RNA

Sequence bias in edited kinetoplastid RNAs

M. L. Burgess and K. Stuart

RNA 2000 6: 1492-1497

References

Article cited in:

<http://www.rnajournal.org/cgi/content/abstract/6/11/1492#otherarticles>

Email alerting service

Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#)

Notes

To subscribe to *RNA* go to:
<http://www.rnajournal.org/subscriptions/>

REPORT

Sequence bias in edited kinetoplastid RNAs

MOFFETT L.K. BURGESS^{1,2} and KENNETH STUART^{1,2}

¹Seattle Biomedical Research Institute, Seattle, Washington 98109, USA

²Department of Pathobiology, University of Washington, Seattle, Washington 98195, USA

ABSTRACT

Uridylate residues (Us) are inserted and deleted at precise positions in mitochondrial transcripts of *Trypanosoma brucei*. These sequence changes are determined by interactions with small guide RNAs (gRNAs) that are complementary to edited sequence. Adenylate (A) and guanylate (G) residues in gRNAs across from editing sites pair with inserted Us. We evaluate whether sequence bias exists in the bases surrounding insertion sites. Upon analyzing all reported insertion sites in *T. brucei*, we find that the predicted base pairs flanking insertion sites show a strong bias. Specifically, guiding As and Gs tend to be flanked by cytosine residues and Us. This bias is expected if precise base-pair interactions at the editing site determine the number of inserted Us.

Keywords: kinetoplastids; RNA editing; sequence bias; Trypanosomes

INTRODUCTION

The single-celled flagellated organisms in the order Kinetoplastidae posttranscriptionally remodel the sequences of many of their mitochondrially encoded transcripts by the precise insertion and deletion of uridylates (Us). These sequence changes are directed by small *trans*-acting guide RNAs (gRNAs) that are complementary (by both Watson–Crick and G:U base pairing) to edited sequences (Blum et al., 1990; for recent reviews, see Alfonzo et al., 1997; Kable et al., 1997; Sloof & Benne, 1997). The insertion and deletion of Us in kinetoplastid RNAs (kRNAs) creates the functional open reading frames as is evident from the homology of the proteins predicted from translation of edited mRNAs to proteins encoded in the mitochondria of numerous other organisms. In the kinetoplastid parasite, *Trypanosoma brucei*, 12 of the 18 mitochondrially encoded pre-mRNAs are edited to varying degrees, many extensively, resulting in a total of 1,362 editing sites (Arts & Benne, 1996, and references therein). This quantity of editing is predicted to require hundreds of gRNAs, each specifying the editing of 4 to 10 sites and resulting in complementarity to an ~45-nt stretch of edited RNA. Although only a small fraction of gRNAs have been sequenced, renaturation kinetic analyses of the coding capacity of the minicircle component of mitochondrial

DNA of *T. brucei* predicts more than enough capacity for the number of different gRNAs required to specify all the observed editing (Stuart, 1993).

The mechanism by which gRNAs direct the insertion and deletion of Us follows a cleavage/U addition or removal/ligation pathway initially proposed by Blum et al. (1990). gRNAs appear to select the site to be processed by forming a short duplex adjacent to the editing site. Editing is initiated by endonucleolytic cleavage of the pre-mRNA 3' (with respect to gRNA) to the terminus of this duplex, generating upstream and downstream pre-mRNA cleavage fragments. For clarity, the positions that are 5' and 3', with respect to pre-mRNA, of the editing site will be referred to as upstream and downstream, respectively. After cleavage, the 3' terminus of the upstream fragment is processed by U addition or removal during insertion and deletion editing, respectively, and then the fragments are rejoined. Each round of editing at a site *in vitro* is specified by the gRNA sequence opposing it (Cruz-Reyes & Sollner-Webb, 1996; Kable et al., 1996; Seiwert et al., 1996). During U deletion, the cleavage occurs downstream of the Us to be deleted (Cruz-Reyes & Sollner-Webb, 1996; Seiwert et al., 1996), and the Us are removed by a U-specific 3' exonuclease until the first non-U nucleotide is encountered (S. Lawson, unpubl. data). Thus, the site of gRNA-directed cleavage and the specificity of the exonuclease can account for the number of Us deleted. How the gRNA directs the addition of the appropriate number of Us during insertion editing appears more complex. Insertion sites are edited by

Reprint requests to: Kenneth Stuart, Department of Pathobiology, University of Washington, Seattle, Washington 98195, USA; e-mail: kstuart@u.washington.edu.

