RNA editing in trypanosome mitochondria: guidelines for models

Paul Sloof and Rob Benne

E. C. Slater Institute, University of Amsterdam, Academic Medical Centre, Amsterdam, The Netherlands

Received 16 March 1993

Mitochondrial RNAs in trypanosomes are post-transcriptionally altered by uridine insertion and deletion. The information for these RNA editing processes, which are essential for the production of functional messengers, is provided by small guide RNAs. This article discusses how features of partially edited RNAs, gRNAs and chimeric RNAs, in which a gRNA is covalently linked to an editing site of pre-mRNA, have been used for the construction of models.

Guide RNA; RNA editing; Trypanosome; Mitochondrion

1. INTRODUCTION

RNA editing is defined as the post-transcriptional alteration of the nucleotide sequence of protein-coding regions of transcripts [1]. A number of mechanistically different forms of RNA editing have recently been discovered in nuclear (reviewed in [2,3]), chloroplast [4], mitochondrial (reviewed in [5-9]) and viral RNAs ([10], reviewed in [11]). In kinetoplastid Protozoa (trypanosomes), RNA editing is essential for the expression of mitochondrial (mt) genes (called cryptogenes) which produce non-functional transcripts, e.g. lacking information for translational initiation and/or termination, encoding reading frame shifts or even complete nonsense. These defective transcripts are edited into functional messengers by altering their uridine sequences, either locally in a few sites or extensively in hundreds of sites throughout the entire length of the transcript (pan-editing) (reviewed in [5-7,9], see Table I). The information required for editing is provided by small, mitochondrially encoded, guide (g) RNAs of 40-70 nucleotides in length which possess a 3'-terminal oligo U extension that is not genomically encoded. The precise mechanism by which gRNAs direct the editing processes is not known. Current models hypothesize that base-pairing interactions between gRNAs and pre-mRNAs, also involving non-Watson/Crick G:U and C:A pairs, define two crucial stages in RNA editing: (i) the association of the gRNA with its target pre-mRNA via the formation of a duplex between an ‘anchor’ sequence of 4-14 nucleotides embedded in the 5’ region of a gRNA and a sequence immediately 3’ of a region to be edited in pre-mRNA, and (ii) the alteration of the U sequence of the editing region, as instructed by the remainder of the complementary part of the gRNA, the informational region.

The identification in mtRNA of chimeric molecules in which a gRNA is hooked up via its 3’ U-tail to an editing site of pre-mRNA prompted the suggestions that inserted U’s are derived from the U tail of a gRNA and that U insertion/deletion processes involve either consecutive RNA-mediated trans-esterifications or enzymatic ‘cut and paste’ reactions. This minireview summarizes the data from three trypanosome species (Trypanosoma brucei, Leishmania tarentolae and Crithidia fasciculata) that have led to these suggestions.

2. CRYPTOGENES AND gRNA GENES

In mtDNA of trypanosomes, approximately 50 maxicircles and 10,000 minicircles are interlocked into a network which inherited the name ‘kinetoplast’ from early microscopical work and is therefore referred to as k(inetoplast) DNA [12]. Within one trypanosome species each mtDNA component is homogeneous in size, but between species the size of maxicircles ranges from 16 to 36 kb and that of minicircles from 0.8 to 2.5 kb. Maxicircles are homogeneous in sequence and they are the trypanosome’s best approximation to the mtDNA of other organisms, since they contain a set of recognizable mt genes, such as genes for the two ribosomal (r) RNAs and genes for a number of subunits of respiratory chain complexes. The requirement for an RNA-editing step in the expression of many of the protein genes has, however, in some cases delayed their identification. In addition, the maxicircles encode a set of
gRNAs (see Table I). All mt transcripts undergo processing at their 3' end: gRNAs and rRNAs are provided with an oligo U tail [13,14] and mt messenger (m) RNAs have an AU extension [1,15,16].

