Editing and Modification of Messenger RNA

J. Scott[1]

1 Introduction

The discovery of messenger RNA (mRNA) more than 30 years ago led to the proposition of the central dogma of molecular biology that DNA makes RNA, and that RNA makes protein. Research has since shown that mRNA takes part in various complex reactions and is subject to a variety of co- and post-transcriptional modifications. RNA can be faithfully copied by reverse transcriptase, edited by the splicing out of intervening sequences, modified by capping and polyadenylation, and can itself catalyze transesterification reactions. None of these discoveries seriously challenges the central dogma of molecular biology. It came therefore as a considerable surprise to find that genetic information not found in the genomic template can be transferred into the mRNA after transcription. Thus, certain trypanosome mitochondrial mRNAs, which are unable to be translated for lack of AUG initiation codons or because of the presence of frameshifts in the coding sequence, are rendered translatable by the introduction or deletion of U residues that are not encoded in the mitochondrial or nuclear genome (Benne et al. 1986; Feagin et al. 1987, 1988; Abraham et al. 1988). Other modifications that alter the coding ability of mRNAs have since been discovered. These include the modification of apolipoprotein (apo)-B mRNA (Chen et al. 1987; Powell et al. 1987) and of plant mitochondrial mRNAs (Covello and Gray 1989; Gualberto et al. 1989; Hiesel et al., in press), in which C to U substitution occurs, and the addition of a nontemplated G residues, which shift the reading frame of paramyxovirus mRNAs (Thomas et al. 1988; Cattaneo et al. 1989; Paterson et al. 1989).

In this review RNA editing is defined as reactions in which nontemplated nucleotides are generated in existing mRNA transcripts in such a way as to alter the protein coding sequence of the mRNA. In addition, certain recently discovered novel forms of RNA modification are discussed (Bass and Weintraub 1987; Rebagliati and Melton 1987; Narayan and Rottman 1988; Wagner and Nishikura 1989; Wagner et al. 1989).

[1] Division of Molecular Medicine, MRC Clinical Research Centre, Watford Road, Harrow, Middx HA1 3UJ, UK

Nucleic Acids and Molecular Biology, Vol. 4
ed. by F. Eckstein and D.M.J. Lilley

2 Editing and Modification of RNA Expressed in the Cytoplasm

2.1 Apolipoprotein-B RNA

The intestinal form of apolipoprotein (apo)-B is required for the absorption of dietary lipid. It is designated apo-B48 because it is 48% of the size of the hepatic apo-B (apo-B100) on SDS-polyacrylamide gels. Analysis of intestinal cDNA clones, direct protein sequencing, and mapping of epitopes for monoclonal and antipeptide antibodies have indicated that apo-B48 is colinear with the amino terminal half of apo-B100 and terminates at the isoleucine residue 2152 of apo-B100 (Chen et al. 1987; Powell et al. 1987; Hospattankar et al. 1987). Therefore apo-B48 lacks the carboxy terminal domain found in apo-B100, which mediates the clearance of apo-B-containing particles from the circulation by the LDL receptor pathway. This process facilitates the rapid delivery of dietary lipids to the liver (Brown and Goldstein 1987).

In apo-B48 cDNAs derived from human, rabbit, rat and mouse a TAA stop translation codon is found at position 2153 (nucleotides 6666 to 6668 in the human mRNA), whereas in cDNAs coding for apo-B100 this codon is CAA for glutamine. In the apo-B gene codon 2153 also specifies glutamine (CAA) (Fig. 1). Codon 2153 is in the middle of the huge 7572 bp exon 26 of the apo-B gene. To explain this discrepancy, conventional mechanisms, such as the presence of two apo-B genes, one coding for apo-B100 and the other for apo-B48, alternative splicing of apo-B100 and apo-B48 exons, and unconventional mechanisms such as rearrangement of the apo-B gene in the cells that make apo-B48, have all been excluded. It has therefore been concluded that the C at position 6666 in the apo-B100 mRNA is modified in the apo-B48 mRNA to form a U or a U-like base (Fig. 1). The conversion of C to U in apo-B RNA has been confirmed by creating transgenic mice expressing segments of apo-B RNA spanning the RNA modification site (Powell, Driscoll, Davies, Wallis, Mullins and Scott, unpublished results) by transfection studies and by conversion in vitro using cytoplasmic S100

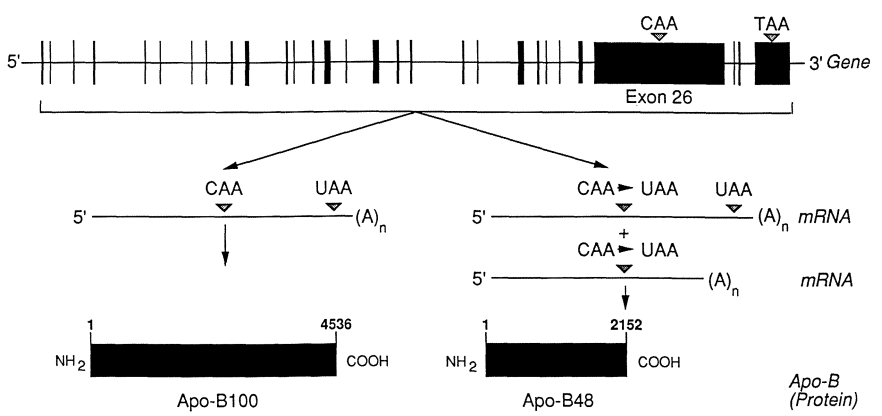

Fig. 1. The intron–exon organization of the apo-B gene and origin of the mRNAs encoding apo-B100 and apo-B48

and nuclear extracts from cells that convert apo-B mRNA (Chen et al. 1989; Davies et al. 1989; Driscoll et al. 1989).

