BIOCHEMISTRY
including biophysical chemistry & molecular biology

Current Topic

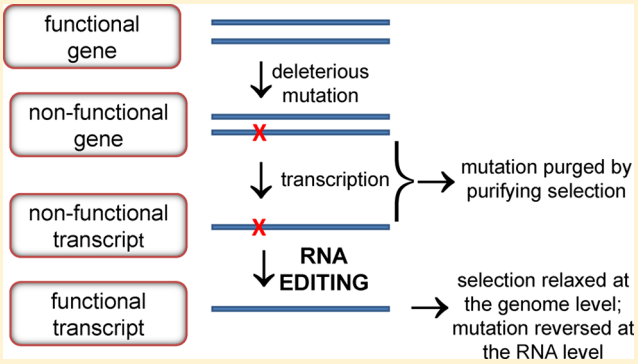
pubs.acs.org/biochemistry

Evolutionary Origin of RNA Editing

Michael W. Gray*

Centre for Comparative Genomics and Evolutionary Bioinformatics, Department of Biochemistry and Molecular Biology, Dalhousie University, Halifax, Nova Scotia B3M 4R2, Canada

ABSTRACT: The term "RNA editing" encompasses a wide variety of mechanistically and phylogenetically unrelated processes that change the nucleotide sequence of an RNA species relative to that of the encoding DNA. Two general classes of editing, substitution and insertion/deletion, have been described, with all major types of cellular RNA (messenger, ribosomal, and transfer) undergoing editing in different organisms. In cases where RNA editing is required for function (e.g., to generate a translatable open reading frame in a mRNA), editing is an obligatory step in the pathway of genetic information expression. How, when, and why individual RNA editing systems originated are intriguing biochemical and evolutionary questions. Here I review briefly what is known about the biochemistry, genetics, and phylogenetics of several very different RNA editing systems, emphasizing what we can deduce about their origin and evolution from the molecular machinery involved. An evolutionary model, centered on the concept of "constructive neutral evolution", is able to account in a general way for the origin of RNA editing systems. The model posits that the biochemical elements of an RNA editing system must be in place before there is an actual need for editing, and that RNA editing systems are inherently mutagenic because they allow potentially deleterious or lethal mutations to persist at the genome level, whereas they would otherwise be purged by purifying selection.



The term "RNA editing" originally described an unusual form of post-transcriptional processing involving the insertion of uridylate (U) residues into a mRNA in the mitochondria of two trypanosomatid protozoans (*Crithidia fasciculata* and *Trypanosoma brucei*).[1] This type of mitochondrial mRNA editing was later found to be widespread throughout Kinetoplastida (a class containing trypanosomatid protozoa, within the eukaryotic phylum Euglenozoa), to affect many mRNAs, and to involve deletion as well as insertion of U. Subsequent to its initial usage, RNA editing has become a generic term applied to a bewildering array of post-transcriptional processes (Table 1) whose effect is to change the nucleotide sequence of a mature RNA species relative to the encoding DNA sequence. RNA editing excludes 5′ capping, splicing, and 3′ polyadenylation of mRNA, as well as formation of modified nucleosides in RNA, although the distinction between "editing" and "modification" is somewhat blurred. For instance, one of the types of editing described below involves deamination of adenosine (A) residues to form inosine (I), resulting in the conversion of the 6-aminopurine ring of A to 6-oxopurine in I. When this change occurs within the coding region of a mRNA, the edited site (I) is recognized as G during translation. However, A residues in the first ("wobble") position of tRNA anticodons also undergo deamination (by an evolutionarily related enzyme[2]) to I, which similarly results in a change in the anticodon pairing properties. Thus, at a certain level, editing and modification might be regarded as two sides of the same (post-transcriptional processing) coin.

RNA editing systems fall into two general classes: insertion/deletion and substitution (Figure 1). In the first class, the nucleotide sequence of the mature RNA product is not collinear with that of its DNA coding sequence: the final RNA product contains extra nucleotides compared to its gene. In the second class, coding sequences of a mature RNA and its gene are collinear: they contain the same number of nucleotides but differ at those positions where editing occurs. It should be emphasized that this classification does not address mechanism but describes only the end result of the editing process: a nucleotide substitution might be the result of an excision repair (i.e., deletion/insertion)-type process, rather than direct base conversion (e.g., deamination). A fundamental question for both classes of editing and for all individual editing types is not only what the underlying biochemistry of the system is that effects the change but also what the specificity determinants are that direct the conversion so that it is position-specific.

The first few RNA editing systems to be described were ones that mediate changes in mRNA, but subsequent discoveries revealed that RNA editing also retailors both rRNA[3] and tRNA[4] molecules, as well as some viral RNAs (Table 1). Editing may act on a single residue whose alteration nevertheless has a profound effect on biological function[5] but may also involve such extensive changes at the RNA level that

Received: April 2, 2012
Revised: June 18, 2012
Published: June 18, 2012

   dx.doi.org/10.1021/bi300419r | *Biochemistry* 2012, 51, 5235−5242

Table 1. Mechanistically Distinct and Phylogenetically Disparate Types of RNA Editing in Eukaryotes[a]

type of editing	edited RNAs	organisms	genetic system
U insertion/deletion	mRNA	kinetoplastid protozoa	mitochondrial
N insertion (C, U, dinucleotides)	mRNA, rRNA, tRNA	*Physarum polycephalum*, other myxomycete slime molds	mitochondrial
N replacement	tRNA (5′-terminal nucleotides)	amoebozoans, chytridiomycete fungi	mitochondrial
	tRNA (3′-terminal nucleotides)	various animals, *Seculamonas ecuadoriensis* (jakobid)	mitochondrial
C-to-U conversion	apolipoprotein B mRNA	various mammals	nuclear
	mRNA, tRNA	land plants	mitochondrial,[b] chloroplast
	$tRNA^{Asp}$ (anticodon)	marsupials	nuclear
	$tRNA^{Trp}$ (anticodon)	kinetoplastid protozoa	nuclear/mitochondrial[c]
	cox1 mRNA	*P. polycephalum*	mitochondrial
	cox1, *cox3* mRNAs	*Naegleria gruberi*	mitochondrial
U-to-C conversion	mRNA	land plants	mitochondrial, chloroplast
	cox1 mRNA	placazoan (*Trichoplax adhaerens*)	mitochondrial
	WT1 mRNA[d]	mammals	nuclear
N substitution (multiple)	mRNA, rRNA	dinoflagellates	mitochondrial, chloroplast
A-to-I conversion	mRNA, miRNA, viral RNA	metazoan animals	nuclear

