon stem/loop (the "bottom half") (see Fig. 1). Remarkably, the covalent connections between the middle of one helical stack and the middle of the other hardly distorted either helical stack: The top and bottom halves of tRNA appeared to be inserted into each other with surgical precision. A great deal of evidence has subsequently shown that the top and bottom halves of tRNA are indeed structurally and functionally independent units. This suggests that the two halves of tRNA could have evolved independently. Here we review the experimental evidence bearing on our hypothesis (Weiner and Maizels 1987) that the top half of tRNA evolved first as a 3′ terminal "genomic tag" that marked single-stranded RNA genomes for replication in what Gilbert was first to call the "RNA World" (Gilbert 1986). The bottom half of tRNA would then have evolved separately as replication in the RNA World became more sophisticated, or as the advent of templated protein synthesis in the RNA World gave birth to the RNP World (Noller 1993; Schimmel et al. 1993).
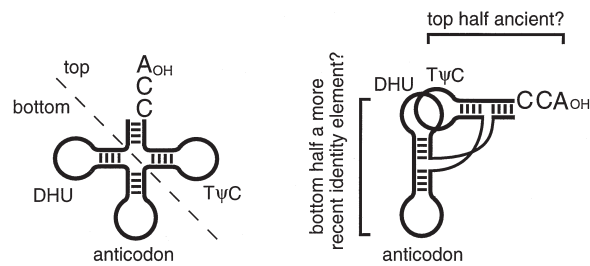
Figure 1   tRNA has two structural domains. The "top half" of tRNA (a coaxial stack of the acceptor stem on the TψC stem) is structurally and functionally independent of the "bottom half" (a coaxial stack of the dihydrouracil and anticodon stems). The top half may also be more ancient, and the bottom half may have evolved as an expansion loop which supplemented the function of the top half without disrupting it.

tRNA Plays a Surprising Number of Roles in Replication

First impressions are often lasting, and it is probably safe to say that most molecular biologists first encounter tRNA as a key component of the translation machinery. This is how tRNA is presented in elementary courses, and this is how the molecule is portrayed in textbooks. Yet because tRNA is commonly introduced as a component of the translation machinery, it is all too easy to think of translation as the *primary* or *proper* function of tRNA. In fact, as we discuss in detail, tRNA and tRNA-like molecules also play key roles in a wide variety of replicative processes including replication of single-stranded RNA viruses of bacteria, plants, and possibly mammals; replication of duplex DNA plasmids of fungal mitochondria; retroviral replication; and replication of modern chromosomal telomeres.

How did tRNA come to have so many different roles in replication? One possibility is that, for reasons which are not yet understood, tRNA or tRNA-like structures have been repeatedly and independently borrowed from translation to serve ad hoc roles in replication. Alternatively, tRNA or tRNA-like structures may be widespread in contemporary replication because tRNA played a central role in the replication of ancient RNA genomes, a role that has been conserved as well as subtly transformed as genomes evolved from RNA to duplex DNA.

We proposed that tRNA-like structural motifs first evolved as 3′-terminal structures that tagged RNA genomes for replication in the RNA World before the advent of protein synthesis (Weiner and Maizels 1987). This hypothesis provides a natural explanation for the ubiquity of tRNA-like motifs by suggesting that tRNA-like structures arose early and played

an essential role in the earliest replicating systems. The central role of the early tRNA-like structures in replication makes it likely that the structural motif was conserved throughout subsequent evolution, and the antiquity of the motif would assure that it was used and reused in many different ways. This simple genomic tag hypothesis has surprising power to organize many apparently unrelated aspects of molecular biology into a coherent whole, and to suggest new relationships between areas of research that appear superficially to be unconnected. As genomic tags enter a second decade, no evidence has appeared that contradicts the hypothesis, and experimental support continues to mount.

Molecular Fossils, Coevolution, and Continuity

The genomic tag hypothesis explicitly argues that molecular evolution has been so conservative as to preserve a role for tRNA-like structures in replication through 3.5 billion years of genomic evolution. Given the sweep of evolution—from the first hopeful prebiotic smudge to the glories and follies of humankind—it may be tempting to believe that the transformations that occur during evolution obliterate the evidence of their own molecular origins, with the result that the most ancient aspects of cellular structure would be *least* likely to survive in recognizable form. However, the more we understand the full complexity of the biological machinery, the clearer it becomes that the components of a living cell interact in so many ways that a change in any key component requires compensatory changes in many others. Thus, most molecules cannot freely evolve to maximize the efficiency of a single function, but must instead *coevolve* with other physically and functionally interacting molecules. When a significant change in one molecule would entail an impossibly large number of simultaneous compensatory changes in others, then the necessity of coevolution can effectively freeze a molecule in time. It was clear even to Lucretius that "Natura non saltus fecit" (Nature never makes a leap). More recently, Orgel (1968) reformulated this old idea as the principle of continuity.

Skeletons, shells, feathers, leaves, and wood are preserved as fossils by death and mineralization. Molecular structures and functions are preserved because the complexity of the living process retards or prevents further significant evolution. White (1976) appears to have been the first to apply the term "fossil" to biochemical processes. We define a *molecular fossil* as any contemporary structure or function that is ancient in origin and provides us with clues about the history of life. We argue here that the ubiquity and conservation of tRNA-like structures strongly suggest

that this motif defines a molecular fossil record of events dating back to the beginnings of life on earth.

The Explanatory Power of the Hypothesis

Molecular biologists are accustomed to experiments that produce instant yes or no answers. The genomic tag hypothesis does not suggest one particular "killer" experiment that could support or falsify it, but the hypothesis has nonetheless inspired a number of experiments and has proved relevant to many others in ways we could not have anticipated. Thus, in the years following the initial genomic tag hypothesis (Weiner and Maizels 1987), a remarkable number of unanticipated experimental results have emerged that directly support, were predicted by, or are consistent with, the hypothesis as originally stated. K.R. Popper (1963) coined the term "explanatory power" to describe the ability of a hypothesis to account for previously unrelated and apparently disparate observations. Not only does the genomic tag hypothesis appear to possess considerable explanatory power, but the hypothesis appears to be robust, comfortably making sense of new data rather than struggling to accommodate unwelcome experimental results. As discussed below, some of these new results include cleavage of modern genomic tags by RNase P (Green et al. 1988; Guerrier-Takada et al. 1988; Mans et al. 1990); identification of major tRNA identity elements within the acceptor stem (Musier-Forsyth and Schimmel 1992); division of tRNA synthetases into at least two unrelated classes (Eriani et al. 1990); a role for tRNA as *template* for reverse transcription of a retroplasmid genome (Akins et al. 1989; Saville and Collins 1990); and the existence of an internal tRNA-like template in telomerase (for review, see Blackburn 1991). A variety of results indicate that RNA can activate and polymerize amino acids: specific binding of an amino acid by RNA (for review, see Yarus et al. 1991); RNA catalysis of reactions at an aminoacylphosphoester center (Piccirilli et al. 1992; Illangesekare and Yarus 1995, 1997); and the ability of RNA to catalyze peptidyl transfer (Lohse and Szostak 1996) as rRNA itself is thought to do (Noller et al. 1992; Green and Noller 1997; Welch et al. 1997). The notion that the top half of tRNA evolved very early is supported by the fact that enzymes as different as RNase P (McClain et al. 1987), tRNA synthetases (Rould et al. 1989), ribosomal RNA (Noller et al. 1992), EF-Tu (Rasmussen et al. 1990), and the archaeal and eubacterial CCA-adding enzymes (Shi et al. 1998b) recognize primarily the top half of tRNA. We take the resilience and predictive power of the hypothesis as evidence that it has been fruitful and may be substantially correct.

THE GENOMIC TAG HYPOTHESIS

The Hypothesis in Outline

We proposed that ancient linear RNA genomes possessed 3′-terminal tRNA-like structures, which we called genomic tags (Weiner and Maizels 1987). Figure 2 shows the simplest form of a tRNA-like genomic tag, a stem and loop immediately followed by a 3′-terminal CCA. This resembles what is sometimes called the top half of tRNA or minihelix, i.e., a coaxial stack of the acceptor stem on the TψC arm (Fig. 1). Like the 3′-terminal tRNA-like motifs of contemporary bacterial and plant RNA viruses (Rao et al. 1989), and possibly animal picornaviruses (Pilipenko et al. 1992), the genomic tag would have served two main roles, providing an initiation site for replication and functioning as a simple telomere.

As an initiation site for replication, the tag would bind to the replicase, ensuring replication of genomic (as opposed to nongenomic and random) RNA molecules. In addition to conferring template specificity, the tag would also sequester subterminal RNA sequences in secondary structure, thereby forcing the replicase to initiate on the 3′-terminal CCA of the genomic RNA. Chemical considerations suggest that the 3′-terminal CC of the CCA sequence could in fact have been *selected* to facilitate efficient and faithful replication. Initiation with guanosine on the penultimate base
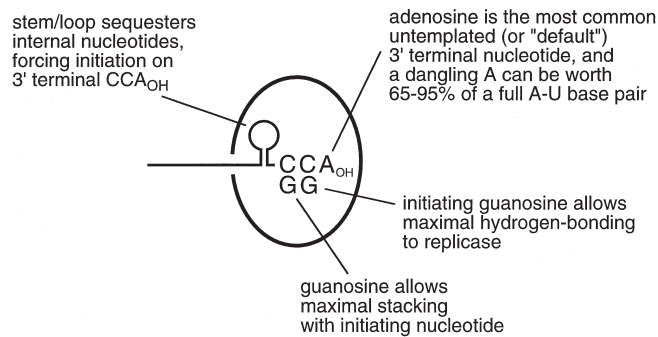
stem/loop sequesters internal nucleotides, forcing initiation on 3′ terminal CCA$_{OH}$

adenosine is the most common untemplated (or "default") 3′ terminal nucleotide, and a dangling A can be worth 65-95% of a full A-U base pair

CCA$_{OH}$
GG

initiating guanosine allows maximal hydrogen-bonding to replicase

guanosine allows maximal stacking with initiating nucleotide

Figure 2   The simplest genomic tag. The tag functions as a simple telomere by sequestering subterminal RNA sequences in secondary structure, thus forcing the replicase to initiate on the 3′-terminal CCA of the genomic RNA. As shown in this figure and described in the text, the 3′-terminal CCA sequence may have been selected to facilitate efficient and faithful initiation of replication with the sequence 5′ GG. The duplex stem of the simplest tag corresponds to the top half of modern tRNA (Fig. 1). Stacking of the 3′-terminal A on the penultimate base pair can account for as much as 65–95% of the stability of a full base pair, and may account for the ability of many polymerases to add an untemplated A (see text).

of the template might be favored because G:C pairs are stronger than A:U pairs; because G has a greater potential than the other bases for hydrogen bonding with the polymerase; and because strong stacking of G on G in the 5′ GG dinucleotide might help to compensate for the absence of a primer (Fig. 2).

