

CHAPTER 19

The Interplay between RNA and DNA Modifications: Back to the RNA World

Patrick Forterre* and Henri Grosjean

Abstract

An RNA/protein world (probably cellular) is widely accepted as a probable step in the early evolution of life. During subsequent life evolution various enzymes emerged that allowed some organisms to generate deoxyribonucleotides from ribonucleotide precursors and to synthesize DNA molecules using ancestral RNA genomes as templates. Later on, once the DNA became the major repository of genetic information, cells and viruses had to develop new strategies to protect their DNA genomes against the aggressive chemical environment and/or destructive enzymes produced by competitors. This was performed mainly by further modifying the DNA genomes after their synthesis through pre and postreplicative enzymatic processes, using enzymes, mostly methyltransferases and deaminases that were probably initially designated to modify primordial RNAs. Relics of this ancient interplay between RNA and DNA modification in cells and viruses are still abundant in our modern 'DNA-makes-RNA-makes-proteins' world. Some enzymes are still able today to modify both DNA and RNA, demonstrating the versatility of the modification apparatus and testifying for the time when proteins of the late RNA world were recruited to work with DNA. Here, we review what is known about these enzymes that were designated to synthesize DNA within the framework of a hypothetical cellular and viral co-evolution. From this analysis, DNA appears as just another type of (hyper)modified RNA polymer, specialized in storing the genetic inheritance of the living organisms.

Introduction

For a long time DNA was viewed as the cornerstone of all biology, the mastermind of life, the aperiodic crystal whose existence had been wisely predicted by Schroedinger in his famous book "What is life?".¹ For Monod, a first meaningful DNA molecule appeared by chance in the primordial ocean and life began to emerge, just as Venus emerged from her shell.² In this scenario, RNA molecules appears only as an intermediates used by DNA to perform its task, connecting informational content of its sequence to biologically active protein tools via passive "messenger" RNAs. Why was DNA not directly translated to proteins then? Why do giant complex ribonucleoproteins (ribosomes) need RNA messenger, not "DNA messenger" to decode the genetic information?

From molecular designer's viewpoint, a DNA-based translation machinery would have been obviously more rational. However, living organisms are not rational machines, but historical products. Nowadays, molecular biologists have completely reevaluated the role of RNA in modern life and in

*Corresponding Author: Patrick Forterre—Institut Pasteur, Département de Microbiologie, 25 rue du Dr Roux, 75015, Paris, France. Email: forterre@pasteur.fr

life history: DNA is not the *deus ex machina* of the living world but, more simply, an RNA offspring and modern RNA molecules (cellular and viral) are now considered as vivid and fascinating relics of our distant past (the RNA world).³ The flow of information in present day 'ribosome encoding organisms',⁴ which has been enshrined in the central dogma 'DNA-makes-RNA-makes proteins',⁵ is itself the product of history.⁶ From an evolutionary point of view, one can identify several steps in this emerging pathway:⁷ first 'RNA-makes-RNA', second 'RNA-makes-proteins', third, "proteins make RNA" and finally "proteins-makes-DNA (from RNA)" (Fig. 1, boxes A-C).

DNA thus appears as a by-product of the RNA world evolution and not the opposite. As a matter of facts, in every cell examined so far, the deoxyribonucleotide precursors of DNA are produced by the enzymatic modification of RNA precursors (ribonucleotides) in which every 2'-hydroxyl group of ribose is reduced to a simple hydrogen and all the uracil bases are methylated to 5-methyl-uracil (thymine). In modern organisms, many cytidines and adenines in DNA are further deaminated to uracil or hypoxanthine, or modified to yield methylated derivatives and the enzymes involved in such postreplicative chemical alterations have many analogies with the homologous posttranscriptional base modification machinery in RNA (see below). Modern DNA thus appears as a heavily (hyper)modified RNA molecule.

In the early RNA world, the intrinsic chemical instability of RNA has probably limited the genome size, base modification can therefore have played very early a role in the evolution of the first living organisms toward more complexity, allowing the formation of longer proto-genomes

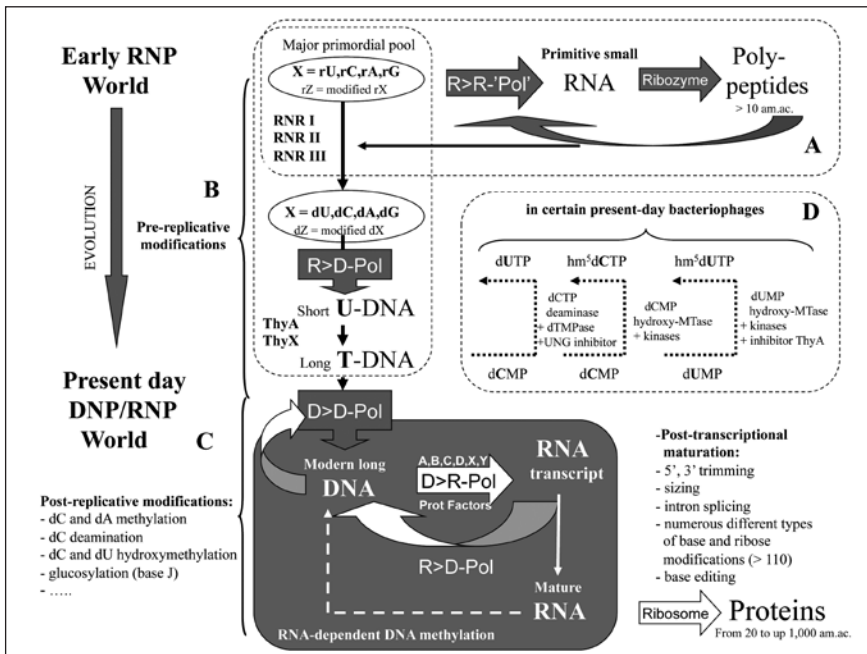


Figure 1. Various steps for the emergence of the contemporary 'DNA-makes-RNA-makes-Proteins' biosynthetic system (box C) from a simpler ancestral 'RNA-directed polypeptides biosynthetic' machinery (primordial step A). In between are the fundamental catalytic processes that allowed to stepwise generate deoxyribonucleotides from ribonucleotides by ribonucleotide reductases (RNR I, II, III), the formation of the first U-containing deoxy-polymer by RNA-dependent DNA polymerase ($R > D\text{-Pol}$) and formation of TMP from dUMP that led to the generation of modern T-containing DNA (box B). In box D are a few reactions that allow the generation of modified derivatives of dCMP and dUMP to be incorporated by a prereplicative process, as exists in certain present-day bacteriophages.

and ribozymes. The fight between parasites and their targets probably started early on (for instance between RNA viruses and RNA cells), also creating many opportunities for ribose and base modifications to be selected in the course of evolution. The fine-tuning of ribozymes by base modification (a possibility presently unexplored in typical in vitro RNA world experiments) might also have been critical for the evolution of complex functions before the advent of modern proteins. The aim of this chapter is to identify in our modern RNA/DNA/Protein world relics of biosynthetic and modification machineries that could testify for ancient processes and could shed light on the RNA to DNA transition.