[a]Citations to the original litrerature may be found in refs 4, 13, 17, 18, and 50. [b]Editing also occurs to a minor extent in intron sequences and untranslated regions of mRNA. [c]The substrate is a nucleus-encoded tRNA that is imported into mitochondria where it undergoes editing.[7] [d]Encodes Wilms' tumor susceptibility factor.

the corresponding gene is not even recognizable as such at the genome level.[6] Different types of editing may exist within a single organism, or organelle (e.g., C-to-U editing of a tRNA as well as U insertion/deletion of mRNAs in *Leishmania*

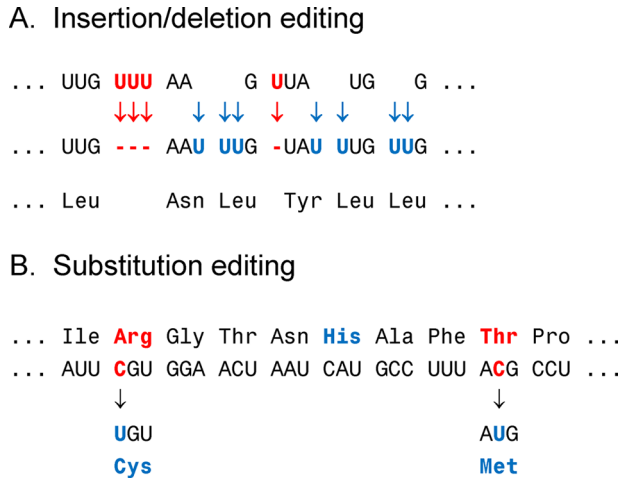


Figure 1. Examples of the two general classes of RNA editing, insertion/deletion and substitution. (A) U insertion/deletion editing of a portion of the transcript encoding the ATPase subunit 6 in the mitochondrion of a kinetoplastid protozoan, *T. brucei*. The top sequence includes U residues (red) that are encoded in the genome, present in the primary transcript, but deleted in the process of editing during maturation of the mRNA. The bottom sequence shows where nonencoded U residues (blue) are added during the editing process. The end result is a translatable open reading frame and the effective elimination of numerous frameshifts in the coding region of the primary transcript. (B) C-to-U editing of a portion of the transcript encoding subunit 2 of cytochrome *c* oxidase (Cox2p) in the mitochondrion of wheat *Triticum aestivum*. This region comprises part of a functionally critical region, the Cu_A binding site, with amino acids colored blue representing three of the five essential copper ligands within this region of Cox2p. At two codons, the *cox2* gene sequence predicts amino acids (red) that differ from the universally conserved consensus amino acids at these positions. However, C-to-U editing changes these codons in the mature mRNA so that they now encode the expected amino acid ligands.

mitochondria[7]), and different RNA editing systems may act upon a given transcript.[8,9]

RNA editing is found almost exclusively within the domain Eucarya (eukaryotes), although C-to-U editing of tRNAs has recently been described in Archaea.[10] Eukaryotic organelles (chloroplasts and particularly mitochondria) harbor the greatest variety of RNA editing systems (Table 1), and the magnitude of sequence changes can be truly remarkable. For example, in the lycophyte *Isoetes engelmannii*, a primitive land plant, 1782 positions of RNA editing (both C-to-U and U-to-C conversions) alter 1406 codon identities in all 25 mRNAs specified by the mitochondrial genome.[11] Why RNA editing systems should so frequently emerge in mitochondria is unclear. Perhaps, given that mitochondrial genomes encode relatively few genes (<100), relaxed constraints on mitochondrial gene expression favor the evolutionary processes that lead to the emergence of RNA editing.

Finally, a hallmark of RNA editing systems is their narrow phylogenetic distribution: they are limited to specific organismal groups or have a highly punctuate (scattered) localization. To date, for example, the trypanosomatid type of U insertion/deletion editing has not been definitively demonstrated outside Kinetoplastida, although there are recent indications that it may be operative on mitochondrial mRNAs in a diplonemid, *Diplonema papillatum*, a specific euglenozoan relative of kinetoplastid protozoa.[12] The lycophyte type of C-to-U/U-to-C substitution editing mentioned above is found throughout land plants (in both mitochondria and chloroplasts) but so far has not been detected in any of the charophyte algae—the specific green algal relatives of land plants—or in any other lineages of green algae. This localized distribution is consistent with the view that RNA editing systems are derived traits within the lineages in which they are found, rather than primitive traits retained from a remote common evolutionary ancestor.[13]

BIOCHEMISTRY, GENETICS, AND PHYLOGENETICS OF RNA EDITING

Although many varieties of RNA editing have been reported (Table 1), only a few systems have been studied in sufficient detail that the editing mechanism is understood and the editing

machinery well-defined. Four of these systems are briefly reviewed here from the perspective of the evolutionary insights they provide. A number of excellent books[14−16] and review articles[17,18] on the subject of RNA editing offer more comprehensive and detailed coverage than is possible here.

A-TO-I AND C-TO-U EDITING IN ANIMALS

A-to-I editing operates in animals on nuclear transcripts and is catalyzed by ADARs, "adenosine deaminases that act on RNA".[19−21] This type of editing enzyme was initially characterized as a double-stranded RNA (dsRNA) unwinding activity, now understood as an ADAR-induced conversion of A-U pairs to less stable I-U pairs. These early observations emphasize that ADARs are dsRNA-binding proteins and that their catalytic activity is directed toward duplex regions in RNA. Although the biologically significant function of ADARs is considered to be site-specific deamination in mRNA, duplex regions in noncoding RNAs—including microRNAs, small interfering RNAs, transcripts containing *Alu* repeats, and viral RNAs—are also ADAR substrates.[20]

The phylogenetic distribution of ADARs suggests an origin early in the metazoan radiation: ADARs are found in most multicellular animals but so far have not been detected outside Metazoa. An ancient gene duplication accounts for the presence of two ADAR paralogs, ADAR1 and ADAR2, found together in many animal species, although one or the other has evidently been lost in some lineages. ADARs share a common structure, consisting of a highly conserved, C-terminal deaminase (catalytic) domain and a variable number of double-stranded RNA-binding domains (dsRBDs) in the N-terminal half of the protein.[19]