As a telomere, the tag would also provide a site for *untemplated* nucleotide addition, ensuring that critical terminal regions of the genome were not lost during replication. For example, if replication initiated on the penultimate G (as Qβ replicase does today; Blumenthal and Carmichael 1979), the 3′-terminal A would be the minimal telomeric sequence, and untemplated addition of this A residue would be required to regenerate the genome before each new round of replication. Remarkably, addition of an untemplated 3′-terminal nucleotide (typically A but occasionally C) is a common activity of both RNA and DNA polymerases made of protein, and thus could also have been a property of an RNA replicase made of RNA. Qβ replicase possesses this terminal transferase activity (Blumenthal and Carmichael 1979); both the bacteriophage T7 (Milligan et al. 1987; Gardner et al. 1997) and SP6 RNA polymerases will add a single untemplated 3′ nucleotide (Melton et al. 1984); and the ability of Taq I DNA polymerase to add an untemplated A to PCR products is notorious (see, e.g., Tse and Forget 1990). In fact, addition of an untemplated (default) 3′-terminal nucleotide may represent an attempt by polymerase to continue polymerization despite the absence of a template nucleotide. Although the incoming nucleotide cannot base-pair with the template strand, the ribose triphosphate moiety can still interact with the enzyme, and the base can still stack on the previous base pair. Stacking of such a "dangling" or untemplated A on the terminal base pair of an RNA duplex can account for as much as 65–95% of the stability contributed by a bona fide base pair (Freier et al. 1985; SantaLucia et al. 1990; Sugimoto et al. 1990; Limmer et al. 1993). Since binding of the ribose triphosphate and stacking of the base on the previous base pair must be universal properties of any replicative polymerase, addition of an untemplated 3′-terminal A (or C) might be expected to occur even if the polymerase were made of RNA.

Finally, although the CC of CCA might be explained by advantageous stacking and hydrogen-bonding interactions during the initiation of replication, and the 3′-terminal A of CCA by default addition of an untemplated nucleotide upon completion of replication, accumulating evidence indicates that the 3′-terminal NCCA sequence of tRNA (where N is the unpaired "discriminator" base) can assume a distinct structure when stacked upon the last base pair of the acceptor stem (Limmer et

al. 1993; Puglisi et al. 1994; Shi et al. 1998a). Indeed, as we discuss below, the *Neurospora* retroplasmid reverse transcriptase mainly recognizes this CCA sequence (Chen and Lambowitz 1997; also see Fig. 4). Thus, the entire NCCA sequence may have been subject to an additional selection for the ability to form a structure immediately adjacent to a 3′-terminal stem/loop that could be easily recognized by the RNA replicase.

Some Implications of the Genomic Tag Hypothesis

The genomic tag hypothesis explains and relates many disparate roles of tRNA motifs in cellular metabolism. In this section, we review some aspects of the hypothesis considered in detail previously (Weiner and Maizels 1987) and discuss data from experiments inspired by or relevant to these facets of the hypothesis. In later sections, we develop new aspects of the hypothesis and discuss relevant new data.

RNase P

RNase P is a ribonucleoprotein enzyme that functions as an endoribonuclease to remove the 5′ leader from tRNA precursors. This processing reaction occurs in all contemporary cells and organelles. In some sense this is a rather surprising reaction, since there is no obvious reason that the 5′ end of tRNA could not be generated directly by transcription, as is known to be the case for *Xenopus laevis* selenocysteine tRNA (Lee et al. 1989).

    We have suggested that the contemporary RNase P activity derives from an ancient activity which arose to convert genomic RNA molecules into functional subgenomic RNAs by removing the 3′-terminal tRNA-like tag, thereby enhancing the structural and enzymatic versatility of functional RNA molecules. Then, in a reversal of fortunes, this enzyme, which had once freed functional RNA from a tRNA-like 3′ tag, survived as an enzyme for removing the nonfunctional 5′ leader from tRNA precursors. Our suggestion that RNase P might have evolved to recognize 3′-terminal tRNA-like genomic tags led directly to experiments that asked, Is the 3′ tRNA-like structure on contemporary plant viruses sufficiently conserved that it can be recognized by *Escherichia coli* RNase P? The answer, remarkably, was yes: *E. coli* RNase P can process the 3′-terminal tRNA-like pseudoknot of turnip yellow mosaic virus (TYMV) at a site corresponding to the 5′ leader of a tRNA precursor (Green et al. 1988; Guerrier-Takada et al. 1988; Mans et al. 1990). The fact that a contemporary eubacterial processing enzyme recognizes the 3′ structures of plant viruses

suggests remarkably conservative coevolution of both the activity itself and the structure of its substrate.

Since the RNA component of RNase P is by itself enzymatically active under appropriate conditions (Guerrier-Takada et al. 1983), it is plausible that early forms of RNase P could have been composed entirely of RNA. Indeed, Alberts (1986) has argued that proteins are such efficient and versatile catalysts that the very presence of RNA in an enzyme may be taken as prima facie evidence that the enzyme is ancient, and might predate the advent of templated protein synthesis. Thus, both the function and composition of RNase P are consistent with preservation of this activity ever since a very early era in molecular evolution. The genomic tag hypothesis plausibly explains why an activity critical for tRNA processing would have arisen so early.

Was the CCA-adding Activity the First Telomerase?

While endonucleolytic cleavage by RNase P generates the mature 5′ end of tRNA, exonucleolytic processing resects the 3′ end of tRNA precursors (Li and Deutscher 1996), and the CCA-adding enzyme [ATP(CTP):tRNA nucleotidyltransferase] then reconstructs the CCA terminus (Sprinzl and Cramer 1979; Deutscher 1982). Once again, this is a rather surprising reaction because the CCA end is destroyed only to be rebuilt. The purpose of this activity becomes clear, however, when the CCA-adding activity is viewed as a molecular fossil. If the 3′ end of a tRNA-like structure served as the initiation site for an early replicase, a CCA-adding activity would have been necessary to restore nucleotides lost as a result of incorrect initiation. The CCA-adding activity therefore would have functioned as the first telomerase, and the modern CCA-adding activity may have evolved from an RNA enzyme into a protein enzyme by stepwise replacement of RNA by protein (White 1976, 1982; Visser and Kellogg 1978a,b). Indeed, if the CCA-adding activity is ancient and essential, it should be present in all three living kingdoms today. Consistent with this hypothesis, we found that archaea, like eubacteria and eukaryotes, also contain a CCA-adding enzyme (Yue et al. 1996). The archaeal, eukaryotic, and eubacterial CCA-adding enzymes all belong to the same nucleotidyltransferase superfamily (Holm and Sander 1995; Martin and Keller 1996; Yue et al. 1996) and are functionally very similar (Shi et al. 1998a), but the archaeal CCA-adding enzyme diverges completely from the closely related eukaryotic and eubacterial enzymes outside the nucleotidyltransferase active-site signature sequence (Yue et al. 1996). One explanation for sequence divergence at the protein level would be descent from a common ancestral ribozyme in the RNA World (Weiner and Maizels 1987; Benner et al. 1989). As discussed

below, use of a telomeric C_mA_n motif by early RNA genomes may also explain why modern DNA telomeres employ very similar sequence motifs.

tRNA Synthetase Function and the Origin of Protein Synthesis

We have argued that the series of reactions required for aminoacylation of tRNA chemically resembles RNA polymerization (Fig. 3) (Weiner and Maizels 1987). This prompted us to suggest that 3′-terminal tRNA-like structures of ancient RNA genomes may have been aminoacylated in a reaction resembling the charging of contemporary tRNAs by aminoacyl tRNA synthetases. In making this suggestion, we explicitly postulated that the enzyme responsible for aminoacylation was an RNA, and that this RNA enzyme could bind both an amino acid and a mononucleotide and could catalyze reactions at an aminoacylphosphoester center. We also suggested that RNA, as a structured polyanion, would most likely bind basic amino acids, and that charging of a genomic tag with a basic amino acid would have the greatest effect on the structure or function of the RNA. The ability of group I introns to bind the mononucleotide guanosine was known at the time (for review, see Michel et al. 1989). It has since been shown that group I introns will bind L-arginine specifically (Hicke et al. 1989); that the group I guanosine-binding site can be rationally redesigned to bind other nucleotides (Michel et al. 1989); that a group I intron can, without significant redesign, function as a credible aminoacylphosphoesterase (Piccirilli et al. 1992); and that RNAs can be selected that resemble the suspected peptidyl transferase site on the large rRNA (Welch et al. 1997), aminoacylate the 3′ hydroxyl of RNA
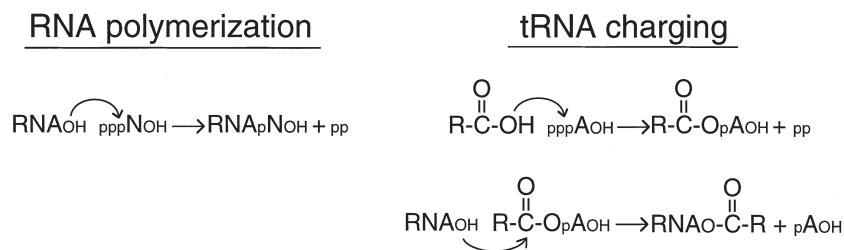
RNA polymerization   tRNA charging

$RNA_{OH} \; pppN_{OH} \longrightarrow RNApN_{OH} + pp$

$\overset{O}{\overset{\|}{R\text{-}C\text{-}OH}} \; pppA_{OH} \longrightarrow \overset{O}{\overset{\|}{R\text{-}C\text{-}O}}pA_{OH} + pp$

$RNA_{OH} \; \overset{O}{\overset{\|}{R\text{-}C\text{-}O}}pA_{OH} \longrightarrow RNAO\text{-}\overset{O}{\overset{\|}{C}}\text{-}R + pA_{OH}$

Figure 3   tRNA charging resembles RNA polymerization. Attack on the phosphoester bond proceeds by an in-line, S_N2 mechanism, resulting in a trigonal bipyramidal transition state, whereas attack on the aminoacylphosphoester bond occurs orthogonal to the π face, resulting in a tetrahedral intermediate. Despite these differences in stereochemistry, a ribozyme that catalyzes phosphoester bond transfer can also catalyze reactions at the carbon center, albeit at reduced efficiency (Piccirilli et al. 1992).

(Illangesekare et al. 1995, 1997), and possess peptidyl transferase activity (Lohse and Szostak 1996).

The genomic tag hypothesis suggests that aminoacylation initially conferred a *replication advantage* on molecules carrying a genomic tag. This could have occurred in any of several ways. Aminoacylation might have facilitated binding of the replicase to the $3'$ end of the genome, perhaps simply by countering the net negative charge of the RNA replicase with a positively charged (basic) amino acid. Aminoacylation could also have served as a regulatory mechanism for withdrawing a genomic RNA from the replicative pool by blocking binding of the tag to the replicase. A third possibility, suggested by Wong (1991), is that aminoacylation might be seen as a form of RNA modification (like methylation, thiolation, isopentenylylation) that would broaden the structural or catalytic range of the RNA bearing it.

