Nucleic Acids Research, 1997, Vol. 25, No. 19 **3751–3759**

The mechanism of U insertion/deletion RNA editing in kinetoplastid mitochondria

Juan D. Alfonzo, Otavio Thiemann and Larry Simpson*

Howard Hughes Medical Institute and Department of Molecular, Cell and Developmental Biology, University of California, Los Angeles, CA 90095-1662, USA

Received June 18, 1997; Revised and Accepted July 29, 1997

ABSTRACT

Recent advances in *in vitro* systems and identification of putative enzymatic activities have led to the acceptance of a modified 'enzyme cascade' model for U insertion/deletion RNA editing in kinetoplastid mitochondria. Models involving the transfer of uridines (Us) from the 3′-end of gRNA to the editing site appear to be untenable. Two types of *in vitro* systems have been reported: (i) a gRNA-independent U insertion activity that is dependent on the secondary structure of the mRNA; (ii) a gRNA-dependent U insertion activity that requires addition of a gRNA that can form an anchor duplex with the pre-edited mRNA and which contains guiding A and G nucleotides to base pair with the added Us. In the case of the gRNA-mediated reaction, the precise site of cleavage is at the end of the gRNA–mRNA anchor duplex, as predicted by the original model. The model has been modified to include the addition of multiple Us to the 3′-end of the 5′-cleavage fragment, followed by the formation of base pairs with the guiding nucleotides and trimming back of the single-stranded oligo(U) 3′-overhang. The two fragments, which are held together by the gRNA 'splint', are then ligated. Circumstantial *in vitro* evidence for involvement of an RNA ligase and an endoribonuclease, which are components of a 20S complex, was obtained. Efforts are underway in several laboratories to isolate and characterize specific components of the editing machinery.

INTRODUCTION

Kinetoplastid protozoa have a single mitochondrion which contains the mitochondrial (kinetoplast) DNA in the form of a nucleoid body situated within the kinetoplast portion of the mitochondrion adjacent to the basal body of the flagellum (1,2). There are two major known taxonomic groups within the kinetoplastids: the trypanosomatids and the bodonids/cryptobiids (3). Kinetoplast DNA (kDNA) has been extensively studied from several trypanosomatid species (*Trypanosoma brucei*, *Leishmania tarentolae*, *Trypanosoma cruzi* and *Crithidia fasciculata*) and from one cryptobiid species (*Trypanoplasma borreli*).

Uridine (U) insertion/deletion RNA editing was first described in the kinetoplast-mitochondrion of *T.brucei* (4). This unusual RNA modification process (5) involves the insertion and, to a lesser extent, the deletion of U residues from transcripts of maxicircle 'cryptogenes' (6–11). The extent of editing varies from a few Us at a few adjacent sites to hundreds of Us at hundreds of sites over the entire gene ('pan-editing') (12). Editing corrects frameshifts, creates translation initiation codons and, in the case of pan-edited genes, converts the transcripts of unrecognizable cryptogenes into translatable mRNAs (13). Editing has been shown to be developmentally regulated in *T.brucei* (9,14,15).

The complementary sequence information for the specific insertion and deletion of U residues resides in a novel class of short 3′-oligo(U) RNAs, which can form 'anchor' duplexes with mRNA just downstream of specific editing blocks (16). These 'guide RNAs' (gRNAs) appear to specify the insertion and deletion of U residues by base pairing. In addition, the observed overall 3′→5′ polarity of editing site selection within an editing domain results from the mediation of multiple overlapping gRNAs, in which upstream anchor sequences are created by downstream editing. gRNAs are transcribed from both the maxicircle and the minicircle components of the kDNA in trypanosomatids (16–20) and from the 180 kb circles in the kDNA of *T.borreli* (21).

The initial hypothesis for the mechanism of RNA editing was the 'enzyme cascade' model (16), in which the gRNA was mostly a passive carrier of the editing information and the act of information transfer was relegated to protein catalyzed cleavage–ligation reactions. The source of U residues was either UTP or, in a variant of this model, the 3′-oligo(U) tail of the gRNA itself (22). The evidence for this model was initially based on the existence in mitochondrial lysates of several enzymatic activities, including a terminal uridylyl transferase activity, RNA ligase activity (23,24) and pre-edited region-specific endoribonuclease activity (25,26). A second model was based on an analogy to Group I and Group II RNA catalyzed splicing reactions. In this 'double transesterification' model the source of U residues is again either UTP (27) or the 3′-oligo(U) tail of the gRNA (28). The evidence for the latter type of transesterification model was the *in vivo* existence of gRNA–mRNA chimeric molecules (28), which were the predicted intermediates of the first transesterification. Chimeric gRNA–mRNA molecules could also be generated *in vitro* by incubation of synthetic RNAs with mitochondrial extracts or with glycerol gradient fractionated extracts (29–32).

*To whom correspondence should be addressed. Tel: +1 310 825 4215; Fax: +1 310 206 8967; Email: simpson@hhmi.ucla.edu

Table 1. Summary of U insertion and U deletion editing events

Maxicircle gene	Species *L.tarentolae*	*T.brucei*	*C.fasciculata*	*T.borreli*	*T.cruzi*
ND1	0[a]–0[b]	0–0	5′–0		
ND4	0–0	0–0			
ND3	176–5	210–13			
ND5	0–0	0–0			
ND7	24–0	553–89	27–0		
ND8	218–40	259–46	nd		
ND9	334–41	345–20	nd		
RPS12	117–32	132–28	135–7	133–32	
A6	106–5	447–28	68–3		446–55
Cytb	39–0	34–0	39–0	5′: 47–4 3′: 144–40	
CO1	0–0	0–0	0–0	5′: 72–11 3′: 177–13	
CO2	4–0	4–0	4–0	0–0	4–0
CO3	29–15	547–41	32–2	0–0	
MURF1	0–0	0–0			
MURF2	28–4	26–4	30–0		
MURF5	nd	nd	nd		
G3	35–14	148–13			
G4	326–5	325–40			
12S rDNA	0–0	0–0	0–0	0–0	
9S rDNA	0–0	0–0	0–0	0–0	

[a]Uridine additions.
[b]Uridine deletions.
nd, not analyzed experimentally.

ORGANIZATION OF THE MITOCHONDRIAL GENOME IN KINETOPLASTIDS

kDNA is a unique biological structure, both physically and genetically. Physically, the kDNA of trypanosomatids consists of 20–50 catenated homoplasmic maxicircle molecules, 20–40 kb in size in different species, and ~5000–12 000 catenated minicircles, 0.8–2.5 kb in size in different species (33). Several small basic proteins from *C.fasciculata* have been isolated which could be cross-linked to kDNA by treatment of cells with formaldehyde (34). These proteins have limited homology with histone H1 sequences and have been proposed to be involved with condensation of the kDNA network *in situ* into the highly structured nucleoid body (35). Evidence has also been presented for a physical rotation of the nucleoid body within the mitochondrion during S phase in *Leishmania* and *Crithidia*, but not in *Trypanosoma* (36,37). Covalently closed minicircles are randomly removed from catenation in the network by a Type II topoisomerase and replication occurs within two replisomes situated at either end of the nucleoid body (38). We have previously speculated that rotation of the nucleoid body aids redistribution within the network of catenated minicircles encoding different gRNAs, to avoid loss of specific minicircle sequence classes upon segregation of the nucleoid body in daughter cells (7).