The first reported instance of C-to-U editing was the site-specific deamination in mammalian intestine that converts a CAA (glutamine) codon to a UAA (stop) codon roughly midway through the mRNA encoding apolipoprotein B (apoB), a component of plasma low-density lipopoprotein (LDL) particles.[22] Generation of the UAA stop codon by this editing event results in premature termination of apoB mRNA translation in intestine, producing a truncated apoB48 rather than the full-length apoB100 produced in liver. Thus, editing effectively removes the C-terminal domain that in apoB100 binds to LDL receptors on cell membranes. Editing of apoB mRNA is mediated by a multicomponent "apoB mRNA editing complex" (APOBEC), comprising a catalytic subunit, APOBEC-1 (a cytidine deaminase having a novel RNA-binding domain[23]), and a required cofactor, "APOBEC-1 complementation factor" (ACF), a protein that functions as an adaptor protein by binding to both APOBEC-1 and the single-stranded RNA substrate.[22] Notably, ACF contains three RNA recognition motifs. Position specificity is provided by *cis*-acting sequence elements that flank the edited C residue.[24] Although APOBEC-type enzymes obviously have a very limited phylogenetic distribution, other "cytidine deaminases acting on RNA" (CDARs) may be more widely distributed within eukaryotes, given that C-to-U editing has also been reported outside animals, most notably in land plant mitochondria and chloroplasts (see below).

An important insight from comparative analysis of ADAR and CDAR sequences is that they belong to a superfamily of RNA-dependent deaminases that includes tRNA-specific deaminases (ADATs),[2] enzymes that are ubiquitous in eukaryotes. A common feature of ADARs, CDARs, and ADATs is the presence in the deaminase domain of conserved residues that are essential for catalysis.[25] This consideration together with phylogenetic analysis suggests that all three types of deaminases likely arose from an ancestral cytidine deaminase by acquisition of RNA-binding domains.[25] More particularly, it has been proposed that ADARs evolved from an ADAT progenitor through fusion of two or more dsRBDs to a common adenosine deaminase catalytic domain.[26] A recent sudy of the evolution of the deaminase fold has identified bacterial toxin deaminases and explored their relationship with the eukaryotic deaminases discussed above, prompting the intriguing suggestion that the latter might have been acquired by eukaryotes from bacterial symbionts on a number of occasions and recruited as organellar or nucleocytoplasmic modifiers of tRNAs and mRNAs.[27]

C-TO-U AND U-TO-C RNA EDITING IN LAND PLANT ORGANELLES

C-to-U editing of mRNAs is rampant in flowering plants (angiosperms), typically targeting some 300−500 sites in mitochondrial transcripts and 30−50 sites in chloroplast transcripts of the same species.[28−31] Editing takes place most frequently in mRNA coding regions, largely at first or second codon positions, and so almost always changes the identity of the amino acid specified by the edited codon relative to its gene-encoded counterpart. Transfer RNAs also undergo editing in plant organelles, as do (to a very limited extent) intron sequences in unspliced transcripts and noncoding regions of mRNA. "Reverse" U-to-C editing constitutes a significant proportion of total editing in some early diverging plant lineages, although it does not seem to occur in seed plants.[29] C-to-U editing in plant mitochondria occurs by deamination, so that one or more cytidine deaminases is presumed to provide the catalytic function of the editing process. However, the actual editing enzyme has remained elusive since the discovery of plant mitochondrial editing in 1989.[28,29] How U-to-C editing occurs is completely unknown and has yet to be studied.

Considering the myriad of editing sites that must be identified in plant organellar transcripts, the question of how site specificity is determined (i.e., how specific C residues are identified for conversion to U) has long been at the forefront of research in this area. Initially, it was assumed that the missing specificity factors might be small, antisense, guide-type RNAs that would target individual editing sites through complementary base pairing (analogous, e.g., to the snoRNAs that mediate modification of eukaryotic rRNA); however, searches for such molecules proved to be fruitless. Instead, and quite surprisingly, pentatricopeptide repeat (PPR) proteins, expressed from nuclear genes and targeted to mitochondria and chloroplasts, have been identified as site specificity factors for RNA editing in plant organelles. The PPR repeat is a degenerate, 35-amino acid RNA-binding motif, and PPR proteins typically contain a number of such motifs in tandem.[32] Genetic and biochemical studies have implicated PPR proteins in a wide range of RNA metabolic processes in plant organelles, including but not limited to editing.[33] A current model involves PPR proteins interacting directly with specific sites in the target transcript, in turn recruiting a generic enzyme responsible for RNA maturation, such as a C deaminase in the case of C-to-U editing. Very recently, this model has been expanded by the characterization of "multiple organellar RNA editing factors" (MORFs), which comprise a new family of proteins required for plant organellar RNA editing.[34] MORFs interact selectively with PPR proteins, suggesting that the editing complex

dx.doi.org/10.1021/bi300419r | *Biochemistry* 2012, 51, 5235−5242

("editosome") in plant organelles may be more complex than previously thought.

A striking feature of PPR proteins is their marked expansion in land plants, with the PPR gene family one of the largest in angiosperms, typically comprising 400−600 members compared to ∼10−20 in other eukaryotes, including green algae. Two classes of PPR proteins, P and PLS, are known, differing in the length and arrangement of tandem PPR repeats, with the PLS class having variant C-terminal domains that define E and DYW subclasses. Although PPR proteins comprising only classical PPR repeats (P class) are phylogenetically widely distributed, genes for the PLS class were until recently thought to be restricted to the land plant lineage (but see below).

The presence of DYW domain-containing PPR proteins appears to be strictly correlated with the presence of RNA editing in land plants, as does the number of such proteins and the number of organellar editing sites. These correlations highlight the PLS class of PPR proteins, and particularly the DYW subclass, as key specificity determinants for RNA editing in plant organelles. Thus, expansion of the PPR gene family in land plants is an important issue for our understanding of the emergence of RNA editing in plant organelles.