Early Pathways from RNA to Modern DNA (T-DNA)

Reduction of ribose to deoxyribose and transformation of dUMP to dTMP (to replace uracil in RNA by thymine in DNA) are performed in modern cells by sophisticated enzymes: ribonucleotide reductases (Fig. 1, box B and Fig. 2, box A) and thymidylate synthases (box B in Figs. 1 and 2). The order of emergence of these two enzymes can be deduced from the fact that dUMP is the substrate of thymidylate synthase. Thus ribonucleotide reductase should have originated first, followed by a polymerase that could synthesise U-containing DNA (U-DNA) using originally an RNA template (reverse transcription) and later on a U-DNA template. Thymidylate synthases originated later on, producing thymidine-containing DNA (T-DNA—Fig. 2, box B). Ribonucleotide reductases and thymidylate synthases are encoded in all cellular genomes and in the genomes of many DNA viruses. The complexity of the reaction catalyzed by the tetrahydrofolate-dependent thymidylate synthases (reviewed in chapter by Myllykallio et al in this volume), as well as the reductant-dependent ribonucleotide reductases—both types of enzymes requiring radical intermediate species—indicates that these reactions could not have been performed by ribozymes, or by polypeptides.⁷⁻⁹ Probably

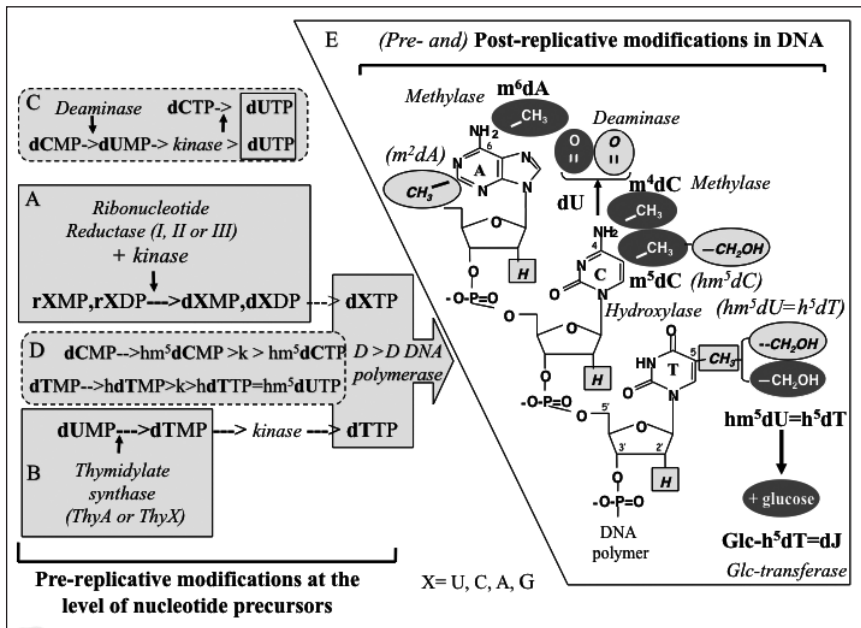


Figure 2. Compilation of the most frequent present-day various types of prereplicative and postreplicative modifications identified DNA. In box E, when in light grey boxes, in italics and not in bold, it means that chemical modification occurs at the nucleotide precursors level (prereplicative process). Detailed information is in text. See also two review papers by Warren²⁵ and by Gommers-Ampt and Borst²⁶ for other rare types of modified nucleotides found mainly in certain viruses.

these reactions have been always performed by very sophisticated proteins. The first ribonucleotide reductases and thymidylate synthases were thus made by ancestral ribosomes containing both RNA and proteins and that were capable to perform already accurate translation. The RNA to DNA transition thus should have occurred in a complex cellular environment suitable for the production of these enzymes. This environment had to be elaborated enough to support the entire metabolism for the production of RNA precursors (rNTPs), including mechanisms for energy production. Hence, the cellular environment in which DNA finally emerged was not as "simple" as sometimes imagined, but was certainly populated by elaborated cells and viruses with an already complex metabolic network and well-organized membrane systems.⁷

Besides ribonucleotide reductase and later on thymidylate synthase, the newly emerging thymidine-containing DNA synthesizing systems also had to involve a primordial RNA-dependent DNA polymerase (a sort of reverse transcriptase) capable of accurately incorporating the newly emerging deoxynucleotide triphosphates into DNA-like polymers. These enzymes probably evolved from RNA-dependent RNA polymerases, since modern reverse transcriptases encoded by retroviruses are evolutionarily related to RNA-dependent RNA polymerases.¹⁰ Later on, bona fide DNA polymerases probably originated from reverse transcriptases. DNA polymerases of the A family are also evolutionarily related to both RNA polymerases and reverse transcriptase. It is likely that ancestral RNA polymerases were quite flexible in their substrate requirement and that this flexibility has facilitated the various enzymatic transitions leading from RNA to DNA. Indeed, some modern polymerases can relatively easily change their substrate; for instance, the DNA-dependent RNA polymerase II that transcribes much of protein-encoding genes in Eukarya, can also be used to replicate and transcribe RNA by some viruses.¹¹ The actors involved in DNA emergence are thus well delineated. The fundamental question to answer is why was DNA invented after all? The usual response is because DNA is chemically more stable than RNA.

Importance of DNA Stability

The presence of the oxygen in the 2' position of the ribose is the major cause of RNA instability when tested *in vitro*, outside a cellular context. Indeed, this reactive oxygen can attack and cleave 5'-3' phosphodiester bonds, a reaction that is considerably enhanced in the presence of bivalent ions and by increasing the temperature.¹² Hence, DNA lacking this 2' hydroxyl group, when tested under identical conditions as RNA of the same size has a much longer lifetime. Moreover, the modification of uracil to thymidine (C5-methylation) further stabilizes the DNA by allowing the formation of rigid stretches of double stranded helices, possibly supercoiled into a very compact tertiary structure, a conformation that is much more stable than any RNA in solution. Lastly, once free rotation between the two DNA strands is prevented by circularization (in covalently closed circular molecules) DNA become fairly resistant to thermodenaturation.¹³

Because of its superior stability over RNA, it is generally assumed that DNA was naturally selected as genetic material, allowing the emergence of larger genomes, a prerequisite for the evolution of life toward more complexity.¹⁴ In this classical scenario of DNA emergence, once a hypothetical primordial cell has acquired this useful polymer by chance, it was selected (by necessity?) to become the forerunner of all subsequent evolution of life on our planet. However in a Darwinian explanation, one has to explain also why the first organism with a DNA genome was selected over its neighbours, independently of the success of its descendants; why a rapidly dividing small RNA cell with a short genome have found an immediate advantage in having a more stable genome?

Some scientists argue that the Last Universal Cellular Ancestor (LUCA) was a thermophilic DNA-based cell because it had to survive the terrible late heavy bombardment (LHB) of meteorites that nearly destroyed our planet 3.9 Gy ago.¹⁵ However, other argued that this bombardment was so drastic that life could have only appeared after it was over.¹⁶ Even if DNA-containing cells equipped to sustain very high temperature survived the LHB, whereas cells with RNA genomes were eliminated, this does not explain how the ancestor of these DNA survivors emerged in the first place.

In fact, the problem raised by RNA instability should not be overestimated, especially if one considers the normal cellular context. After all, modern hyperthermophiles can thrive at temperatures as high as 110°C, despite the thermosensitivity of their numerous and complex RNA molecules. In such contemporary living cells, RNA molecules are protected against the effect of temperature by various tricks: interactions with monovalent cations, polyamines and other macromolecules, including small RNAs (reviewed in refs. 17, 18). Most of these stabilizing devices were probably present at the time of DNA emergence, i.e., in the late step of the RNA world. Also, simple modified nucleotide precursors, including methylated base and/or ribose, pseudouridine, inosine ..., possibly produced by ancestral active ribozymes (see symbol 'Z' in Fig. 1, box A), were almost certainly present in the RNA world,¹⁹ thus allowing the generation of chemically modified and hence more stable RNA molecules.

As an alternative to the "stability theory" for the origin of DNA, one of us has suggested that DNA emerged as a byproduct of the competition between RNA viruses and DNA cells.^{7,20} Below, we discuss this problem, first in the light of strategies used by modern cells and viruses and on analogies between a few modification enzymes acting on DNA and RNA that might have a common origin.

Nucleotide Modifications in the Context of Present-Day Viruses/Cell Competition

A classical way for a virus to kill its host cell (and conversely for the cell to get rid of the virus) is to degrade the genome of the enemy. Viruses and cells have therefore developed various strategies in a dramatic arms race to prevent this process. One of these strategies is to physically separate the genomes from the proteins of the adversary. For instance many RNA viruses confine their genomes during the infection process either inside a viral core particle or inside a viral factory physically separated from the cytoplasm of the infected cells, in order to protect their genomes against the nucleases of their unwilling "host". However the major strategy used by both viruses and cells of all domains of life is to differentiate their genomes from those of their "host" by chemically modifying selected bases at specific crucial locations (molecular shields). This "molecular 'apartheid'" can take place either before or after DNA replication, using specific DNA modification enzymes (Fig. 2, boxes D, E). This is exemplified by the restriction-modification systems widely present in the domains Archaea and Bacteria (see chapters by White et al and by Coffin et al in this volume).