The kDNA of the cryptobiid *T.borreli* consists of two classes of large circular molecules, Component II, varying from 40 to 80 kb in two strains, and Component I, 180–200 kb in size, which do not appear to be catenated into a network (39,40).

Genetically, the kDNA genomes from both trypanosomatids and cryptobiids are similar in that they are composed of two separate but interacting genomes, one of which encodes rRNAs, structural genes and cryptogenes and the other of which encodes the gRNAs. In the trypanosomatids the maxicircle DNA encodes the cryptogenes (and a few gRNAs) and the minicircle DNA encodes the majority of the gRNAs. In *T.borreli* the 40–80 kb circles represent the maxicircle homologs and the 180 kb circles the minicircle homologs (21). The question of the evolutionary origin of the minicircle awaits investigation of additional bodonid/cryptobiid species, but the simplest scenario is that an ancestral kinetoplastid contained gRNA genes segregated on large circles which were excised, linked to autonomous replication elements and circularized to yield minicircles.

The location and polarity of the rRNA and structural genes in the maxicircle genome are conserved in all trypanosomatid species analyzed. This is shown in Figure 1 for the sequenced regions of the *L.tarentolae*, *T.brucei* and *C.fasciculata* maxicircles. The 9S and 12S rRNAs are not edited, but the RNA transcripts possess 3′-oligo(U) tails (41); the 3′-tail of the 12S rRNA is heterogeneous in length but that of the 9S rRNA has 11 Us. These non-encoded 3′-oligo(U) sequences are probably added by the known mitochondrial terminal uridylyl transferase (TUTase) activity, but whether this represents an adventitious by-product of the editing machinery or part of the normal maturation process is not known. gRNAs also have non-encoded 3′-oligo(U) tails ranging from 5 to 30 Us (42). mRNA transcripts of both genes and cryptogenes have non-encoded 3′-oligo(AU)

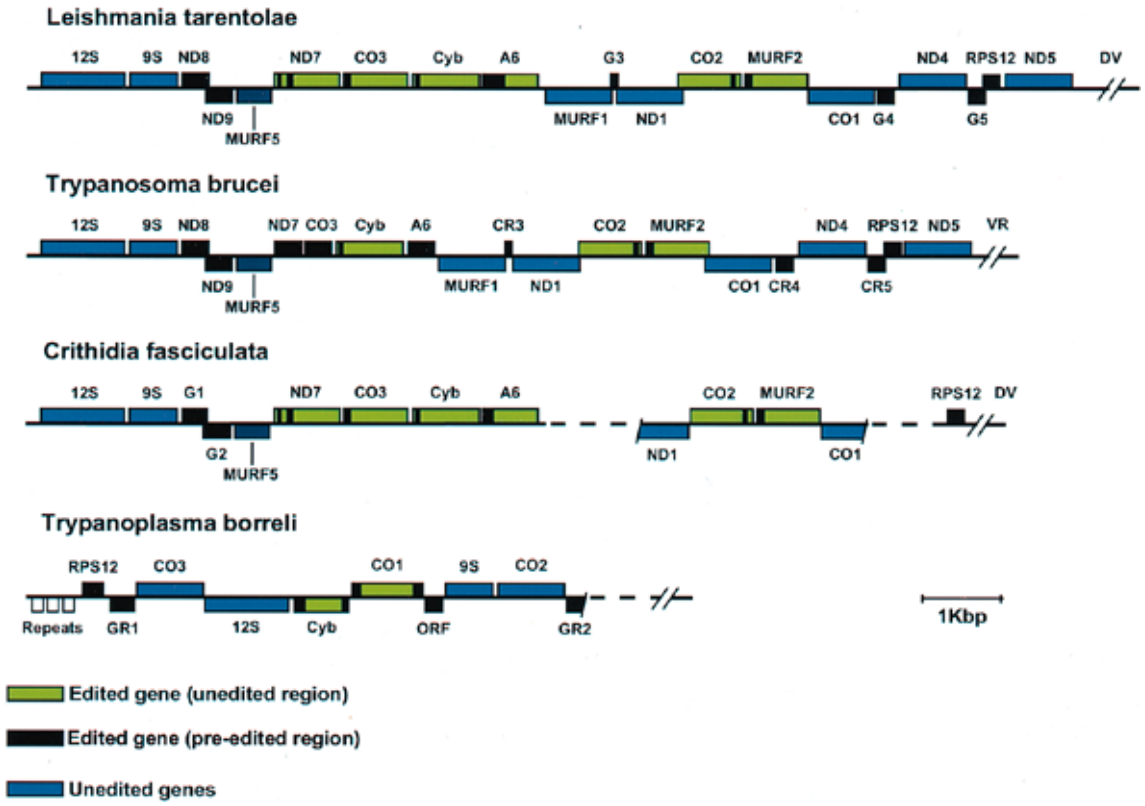

Figure 1. Comparative organization of maxicircle genomes in four kinetoplastid species. The maxicircles are shown linearized, with the genes above the line 5′→3′ left to right and the genes below the line 5′→3′ right to left. Unedited genes and pre-edited and unedited regions of edited genes are indicated as shown. DV, divergent region of *L.tarentolae* and *C.fasciculata*. VR, variable region in *T.brucei*. The unsequenced portions of the *C.fasciculata* genome are indicated by ----. G1, G2, G-rich regions in *C.fasciculata*, which are putative pan-edited ND8 and ND9 cryptogenes (from analysis of G-rich sequences in the genomic sequence; O.Thiemann and L.Simpson, unpublished results). MURF5 in *C.fasciculata* is an open reading frame homologous to the MURF5 genes in *L.tarentolae* and *T.brucei* (O.Thiemann and L.Simpson, unpublished results).

tails, presumably resulting from interaction of the 3′ polyadenylation machinery with TUTase activity (43).

The available sequence information from the 40–80 kb component of the *T.borreli* mitochondrial genome reveals a quite distinct genomic organization (Fig. 1). The identified gene homologs differ in relative location and polarity from those in trypanosomatids (39,40).

Eighteen structural genes have been identified in the maxicircle sequences of both *L.tarentolae* and *T.brucei*. The transcripts of six genes do not undergo editing and appear to be functional mRNAs: maxicircle unidentified reading frames 1 and 5 (MURF1 and MURF5), NADH dehydrogenase subunits 1, 4 and 5 (ND1, ND4 and ND5) and cytochrome oxidase subunit 1 (CO1). Transcripts of the cytochrome b (Cyb), MURF2, CO2 and ND7 and CO3 cryptogene in *L.tarentolae* undergo a limited form of editing at the 5′-end of editing domains. Transcripts of the ND3, ND8, ND9, G3 (CR3), G4 (CR4) and A6 cryptogenes in both species and the ND7 and CO3 cryptogenes in *T.brucei* undergo pan-editing. For example, in *L.tarentolae* up to 335 U additions and 40 deletions occur in the ND9 transcripts (44) and in *T.brucei* 547 U additions and 89 deletions occur in the ND7 transcripts (45). A novel type of edited Cyb gene is found in *T.borreli*, in which both 3′ as well as 5′ sequences are pan-edited (39). (See Table 1 for a summary of all known editing events.)