A number of studies suggest that amplification by retrotransposition is likely to have been the mechanism for this gene expansion early in plant evolution, with overall genome duplication events (known to have occurred multiple times in angiosperms) playing little or no role.[32] Fujii and Small[32] argue that any selective pressures acting to enhance RNA editing in early land plants would also have selected for increasing numbers of PPR genes, although the reverse argument is also possible: increasing numbers of PPR genes might have acted to enhance RNA editing[35] (see below). A drive from relatively AT-rich organellar genomes in green algae to relatively GC-rich organellar genomes in land plants provides a potential link with RNA editing, which would have the effect of reversing, at the RNA level, genomic T-to-C mutations. Several studies strongly correlate the number of RNA editing sites with genome GC content.[32]

There is little to suggest that RNA editing sites have been strongly conserved in either mitochondrial or chloroplast transcripts throughout the evolution of the land plant lineage. Thus, we cannot infer with any certainty what the ancestral RNA editing pattern might have been, either at the base of land plants per se or within any particular clade within this lineage. Data do, however, suggest a progressive loss of editing sites in chloroplasts, with a substantially larger number of such sites in early diverging plants such as lycophytes and bryophytes compared to their later diverging angiosperm relatives. Notably, the number of such editing sites is correlated with the mutational pattern, with the number of C-to-T transitions exceeding the number of T-to-C transitions in angiosperm chloroplast genomes, whereas the reverse is the case in early diverging land plants.[32]

Finally, DYW-PLS PPR genes have recently been identified in a heterolobosean protozoan, *Naegleria gruberi*, for which evidence for C-to-U mitochondrial mRNA editing was also reported.[36] One interpretation of this important finding is a possible horizontal gene transfer of DYW-PLS genes between a *Naegleria*-like protozoan and a primitive land plant progenitor.[37] The remarkable conservation of PPR and DYW domains in the plant and *Naegleria* versions, in both organization and sequence, would seem to argue against the possibility of independent emergence of DYW-PLS genes in the two lineages (convergent evolution). A third possibility—ancestral presence of DYW-PLS genes early in eukaryotic cell evolution, followed by multiple subsequent losses (of at least the DYW domain) in diverse eukaryotic lineages—would be consistent with the observation that P-type (i.e., non-PLS) PPR proteins are rather more widely distributed in various lineages of eukaryotic microbes (protists).[37] Whatever the explanation, this intriguing finding emphasizes the importance of broad phylogenetic sampling, especially within protists, in uncovering novel examples of RNA editing.

TRANSFER RNA 5′ EDITING IN MITOCHONDRIA

One of the first examples of tRNA editing emerged from studies of an amoeboid protozoan, *Acanthamoeba castellanii*, in which the mitochondrial DNA (mtDNA) encodes 15 tRNAs, 12 of which were predicted to have base pair mismatches at one or more of the first three positions of the acceptor stem[38] (Figure 2). Sequence analysis of the corresponding mature

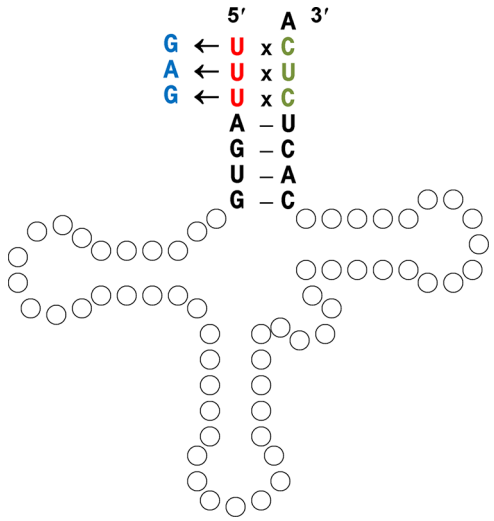


Figure 2. Mitochondrial tRNA 5′ editing in an amoeboid protozoan, *A. castellanii*. The sequence of the acceptor stem of the mtDNA-encoded $tRNA^{Asp}$ is shown, with standard base pairs indicated by a dash (−) and mismatches by ×. As encoded in the mitochondrial genome, the predicted cloverleaf structure of the tRNA is completely conventional except for the three indicated mismatches in the first three positions of the acceptor stem. Post-transcriptional editing occurs by removal of the first three nucleotides (red) on the 5′ side of the acceptor stem, and their replacement with nucleotides (blue) that form standard base pairs with their partner internal guide nucleotides (green) at the corresponding positions in the 3′ half of the acceptor stem.

tRNAs demonstrated that these mismatches were corrected either co- or post-transcriptionally, resulting in fully base paired tRNA acceptor stems. A working model proposed that this tRNA editing proceeded by removal of nucleotides from the 5′ side of the acceptor stem followed by their replacement with nucleotides that formed standard base pairs with their partner nucleotides on the 3′ side of the stem. This type of editing appears to be restricted to the first three acceptor stem positions; involves substitution with either purine or pyrimidine nucleotides, as specified by the residue on the 3′ side of the stem, which serves as an "internal guide" at each editing site; and evidently requires the combined activity of a nuclease (to remove mismatched 5′ nucleotides) and a nucleotide replacement component. Subsequently, the same type of

mitochondrial tRNA editing was reported in a chytridiomycete fungus, *Spizellomyces punctatus*.[39] Partially purified mitochondrial extracts from *A. castellanii*[40] and *S. punctatus*,[41] having virtually indistinguishable biochemical properties, were able to sequentially add the expected nucleotides to a 5′-truncated tRNA substrate, implicating a novel 3′-to-5′ nucleotidyltransferase (effectively a "reverse" RNA-dependent RNA polymerase) as a likely component of the tRNA editing activity.