Likewise, specific enzymes such as Apobec3 (DNA-Cytidine Deaminase, see below) or AID (Activation-Induced Deaminase) exist in mammalian cells that deaminate selected cytidine residues to uracils in the invading viruses. This mechanism allows protection against viruses (see chapter by Smith) or of the genomic DNA (DNA editing, somatic hypermutation), thus allowing the generation of immunoglobulins diversity that also serves in antiviral host defence (see chapter by Parisien and Bhagwat). In Eukarya, RNAs produced by viruses, transgenes and transposons and host DNA methyltransferases are used to develop a variety of defence mechanisms including RNA degradation, gene silencing by small noncoding (nc) RNAs and heterochromatin formation through cytosine methylation and demethylation of DNA (see refs. 21-33 for examples; and chapter by Roldan-Arjona and Ariza). This seems also the case in Archaea.²⁴

In cellular organisms, all postreplicative modifications of DNA concern only a few percent of the DNA bases (box E in Fig. 2). However in the genome of some bacteriophages, DNA can be exhaustively modified up to 100% for a given type of base. In these cases, the route of their biosynthesis is prereplicative (Fig. 1, box D and Fig. 2, boxes C, D). For example, in the DNA of bacteriophage T4, all cytosines are replaced by 5-hydroxymethyl-cytosines (hm⁵dC; HMC-DNA) and in the case of DNA of the virus PBS2 infecting *Bacillus subtilis*, all thymines are replaced by uracils (U-DNA). In the case of hm⁵dC-containing DNA, modified hm⁵dCTP precursor is first synthesized by a specific dCTP hydroxymethylase (encoded by T4 bacteriophage), while in the case of U-containing DNA, the dUTP precursor is formed by the viral-coded CMP/CTP deaminase/kinase system before being subsequently incorporated into DNA by a DNA polymerase that is specific of the virus (reviewed in refs. 25, 26). To escape the DNA repair system of the cell host

(here *B. Subtilis*) that normally eliminates uracil residues from DNA by uracil-DNA glycosylase (UNGase), the bacteriophage PB2 induces a viral-encoded specific host UNGase inhibitor and a specific dTMPase to eliminate the competing dTMP produced by the host cell (25 and references therein). These examples illustrate that simply using another DNA precursor (here dUTP or hm⁵dUTP—box D in Fig. 2) implies the setting up of a coordinate network of different enzymatic activities that obviously cost a lot of energy and tricky tinkering during evolution of the cell or the virus. It is worth mentioning that a glucosylated derivative of the same hm⁵dU as mentioned above (designated deoxynucleoside dJ, box E in Fig. 2) is scarcely present in DNA (a few % of the total thymidine) of all unicellular flagellated protozoa of the order of the Kinetoplastida and in the closely related unicellular *Euglena gracilis*. However, at variance with the case of the bacteriophages, both the enzymatic hydroxylation of thymidine and the enzymatic transfer of the glycosyl group to hm⁵dU are formed after the synthesis of the DNA (post-replicative events—see chapter by Sabatini et al).

The presence in viral DNA of modified nucleotides and/or genes coding for DNA or RNA modification may be ubiquitous. Unfortunately, a systematic analysis of huge genomic polymers to detect the presence of only trace amounts of modified nucleotides is technically extremely difficult and probably for that reason has been largely neglected during these last two or three decades. Also, in the case of RNA viruses, such as the plant TMV, TYMV and BMV, the presence of modified nucleosides has long been suspected but not properly investigated. When fragments of such plant genomic RNAs, encompassing characteristic tRNA-like structures, were incubated with cell extracts or with purified recombinant tRNA modification enzymes, m⁵C, m⁵U and pseudouridine were demonstrated to occur.²⁷⁻³⁰ Moreover, since RNA virus genomes contain fairly long stretches of double-stranded RNA, they are also targets of mammalian enzymes of the family ADAR (Adenosine Deaminases acting on RNA) that catalyzes A-to-I editing, as demonstrated both in vivo and in vitro with various animal RNA viruses (see refs. 31, 32 and chapter by Heale and O'Connell).

Thus our knowledge about the presence of very diverse enzymatically modified nucleotides in viral genomes (DNA and RNA) is probably only the tip of the iceberg. Although there are presently only very few examples of modifications in RNA genomes, there is no good reason for the 'self-defence' strategy of genome modification to be an exclusive property of DNA viruses and of modern times.

Versatility of the Modification Apparatus

In this section we examine the origin of the 'pre-replicative' modifying enzymes ribonucleotide reductases (RNR) and thymidylate synthases (Tds) and of a few 'post-replicative' modifying enzymes acting on DNA that belong to the same family as the ones catalyzing the modification of RNA. These modern enzymes might have emerged at the same period in the late RNA/Proteins world and possibly still contain cryptic imprints of the period when DNA genomes emerged and finally replaced RNA genomes in the cellular world.

Ribonucleotide Reductase (RNR)

Ribonucleotide reductases (RNR) are obvious relics of the RNA/Protein world since, producing DNA precursors from RNA precursors, their substrates were present before the emergence of DNA. There presently exist three separate classes of RNA (I, II, III; box A in Fig. 3) with widely divergent amino acid sequences, distinct quaternary structures and different mechanisms (including different cofactors) for the generation of the free radical required for catalysis.³³⁻³⁵ RNRs of class I are dimeric and require oxygen for the generation of a stable tyrosyl radical by a Fe-O-Fe center. The electrons required for the reaction are provided from redox-active cysteines of small proteins, such as thioredoxin or glutaredoxin. They probably originated quite late in evolution since they require oxygen for their catalytic function. In contrast class II RNRs are indifferent to oxygen, whereas class III RNRs are anaerobic enzymes that are inactivated by oxygen. Class II RNRs, which are present in both aerobic and anaerobic organisms (either Archaea and Bacteria) contain

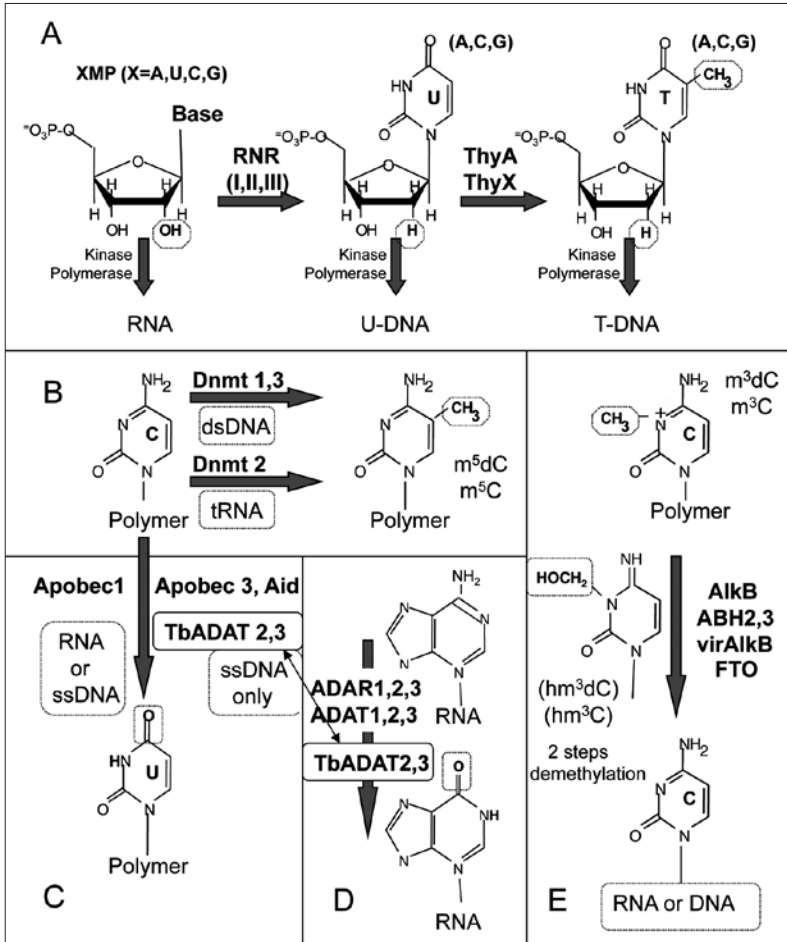


Figure 3. Detailed chemical transformation of bases within nucleotide monomers (prereplicative process—box A) or RNA/DNA polymer (boxes B, C, D, E). The enzymes catalyzing these reactions are indicated by their conventional symbols in bold and their polymer substrates (DNA and/or RNA) are indicated in dashed boxes. ‘ds’ means double-stranded, ‘ss’ means single-stranded, ‘t’ means transfer. All other explanations are in the text.

a single subunit and are isolated either as monomers or as dimers. They generate their tyrosyl radical with the aid of adenosyl-cobalamin. Class III RNR’s are found in both strict anaerobic and facultative anaerobic organisms (either Archaea or Bacteria). They are formed by the association of two dimeric proteins; one of them contains a redox-active (4Fe-4S) center, which together with S-adenosyl-L-methionine and reduced flavodoxin generate a stable but oxygen-sensitive glycy radical. Several bacteria and archaea harbor more than one class of RNR’s and many DNA viruses encode their own RNR’s.

