Methyl-RNA: an evolutionary bridge between RNA and DNA?

Anthony Poole [1,2], David Penny [1] and Britt-Marie Sjöberg [2]

Given the apparent limitation of double-stranded RNA (dsRNA) genomes to about 30 kb, together with the complexity of DNA synthesis, it appears difficult for a dsRNA genome to encode all the information required before the transition from an RNA to a DNA genome. Ribonucleotide reductase itself, which synthesises deoxyribonucleotides from ribonucleotides, requires complex protein radical chemistry, and RNA world genomes may have reached their limits of coding capacity well before such complex enzymes had evolved. The transition from RNA to DNA thus appears to require intermediate steps, and we suggest that the naturally occurring 2′-*O*-methylated RNA, with chemical properties intermediate between RNA and DNA, is a suitable candidate.

[1]Institute of Molecular BioSciences, Massey University, Private Bag 11-222, P.O. Box 11-222, Palmerston North, New Zealand
[2]Department of Molecular Biology, Stockholm University, 106 91 Stockholm, Sweden

Correspondence: Anthony Poole
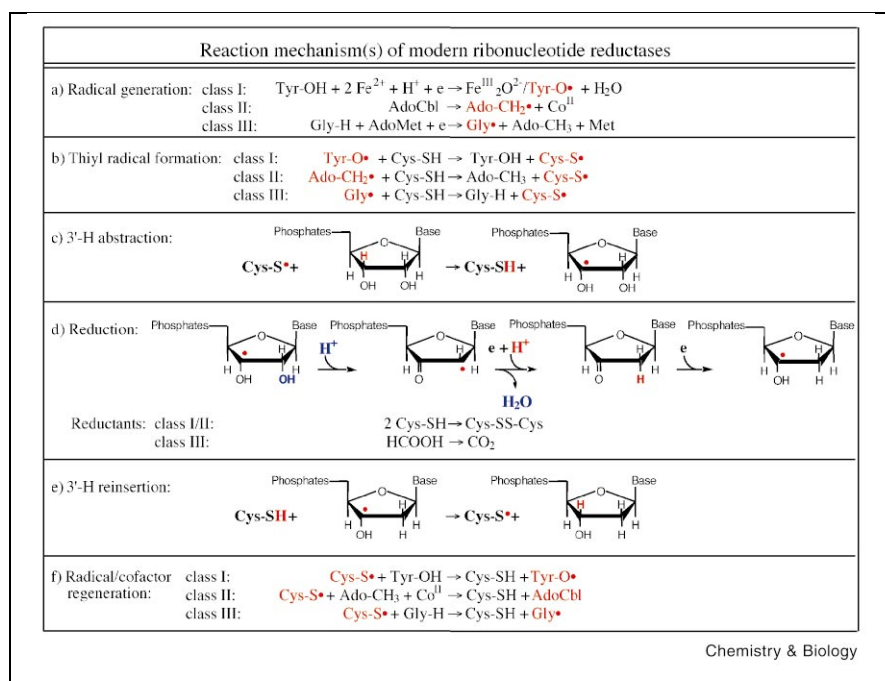E-mail: a.m.poole@massey.ac.nz

Introduction

Before genetically encoded proteins and DNA, there is thought to be a period in the evolution of life where RNA was both catalyst and genetic material. Some 'relics' of this RNA world [1] period remain as a handful of catalytic and evolutionarily conserved RNAs, and RNA viruses [2–4]. The most spectacular is the ribosome with a core of RNAs necessary, and perhaps sufficient, for protein synthesis [5–7]. DNA synthesis is not carried out by any such RNA machine, but traces of an RNA world origin are nevertheless evident [8,9]. DNA synthesis cannot proceed without an RNA primer [2], and that deoxyribonucleotides are synthesised from ribonucleotides (by ribonucleotide reductases) also argues for the evolutionary transition from RNA to DNA [10].

These enzymes (ribonucleotide reductases) are one of our research topics, and a good place to start the discussion. On the basis of structure [11] and suggested sequence homology [12,13] all three classes appear to share a common origin. All require complex protein radical chemistry to reduce ribonucleotides to deoxyribonucleotides, though the radical generation mechanism differs between the classes (Figure 1) [10,14]. The highly complex nature of the reduction suggests that catalytic proteins had to arise before the transition to DNA could occur. Catalytic RNA (ribozymes), almost certainly could not have carried out such free radical chemistry – even if a radical could have been generated it is unlikely it could have been controlled (Table 1) [4,15]. A salient example of this is the use of radicals in probing of nucleic acid structure. Fe(II)–EDTA generated radicals cleave RNA and DNA non-specifically, regardless of whether single-stranded or double-stranded, and irrespective of local structure [15]. While class II and III reductases use nucleotide cofactors, this does not mean that the reactions would have occurred in the RNA world, though these cofactors may date from the RNA world. It is the cofactors themselves [16], not the diversity of reactions they are involved in, that date back to the RNA world – they have been recruited into many new metabolic pathways during evolution. Given known RNA chemistry, ribonucleotide reduction appears impossible until the development of protein catalysts.

Other enzymes required for the transition to DNA are polymerases and helicases. DNA and RNA polymerases may use a common catalytic mechanism [17], and mutagenesis studies indicate that substrate selection can be discriminated by a single residue, a single point mutation permits DNA polymerase to accept ribonucleotides [18,19]. Furthermore, DNA and RNA helicases belong to a single superfamily, [20], thus much of the DNA synthesis machinery may have been recruited from RNA synthesis.

Although ribonucleotide reduction is a crucial link in the transition from RNA to DNA, is this single step sufficient? In this paper we review the relevant aspects of the RNA world, RNA chemistry, the chemistry of ribonucleotide reduction and the coding capacity of double-stranded RNA (dsRNA), and overall we conclude that it is not. We suggest a possible bridge between RNA and DNA that has a basis in known RNA biochemistry.

Reaction mechanism(s) of modern ribonucleotide reductases

a) Radical generation: class I: $Tyr\text{-}OH + 2\,Fe^{2+} + H^+ + e \rightarrow Fe^{III}{}_2O^{2-}/Tyr\text{-}O\bullet + H_2O$
class II: $AdoCbl \rightarrow Ado\text{-}CH_2\bullet + Co^{II}$
class III: $Gly\text{-}H + AdoMet + e \rightarrow Gly\bullet + Ado\text{-}CH_3 + Met$

b) Thiyl radical formation: class I: $Tyr\text{-}O\bullet + Cys\text{-}SH \rightarrow Tyr\text{-}OH + Cys\text{-}S\bullet$
class II: $Ado\text{-}CH_2\bullet + Cys\text{-}SH \rightarrow Ado\text{-}CH_3 + Cys\text{-}S\bullet$
class III: $Gly\bullet + Cys\text{-}SH \rightarrow Gly\text{-}H + Cys\text{-}S\bullet$

c) 3'-H abstraction: $Cys\text{-}S\bullet +$ [ribose structure] $\rightarrow Cys\text{-}SH +$ [ribose structure]

d) Reduction: [ribose structures with H^+, $e + H^+$, H_2O, e]