A candidate mitochondrial tRNA 5′ editing enzyme was identified via characterization of the yeast (*Saccharomyces cerevisiae*) Thg1 gene,[42] which specifies tRNA histidine guanylyltransferase, an enzyme that adds an extra G residue specifically to the 5′ end of histidine tRNA in eukaryotes (i.e., Thg1 is a 3′-to-5′ nucleotidyltransferase). The availability of the Thg1 sequence quickly led to the identification of related but clearly distinct proteins (designated Thg1-like proteins, or TLPs[43]) in Archaea and Bacteria, which were shown to function as 3′-to-5′ polymerases in the repair of 5′-truncated tRNA substrates.[44] Genes encoding TLPs are also found in eukaryotes, in addition to or instead of Thg1, and distributed in a punctate fashion;[43] in particular, three TLP genes in addition to Thg1 have been identified in *Dictyostelium discoideum*, an amoebozoan relative of *A. castellanii* that also conducts mitochondrial tRNA 5′ editing. Biochemical characterization has demonstrated the ability of two of these TLPs, one of which has an N-terminal mitochondrial targeting peptide (MTS), to function in tRNA 5′ repair.[45]

The phylogenetic distribution of Thg1/TLP homologues is puzzling.[43,46] Authentic Thg1 orthologs are monophyletic and found only in the domain Eucarya (eukaryotes), whereas the domains Bacteria and Archaea contain only TLPs. Bacterial TLPs have a much more limited distribution than archaeal TLPs, and the former are clearly divided into two separate clades in phylogenetic analysis, with eukaryotic TLPs (sparsely distributed within eukaryotes) affiliated with one of the two bacterial groups. The pathway of evolutionary diversification of members of the Thg1 superfamily is difficult to discern, although it appears to have involved several horizontal gene transfers between the various groups, with the eukaryotic TLPs likely resulting from a bacterial-to-eukaryote horizontal gene transfer(s), rather than emerging directly out of the Thg1 clade per se.[43,46] With regard to mitochondrial tRNA 5′ editing, the salient point is that an ancient activity, likely functioning in generalized 5′ repair of tRNA acceptor stems, has been recruited to serve in mitochondrial tRNA 5′ repair in a number of eukaryotes.

WHY RNA EDITING?

The snapshots of several functionally and mechanistically different RNA editing systems given above illustrate how detailed biochemical, genetic, and phylogenetic investigations have helped answer questions about how (in a biochemical sense) and when these different systems emerged in evolution. ADARs likely evolved from an adenosine deaminase progenitor (ADAT?) through acquisition of dsRBDs that ultimately allowed them to work on a macromolecular substrate. Similarly, APOBEC evolved from a cytidine deaminase precursor through acquisition of a single-stranded RNA binding domain and interaction with a partner RNA-binding protein (ACP). In land plants, specificity factors for C-to-U editing in mitochondria and plastids were provided through a marked expansion of the DYW-PLS subclass of RNA-binding PPR proteins. Finally, a reverse (3′-to-5′) RNA-dependent RNA polymerase, of likely bacterial origin and capable of generalized repair of 5′-truncated tRNAs, has evidently been recruited to function in mitochondrial tRNA 5′ editing.

What these sorts of studies do not tell us is why any of these RNA editing systems should have emerged in the first place. Phylogenetic considerations strongly indicate that RNA editing systems are derived traits arising within lineages with conventional pathways of genetic information expression, in which case RNA editing would seem superfluous, seemingly adding unnecessary complexity to the process. RNA editing systems are often viewed as RNA repair systems, which is indeed how many of them now function, but arguments that they arose or were selected to correct deleterious mutations fail to take into account the fact that such mutations are likely to have been purged by purifying selection long before an RNA editing system could have been elaborated to correct them. We might more productively approach the "why" question by asking whether there is a general principle that could explain the evolutionary emergence of RNA editing as a biological phenomenon, as well as accommodate the origin of individual, mechanistically diverse RNA systems in different organisms.

A NEUTRAL EVOLUTIONARY VIEW OF THE ORIGIN OF RNA EDITING

One such model has been dubbed "constructive neutral evolution" (CNE).[35,47,48] This model, an extension of an earlier proposal for the origin of RNA editing,[49] describes how complexity might evolve in the absence of positive selection through a process of "presuppression", in which an autonomously functioning cellular component acquires mutations that make it dependent for function on another, preexisting component or process. Importantly, when there are multiple ways in which such dependency arises, CNE is seen to act as a unidirectional evolutionary ratchet, with increasing complexity an inevitable consequence of what is an inherently neutral evolutionary process.

Within a CNE context, the following points merit emphasis. (1) RNA editing systems emerge in a neutral fashion, before there is any need for editing. (2) They evolve or are assembled from extant proteins that already have a metabolic role in the cell. (3) Their capacity to reverse potentially deleterious mutations at the RNA level relaxes functional constraints at the gene level, allowing mutations to accumulate that would otherwise be purged. (4) RNA editing may be lost when the number of such mutations is sufficiently low that there is a finite probability that they can be simultaneously reversed at the genome level. (5) RNA editing persists and becomes an indispensable part of the genetic information expression pathway when mutations accumulate to such an extent that the probability of their simultaneous reversal becomes vanishingly small (the consequence of a unidirectional evolutionary ratchet).

As an example of how CNE might work, consider the case of mitochondrial tRNA 5′ editing.[50] In the absence of a mechanism for correcting mismatched acceptor stem base pairs, functional constraints on mitochondrial tRNA structure would effectively eliminate mutations that result in such mismatches. In the presence of such an activity, positions that can be edited at the RNA level are free to change according to the overall mutational pattern of the mtDNA, without regard to maintenance of tRNA secondary structure. For example, in an organism experiencing a drive toward a more AT-rich

mitochondrial genome (frequently seen in mtDNAs), editable positions would be expected to become more AT-rich.

In *A. castellanii* mtDNA, AT drive has resulted in a genome that is 71% A+T overall and 76% A+T in noncoding spacer regions. In the 12 tRNA species in which editing has been shown to occur, the A+U content of the 36 editable positions (the first three nucleotides at the 5′ end of each of the 12 tRNAs) is 72% before editing, but only 31% after editing. Considering only the 23 editable positions at which a nucleotide change actually occurs, the corresponding values are 87% prior to editing and 17% after editing. Most of the editing events that result in a nucleotide substitution (20 of 23) involve a change from A or U to another nucleotide. In fact, the base composition at these edited positions shifts dramatically: from 14 U residues but no G residues before editing to no U residues but 17 G residues after editing. If we allow for the fact that the numbers involved in these calculations are small, the observations are consistent with the expectation that the presence of a tRNA editing system allows editable sites in *A. castellanii* mitochondrial tRNA genes to become more A/T-rich than would otherwise be possible in the absence of editing.