Although very divergent, Class II and III RNR’s exhibit enough sequence similarities to be aligned for phylogenetic analysis. Class I RNR’s appear to be also evolutionary related to the two others, but similarities can only be detected at the structural and mechanistic levels. However, the three classes of RNR’s are built around a homologous protein fold and exhibit striking similarities in the catalytic mechanisms involved in the reduction of ribose.³³ It has been therefore suggested

that the three classes of RNRs arose by divergent evolution from a common ancestral protein that existed before the transition of the RNA to DNA world. However, it is unclear if this protein already exhibited RNR activity. Indeed, the common fold of the various RNRs is also present in pyruvate formate lyase, an anaerobic enzyme of intermediary metabolism that appears also very ancient. It might well be that class III RNRs are the ancestral enzymes, partly because it is more related to pyruvate formate lyase than the two other classes and because it also uses formate, the product of the pyruvate formate lyase reaction, for the reduction of ribonucleotides. If class III RNRs originated first, then RNRs of class II and class I arose later, the later only after accumulation of atmospheric oxygen. Alternatively, the three classes of RNRs might have emerged independently from a protein without RNR activity resembling pyruvate formate lyases (discussed in ref. 9).

Thymidylate Synthases (*ThyA*, *ThyX*)

Thymidylate synthases (box A in Fig. 3) catalyze the formation of ribothymine from the 5'-deoxyuridine monophosphate (dUMP), which itself originates from the deamination of 5'-deoxycytidine monophosphate (dCMP, box D in Fig. 1 and box C in Fig. 2). These enzymes (methyltransferases and deaminases) are obviously modifying enzymes of DNA metabolism acting at the precursor level. In the case of thymidylate synthases, two different, phylogenetically unrelated and mutually exclusive types of enzymes have been discovered in modern organisms (conversion type of evolution). In one case, thymidylate synthase A (*ThyA*) uses methylene tetrahydrofolate as carbon source and an external reducing agent to form thymidylate DNA precursor and dihydrofolate, while in the second case, *ThyX* is a flavoprotein (FADH₂-containing enzyme) using the same methylene tetrahydrofolate to directly form thymidylate and tetrahydrofolate (see chapter by Myllykallio et al).

Interestingly, a single change of a conserved Asn to Asp in the active site of *ThyA* of *Escherichia coli* or *Lactobacillus casei* allowed these methylases to methylate also dCMP in addition of the normal dUMP substrate.^{36,37} Furthermore, when this mutation was coupled with another mutation in the same active site of *ThyA* (a conserved His), the enzyme lost all its capability to methylate the C5 atom of dUMP and selectively methylated only dCMP into m⁵dCMP.³⁸ These experiments demonstrate the easiness of this particular C5-methylase to shift specificity for one type of pyrimidine to another, a situation that might have occurred during cell evolution.

DNA Methyltransferase (*Dnmt2*)

The AdoMet-dependent DNA methyltransferases, Dnmts (for DNA methyltransferases) are widely distributed in mammals and vertebrates. They catalyze the site-specific formation of 5-methyldeoxycytosine (m⁵dC) mostly in dCpdG dideoxynucleotides of double-stranded DNA substrates (Fig. 3, box B). They can be divided into three major classes based on substrate specificity: Dnmt1, Dnmt2 and Dnmt3; the latter itself is subdivided into three subfamilies: 3a, 3b and 3L (see chapters by Cheng and Blumenthal and by Coffin et al in this volume).

Mammalian Dnmt2 family is relatively small in size (391 amino acids as compared to 850-910 for Dnmt3ab and 1616 for Dnmt1). It is also the most conserved and the best representative of the original protein from which all other Dnmt enzymes probably derived (reviewed in 39). Ironically, at variance with the other Dnmt enzymes for which the ability for the formation of 5-methyldeoxycytosine (m⁵dC) at specific sites in double-stranded DNA has been clearly demonstrated, the function of Dnmt2 long remained enigmatic due to its poor activity with double-stranded DNA. Only recently, by examining Dnmt2 knockout mutants in diverse eukaryotic cells, it was discovered that the main function of this putative DNA-specific enzyme is rather to be a tRNA modifying enzyme that catalyzes the specific methylation of a ribocytosine located in the CpG diribonucleotides of the anticodon branch (position C₃₈G₃₉) of mammalian tRNA-Asp.⁴⁰ Comparison of crystal structures of human Dnmt2 and the genuine bacterial DNA:m⁵C forming methyltransferase M.HhaI shows that they are structurally very similar over the large domain and that all the conserved Mtase motifs needed for the catalytic function of the enzyme are present in a well-organized active site with every side chain in the correct orientation.⁴¹ One main difference however between the evolutionarily divergent Dnmt1/3 and Dnmt2 protein family is the

presence in the former of an additional N-terminal extension that is systematically missing in the latter (see Fig. 2 in chapter by Cheng and Blumenthal). This N-terminal part of Dnmt1, 3a and 3b is important for targeting specifically the DNA substrate.⁴²

Bioinformatics and biochemical analyses suggest that Dnmt2 evolved from a hypothetical very ancient RNA:(pyrimidine, C5) methyltransferase, which was also the ancestor for a majority, if not all, of the presently known DNA:(m⁵C-restriction) methyltransferases, RNA:(m⁵C) methyltransferases and even the RNA:(m⁵U) methyltransferases (43-45; see also chapter by Czerwoniec et al in this volume). It remains to demonstrate whether the low activity for DNA substrates by Dnmt2 enzymes detectable in vitro is physiologically relevant, or if it represents an extreme case in which the enzyme is forced under nonphysiological conditions to perform a reaction that it will not perform in vivo. Nevertheless it illustrates how easily a particular enzyme may change its specificity from one type of nucleic acid substrate to another.

Interestingly, in the absence of S-adenosyl-L-methionine in the in vitro reaction mixture, a DNA (cytosine, C5) methyltransferase can also become an efficient C-to-U deaminase (Fig. 4, box A). After a normal attack of the C6 atom of the cytidyl ring by a cysteine side chain as the active site of the enzyme, an anion is created allowing the C5 atom to be protonated. This is followed by the attack of a water molecule at C4 and subsequent hydrolytic deamination of the exocyclic amino group of the target cytosine to produce uracil.⁴⁶ Again, this observation demonstrates the ease with which a methylase may 'switch' its reaction type and mechanism and has interesting implication for enzyme evolution (ref. 47, see also chapter by Parisien and Bhagwat in this volume).

dC-to-dU and rC-to-rU Deaminases (Apobec-1)

In mammals, there exist a large family of dC-to-dU deaminases acting on DNA (see chapter by Smith). The first member of this family was originally identified as a rC-to-rU deaminase acting on Apolipoprotein B mRNA and, because it was found to be part of the Editing Catalytic subunit 1 of a multiprotein complex, it was designated Apobec-1. Among the 12 members of the subsequently identified Apobec superfamily of deaminases (including the enzyme Aid, for Activation-Induced Deaminase), Apobec-1 is the only one that has been shown to work on both RNA (mRNA) and on single-stranded DNA (Fig. 3, Box C). They all share a signature of amino acid sequence characteristics of a zinc-dependent enzyme and, among those that have been crystallized, they also share characteristic three-dimensional folds (see chapter by Wedekind and Beal). Phylogenetic modelling of this huge family of Apobec/Aid-type of enzymes acting on nucleic acid polymers suggests they all have evolved from a primordial cytidine deaminase active on free (deoxy)ribonucleotides (d)CMP and (d)CTP (the Cytidine deaminases—Cdd), a property that has been verified in vitro in the case of the catalytic subunit-1 of Apobec-1⁴⁸⁻⁵⁰ (Fig. 1, box D and Fig. 2, box C). The main differences between the present day enzymes working on DNA/RNA polymers and the phylogenetically related enzymes Cdd acting exclusively on free nucleotides is the acquisition during evolution of additional modules on the N-side of the catalytic core module (Fig. 4, box B) that allow the Apobec proteins to associate with several other proteins forming multiprotein complexes ('editosomes') and finally to bind to specific sites of DNA (also of RNA in the case of Apobec-1). Apobec-1 is therefore a true 'dual' enzyme able to work on both single-stranded DNA and RNA and this has been verified in vitro and in vivo.^{51,52} This contemporary enzyme is probably the best example of a relic of a primordial ancient deaminase acting probably initially on free nucleotides or nucleosides (as in the RNA world, then later on RNA and/or DNA).