Our model for the origin of protein synthesis is unique in postulating that key components of the translation apparatus—tRNA and tRNA aminoacylation activity—first evolved as essential components of the *replication* machinery before the advent of protein synthesis. With these two key components of the translation apparatus in place, the scene was set for the interdependent coevolution of replication and templated protein synthesis. Perhaps at first, random condensation of aminoacylated tRNAs generated short polycations. These simple polymers could have facilitated RNA-catalyzed reactions in a manner analogous to modern polyamines (Jay and Gilbert 1987; Maizels and Weiner 1987; Weiner and Maizels 1987) or they could have stabilized or promoted a particular RNA structure, much as the tract of basic amino acids within the HIV Tat protein shapes the structure of the TAR element (Puglisi et al. 1992). The ability of polyamines to regulate translational frameshifting on the ornithine decarboxylase antizyme mRNA also provides additional, albeit less direct, evidence that short polycations might have been useful in an RNA World (Matsufuji et al. 1995). Without several different species of charged tRNA, however, no mRNA would be necessary or (for that matter) useful. We therefore speculated that the original protoribosome evolved to facilitate synthesis of specific peptides by aligning charged tRNAs *before* the advent of mRNA, and that templating of peptide synthesis by mRNA was one of the *last* steps in the evolution of the modern translation apparatus (Maizels and Weiner 1987; also see Schimmel and Henderson 1994). This scenario is consistent with recent evidence that contemporary tRNAs comprise two separate structural domains (see below). We also emphasize that borrowing tRNA from replication for protein synthesis might constrain, but would not preclude, further evolution of the role of tRNA in replication.

The Unexpected Diversity of tRNA Synthetases

A further prediction of the genomic tag model for the origin of protein synthesis is that at least some tRNA synthetases were originally RNA enzymes (Weiner and Maizels 1987). We originally suggested that this might account for the remarkable diversity in structure of modern synthetases already apparent at the time (Schimmel 1987) because there is unlikely to be a unique pathway by which an RNA enzyme would evolve into a protein enzyme by gradual replacement of RNA structures with protein structures (White 1976, 1982; Visser and Kellogg 1978a,b). Thus, we explicitly proposed that the existence of apparently unrelated protein enzymes carrying out the same essential function could be taken as strong evidence for an ancestral RNA enzyme. This notion was further developed by Benner et al. (1989).

Despite the apparent diversity of tRNA synthetase subunit structure and sequences (Schimmel 1987), the suggestion that there might be more than one class of synthetase was not well received, perhaps because it was so difficult to imagine that 20 tRNA synthetases doing the same enzymatic job could have descended from more than one ancestral form. However, Eriani et al. (1990) subsequently found that tRNA synthetases could be partitioned into at least two classes based on mutually exclusive sets of sequence motifs. Class I has the Rossmann fold for nucleotide binding, class II an antiparallel β sheet (Cusack et al. 1990); class I acylates the $2'$ hydroxyl, class II the $3'$ hydroxyl; yet the two classes employ virtually identical reaction mechanisms involving an enzyme-bound aminoacyl-adenylate intermediate derived from ATP. From these structural and functional data one must conclude either that (1) tRNA synthetases made of protein evolved twice but independently adopted the same reaction mechanism; or, as we proposed originally, (2) the first tRNA synthetases were made of RNA and evolved stepwise by distinct pathways through RNP intermediates into two ancestral tRNA synthetases made of protein. In the second scenario, evolution of an RNA enzyme into a protein enzyme would have been constrained to preserve the two-step charging mechanism that first evolved in the RNA World (White 1976, 1982), but not charging of the $2'$ or $3'$ hydroxyl, as these two aminoacylated products equilibrate quickly.

tRNA as Genetic Punctuation in RNA Processing and Translation

If tRNA or tRNA-like structures functioned as $3'$-terminal genomic tags in an RNA World, the principle of continuity suggests that tRNA would continue to play a role in genomic organization as RNA genomes evolved into transitional genomes made of both RNA and DNA, and eventually

into DNA genomes. In fact, tRNA coding regions are well known to serve as intergenic "punctuation" in eubacterial rRNA genes and mitochondrial genomes: RNase P then cleaves the primary transcripts at the mature 5′ end of each tRNA. This form of genomic organization could be either ancestral (reflecting an ancient origin) or derived (more recently acquired). tRNA punctuation in eubacterial rRNA genes would be ancestral if the 16S, 5S, and 23S components of the rRNA arose as separate RNA genomes each bearing a 3′-terminal genomic tag that was removed by RNase P. Genetic linkage of these Ur-rRNA subunits end-to-end either as one large RNA, or ultimately as DNA, might then preserve the original RNA processing strategy. On the other hand, tRNA punctuation in eubacterial organellar genomes may be more likely to represent "devolution" in which the complete genome of a free living endosymbiont was gradually reduced to rudimentary status, and RNA processing came to rely on a handful of essential enzymes.

tRNA can also serve as punctuation in translation. 10Sa RNA is a small, stable, highly conserved RNA found in diverse eubacteria and some eubacterial endosymbionts (Subbarao and Apirion 1989; Brown et al. 1990; Tyagi and Kinger 1992; Williams and Bartel 1996). The mature 10Sa molecule is generated by RNA processing from a larger transcript, and the 3′ end of 10Sa RNA almost perfectly matches the conserved residues in the top half of tRNA including the CCA (Tyagi and Kinger 1992). This originally suggested the intriguing possibility that 10Sa RNA might be the genomic or subgenomic transcript of a new retroelement. However, 10Sa RNA was subsequently found to be a "tmRNA"—a single molecule combining both tRNA and mRNA functions (Komine et al. 1994; Tu et al. 1995; Keiler et al. 1996; Williams and Bartel 1996; Felden et al. 1997; Himeno et al. 1997). tmRNA is charged with alanine by alanine tRNA synthetase, enters the A site of a ribosome stalled at the 3′ end of a broken mRNA, and adds a single untemplated alanine residue; the tmRNA then undergoes a conformational change that enables it to function as a bona fide mRNA encoding a 10-residue carboxy-terminal peptide tag that targets the potentially harmful amino-terminal protein fragment for rapid degradation. The 5′ and 3′ ends of 10Sa RNA come together to form the tRNA[Ala]-like structure, with the decapeptide coding region in between (Williams and Bartel 1996; Felden et al. 1997). The remarkable ability of this tRNA-like structure to enter the ribosomal decoding site in the absence of the corresponding anticodon suggests that pairing between a 3′ genomic tag and the 5′ end of the same RNA could have served in the RNA World as a signal for the initiation of translation, and may have been the predecessor of modern initiator tRNA.

The Two Structural Domains of Contemporary tRNA May Have Evolved Independently

Contemporary tRNAs are composed of two structural domains, a top half consisting of a coaxial stack of the acceptor stem on the TψC arm, and a bottom half consisting of a coaxial stack of the dihydrouracil arm on the anticodon arm (Fig. 1). The genomic tag hypothesis suggested the intriguing possibility that these two structural domains evolved independently. When, how, or why the anticodon domain evolved is currently one—if not *the*—unsolved mystery of molecular evolution.

At present, all available experimental evidence is consistent not only with independent evolution of the top and bottom half domains, but also with the top half of the molecule arising early and the bottom half arising later. The fact that enzymes as different as RNase P (McClain et al. 1987), tRNA synthetases (Rould et al. 1989), ribosomal RNA (Noller et al. 1992), EF-Tu (Rasmussen et al. 1990), and the archaeal and eubacterial CCA-adding enzymes (Shi et al. 1998b) recognize primarily the top half of tRNA suggests that this is, indeed, the more ancient half of the molecule containing the most essential identity elements. Presuming that RNase P is ancient, the ability of RNase P to recognize the top half of tRNA alone (McClain et al. 1987) can be interpreted as evidence that the original genomic tags may have been as simple as the top half of the molecule (Fig. 2). Charging of such a simple genomic tag with different amino acids might have arisen initially to facilitate differential replication, modification, or processing of genomic RNAs; alternatively, differential charging could have arisen to facilitate synthesis of new (or more precise) peptides. In either case, there would have been selective pressure to distinguish different tags from each other, and this is consistent with functional (Francklyn et al. 1992) and structural studies (Rould et al. 1989) demonstrating that critical tRNA identity elements can lie within the top half of the molecule. In particular, minihelices corresponding to the top half of tRNA are sufficient in some cases for highly specific charging (Francklyn et al. 1992; Musier-Forsyth and Schimmel 1992), and a fragment of the top half is a substrate for the protein-free peptidyl transferase activity (Noller et al. 1992). Further evidence for the potential complexity of tRNA identity elements within the top half of the molecule comes from crystallographic work showing that critical identity elements are revealed by partial melting of the acceptor stem when *E. coli* glutamine tRNA binds to the cognate synthetase (Rould et al. 1989).

The bottom half of tRNA appears to be a more recent addition, used both by synthetases and by mRNAs to distinguish one species of tRNA from another. In this view, the bottom half of tRNA can be thought of as

an expansion "loop," similar to those found in ribosomal RNA (see chapters 8 and 15). The notion that the bottom half of the tRNA molecule evolved after the top half is also consistent with our suggestion that templated protein synthesis, requiring tRNA–mRNA recognition mediated by the anticodon loop, could not have been selected until relatively late in the evolution of the translation apparatus when there were already several different species of charged tRNA to read the mRNA (Maizels and Weiner 1987).

TRACING GENOMIC TAGS FROM RNA PHAGE TO MODERN CHROMOSOMAL TELOMERES: THE NOTION OF TRANSITIONAL GENOMES

Although we suspected that tRNA-like genomic tags survived from an RNA World into a DNA World and were, in the process, transformed into the tRNA primers of retrovirus reverse transcription and the terminal C_mA_n motifs of modern chromosomal telomeres (Weiner and Maizels 1987), we were unable to make these connections explicit because there appeared to be one or more missing links in the molecular fossil record. Stunning and completely unanticipated results from the groups of Lambowitz (Kuiper and Lambowitz 1988; Akins et al. 1989; Chen and Lambowitz 1997), Saville and Collins (1990), Blackburn (for review, see Blackburn 1991), and Cech (Lingner et al. 1997) subsequently provided key molecular fossil evidence for genomes and genomic tags in transition from single-stranded RNA to double-stranded DNA. In this section we discuss the *Neurospora* retroplasmid and the *Tetrahymena* telomerase, two critical links that were missing from the molecular fossil record as it was known when we first formulated the genomic tag hypothesis (Weiner and Maizels 1987).