Reductants: class I/II: $2\,Cys\text{-}SH \rightarrow Cys\text{-}SS\text{-}Cys$
class III: $HCOOH \rightarrow CO_2$

e) 3'-H reinsertion: $Cys\text{-}SH +$ [ribose structure] $\rightarrow Cys\text{-}S\bullet +$ [ribose structure]

f) Radical/cofactor regeneration: class I: $Cys\text{-}S\bullet + Tyr\text{-}OH \rightarrow Cys\text{-}SH + Tyr\text{-}O\bullet$
class II: $Cys\text{-}S\bullet + Ado\text{-}CH_3 + Co^{II} \rightarrow Cys\text{-}SH + AdoCbl$
class III: $Cys\text{-}S\bullet + Gly\text{-}H \rightarrow Cys\text{-}SH + Gly\bullet$

Chemistry & Biology

Figure 1. Modern protein ribonucleotide reductases. A simplified scheme summarising the chemistry of ribonucleotide reduction for each of the three classes of ribonucleotide reductase. In the first step, a radical is generated at some distance from the active site. In class I and III, radical generation and catalysis occur on different subunits. The class I reductases produce a stable tyrosine radical with the help of an iron-oxygen centre, and this is transferred to the active site, 30–40 Å away. In classes II and III, a nucleotide cofactor (adenosylcobalamin (AdoCbl) and AdoMet, respectively) is cleaved to generate a deoxyadenosyl radical (AdoCH$_2\bullet$), which subsequently forms a stable glycine radical in the class III reductase. In the next step the radical in all three classes is transferred to an active site cysteine, forming a thiyl radical. This is then transferred to the 3′ position of the ribose, and two cysteines act as reductants in classes I and II, while formate is the reductant in the class III reaction. Once the 2′-OH has been reduced to 2′-H, with regeneration of the thiyl radical, this is transferred back to the tyrosine in class I, to the glycine in class III, while in class II AdoCbl is regenerated. In the class III reaction, AdoMet is consumed in generating the glycine radical, but like the tyrosine radical in the class I reaction, the glycine radical is stable and can presumably support several enzymatic turn-over cycles.

RNA coding capacity

DNA can retain much larger quantities of genetic information than RNA, and this is probably why the former was selected over the latter during evolution. Coding capacity is described by the 'error threshold' or 'Eigen limit'. Put simply, the accuracy of copying information places a limit on a genome's coding capacity. If accuracy is high, large amounts of information can be maintained; if accuracy is poor, the amount of information maintained is low (see Eigen, [21]). The Eigen limit describes coding capacity in terms of the overall fidelity of replication, and for RNA viruses, which lack proof-reading and repair, genome size may offer a good measure of this and provides a reasonable indication of genome size in the RNA world [4]. If RNA genomes were subject to proof-reading and repair, the information content would be greater than that dictated by the spontaneous mutation rate.

Coronaviruses have the largest viral RNA genomes identified, reaching 30 kb [22]. It is reasonable to assume that the largest modern RNA viruses are indicative of the upper size limit for an RNA world genome, and perhaps even exceed it [4,9]. If larger RNA genomes are possible without proof-reading and repair, neither theoretical treatments nor viral genome sizes predict it, and there is currently no reason to suspect proof-reading and repair were properties of RNA genomes.

In addition, in the RNA world, coding capacity was probably small since the genome was replicated by a ribozyme RNA polymerase, which is expected to lack the accuracy of modern RNA polymerases. If repair and proof-reading functions arose after the advent of DNA, the spontaneous mutation rate in the RNA world would have been approximately 10^{-4} [8].

The numerous RNA world relics identified in modern organisms give a partial picture of the RNA world. Many traits presumably essential to RNA-based life left no evidence of their existence, other than the handful of nucleotide cofactors [16]. Even with this gap in the RNA world model, current knowledge suggests the RNA organism that evolved protein synthesis was likely to have been dangerously close to the Eigen limit for RNA, and probably em-

ployed strategies, such as genetic redundancy and recombination, to 'buffer' against information loss [4].

Proteins increase fidelity

Protein synthesis is expected to have had a two-fold effect on RNA organisms: improvement of RNA-catalysed reactions (including replication) which in turn allowed an increase in genome coding capacity. The first proteins are expected to be non-specific, short basic proteins with few constraints on primary structure, selected for their propensity to bind, stabilise and globally improve ribozyme efficiency [23]. Any protein that improved replication/translation by stabilising rRNA would, in the next round of translation, be translated with greater efficiency, improving both replication and translation. This positive feedback cycle would allow larger amounts of information to be maintained, allowing new innovations to arise (see Figure 4 legend) [9,24].

With a ribozyme RNA polymerase, and RNA as genetic material, such feedback would have edged the genome size closer to the presumed modern maximum for RNA genomes. However, the gap between these first proteins and the first catalytic proteins is large. This is especially clear when we examine the ribonucleotide reductases. In all three cases the enzyme is large (radical storage/handling requires this) and radical generation is separate from catalysis (Figure 1) – the radical must be contained in the absence of substrate and directed specifically to the ribose 2′-hydroxyl during catalysis. Such a complex reaction was almost certainly not carried out by the first catalytic proteins, which probably tackled less complex reactions. Significantly, it has been noted that, in contrast to many other biological reactions, the expectation on chemical grounds is that ribonucleotide reduction could not possibly be carried out in the absence of an enzyme [25]. As in modern organisms, free radicals could have caused considerable damage to the cell.
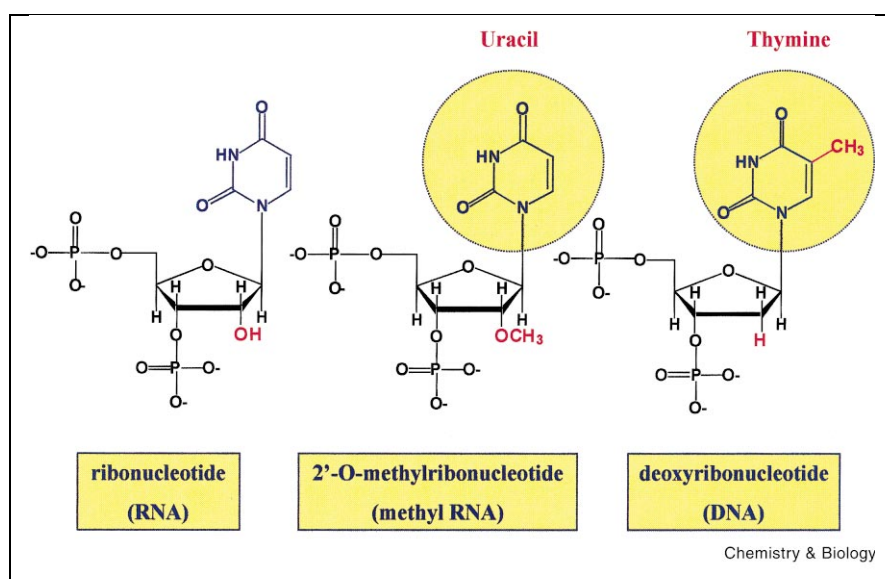
RNA vs DNA

The major difference between DNA and RNA is that the former has hydrogen at the 2′ position of the ribose while the latter has hydroxyl (Figure 2). The 2′-hydroxyl is crucial for RNA tertiary structure, and is also the nucleophile in self-cleavage by several ribozymes [26]. Self-cleavage is problematic for genomic RNA as there is potential for the 2′-OH of any ribose to attack the adjacent 3′-phosphate, breaking the phosphodiester bond. With hydrogen at the 2′ position DNA cannot self-cleave, and is consequently a poor catalyst, but more stable a genetic material than RNA.