rA-to-rI Deaminase (tbAdat 2,3)

The ADAT and ADAR family of enzymes catalyze the site-specific deamination of adenine-to-inosine exclusively in RNA: double-stranded RNA in the case of the eukaryotic ADAR family of enzymes and tRNA (anticodon loop) in the case of eukaryotic and bacterial ADAT (Fig. 3, box D—see chapters by Wedekind and Beal and by Haele and O'Connell). Phylogenetic analysis of this large family of A-deaminases indicates that they belong to the same superfamily of Apobec C-deaminases as mentioned above and therefore have evolved from the same primordial C-deaminases active on free (deoxy)ribonucleotides (d)CMP and (d)CTP^{50,53,54} (Fig. 4, box B).

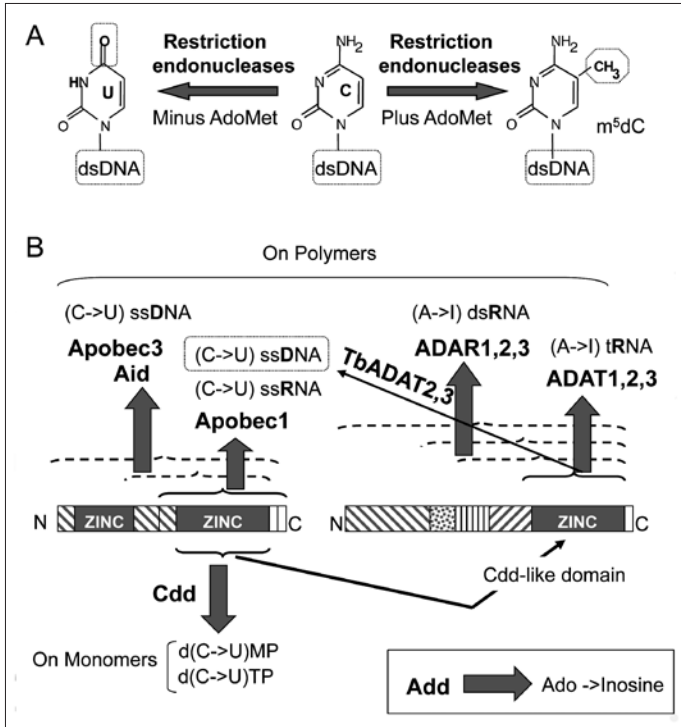


Figure 4. Detailed chemical transformation of bases within RNA/DNA polymers (continued). In box A is the situation that occurs when DNA is incubated in vitro with a purified restriction endonuclease (for examples EcoRII, HhaI, HpaII, SssI, MspI, PspGI, see refs. 46, 47) in the presence or absence of S-adenosyl-L-methionine. See also chapter by Parisien and Bhagwat in this volume. In box B, is a schematic view of relationships between the length of modification enzymes (including additional motifs besides the minimal core catalytic motif) and the type of monomers/polymers they modify (for more information see chapters by Wedekind and Beal, by Haele and O'Connell and by Smith). An identical situation is described for the Dnmt family of enzymes (see Fig. 2 in chapter by Cheng and Blumenthal and in chapter by Coffin et al in this volume). The motif indicated with zinc, is the catalytic deaminase motif that by itself can catalyze the deamination of the cytosine monomer (for details and references, see text). In the small box at the bottom right in panel B, is the reaction catalyzed by adenosine deaminase (Add, sometimes also referred as Ada), an enzyme that is evolutionarily unrelated to the catalytic domain of Cdd in C-deaminases or in A-deaminases.

Here it was demonstrated that one member of this rA-to-rI tRNA deaminases (I stands for inosine), TbADAT 2,3 of trypanosomatids, can also perform dC-to-dU deamination in single-stranded DNA in vitro, but not both reactions with either substrate.⁵⁵ The ability of a single rA-to-rI deaminase to work on tRNA to perform a different deamination reaction in a different type of substrate (single stranded DNA) is of course unexpected but nevertheless is reasonable if we take into account the 'history' of the enzyme, that is its phylogenetic origin from a simple ancestral (d) C-to-(d)U deaminase and not, as one would naively expect, from an ancient adenosine deaminase catalyzing the formation of inosine from adenosine (or hypoxanthine from adenine—Fig. 4, box B). Evolutionarily, the observed flexibility may be a reflection of the eukaryotic enzyme's need to acquire multi-substrate specificity. Indeed, while the bacterial ADAT only deaminates a single substrate in vivo, the eukaryotic counterpart targets 7-8 different substrates depending on the organism. This leads to the proposal that active site changes occur to evolution to facilitate recognition of

multiple substrates which in turn led to a repositioning of the RNA binding domain away from the active site, allowing the enzyme to catalyze more than one reaction.

As in the case of Dnmt2, it is not sure that TbADAT 2,3 displays activity on both tRNA (probably the major function) and single-stranded DNA *in vivo*, but the possibility for a modern rA-to-rI deaminases to catalyze a (d)C-to-(d)U deaminase reaction *in vitro*, suggests that this activity might have been an ancient function of the ancestor of TbADAT 2,3.

DNA/RNA Oxidative Demethylases/Dealkylases (AlkB)

Accidental non enzymatic methylation (or alkylation) of a base in a nucleic acid (DNA or RNA), in addition to the normal enzymatic methylation processes, can pose serious problems for modern cells, especially in the case of DNA for which abnormal methylation/alkylation can be mutagenic. To overcome this problem, living cells have developed several repair mechanisms devoted to eliminate such damages (reviewed in 56). Among them is a family of template-independent type of repair enzyme AlkB (Alk for alkylate) and FTO (Fat mass and Obesity-associated gene product) able to remove the adduct of the cytosine base (usually a simple methyl group, but also more bulky alkyl groups) without removing the base itself as do most other repair enzymes of the group of template-dependent DNA glycosylases (AlkA) or photolyases. At variance with AlkA enzymes, the AlkB enzymes belong to the 2-oxoglutarate- and iron-dependent oxygenase superfamily, producing hydroxymethyl intermediate products (for example hm³dC from m³dC), of which the methylenehydroxyl group is further eliminated in the form of formaldehyde (Fig. 3, Box E). AlkB proteins were initially discovered in bacteria (*E. coli*) however have homologues in genomes of eukaryotes and of plant-infecting RNA viruses, which leads to the hypothesis that these enzymes are also involved in RNA demethylation.⁵⁷ This was thereafter demonstrated in the case of a human homologue hABH3,⁵⁸ *E. coli* AlkB⁵⁹ and more recently with several plant viral AlkB⁶⁰ and one of the FTO-associated gene product.⁶¹ Interestingly, when tested *in vitro*, *E. coli* AlkB demethylates single-stranded DNA (from bacteriophage M13mp18) with better efficiency than RNA (from bacteriophage MS2), human ABH3 works equally well on DNA and RNA, while viral AlkB homologues clearly prefer RNA over DNA. In agreement with the biochemical data, it was demonstrated that expression of *E. coli* AlkB, hABH3 or plant viral AlkB in an *E. coli* alkB mutant reactivate single-stranded RNA bacteriophages pretreated with the methylating agent methylmethane-sulfonate (see chapter by Falnes et al). This is another clear example of a contemporary enzyme able to work on both types of nucleic acids, although with a preference for DNA or RNA depending on the type of genome (DNA as in *E. coli*) or RNA, as in plant, on which the enzyme acts. Again, this illustrates well the versatility of the enzymes involved in the metabolism of DNA and RNA.