As illustrated in Table I, in T. brucei 12 of the 17 protein encoding maxicircle genes are cryptic, with local editing in 3 RNAs (varying from 4 inserted U's in cytochrome c oxidase (cox) subunit 2 RNA to 34 U's in apocytochrome b (CYb) RNA, whereas the remainder of the RNAs are pan-edited. Some of the pan-edited RNAs encode proteins homologous to known proteins in other organisms, such as Cox3 [17], two putative subunits of NADH dehydrogenase (called ND7 and 8 [18,19]) and a small ribosomal subunit protein (RPS12 [20]), whereas the homology of MURF4 to the ATPase subunit 6 sequence is controversial [7,21]. The proteins encoded by other pan-edited RNAs have not been identified as yet, and the corresponding DNA sections are being referred to as G-regions to indicate the fact that their primary transcripts are extremely G-rich (G2-5, Table I [6]). G-Regions also occur in maxicircles of L. tarentolae and C. fasciculata, albeit in reduced numbers, since the Cox3 and ND7 genes are much less cryptic in these species and only limited editing is required to transform the precursor RNAs into functional messengers (e.g. 27 inserted U's in C. fasciculata ND7 instead of the 553 inserted and 89 deleted U's in T. brucei ND7).

Mitochondrial transcripts in these trypanosomes mostly display limited editing; pan-editing e.g. in cultured L. tarentolae is confined to the MURF 4 and RPS12 transcripts. Transcripts of the other G-rich regions (G1-5) are most likely not edited and therefore not expressed ([7,22], Arts et al. unpublished observations). The abundant editing in T. brucei mitochondria requires a huge collection of gRNAs, the current estimate being approximately 240 [6]. The bulk of the gRNAs appears to be encoded in minicircle DNA and, although only a limited number of minicircle gRNA genes has been identified thus far, the fact that T. brucei possesses approximately 400 different minicircle sequence classes that could encode three or four gRNAs each [6,23], seems to guarantee sufficient coding capacity. So far, only three gRNA genes have been proposed to reside in T. brucei maxicircle DNA [24]. Since editing in L. tarentolae and C. fasciculata appears to be much less extensive, the number of gRNAs required would be much lower. The experimental results seem to confirm this. An extensive search in L. tarentolae has resulted in the identification of 9 maxicircle- and 17 minicircle-encoded gRNAs, all but three of which can be assigned to known edited transcripts [22] (see Table I). These 23 gRNAs provide the complete information for all the known edited transcripts. It has therefore been proposed that the loss of the potential to edit most of the G-rich RNAs in L. tarentolae is due to the loss of minicircles encoding the necessary gRNAs during years of cultivation in rich media, the corresponding proteins obviously not being required under such conditions [22]. A similar loss of gRNA genes in C. fasciculata may explain the lack of editing of the G2 transcript and insufficient editing in

<table>
<thead>
<tr>
<th>RNA</th>
<th>T. brucei</th>
<th>L. tarentolae</th>
<th>C. fasciculata</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>+</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>cox1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>cox2</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>cox3</td>
<td>547</td>
<td>41</td>
<td>29</td>
</tr>
<tr>
<td>MURF1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>MURF2</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>MURF4</td>
<td>26</td>
<td>4</td>
<td>28</td>
</tr>
<tr>
<td>CYb</td>
<td>34</td>
<td>0</td>
<td>39</td>
</tr>
<tr>
<td>ND1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>ND4</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>ND5</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>ND7</td>
<td>553</td>
<td>89</td>
<td>25</td>
</tr>
<tr>
<td>ND8</td>
<td>259</td>
<td>46</td>
<td>0</td>
</tr>
<tr>
<td>RPS12</td>
<td>132</td>
<td>28</td>
<td>117</td>
</tr>
<tr>
<td>G2</td>
<td>345</td>
<td>20</td>
<td>0</td>
</tr>
<tr>
<td>G3</td>
<td>148</td>
<td>13</td>
<td>0</td>
</tr>
<tr>
<td>G4</td>
<td>325</td>
<td>40</td>
<td>0</td>
</tr>
<tr>
<td>G5</td>
<td>210</td>
<td>13</td>
<td>0</td>
</tr>
</tbody>
</table>

The extent of RNA editing is indicated by the number of inserted (+) and deleted (-) uridines into/from transcripts of maxicircle protein genes: cox, cytochrome oxidase; MURF, maxicircle unassigned reading frame; CYb, apocytochrome b; ND, NADH dehydrogenase; RPS, protein of the small ribosomal subunit. Some mt transcripts have not been investigated (n.i.). A complete picture of editing is available for L. tarentolae [22] and therefore the number of gRNAs and the location of the gRNA genes required for editing of a specific transcript have been summarized.
ND8 RNA (Table I; Arts et al. unpublished observations).