The sequence requirements for apo-B mRNA editing have been localized within 55 nucleotides spanning the RNA modification site (Davies et al. 1989; Driscoll et al. 1989). This sequence forms a stem-loop, conserved in four mammalian species, with the edited base always occurring at the same position in an AU-rich 8-base loop. Disruption of the stem-loop abolishes editing in vitro, but not in vivo. Therefore the precise primary and higher-order structural requirements for recognition of the modification site have still to be established. AU-rich sequences are generally lost from mRNAs expressed in the cytoplasm because they are sensitive to AU-specific nucleases (Beutler et al. 1989). It is likely therefore that the AU-rich sequence in apo-B is protected from nuclease attack for functional reasons. This view is supported by the finding that cryptic polyadenation signals downstream of nucleotide 6666 are exposed by the editing process with the production of 7 kb apo-B truncated mRNAs representing the 5′ half of the full length 15 kb mRNA (Powell et al. 1987); this finding has been reproduced experimentally by transfection studies with constructs spanning the editing site (Bostrom et al. 1989).

RNA modification in vitro is protease-sensitive, but resistant to micrococcal nuclease and therefore does not require exposed nucleic acid for recognition of the editing site (Wynne, Driscoll and Scott, unpublished results; Driscoll et al. 1989). Additional ATP and divalent cations are not necessary for modification, and treatment of the S100 extracts with ATPase does not destroy conversion (Wynne, Driscoll and Scott, unpublished results). These results suggest that editing of apo-B mRNA may be similar to the enzyme-mediated modifications that occur to tRNA.

A variety of enzymes can convert C to U. Cytidine deaminase is an enzyme involved in pyrimidine metabolism that is responsible for the deamination of cytidine to form uridine (Wentworth and Wolfenden 1978). It is widely distributed in mammalian tissues. The enzyme is one candidate for the activity which modifies apo-B mRNA. However, this enzyme acts on free cytidine and is not known to act on cytidine in phosphodiester linkage in RNA. The C to U conversion could also be achieved by transglycosylation, in which C is replaced by U, without cleavage of the phosphodiester linkage in the polynucleotide chain. Transglycosylation is responsible for the substitution of G with the hypermodified nucleotide queuosine (Q) and A with inosine (I), the nucleotide of the base hypoxanthine, at wobble base position 34 of the anticodon loop of certain tRNAs (Okada and Nishimura 1979; Okada et al. 1979; Elliot and Trewyn 1984). Alternatively, base replacement might occur by the more general activities that catalyze exonucleocytic excision, gap-filling and ligation that occur during DNA repair (Grossman 1981). It has also been suggested that apo-B RNA could be modified by a mechanism similar to the unwinding and modifying activity that substitutes A for I in double-stranded RNA (dsRNA) (see below; Bass and Weintraub 1988; Simpson and Shaw 1989). This is unlikely, because S100 and nuclear extracts from tissues that modify apo-B mRNA are not active against dsRNA (Wynne and Scott, unpublished results). A further possible mechanism for the modification of apo-B mRNA is the formation of a hypermodified base

such as lysidine (L), a C residue in which the 0 in the 2 position is substituted with the ε-NH_2 of L-lysine (Muramatsu et al. 1988a, b). L is found in the wobble position in a minor species of isoleucine tRNA and is read as a U. A hypermodified base like L, if it occurred in apo-B mRNA, would have to be read as a U by both tRNA on the ribosome and by reverse transcriptase.

The apo-B RNA editing activity is under developmental, tissue-specific and metabolic regulation. In adult rodents both apo-B100 and apo-B48 mRNA are produced in the liver, and in the liver thyroxine produces a striking modulation of the editing reaction (Davidson et al. 1988). Adult mice also express and edit apo-B mRNA in a wide variety of other tissues, including the kidney, brain, testes and spleen (Driscoll, Davies, Wynne and Scott, unpublished results). Both forms of apo-B mRNA are also widely expressed in fetal tissues (Teng and Davidson, pers. comm.). It has yet to be established whether the editing activity is present in many cell types or only in differentiated cells such as fibroblasts or macrophages which are present in most tissues. The finding of apo-B in tissues like the brain and testis also suggests that the editing activity may be expressed more widely to regulate the expression of genes other than apo-B. Indeed, preliminary evidence that other mRNAs potentially undergo editing by the same mechanism as that for apo-B mRNA has recently been presented (Chen et al. 1989).

2.2 N^6methyladenosine in mRNA

Apart from 5′ capping nucleotides mRNAs contain a number of other modified nucleotides, particularly in their 5′ and 3′ untranslated regions (Adams and Cory 1975; Desrosiers et al. 1975; Narayan and Rottman 1988). Most prominent amongst these is N^6methyladenosine (m^6A). m^6A is a post-transcriptional modification of adenosine that occurs in most eukaryotic cytoplasmic and viral mRNAs with 1–3 m^6A residues occurring per 1000 nucleotides. m^6A is generally observed within an AAC or GAC consensus sequence and is nonrandom in distribution. In Rous sarcoma virus RNA and bovine prolactin mRNA most of the m^6A residues are located in the 3′ untranslated region where only 3 of 27 consensus sequences occur in the RNA. This suggests that the site of methylation is specific and directed by more than just the 3-nucleotide consensus sequence. Indeed, PuGACU sequences are methylated more frequently than simple AAC sequences. Further details concerning the possible consensus sequence and secondary structure requirements for the formulation of m^6A have yet to be determined, as has the function of this nucleotide. However, m^6A occurs just downstream of the poly-A addition signal and could be involved in polyadenylation. A system in vitro for methylation of adenosine residues has been established using Hela cell nuclear extracts (Narayan and Rottman 1988). Methylation is believed to occur before splicing in vivo, but mature mRNA is accurately methylated using the extract in vitro. The formation of m^6A is likely to be catalyzed by an RNA methyltransferase activity, but other mechanisms such as base exchange cannot be excluded. m^6A is also found in eubacterial tRNAs (Björk et al. 1987).