The Virogenesis Hypothesis for the Origin of DNA

Considering the universal and important role that nucleotide modification actually plays in the competition between cells and viruses (see above), it is tempting to speculate that a few of the corresponding modifying enzymes arose in the RNA/Proteins world, hence possibly leading to the emergence of DNA itself. Indeed, one of us suggested that the first organism with an uracil-containing genome (U-DNA, Fig. 1, box B) was a virus that was selected because the modifications of its genome immunized it against the nucleases of the host.²⁰ This could have occurred for each of the two stepwise transitions in the formation of U-DNA and later of T-DNA. If this hypothesis is correct, the RNA to DNA transition occurred indeed in the viral factory of a DNA virus, i.e., inside an infected cell. The suggestion that the transition from RNA to DNA genomes occurred first in the virosphere and that DNA was only later on transferred to cells is in line with the observation that modern viruses exhibit a greater variety of genome types and structure than modern cells, implying a greater tolerance to genome modification. Moreover, the extent of base modification and thus, the capacity to explore diversity of combinatory solutions is greater when it occurs at the prereplication level, as explained above in the case of bacteriophages, compared to the relative scarcity of the modifications observed in DNA genomes when they occur at the postreplicative level.

Thus, except for the need of selected inhibitor of UNGase for destruction of dTMP (as in the case of bacteriophages PBS2, see above) that have evolved later, the deaminase specific for CMP or CTP and dCMP/dUMP hydroxy-methyltransferases as it exists in certain present day viruses (box D in Fig. 1 and box C in Fig. 2) might correspond to relics of primordial enzymes that might have been used more systematically in early stages of cell evolution, during and/or after the RNA/Proteins world but before the present day T-DNA. This type of strategy is still used today by mammalian cells to generate antibody diversity (reviewed in ref. 62 and chapter by Parisien and Bhagwat) and by viruses to escape self-defence by host cells.^{31,32} Although it will never be possible to know what exactly happened, exploration of RNA and DNA biogenesis and maturation (especially modifying enzymes) in our modern RNA world could help us to give some consistence to the virogenesis hypothesis for the origin of nucleotides or of base modifications.

It is worth remembering that reverse transcriptase, a typical viral enzyme, was an essential component of the early RNA-to-DNA transition. Many modern cells, especially in the domain Eukarya, still encode reverse transcriptases of viral origin and the RNA-to-DNA transition still occurs in modern organisms, a possible relic of the role that reverse transcriptase initially played in the origin of DNA.⁶³ Likewise, cellular ribonucleotide reductases (RNR), thymidylate synthases (Tds) and possibly some modification/repair enzymes also could have derived from ancestral viral genes that were transferred at different times of life history into various cellular lineages.⁶⁴ However, in this virogenesis hypothesis for the origin of DNA, the first appearance of RNR remains a problem because of the probable absence of a reverse transcriptase suitable to polymerize the products of their reactions. Before dNTP's were produced, what was the role of ancestral RNR? One possibility is that RNR originated first to inactivate ribozymes, that often rely on the 2'-hydroxyl of the ribose to perform their catalytic task.⁶⁵ This can again be rationalized in the framework of the conflict between viruses and host cells, each partner trying to inhibit ribozymes of their adversary. Another possibility could be that RNR was first used to deplete the NTP pool of the adversary. This is a real possibility if cell and/or viruses of that time were already able to confine their own NTP's into separate pools. The possibility for a virus to make use later on of dNTP's (originally by products of RNR anti-ribozyme activity) to modify its genome would have been a tremendous and direct selective advantage for this virus.

The First Modifications, Back to the RNA World, Beyond and After

We still do not know how the first ribonucleotides were synthesized.^{19,20} The canonical bases or derivatives can be produced by prebiotic chemistry and phosphate could have been produced in a volcanic setting. However, although recent experiments suggest tricks to select ribose over the competing sugars in the first steps of life evolution,^{66,67} the preferential selective abiotic synthesis of ribose has not yet been achieved. Furthermore, we have no idea how to put the base, the ribose and the activated-phosphoric acid together *in vitro* to get a bona fide ribonucleotides.²⁰ In fact, a form of early metabolism with a continuous energy source operating in "open vesicles" (an open thermodynamic system) was probably already in place when the first ribonucleotides started to accumulate and to be polymerized forming ever longer oligonucleotides. Such primitive autocatalytic cellular systems, that probably already contained catalytic peptides, peptidyl-nucleic acids and simple lipids (for reviews, see refs. 20 and 68), were probably able to produce a large number of derivatives/variants of the four canonical ribonucleotides. Later on, a few 'modified' RNA biomolecules were possibly selected¹⁹ (symbolized by 'Z?' in Fig. 1, box A), together with the canonical biomolecules, to built up the primordial functional ribozyme.

The first modifications might have been simple methylated bases or even pseudouridine, an isomer of uridine. Later on, ribozymes capable of catalyzing other types of modified nucleosides might have evolved, possibly in conjunction with selected polypeptides (acting as chaperones) and/or cofactors, such as S-adenosyl-methionine, thiamine pyrophosphate, flavin nucleotide, cobalamin (as in present-day riboswitches (reviewed in ref. 69)). Very early on, the ancestors of modern organisms probably started to preferentially incorporate modified nucleotides at selected positions, first in RNA by posttranscriptional processes and later in DNA by pre and postreplicative processes.

The very first early such ‘decorations’ of the nucleic acids should have corresponded to simple chemical modifications at selected positions of newly synthesized nucleic acids, which in turn might probably have had beneficial incidences on the structure and/or the function of the polymer itself and consequently become selected during evolution. Simple chemical reactions, such as methylation of an exocyclic nitrogen atom of adenine, cytosine or guanine and acceleration of naturally occurring hydrolytic deamination of a cytosine or an adenosine in RNA might have been easy to create with primitive catalysts. In the case of methylation, a sort of ‘riboswitch’ with a precursor of S-adenosyl-methionine (see for example ref. 70) and in the case of deaminase, a zinc ion trapped by a cluster of appropriate counterions allowing the water molecule to access and attack the amine group of the cytidine or adenine (see chapter by Wedekind and Beal) should not have been too difficult to perform.

It is worth mentioning that the minimum number of amino acids of a polypeptide that can fold like in a protein is around 10 amino acids (1-2 kDa).⁷¹ Also, based on the use of only one of the most common cofactors (besides ATP) i.e., S-adenosyl-L-methionine, a large array of different reactions can be generated, not only simple methylation of RNAs, but also reactions involving transfer of a ribosyl-group, an aminoalkyl group and of a 5'-deoxyadenosyl radicals on selected bases of RNA, as well as many other biomolecules (nicely reviewed in refs. 72-74), thus demonstrating the huge potentialities of the living cells to create ‘bio’-catalysts.

As a rule, the more cells became complex during evolution, the repertoire of enzymes acting on their nucleic acids diversified became more diversified. From systematic analyses of amino acid sequences, as well as of 3D-structures obtained from X-ray crystallography or NMR studies, we know more and more about enzyme architectures. It clearly appears that these biocatalysts are generally composed of small modules with distinct characteristic motifs that were initially encoded and assembled at the genome level (ref. 45 and chapter by Czerwoniec et al in this volume). Also, ‘analogous’ enzymes catalyzing the same reactions can emerge (independent invention) during evolution.^{75,76} Along millions of years of cell evolution such a vertical type of evolution progressively generated the vast majority of these modification enzymes as we know to date. Exchanging genetic materials by lateral gene transfer between organisms and viruses thriving within the same ecological niches (horizontal type of evolution or LGT) is another important way of diversifying the panoply of biocatalysts and modification enzymes in particular, obviously faster than by vertical heritage.⁷⁷⁻⁷⁹ This process was possibly much more frequent in the ‘very early days’ of cellular evolution, when primitive cells had not yet learnt how to defend themselves against illegitimate interspecies DNA/RNA recombination (see above).

Conclusion and Future Prospects

As made explicit in the Introduction and in many sections in this chapter, modern processes are historical products and enzyme versatility (the possibility to change of substrates) has been an important factor that makes this history possible. While RNA dominated ancestral cellular metabolic systems (in very different contexts and environments), the present-day biological systems are now dominated by both RNA and Proteins (at least for all the metabolic and energetic supplies of the cell). In the modern cellular world, the storage of genetic information has been mainly delegated to genomic DNA. However, interdependence between the two types of nucleic acids remains important as attested by the large number of metabolic processes at the DNA level that still remain under the control of RNA (for more information consult refs. 80, 81).