3. THE SEQUENCE OF EVENTS IN RNA EDITING

3.1. Initiation

The potential of gRNAs to anneal with edited mRNA regions constitutes the heart of all models for RNA editing, including the one pictured in Fig. 1. This complementarity seems to be crucial, since all sequence differences between the edited regions of RNAs from *L. tarentolae* and *C. fasciculata* are accompanied by compensatory base changes in the corresponding gRNA genes, in such a way that the capacity to form gRNA-mRNA hybrids is fully maintained in both organisms [24]. All gRNAs possess an anchor sequence which can basepair with a target sequence 3' of an editing region. The proposed role of these interactions in the initial phases of editing is supported by recent studies using in vitro systems in which it was shown that anchor sequences are essential for the formation of chimeric molecules [25]. The strength of the anchor duplex, however, is clearly not the sole determinant of the efficiency with which a region is edited, since relatively poor anchors are found for editing regions that are efficiently edited (e.g. cox2, [1,24]) and vice versa [26]. Other features, such as gRNA concentration and auxiliary factors, may help to determine the efficiency of editing, but see below.

3.2. Polarity of pan-editing

For pan-edited RNAs, the characteristics of partially edited transcripts, in which edited 3' regions are combined with unedited 5' segments, suggest an overall 3'-to-5' polarity of editing [17,27]. For some transcripts, however, editing can be initiated independently in different regions, e.g. RPS12 in *L. tarentolae* has three of these so-called editing 'domains' [28] and *T. brucei* ND7 has two [18]. For pan-editing of the MURF4 and RPS12 transcripts of *L. tarentolae*, complete sets of gRNAs have been identified [22] (see Table I), which showed that only the first gRNA (=most 3' acting) can basepair with unedited pre-mRNA. The anchors of all
the other gRNAs are created during editing by their upstream neighbour, in line with the predicted 3'-to-5' direction of editing. Extensive analysis of gRNA:mRNA duplexes for pan-edited RNAs seems to indicate that that non-Watson/Crick basepairs are much less frequent in the anchor duplexes than in the remainder of the basepaired region [22]. It has been postulated that this unequal distribution allows the downstream part of the duplex to breathe, facilitating the formation of the anchor duplex with the next incoming gRNA which, with fewer G:U pairs, has a higher thermodynamic stability.

In C. fasciculata, editing of the 5' and frameshift (FS) regions of the ND7 transcript is tightly coupled. These editing regions are 172 nucleotides apart, each being edited by one gRNA (Fig. 2). They are edited with the same efficiency (around 50%), although the 5' gRNA is present in mtRNA in an 8-fold higher concentration than FS gRNA, and the anchor duplex stability at the 5' region is considerably higher than that at the FS region (−14.3 vs. −9.8 kcal/mol; see Fig. 2 and [26]). Furthermore, all 16 cDNAs and chimeric molecules investigated in which the 5' region has been or is being edited show complete editing of the FS region [26,29]. These observations imply that the 5' region is not edited independently of the FS region, even though the target sequence required for the formation of the 5' anchor duplex is already present in the pre-mRNA. Although other explanations are feasible, these data are in line with an overall 3'-to-5' polarity of editing also for the ND7 transcript, and suggest some sort of tracking mechanism in which the editing machinery moves from 3'-to-5' over the pre-mRNA, thereby serving the FS region ahead of the 5' region. If so, the two editing regions of this RNA cannot be considered as independent editing domains in spite of the relatively long distance that separates them.