2.3 RNA Unwinding and Modification

A dsRNA unwinding and modifying activity has been detected in *Xenopus laevis, Caenorhabditis elegans* and a variety of mammalian tissue culture cells (Bass and Weintraub 1987, 1988; Rebagliati and Melton 1987; Kimelman and Kirschner 1989; Wagner and Nishikura 1989; Wagner et al. 1989). It is developmentally regulated and the activity is high in cells stimulated to regrow with serum. The activity is specific for intermolecular RNA/RNA interactions, but will unwind many different duplexes of different sequences including "killer" dsRNA of yeast, sense and anti-sense transcripts of chloramphenicol acetyl transferase, actin, neomycin, β-globin, cMyc and basic fibroblast growth factor (FGF). It is not active against single-stranded RNA, single-stranded DNA, double-stranded DNA or in dsRNA in intramolecular interactions. Many but not all of the A nucleotides were modified to I during the unwinding reaction, so as to permanently affect the I–U base pairing properties of the RNA, which is unable to rehybridize in its original duplex form. The I–U mismatch exposes the U base to cleavage by RNase A, which attacks mismatched pyrimidines. This modification has the nearest neighbour preference $G = C \geq U > A$ (Wagner et al. 1989), but there is a preference for modification of A with an upstream A or U (Bass, pers. comm.; Kimelman and Kirschner 1989), suggesting that there may be some sequence selectivity of modification.

The activity that modifies dsRNA is unusual, because it also apparently involves an RNA helicase activity. Enzymes that unwind both RNA and DNA are well known. A large family of ATP-dependent RNA helicases has been described; it includes the eukaryotic translation initiation factor e1F-4A, SV40 large T antigen and p68, a nuclear protein involved in the control of growth and division (Ford et al. 1988; Hirling et al. 1989; Scheffner et al. 1989). None of these activities is known to unwind and modify dsRNA. Moreover, the dsRNA unwinding and modifying activity does not apparently require ATP as an energy source (Bass and Weintraub 1988). While it is possible that the unwinding of dsRNA involves an RNA helicases, A to I modification per se would produce an unwound molecule, because of the inability to base-pair. The unwinding and modification of dsRNA activity would then require only adenosine deamination or base exchange.

Several enzymes have been characterized that catalyze the conversion of A to I. Adenosine deaminase and adenylic deaminase are enzymes of purine metabolism that deaminate at the N6 position of the adenine ring and convert adenine to hypoxanthine (Zielke and Suelter 1971). However, adenosine deaminase and adenylic deaminase do not unwind and modify double-stranded RNA, and *Xenopus* extracts do not modify adenosine or adenosine monophosphate, the normal substrates for these enzymes (Bass and Weintraub 1988). Moreover, the unwinding and modification activity in *Xenopus* extracts is not inhibited by deoxycoformycin, a potent inhibitor of adenosine and adenylic deaminases. Another activity that could mediate this conversion is the enzyme tRNA hypoxanthine ribosyltransferase. This enzyme catalyzes the replacement of adenine with hypoxanthine, in position 34 of the anti-codon loop of certain tRNAs. It is not yet known if this enzyme can perform this function.

The function of the dsRNA unwinding and modifying activity could be to reverse or modulate specific RNA/RNA interactions which would alter the expression of certain RNAs in a post-transcriptional fashion. Alternatively, it could lead to the degradation of dsRNA in a physiological role or as a cellular defence system against double-stranded RNA viruses. In *Xenopus* basic FGF and an FGF anti-sense RNA that encodes a highly conserved 21 kD protein are together a natural substrate for the unwinding and modifying activity, providing the most direct link to growth and development (Kimelman and Kirschner 1989). A dsRNA unwinding and modifying activity, which is ATP-independent has recently been described, which is abundant in natural killer cells, but absent from primary glial cells such as astrocytes and oligodendrocytes (Wagner and Nishikura 1989).

The unwinding and modifying activity has been implicated in measles inclusion body encephalitis (Bass et al. 1989). In the matrix gene of a defective measles virus isolated from a patient with measles inclusion body encephalitis, 50% (132 of 266) of the U residues were found to be changed to C. This modification was described as "biased hypermutation". dsRNA unwinding and conversion of A residues to I has been postulated as an explanation for this phenomenon. Thus, if nascent transcripts or completed measles virus mRNAs were to undergo hybridization to genomic RNA (Fig. 2), then under these circumstances the RNA unwinding and modifying activity would introduce A to I mutations in both strands of the duplex partly detaching the mRNA from the genomic RNA. In the event that this RNA could undergo replication, I residues would direct the incorporation of C, giving rise to U to C mutations in the positive strand with A to G changes in the negative strand. In slow viral infection

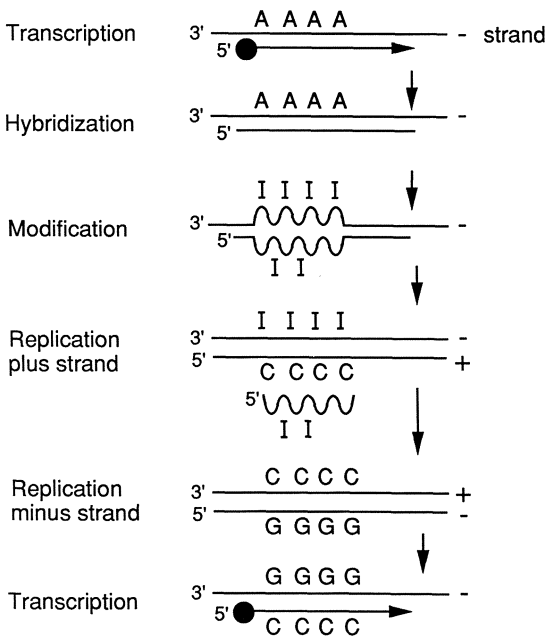

Fig. 2. Scheme for biased hypermutation of viral RNA genomes by unwinding and modification of dsRNA (Cattaneo et al. 1989)

inactivation of the matrix gene might have selective advantage. Additional examples of A to G mutations in minus-strand RNA viruses give support to this hypothesis (Bass et al. 1989; Cattaneo et al. 1989 and in press).