Neurospora Mitochondrial Retroplasmids: Evidence That Genomic Tags Survived the Transition from RNA to DNA Genomes

Mitochondria of some strains of *Neurospora crassa* contain double-stranded DNA plasmids that were originally detected because their replication causes a respiratory-deficient phenotype by competing with replication of mitochondrial DNA. Most surprisingly, replication of these plasmids requires production of a full-length RNA transcript of genomic DNA (Figs. 4 and 5) (Kuiper and Lambowitz 1988; Akins et al. 1989), and this RNA carries a 3′-terminal tRNA-like structure ending in CCACCA, implying that it functions as a tRNA-like genomic tag (Akins et al. 1989). An open reading frame within the genomic RNA encodes a
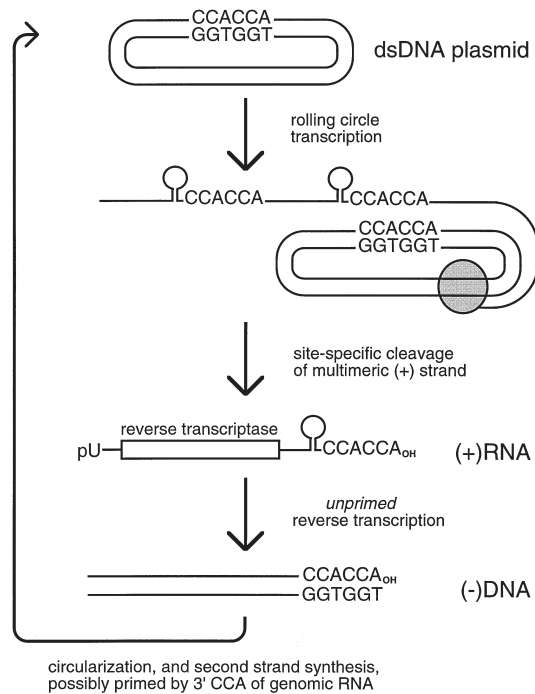
Figure 4   The *Neurospora* mitochondrial Varkud plasmid is a simple retroviral-like element. Note that first-strand cDNA synthesis uses the tRNA-like structure as template, whereas second-strand synthesis and/or circularization may reuse the tRNA-like structure as primer.

functional reverse transcriptase, and mitochondrial extracts of plasmid-containing strains can produce a full-length cDNA copy of the genomic RNA by de novo initiation of cDNA synthesis using the terminal A of the CCACCA as template (Chen and Lambowitz 1997). The 3′-terminal genomic tag therefore serves as the initiation site for replication as in Qβ, but Qβ replicase is replaced by the plasmid reverse transcriptase (Fig. 5). Whether the full-length RNA/DNA hybrid produced after the first step in plasmid replication circularizes directly, or is first converted into double-stranded DNA, is not yet known; the ability of the *Neurospora* reverse transcriptase to use both specific and nonspecific DNA primers (Chen and Lambowitz 1997) suggests that the enzyme itself might be responsible for circularizing the linear replicative intermediate. Once a circular, duplex DNA replicative intermediate has formed, rolling circle *transcription* yields multimeric plus strands, which are then cleaved into monomeric genomic RNAs. At least in the case of the Varkud-associated VSDNA
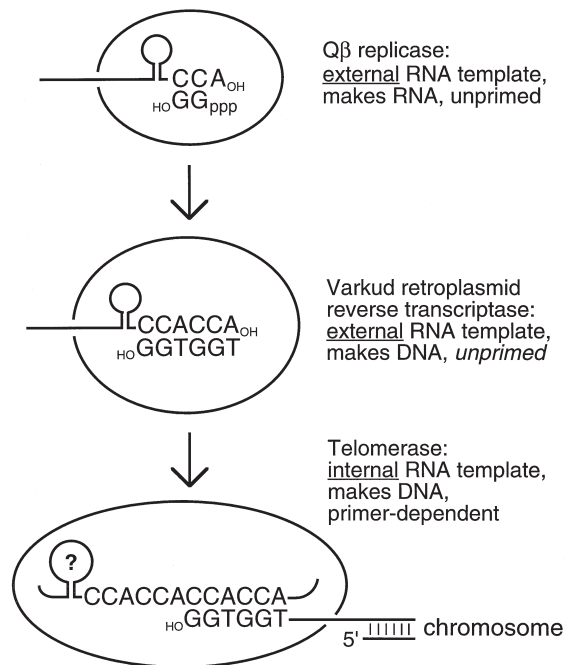
Figure 5 From Qβ replicase to telomerase. Qβ replicase, the *Neurospora* retroplasmid reverse transcriptase, and *Tetrahymena* telomerase all initiate on a tRNA-like template, suggesting that the ancestral forms of these enzymes represent a single line of evolutionary descent.

retroplasmid, a ribozyme contained within the RNA genome itself appears to perform this cleavage reaction (Saville and Collins 1990).

The fungal mitochondrial retroplasmids can be thought of either as linear RNA genomes that replicate through a circular DNA intermediate, or circular DNA genomes that replicate through a linear RNA intermediate. We therefore consider these retroplasmids to be examples of *transitional genomes*, contemporary representatives of an era in which genomic tags were being transformed from templates for RNA synthesis to templates for DNA synthesis.

Telomerase: A Genomic Tag Carried by the RNA Component of a Reverse Transcriptase

When we first suggested that the $3'$-terminal CCA motif of tRNA-like genomic tags was related to the nearly universal C_mA_n motif of eukaryotic

nuclear telomeres (Weiner and Maizels 1987), we were unable to discern any hints in the molecular fossil record regarding the mechanism or chain of events by which a $3'$-terminal CCA motif in RNA could be transformed into a $5'$-terminal motif in DNA. Another missing link in the molecular fossil record emerged from detailed characterization of the *Tetrahymena* telomerase. This enzyme, first described by Greider and Blackburn (1985, 1989), possesses an unusual terminal nucleotidyltransferase activity that adds the species-specific T_nG_m repeats, one nucleotide at a time, to an appropriate T_nG_m primer (for review, see Blackburn 1991). Remarkably, telomerase is a ribonucleoprotein, and the RNA component is in fact an *internal* template for synthesis of the species-specific T_nG_m repeat. (For example, the telomeric repeat in the ciliate *Tetrahymena* is T_2G_4, and the internal template sequence in the *Tetrahymena* telomerase is $5'$-CAACC-CCAA-$3'$.) One of the protein components of telomerase, known in *Saccharomyces cerevisiae* as *EST2* (*e*ver *s*horter *t*elomeres), is homologous to retroviral reverse transcriptases (Lingner et al. 1997). Telomerase can thus be seen as a specialized reverse transcriptase with an internal tRNA-like template (for review, see Blackburn 1991 and this volume). The secondary and tertiary structures of telomerase RNA have been studied in detail (Romero and Blackburn 1991; Bhattacharyya and Blackburn 1994; Lingner et al. 1994), and all known telomerase RNAs share a repetitive CCA-like internal template sequence; however, the rest of telomerase RNA is not well conserved across large evolutionary distances.

The Unusual Properties of G and G-rich DNAs

The unusual properties of G and of G-rich DNA suggest that chemical determinism may explain both the C_nA_m motif in RNA telomeres and use of the T_mG_n motif in modern DNA telomeres (Weiner and Maizels 1987). Telomeric T_nG_m repeats and other G-rich DNAs can spontaneously form unusual four-stranded structures in vitro that contain *intra*molecular non-Watson-Crick base pairs (Henderson et al. 1987). These structures, dubbed G quartets, may play a role in telomere aggregation or function in vivo (Sen and Gilbert 1988; Williamson et al. 1989; Sundquist and Klug 1989; Fang and Cech 1993a,b; for review, see Blackburn 1991; Zakian 1996). As detailed in Figure 2, the sequence CCA might have been selected in an RNA World as an efficient initiation site because the initiating guanosines stack more strongly and have more potential hydrogen-bonding interactions than any other bases. In effect, the CC of CCA would have been selected for the unusual stacking and hydrogen-bonding properties of its G-rich complement. Viewed in this way, the complementary

T_mG_n motif of modern telomeres was selected in the RNA World, but more fully exploited in the DNA World.

Functional Similarities between Telomerase and the Retroplasmid Reverse Transcriptase

The *Neurospora* retroplasmid reverse transcriptase copies genomic RNA into a cDNA by de novo initiation on the terminal nucleotide of the 3′-terminal CCACCA template sequence (Chen and Lambowitz 1997; see also Fig. 4). Telomerase uses telomeric DNA as primer to copy a CCACCA-like template within the enzyme itself into a cDNA. This provides a striking link between telomerase and the retroplasmid reverse transcriptase (Fig. 5). Both enzymes copy RNA into DNA and, like Qβ replicase, both initiate on the CCA motif of a tRNA-like template.

Might there be a single line of evolutionary descent for these enzymes and the genomes they replicate? Figure 5 shows how a tRNA recognition domain that is critical for initiation may be conserved among these enzymes, and how the polymerase domain would interact with this recognition complex. Figure 5 also suggests a simple pathway by which an enzyme like the retroplasmid reverse transcriptase could evolve into a specialized reverse transcriptase like telomerase. The retroplasmid enzyme would capture an internal template by stably binding a tRNA or tRNA-like structure; any primer terminus that could pair with this internal template would then allow the reverse transcriptase to enter directly into elongation mode. Indeed, as mentioned above, the *Neurospora* reverse transcriptase can bind the 3′-terminal CCA (inboard sequences help but are not essential) and can use specific as well as nonspecific primers (Chen and Lambowitz 1997). Thus, telomerase RNA appears to be derived from a captured 3′-terminal genomic tag, and telomerase action can therefore be regarded as abortive replication of a genome that has been reduced to a bare 3′-terminal C_nA_m motif. This may be mechanistically similar to abortive initiation by many RNA polymerases—the repeated synthesis of short initiating oligonucleotides before the polymerase enters elongation mode (McClure 1980). The ability of the *Neurospora* reverse transcriptase to initiate de novo, as well as to use specific and nonspecific primers (Chen and Lambowitz 1997), suggests that primer dependence would not be a major constraint on the evolutionary scenario (Fig. 5). Although retroviral reverse transcriptases require a primer and Qβ replicase does not, the initiating nucleotide can plausibly be regarded as a very short primer because it must occupy the same site on the enzyme as the primer terminus during elongation. Indeed, early work on *E. coli* RNA

polymerase showed that a primer as short as a dinucleotide dramatically increases the efficiency and specificity of initiation (Downey et al. 1971; Maizels 1973), and more recent structural and mechanistic studies on DNA polymerase suggest that the initiating nucleotide occupies the same site on the enzyme as the primer terminus during elongation (Polesky et al. 1992).

Nonviral Retrotransposons: Parents or Children of Telomerase?

Eukaryotes from yeasts to humans have telomeric T_mG_n repeats templated by telomerase RNA, but diptera are a surprising and curious exception in which telomerase activity has apparently been lost (Levis et al. 1993; Biessmann and Mason 1997; Pardue et al. 1997). Instead, the chromosomes in *Drosophila melanogaster* are capped by either of two different kinds of nonviral retrotransposons, HeT-A (*het*erochromatin-*a*ssociated) and TART (*t*elomere-*a*ssociated *retro*transposon). TART encodes a reverse transcriptase, but HeT-A does not; both retroelements are also found in pericentromeric heterochromatin. Complete Het-A and TART elements are transcribed to yield full-length polyadenylated genomic RNAs, and these in turn retropose to the 3′ end of the chromosomes. The 3′ end of chromosomal DNA may serve repeatedly as primer for reverse transcription of these polyadenylated genomic RNAs, thus generating tandem telomeric arrays of HeT-A and TART elements. Occasional retroposition of these 6-kb elements continually replenishes terminal chromosome sequences, balancing loss of about 50–100 bp of telomeric sequence per fly generation (Levis 1989). Thus, diptera have devised a mechanism for telomere maintenance that is strikingly similar to telomerase; in each case, terminal DNA sequences are regenerated by copying an RNA template—a complete genomic RNA in diptera, or an isolated genomic tag in telomerase.