Since in biochemical terms the leap to DNA is so large, there must have been strong selection to make the transition from RNA to DNA. A low-fidelity RNA genome is not expected to have been able to encode the large number of biochemical processes that even the smallest genomes [27] of today do, yet ribonucleotide reductases provide the only link in the transition from RNA to DNA genomes.

With catalytic RNAs and multifunctional RNA-binding proteins, a 30 kb RNA genome (see above) would be stretched to its limit. Where is there sufficient space to encode simple proteins, then more complex catalytic proteins capable of generating and controlling protein radicals? Without proof-reading and repair, it seems highly unlikely that an RNA genome could have been large enough to support such a large collection of protein genes, making it hard to see how ribonucleotide reductase could have arisen in an RNA world. This is the central problem we address. Intermediates between RNA and DNA, with improved fidelity of replication (and thus larger coding capacity), are the obvious way of solving the conundrum.

Bridging the gap? The methyl-RNA world hypothesis

The possibility that an intermediate genetic material ex-

Table 1

General steps required for ribonucleotide reduction vs known RNA chemistry.

Steps carried out by RNR	Likely on basis of known RNA chemistry?
Binding to substrate and cofactor	Yes
Cleavage of cofactor (C–Co(III) bond or C–S$^+$ bond) to generate a radical	Possibly
Containment and control of radical[a]	No
Transfer of radical to substrate	No
Radical/cofactor regeneration	No
Known reactions carried out by RNA[b]	
Transesterification (cleavage, ligation)	
N-Alkylation	
Peptide bond formation	
Acyl transfer	
Isomerisation	
N-Glycosidic bond formation	

[a]RNA is non-specifically cleaved by free radicals, as shown by nucleic acid structure probing with radicals generated by Fe(II)–EDTA cleavage [15] or radiolysis of water [52].
[b]Reviewed in Yarus [53] and Carola and Eckstein [54].

Figure 2. Nucleotides as they would appear within a nucleic acid chain. The difference between RNA, 2′-*O*-methylated RNA and DNA is the nature of the group at the 2′ position of the ribose (highlighted in red). In RNA the 2′ position carries a hydroxyl. The reactivity of this group is a major determinant of RNA catalytic activity. 2′-*O*-Methyl RNA occurs extensively on eukaryotic and archaeal RNAs, particularly ribosomal RNA, where the reaction is snoRNA-mediated. Site-specific protein methylases catalyse the rRNA methylation in bacteria. Methylation eliminates the propensity for 2′-OH-mediated catalytic activity. In DNA, hydrogen is at the 2′ position on the ribose. Ribonucleotide reductases utilise free radical chemistry to reduce the 2′-OH of free ribonucleotides to a 2′-H, forming free deoxyribonucleotides. Bases are shown in blue: uracil in RNA and 2′-*O*-methyl RNA and thymine in DNA.

isted in the transition from the RNA to the DNA world has not been considered, primarily since no such material has been identified. Most functional RNAs however carry modifications, many of which are to the bases [28]. In addition, methylation of ribose at the 2′-OH (Figure 2) is both common and universal in all organisms [29].

The advantage that ribose methylation offers RNA is two-fold: silencing 2′-OH moieties within RNA, independent of sequence, ensures only specific residues are catalytic (thus preventing unwanted side reactions), and, preventing 2′-OH from forming H-bonds in the tertiary structure, thus perhaps favouring a particular folding pathway. Both (silencing of catalysis and effects on folding) may also be relevant to RNA as genetic material. Starting from the assumption that 2′-OH silencing is the basis of improved stability in nucleic acids, we will first argue that 2′-*O*-methylation dates back to an RNA world. Then we examine the possibility that, in terms of genomic stability and coding capacity, 2′-*O*-methylation of an RNA genome may have provided a chemically simple stepping-stone from RNA to DNA in genome evolution (Table 2). Others have argued for an ancient origin for modified nucleotides in general [30,31], but we restrict our discussion to 2′-*O*-methylation.

Before describing the theory in detail, it is worth considering that it appears to add an extra step to the RNA to DNA transition. Our rationale for doing so is that the strong limits to RNA coding capacity imply that complex protein-catalysed metabolic processes could not have evolved in an RNA genome in the absence of proof-reading and repair, even though some increase in fidelity comes from maintaining the genome in multiple copies [32]. Additionally, ribonucleotide reduction most probably could not have been ribozyme-catalysed (Table 1). Given that it is among the most complex of reactions catalysed by protein enzymes, we consider it to have arisen relatively late in protein evolution, and could only occur in organisms with quite large genomes with a large coding capacity.

The main advantage of an extra stage in the RNA to DNA transition is that it does not require large, extremely unlikely, steps. Each step relies only on stages that are consistent with established mechanisms, and in this sense is the simplest (or maximum likelihood) explanation [33]. A similar argument for an intermediate stage has been used for the origins of the RNA world itself. On current knowledge of prebiotic chemistry, β-D-ribose was not readily produced on the early Earth and this has led [34] to a search for intermediate genetic materials that might have pre-

Table 2

The pros and cons of 2′-*O*-methylated RNA as a genetic material.

Better than RNA	Worse than DNA
The reactivity of the 2′-OH is eliminated.	Modification of the ribose is post-transcriptional.
2′-OCH$_3$ confers qualities similar to DNA as a genetic material.	2′-*O*-methyl ribose would produce a hydrophobic cushion in the deep groove of the helix – full methylation may compromise genome functionality.

ceded RNA. Nelson et al. [35] suggest PNA (peptide nucleic acid) as a candidate first genetic material. Though there is no evidence for PNA having ever been a biologically relevant molecule, searching for intermediates in the origin of life can aid in uncovering simpler solutions to the problem. In the current case, we approach the question of the origins of DNA within the framework of the accepted model for the RNA world, and in doing so we identify a number of problems with this model in understanding the origin of DNA. We suggest that adding an extra step is currently the simplest way [33] of treating the problem, and it leads to potential experimental tests.