The majority of the present day biosynthetic and processing enzymes acting on RNA and on DNA appear to belong to two different ‘worlds’, exactly as the scientists working on DNA modification enzymes belong to distinctive castes having their own workshops and even their specialized Journals such as “Genome Research” and “*DNA*”, or “*RNA*” and “*RNA Biology*”. This is quite an unfortunate situation, indeed although most of the modification/editing enzymes studied so far are claimed to be specific for either DNA (usually double-stranded DNA but also single-stranded DNA for the repair enzymes) or RNA (single-stranded RNA, double-stranded RNA or even compact-3D architectural RNA or RNA already associated with proteins), it is not sure that the

appropriate experiments testing both types of substrates have been carefully performed (if tested). Not too many biochemists systematically test in parallel the two types of substrates and in the few cases when this was done, the results have been surprising (see above). It is therefore possible that more of the many present-day enzymes have true 'dual' specificity for both types of nucleic acids, or apparent 'dual' specificity enzymes. This might be the case for enzymes that have shifted specificity from one type of nucleic acid to another during evolution but still harbour to date the 'memory' of their ancient activity that might be revealed under special conditions *in vitro*. More surprise thus may come from more systematic detection of modifications in DNA as well as in noncoding RNAs and from an open-minded study of the enzymes involved in these processes within an evolutionary perspective.

Acknowledgement

Patrick Forterre thanks the 'Institut Universitaire de France' and the 'Association de la Recherche contre le Cancer (ARC)' for financial supports. Henri Grosjean (emeritus scientist) thanks Prof. Jean-Pierre Rousset from Université Paris-Sud for providing facilities to continue working in his laboratory. We thank Anne-Lise Haenni (IJM- Université Paris 7-France) Christian Rother (Inst Mol Cell Biol / Warsaw-Poland) and Nicolas Glansdorff, (Université de Bruxelles-Belgium) for useful comments on the manuscript.

References

- Schroedinger E. What is life?:With Mind and Matter and Autobiographical Sketches, Cambridge University Press, 1944;
- Monod J. Chance and Necessity: An Essay on the Natural Philosophy of Modern Biology. New York: Alfred A. Knopf, 1971;
- Gilbert W. The 'RNA world'. *Nature* 1986; 319, 618.
- Raoult D, Forterre P. Redefining viruses: lessons from mimivirus. *Nature Reviews Microbiology* 2008; 6:315-319.
- Crick FHC. On protein synthesis. *Symp Soc Exp Biol* 1958; 12:139-169.
- Morange M. What history tells us XIII. Fifty years of the Central Dogma. *J Biosci* 2008; 33:171-175.
- Forterre P. The two ages of the RNA world and the transition to the DNA world: a story of viruses and cells. *Biochimie* 2005; 87:793-803.
- Freeland SJ, Knight RD, Landweber LF. Do proteins predate DNA? *Science* 1999; 286:690-692.
- Forterre P, Filé J, Myllykallio H. Origin and evolution of DNA and DNA replication machineries. In: Ribas L, ed. "The Genetic Code and the Origin of Life" Landes Bioscience, 2004:145-168.
- Delarue M, Poch O, Tordo N et al. An attempt to unify the structure of polymerases. *Protein Eng* 1990; 3:461-467.
- Chang J, Nie X, Chang HE et al. Transcription of hepatitis delta virus RNA by RNA polymerase II. *J Virol* 2008; 82:1118-1127.
- Butzow JJ, Eichhorn GL. Different susceptibility of DNA and RNA to cleavage by metal ions. *Nature* 1975; 254(5498):358-359.
- Marguet E, Forterre P. DNA stability at temperatures typical for hyperthermophiles. *Nucl Acid Res* 1994; 22:1681-1686.
- Lazcano A, Llaca V, Cappello R et al. The origin and early evolution of nucleic acid polymerases. *Adv Space Res* 1992; 12:207-16.
- Boussau B, Blanquart S, Necsulea A et al. Parallel adaptations to high temperatures in the Archaea. *Nature* 2008; 456:942-945.
- Forterre P, Gribaldo S. The origin of modern terrestrial life. *HFSP* 2007; 1:156-168.
- Forterre P. Strategies of hyperthermophiles in adaptation of nucleic acids at high temperatures. In: Gerday C, Glansdorff N, eds. *Extremophiles in Encyclopedia of Life Support Systems (EOLSS)*, Oxford UK: EOLSS Publishers, 2003 (<http://www.eolss.net>).
- Grosjean H, Oshima T. How nucleic acids cope with high temperature. In: Gerday C, Glansdorff N, eds. *Physiology and Biochemistry of Extremophiles*, Washington, DC: ASM Press, 2007:39-56.
- Cermakian N, Cedergren R. Modified Nucleosides always were: an evolutionary model. In: Grosjean H, Benne R, eds. *Modification and Editing of RNA*, Washington DC: ASM Press, 1998:535-541.
- Forterre P. The origin of DNA genomes and DNA replication proteins. *Curr Opin Microbiol* 2002; 5:525-532.
- Huetzel B, Kanno T, Daxinger L et al. RNA-directed DNA methylation mediated by DRD1 and Pol IVb: a versatile pathway for transcriptional gene silencing in plants. *Biochem Biophys Acta* 2007; 1769:358-374.

22. Hartig JV, Tomari Y, Förstemann K. PiRNAs—The ancient hunters of genome invaders. *Genes and Develop* 2007; 21:1701-1713; see also Makarva KS, Grishin NV, Shabalina SA et al. A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi and hypothetical mechanism of action. *Biology Direct* 2006; 1(7):1-26.
23. Goll MG, Bestor TH. Eukaryotic cytosine methyltransferases. *Ann Rev Biochem* 2005; 74:481-514.
24. Sorek R, Kunin V, Hugenholtz P. CRISPR—a widespread system that provides acquired resistance against phages in bacteria and Archaea. *Nat Rev Microbiol* 2008; 6:181-186.
25. Warren RAJ. Modified bases in bacteriophage DNAs. *Ann Rev Microbiol* 1980; 34:137-158.
26. Gommers-Ampt JH, Borst P. Hypermodified bases in DNA. *FASEB J* 1995; 9:1034-1042.
27. Lesiewicz J, Dudock B. In vitro methylation of tobacco mosaic virus RNA with ribothymidine-forming tRNA methyltransferase. Characterization and specificity of the reaction. *Biochem Biophys Acta* 1978; 520:411-418.
28. Keith JM, Winters EM, Moss B. Purification and characterization of a HeLa cell transfer RNA(cytosine-5)-methyltransferase. *J Biol Chem* 1980; 255:4636-4644.
29. Becker HF, Motorin Y, Florentz C et al. Pseudouridine and ribothymidine formation in the tRNA-like domain of turnip yellow mosaic virus RNA. *Nucleic Acids Res* 1998; 26:3991-3997.
30. Brulé H, Grosjean H, Giegé R et al. A pseudoknotted tRNA variant is a substrate for tRNA (cytosine-5)-methyltransferase from *Xenopus laevis*. *Biochimie* 1998; 80:977-985.
31. Cattaneo R. Biased A-to-I hypermutation of animal RNA virus genomes. *Curr Opin Genet Dev* 1994; 4:895-990.
32. Taylor DR, Puig M, Darnell ME et al. New antiviral pathway that mediates hepatitis C virus replicon interferon sensibility through ADAR1. *J Virol* 2005; 79:6291-6298.
33. Nordlund P, Reichard. Ribonucleotide reductases. *Ann Rev Biochem* 2006; 75:681-706.
34. Reichard P. Ribonucleotide reductases: the evolution of allosteric regulation. *Arch Biochem Biophys* 2002; 397:149-155.
35. Riera J, Robb FT, Weiss R et al. Ribonucleotide reductase in the archaeon *Pyrococcus furiosus*: a critical enzyme in the evolution of DNA genomes. *Proc Natl Acad Sci USA* 1997; 94:475-478.
36. Liu L, Santi DV. Mutation of asparagine 229 to aspartate in thymidylate synthase converts the enzyme to deoxycytidylate methylase. *Biochem* 1992; 31:5100-5104.
37. Hardy LW, Nalivaika E. Asn 177 in *Escherichia coli* thymidylate synthase is a major determinant of pyrimidine specificity. *Proc Natl Acad Sci USA* 1992; 89:9725-9729.
38. Agarwalla S, LaPorte S, Liu L et al. A novel dCMP methylase by engineering Thymidylate Synthase. *Biochem* 1997; 36:15909-15917.
39. Goll MG, Bestor TH. Eukaryotic cytosine methyltransferases. *Ann Rev Biochem* 2005; 74:481-514.
40. Goll MG, Kirpekar F, Maggert KA et al. Methylation of tRNA-Asp by the DNA methyltransferase homolog Dnmt2. *Science* 2006; 311:395-398.
41. Dong A, Yoder JA, Zhang X et al. Structure of human DNMT2, an enigmatic DNA methyltransferase homolog that displays denaturant-resistant binding to DNA. *Nucleic Acids Res* 2001; 29:439-448.
42. Jeltsch A, Nellen W, Lyko F. Two substrates are better than one: dual specificities for Dnmt2 methyltransferases. *Trends Biochem Sci* 2006; 31:306-308.
43. Bujnicki JM, Feder M, Ayres CL et al. Sequence-structure-function studies of tRNA:m⁵C methyltransferase Trm4p and its relationship to DNA:m⁵C and RNA:m⁵U methyltransferases. *Nucleic Acids Res* 2004; 32:2453-2463.
44. Sunita S, Tkaczuk KL, Purta E et al. Crystal structure of the *Escherichia coli* 23S rRNA:m⁵C methyltransferase RlmI (YceW) reveals evolutionary links between RNA modification enzymes. *J Mol Biol* 2008; 383:652-666.
45. Anantharaman V, Koonin EV, Aravind L. Comparative genomics and evolution of proteins involved in RNA metabolism. *Nucl Acids Res* 2002; 30:1427-1464.
46. Sharath AN, Weinhold E, Bhagwat AS. Reviving a dead enzyme: cytosine deaminations promoted by an inactive DNA methyltransferase and an S-adenosylmethionine analogue. *Biochem* 2000; 39:14611-14616.
47. Carpenter M, Divvela P, Pingoud V et al. Sequence-dependent enhancement of hydrolytic deamination of cytosines in DNA by the restriction enzyme PspGI. *Nucl Acids Res* 2006; 34:3762-3770.
48. Navaratnam N, Bhattacharya S, Fujino T et al. Evolutionary origins of apoB mRNA editing: catalysis by a cytidine deaminase that has acquired a novel RNA-binding motif at its active site. *Cell* 1995; 81:187-195.
49. Sawyer SL, Emerman M, Malik HS. Ancient adaptive evolution of the primate antiviral DNA editing enzyme APOBEC3G. *PLoS Biol* 2004; 2(9):E275.
50. Conticello SG, Thomas CJ, Petersen-Mahrt SK et al. Evolution of the AID/APOBEC Family of Polynucleotide (Deoxy)cytidine deaminases. *Mol Biol Evol* 2005; 22:367-377.