A comparative analysis of the extent of editing of homologous cryptogenes in various kinetoplastid species showed that in the evolution of these cells there was a retention of editing at the 5′-ends of editing domains and that the extent of editing in any species is dependent on the presence of minicircle-encoded gRNAs for the overlapping editing blocks (46). These data led to the suggestion that pan-edited genes are replaced in evolution with partially 3′-edited genes by a retroposition mechanism, leading to the observed stepwise decrease in the extent of edited regions (47,48).

Additional evidence for the retroposition model came from the observation that transcripts of the pan-edited cryptogenes G1–G5 are not edited in an old laboratory strain of *L.tarentolae* but are productively edited in a recently isolated strain (44). In the old laboratory strain the loss of editing was correlated with a loss of the minicircle sequence classes encoding the required gRNAs for these editing cascades. It was speculated that an absence of a selective pressure for the protein products of specific edited mRNAs could lead to the loss, perhaps by missegregation at mitochondrial division of entire low copy number minicircle sequence classes encoding gRNAs (48). In nature these proteins would be required for survival of the organism, leading to selection for cells in which a partially edited gene replaced a pan-edited gene.

The maxicircle genome also contains a subset of the gRNA genes: 15 maxicircle-encoded gRNA genes have been identified in *L.tarentolae*, seven in *C.fasciculata* and three in *T.brucei*. The gMURF2-I, gMURF2-II and gCO3 gRNA genes are conserved in these three species in location and polarity, although not in sequence.

The size, organization and complexity of the minicircle DNA molecules differ from species to species in the trypanosomatids (see 49 for a recent review). In *L.tarentolae* the minicircles are ~800–950 bp in size and contain a single conserved region ~170 bp in size and a single variable region. A 12mer conserved sequence (CSB-3) is present within the conserved region in all trypanosomatid minicircles. The variable region defines the specific minicircle sequence class and contains the encoded gRNA gene. The gRNA gene in *L.tarentolae* is located ~250–300 bp from the CSB-3 sequence downstream of an intrinsic DNA bend region of unknown function (50).

The complexity of the minicircle DNA in *L.tarentolae* was shown to vary dramatically between an old laboratory strain (UC) and a recently isolated strain (LEM125) (44). The UC strain contains a total of 17 minicircle sequence classes, whereas the LEM125 strain was shown to contain 32 additional classes and probably contains up to ~55 additional yet uncharacterized classes. The absolute number of minicircles per network is conserved between these two strains, but the frequency of individual minicircle sequence classes varies dramatically (O.Thiemann and L.Simpson, unpublished results). There is no correlation between minicircle copy number and steady-state abundance of the encoded gRNAs, suggesting that gRNA abundance is controlled by either relative promoter strength or turnover (51).

In *T.brucei* the minicircles are ~1 kb in size and also contain a single conserved region of ~120 bp and an associated DNA bend. They differ from the *Leishmania* minicircles, however, in terms of coding capacity; there are usually three gRNA genes flanked by 18mer inverted repeats within the variable region (18). Only three examples of gRNA coding genes localized outside inverted repeats have been reported (52). The number of minicircle sequence classes has been estimated to be 200–300 from DNA renaturation kinetic data (53). Multiple redundant gRNAs have been identified, which are defined as gRNAs of different sequence encoding the identical editing information due to the presence of G-U wobble base pairing (19).

The minicircles of *C.fasciculata* are ~2.5 kb in size and contain two antipodal conserved regions and one DNA bend sequence situated in the center of one of the variable regions (54). Five gRNA genes have been identified, which are localized to one of the variable regions ~60 bp from the bend (55). Interestingly, the kDNA of *C.fasciculata* is composed of a major minicircle class comprising >90% of the total minicircle content of this species (56); this class encodes a yet unassigned gRNA (55). The remaining classes represent minor components of the kDNA network. The presence of a major sequence class provides further evidence for plasticity of the minicircle genome in trypanosomatids.

The *T.cruzi* kDNA minicircle is ~1.5 kb in size and contains four 120 bp conserved sequences and four variable regions (57). Single gRNA genes are situated within each variable region ~100 bp from the CSB-3 conserved block (58). The genomic complexity of the *T.cruzi* minicircle population is very large and appears to rival that of *T.brucei*. Consistent with this is the presence of a large number of redundant gRNAs (58).

In *T.borreli* the 40–80 kb circle represents the maxicircle homolog and the 180 kb circle the minicircle homolog. Consistent with the phylogenetic separation of the cryptobiid lineage from the trypanosomatid lineage, the relative gene order and polarity of the rRNA genes, structural genes and cryptogenes differs from that found in the trypanosomatids (39,40). However, similar 3′→5′ U insertion/deletion editing occurs, along with misediting at junction regions in the partially edited transcripts. gRNA-like molecules were detected by capping total RNA with GTP and guanylyltransferase (21,39). This RNA hybridized to a 1 kb *Sca*I repeat present in tandem arrays in the 180 kb circle; the editing role of the encoded gRNA is unknown. Several additional identified gRNAs were obtained from a gRNA library and found also to be encoded in the 180 kb circles. The gRNAs are unusual in that, in addition to the 3′-oligo(U) sequences, they possess short non-encoded 5′-oligo(U) sequences of unknown function.

THE MECHANISM OF U INSERTION/DELETION RNA EDITING

Several variations of the two original models for RNA editing have been progressively made as new features of the RNA and protein components have accrued (59–62). All editing models share a common need to accommodate gRNA and mRNA interactions as a prelude to information transfer. The models diverge in regard to the underlying chemistry of the process and the roles played by protein components in catalyzing information transfer.

In the double transesterification model either the 3′-end of the gRNA (28) or free UTP (27) acts as the nucleophile that attacks the phosphodiester bond between the last nucleotide of the gRNA–mRNA anchor and the nucleotide following the first editing site on the mRNA (Fig. 2A). This nucleophilic attack leads to transesterification of the attacking group onto the mRNA and the production of a free mRNA 5′-fragment. The 3′-end of the newly produced 5′-fragment of the mRNA will subsequently act as the nucleophile in the second transesterification reaction, leading to regeneration of a partially edited mRNA (Fig. 2A). The existence of the predicted gRNA–mRNA chimeric molecules and the similarity to Group I and Group II intron RNA catalyzed splicing reactions provided the appeal of this model.

In recent years, however, evidence has been accumulating in support of the cleavage–ligation model for RNA editing (42), in which a series of protein-mediated cleavages, uridine insertions and ligations leads to the production of edited mRNA (Fig. 2B).