insertion of anywhere from 1 to 13 Us (Benne, 1994). Accuracy in the number of Us inserted is expected to result from specificity of the U addition activity, interaction of added Us with the As and Gs in the gRNA, and selectivity of the RNA ligase (Burgess et al., 1999; Estevez & Simpson, 1999). How the appropriate register between gRNA and mRNA is achieved is not known. To explore how gRNAs interact with pre-mRNAs to generate RNA that is accurately edited by U insertion, we analyzed the composition of base pairs that define the boundaries of the editing site: those flanking inserted Us. If specificity in the number of Us inserted at each site is guided by the number of As and Gs opposing the site, the gRNA sequence will tend to be constrained to the number of As and Gs equal to the number of Us inserted at that site. Thus, guiding ribonucleotides will tend to be flanked by ribonucleotides that do not pair with U, resulting in a bias in the nucleotides between which Us are inserted. The strength of this bias likely reflects the importance of base-pairing interactions during the creation of accurately edited RNA.

RESULTS AND DISCUSSION

To evaluate sequence bias at insertion sites, we tabulated the frequencies at which each ribonucleotide occurs at the border of an editing site based on the known sequences of *T. brucei* pre-edited and edited RNAs. We examined the 1,220 positions at which 3,037 Us are inserted. Of the 2,453 internucleotide positions in edited domains of RNA, 49.7% are insertion sites. Therefore, if there were no ribonucleotide bias between which Us are inserted, one-half of all the As, Cs, Gs, and Us encoded in editing domains would be found in each context. As detailed below, this is not the case.

Nomenclature

The ribonucleotide in the pre-mRNA immediately upstream of an editing site is referred to as X_m , and the ribonucleotide downstream of an editing site as Z_m . The ribonucleotides in the gRNA that would be predicted to pair with X_m and Z_m in edited RNA are designated X_g and Z_g , respectively. The designation of R_g is given for guiding ribonucleotides (i.e., Gs and As in the gRNA that pair with inserted Us). The nomenclature for description of insertion sites is summarized in Figure 1.

Position $Z_m:Z_g$

Analysis of ribonucleotides downstream of insertion sites indicates that this position is heavily biased against C and U. Specifically, although 57.8% of As and 60.1% of Gs are downstream of inserted Us, only 17.2% of Cs

FIGURE 1. Diagram of an insertion site. A region of an mRNA into which one U has been inserted (top) is shown paired to the gRNA that directs this insertion (bottom). The ribonucleotides upstream and downstream of the inserted U are represented as X_m and Z_m , respectively. The inserted U pairs with a guiding ribonucleotide, which is universally an A or a G, and here is represented by the symbol R_g . The Z_g in the gRNA pairs with Z_m , and X_g pairs with X_m .

and 16.8% of encoded Us are at this position (Fig. 2A). Because the gRNAs and edited sequence are complementary, this indicates a bias against As and Gs at position Z_g in the gRNA, and suggests that there may have been selection against ribonucleotides that have the capacity to pair with U in the position Z_g of gRNA, perhaps because they may interfere with accurate processing of the upstream pre-mRNA fragment. The base pair formed between Z_g and Z_m is at the terminus of the duplex between gRNA and pre-mRNA before editing at that site. Breathing at the end of the duplex between the gRNA and downstream pre-mRNA fragment may allow any added Us to pair interact with Z_g and adversely affect the accuracy of U insertion during editing of that site.

Position $X_m:X_g$

Us do not occur at position X_m in reported edited sequence, because editing is expected to occur upstream of encoded Us. Evidence from studies of in vitro editing support this prediction, as As and Gs specify cleavage and editing 5' of encoded Us (Seiwert et al., 1996; Cruz-Reyes et al., 1998). Thus, position X_m is effectively restricted to A, C, and G. Of these three ribonucleotides, as shown in Figure 2B, only Cs are uncommon at this position. Strikingly, although 62% of Gs, and 64% of As in edited regions are at position X_m , only 5% of Cs are at this position. Because edited sequences of mRNA are complementary to the guiding regions of gRNAs, a C at position X_m indicates a G at position X_g . Position X_g is adjacent to the guiding ribonucleotides for that site and a G at position X_g may be difficult to distinguish from a guiding purine. Studies of in vitro editing indicate that additional As and Gs across from the editing site specify the insertion of additional Us (Byrne et al., 1996; Kable et al., 1996). Thus, Gs at X_g may affect the specificity of the number of Us inserted during editing, and therefore have been selected against. It is also possible that a C at X_m could interfere with endonucleolytic cleavage if the 5'-most guiding ribonucleotide is a G, because the duplex would be extended by pairing