Pardue et al. (1997) argue that cellular telomerase may have given rise to parasitic Het-A, TART, and similar non-LTR retrotransposons (Danilevskaya et al. 1997; Pardue et al. 1997). We favor the alternative scenario, equally consistent with the phylogenetic data (Eickbush 1997; Nakamura et al. 1997), that these retrotransposons are descendants of *autonomous* transitional genomes which had both RNA and DNA replicative forms (Fig. 6). Modern eukaryotic DNA chromosomes would then have been built by stepwise assembly of these smaller *independent* genomic elements. Indeed, the ability of Het-A and TART elements to generate tandem linear multimers (Danilevskaya et al. 1997) could be viewed as a continuation of this ancient chromosome assembly process. The his-
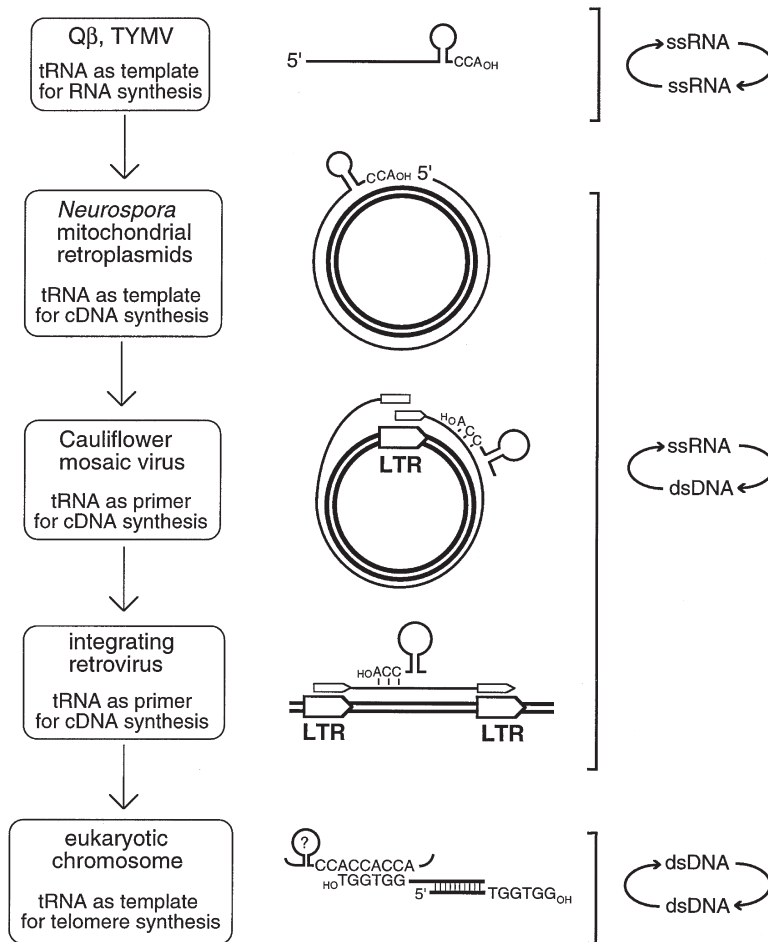
Figure 6   A phylogenetic tree for replication strategies based on conservation of tRNA-like structures in the initiation of genomic replication.

torical roots of modern DNA chromosomes in such transitional genomes would then be apparent in two conspicuous molecular fossils: RNA serves as *primer* for the initiation of DNA synthesis (at least in eubacteria), and RNA serves as *template* for the completion of chromosomal replication (eukaryotic telomeres). Indeed, as mentioned earlier, Alberts (1986) has argued that the presence of RNA in an enzyme suggests that the enzyme is ancient. A friendly amendment might be that involvement of RNA in any DNA transaction (replication, recombination, repair, or modification) suggests that the process is ancient, dating back to an RNA World or a world in transition from RNA to DNA genomes.

FROM TEMPLATE TO PRIMER

tRNAs Prime Retroviral Reverse Transcription

Initiation of retrovirus replication normally requires a tRNA primer (see Fig. 6); the hepadnaviruses are the only conspicuous exception (Ganem and Varmus 1987). tRNA priming of reverse transcription occurs not only in prototypical avian and mammalian retroviruses, but also in such lower eukaryotic retroviral elements as Ty1 in the yeast *S. cerevisiae* (Chapman et al. 1992), and *copia* in *D. melanogaster* (Kikuchi et al. 1986). To function as a primer, the 3′ end of the tRNA partially unfolds to base-pair with the primer-binding site on the genomic RNA (Fig. 6). Additional interactions between the tRNA primer and nearby genomic sequences stabilize this initiation complex, providing multiple sequence determinants that restrict each retroviral element to a specific tRNA primer (Isel et al. 1995; Lanchy et al. 1996). As discussed below, tRNA also primes reverse transcription of circular extrachromosomal retroviruses such as cauliflower mosaic virus (CaMV) (Hohn et al. 1985; Covey and Turner 1986).

When we originally proposed that tRNA priming of retroviral replication was derived from the use of tRNA-like structures as templates for the initiation of RNA replication (Weiner and Maizels 1987), we were unable to cite any transitional forms in the molecular fossil record. The life cycle of the *Neurospora* retroplasmid has now revealed plausible missing links. Initially, a connection between tRNA and reverse transcription was established by the discovery that a 3′-terminal tRNA-like structure serves as template for first-strand cDNA synthesis on the retroplasmid genomic RNA (Figs. 4 and 5). More recently, Chen and Lambowitz (1997) established a connection between tRNA as template and tRNA as primer: The *Neurospora* retroplasmid reverse transcriptase can also use the 3′-terminal CCA of the RNA genome as a specific (base paired) or nonspecific (unpaired) primer (Fig. 7). Thus, first-strand synthesis by the *Neurospora* reverse transcriptase resembles Qβ replicase in using a tRNA-like structure as template; during second-strand synthesis and/or circularization, the same reverse transcriptase apparently reuses the tRNA-like structure as primer.

**In Most Polymerases, Distinct Protein Domains Are Responsible
for Template Specificity and Catalysis**

How surprised should we be that one enzyme can use the same molecule as both template and primer? A wealth of molecular and structural data is consistent with the notion that the two key functions of any polymerase—template specificity and catalysis—are typically carried out by distinct

tRNA-like structure serves as template for first strand cDNA synthesis in the retroplasmid life cycle

tRNA-like structure may serve as primer for second strand cDNA synthesis
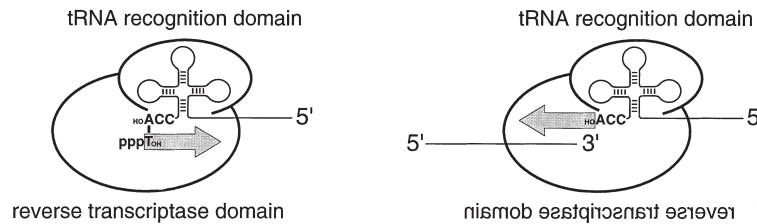
Figure 7 From template to primer. The transformation of tRNA from template for initiation of replication to primer requires that the tRNA-binding domain be functionally separate from, or flexibly tethered to, the catalytic domain. Reverse lettering in the right-hand panel indicates the opposite orientation of the tRNA recognition and reverse transcriptase domains. Note also that the *Neurospora* retroplasmid reverse transcriptase uses a tRNA-like structure both as template and as primer (Chen and Lambowitz 1997; also see Fig. 4).

protein domains. The classic examples are σ factors, which confer promoter specificity on eubacterial RNA polymerases (see, e.g., Decatur and Losick 1996). Separability of template specificity and catalysis is also a property of Qβ replicase and of the generic RNA replicase encoded by poliovirus. The Qβ replicase holoenzyme consists of four subunits (Blumenthal and Carmichael 1979), a phage-encoded replicase (subunit II) and three host-encoded subunits (EF-Tu, EF-Ts, and ribosomal protein S1). Subunit II exhibits a generic RNA replicase activity, whereas template specificity lies entirely within the other subunits, with EF-Tu presumably recognizing the $3'$-terminal tRNA-like structure just as it recognizes tRNA during protein synthesis (Weiner and Maizels 1987, but see Brown and Gold 1996). Similarly, the poliovirus replicase polypeptide $3D^{pol}$ appears to require an additional specificity factor for initiation on the polioviral minus-strand template (Andino et al. 1990; R. Andino, pers. comm.).

Additional evidence that the template specificity and catalytic functions of many polymerases are separable comes from analysis of sequence motifs shared by viral RNA-dependent RNA polymerases (replicases) and retroelement RNA-dependent DNA polymerases (reverse transcriptases). These enzymes display a wide range of template and primer specificity, but nonetheless share essentially invariant signature sequence motifs which are located at the active site rather than in the template recognition domains (Poch et al. 1989; Steitz 1998). Indeed, if template specificity

and catalytic function were not at least partially separable, changes in template specificity would have obliterated these characteristic signatures. In contrast, a polymerase in which template specificity and catalysis are separate functions can easily adapt to new templates, thereby diversifying and ultimately producing more molecular descendants. Thus, as long as the tRNA recognition domain of reverse transcriptase is flexibly tethered to the active site, the enzyme might readily evolve from using tRNA as template to using it as primer. Although HIV reverse transcriptase is only distantly related to the *Neurospora* enzyme, the ability of the HIV enzyme to form a binary complex with primer tRNA (Barat et al. 1989) is consistent with the idea that separation of tRNA recognition and the polymerase module may be a general phenomenon.

Use of tRNA as primer also necessitates RNA helicase activity, because the enzyme would have to melt the top half of tRNA (a coaxial stack of the acceptor stem on the TψC arm) (Figs. 1 and 2) in order to allow the primer to base-pair with the genomic primer-binding site (see Fig. 6). However, this helicase activity would already be in place, because any enzyme that uses a genomic tag as template must possess an RNA helicase activity that can melt the top half of tRNA once initiation has occurred on the 3′-terminal CCA. Extensive complementarity between the tRNA primer and the template (Isel et al. 1995; Lanchy et al. 1996) may be a later refinement to assure a unique site of initiation and to stabilize the initiation complex.

Caulimoviruses Are Non-integrating One-LTR Retroviruses

We have discussed several reasons for thinking that mitochondrial retroplasmids may represent the ancestral form of modern retroviruses, but there is yet another missing link between duplex DNA plasmids and the prototypical integrating retrovirus with long terminal direct repeats (LTRs). A plant retrovirus, cauliflower mosaic virus (CaMV), may provide this connection. The CaMV genome is an extrachromosomal circular duplex DNA that lacks direct repeats and never integrates into chromosomal DNA. Transcription of viral DNA generates a full-length genomic RNA, but because the polyadenylation site for this transcript is located 180 bp downstream from the viral promoter, the transcript contains a 180-nucleotide terminal redundancy (Hohn et al. 1985; Covey and Turner 1986; also see Fig. 6).