Frech *et al*. (63) described an *in vitro* editing-like activity in which a synthetic pre-edited mRNA substrate could be labeled internally with [α-^{32}P]UTP by incubation with a *L.tarentolae* mitochondrial extract. Although the insertion of Us occurs predominantly within the pre-edited region, addition of exogenous gRNA was not required and, in fact, inhibited the reaction (63), possibly by titrating away [α-^{32}P]UTP for the more favored gRNA 3′-uridylylation reaction (G.Connell and L.Simpson, unpublished results). Evidence against a transfer of Us from the gRNA 3′-oligo(U) tail to the insertion site in this reaction was provided by showing that internal incorporation of (Sp)-α-S-UTP proceeded by an inversion of the stereoconfiguration, as predicted by a ligase-mediated reaction, and not by the net retention of the stereoconfiguration predicted by a transfer model (64). Connell *et al*. (65) also showed, by primer extension assay, that the *in vitro* U insertion reaction occurring in

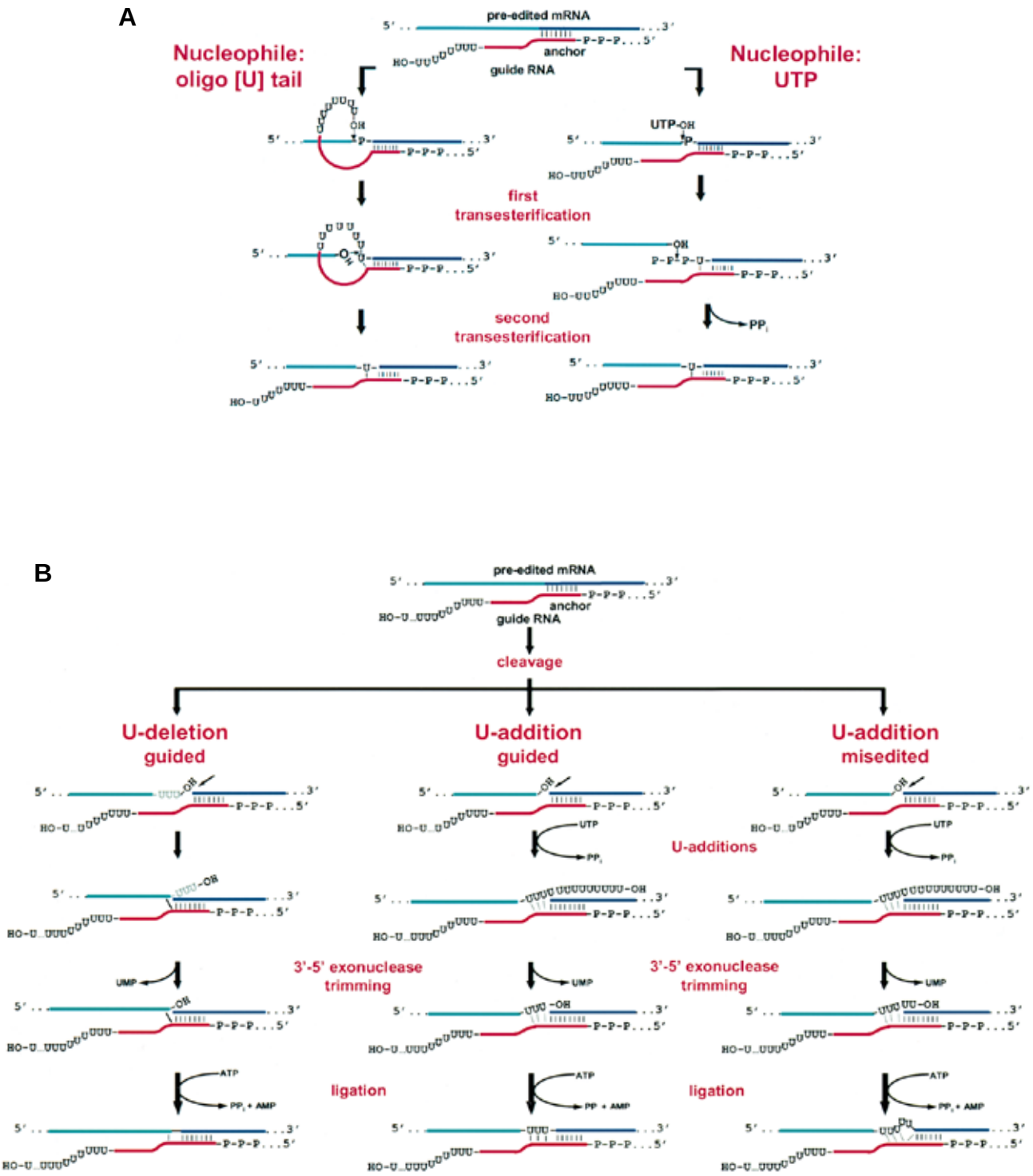

Figure 2. Diagram of models for U insertion/deletion RNA editing. The vertical lines indicate base pairs. The arrowheads indicate sites of cleavage. (**A**) Double transesterification models (27,28). Only U insertions are shown. (**B**) Modified enzyme cascade model (22). The 3′-oligo(U) tail of the gRNA is shown as a single-stranded overhang, but it is possible that the tail can interact with the purine-rich pre-edited sequence (16,61) and the gRNA may have secondary structure (81). In the U deletion model three unpaired Us (in gray) to be deleted are shown as an example. It is possible that the U addition activity adds Us to the 3′-end of the cleavage fragment at the deletion site, which are then trimmed back, but this scenario is not indicated. In the U addition model 13 Us are shown added to the 5′-fragment, but the evidence indicates that the number of added Us is actually heterogeneous (60). In the 'guided' diagrams the exonuclease nucleotide trimming is complete, yielding the correct –3 or +3 guided products. If trimming is incomplete or excessive prior to ligation, gRNA-dependent misedited products are produced, as shown in the 'misedited' diagram.

sites 1 and 2 of a Cyb pre-edited mRNA substrate was unaffected by mutating the anchor sequence of the mRNA substrate and therefore was also independent of endogenous gRNA. Nevertheless, the fact that U is the predominant nucleotide incorporated and that its incorporation is limited to the pre-edited region suggests a connection between this type of U insertion and bona fide RNA editing activity. This U insertion requires free UTP and ATP. ATP analogs non-hydrolyzable at the α-β phosphate could not satisfy the ATP requirement.

Connell *et al*. (65) also showed that the *in vitro* gRNA-independent U insertion activity in *L.tarentolae* is dependent on the presence of an intramolecular RNA duplex formed by complementary sequences upstream and downstream of the mRNA pre-edited region. They proposed that this duplex may structurally and functionally mimic the gRNA–mRNA anchor duplex in a gRNA-mediated system *in vitro*, but left the question open as to whether this also occurs *in vivo*.

An intriguing observation made by Frech *et al*. (63) was that U insertion activity was selectively inhibited by digestion of the extract with micrococcal nuclease. Since the insertion activity is independent of endogenous gRNA, at least for guiding the insertions, the suggestion is that either the endogenous gRNA is required for a non-coding function or that there is another RNA component in the editing complex required for U insertions. This remains an open question and requires confirmation by identification and characterization of the putative RNA component(s).