FIGURE 2. Percent of each ribonucleotide flanking of an inserted U. **A:** The percent of each ribonucleotide at position Z_m is shown in the bar graph and in the bottom row of the table. These values were calculated from the number of times A, G, C, and U are each found downstream of an inserted U (first row of the table) and the number of each encoded ribonucleotide in edited domains (second row of the table). Numbers in parentheses were calculated assuming that Us are inserted downstream instead of upstream of encoded Us and are represented by the symbol (—●) in the bar graph. **B:** The same analysis performed for ribonucleotides at position X_m . A comparison of population proportions for data on both graphs indicates that the percentages for each ribonucleotide are significantly different than the percentage expected if editing sites were randomly distributed (all p -values < 0.0001).

at this position. However, this would not happen if the guiding ribonucleotide was an A, making it unlikely that this could account for the entire strength of the bias, particularly considering that nearly three-quarters of guiding ribonucleotides are As. We examined the context of the rare (12) editing sites in *T. brucei* at which Us are inserted downstream of a C and found no obvious sequence bias surrounding the editing site. The three gRNAs identified in *T. brucei* that appear to direct insertion of a U downstream of a C [gCR1[164], gCR1[219], gCOIII[311] (see Simpson et al., 1998)] show no obvious features, such as alternatively base-paired anchors, that differentiate them from other insertion sites.

Nucleotide composition in edited regions

The encoded sequence of edited RNAs contains 45.0% G, 32.5% A, 12.8% U, and 9.7% C. The high frequency of Gs (and a corresponding lack of Cs in the coding

strand) was one means by which the coding sequences for extensively edited RNAs were identified (Bhat et al., 1990; Maslov et al., 1992; Souza et al., 1992), although the underlying reason for this bias was unknown. An obvious possible basis for the low content of encoded Us is that this content increases to nearly 60% upon editing of the RNA. Selection against encoded Us adjacent to insertion sites could also contribute to the bias against Us. Selection against Cs at insertion sites would explain their paucity in pre-edited RNAs. Thus, selection acting on nucleotide content at U insertion sites may, to some degree, provide a basis for the overall nucleotide bias that has long been observed in regions encoding mRNAs that are extensively edited. The bias against insertion between pyrimidines is even more striking if calculated without correcting for nucleotide composition. The raw statistics show that only 3.4% and 4.3% of inserted Us are upstream of Cs and Us, respectively. Less than 1% of Us are inserted downstream of a C. Thus, editing sites at which the number

of purines in the gRNA across from an editing site exceeds the number of inserted Us are rare.

Ribonucleotides adjacent to deletion sites

The profile of ribonucleotides adjacent to deletion sites is unlike what is seen at insertion sites. As a point of reference, the baseline percentage of ribonucleotides adjacent to deletion sites is lower because there are many fewer deletion sites than insertion sites in edited regions of RNA: 5.79% of ribonucleotides in edited regions are located adjacent to a deleted U. The proportions of As, Gs, and Cs upstream of inserted Us vary less than at insertion sites, and are not significantly different from random (see Fig. 3B). Us adjacent to deletion sites, consistent with the current mechanistic model of U deletion (Hajduk et al., 1997; Stuart et al., 1997; Estevez & Simpson, 1999), are considered to be downstream of deleted Us. This is supported by stud-

ies of in vitro U deletion, which indicate that all Us 5' to the position of cleavage at the editing site are removed down to the first non-U ribonucleotide, and thus, that nondeleted Us are not upstream of deleted Us (Cruz-Reyes & Solner-Webb, 1996; Seiwert et al., 1996; Lawson et al., 2000).

The relatively high proportion of Cs upstream of deletion sites contrasts with the strong bias against C upstream of insertion sites. The finding that Cs are well represented in this position at deletion sites is surprising. This context was suggested to pose a particular difficulty because the complementary G in the gRNA would be predicted to pair with the U to be deleted (Feagin, 1990; Sturm et al., 1992). This was termed the "CU paradox" (Sturm et al., 1992), and was suggested to possibly be a contributing factor to misediting at deletion sites. However, the relatively favorable probability that a C in an edited region will be found in this context suggests that it is not selected against, and raises the pos-

FIGURE 3. Percent of each ribonucleotide flanking a deleted U. **A:** The percent of each ribonucleotide downstream of a deleted U is shown in the bar graph and in the bottom row of the table. These values were calculated from the number of times A, G, C, and U are each found downstream of an inserted U (first row of the table) and the number of each encoded ribonucleotide in edited domains. The proportion of Gs and Cs in this position is not statistically different than expected if deletion site placement was random (p -values of 0.263 and 0.955). The lack of statistical significance is indicated in the chart by the symbol ([†]). The proportion of As is lower and the proportion of Us is higher in this position (both p -values <0.0001). **B:** The same analysis performed for ribonucleotides upstream of deleted Us. For this analysis, encoded Us are considered never to be upstream of deleted Us. None of the percentages are statistically different from random at the level of 95% confidence: p -values for A, G, and C are 0.207, 0.052, and 0.128, respectively.

sibility that the extension of the anchor duplex by a G:U pair does not negatively influence U deletion at that site.