2.4 Paramyxovirus mRNA Editing

The paramyxovirus simian virus 5 (SV5) has a single-stranded anti-sense genomic RNA of 15 kb (Thomas et al. 1988). Viral mRNAs are transcribed in vivo by a viral RNA transcriptase. The SV5 P gene has been shown to encode two proteins: the P protein of 44 kD and the V protein of 24 kD. The P and V proteins have tryptic peptides in common, but no precursor/product relationship has been demonstrated between them. Analysis of the complete nucleotide sequence of the P gene has shown it to contain two open reading frames of 222 and 250 residues (Fig. 3). The V protein is coded for by the 222 residue reading frame, but neither of these is sufficient to encode the P protein. The P and V proteins are translated from independent mRNAs that are synthesized in SV5-infected cells and have been found to differ from each other by the presence in the P mRNA of two additional G residues which are not templated by the SV5 viral RNA. The two additional Gs follow a stretch of four genomically encoded Cs. These two additional nucleotides convert the two open reading frames of the gene to a single open reading frame coding for the 392 amino acid residue P protein, so that the P and V proteins have the same 164 amino acid residues

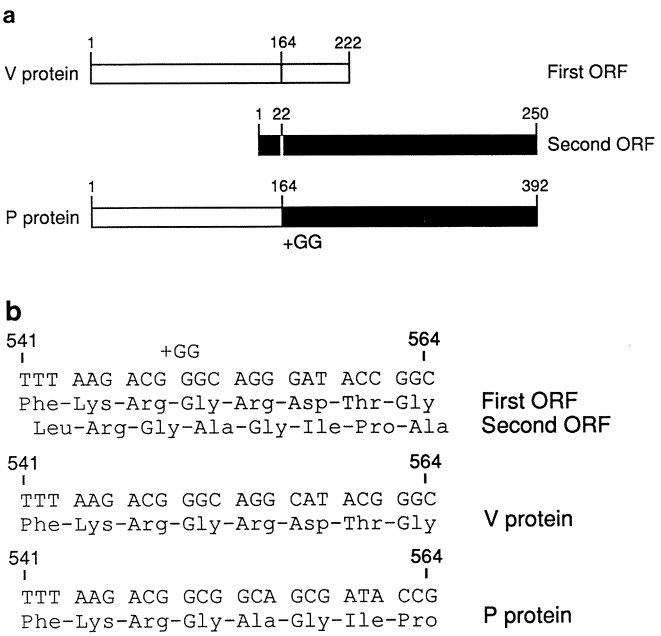

Fig. 3. **a** Origin of V and P proteins from simian virus 5 P gene. **b** Sequences of simian virus 5 genomic RNA at site of G residue insertion (Thomas et al. 1988)

amino terminal. The P protein of mumps virus is made in the same way (Paterson et al. 1989).

The most striking characteristic of either protein is the unique carboxy terminal portion of the V protein that contains a cysteine-rich region bearing close resemblance to the cysteine-rich regions found in adenovirus E1A protein, yeast transcription factor GAL4, and the steroid hormone receptor superfamily. Open reading frames that encode the cysteine-rich region are found in all other paramyxovirus P genes, suggesting that they have been conserved for functional reasons, possibly concerning the control of transcription or replication (Paterson et al. 1989).

The measles virus P protein gene encodes a P protein of 70 kD and a V protein of 20 kD (Cattaneo et al. 1989). Examination of cDNAs and RNA sequences representing the P and V mRNAs and the RNA genome has shown that the measles virus P gene also encodes a third product, the V protein of 46 kD containing the amino terminal region of the P protein, but a different cysteine-rich carboxy terminal domain. The V protein is translated from mRNAs containing one extra G residue that has been inserted after three Cs in the genome. This situation is also found in the Sendai virus (Paterson et al. 1989). It is the reverse of that in SV5 and mumps virus. About 50% of the P mRNAs contain the modification and, in addition, a small fraction contains several additional G residues. Therefore, a significant difference between the SV5 and measles virus editing is that the SV5 editing appears to be precise, whereas that in the measles virus is imprecise with regard to the number of G residues added.

The mechanism by which the untemplated G residues are added to the paramyxovirus mRNAs is not known. However, the viral RNA polymerase of negative strand RNA viruses also adds the poly-A tail to viral mRNAs (Thomas et al. 1988; Paterson et al. 1989). This is considered to occur due to "slippage" over the copying of a stretch of templated U residues at the 3′ end of each gene. Since the nontemplated G residues are added to the P RNA after a stretch of C residues in the genome, it is possible that the viral polymerase slips as it copies this region of the genome and adds nontemplated nucleotides. Immediately upstream of the 4 C residues in the SV5 genomic RNA is a sequence 3′-AAAAUUCU-5′, which resembles the presumptive poly-A addition signal found at the end of SV5 genes. Possibly this sequence promotes mRNA editing by a mechanism similar to viral polyadenylation. A stable secondary structure has been identified in the region that undergoes RNA editing, and it has been noted that this structure inhibits processive synthesis by T, DNA polymerase, suggesting that there is a native secondary structure in this region and that this could also be involved in editing. It is not known whether the measles virus transcripts with an excess of Gs are wasted or trimmed to the correct sequence.

2.5 Other Examples of Viral RNA Editing

Viruses have evolved other forms of RNA editing for the production of mRNA. A cotranslational RNA modification has been described in *Vaccinia* virus late

transcripts, which contain a poly-A leader at the 5′ end which is not encoded by the viral genome (Schwer and Stunnenberg 1988). The 5′ polyadenylated region is considered to be added by "slippage" of the *Vaccinia* virus polymerase at a series of T residues found in the DNA template. Another example of modification of RNA sequences has been called jumping RNA transcription, to explain unusual aspects of *Corona* virus mRNAs (Makino et al. 1988a, b). Thus, a defective mouse hepatitis virus RNA is made up of three separate genomic regions formed by discontinuous RNA synthesis and fused into a single reading frame. It has been suggested that a leader RNA is synthesized, disassociates from its template and reassociates at a downstream initiation site to act as a primer for further transcription.