The first direct evidence for the mechanism of the editing process came from the work of Seiwert *et al*. (66), who initially visualized gRNA-directed U deletions at site 1 of the A6 mRNA by a primer extension assay using unfractionated *T.brucei* mitochondrial extract. This was the first confirmation of the hypothesis that gRNA mediated U deletion editing by base pairing. Seiwert *et al*. (61) then directly visualized putative intermediates in U deletion editing using a 73 nt end-labeled synthetic pre-edited A6 mRNA substrate incubated in a gradient-fractionated mitochondrial extract in the presence of synthetic cognate gRNA. The number of residues deleted from the mRNA could be manipulated in a predictable manner by affecting the extent of base pairing with the guiding nucleotides in the added gRNA. When 3′-end-labeled substrate RNA was incubated with the 20S glycerol gradient fraction four minor products were observed, which were shown to represent a molecule with the expected U insertions at site 1, a fragment representing the 3′-half of the substrate RNA cleaved adjacent to the Us to be deleted and two types of gRNA–mRNA chimeric molecules with different lengths of connecting U residues. In a time course experiment the 3′-cleavage fragment and edited product appeared a few minutes prior to the chimeric molecules and this was taken as evidence for the chimeric molecules not representing intermediates but by-products of the editing reaction. However, kinetics are difficult to interpret in a system in which all the components are not characterized and identified. The 5′-cleavage products were detected by 5′-end-labeling the substrate RNA (61). However, the major initial cleavage fragment visualized after 10 min incubation already had the –4 U deletion, but with further time of incubation molecules with four, three, two and one U were also visible, suggesting that these may not represent intermediates in successive U deletions but rather aberrant products of the reaction.

Mutations in the added gRNA were used to define important sequence elements required for the *in vitro* reaction. The anchor sequence was required, as was the 3′-oligo(U) tail, which had previously been proposed to hybridize with the purine-rich pre-edited sequence and stabilize the initial gRNA–mRNA interaction (42). However, artificially increasing the stability of the interaction between the 3′-tail of the gRNA and the mRNA prevented chimera formation but had no effect on generation of the edited mRNA, suggesting that chimeras represent aberrant by-products of the editing reaction (61). However, definitive evidence for this could not be obtained since blockage of the 3′-end of the gRNA by periodation, which would be expected to have no effect, inhibited the production of edited mRNA, as well as chimeras.

Cruz-Reyes *et al*. (67) analyzed the gRNA-directed U deletion mechanism in the *T.brucei* system in more detail. They showed that the cleavage occurs precisely at the first mismatched nucleotide upstream of the gRNA anchor duplex, as predicted by the enzyme cascade model; this cleavage site is incompatible with a chimera-based model mechanism. Evidence was presented that the 3′-exonuclease activity is not just a reversal of the TUTase activity and appears to be U-specific, although this must be confirmed with purified enzyme and model substrates.

Soon after the report of U deletion editing, gRNA-dependent U insertion editing was demonstrated to occur *in vitro* in *T.brucei* and *L.tarentolae* mitochondrial extracts (60,62). U insertion requires UTP, ATP, pre-edited mRNA, gRNA, Mg^{2+} and mitochondrial extract (60,62). Hydrolysis of the α-β bond of ATP is required for the gRNA-dependent editing reaction (60,62), as is the case for the *in vitro* gRNA-independent U insertion activity in *L.tarentolae* extracts (63). In spite of an earlier report to the contrary (66), hydrolysis at the β-γ bond of ATP does not seem to be required for U deletion editing *in vitro*. The α-β bond hydrolysis requirement is entirely consistent with an RNA ligation reaction, in which the charged intermediate has a covalently linked AMP residue, and is inconsistent with a transesterification model, in which the energy derived from breakage of the phosphodiester bond at the editing site is used to form the new bond with the incoming uridine. It should be cautioned, however, that the use of ATP analogs is only suggestive of a requirement for hydrolysis. A direct measurement of AMP production in the same stoichiometry as product formation would be required as definitive evidence.

The requirement for exogenously added UTP for *in vitro* gRNA-dependent U insertion activity, in both the *T.brucei* and the *L.tarentolae* systems, also appears to rule out models proposing that the non-templated U tails of the gRNAs are the reservoir for the Us to be inserted into the pre-edited mRNA, but does not exclude UTP acting as the nucleophile in RNA catalyzed editing, as proposed by Cech (27). However, the requirement for exogenous UTP could also reflect the normal maturation of the gRNA 3′-ends, leaving open the possibility that the inserted uridines are derived from the gRNA U tail. Definitive evidence against this has been provided by data from the *L.tarentolae in vitro* system, in which chemically blocking the 3′-end of the added gRNA by periodation did not interfere with U insertion into the pre-edited mRNA (60). In these experiments controls showed that ~70% of the 3′-block was retained after the *in vitro* reaction. This indicated that the lack of inhibition of *in vitro* U insertion was

not due to removal of the 3′-block by nuclease activity during the reaction.

In the direct U deletion assay the addition of UTP was reported to inhibit accumulation of the mRNA 5′-cleavage fragment that completely lacks the 3′-terminal Us (67). However, fragments with one, two and three 3′-Us were present, suggesting that there was an equilibrium between addition of Us to the 3′-end of the 5′-fragment and exonuclease trimming of Us from this fragment or that UTP competitively inhibits the exonuclease activity. Of course, the *in vivo* concentrations of mRNA substrate and UTP are unknown and the *in vitro* situation may not reflect the *in vivo* conditions.

Kable *et al*. (62) examined U insertion editing at A6 editing site 2 in a substrate in which the Us at site 1 were already deleted. In the absence of UTP they saw accumulation of the 5′-fragment of the mRNA cleaved at the editing site dictated by the gRNA, but no uridines were incorporated into the mRNA. Upon addition of [α-^{32}P]UTP the 5′-fragment of the mRNA became labeled at its 3′-end. This observation is inconsistent with a double transesterification, which predicts uridine incorporation at the site of nucleophilic attack (i.e. the 5′-end of the 3′-fragment).

A gRNA-dependent U insertion activity was also demonstrated using *L.tarentolae* mitochondrial extracts (60). Due to the fact that the level of this activity was even lower than that reported using the *T.brucei* extracts, an indirect primer extension PCR assay had to be used. The substrate RNA contained the 5′-domain of the ND7 pre-edited mRNA, which normally has seven editing sites, the editing of which is mediated by a single gRNA. The mRNA substrate was mutated in the anchor region and a synthetic gRNA with compensatory mutations was used, to avoid interference with the endogenous wild-type edited RNA. U insertions at site 1 were assayed by PCR, amplifying the modified substrate using a 3′-primer specific to the mutated anchor sequence and a 5′-primer specific to the upstream mRNA sequence and then performing a primer extension through site 1 with [α-^{32}P]ATP as the only nucleotide provided.

Both gRNA-independent U insertions of 1–13 Us at site 1 and gRNA-dependent U insertions were detected by this assay (60). As was the case with the Cyb mRNA substrate (65), the extent of the gRNA-independent activity was dependent on 5′ and 3′ sequences in the mRNA. Addition of cognate gRNA directed the insertion of up to 13 Us in site 1, with the predominant number being determined by the number of guiding nucleotides. Use of a gRNA with no guiding nucleotides produced a decrease in the background ladder without any enhancement of specific bands.