A relatively high proportion of encoded Us in edited regions are located downstream of deletion sites (Fig. 3A). The fraction of deletion sites with a G or a C downstream is nearly what would be expected by chance. As are present in this position slightly less often than would be expected by a random distribution. This is unlike the distribution of ribonucleotides downstream of insertion sites, at which As and Gs predominate. Both upstream and downstream of insertion sites, the ribonucleotides that are found more often than would be expected by chance are those that imply complementary gRNA bases that do not pair with U. The profile of ribonucleotides adjacent to deletion sites indicates no such bias.

Accuracy of Insertion

A context in which the guiding purine ribonucleotides are flanked by pyrimidine ribonucleotides (which cannot pair with inserted Us) predominates at insertion sites. Consequently, Cs and Us that are encoded in the pre-mRNA are more likely to be adjacent to deletion sites or sites where editing does not occur. These biases may relate to the manner in which accuracy is obtained during U insertion. Purines in the gRNA that are adjacent to the purines that guide U insertion could affect the specificity of U addition and/or ligation. Pairing of added Us with these flanking purines at position X_g or Z_g (see Fig. 4A) may result in inaccuracy in the insertion of Us into the edited product. The strengths of the biases upstream and downstream of inserted Us are also correlated with the locations of the editing sites. The bias is weaker immediately downstream of the editing sites, which is predicted to be at the terminus of a duplex, and thus less available to pair with added Us. A similar bias against purines is not observed at positions adjacent to U deletion sites (Fig. 3), supporting the possibility that the bias reflects selection for sequences that avoid pairing with added Us.

The appropriate register between the added Us and guiding ribonucleotides is presumably necessary to achieve the accurate editing observed during *in vitro* U insertion. If X_g is not aligned with X_m during U insertion, too few or too many Us could pair with guiding ribonucleotides and promote inaccurate insertion (Fig. 3B). X_m may be aligned with X_g after cleavage because it is the next complementary ribonucleotide upstream of the editing site (Fig. 4C). Restriction of position X_m to A or G would ensure that this base pair would never be with a guiding ribonucleotide. This single base pair would not be thermodynamically stable and thus, one function of the multiprotein complex that catalyzes editing may be to stably align the terminal nucleotide of the upstream cleavage fragment with the first complementary ribonucleotide in the gRNA. The base pair upstream of the editing site is predicted to be an mA/gU

FIGURE 4. Potential base pairing interactions at an insertion site. **A:** An added U can pair with the guiding ribonucleotide at position R_g (bold arrow), or could pair with positions X_g or Z_g if they are As or Gs (smaller arrows). **B:** A site where four guiding As specify the insertion of four Us is depicted. If the correct register between the gRNA and mRNA is not achieved, fewer (top), or more (bottom) Us than specified could be inserted. **C:** Upon accurate editing, X_m and Z_m are paired, respectively, with X_g and Z_g .

or mG/gCorU at nearly 99% of insertion sites, indicating that a preference for alignment of these specific (and specifically oriented) base pairs may exist. It is possible that, similar to template-driven polymerization, Us would not be added unless the substrate and template were aligned by a base pair. Proper alignment of the upstream cleavage fragment with the gRNA may be one means by which accuracy of U insertion is achieved. Misedited sites, if not corrected before translation, would create frameshifts in the mRNA and hence not encode functional proteins.

From the available database of *Leishmania tarentolae* and *T. brucei* gRNA alignments with edited sequence, three sites at which insertion 3' of C appears to be guided by a G have been identified. In these three cases, a strictly processive model for editing cannot account for the observed editing, because the G in the gRNA would be predicted to base pair with the C upstream of the editing site. Similarly, indeed more frequently, deletion of a U does occur downstream of a C, although the G at position X_g is predicted to pair with the U to be deleted. This implies that editing sites may not be selected solely because they are at the terminus of the mRNA:gRNA duplex and hence that the accu-

racy of the final edited sequence may involve more than just extension of the duplex between mRNA and gRNA during each round of editing. It is possible that base modification within mRNA and/or gRNA occurs at specific sites, creating base pairing that allows proper alignment for editing. Base change by C to U modification has been observed in the anticodon of a mitochondrially imported tRNA in the kinetoplastid species, *L. tarentolae*, suggesting that the machinery to perform selective base editing exists in kinetoplastids (Alfonzo et al., 1999).