2′-*O*-Methylation in the RNA world?

Returning to the main theme, there are many pointers to an RNA world origin for 2′-*O*-methylation. In all organisms, ribosomal RNA transcripts are 2′-*O*-methylated as part of the process that produces mature rRNA. Ribosomal processing in eukaryotes uses a number of *small nucleolar* RNAs (snoRNAs) required for 2′-*O*-methylation of rRNAs at specific sites [29,36]. In bacteria, snoRNAs are apparently absent, and site-specific protein enzymes methylate rRNA [29,39]. In support of the ancient origin of snoRNA-mediated methylation, homologues of methylation snoRNAs have been also identified in archaea [37], suggesting snoRNAs (and methylation) pre-date the bacterial protein–enzyme system [9].

The 'bare-bones' rRNAs, as they presumably were in the RNA world, would have greatly benefited from silencing of 2′-OH groups by methylation as large RNAs would have had a high propensity for misfolding, and probably contained many sites of potential catalytic activity. This may also have been the case for other RNAs, so snoRNA-mediated post-transcriptional methylation could have been of global selective advantage. A recent report describes two C/D box snoRNAs that guide 2′-*O*-methylation of U6 snRNA [38] – one of these also directs methylation of 28S rRNA. We take this as support that methylation could have been widespread in the RNA world. We suggest that once RNA-binding proteins arose, the role that methylation played in folding and stabilising functional RNA was gradually replaced by protein–RNA interactions. An explanation for the lack of snoRNA in bacteria is based on this premise [9,23].

With only snoRNA-guided RNA modification in the RNA world, the degree of methylation of the RNA genome, if any, would be limited to sites of maximal benefit, such as those that confer an advantage on both the functional RNA and its gene. The reason is straightforward: each modification is guided by a site-specific snoRNA approximately 70–100 nt long [29] and, hence, unless each methylation translates to an increase in coding capacity of 70–100 nt or more, the genome gets filled up faster than it can increase in size! Some snoRNAs guide more than one modification,

but even so, the implication of snoRNA-mediated modification is that the number of modifications that could have been maintained on functional RNA was small and limited to the most crucial sites. The large diversity of snoRNAs found in modern eukaryotes may thus have arisen later in evolution [36]. The possibility of a 2′-*O*-methylated RNA genome as an intermediate in the evolution of DNA requires a means of methylation that does not exponentially eat up coding capacity.

Can 2′-*O*-methylation be carried out by a ribozyme?

Before examining a possible methyl-RNA genome we need to consider the question whether RNA itself can carry out 2′-*O*-methylation. In principle, a methyl-RNA genome could have arisen either before, or after, protein synthesis. The methylase which putatively interacts with snoRNAs in modern eukaryotic rRNA processing has been partially purified, and shown to use the pyridine nucleotide cofactor *S*-adenosyl-L-methionine (AdoMet) as methyl donor [39], so we will begin by looking at the role of AdoMet in methylation.

The crystal structures of two methylases that transfer a methyl group from AdoMet to a hydroxyl oxygen have been solved: catechol-*O*-methyltransferase [40], and VP39 from vaccinia virus (which 2′-*O*-methylates the residue adjacent to the cap in mRNA) [41]. While only the latter acts on the 2′-hydroxyl of ribose, both are instructive in trying to understand the possible mechanism of this reaction. Biochemical studies of catechol-*O*-methyltransferase reveal that the methyl group is almost certainly transferred directly from AdoMet to the oxygen by an S_N2 mechanism, with no methylated enzyme intermediate [42]. Positive charges (Lys-144, Mg^{2+}, and S^+ of AdoMet) close to the hydroxyl are predicted to promote leaving of the hydrogen, upon which the oxygen would make a nucleophilic attack on the reactive $H_3C–S^+$ bond on AdoMet (Figure 3A), producing *S*-adenosyl-L-homocysteine [40,42]. The data for VP39 [41] likewise suggest the reaction is coaxed along simply by the coordination of positively charged residues in the vicinity of the 2′-hydroxyl.

AdoMet is widely used as a methyl donor in biochemical reactions to a broad variety of acceptors. The positively charged sulphonium ion, to which the methyl group is linked, makes AdoMet highly reactive, and hence a good methyl donor. So reactive is AdoMet that it also methylates both proteins and DNA non-enzymatically [43,44]. In DNA, AdoMet spontaneously methylates adenine and guanine, forming 3-methyladenine and 7-methylguanine. In both these cases, methylation is on a nitrogen moiety, though there is tentative evidence that O_6-methylguanosine is also produced in trace amounts [45].

As yet, the possibility that 2′-*O*-methylations might arise

spontaneously has not been examined, but the work on methylases and AdoMet suggests a methylase ribozyme could be selected for in vitro. Certainly available evidence suggests *2′-O*-methylation is a simple enough reaction to be placed in the RNA world.

A methyl-RNA genome?