3.3. Polarity of editing with a single gRNA

Although there is general agreement on the overall 3'-to-5' polarity of pan-editing, the polarity of editing within the region of action of a single gRNA is not completely clear. Partially edited CYb and cox3 transcripts of L. tarentolae give the impression that the editing machinery moves from 3'-to-5', inserting one U at a time in one editing site at a time [30]. It is on this type of molecules that the model in Fig. 1 is based. In theory, however, they could also be created by editing under the direction of a nested set of truncated versions of a gRNA [26], or by gRNAs with an insufficient number of U's in the tail [31]. Truncated gRNAs are abundantly present in mtRNA from C. fasciculata [26]. It is unknown how they are generated, but their presence in chimeric molecules may indicate that they do indeed participate in the editing process [26,31–33]. At junctions of edited and unedited regions in other partially edited transcripts, 'unexpected' sequences are found which at first sight cannot be produced by a strictly processive type of mechanism. For L. tarentolae the creation of unexpected sequences has been explained by assuming that unrelated, or even cognate gRNAs, can form 'false' anchors, followed by 'misediting via misguiding' with the same strict 3'-to-5' polarity [34]. Indeed, in some cases unexpected sequences are found that are complementary to (parts of) an unrelated gRNA, and chimeric molecules between non-cognate partners have been identified. Another possible model, however, infers that abnormal sequences are created by random U insertion and deletions in pre-mRNA. In such a scheme the correctly edited sequence, whenever it arises by coincidence, is shielded from further editing by duplex formation with the gRNA ('match protection' [35]). Alternatively, a gRNA-directed 'progressive re-alignment' mechanism has been proposed in which a region may undergo various rounds of editing producing unexpected sequences in the process until the gRNA:mRNA duplex has attained the highest possible thermodynamic stability [36]. The basic differences between the latter two models and the one of Fig. 1 are that (i) editing does not have a strict 3'-to-5' polarity, (ii) it may start anywhere in an editing region at more than one site at a time, and (iii) unexpected sequences are not the result of false editing but rather represent bona fide intermediates of the editing process.

In resolving the matter of the polarity of editing in the region of one gRNA, the chimeric molecules provide little additional insight. In some of them, the 3' end of the gRNA is hooked up to the first editing site of a region next to the anchor duplex (site 1 in Fig. 1). These molecules could be considered as the result of the first step in editing with a 3'-to-5' polarity (see next section). In those molecules, however, that could represent later stages of the process, i.e. in which the gRNA is linked to a more downstream editing site (sites 2 or 3 in Fig. 1), upstream sites are not always fully edited. In some chimeric molecules unexpected patterns of editing are found [6,31], in others upstream sites are not edited at all [25,37]. In the latter category of chimeras, which are the predominant product of in vitro systems for chimeric formation, the distance between the linkage point and the anchor duplex can be considerable, i.e. up to 50 nucleotides. Without reliable in vitro editing systems that can establish possible precursor–product relationships, it is therefore difficult to decide conclusively between the different scenarios for editing in the region of a single gRNA. In spite of the attractive simplicity and order of the 3'-to-5' models, for the time being it is unclear how the different types of chimeric molecules and partially edited pre-mRNAs are exactly created, which of them are functional intermediates of the editing process, and which are dead-end side products.

3.4. Termination

Another 'white area' in the scheme presented in Fig.
1 is the sequence of events that lead to termination of editing and dissociation of the duplexes that are created by it. Since a number of gRNAs contain a segment of unique sequence 3' of the informational region, in theory, editing of the pre-mRNA could proceed beyond the editing domain in those cases in which the area of complementarity can be extended by U insertion and/or deletion. However, extended editing patterns have not been observed in the corresponding mRNAs, and thus a mechanism for termination of editing at specific sites must exist. It could be envisaged that primary and/or secondary structure motifs are involved, but these are not immediately obvious from inspection of the available sequences [24]. Also the mechanism by which the dissociation of gRNA : mRNA duplexes takes place is unknown. It has been speculated that the translational machinery is capable of doing that, either by the impact of the moving ribosome or with the aid of specific unwindases comparable to eukaryotic initiation factor eIF-4A [38]. However, such unwindases have not been identified so far.