3 RNA Editing in Mitochondria

3.1 Trypanosomes

The mitochondria of kinetoplastid trypanosomes (*T. brucei, Leishmania tarentolae* and *Crithidia fasciculata*) contain two types of circular DNA molecules called mini- and maxicircles (Benne 1989; Simpson and Shaw 1989; Stuart 1989; Stuart et al. 1989; Stuart, pers. comm.). The function of the minicircle DNA is not known. The maxicircles contain the genes of the respiratory chain complexes and small mitochondrial ribosomal RNAs. In addition, there are a number of maxicircle, unidentified, open reading frames (MURFs). The respiratory chain subunits include apo cytochrome B (CYb), cytochrome oxidase subunits I, II and III (COI, COII and COIII), and NADH dehydrogenase subunits 1, 4 and 5 (ND1, ND4 and ND5). The maxicircle genes are unusual in that several lack ATG initiation codons and some contain internal reading frameshifts, which are conserved between different trypanosome species. Comparison of the mitochondrial genomic DNA sequence with that of mitochondrial cDNAs led to the remarkable discovery that the RNA is remodelled by the insertion and deletion of U residues (Table 1; Benne et al. 1986; Feagin et al. 1987, 1988; Abraham et

Table 1. Comparison of COIII mRNA and genes in trypanosomes[a]

COIII mRNA	LeuCysPheTrpPheArgPheCysCys CysCysPheValLeuTrpLeuSer
T. bruce DNA	G GTTTTTGG AGG G GTTTTG G G A A GA GAG
T. bruce mRNA	uuGuGUUUUUGGuuuAGGuuuuuuuGuuG UUGuuGuuuuGuAuuAuGAuuGAGu
C. fasiculata DNA	TTTTTATTTTGATTTCGTTTTTTTTTATG TGTATTATTTGTGCTTTGATCCGCT
L. tarentolae DNA	TTTTTATTTTGATTTCGTTTTTTTTTATG TGTTTTATTTATGTTATGAGTAGCA

[a] DNA sequences are shown in upper case; uridine sequences in mRNA that are not encoded by DNA in lower case (insertions); thymidines that do not correspond to uridines in mature COIII mRNA by gaps (deletions). Differences between DNA and mRNA can be accounted for by insertion and deletion of uridines; cytidine, guanidine and adenosine residues are identical (Feagin et al. 1988; Maizels and Weiner 1988).

al. 1988). This process generates translatable RNA by the production of in-frame AUG initiation codons at the 5′ end of the coding sequence and by the correction of internal frameshifts. Evidence from Western blotting indicates that edited mRNAs are translated into protein. An inactive MURF3 RNA is also produced by insertion of a U that moves the AUG codon out of frame. In addition, editing occurs in 3′ poly-A extensions since many transcripts have been found to contain poly-(AU) tails. In the COIII, MURF3 and MURF4 mRNAs of *T. brucei* 50% of the sequence information has been produced by RNA editing. The COIII mRNA is not so extensively edited in *Leishmania* or *Crithidia.* In MURF3 mRNA editing occurs in two domains, but editing is not generally limited to special regions of the mRNA. Editing is usually found 3′ to a purine in purine-rich sequences.

Mitochondria contain fully edited, partly edited and unedited transcripts. Invariably, partly edited transcripts contain an unedited 5′ region and edited 3′ sequences. Editing does not progress in a strict 3′ to 5′ order, but overall 5′ to 3′ editing is not found. The junction between edited and unedited regions varies in size, and the extent of partial editing of the two domains of different MURF3 mRNAs is out of phase. Some partly edited transcripts contain more or less Us than are found in mature mRNA, and such partly edited sites are often contiguous.

From all of these observations, a series of several deductions have been made about the mechanism of trypanosome mitochondrial RNA editing (Maizels and Weiner 1988; Benne 1989; Simpson and Shaw 1989; Stuart 1989; Stuart et al. 1989; Stuart, pers. comm.). The presence of U residues in the poly-A tail indicates that it is a post-transcriptional event. Editing is unlikely to be cotranslational as U residues are added in a 3′ to 5′ direction. The presence of out of phase domains of editing indicates that initiation occurs at more than one site. The presence of partly edited transcripts with more or less Us than the final mRNA suggests that cleavage, poly-U polymerase and exonuclease, and RNA ligase activities are required for editing. Poly-U polymerase and RNA ligase activities have been found in trypanosome mitochondria.

The remarkable specificity and extensiveness of editing in some transcripts have led to the proposal that it is templated by either DNA or RNA. The presence of numerous, different, partially edited transcripts affecting various sites and failure to find a template argue against its existence. Moreover, the variation in size of the junction between edited and unedited mRNA, the observation that partly edited sites do not appear to be in an exact 3′ to 5′ order, and that there may be more or less Us in the partially edited RNAs than in the final mRNA indicate that editing may jump, and that re-editing may occur. These observations favour the involvement of a code involving secondary or higher order structure in the primary transcript that is used to recognize the editing sites. A macromolecular complex called the "editosome" which behaves like the spliceosome and possibly uses small RNAs has been postulated to be involved in this process.

The editing mechanism is under developmental control. It varies with the life cycle of the parasites, suggesting that RNA editing is involved in the regulation of maxicircle gene expression by the production of translatable mRNAs.

Table 2. Probable sites of RNA editing in plant cytochrome *c* oxidase (COII) proteins[a]

Position	Amino acid: Plant	Amino acid: Non-plant	Plant codon	Closest consensus codon(s)	Species: Monocots: Maize	Monocots: Rice	Monocots: Wheat	Dicots: Pea	Dicots: Soybean	Dicots: *Oonethera*
87	Arg	Trp	*C*GG	*U*GG	+	+	+	+	+	+
95	Pro	Leu	CC*N*[b]	C*U*N	+	+		+	+	+
129	Arg	Trp	*C*GG	*U*GG	+	+	+	+	+	+
156	Pro	Leu	C*C*A	C*U*A	+	+	+			
196	Ser	Leu	U*C*A	U*U*A	+	+	+			+
207	Ser	Leu	U*C*A	U*U*A	+	+	+			
228	Arg	Cys	*C*GU	*U*GU			+			
235	Thr	Met	A*C*G	A*U*G	+	+	+	[c]	+	
250	Arg	Trp	*C*GG	*U*GG				+	+	+

[a] Inferred amino-acid substitutions in plant COII sequences at positions that are invariant in five non-plant species are listed. Amino acid positions are numbered according to the wheat sequence. The plant codons and the closest codon in the non-plant sequences are shown, together with the nucleotide change that would be required in the mRNA to convert the plant codon to the non-plant codon. Non-conservative amino acid substitutions are included. +, Site at which a particular plant COII protein differs from the non-plant consensus (that is, a possible site of RNA editing) (Covello and Gray 1989).