Addition of low concentrations of heparin selectively inhibited the gRNA-independent background ladder but had no effect on either the gRNA-guided predominant band or on the gRNA-dependent background ladder (60). This suggests that the mechanism for creation of the background ladder in the presence of added gRNA differs from the mechanism for creation of the apparently similar ladder occurring in the absence of gRNA, possibly in the affinity of protein–RNA interactions. The authors speculated that the heparin-resistant gRNA-dependent non-specific U insertion activity may represent an integral part of the mechanism of editing. They suggested that multiple Us are first added to the 3′-end of the mRNA 5′-cleavage fragment at a U insertion site in a template-independent fashion. It is possible that this also occurs at U deletion sites, but as yet there is no evidence for this. The next event would be the formation of base pairs with the guiding nucleotides in the gRNA, followed by 3′→5′ exonucleolytic trimming of the 3′-overhang, assuming the 5′-cleavage fragment remains in the editing complex. The trimming is either U-specific (67) or terminates at the adjacent upstream base pair. The gRNA essentially functions as a ‘splint’ and facilitates ligase-mediated joining of the two fragments when there is no bulge (62). The presence of a gRNA-dependent ladder of 1–13 Us is attributed to ligation occurring prior to complete trimming or after excessive trimming (60). One could argue that the gRNA-dependent ladder is an artifact of this *in vitro* system, insofar as the assay is indirect and the system is inefficient. However, direct evidence for a non-templated 3′-terminal addition of Us to the 5′-cleavage fragment was obtained in the *T.brucei* A6 mRNA *in vitro* system; 5′-cleavage fragments terminating in one to four Us were visualized even though the added gRNA only templated two U insertions (62). This suggests that the *L.tarentolae in vitro* data is meaningful.

The idea that multiple Us are added to the 5′-cleavage fragment and then the Us are trimmed back is the major conceptual modification of the original enzyme cascade model, which proposed the addition of a single U per cycle (16). It is an important concept since it ties together U deletions and U insertions into one mechanism and, in addition, predicts the production of a certain proportion of misedited sequences as a normal consequence of the editing process. The effective UTP concentration at the editing site appears critical in this model, since U insertions are completely dependent on the presence of at least 30–50 μM UTP (E.Byrne and L.Simpson, unpublished results). In the case of a U deletion site the data from Cruz-Reyes *et al*. (67) suggests that UTP may inhibit exonuclease trimming at a U deletion site. It is possible that the effective concentration of UTP at an editing site is somehow regulated, perhaps by UTP binding proteins (L.Simpson, unpublished results). Another possibility is that the 3′-exonuclease activity *in vivo* has a specificity for single-stranded 3′-oligo(U) overhangs, whereas the *in vivo* U insertion activity prefers to add Us to a nick in a duplex RNA.

The requirement for exogenous UTP, the apparent requirement for α-β bond ATP hydrolysis, the observed patterns of pre-mRNA cleavage and U insertion and the lack of inhibition of the gRNA-mediated U insertion process by chemical blockage of the 3′-end of the gRNA are all in agreement with a protein-mediated process. The chemistry of the process is in question, as the possibility of a protein-mediated transesterification mechanism, as suggested by Cech (27), is still open. In fact, the possible involvement of an RNA ligase in editing implies that at least part of the editing chemistry will include a transesterification step, since most RNA ligases studied to date proceed by a transesterification mechanism (68).

Regardless of which model proves to be the correct one, the various editing models have been pivotal in providing hypotheses for probing the mechanism of editing *in vitro*. Ultimately, it is careful biochemical studies in conjunction with genetic analysis that will conclusively answer the remaining questions of the actual chemistry of the editing process.

THE ACTIVITIES

Evidence for a protein-mediated editing mechanism, albeit circumstantial, has also been provided by the detection of various enzymatic activities in mitochondrial extracts from trypanosomatids. Initially, a TUTase (23), an endoribonuclease (25,26) and an

RNA ligase activity (23) were reported. All of these activities seem to possess some sort of specificity for pre-edited mRNA and/or gRNA and this specificity suggested some involvement of these activities in RNA editing. In fact, the presence of these activities provided the initial stimulus for the development of the enzyme cascade model of editing.

These activities have been fractionated to different extents. Initially, Pollard *et al*. (69) found that the three activities of *T.brucei* extracts as well as the mRNA and gRNA could be separated into two distinct peaks by glycerol gradient sedimentation. One peak sedimenting at 19S contained TUTase, RNA ligase and gRNAs, while mRNA and also gRNA were found in a second peak in the 35–40S region of the gradient. Since then, co-fractionating putative editing activities have been obtained in several laboratories, both with *T.brucei* and with *L.tarentolae* extracts (70–74). Peris *et al*. (75) recently reported the separation of potential editing activities from both species into two peaks; a major peak sedimenting at 10S, which contains TUTase and gRNAs, and a minor peak sedimenting at 20S, which contains the RNA ligase and some TUTase. The TUTase-containing and ligase-containing complexes could be visualized on native gels and were termed the T-IV complex and the ligase complex. In addition, a heterodisperse series of gRNA- and mRNA-containing complexes were found throughout the gradient. The main difference between the *T.brucei* and *L.tarentolae* extracts is that the 20S peak of *T.brucei* contains as much TUTase activity as the 10S peak. Both the gRNA-independent U insertion activity in *L.tarentolae* (70) and the gRNA-dependent U deletion and U insertion activities from *T.brucei* (67,71) were found to sediment at ~20S, together with the ligase complex, suggesting that this complex may represent a minimal editing enzymatic unit (76).

Although the data from various laboratories seems to differ in the number and/or size of the putative editing complexes, this might simply reflect subtle differences due to variations in the fractionation procedures and/or the species. However, given the high hydrophobicity of many mitochondrial proteins, it is probable that the observed apparent sizes of the editing complexes may be influenced by interaction with proteins that bind editing complexes non-specifically. Thus the actual size of the editing complexes under non-denaturing conditions could easily be overestimated. However, from the glycerol gradient fractionation data it is clear that RNA editing does take place within the realm of a multiprotein complex. It is expected that for complex assembly not only catalytic components of editing will be required, but also a set of structural proteins to provide a scaffold within which the actual catalysis occurs. The ultimate answer to the question of the size and/or nature of the editing complex will be through the establishment of a highly sensitive and quantitative *in vitro* editing assay, purification of the individual components and reconstitution of the active editing complex from the individual parts.

Studies on the individual components of the editing complexes are beginning to accumulate. An earlier report of a putative editing endoribonuclease failed to show a gRNA dependence of cleavage. The first experimental evidence for gRNA-dependent cleavage came in the *in vitro* system described by Seiwert *et al*. (61), Cruz-Reyes *et al*. (67) and Piller *et al*. (77) for U deletion editing. Piller *et al*. (77) were able to resolve three endoribonuclease activities from the *T.brucei* mitochondrial extract, by sensitivity to DTT, cleavage specificity and a requirement for gRNA. One activity cleaves Cyb mRNA within the pre-edited region and the specificity is unaffected by addition of cognate gRNA. A second activity is single-strand specific and cleaves Cyb mRNA in the 3′-end of the editing domain. A third activity has all the features of the predicted gRNA-dependent editing nuclease. It cleaves immediately 5′ of a duplex but is also specific for the gRNA-directed editing site; digestion yields a 5′-P and a 3′-OH, consistent with the cleavage fragments being substrates for RNA ligase. This activity also sediments at ~20S. Isolation and characterization of purified nucleases, however, is necessary to confirm these assignments of activities to different enzymes. Alfonzo and Simpson (unpublished results) have recently purified to homogeneity, cloned and expressed a 22 kDa nuclease from *L.tarentolae* extracts with some specificity towards pre-edited RNAs and which appears to be gRNA independent; this probably corresponds to the nuclease activity previously described by Simpson *et al*. (25).