51. Petersen-Mahrt SK, Neuberger MS. In vitro deamination of cytidine to uracil in single-stranded DNA by Apolipoprotein B editing complex catalytic subunit I APOBEC-1. *J Biol Chem* 2003; 278:19583-19586.
52. Harris RS, Petersen-Mahrt SK, Neuberger MS. RNA editing enzyme APOBEC-1 and some of its homologs can act as DNA mutators. *Mol Cell* 2002; 10:1247-1253.
53. Keegan LP, Gerber AP, Brindle J et al. The properties of a tRNA-specific adenosine deaminase from *Drosophila melanogaster* support an evolutionary link between pre-mRNA editing and tRNA modification. *Mol Cell Biol* 2000; 20:825-833.
54. Gerber AP, Keller W. RNA Editing by base deamination: more enzyme, more targets, new mysteries. *Trends Biochem Sci* 2001; 26:376-384.
55. Rubio MA, Pastar I, Gaston KW et al. An adenosine-to-inosine tRNA-editing enzyme that can perform C-to-U deamination of DNA. *Proc Natl Acad Sci USA* 2007; 104:7821-6.
56. Sedgwick B, Bates PA, Paik J et al. Repair of alkylated DNA: recent advance. *DNA Repair* 2007; 6:429-442.
57. Aravind L, Koonin EV. The DNA-repair protein AlkB, EGL-9 and leprecan define a new family of 2-oxoglutarate- and iron-dependent dioxygenases. *Genome Biol* 2001; 2:3 RESEARCH0007.
58. Aas PA, Otterlei M, Falnes PO et al. Human and bacterial oxidative demethylases repair alkylation damage in both RNA and DNA. *Nature* 2003; 421:859-863.
59. Ougland R, Zhang C-M, Liiv A et al. AlkB restores the biological function of mRNA and tRNA inactivated by chemical methylation. *Mol Cell* 2004; 16:107-116.
60. Van den Born E, Omelchenko MV, Bekkelund A et al. Viral AlkB proteins repair RNA damage by oxidative demethylation. *Nucleic Acid Res* 2008; 36:5451-5461.
61. Gerken T, Girard CA, Tung YC et al. The obesity-associated FTO gene encodes a 2-oxoglutarate-dependent nucleic acid demethylase. *Science* 2007; 318:1469-1472.
62. Di Noia JM, Neuberger MS. Molecular mechanisms of antibody somatic hypermutation. *Annu Rev Biochem* 2007; 76:1-22.
63. Brosius J. The contribution of RNAs and retroposition to evolutionary novelties. *Genetica* 2003; 118:99-116.
64. Filée J, Forterre P, Laurent J. The role played by viruses in the evolution of their hosts: a view based on informational protein phylogenies. *Res Microbiol* 2003; 154:237-243.
65. Leon PE. Inhibition of ribozymes by deoxyribonucleotides and the origin of DNA. *J Mol Evol* 1998; 47:122-126.
66. Sacerdote MG, Szostak JW. Semipermeable lipid bilayers exhibit diastereoselectivity favoring ribose *Proc Natl Acad Sci USA* 2005; 102:6004-6008.
67. Ricardo A, Carrigan MA, Olcott AN et al. Borate minerals stabilize ribose. *Science* 2004; 303:196.
68. Glandsdorff N, Xu Y, Labeledan B. The last universal common ancestor: emergence, constitution and genetic legacy of an elusive forerunner. *Biol Direct* 2008; 3:29.
69. Cochrane JC, Strobel SA. Riboswitch effectors as protein enzyme cofactors. *RNA* 2008; 14:993-1002.
70. Winkler WC, Nahvi A, Sudarsan N et al. An mRNA structure that controls gene expression by binding S-adenosylmethionine. *Nature Struct Biol* 2003; 10:701-707.
71. Fan K, Wang W. What is the minimum number of letters required to fold a protein. *J Mol Biol* 2003; 328:921-926.
72. Schubert HL, Blumenthal RM, Cheng X. Many paths to methyltransferase: a chronicle of convergence. *Trends Biochem Sci* 2003; 28:329-335.
73. Fontecave M, Atta M, Mulliez E. S-adenosylmethionine: nothing goes to waste. *Trends Biochem Sci* 2004; 29:243-249.
74. Wang SC, Frey PA. S-adenosylmethionine as an oxidant: the radical SAM superfamily. *Trends Biochem Sci* 2007; 32:101-110.
75. Suck D. Common fold, common function, common origin? *Nature Struct Biol* 1997; 4:161-165.
76. Galperin MY, Walker DR, Koonin EV. Analogous enzymes: independent inventions in enzyme evolution. *Genome Research* 1998; 8:779-790.
77. Lawrence JG. Gene transfer, speciation and the evolution of bacterial genomes. *Curr Opin Microbiol* 1999; 2:519-523.
78. Yanai I, Wolf YI, Koonin EV. Evolution of gene fusions: horizontal transfer versus independent events. *Genome Biology* 2002; 3: research0024.
79. Zhaxybayeva, Lapierre P, Gorgarten JP. Genome mosaicism and organismal lineages. *Trend Gene* 2004; 20:254-266.
80. Jeffares DC, Poole AM, Penny D. Relics from the RNA world. *J Mol Evol* 1998; 46:18-36.
81. Mathieu O, Bender J. RNA-directed DNA methylation/commentary. *J Cell Sci* 2004; 117:4881-4888.