[b] Dicots: N denotes G; maize: N denotes A; rice: N denotes U.

[c] Codon not present in published pea COII gene sequence.

3.2 Plants

It has now been established for a variety of plants including wheat, maize, rice, pea, soya bean, petunia and *Oenothara* that the mitochondrial genes COII, COIII, ND4 and CYb undergo RNA editing at multiple sites at which a genomic C is converted to U in the mRNA (Table 2; Corvello and Gray 1989; Gualberto et al. 1989; Hiesel et al., in press). The mitochondrial genome in plants had previously been presumed to depart from the universal genetic code by the use of CGG to specified tryptophan instead of arginine. Thus RNA editing alters the coding sequence of plant mitochondrial mRNAs from the genomic sequence of plant mitochondrial DNA to that of non-plant mitochondrial DNA. Examination of the primary sequence and potential secondary structure in the vicinity of the edited nucleotides in the plant mitochondrial mRNA has revealed no clues as to motifs which the editing mechanism might recognize.

3.3 Other Forms of RNA Editing in Mitochondria

A cDNA with homology to the amino terminal portion of the alpha subunit of ATP synthetase has been described in the acellular slime mould *Physarum polycephalum,* which contains more than 50 uniformly spaced frameshifts minus one in the mitochondrial DNA sequence compared to the cDNA sequence (discussed by Simpson and Shaw 1989). The frameshifts are corrected by insertion of a single C residue. The formation of TAA stop translation codons by the polyadenylation of certain mammalian mitochondrial mRNAs, which would otherwise end in T, could also be classed as RNA editing, since translation is dependent on appropriate polyadenylation (Attardi and Schatz 1989).

4 Perspective

The protein coding sequence of mRNA was, until 2 years ago, considered inviolable. Now, four apparently disparate types of RNA modification that change the coding capacity of mRNA have been described in different species. Yet, with the possible exception of the editing of trypanosome mitochondrial mRNA, none of these processes requires novel machinery. The modification of nucleosides and nucleotides during purine and pyrimidine catabolism and in nucleic acid has been extensively documented (Zilke and Suelter 1971; Grossman 1981; Lindahl 1982; Björk et al. 1987; Normanly and Abelson 1989). A brief consideration of these processes may provide insight into some of the changes that affect mRNA.

More than 50 modified nucleotides have been characterized in tRNAs. Base modification in tRNA ranges from simple reactions like methyl transfer and transglycosylation to complex sequences which produce hypermodified bases. Distinct enzymes that recognize different tRNA primary and higher order structural motifs can produce the same base at different positions. Some of the modified nucleotides in tRNA are the same as those found in mRNA. m^6A is found in both tRNA and mRNA and may be generated by a similar methyl

transfer reaction (Björk et al. 1987; Narayan and Rottman 1988). The substitution of A with I occurs in the wobble base of tRNA and during the unwinding and modification of dsRNA. The conversion of A to I could be achieved by a simple deamination, but in the wobble base of tRNA it is mediated by the enzyme hypoxanthine ribosyltransferase. This enzyme exchanges adenine for hypoxanthine, the base of inosine, on the sugar phosphate without cleavage of the phosphodiester linkage in RNA (Elliott and Trewyn 1984). Although the modification of tRNA and dsRNA achieves the same base exchange, they act on RNAs with very different secondary structure. The enzyme hypoxanthine ribosyltransferase acts on the single-stranded RNA of the anticodon loop of tRNA, and does not apparently modify other adenosines (Elliott and Trewyn 1984). It seems unlikely therefore that hypoxanthine ribosyltransferase per se acts on dsRNA. Either deamination or transglycosylation could also mediate the C to U substitution that occurs in apo-B and plant mitochondrial RNA, but neither reaction is known to produce this change in tRNA. Another reaction that produces an apparent C to U conversion in tRNA is the biosynthesis of the hypermodified base lysidine (Muramatsu et al. 1988a, b; Normanly and Abelson 1989). Thus, few of the activities that modify bases in tRNA seem good candidates for the modification of mRNA, and it is probable that new enzymes will be found that act on mRNA.

Nature uses a variety of other strategies of base modification which may be related to mRNA modification. In the metabolism of purines and pyrimidines several enzymes are involved in deamination. Two enzymes of purine metabolism deaminate adenine nucleosides and nucleotides to form hypoxanthine, but these enzymes do not act on dsRNA, and the dsRNA unwinding and modifying activity is not active in purine catabolism. The enzyme of pyrimidine metabolism, cytidine deaminase, mediates the conversion of cytidine to uridine (Wentworth and Wolfenden 1978). It has yet to be established whether this enzyme can produce C to U conversion in apo-B mRNA or in plant mitochondrial mRNA.

Other mechanisms for changing bases are used by the DNA modification and repair enzymes, and these might also have parallel(s) in the modification of mRNA. DNA methyltransferase is active in the formation of 5-methylcytidine (Bestor and Ingram 1983). A related methyltransferase could be active in the formation of m^6A. The proofreading and repair of active genes and of DNA after replication is performed by a variety of enzymes. The first step in repair of G:T and G:U mismatched base pairs is mediated by glycosylases followed by excision of the apyrimidinic sugar phosphate and repair (Lindahl 1982; Wiebauer and Jiricny 1989). The more general strategy of exonucleolytic excision, gap-filling and ligation is also used in the DNA repair process (Grossman 1981). These types of reactions are not known to be involved in the modification of tRNA, but could be used in RNA editing. The task is now to purify the activities involved in RNA modification, define their mechanisms of action and discover their relationship to other biological processes. As yet it cannot be excluded that some of the simple hydrophilic reactions involved in RNA base modification, like methylation and deamination, are mediated by catalytic RNA or templated by RNA.