Two proteins of 50 and 45 kDa (reported as 57 and 50 kDa from *T.brucei*) have been identified as putative RNA ligase intermediates, since they can be covalently charged with AMP (24,75). The involvement of a particular RNA ligase in editing will require further purification and characterization. So far, the only evidence that this RNA ligase might be involved in editing is the finding that it co-fractionates in glycerol gradients with other putative editing components.

If indeed editing happens within a ribonucleoprotein complex, the process will require structural components that will not be directly involved in catalysis. It is presumed these structural factors should have specificity and high affinity for either the mRNA and/or gRNA. It is appealing to think that some interactions between editing components will help in assembly of an active complex and also help stabilize the complex once formed. The best candidate so far for a structural editing factor is a gBP21 protein purified from *T.brucei* by Köller *et al*. (78). gBP21 was identified by UV crosslinking to gRNA. Further analysis showed that gBP21 could bind gRNAs strongly *in vitro* (K_d = 5–10 nM). Although no biological significance has been proven for the strong interaction between gRNA and gBP21, in view of the low dissociation constant it may play some role in RNA editing. However, the specific role of gBP21 in editing remains an open question.

Another activity which was predicted to explain opening of the duplex formed by the 5′-most gRNA in an editing domain and the fully edited mRNA, to allow interaction with ribosomes, is an RNA helicase. This activity could also be invoked, in addition to breathing of the less thermodynamically stable portion of the duplex (51), to explain the melting of downstream gRNA–edited mRNA duplexes to allow formation of the anchor duplex with the adjacent upstream gRNA. A putative RNA helicase of the DEAD box family was isolated from mitochondrial extracts of *T.brucei* (79,80). Again, the specific role if any of this helicase in RNA editing remains an open question.

CONCLUSIONS

The overall evidence strongly indicates that U insertion/deletion RNA editing in trypanosomatids involves RNA–protein complexes and involves a series of successive specific gRNA-directed cleavages, 3′-terminal U addition, 3′-exonuclease trimming and re-ligation, almost exactly as predicted by the original enzyme cascade model (16). The field has progressed to the stage of biochemical characterization of individual components and

reconstitution of editing activities. Major advances have been made in obtaining *in vitro* systems in which a complete cycle of U deletion and U insertion editing occurs, but the systems are extremely inefficient and do not progress past a single cycle of editing. There is clearly a need to optimize and reconstitute the *in vitro* systems with recombinant proteins and obtain progressive editing in multiple sites. Antisera against purified putative editing components should be generated and tested for immunodepletion activity and immunoprecipitation of editing complexes. Finally, gene disruption data must be obtained to confirm a role of each component in the editing process.