Here we will argue expansion from methylation of functional RNAs by a limited repertoire of snoRNAs (and perhaps a methylase ribozyme) to genome methylation could have come about with the origin of a non-specific dsRNA methylase which methylated specific sites by interaction with snoRNA–pre-rRNA duplexes. Based on the crystallographic data above, minimal requirements for a simple methylase would be positive charge, dsRNA-binding activity plus snoRNA–rRNA recognition, and AdoMet-binding activity, though the snoRNA/ribozyme could have bound AdoMet (Figure 3B). Such simple features are reminiscent of the expected properties of the earliest proteins [23]. We shall first describe the model, then present data suggesting its feasibility.
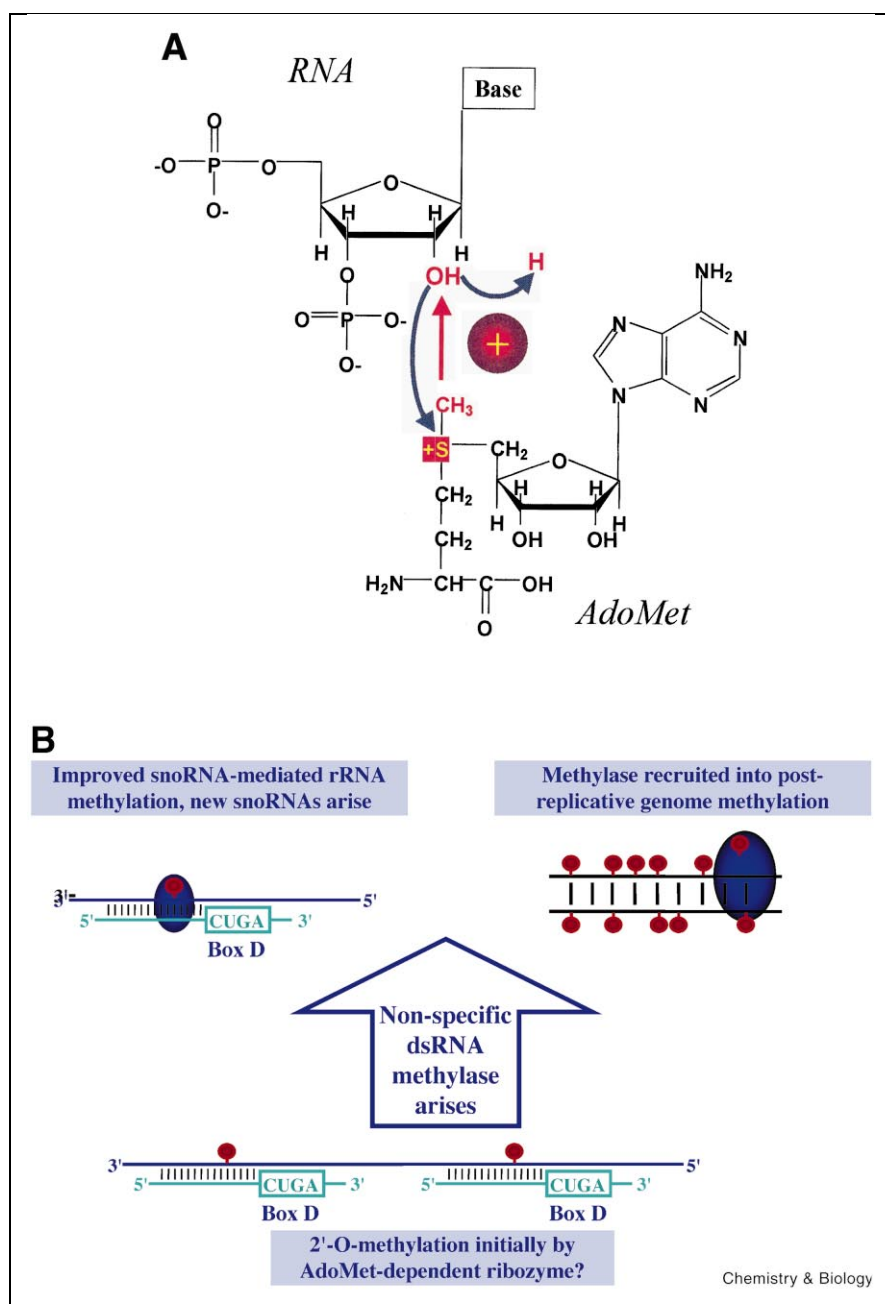
Figure 3. Methylase recruitment into RNA genome modification. **(A)** Predicted minimal requirements for methylation by an ancient methylase. Methylation on hydroxyl oxygen may require AdoMet as methyl group donor, a means of bringing the 2′-hydroxyl and methyl moieties into close contact, and positive charges to facilitate leaving of the hydrogen (see text). The role of the protein enzyme may be simply to augment methylation rather than being directly involved in the catalytic step. **(B)** SnoRNA-dependent modification of pre-rRNA dates back to the RNA world. As in the modern system, methylation may originally have occurred concurrent with transcription. With the advent of protein synthesis very simple catalysts arose, among these a non-specific dsRNA methylase. The methylase interacted with snoRNAs, being guided or tethered by these to specific sites on the pre-rRNA. The expectation is that this methylase has been retained in the modern eukaryotes and archaea but lost from bacteria. By virtue of its non-specific dsRNA methylating activity, this enzyme is hypothesised to have been recruited into methylation of genomic RNA, improving the stability of the genetic material prior to the origin of DNA. Methylation may initially have been carried out by a ribozyme that associated with each of the methylating snoRNAs, or alternatively methylation could have been carried out by the snoRNAs themselves. Methylase: blue ellipse; methyl groups: red lollipops.

Figure 4. Stepwise increases in genome coding capacity during evolution. The events leading to the origin of DNA can be broken down into a number of steps, with selection acting at each step. This Darwinian approach to understanding complex systems can be further understood in terms of the Eigen limit (see text). In the case of genome evolution, it can be described in terms of a positive-feedback cycle: (1) Each newly selected function permits a greater potential coding capacity. (2) In turn, this permits new functions to arise. (3) These functions are selected for, thus permitting a greater potential coding capacity, and so the process is a positive feedback cycle (This we call the Darwin–Eigen cycle [9], depicted schematically in pink). Our model cannot determine whether the RNA polymerase arose before the dsRNA methylase or vice versa; we consider both to be essential steps in the origin of DNA, irrespective of order. The later stages, where U is replaced by T, after the evolution of ribonucleotide reduction, are included for completeness and will be discussed elsewhere (Poole, Penny and Sjöberg, submitted). Abbreviations: 2′-O-Me, ribose 2′-O-methylation; RNA/DNA pol., RNA/DNA polymerase; Ur-RNR, ancestral ribonucleotide reductase; U-N-glycosylase, Uracil-N-glycosylase; TS, thymidylate synthase.
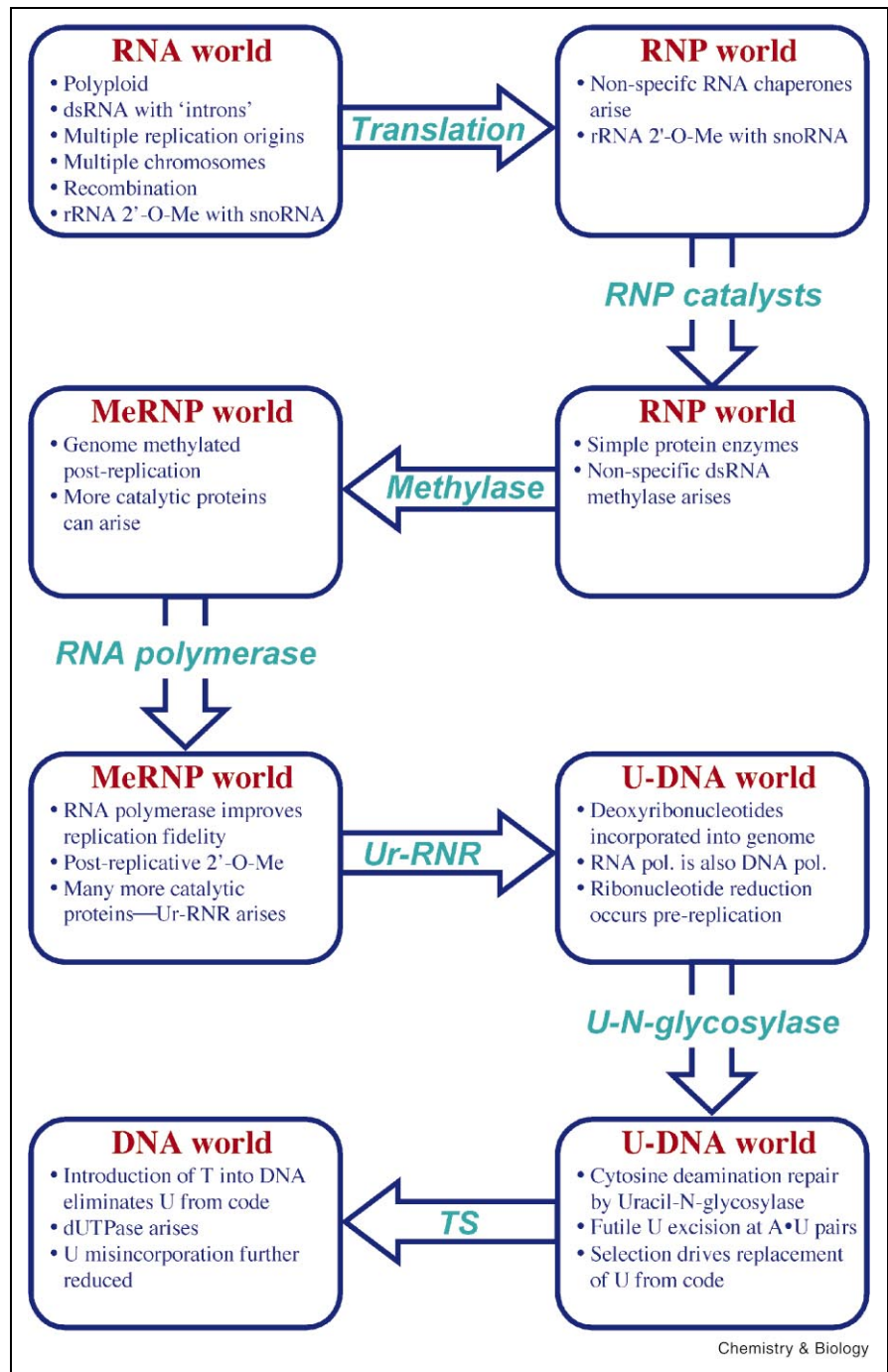
By virtue of its recognition of dsRNA, such a methylase might also have been able to modify genomic dsRNA (Figure 3B). The predicted stabilising effect of 2′-O-methylation on genomic RNA argues that the coding capacity of the genome would have increased sufficiently for genuinely catalytic proteins such as RNA polymerase, and later ribonucleotide reductase, to arise, paving the way to the DNA world (Figure 4). A crucial point is that the methylase could recognise dsRNA in a non-specific manner, allowing it to act on the, presumably, dsRNA of the genome. This amounts to recruitment from RNA post-transcriptional methylation into RNA post-replicative methylation. In this model, no snoRNAs are required for genome methylation, and because the position of methylation is less crucial than for functional RNA, methylation at many points on the genome would be advantageous (Figure 3B). In