Trypanosome mitochondrial RNA editing is much more extensive than that of other forms of RNA editing, and appears to be quite distinct. However, like the general strategies of DNA repair, it involves cleavage and ligation and a U-specific polymerase (gap-filling) function (Benne 1989; Simpson and Shaw 1989; Stuart 1989; Stuart et al. 1989). The last two of these activities have been detected in trypanosome mitochondria. To analyze these activities further it is essential to set up systems for editing trypanosome mRNA in vitro so that the process can be taken apart and reassembled. A major question with the extensive editing in trypanosomes is whether it is templated by RNA or DNA, or whether the editing signals are in the primary or higher order structure of unedited RNA transcript. For either model it is probable that editing, like splicing and translation, is carried out by macromolecular complexes involving both enzymes and RNAs. The central issues are therefore to find the enzymes and the code.

Although the process of plant mitochondrial mRNA editing appears quite different from that which occurs in trypanosomes, both processes affect the same group of genes – albeit the only ones in the mitochondrial genome – and effect changes to the same nucleotide. If C to U conversion in plants were found to be mediated by base excision and replacement, then the extensive editing of trypanosome mitochondrial RNA could have evolved from these simple activities. A further intriguing possibility is that plant mitochondrial RNA editing might have evolved from editing in purple bacteria, which are considered to be the eubacterial ancestor of the plant mitochondrion (Covello and Gray 1989). If this is so, a tantalizing speculation is that bacterial, plant, apo-B and even trypanosome mRNA editing might be related. Whether the different forms of RNA editing are related or not, it should be asked why such diverse mechanisms of RNA editing have evolved. One feature in common between apo-B, plant and trypanosome mitochondrial, paramyxovirus and FGF mRNAs is that editing or modification interposes another layer for the regulation of gene expression, and this would have survival value.

Acknowledgments. I wish to acknowledge gratefully Brenda Bass, Axel Brennicke, Roberto Cattaneo, Nick Davidson, Tom Innerarity, Marc Kirschner, Robert Lamb, Reay Paterson and Kenneth Stuart for allowing me access to their work prior to publication.

References

Abraham JM, Feagin JE, Stuart K (1988) Characterization of cytochrome C oxidase III transcripts that are edited only in the 3′ region. Cell 55:267–272

Adams JM, Cory S (1975) Modified nucleosides and bizarre 5′-termini in mouse myeloma mRNA. Nature 255:28–33

Attardi G, Schatz G (1989) Biogenesis of mitochondria. Ann Rev Cell Biol 4:289–333

Bass BL, Weintraub H (1987) A developmentally regulated activity that unwinds RNA duplexes. Cell 48:607–613

Bass BL, Weintraub H (1988) An unwinding activity that covalently modifies its double-stranded RNA substrate. Cell 55:1089–1098

Bass BL, Weintraub H, Cattaneo R, Billeter MA (1989) Biased hypermutation of viral RNA genomes could be due to unwinding/modifying activity of double-stranded RNA. Cell 56:331

Benne R (1989) RNA-editing in trypanosome mitochondria. Biochem Biophys Acta 1007:131–139

Benne R, Van den Burg J, Brakenhoff JPJ, Sloof P, Van Boom JH, Tromp MC (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

Bestor TH, Ingram VM (1983) Two DNA methyltransferases from murine erythroleukemia cells: purification, sequence specificity, and mode of interaction with DNA. Proc Natl Acad Sci USA 80:5559–5563

Beutler E, Gelbart T, Han J, Koziol JA, Beutler B (1989) Evolution of the genome and genetic code. Selection at the dinucleotide level in methylation and polyribonucleotide cleavage. Proc Natl Acad Sci USA 86:192–196

Björk GR, Ericson JU, Gustafsson CED, Hagervall TG, Jönsson YH, Wikström PM (1987) Transfer RNA modification. Annu Rev Biochem 56:263–287

Bostrom K, Lauer SJ, Poksay KS, Garcia Z, Taylor JM, Innerarity TL Apolipoprotein B48 RNA editing in chimeric apolipoprotein EB. J Biol Chem 264 (26):15701–15708

Brown MS, Goldstein JL (1987) Teaching old dogmas new tricks. Nature 330:113–114

Cattaneo R, Kaelin K, Baczko K, Billeter MA (1989) Measles virus editing provides an additional cysteine-rich protein. Cell 56:759–764

Cattaneo R, Schmid A, Spielhofer P, Kaelin K, Baczko K, Ter Meulen V, Pardowitz J, Flanagan S, Rima BK, Udem SA, Billeter MA Mutated and hypermutated genes of persistent measles viruses which caused lethal human brain diseases. Virology (in press)

Chen SH, Habib G, Yang CY, Gu ZW, Lee BR, Weng SA, Silberman SR, Cai SJ, Deslypere JP, Rosseneu M, Gotto AM Jr, Li WH, Chan L (1987) Apolipoprotein B-48 is the product of a messenger RNA with an organ-specific in-frame stop codon. Science 238:363–366

Chen SH, Wu JH, Li X, Liao WSI, Chan L (1989) RNA editing in mammals is not confined to apolipoprotein B. Cell Biol 109:96a (abstr)

Covello PS, Gray MW (1989) RNA editing in plant mitochondria. Nature 341:662–666

Davidson NO, Powell LM, Wallis SC, Scott J (1988) Thyroid hormone modulates the introduction of a stop codon in rat liver apolipoprotein B messenger RNA. J Biol Chem 263:13482–13485

Davies MS, Wallis SC, Driscoll DM, Wynne JK, Williams GW, Powell LM, Scott J (1989) Sequence requirements for apolipoprotein B RNA editing in transfected rat hepatoma cells. J Biol Chem 264:13395–13398

Desrosiers RC, Friderici KH, Rottman FM (1975) Characterization of Novikoff hepatoma mRNA methylation and heterogeneity in the methylated 5′ termini. Biochemistry 14:4367–4374

Driscoll DM, Wynne JK, Wallis SC, Scott J (1989) An in vitro system for the editing of apolipoprotein B mRNA. Cell 58:519–525