REFERENCES

1 Simpson,L. (1972) *Int. Rev. Cytol.*, **32**, 139–207.
2 Simpson,L. (1986) *Int. Rev. Cytol.*, **99**, 119–179.
3 Vickerman,K. (1976) In Lumsden,W.H.R. and Evans,D.A. (eds), *Biology of the Kinetoplastida*. Academic Press, London, UK, pp. 1–34.
4 Benne,R., Van den Burg,J., Brakenhoff,J., Sloof,P., Van Boom,J. and Tromp,M. (1986) *Cell*, **46**, 819–826.
5 Bass,B.L. (1993) In Gesteland,R.F. and Atkins,J.F. (eds), *The RNA World*. Cold Spring Harbor Laboratory Press, Plainview, NY, pp. 383–418.
6 Simpson,L. and Emeson,R.B. (1996) *Annu. Rev. Neurosci.*, **19**, 27–52.
7 Simpson,L. and Thiemann,O.H. (1995) *Cell*, **81**, 837–840.
8 Simpson,L., Maslov,D.A. and Blum,B. (1993) In Benne,R. (ed.), *RNA Editing—The Alteration of Protein Coding Sequences of RNA*. Ellis Horwood, New York, NY, pp. 53–85.
9 Stuart,K. (1993) In Benne,R. (ed.), *RNA Editing—The Alteration of Protein Coding Sequences of RNA*. Ellis Horwood, New York, NY, Vol. 1, pp. 25–52.
10 Hajduk,S.L., Harris,M.E. and Pollard,V.W. (1993) *FASEB J.*, **7**, 54–63.
11 Benne,R. (1994) *Eur. J. Biochem.*, **221**, 9–23.
12 Feagin,J.E., Abraham,J. and Stuart,K. (1988) *Cell*, **53**, 413–422.
13 Simpson,L. and Shaw,J. (1989) *Cell*, **57**, 355–366.
14 Feagin,J., Jasmer,D. and Stuart,K. (1987) *Cell*, **49**, 337–345.
15 Read,L.K., Stankey,K.A., Fish,W.R., Muthiani,A.M. and Stuart,K. (1994) *Mol. Biochem. Parasitol.*, **68**, 297–306.
16 Blum,B., Bakalara,N. and Simpson,L. (1990) *Cell*, **60**, 189–198.
17 Sturm,N.R. and Simpson,L. (1990) *Cell*, **61**, 879–884.
18 Pollard,V.W., Rohrer,S.P., Michelotti,E.F., Hancock,K. and Hajduk,S.L. (1990) *Cell*, **63**, 783–790.
19 Corell,R.A., Feagin,J.E., Riley,G.R., Strickland,T., Guderian,J.A., Myler,P.J. and Stuart,K. (1993) *Nucleic Acids Res.*, **21**, 4313–4320.
20 Van der Spek,H., Arts,G.-J., Zwaal,R.R., Van den Burg,J., Sloof,P. and Benne,R. (1991) *EMBO J.*, **10**, 1217–1224.
21 Yasuhira,S. and Simpson,L. (1996) *RNA*, **2**, 1153–1160.
22 Sollner-Webb,B. (1992) *Nature*, **356**, 743–744.
23 Bakalara,N., Simpson,A.M. and Simpson,L. (1989) *J. Biol. Chem.*, **264**, 18679–18686.
24 Sabatini,R. and Hajduk,S.L. (1995) *J. Biol. Chem.*, **270**, 7233–7240.
25 Simpson,A.M., Bakalara,N. and Simpson,L. (1992) *J. Biol. Chem.*, **267**, 6782–6788.
26 Harris,M., Decker,C., Sollner-Webb,B. and Hajduk,S. (1992) *Mol. Cell. Biol.*, **12**, 2591–2598.
27 Cech,T.R. (1991) *Cell*, **64**, 667–669.
28 Blum,B., Sturm,N.R., Simpson,A.M. and Simpson,L. (1991) *Cell*, **65**, 543–550.
29 Koslowsky,D.J., Goringer,H.U., Morales,T.H. and Stuart,K. (1992) *Nature*, **356**, 807–809.
30 Blum,B. and Simpson,L. (1992) *Proc. Natl. Acad. Sci. USA*, **89**, 11944–11948.
31 Harris,M.E. and Hajduk,S.L. (1992) *Cell*, **68**, 1091–1099.
32 Piller,K.J., Decker,C.J., Rusche,L.N. and Sollner-Webb,B. (1995) *Mol. Cell. Biol.*, **15**, 2925–2932.
33 Simpson,L. (1987) *Annu. Rev. Microbiol.*, **41**, 363–382.
34 Xu,C. and Ray,D.S. (1993) *Proc. Natl. Acad. Sci. USA*, **90**, 1786–1789.
35 Xu,C.W., Hines,J.C., Engel,M.L., Russell,D.G. and Ray,D.S. (1996) *Mol. Cell. Biol.*, **16**, 564–576.
36 Carpenter,L.R. and Englund,P.T. (1995) *Mol. Cell. Biol.*, **15**, 6794–6803.
37 Robinson,D.R. and Gull,K. (1994) *J. Cell Biol.*, **126**, 641–648.
38 Perez-Morga,D.L. and Englund,P.T. (1993) *Cell*, **74**, 703–711.
39 Maslov,D.A. and Simpson,L. (1994) *Mol. Cell. Biol.*, **14**, 8174–8182.
40 Lukes,J., Arts,G.J., Van den Burg,J., De Haan,A., Opperdoes,F., Sloof,P. and Benne,R. (1994) *EMBO J.*, **13**, 5086–5098.
41 Adler,B.K., Harris,M.E., Bertrand,K.I. and Hajduk,S.L. (1991) *Mol. Cell. Biol.*, **11**, 5878–5884.
42 Blum,B. and Simpson,L. (1990) *Cell*, **62**, 391–397.
43 Campbell,D.A., Spithill,T.W., Samaras,N., Simpson,A. and Simpson,L. (1989) *Mol. Biochem. Parasitol.*, **36**, 197–200.
44 Thiemann,O.H., Maslov,D.A. and Simpson,L. (1994) *EMBO J.*, **13**, 5689–5700.
45 Koslowsky,D.J., Bhat,G.J., Perrollaz,A.L., Feagin,J.E. and Stuart,K. (1990) *Cell*, **62**, 901–911.
46 Maslov,D.A., Avila,H.A., Lake,J.A. and Simpson,L. (1994) *Nature*, **365**, 345–348.
47 Landweber,L.F. (1992) *BioSystems*, **28**, 41–45.
48 Simpson,L. and Maslov,D.A. (1994) *Science*, **264**, 1870–1871.
49 Simpson,L. (1997) *Mol. Biochem. Parasitol.*, **86**, 133–141.
50 Sturm,N.R. and Simpson,L. (1991) *Nucleic Acids Res.*, **19**, 6277–6281.
51 Maslov,D.A. and Simpson,L. (1992) *Cell*, **70**, 459–467.
52 Riley,G.R., Corell,R.A. and Stuart,K. (1994) *J. Biol. Chem.*, **269**, 6101–6108.
53 Steinert,M. and Van Assel,S. (1980) *Plasmid*, **3**, 7–17.
54 Ray,D., Hines,J., Sugisaki,H. and Sheline,C. (1986) *Nucleic Acids Res.*, **14**, 7953–7965.
55 Yasuhira,S. and Simpson,L. (1995) *RNA*, **1**, 634–643.
56 Birkenmeyer,L., Sugisaki,H. and Ray,D.S. (1985) *Nucleic Acids Res.*, **13**, 7107–7118.
57 Degrave,W., Fragoso,S., Britto,C., Van Heuverswyn,H., Kidane,G., Cardoso,M., Mueller,R., Simpson,L. and Morel,C. (1988) *Mol. Biochem. Parasitol.*, **27**, 63–70.
58 Avila,H. and Simpson,L. (1995) *RNA*, **1**, 939–947.
59 Rusche,L.N., Piller,K.J. and Sollner-Webb,B. (1995) *Mol. Cell. Biol.*, **15**, 2933–2941.
60 Byrne,E.M., Connell,G.J. and Simpson,L. (1996) *EMBO J.*, **15**, 6758–6765.
61 Seiwert,S.D., Heidmann,S. and Stuart,K. (1996) *Cell*, **84**, 831–841.
62 Kable,M.L., Seiwert,S.D., Heidmann,S. and Stuart,K. (1996) *Science*, **273**, 1189–1195.
63 Frech,G.C., Bakalara,N., Simpson,L. and Simpson,A.M. (1995) *EMBO J.*, **14**, 178–187.
64 Frech,G.C. and Simpson,L. (1996) *Mol. Cell. Biol.*, **16**, 4584–4589.
65 Connell,G.J., Byrne,E.M. and Simpson,L. (1997) *J. Biol. Chem.*, **272**, 4212–4218.
66 Seiwert,S.D. and Stuart,K. (1994) *Science*, **266**, 114–117.
67 Cruz-Reyes,J. and Sollner-Webb,B. (1996) *Proc. Natl. Acad. Sci. USA*, **93**, 8901–8906.
68 Harnett,S.P., Lowe,G. and Tansley,G. (1985) *Biochemistry*, **24**, 7446–7449.
69 Pollard,V.W., Harris,M.E. and Hajduk,S.L. (1992) *EMBO J.*, **11**, 4429–4438.
70 Peris,M., Frech,G.C., Simpson,A.M., Bringaud,F., Byrne,E., Bakker,A. and Simpson,L. (1994) *EMBO J.*, **13**, 1664–1672.
71 Corell,R.A., Read,L.K., Riley,G.R., Nellissery,J.K., Allen,T.E., Kable,M.L., Wachal,M.D., Seiwert,S.D., Myler,P.J. and Stuart,K.D. (1996) *Mol. Cell. Biol.*, **16**, 1410–1418.
72 Read,L.K., Goringer,H.U. and Stuart,K. (1994) *Mol. Cell. Biol.*, **14**, 2629–2639.
73 Köller,J., Nörskau,G., Paul,A.S., Stuart,K. and Goringer,H.U. (1994) *Nucleic Acids Res.*, **22**, 1988–1995.
74 Goringer,H.U., Köller,J. and Shu,H.H. (1995) *Parasitol. Today*, **11**, 265–267.
75 Peris,M., Simpson,A.M., Grunstein,J., Liliental,J.E., Frech,G.C. and Simpson,L. (1997) *Mol. Biochem. Parasitol.*, **85**, 9–24.
76 Rusche,L.N., Cruz-Reyes,J., Piller,K.J. and Sollner-Webb,B. (1997) *EMBO J.*, **16**, 4069–4081.
77 Piller,K.J., Rusche,L.N., Cruz-Reyes,J. and Sollner-Webb,B. (1997) *RNA*, **3**, 279–290.
78 Köller,J., Müller,U.F., Schmid,B., Missel,A., Kruft,V., Stuart,K. and Göringer,H.U. (1997) *J. Biol. Chem.*, **272**, 3749–3757.
79 Missel,A. and Goringer,H.U. (1994) *Nucleic Acids Res.*, **22**, 4050–4056.
80 Missel,A., Nörskau,G., Shu,H.H. and Göringer,H.U. (1995) *Mol. Biochem. Parasitol.*, **75**, 123–126.
81 Schmidt,B., Riley,G.R., Stuart,K. and Goringer,H.U. (1995) *Nucleic Acids Res.*, **23**, 3093–3102.