Although CaMV (a pararetrovirus) differs from prototypical retroviruses in having one rather than two LTRs in the duplex DNA form, CaMV uses exactly the same strategy to generate a terminally redundant genomic RNA. In both cases, a polyadenylation site is ignored when it oc-

curs too near the 5′ end of the RNA, apparently because RNA processing of the nascent transcript is inhibited by proximity to the promoter (Sanfacon and Hohn 1990; but see DeZazzo et al. 1992). As in prototypical two-LTR retroviruses, the tRNA primer binding site in CaMV genomic RNA lies downstream from the 5′-terminal redundancy, so reverse transcriptase can effectively circularize the genome by switching templates from the 5′ end of the genomic RNA to the corresponding position in the 3′-terminal redundancy.

These similarities in genomic structure and replication strategy imply that modern integrating retroviruses descended from an extrachromosomal retroviral element resembling CaMV. The obvious advantage of chromosomal integration is that it assures perpetuation of the element. The existence of two LTRs in a prototypical retroviral provirus could then be interpreted as an invention that preserved the established replication strategy by counterfeiting the circular topology of the ancestral genome.

The origin of LTRs remains mysterious. An intriguing scenario advanced by Danilevskaya et al. (1997) is that LTRs were first generated by repeated retroposition of a linear polyadenylated RNA such as HeT-A. As described above, contemporary HeT-A elements integrate head to tail, generating telomeric tandem repeats that are topologically equivalent to a circular HeT-A genome. Remarkably, the HeT-A promoter is located near the 3′ end of the element, just upstream of the polyadenylation signal, in an arrangement reminiscent of modern LTRs (Fig. 6). As a result, each HeT-A element in a tandem array provides the promoter for the next element downstream (promoting in tandem) and generates a terminally redundant, polyadenylated primary transcript. Circularization of a single repeat unit from the tandem array could then generate an extrachromosomal LTR-containing CaMV-like retroelement. Circularization might occur at the DNA level by homologous recombination, or at the level of a hybrid RNA/DNA replicative intermediate as suspected for the *Neurospora* retroplasmid (Fig. 4). The key question, however, is the direction of time's arrow. Did linear HeT-A-like elements give rise to circular retroelements, or did autonomous circular elements generate multimers (perhaps by homologous recombination or runaround transcription) that then acquired the ability to retropose as linears? We favor the view that large DNA chromosomes were assembled stepwise from smaller autonomously replicating units carrying one or a few linked genes. In this view, small autonomous genomes—especially transitional genomes that replicate through both RNA and DNA intermediates—are more likely to be precursors than progeny of large modern chromosomes.

**ARE MODERN VIRUSES MOLECULAR FOSSILS OF ANCIENT STRATEGIES
FOR GENOMIC REPLICATION?**

Most of us have come to have considerable confidence in the validity of
sequence-based phylogenies relating organisms and organelles (Pace et
al. 1986). Viruses, however, are quite another story. No single molecule,
product, or function is common to all viruses, so there is no universal
standard for evolutionary comparisons that is useful in quite the same
way as small subunit ribosomal RNA sequences have been for organis-
mal phylogenies. Furthermore, the interchangeability of functionally
similar viral modules can undermine viral phylogenies based on a single
molecule such as reverse transcriptase (but see Xiong and Eickbusch
1990; Eickbush 1997; Nakamura et al. 1997). Simple assumptions can-
not be made about the regularity of viral molecular clocks, because gen-
eration times are short, burst sizes are large, and there is no single typi-
cal genome, but rather a population of quasispecies and defective
interfering particles all propagating simultaneously (Eigen et al. 1981).
Compounding these formidable problems, it can be difficult or impossi-
ble to trace the source of new viral genetic information, because viruses
move horizontally between hosts, exchange genetic information with
other viruses by recombination, and acquire new characteristics in each
of a succession of hosts.

   Although many aspects of viral life cycles violate the assumptions of
sequence-based phylogenetic methods, viruses also present a unique op-
portunity for evolutionary analysis because certain aspects of viral life
cycles appear to resist rapid change. Gene order and replication strategy,
for example, are remarkably conserved between some plant and animal
viruses. The stability of viral strategies for genomic replication and gene
expression led Strauss and Strauss (1983) to propose that these proper-
ties of the viral life cycle are valuable and appropriate markers for virus
evolution.

   The attitude that viruses are cellular parasites, assembled in relatively
recent evolutionary time from preexisting parts of the cell (Benner and
Ellington 1988), is difficult to reconcile with data showing that many as-
pects of viral life cycles are stable over evolutionary time. In fact, the ge-
nomic tag hypothesis suggests an alternative view of viruses, as fossils
that reveal the diversity of ancient replication pathways (see also Win-
tersberger and Wintersberger 1987). Considered in this way, viruses have
not devised novel and subversive replication strategies; rather, they have
conserved the useful features of more ancient forms of cellular chromo-
somal replication, even as cellular replication strategies continued to

evolve. Contemporary viruses may thus preserve a record of the replication strategies used by more ancient chromosomes.

To illustrate how this view of contemporary viruses may be informative about molecular evolution, we show how RNA viruses, transitional genomes, and contemporary DNA genomes might be related using a phylogenetic tree based on the use of tRNA-like structures in replication (Fig. 6). The plus-strand RNA bacteriophages (Qβ) and plant viruses (brome mosaic virus) may represent molecular fossils of an RNA World in which cellular chromosomes were single-stranded RNA molecules with a 3′-terminal genomic tag. The *Neurospora* mitochondrial retroplasmid may represent a transitional stage from RNA to DNA genomes, in which duplex DNA functioned as the storage form for genetic information while genomic RNA with a 3′-terminal genomic tag served as a replicative intermediate. Cauliflower mosaic virus (CaMV) may be a molecular fossil of a more advanced transitional genome, which still replicated through a genomic RNA but in which the genomic tag had been transformed into a tRNA primer.

This view of viruses further suggests that modern duplex DNA genomes might be usefully regarded as retroviruses writ large. Eigen et al. (1981) originally observed that the role of RNA priming in the initiation of DNA synthesis might be a remnant of earlier RNA genomes. We are intrigued by the more specific possibility that the RNA primer for bacterial DNA replication is a degenerate form of the retroviral tRNA primer (see also Wintersberger and Wintersberger 1987), although it must be noted that eukaryotes generally use DNA primases that are capable of de novo initiation (for recent references, see Marini et al. 1997). RNA primases, for example, might turn out to be more closely related to reverse transcriptases than to RNA polymerases. Similarly, as discussed above, the telomerase responsible for completing replication of eukaryotic chromosomes is a specialized form of reverse transcriptase with an internal tRNA-like template (Fig. 5). Taken together, this evidence for the key role of RNA in both initiation and completion of chromosomal replication supports the notion that retroelements are the ancestors of modern eukaryotic chromosomes.

In arranging these selected viruses in this way, we in no sense imply that any one of these viruses is unchanged from ancient times. We do wish to point out that essential aspects of a viral replication strategy may remain stable over billions of years, and that the genomes of ancient organisms may survive in the form of viruses as the host replication strategy evolves (Maizels and Weiner 1994). This is of course consonant with the genomic tag hypothesis itself, which credits plus-strand RNA genomes

similar to Qβ and modern plant viruses with the invention of tRNA and aminoacylation.

Viewed as predecessors rather than derivatives of modern chromosomes, viruses may have much more to tell us about molecular evolution than is commonly appreciated.

CONCLUSION

We originally conceived of the genomic tag hypothesis to explain the origin of protein synthesis (Weiner and Maizels 1987), but with time it became clear that this hypothesis had equally distinct implications for the evolution of replicative mechanisms, beginning in an RNA World and continuing to the present day. The new evidence that we have discussed for the central role of genomic tags in the evolution of RNA to DNA genomes strengthens the case that tRNA-like genomic tags arose early in an RNA World.

Translation today is a complex and sophisticated process, involving at least 2 ribosomal RNAs, more than 50 ribosomal proteins, 20 synthetases, tRNAs, initiation factors, elongation factors, etc. It is clear that the translation apparatus must have arisen stepwise, but it has been difficult to imagine how any single component could be useful by itself, or how additional components could each individually confer a further selective advantage. Central to the genomic tag hypothesis is the suggestion that the two key components of the translation apparatus—tRNA and tRNA aminoacylation activity—first evolved as essential components of the replication apparatus, and were subject to selection *before* the advent of protein synthesis. Once in place, these two key components of the translation apparatus could be coopted for other purposes, with the result that replication and templated protein synthesis were fated to coevolve forever after.

REFERENCES

Akins R.A., Kelley R.L., and Lambowitz A.M. 1989. Characterization of mutant mitochondrial plasmids of *Neurospora spp.* that have incorporated tRNAs by reverse transcription. *Mol. Cell. Biol.* **9:** 678–691.

Alberts B.M. 1986. The function of the hereditary materials: Biological catalyses reflect the cell's evolutionary history. *Am. Zool.* **26:** 781–796.

Andino R., Rieckhof G.E., and Baltimore D. 1990. A functional ribonucleoprotein complex forms around the 5′ end of poliovirus RNA. *Cell* **63:** 369–380.

Barat C., Lullien V., Schatz O., Keith G., Nugeyre M.T., Gruninger-Leitch F., Barré-Sinoussi F., LeGrice S.F., and Darlix J.L. 1989. HIV-1 reverse transcriptase specifi-

cally interacts with the anticodon domain of its cognate primer tRNA. *EMBO J.* **8:** 3279–3285.

Benner S.A. and Ellington A.D. 1988. Return of the "last ribo-organism." *Nature* **332:** 688–689.

Benner S.A., Ellington A.D., and Tauer A. 1989. Modern metabolism as a palimpsest of the RNA world. *Proc. Natl. Acad. Sci.* **86:** 7054–7058.

Bhattacharyya A. and Blackburn E.H. 1994. Architecture of telomerase RNA. *EMBO J.* **13:** 5721–5723.

Biessmann H. and Mason J.M. 1997. Telomere maintenance without telomerase. *Chromosoma* **106:** 63–69.

Blackburn E.H. 1991. Structure and function of telomeres. *Nature* **350:** 569–573.

Blumenthal T. and Carmichael G.C. 1979. RNA replication: Function and structure of Qβ-replicase. *Annu. Rev. Biochem.* **48:** 525–548.

Brown D. and Gold L. 1996. RNA replication by Qβ replicase: A working model. *Proc. Natl. Acad. Sci.* **93:** 11558–11562.

Brown J.W., Hunt D.A., and Pace N.R. 1990. Nucleotide sequence of the 10Sa RNA gene of the β-purple eubacterium *Alcaligenes eutrophus. Nucleic Acids Res.* **18:** 2820.