contrast, non-specific methylations of functional RNAs could result in silencing of residues crucial for catalysis and loss of function [46].

Structural studies of 2′-O-methyl RNA [47,48] provide insight into the feasibility of a methyl-RNA genome, demonstrating that methylation has a minimal effect on the structure of duplex RNA, and that it is A-form, suggesting that partially or fully methylated dsRNA might have the capacity to carry genetic information.

A non-specific methylase would have offered an improved system for post-transcriptional modification of pre-rRNA at regions specified by the binding of complementary snoRNA. The methylase we envisage is tethered at the snoRNA–rRNA duplex by some feature of the snoRNA, possibly the C and/or D box [29,49], forcing methylation 5 nt from the D box as per the contemporary system (Figure 3B).

The implication of our model is that the function of this enzyme has been preserved from before the origin of DNA, and we predict that it will be identified in all eukaryotes and archaea that use snoRNAs. We suggest that the methylase that acts in concert with snoRNAs is also non-specific for dsRNA, and may not in itself be catalytic. Our model proposes a function that is a consequence of snoRNA–pre-rRNA interaction, and the difficulty in isolating this methylation activity may in part be due to its non-specific affinity for duplex RNA along with a 'chaperone-like' (as defined in [23]) role in catalysis. A crucial test for feasibility of our model would be to examine whether or not the methylase can carry out non-specific methylation of duplex RNA.

Eichler et al. [50] have partially purified a nucleolar 2′-O-methyltransferase, demonstrating that it acts on ribosomal RNA [50,51]. Furthermore, it was shown to carry out non-specific methylation in vitro; the external transcribed spacer of 18S rRNA, which is not modified in vivo, also becomes methylated, suggesting that there are factors present in vivo (possibly snoRNAs) that limit or direct methylation. Furthermore, under in vitro conditions, approximately 6% of all residues on the pre-18S rRNA were methylated after 1 h, in contrast to 1.2% modification observed in vivo [51]. Most strikingly, the partially purified methyltransferase also non-specifically methylates double-stranded poly(A)•poly(U) RNA (as well as single-stranded poly(C) and poly(A) RNA) [50]). Once the nucleolar 2′-O-methyltransferase has been fully purified, it will be of great interest to confirm these results, verifying AdoMet requirement [39] and the non-specific RNA methylation activity reported [51].

Why DNA?
If 2′-O-methyl RNA is a candidate intermediate between

RNA and DNA, what selection was there to go from methyl-RNA to DNA? If silencing of the 2′-OH provides improved stability there must be an additional explanation for the origin of deoxyribonucleotides. One explanation is that the genomic RNA methylation we describe is post-replicative (Table 2). Ribonucleotide reductase, however, performs reduction of the 2′-OH to 2′-H pre-replicatively. The advantage is clear: guaranteed complete modification, and at no stage is the genome susceptible to 2′-OH-mediated self-cleavage.

Completeness of genomic RNA methylation depends on the processivity of the methylase, and if it is non-specific, it may not necessarily methylate every residue. Certainly it seems improbable that the methylase reached the level of processivity of RNA polymerases. Complete methylation may in fact be a hindrance, since a hydrophobic 'cushion' would form in the deep groove of the RNA [47,48]. Indeed the partially purified 2′-O-methyltransferase shows preference for unmethylated RNA over methylated RNA [51], and increased hydrophobicity might be one explanation for this. Lastly, incomplete, non-specific methylation would have its drawbacks, in that there would be no means of repeatedly methylating the most susceptible sites. The evolution of ribonucleotide reduction would solve these problems (Table 2): complete pre-replicative modification of the 2′-OH is possible; deep groove hydrophobicity is avoided; B form DNA (where bases are buried in the centre of the double helix) may be less likely to suffer base damage than unmethylated or partially methylated duplex RNA.

Conclusions
In this article we have reviewed our current understanding of the evolutionary transition from RNA to DNA, and have identified a gap: information theory argues that the limited coding capacity of RNA is the major obstacle in the transition to the DNA world. We have described one possible solution to this problem. By introducing an intermediate step, post-replicative 2′-O-methylation, coding capacity is increased via a positive feedback loop (Figure 4).