In *A. castellanii*, the discovery of mitochondrial tRNA 5′ editing posed two puzzling questions. (1) Why is editing confined to the first three positions on the 5′ side of the acceptor stem? (2) Why do most (12 of 15) mtDNA-encoded tRNAs in this organism require editing? The answer would seem to be that the responsible enzymatic activity (a combination of nuclease and 3′-to-5′ polymerase) is inherently restricted to act on the first three 5′ positions of the acceptor stem, so these are the only sites at which mutations can accumulate. Also, as long as the editing system is a generalized one that is able to act on all (or at least most) of the mitochondrial tRNAs, relaxed functional constraints will ensure that mutations accumulate at these particular sites in the genes for all those tRNAs that are competent substrates for editing.

A CNE model argues that mutation fixation does not occur unless and until an appropriate tRNA editing system is already in place (i.e., before there is a need for editing). How might this have occurred? Although the putative nuclease component of the tRNA editing activity has yet to be characterized, the 3′-to-5′ RNA polymerase component has been identified as a TLP, as described above. Bacteria-to-Eucarya horizontal gene transfer appears to have been the most likely route for the acquisition of TLPs by eukaryotes. Once TLPs were established as nuclear genes, subsequent acquisition of a MTS is all that would be required to target a TLP to mitochondria for import. Independent acquisition of a MTS by TLP genes in different eukaryotes could then account for the punctate phylogenetic distribution of mitochondrial tRNA 5′ editing within the domain Eucarya. Although data are limited, it is notable that a correlation seems to be emerging between the presence of genes encoding mitochondrion-targeted TLPs and the presence of mitochondrial tRNA 5′ editing, either demonstrated or inferred from mtDNA sequences.[43]

It should be noted that a number of examples of 3′ editing of metazoan mitochondrial tRNAs have been reported, and in these instances, also, a ratchet-type mechanism has been invoked to explain the accumulation of potentially deleterious mutations in mitochondrial tRNA genes, with tRNA 3′ editing evolving "as a result of a need to escape" from the effects of such a ratchet.[51] In these cases of tRNA 3′ editing, the prior availability of a 5′-to-3′ nucleotidyltransferase/polymerase would presumably have been necessary to "preadapt" the system to conduct such editing, allowing the ratcheted accumulation of mutations that could be "corrected" at the RNA level.

With regard to the origin of plant organellar C-to-U editing, the pivotal evolutionary event is seen to be the expansion of the PLS class of PPR protein genes at the base of land plants. This multiplication of PPR proteins is presumed to have occurred before the emergence of RNA editing and in fact is proposed to have allowed subsequent, deleterious T-to-C mutations (occurring perhaps as a result of a drive toward more GC-rich genomes) to be fixed in plant organellar genes. Elsewhere,[35] colleagues and I have argued that the burgeoning complement of organellar-targeted, RNA-binding PPR proteins would essentially preadapt other proteins with which they were able to interact to function in a variety of roles having to do with RNA metabolism, including C-to-U RNA editing. If a PPR protein having a binding site in the vicinity of a potentially editable site were to recruit an activity (say, a C deaminase) capable of reversing the deleterious effect of the mutation at that particular position, that mutation could become fixed in the mitochondrial genome. Each individual edited site is potentially revertible, in which case the cognate PPR protein would no longer be essential. However, as the number of such sites increases, a ratchet-type mechanism ensures that there is virtually no possibility of a return to a state in which editing is no longer required, thereby effectively locking in the editing system as an integral part of the genetic information pathway.

Similar CNE models have been proposed to account for the origin of other RNA editing systems, including U insertion/deletion editing in kinetoplastid protozoa.[35,52] Alternative models invoke adaptive forces driven by positive selection in the origin of particular RNA editing systems.[53,54] Considering that evolutionary inferences are inherently speculative, it is not surprising there has been considerable debate about the relative merits of nonadaptive versus adaptive models in the evolution of RNA editing.[55−57] One of the common critiques of the CNE model is that because it invokes a unidirectional evolutionary ratchet, RNA editing cannot be lost once established. However, as colleagues and I have emphasized,[35,57] such a ratchet comes into play only when the number of editing events becomes large. Individual editing sites can and evidently have been lost in a number of systems, by mechanisms such as retro-transposition of edited mRNA. The secondary loss of the plant mitochondrial type of RNA editing does appear to have occurred at the base of the land plant lineage, within the marchantiid liverworts.[58,59] In this case, a CNE scenario would presume that the number of editing sites and/or number of transcripts requiring editing remained sufficiently low in a specific evolutionary ancestor of the marchantiid liverworts that reversion at the genome level to a state not requiring editing could still occur within this lineage, with consequent wholesale loss of the RNA editing machinery per se. Nevertheless, when the number of editing sites becomes very large, and/or when a diverse assemblage of transcripts must undergo editing, simultaneous reversion with elimination of the RNA editing machinery effectively does become impossible.

In conclusion, a CNE perspective on the evolutionary origin of RNA editing might be summed up as follows:[50] "...RNA editing systems, far from evolving in response to a need to 'correct' a problem, actually allow the problem to emerge in the first place, i.e., they permit DNA-encoded genetic information to degenerate progressively. Viewed in this light, RNA editing

systems are part and parcel of both the problem and its solution."

AUTHOR INFORMATION

Corresponding Author

*Telephone: (902) 494-2521. Fax: (902) 494-1355. E-mail: m.w.gray@dal.ca.

Notes

The author declares no competing financial interest.

ABBREVIATIONS

ACF, APOBEC-1 complementation factor; ADAR, adenosine deaminase acting on RNA; ADAT, adenosine deaminase acting on tRNA; apoB, apolipoprotein B; APOBEC, apoB editing complex; CDAR, cytidine deaminase acting on RNA; CNE, constructive neutral evolution; dsRBD, double-stranded RNA-binding domain; dsRNA, double-stranded RNA; MORF, multiple organellar RNA editing factor; mtDNA, mitochondrial DNA; MTS, mitochondrial targeting sequence; PPR, pentatricopeptide repeat; Thg1, histidine tRNA guanylyltransferase; TLP, Thg1-like protein.

REFERENCES

(1) Benne, R., Van den Burg, J., Brakenhoff, J. P., Sloof, P., Van Boom, J. H., and Tromp, M. C. (1986) Major transcript of the frameshifted coxII gene from trypanosome mitochondria contains four nucleotides that are not encoded in the DNA. *Cell 46*, 819−826.

(2) Gerber, A., Grosjean, H., Melcher, T., and Keller, W. (1998) Tad1p, a yeast tRNA-specific adenosine deaminase, is related to the mammalian pre-mRNA editing enzymes ADAR1 and ADAR2. *EMBO J. 17*, 4780−4789.