Chapman K.B., Bystrom A.S., and Boeke J.D. 1992. Initiator methionine tRNA is essential for Ty1 transposition in yeast. *Proc. Natl. Acad. Sci.* **89:** 3236–3240.

Chen B. and Lambowitz A.M. 1997. De novo and DNA primer-mediated initiation of cDNA synthesis by the Mauriceville retroplasmid reverse transcriptase involve recognition of a 3′ CCA sequence. *J. Mol. Biol.* **271:** 311–332.

Covey S.N. and Turner D.S. 1986. Hairpin DNAs of cauliflower mosaic virus generated by reverse transcription *in vivo. EMBO J.* **5:** 2763–2768.

Cusack S., Berthet-Colominas C., Hartlein M., Nassar N., and Leberman R. 1990. A second class of synthetase structure revealed by X-ray analysis of *Escherichia coli* seryl-tRNA synthetase at 2.5 Å. *Nature* **347:** 249–255.

Danilevskaya O.N., Arkhipova I.R., Traverse K.L., and Pardue M.L. 1997. Promoting in tandem: The promoter for telomere transposon HeT-A and implications for the evolution of retroviral LTRs. *Cell* **88:** 647–655.

Decatur A.L. and Losick R. 1996. Three sites of contact between the *Bacillus subtilis* transcription factor σF and its antisigma factor SpoIIAB. *Genes Dev.* **10:** 2348–2358.

Deutscher M.P. 1982. tRNA nucleotidyltransferase. *Enzymes* **15:** 183–215.

DeZazzo J.D., Scott J.M., and Imperiale M.J. 1992. Relative roles of signals upstream of AAUAAAA and promoter proximity in regulation of HIV-1 mRNA 3′ end formation. *Mol. Cell. Biol.* **12:** 5555–5562.

Downey K.M., Jurmark B.S., and So A.G. 1971. Determination of nucleotide sequences at promoter regions by the use of dinucleotides. *Biochemistry* **10:** 4970–4975.

Dudock B.S., Katz G., Taylor E.K., and Holley R.W. 1969. Primary structure of wheat germ phenylalanine transfer RNA. *Proc. Natl. Acad. Sci.* **62:** 941–945.

Eickbush T.H. 1997. Telomerase and retrotransposons: Which came first? *Science* **277:** 911–912.

Eigen M., Gardiner W., Schuster P., and Winkler-Oswatitsch H. 1981. The origin of genetic information. *Sci. Am.* **244:** 88–118.

Eriani G., Delarue M., Poch O., Gangloff J., and Moras D. 1990. Partition of tRNA synthetases into two classes based on mutually exclusive sets of sequence motifs. *Nature* **347:** 203–206.

Fang G. and Cech T.R. 1993a. The β-subunit of *Oxytricha* telomere-binding protein promotes G-quartet formation by telomeric DNA. *Cell* **74:** 875–885.

———. 1993b. Characterization of a G-quartet formation reaction promoted by the β-subunit of the *Oxytricha* telomere-binding protein. *Biochemistry* **32:** 11646–11657.

Felden B., Himeno H., Muto A., McCutcheon J.P., Atkins J.F., and Gesteland R.F. 1997. Probing the structure of the *Escherichia coli* 10Sa RNA (tmRNA). *RNA* **3:** 89–103.

Francklyn C., Shi J.P., and Schimmel P. 1992. Overlapping nucleotide determinants for specific aminoacylation of RNA microhelices. *Science* **255:** 1121–1125.

Freier S.M., Alkema D., Sinclair A., Neilson T., and Turner D.H. 1985. Contributions of dangling end stacking and terminal base-pair formation to the stabilities of XGGCCp, XCCGGp, XGGCCYp, and XCCGGYp helixes. *Biochemistry* **24:** 4533–4539.

Ganem D. and Varmus H.E. 1987. The molecular biology of the hepatitis B viruses. *Annu. Rev. Biochem.* **56:** 651–693.

Gardner L.P., Mookhtiar K.A., and Coleman J.E. 1997. Initiation, elongation, and processivity of carboxyl-terminal mutants of T7 RNA polymerase. *Biochemistry* **36:** 2908–2918.

Gilbert W. 1986. The RNA world. *Nature* **319:** 618.

Green C.J., Vold B.S., Morch M.D., Joshi R.L., and Haenni A.L. 1988. Ionic conditions for the cleavage of the tRNA-like structure of turnip yellow mosaic virus by the catalytic RNA of RNase P. *J. Biol. Chem.* **263:** 11617–11620.

Green R. and Noller H.F. 1997. Ribosomes and translation. *Annu. Rev. Biochem.* **66:** 679–716.

Greider C.W. and Blackburn E.H. 1985. Identification of a specific telomere terminal transferase activity in *Tetrahymena* extracts. *Cell* **43:** 405–413.

———. 1989. A telomeric sequence in the RNA of *Tetrahymena* telomerase required for telomere repeat synthesis. *Nature* **337:** 331–337.

Guerrier-Takada C., van Belkum A., Pleij C.W.A., and Altman S. 1988. Novel reactions of RNAase P with a tRNA-like structure in turnip yellow mosaic virus RNA. *Cell* **53:** 267–272.

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

Henderson E., Hardin C.C., Walk S.K., Tinoco I. Jr., and Blackburn E.H. 1987. Telomeric DNA oligonucleotides form novel intramolecular structures containing guanine-guanine base pairs. *Cell* **51:** 899–908.

Hicke B.J., Christian E.L., and Yarus M. 1989. Stereoselective arginine binding is a phylogenetically conserved property of group I self-splicing RNAs. *EMBO J.* **8:** 3843–3851.

Himeno H., Sato M., Tadaki T., Fukushima M., Ushida C., and Muto A. 1997. In vitro *trans* translation mediated by alanine-charged 10Sa RNA. *J. Mol. Biol.* **268:** 803–808.

Hohn T., Hohn B., and Pfeiffer P. 1985. Reverse transcription in CaMV. *Trends Biochem. Sci.* **10:** 205–209.

Holm L. and Sander C. 1995. DNA polymerase β belongs to an ancient nucleotidyltransferase superfamily. *Trends Biochem. Sci.* **20:** 345–347.

Illangasekare M., Kovalchuke O., and Yarus M. 1997. Essential structures of a self-aminoacylating RNA. *J. Mol. Biol.* **274:** 519–529.

Illangasekare M., Sanchez G., Nickles T., and Yarus M. 1995. Aminoacyl-RNA synthesis catalyzed by an RNA. *Science* **267:** 643–647.

Isel C., Ehresmann C., Keith G., Ehresmann B., and Marquet R. 1995. Initiation of reverse transcription of HIV-1: Secondary structure of the HIV-1 RNA/tRNA(3Lys) (template/primer). *J. Mol. Biol.* **247:** 236–250.

Jay D.G. and Gilbert W. 1987. Basic protein enhances the incorporation of DNA into lipid vesicles: A model for the formation of primordial cells. *Proc. Natl. Acad. Sci.* **84:** 1978–1980.

Keiler K.C., Waller P.R.H., and Sauer R.T. 1996. Role of a peptide tagging system in degradation of proteins synthesized from damaged messenger RNA. *Science* **271:** 990–993.

Kikuchi Y., Ando Y., and Shiba T. 1986. Unusual priming mechanism of RNA-directed DNA synthesis in copia retrovirus-like particles of *Drosophila. Nature* **323:** 824–826.

Komine Y., Kitabatake M., Yokogawa T., Nishikawa K., and Inokuchi H. 1994. A tRNA-like structure is present in 10Sa RNA, a small stable RNA from *Escherichia coli. Proc. Natl. Acad. Sci.* **91:** 9223–9227.

Kuiper M.T.R. and Lambowitz A.M. 1988. A novel reverse transcriptase activity associated with mitochondrial plasmids of *Neurospora. Cell* **55:** 693–704.

Lanchy J.M., Isel C., Ehresmann C., Marquet R., and Ehresmann B. 1996. Structural and functional evidence that initiation and elongation of HIV-1 reverse transcription are distinct processes. *Biochimie* **78:** 1087–1096.

Lee B.J., Kang S.G., and Hatfield D. 1989. Transcription of *Xenopus* selenocysteine tRNA Ser (formerly designated opal suppressor phosphoserine tRNA) gene is directed by multiple 5′-extragenic regulatory elements. *J. Biol. Chem.* **264:** 9696–9702.

Levis R.W. 1989. Viable deletions of a telomere from a *Drosophila* chromosome. *Cell* **58:** 791–801.

Levis R.W., Ganesan R., Houtchens K., Tolar L.A., and Sheen F.M. 1993. Transposons in place of telomeric repeats at a *Drosophila* telomere. *Cell* **75:** 1083–1093.

Li Z. and Deutscher M.P. 1996. Maturation pathways for *E. coli* tRNA precursors: A random multienzyme process *in vivo. Cell* **86:** 503–512.

Limmer S., Hofmann H.P., Ott G., and Sprinzl M. 1993. The 3′-terminal end (NCCA) of tRNA determines the structure and stability of the aminoacyl acceptor stem. *Proc. Natl. Acad. Sci.* **90:** 6199–6202.

Lingner J., Hendrick L.L., and Cech T.R. 1994. Telomerase RNAs of different ciliates have a common secondary structure and a permuted template. *Genes Dev.* **8:** 1984–1998.

Lingner J., Hughes T.R., Shevchenko A., Mann M., Lundblad V., and Cech T.R. 1997. Reverse transcriptase motifs in the catalytic subunit of telomerase. *Science* **276:** 561–567.

Lohse P.A. and Szostak J.W. 1996. Ribozyme-catalysed amino-acid transfer reactions. *Nature* **381:** 442–444.

Maizels N. 1973. The nucleotide sequence of the lactose mRNA transcribed from the UV5 promoter mutant of *E. coli. Proc. Natl. Acad. Sci.* **70:** 3585–3589.

Maizels N. and Weiner A.M. 1987. Peptide-specific ribosomes, genomic tags and the origin of the genetic code. *Cold Spring Harbor Symp. Quant. Biol.* **52:** 743–749.

———. 1994. Phylogeny from function: Evidence from the molecular fossil record that tRNA originated in replication, not translation. *Proc. Natl. Acad. Sci.* **91:** 6729–6734.

Mans R.M., Guerrier-Takada C., Altman S., and Pleij C.W. 1990. Interaction of RNase P from *Escherichia coli* with peudoknotted structures in viral RNAs. *Nucleic Acids Res.* **18:** 3479–3487.

Marini F., Pellicioli A., Paciotti V., Lucchini G., Plevani P., Stern D.F., and Foiani M. 1997. A role for DNA primase in coupling DNA replication to DNA damage response. *EMBO J.* **16:** 639–650.

Martin G. and Keller W. 1996. Mutational analysis of mammalian poly(A) polymerase identifies a region for primer binding and catalytic domain, homologous to the family X polymerases, and to other nucleotidyltransferases. *EMBO J.* **15:** 2593–2603.

Matsufuji S., Matsufuji T., Miyazaki Y., Murakami Y., Atkins J.F., Gesteland R.F., and Hayashi S. 1995. Autoregulatory frameshifting in decoding mammalian ornithine decarboxylase antizyme. *Cell* **80:** 51–60.