Elliott MS, Trewyn RW (1984) Inosine biosynthesis in transfer RNA by an enzymatic insertion of hypoxanthine. J Biol Chem 259:2407–2410

Feagin JE, Jasmer DP, Stuart K (1987) Developmentally regulated addition of nucleotides within apocytochrome b transcripts in *Trypanosoma brucei*. Cell 49:337–345

Feagin JE, Abraham JM, Stuart K (1988) Extensive editing of the cytochrome c oxidase III transcript in *Trypanosoma brucei*. Cell 53:413–422

Ford MJ, Anton IA, Lane DP (1988) Nuclear protein with sequence homology to translation initiation factor eIF-4A. Nature 332:736–738

Grossman L (1981) Enzymes involved in the repair of damaged DNA. Arch Biochem Biophys 211:511–522

Gualberto JM, Lamattina L, Bonnard G, Weil JH, Grienenberger JM (1989) RNA editing in wheat mitochondria results in the conservation of protein sequences. Nature 341:660–662

Hiesel R, Wissinger B, Schuster W, Brennicke A RNA editing in plant mitochondria. Science (in press)

Hirling H, Scheffner M, Restle T, Stahl H (1989) RNA helicase activity associated with the human p68 protein. Nature 339:562–564

Hospattankar AV, Higuchi K, Law SW, Meglin NM, Brewer HB Jr (1987) Identification of a novel in-frame translational stop codon in human intestine apo-B mRNA. Biochem Biophys Res Commun 148:279–285

Kimelman D, Kirschner MW (1989) An antisense mRNA directs the covalent modification of the transcript encoding fibroblast growth factor in *Xenopus* oocytes. Cell 59 (4):687–696
Lindahl T (1982) DNA repair enzymes. Ann Rev Biochem 51:61–87
Maizels N, Weiner A (1988) In search of a template. Nature 334:469–470
Makino S, Soe LH, Shieh CK, Lai MMC (1988a) Discontinuous transcription generates heterogeneity at the leader fusion sites of coronavirus mRNAs. J Virol 62:3870–3873
Makino S, Shieh S, Soe L, Baker S, Lai M (1988b) Primary structure and translation of a defective interfering RNA of murine coronavirus. Virology 166:550–560
Muramatsu T, Yokoyama S, Horie N, Matsuda A, Ueda T, Yamaizumi Z, Kuchino Y, Nishimura S, Miyazawa T (1988a) A novel lysine-substituted nucleoside in the first position of the anticodon of minor isoleucine tRNA from *Escherichia coli.* J Biol Chem 263:9261–9267
Muramatsu T, Nishikawa K, Nemoto F, Kuchino Y, Nishimura S, Miyazawa T, Yokoyama S (1988b) Codon and amino-acid specificities of a transfer RNA are both converted by a single post-transcriptional modification. Nature 336:179–181
Narayan P, Rottman FM (1988) An in vitro system for accurate methylation of internal adenosine residues in messenger RNA. Science 242:1159–1161
Normanly J, Abelson J (1989) tRNA Identity. Ann Rev Biochem 58:1029–1049
Okada N, Nishimura S (1979) Isolation and characterization of a guanine insertion enzyme, a specific tRNA transglycosylase, from *Escherichia coli.* J Biol Chem 254:3061–3066
Okada N, Noguchi S, Kasai H, Shindo-Okada N, Ohgi T, Goto T, Nishimura S (1979) Novel mechanism of post-transcriptional modification of tRNA. J Biol Chem 254:3067–3073
Paterson RG, Thomas SM, Lamb RA (1989) Specific non-templated nucleotide addition to a simian virus 5 mRNA: prediction of a common mechanism by which unrecognized hybrid P-cysteine-rich proteins are encoded by paramyxovirus "P" genes. In: Kolakofsky D, Mahy BWJ (eds) Genetics and pathogenicity of negative strand viruses. Elsevier, London, pp 232–245
Powell LM, Wallis SC, Pease RJ, Edwards YH, Knott TJ, Scott J (1987) A novel form of tissue-specific RNA processing produces apolipoprotein-B48 in intestine. Cell 50:831–840
Rebagliati MR, Melton DA (1987) Antisense RNA injections in fertilized frog eggs reveal an RNA duplex unwinding activity. Cell 48:599–605
Scheffner M, Knippers R, Stahl H (1989) RNA unwinding activity of SV40 large T antigen. Cell 57:955–963
Schwer B, Stunnenberg HG (1988) Vaccinia virus late transcripts generated in vitro have a poly(A) head. EMBO J 7:1183–1190
Simpson L, Shaw J (1989) RNA editing and the mitochondrial cryptogenes of kinetoplastid protozoa. Cell 57:355–366
Stuart K (1989) Minireview – trypanosomatids: mitochondrial RNA editing. Exp Parasitol 68:486–490
Stuart K, Feagin JE, Abraham JM (1989) RNA editing: the creation of nucleotide sequences in mRNA – a minireview. Gene 82 (1):155–160
Thomas SM, Lamb RA, Paterson RG (1988) Two mRNAs that differ by two nontemplated nucleotides encode the amino coterminal proteins P and V of the paramyxovirus SV5. Cell 54:891–902
Wagner RW, Nishikura K (1989) An activity in mammalian cells which modifies adenosines to inosines in double-stranded RNA. J Cell Biol 109:96a (abstr)
Wagner RW, Smith JE, Cooperman BS, Nishikura K (1989) A double-stranded RNA unwinding activity introduces structural alterations by means of adenosine to inosine conversions in mammalian cells and *Xenopus* eggs. Proc Natl Acad Sci USA 86:2647–2651
Wentworth DF, Wolfenden R (1978) Cytidine deaminases (from *Escherichia coli* and human liver). Meth Enzymol LI:401–408
Wiebauer K, Jiricny J (1989) In vitro correction of G·T mispairs to G·C pairs in nuclear extracts from human cells. Nature 339:234–236
Zielke CL Suelter CH (1971) Purine, purine nucleoside and purine nucleotide aminohydrolases. In: Boyer PD (ed) The enzymes, vol 4. Academic Press, New York, pp 47–78