(3) Mahendran, R., Spottswood, M. R., and Miller, D. L. (1991) RNA editing by cytidine insertion in mitochondria of *Physarum polycephalum*. *Nature 349*, 434−438.

(4) Paris, Z., Fleming, I. M. C., and Alfonzo, J. D. (2011) Determinants of tRNA editing and modification: Avoiding conundrums, affecting function. *Semin. Cell Dev. Biol. 23*, 269−274.

(5) Sommer, B., Kohler, M., Sprengel, R., and Seeburg, P. H. (1991) RNA editing in brain controls a determinant of ion flow in glutamate-gated channels. *Cell 67*, 11−19.

(6) Feagin, J. E., Abraham, J. M., and Stuart, K. (1988) Extensive editing of the cytochrome c oxidase III transcript in *Trypanosoma brucei*. *Cell 53*, 413−422.

(7) Alfonzo, J. D., Blanc, V., Estévez, A. M., Rubio, M. A. T., and Simpson, L. (1999) C to U editing of the anticodon of imported mitochondrial $tRNA^{Trp}$ allows decoding of the UGA stop codon in *Leishmania tarentolae*. *EMBO J. 18*, 7056−7062.

(8) Gott, J. M., Visomirski, L. M., and Hunter, J. L. (1993) Substitutional and insertional RNA editing of the cytochrome *c* oxidase subunit 1 mRNA of *Physarum polycephalum*. *J. Biol. Chem. 268*, 25483−25486.

(9) Gott, J. M., Somerlot, B. H., and Gray, M. W. (2010) Two forms of RNA editing are required for tRNA maturation in *Physarum* mitochondria. *RNA 16*, 482−488.

(10) Randau, L., Stanley, B. J., Kohlway, A., Mechta, S., Xiong, Y., and Söll, D. (2009) A cytidine deaminase edits C to U in transfer RNAs in Archaea. *Science 324*, 657−659.

(11) Grewe, F., Herres, S., Viehöver, P., Polsakiewicz, M., Weisshaar, B., and Knoop, V. (2011) A unique transcriptome: 1782 positions of RNA editing alter 1406 codon identities in mitochondrial mRNAs of the lycophyte *Isoetes engelmannii*. *Nucleic Acids Res. 39*, 2890−2902.

(12) Kiethega, G. N., Turcotte, M., and Burger, G. (2011) Evolutionarily conserved *cox1 trans*-splicing without *cis*-motifs. *Mol. Biol. Evol. 28*, 2425−2428.

(13) Gray, M. W. (2003) Diversity and evolution of mitochondrial RNA editing systems. *IUBMB Life 55*, 227−233.

(14) Grosjean, H., and Benne, R., Eds. (1998) *Modification and Editing of RNA*, ASM Press, Washington, DC.

(15) Bass, B. L., Ed. (2001) *RNA Editing*, Vol. *34*, Frontiers in Molecular Biology (Hames, B. D., and Glover, D. M., Series Eds.) Oxford University Press, Oxford, U.K.

(16) Göringer, H. U., Ed. (2008) Nucleic Acids and Molecular Biology. In *RNA Editing*, Vol. *20* (Bujnicki, J. M., Series Ed.); Springer, Berlin.

(17) Gott, J. M., and Emeson, R. B. (2000) Functions and mechanisms of RNA editing. *Annu. Rev. Genet. 34*, 499−531.

(18) Knoop, V. (2011) When you can't trust the DNA: RNA editing changes transcript sequences. *Cell. Mol. Life Sci. 68*, 567−586.

(19) Hogg, M., Paro, S., Keegan, L. P., and O'Connell, M. A. (2011) RNA editing by mammalian ADARs. *Adv. Genet. 73*, 87−120.

(20) Wulff, B.-E., and Nishikura, K. (2010) Substitutional A-to-I RNA editing. *Wiley Interdiscip. Rev.: RNA 1*, 90−101.

(21) Jin, Y., Zhang, W., and Li, Q. (2009) Origins and evolution of ADAR-mediated RNA editing. *IUBMB Life 61*, 572−578.

(22) Blanc, V., and Davidson, N. O. (2010) APOBEC-1-mediated RNA editing. *Wiley Interdiscip. Rev.: Syst. Biol. Med. 2*, 594−602.

(23) Navaratnam, N., Bhattacharya, S., Fujino, T., Patel, D., Jarmuz, A. L., and Scott, J. (1995) Evolutionary origins of *apoB* mRNA editing: Catalysis by a cytidine deaminase that has acquired a novel RNA-binding motif at its active site. *Cell 81*, 187−195.

(24) Blanc, V., and Davidson, N. O. (2003) C-to-U RNA editing: Mechanisms leading to genetic diversity. *J. Biol. Chem. 278*, 1395−1398.

(25) Gerber, A. P., and Keller, W. (2001) RNA editing by base deamination: More enzymes, more targets, new mysteries. *Trends Biochem. Sci. 26*, 376−384.

(26) Keegan, L., Leroy, A., Sproul, D., and O'Connell, M. (2004) Adenosine deaminases acting on RNA (ADARs): RNA-editing enzymes. *Genome Biol. 5*, 209.

(27) Iyer, L. M., Zhang, D., Rogozin, I. B., and Aravind, L. (2011) Evolution of the deaminase fold and multiple origins of eukaryotic editing and mutagenic nucleic acid deaminases from bacterial toxin systems. *Nucleic Acids Res. 39*, 9473−9497.

(28) Gray, M. W. (2009) RNA editing in plant mitochondria: 20 years later. *IUBMB Life 61*, 1101−1104.

(29) Chateigner-Boutin, A.-L., and Small, I. (2010) Plant RNA editing. *RNA Biol. 7*, 213−219.

(30) Chateigner-Boutin, A.-L., and Small, I. (2011) Organellar RNA editing. *Wiley Interdiscip. Rev.: RNA 2*, 493−506.

(31) Castandet, B., and Araya, A. (2011) RNA editing in plant organelles. Why make it easy? *Biochemistry (Moscow, Russ. Fed.) 76*, 924−931.

(32) Fujii, S., and Small, I. (2011) The evolution of RNA editing and pentatricopeptide repeat genes. *New Phytol. 191*, 37−47.