While we do not hold that the model is necessarily correct, it is falsifiable and draws attention to a problem with the RNA world model that is often overlooked. Our model suggests 2′-O-methylation as an early alternative to reduction of the reactive 2′-hydroxyl of RNA. The methylase we describe is predicted to augment a spontaneous reaction, and a simple protein, rich in basic residues, may possibly fulfil this role. A simple system, which offers some of the benefits of DNA, would permit a larger genome size, paving the way for more complex catalysts such as ribonucleotide reductase to evolve.

How can the model be tested? If it can be demonstrated that a ribozyme can carry out ribonucleotide reduction,

solving the chemical problem of reducing the ribosyl 2′-OH within an RNA world, the model is falsified. Likewise, demonstration of a less chemically complex deoxyribonucleotide synthesis pathway, such as from acetaldehyde and glyceraldehyde-3-phosphate (degradation products in deoxyribonucleoside salvage), that abrogates the chemistry of ribonucleotide reduction, would falsify the hypothesis. Finally, a demonstration that early RNA genomes could have held much larger amounts of genetic information than currently supposed would also be a falsification of the hypothesis.

Note added in proof: Since the writing of this article, the crystal structure of an archaeal fibrillarin-homologue has been determined [55]. The C-terminal domain is similar in folding to AdoMet-dependent methyltransferases, with greatest similarity to catechol *O*-methyltransferase. Methylating snoRNAs have been detected in Archaea [37] so the fibrillarin structure is particularly relevant to our hypothesis. The structure of the fold is consistent with our contention that the catalytic role of the methyltranferase may simply be to provide a positive charge to aid leaving of the methyl group from AdoMet.

Acknowledgements

References
1. Gilbert, W. (1986). The RNA world. *Nature* **319**, 618.
2. Maizels, N. & Weiner, A.M. (1999). The genomic tag hypothesis: what molecular fossils tell us about the evolution of tRNA. In *The RNA World*, 2nd edn. (Gesteland, R.F., Cech, T.R. & Atkins, J.F., eds), pp. 79–111, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
3. Benner, S.A., Ellington, A.D. & Tauer, A. (1989). Modern metabolism as a palimpsest of the RNA world. *Proc. Natl. Acad. Sci. USA* **86**, 7054–7058.
4. Jeffares, D.C., Poole, A.M. & Penny, D. (1998). Relics from the RNA world. *J. Mol. Evol.* **46**, 18–36.
5. von Ahsen, U. & Noller, H.F. (1995). Identification of bases in 16S rRNA essential for tRNA binding at the 30S ribosomal P site. *Science* **267**, 234–237.
6. Lodmell, J.S. & Dahlberg, A.E. (1997). A conformational switch in Escherichia coli 16S ribosomal RNA during decoding of messenger RNA. *Science* **277**, 1262–1267.
7. Khaitovich, P., Mankin, A.S., Green, R., Lancaster, L. & Noller, H.F. (1999). Characterization of functionally active subribosomal particles from *Thermus aquaticus*. *Proc. Natl. Acad. Sci. USA* **96**, 85–90.
8. Lazcano, A., Guerrero, R., Margulis, L. & Oró, J. (1988). The evolutionary transition from RNA to DNA in early cells. *J. Mol. Evol.* **27**, 283–290.
9. Poole, A.M., Jeffares, D.C. & Penny, D. (1999). Early evolution: prokaryotes, the new kids on the block. *BioEssays* **21**, 880–889.
10. Reichard, P. (1997). The evolution of ribonucleotide reduction. *Trends Biochem. Sci.* **22**, 81–85.
11. Logan, D.T., Andersson, J., Sjöberg, B.-M. & Nordlund, P. (1999). A glycyl radical site in the crystal structure of a class III ribonucleotide reductase. *Science* **283**, 1499–1504.
12. Tauer, A. & Benner, S.A. (1997). The B12-dependent ribonucleotide reductase from the archaebacterium *Thermoplasma acidophila*: an evolutionary solution to the ribonucleotide reductase conundrum. *Proc. Natl. Acad. Sci. USA* **94**, 53–58.
13. Riera, J., Robb, F.T., Weiss, R. & Fontecave, M. (1997). Ribonucleotide reductase in the archaeon *Pyrococcus furiosus*: a critical enzyme in the evolution of DNA genomes? *Proc. Natl. Acad. Sci. USA* **94**, 475–478.
14. Sjöberg, B.-M. (1997). Ribonucleotide reductases – a group of enzymes with different metallosites and a similar reaction mechanism. *Struct. Bond.* **88**, 139–173.
15. Celander, D.W. & Cech, T.R. (1990). Iron(II)–ethylenediaminetetraacetic acid catalysed cleavage of RNA and DNA oligonucleotides: similar reactivity toward single- and double-stranded forms. *Biochemistry* **29**, 1355–1361.
16. White, H.B. (1982). Evolution of coenzymes and the origin of pyridine nucleotides. In *The Pyridine Nucleotide Coenzymes* (Everse, J., Anderson, B. & You, K., eds), pp. 1–17, Academic Press, New York.
17. Steitz, T.A. (1998). A mechanism for all polymerases. *Nature* **391**, 231–232.
18. Gao, G., Orlova, M., Georgiadis, M.M., Hendrickson, W.A. & Goff, S.P. (1997). Conferring RNA polymerase activity to a DNA polymerase: a single residue in reverse transcriptase controls substrate selection. *Proc. Natl. Acad. Sci. USA* **94**, 407–411.
19. Astatke, M., Ng, K., Grindley, N.G.F. & Joyce, C.M. (1998). A single side chain prevents *Escherichia coli* DNA polymerase I (Klenow fragment) from incorporating ribonucleotides. *Proc. Natl. Acad. Sci. USA* **95**, 3402–3407.
20. Gorbalenya, A.E. & Koonin, E.V. (1993). Helicases: amino acid sequence comparisons and structure–function relationships. *Curr. Opin. Struct. Biol.* **3**, 419–429.
21. Eigen, M. (1993). The origin of genetic information: viruses as models. *Gene* **135**, 37–47.
22. Atkins, J.F. (1993). Contemporary RNA genomes. In *The RNA world* (Gesteland, R.F. & Atkins, J.F. eds), pp. 535–556, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
23. Poole, A.M., Jeffares, D.C. & Penny, D. (1998). The path from the RNA world. *J. Mol. Evol.* **46**, 1–17.
24. Hasegawa, M., Yano, T. & Miyata, T. (1984). Evolutionary implications of error amplification in the self-replicating and protein-synthesizing machinery. *J. Mol. Evol.* **20**, 77–85.
25. Sutherland, J.D. & Whitfield, J.N. (1997). Prebiotic chemistry: a bioorganic perspective. *Tetrahedron* **53**, 11493–11527.
26. Scott, W.G. & Klug, A. (1996). Ribozymes: structure and mechanism in RNA catalysis. *Trends Biochem. Sci.* **21**, 220–224.
27. Fraser, C.M., Gocayne, J.D., White, O., Adams, M.D., Clayton, R.A., Fleischmann, R.D., Bult, C.J., Kerlavage, A.R., Sutton, G., Kelley, J.M., Fritchman, J.L., Weidman, J.F., Small, K.V., Sandusky, M., Fuhrmann, J., Nguyen, D., Utterback, T.R., Saudek, D.M., Philips, C.A., Merrick, J.M., Tomb, J.-F., Dougherty, B.A., Bott, K.F., Hu, P.-C., Lucier, T.S., Peterson, S.N., Smith, H.O., Hutchinson III, C.A. & Venter, J.C. (1995). The minimal gene complement of *Mycoplasma genitalium*. *Science* **270**, 397–403.
28. Agris, P.F. (1996). The importance of being modified: roles of modified nucleosides and Mg^{2+} in RNA structure and function. *Prog. Nucleic Acid Res. Mol. Biol.* **53**, 79–129.
29. Bachellerie, J.-P. & Cavaillé, J. (1998). Small nucleolar RNAs guide the ribose methylations of eukaryotic rRNAs. In *Modification and Editing of RNA* (Grosjean H & Benne R., eds), pp. 255–272, ASM Press, Washington, DC.
30. Martínez Giménez, J.A., Sáez, G.T. & Seisdedos, R.T. (1998). On the function of modified nucleosides in the RNA world. *J. Theor. Biol.* **194**, 485–490.
31. Cermakian, N. & Cedergren, R. (1998). Modified nucleotides always were: an evolutionary model. In *Modification and Editing of RNA* (Grosjean, H. & Benne, R., eds), pp. 535–541, ASM Press, Washington, DC.
32. Pressing, J. & Reanney, D.C. (1984). Divided genomes and intrinsic noise. *J. Mol. Evol.* **20**, 135–146.
33. Steel, M. & Penny, D. (2000). Parsimony, likelihood and the role of models in molecular phylogenetics. *Mol. Biol. Evol.* **17**, 839–850.
34. Joyce, G.F. & Orgel, L.E. (1999). Prospects for understanding the origin of the RNA world. In *The RNA World*, 2nd edn (Gesteland R.F., Cech, T.R. & Atkins, J.F. eds), pp. 49–77, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
35. Nelson, K.E., Levy, M. & Miller, S.L. (2000). Peptide nucleic acids rather than RNA may have been the first genetic molecule. *Proc. Natl. Acad. Sci. USA* **97**, 3868–3871.
36. Lafontaine, D.L.J. & Tollervey, D. (1998). Birth of the snoRNPs: the evolution of the modification-guide snoRNAs. *Trends Biochem. Sci.* **23**, 383–388.
37. Omer, A.D., Lowe, T.M., Russell, A.G., Ebhardt, H., Eddy, S.R. &