McClain W.H., Guerrier-Takada C., and Altman S. 1987. Model substrates for an RNA enzyme. *Science* **238:** 527–530.

McClure W.R. 1980. Rate-limiting steps in RNA chain initiation. *Proc. Natl. Acad. Sci.* **77:** 5634–5638.

Melton D.A., Krieg P.A., Rebagliati M.R., Maniatis T., Zinn K., and Green M.R. 1984. Efficient *in vitro* synthesis of biologically active RNA and RNA hybridization probes from plasmids containing a bacteriophage SP6 promoter. *Nucleic Acids Res.* **12:** 7035–7056.

Michel F., Hanna M., Green R., Bartel D.P., and Szostak J.W. 1989. The guanosine binding site of the *Tetrahymena* ribozyme. *Nature* **342:** 391–395.

Milligan J.F., Groebe D.R., Witherell G.W., and Uhlenbeck O.C. 1987. Oligoribonucleotide synthesis using T7 RNA polymerase and synthetic DNA templates. *Nucleic Acids Res.* **15:** 8783–8798.

Musier-Forsyth K. and Schimmel P. 1992. Functional contacts of a transfer RNA synthetase with 2′-hydroxyl groups in the RNA minor groove. *Nature* **357:** 513–514.

Nakamura T.M., Morin G.B., Chapman K.B., Weinrich S.L., Andrews W.H., Lingner J., Harley C.B., and Cech T.R. 1997. Telomerase catalytic subunit homologs from fission yeast and human. *Science* **277:** 955–959.

Noller H.F. 1993. On the origin of the ribosome: Coevolution of subdomains of tRNA and rRNA. In *The RNA world* (ed. R.F. Gesteland and J.F. Atkins), pp. 137–156. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Noller H.F., Hoffarth V., and Zimniak L. 1992. Unusual resistance of peptidyl transferase to protein extraction procedures. *Science* **256:** 1416–1419.

Orgel L.E. 1968. Evolution of the genetic apparatus. *J. Mol. Biol.* **38:** 381–393.

Pace N.R., Olsen G.J., and Woese C.R. 1986. Ribosomal RNA phylogeny and the primary lines of evolutionary descent. *Cell* **45:** 25–326.

Pardue M.L., Danilevskaya O.N., Traverse K.L., and Lowenhaupt K. 1997. Evolutionary links between telomeres and transposable elements. *Genetica* **100:** 73–84.

Piccirilli J.A., McConnell T.S., Zaug A.J., Noller H.F., and Cech T.R. 1992. Aminoacyl esterase activity of the *Tetrahymena* ribozyme. *Science* **256:** 1420–1424.

Pilipenko E.V., Maslova S.V., Sinyakov A.N., and Agol V.I. 1992. Towards identification of *cis*-acting elements involved in the replication of enterovirus and rhinovirus RNAs: A proposal for the existence of tRNA-like terminal structures. *Nucleic Acids Res.* **20:** 1739–1745.

Poch O., Sauvaget I., Delarue M., and Tordo N. 1989. Identification of four conserved motifs among the RNA-dependent polymerase encoding elements. *EMBO J.* **8:** 3867–3874.

Polesky A.H., Dahlberg M.E., Benkovic S.J., Grindley N.D.F., and Joyce C.M. 1992. Side chains involved in catalysis of the polymerase reaction of DNA polymerase I from *Escherichia coli. J. Biol. Chem.* **267:** 8417–8428.

Popper K.R. 1963. *Conjectures and refutations*, 4th edition. Routledge & Kegan Paul, London, United Kingdom.

Puglisi J.D., Tan R., Calnan B.J., Frankel A.D., and Williamson J.R. 1992. Conformation of the TAR RNA-arginine complex by NMR spectroscopy. *Science* **257:** 76–80.

Puglisi E.V., Puglisi J.D., Williamson J.R., and RajBhandary U.L. 1994. NMR analysis of tRNA acceptor stem microhelices: Discriminator base change affects tRNA conformation at the 3′ end. *Proc. Natl. Acad. Sci.* **91:** 11467–11471.

Quigley G.J. and Rich A. 1976. Structural domains of transfer RNA molecules. *Science* **194:** 796–806.

Rao A.L.N., Dreher T.W., Marsh L.E., and Hall T.C. 1989. Telomeric function of the tRNA-like structure of brome mosaic virus RNA. *Proc. Natl. Acad. Sci.* **86:** 5335–5339.

Rasmussen N.J., Wikman F.P., and Clark B.F. 1990. Crosslinking of tRNA containing a long extra arm to elongation factor Tu by *trans*-diamminedichloroplatinum(II). *Nucleic Acids Res.* **18:** 4883–4890.

Romero D.P. and Blackburn E.H. 1991. A conserved secondary structure for telomerase RNA. *Cell* **67:** 343–353.

Rould M.A., Perona J.J., Söll D., and Steitz T.A. 1989. Structure of *E. coli* glutaminyl-tRNA synthetase complexed with tRNA(Gln) and ATP at 2.8 Å resolution. *Science* **246:** 1135–1142.

Sanfacon H. and Hohn T. 1990. Proximity to the promoter inhibits recognition of cauliflower mosaic virus polyadenylation signal. *Nature* **346:** 81–84.

SantaLucia Jr. J., Kierzek R., and Turner D.H. 1990. Effects of GA mismatches on the structure and thermodynamics of RNA internal loops. *Biochemistry* **29:** 8813–8819.

Saville B.J. and Collins R.A. 1990. A site-specific self-cleavage reaction performed by a novel RNA in *Neurospora* mitochondria. *Cell* **61:** 685–696.

Schimmel P. 1987. Aminoacyl tRNA synthetases: General scheme of structure-function relationships in the polypeptides and recognition of transfer RNAs. *Annu. Rev. Biochem.* **56:** 125–158.

Schimmel P. and Henderson B. 1994. Possible role of aminoacyl-RNA complexes in non-coded peptide synthesis and origin of coded synthesis. *Proc. Natl. Acad. Sci.* **91:** 11283–11286.

Schimmel P., Giegé R., Moras D., and Yokoyama S. 1993. An operational RNA code for amino acids and possible relationship to genetic code. *Proc. Natl. Acad. Sci.* **90:** 8763–8768.

Sen D. and Gilbert W. 1988. Formation of parallel four-stranded complexes by guanine-rich motifs in DNA and its implications for meiosis. *Nature* **334:** 364–366.

Shi P.-Y., Maizels N., and Weiner A.M. 1998a. CCA addition by tRNA nucleotidyltransferase: Polymerization without translocation? *EMBO J.* **17:** 3188–3196.

Shi P.-Y., Weiner A.M., and Maizels N. 1998b. A top-half tDNA minihelix is a good substrate for the eubacterial CCA-adding enzyme. *RNA* **4:** 276–284.

Sprinzl M. and Cramer F. 1979. The -C-C-A end of tRNA and its role in protein biosynthesis. *Prog. Nucleic Acid Res. Mol. Biol.* **22:** 1–69.

Steitz T.A. 1998. A mechanism for all polymerases. *Nature* **391:** 231–232.

Strauss E.G. and Strauss J.H. 1983. Replication strategies of the single stranded RNA viruses of eukaryotes. In *Microbiology and immunology* (ed. M. Cooper et al.), vol. 105, pp. 2–98. Springer-Verlag, Berlin.

Subbarao M.N. and Apirion D. 1989. A precursor for a small stable RNA (10Sa RNA) of *Escherichia coli*. *Mol. Gen. Genet.* **217:** 499–504.

Sugimoto N., Hasegawa K., and Sasaki M. 1990. Stability of 3′ dangling ends on the core helix of AUGCAU at various Na⁺ concentrations. *Nucleic Acids Symp. Ser.* **22:** 107–108.

Sundquist W.I. and Klug A. 1989. Telomeric DNA dimerizes by formation of guanine tetrads between hairpin loops. *Nature* **342:** 825–829.

Tse W.T. and Forget B.G. 1990. Reverse transcription and direct amplification of cellular RNA transcripts by Taq polymerase. *Gene* **88:** 293–296.

Tu G.F., Reid G.E., Zhang J.G., Moritz R.L., and Simpson R.J. 1995. C-terminal extension of truncated recombinant proteins in *Escherichia coli* with 10Sa RNA decapeptide. *J. Biol. Chem.* **270:** 9322–9326.

Tyagi J.S. and Kinger A.K. 1992. Identification of the 10Sa RNA structural gene of *Mycobacterium tuberculosis. Nucleic Acids Res.* **20:** 138.

Visser C.M. and Kellogg R.M. 1978a. Biotin. Its place in evolution. *J. Mol. Evol.* **11:** 171–187.

———. 1978b. Bioorganic chemistry and the origin of life. *J. Mol. Evol.* **11:** 163–169.

Weiner A.M. and Maizels N. 1987. 3′ Terminal tRNA-like structures tag genomic RNA molecules for replication: Implications for the origin of protein synthesis. *Proc. Natl. Acad. Sci.* **84:** 7383–7387.

Welch M., Majerfeld I., and Yarus M. 1997. 23S rRNA similarity from selection for peptidyl transferase mimicry. *Biochemistry* **36:** 6614–6623.

White III, H.B. 1976. Coenzymes as fossils of an earlier metabolic state. *J. Mol. Evol.* **7:** 101–104

———. 1982. Evolution of coenzymes and the origin of pyridine nucleotides. In *The pyridine coenzymes* (ed. J. Everse et al.), pp. 1–17. Academic Press, New York.

Williams K.P. and Bartel D.P. 1996. Phylogenetic analysis of tmRNA secondary structure. *RNA* **2:** 1306–1310.

Williamson J.R., Raghuraman M.K., and Cech T.R. 1989. Monovalent cation-induced structure of telomeric DNA: the G-quartet model. *Cell* **59:** 871–880.

Wintersberger U. and Wintersberger E. 1987. RNA makes DNA: A speculative view of the evolution of DNA replication mechanisms. *Trends Genet.* **3:** 198–202.

Wong J.-T. 1991. Origin of genetically encoded protein synthesis: A model based on selection for RNA peptidation. *Origins Life Evol. Biosph.* **21:** 165–176.

Xiong Y. and Eickbush T.H. 1990. Origin and evolution of retroelements based upon their reverse transcriptase sequences. *EMBO J.* **9:** 3353–3362.

Yarus M., Illangesekare M., and Christian E. 1991. An axial binding site in the *Tetrahymena* precursor RNA. *J. Mol. Biol.* **222:** 995–1012.

Yue D., Maizels N., and Weiner A.M. 1996. CCA-adding enzymes and poly(A) polymerases are all members of the same nucleotidyltransferase superfamily: Characterization of the CCA-adding enzyme from the archaeal hyperthermophile *Sulfolobus shibatae. RNA* **2:** 895–908.

Zakian V.A. 1996. Structure, function, and replication of *Saccharomyces cerevisiae* telomeres. *Annu. Rev. Genet.* **30:** 141–172.