(33) Schmitz-Linneweber, C., and Small, I. (2008) Pentatricopeptide repeat proteins: A socket set for organelle gene expression. *Trends Plant Sci. 13*, 663−670.

(34) Takenaka, M., Zehrmann, A., Verbitskiy, D., Kugelmann, M., Härtel, B., and Brennicke, A. (2012) Multiple organellar RNA editing factor (MORF) family proteins are required for RNA editing in mitochondria and plastids of plants. *Proc. Natl. Acad. Sci. U.S.A. 109*, 5104−5109.

(35) Lukeš, J., Archibald, J. M., Keeling, P. J., Doolittle, W. F., and Gray, M. W. (2011) How a neutral evolutionary ratchet can build cellular complexity. *IUBMB Life 63*, 528−537.

(36) Rüdinger, M., Fritz-Laylin, L., Polsakiewicz, M., and Knoop, V. (2011) Plant-type mitochondrial RNA editing in the protist *Naegleria gruberi*. *RNA 17*, 2058−2062.

(37) Knoop, V., and Rüdinger, M. (2010) DYW-type PPR proteins in a heterolobosean protist: Plant RNA editing factors involved in an ancient horizontal gene transfer? *FEBS Lett. 584*, 4287−4291.

(38) Lonergan, K. M., and Gray, M. W. (1993) Editing of transfer RNAs in *Acanthamoeba castellanii* mitochondria. *Science 259*, 812−816.

(39) Laforest, M. J., Roewer, I., and Lang, B. F. (1997) Mitochondrial tRNAs in the lower fungus *Spizellomyces punctatus*: tRNA editing and

UAG 'stop' codons recognized as leucine. *Nucleic Acids Res. 25*, 626−632.
(40) Price, D. H., and Gray, M. W. (1999) A novel nucleotide incorporation activity implicated in the editing of mitochondrial transfer RNAs in *Acanthamoeba castellanii*. *RNA 5*, 302−317.
(41) Bullerwell, C. E., and Gray, M. W. (2005) *In vitro* characterization of a tRNA editing activity in the mitochondria of *Spizellomyces punctatus*, a chytridiomycete fungus. *J. Biol. Chem. 280*, 2463−2470.
(42) Gu, W., Jackman, J. E., Lohan, A. J., Gray, M. W., and Phizicky, E. M. (2003) $tRNA^{His}$ maturation: An essential yeast protein catalyzes addition of a guanine nucleotide to the 5′ end of $tRNA^{His}$. *Genes Dev. 17*, 2889−2901.
(43) Jackman, J. E., Gott, J. M., and Gray, M. W. (2012) Doing it in reverse: 3′-to-5′ polymerization by the Thg1 superfamily. *RNA 18*, 886−899.
(44) Rao, B. S., Maris, E. L., and Jackman, J. E. (2011) tRNA 5′-end repair activities of $tRNA^{His}$ guanylyltransferase (Thg1)-like proteins from Bacteria and Archaea. *Nucleic Acids Res. 39*, 1833−1842.
(45) Abad, M. G., Long, Y., Willcox, A., Gott, J. M., Gray, M. W., and Jackman, J. E. (2011) A role for $tRNA^{His}$ guanylyltransferase (Thg1)-like proteins from *Dictyostelium discoideum* in mitochondrial 5′-tRNA editing. *RNA 17*, 613−623.
(46) Heinemann, I. U., Randau, L., Tomko, J. R. J., and Söll, D. (2010) 3′-5′ $tRNA^{His}$ guanylyltransferase in bacteria. *FEBS Lett. 584*, 3567−3572.
(47) Stoltzfus, A. (1999) On the possibility of constructive neutral evolution. *J. Mol. Evol. 49*, 169−181.
(48) Gray, M. W., Lukeš, J., Archibald, J. M., Keeling, P. J., and Doolittle, W. F. (2010) Irremediable complexity? *Science 330*, 920−921.
(49) Covello, P. S., and Gray, M. W. (1993) On the evolution of RNA editing. *Trends Genet. 9*, 265−268.
(50) Gray, M. W. (2001) Speculations on the origin and evolution of editing. In *RNA Editing* (Bass, B. L., Ed.), pp 160−184. Oxford University Press, Oxford, U.K.
(51) Börner, G. V., Yokobori, S.-i., Mörl, M., Dörner, M., and Pääbo, S. (1997) RNA editing in metazoan mitochondria: Staying fit without sex. *FEBS Lett. 409*, 320−324.
(52) Lukeš, J., Leander, B. S., and Keeling, P. J. (2009) Cascades of convergent evolution: The corresponding evolutionary histories of euglenozoans and dinoflagellates. *Proc. Natl. Acad. Sci. U.S.A. 106*, 9963−9970.
(53) Speijer, D. (2006) Is kinetoplastid pan-editing the result of an evolutionary balancing act? *IUBMB Life 58*, 91−96.
(54) Castandet, B., and Araya, A. (2012) The nucleocytoplasmic conflict, a driving force for the emergence of plant organellar RNA editing. *IUBMB Life 64*, 120−125.
(55) Speijer, D. (2010) Constructive neutral evolution cannot explain current kinetoplastid panediting patterns. *Proc. Natl. Acad. Sci. U.S.A. 107*, E25.
(56) Speijer, D. (2011) Does constructive neutral evolution play an important role in the origin of cellular complexity? *BioEssays 33*, 344−349.
(57) Doolittle, W. F., Lukeš, J., Archibald, J. M., Keeling, P. J., and Gray, M. W. (2011) Comment on "Does constructive neutral evolution play an important role in the origin of cellular complexity?". *BioEssays 33*, 427−429.
(58) Groth-Malonek, M., Wahrmund, U., Polsakiewicz, M., and Knoop, V. (2007) Evolution of a pseudogene: Exclusive survival of a functional mitochondrial *nad7* gene supports *Haplomitrium* as the earliest liverwort lineage and proposes a secondary loss of RNA editing in Marchantiidae. *Mol. Biol. Evol. 24*, 1068−1074.
(59) Rüdinger, M., Volkmar, U., Lenz, H., Groth-Malonek, M., and Knoop, V. (2012) Nuclear DYW-type PPR gene families diversify with increasing RNA editing frequencies in liverwort and moss mitochondria. *J. Mol. Evol. 74*, 37−51.