Dennis, P.P. (2000). Homologues of small nucleolar RNAs in Archaea. *Science* **288**, 517–522.

38. Tycowski, K.T., You, Z.-H., Graham, P.J. & Steitz, J.A. (1998). Modification of U6 spliceosomal RNA is guided by other small RNAs. *Mol. Cell* **2**, 629–638.

39. Segal, D.M. & Eichler, D.C. (1989). The specificity of interaction between *S*-adenosyl-L-methionine and a nucleolar 2′-*O*-methyltransferase. *Arch. Biochem. Biophys.* **275**, 334–343.

40. Vidgren, J., Svensson, L.A. & Liljas, A. (1994). Crystal structure of catechol *O*-methyltransferase. *Nature* **368**, 354–358.

41. Hodel, A.E., Gershon, P.D., Shi, X. & Quiocho, F.A. (1996). The 1.85 Å structure of Vaccinia protein VP39: a bifunctional enzyme that participates in the modification of both mRNA ends. *Cell* **85**, 247–256.

42. Woodward, R.W., Tsai, M.-D., Floss, H.G., Crooks, P.A. & Coward, J.K. (1980). Stereochemical course of the transmethylation catalysed by catechol *O*-methyltransferase. *J. Biol. Chem.* **19**, 9124–9127.

43. Lindahl, T. (1993). Instability and decay of the primary structure of DNA. *Nature* **362**, 709–715.

44. Kimzey, A.L. & McFadden, P.N. (1994). Spontaneous methylation of hemoglobin by *S*-adenosyl-methionine by a specific and saturable mechanism. *J. Protein Chem.* **13**, 537–546.

45. Barrows, L.R. & Magee, P.N. (1982). Nonenzymatic methylation of DNA by *S*-adenosylmethionine in vitro. *Carcinogenesis* **3**, 349–351.

46. Scott, W.G., Finch, J.T. & Klug, A. (1995). The crystal structure of an all-RNA hammerhead ribozyme: a proposed mechanism for RNA catalytic cleavage. *Cell* **81**, 991–1002.

47. Popenda, M., Biala, E., Milecki, J. & Adamiak, R.W. (1997). Solution structure of RNA duplexes containing alternating CG base pairs: NMR study of r(CGCGCG)2 and 2′-*O*-Me(CGCGCG)2 under low salt conditions. *Nucleic Acids Res.* **25**, 4589–4598.

48. Adamiak, D.A., Milecki, J., Popenda, M., Adamiak, R.W., Dauter, Z. & Rypniewski, W.R. (1997). Crystal structure of 2′-*O*-Me-(CGCGCG)2, an RNA duplex at 1.30 Å resolution. Hydration pattern of 2′-*O*-methylated RNA. *Nucleic Acids Res.* **25**, 4599–4607.

49. Kiss-László, Z., Henry, Y. & Kiss, T. (1998). Sequence and structural elements of methylation guide snoRNAs essential for site-specific ribose methylation of pre-rRNA. *EMBO J.* **17**, 797–807.

50. Eichler, D.C., Raber, N.K., Shumard, C.M. & Eales, S.J. (1987). Isolation and characterisation of a nucleolar 2′-*O*-methyltransferase from Ehrlich ascites tumor cells. *Biochemistry* **26**, 1639–1644.

51. Eichler, D.C. (1994). Characterisation of a nucleolar 2′-*O*-methyltransferase and its involvement in the methylation of mouse precursor ribosomal RNA. *Biochimica* **76**, 1115–1122.

52. Sclavi, B., Woodson, S., Sullivan, M., Chance, M.R. & Brenowitz, M. (1997). Time-resolved synchotron X-ray 'footprinting', a new approach to the study of nucleic acid structure and function: application to protein–DNA interactions and RNA folding. *J. Mol. Biol.* **266**, 144–159.

53. Yarus, M. (1999). Boundaries for an RNA world. *Curr. Opin. Chem. Biol.* **3**, 260–267.

54. Carola, C. & Eckstein, F. (1999). Nucleic acid enzymes. *Curr. Opin. Chem. Biol.* **3**, 274–283.

55. Wang, H., Boisvert, D., Kim, K.K., Kim, R. & Kim, S.-H. (2000). Crystal structure of a fibrilarin homologue from *Methanococcus jannaschi*, a hyperthermophile, at 1.6 Å. *EMBO J.* **19**, 317–323.