3.5. The mechanism of editing

What, then, is the mechanism of RNA editing? Two different models have been proposed. In one of them, U's are inserted or deleted by two consecutive RNA-mediated trans-esterification reactions, reminiscent of self-splicing [33,39]. In the first step, a chimeric molecule is created by an attack of the 3'-OH of the gRNA at the phosphodiester bond in pre-mRNA, 3' of the first mismatching residue. Basepairing between one or more U's of the U tail with the informational part of the gRNA extends the duplexed region (step 1, Fig. 1). In the next step, the 3'-OH of the 5' moiety of the pre-mRNA initiates the second trans-esterification, which results in a partially edited molecule (step 2, Fig. 2). More cycles of this type of reaction are needed for the production of a fully edited mRNA. Although this model has the interesting implication that RNA editing and splicing are mechanistically similar and may have the same evolutionary background, the fact that the trans-esterification reactions that insert U's are targeted at another phosphodiester bond than those that delete U's, poses a conceptual problem [39]. In a refinement of this model, it was therefore proposed that deletion and insertion-type trans-esterifications are aimed at the same phosphodiester bond (the one 3' of the mismatch), but that the U deletion processes involve the action of an U-exonuclease to remove U's from the 3' end of the 5' moiety of the pre-mRNA following the first trans-esterification step [33]. In both papers [33,39] it was speculated that terminal uridylyl transferase (TUTase) could function to (re)load the tail of the gRNAs with U's. In the second model an all-enzyme scenario was proposed [9,33], in which an endonuclease and an RNA ligase are involved in two consecutive cleavage/ligation reactions together with the TUTase and the U-exonuclease to add or to remove U's. This scenario yields the same type of molecules as those produced by the trans-esterifications discussed above. A number of observations argue in favour of an enzymatic cut and paste pathway of editing. First, in contrast to splicing in which the 'incoming' nucleotide is of preferred identity (a G in Group I splicing and an A in Group II; reviewed in [40]), the identity of the last nucleotide of the gRNA is irrelevant for chimer formation. Chimeric molecule formation in vitro requires a free OH at the 3' end of the gRNA, but is unaffected by the addition of a C [37,41]. Chimeric molecules present in mtRNA from C. fasciculata and those produced in vitro in T. brucei frequently lack U residues in the connecting sequence [6,37]. Second, most of the enzymes mentioned above have been found in mitochondrial extracts of trypanosomes [42-44]. Furthermore, RNA ligase and TUTase have been shown to co-sediment with gRNAs and pre-mRNAs during glycerol gradient centrifugation of mt lysates in high-molecular-weight particles (35-40 S), although most of the endonuclease activity was found in lighter fractions [45]. Finally, chimeric molecules can be made in vitro with heterologous enzymes, such as mung bean nuclease and RNA ligase [46]. In a number of chimeras present in our ND7 collection from C. fasciculata the gRNA appeared to be hooked up to the 5' end of pre-mRNA, most likely as a result of end-to-end ligation [26].

4. PERSPECTIVES

After a period of characterizing the end products of RNA editing and making a catalogue of the U sequence alterations in mt transcripts, research in this field has moved to a stage in which the features of partially edited RNAs, gRNAs and chimeric molecules are being analysed. The expectation that their structures would provide...
insight into the molecular interactions involved in RNA editing has, indeed, partially been fulfilled. Remaining problems concern the fact that the functional relevance in the editing process of many of these molecules cannot be assessed without an efficient in vitro editing system. That means that clarification of certain mechanistic aspects of RNA editing, such as its polarity and the precise nature of the molecular reactions involved in the breaking and making of phosphodiester bonds, must await further developments. Fortunately, chimer formation can be studied in crude mt extracts ([25,37,41], own unpublished observations). The strategy for the immediate future adopted by a number of groups is therefore to make further use of the in vitro chimer formation assay to identify and to characterize other components of the editing machinery ([6,7,45], our own unpublished results). Although at present this seems remote, one day a complete in vitro editing system may await further developments. Fortunately, chimer formation can be studied in crude mt extracts ([25,37,41], own unpublished observations).

Acknowledgements. We thank the members of our RNA editing group for their enthusiastic participation and valuable contributions. Special thanks are due to D. Speijer, P. Leegwater and G.J. Arts for sharing unpublished data, and to W. van Noppen and D. Zonneveld for expert help in the preparation of the manuscript. We apologize for the biased selection of references, owing to space limitations. The research is supported by the Netherlands Foundation for Chemical Research (SON), which is subsidized by the Netherlands Foundation for Scientific Research (NWO).

REFERENCES