The high density of editing sites and the apparent lack of signal sequences in kRNAs to demarcate editing sites has tended to suggest that there is nothing to be learned by close analysis of edited sequence. However, statistical analysis indicates that the sequence contexts at which Us are inserted are far from random, and that the base pairs flanking editing sites are biased in composition and orientation, such that the vast majority of guiding ribonucleotides are flanked by Cs and Us. This ensures that the number of Gs and As across from an editing site matches the number of inserted Us, a property that is likely to be important for accurate U insertion.

MATERIALS AND METHODS

The reported consensus sequence for all RNAs edited in *T. brucei* was used to analyze the ribonucleotides between which Us are inserted and deleted. The consensus sequence used matched the edited sequence in the Edited Sequence Database (Simpson et al., 1998) with the exception of the very 3' terminus of cytochrome oxidase III, which included two editing sites reported in Feagin et al. (1988). The number of times each ribonucleotide occurs immediately 5' and 3' of inserted and deleted Us was scored for each editing site. An edited domain is defined as the region in pre-edited RNA beginning with the ribonucleotide immediately downstream of the first editing site through the ribonucleotide upstream of the last editing site.

The raw frequencies of ribonucleotides adjacent to inserted and deleted Us reflect a bias in overall nucleotide composition, and not necessarily positional bias relative to insertion sites because of the predominance of encoded As and Gs in edited regions. Therefore, the frequencies were calculated relative to the number of each encoded ribonucleotide in edited domains. Data are expressed as the proportion of a particular ribonucleotide that occurs adjacent to an inserted or deleted U. A comparison of population proportions was made, comparing each of the actual proportions of ribonucleotides in each position against the proportion in each position if the distribution of editing sites was random. This analysis was used to generate a z-score which was converted to a p-value based on a two-tailed distribution.

ACKNOWLEDGMENTS

We thank members of the Stuart laboratory for helpful discussions. We also thank Peter Myler for compiling information of published sequences and Larry Simpson and the Simpson laboratory for providing the extensive sequence information in the Insertion/Deletion Edited Sequence Data-

base. This work was supported by NIH grants GM42188 and AI14102.

Received January 7, 2000; returned for revision February 15, 2000; revised manuscript received August 18, 2000

REFERENCES

- Alfonzo JD, Thiemann O, Simpson L. 1997. The mechanism of U insertion/deletion RNA editing in kinetoplastid mitochondria. *Nucleic Acids Res* 25:3751–3759.
- Arts GJ, Benne R. 1996. Mechanism and evolution of RNA editing in kinetoplastids. *Biochim Biophys Acta* 1307:39–54.
- Benne R. 1994. RNA editing in trypanosomes. *Eur J Biochem* 221:9–23.
- Bhat GJ, Koslowsky DJ, Feagin JE, Smiley BL, Stuart K. 1990. An extensively edited mitochondrial transcript in kinetoplastids encodes a protein homologous to ATPase subunit 6. *Cell* 61:885–894.
- Blum B, Bakalara N, Simpson L. 1990. A model for RNA editing in kinetoplastid mitochondria: "Guide" RNA molecules transcribed from maxicircle DNA provide the edited information. *Cell* 60:189–198.
- Burgess ML, Heidmann S, Stuart K. 1999. Kinetoplastid RNA editing does not require the terminal 3' hydroxyl of guide RNA, but modifications to the guide RNA terminus can inhibit in vitro U insertion. *RNA* 5:883–892.
- Byrne EM, Connell GJ, Simpson L. 1996. Guide RNA-directed uridine insertion RNA editing in vitro. *EMBO J* 15:6758–6765.
- Cruz-Reyes J, Rusche LN, Piller KJ, Sollner-Webb B. 1998. *T. brucei* RNA editing: Adenosine nucleotides inversely affect U-deletion and U-insertion reactions at mRNA cleavage. *Mol Cell* 1:401–409.
- Cruz-Reyes J, Sollner-Webb B. 1996. Trypanosome U-deletional RNA editing involves guide RNA-directed endonuclease cleavage, terminal U exonuclease, and RNA ligase activities. *Proc Natl Acad Sci USA* 93:8901–8906.
- Estevez AM, Simpson L. 1999. Uridine insertion/deletion RNA editing in trypanosome mitochondria—A review. *Gene* 240:247–260.
- Feagin JE. 1990. RNA editing in kinetoplastid mitochondria. *J Biol Chem* 265:19373–19376.
- Feagin JE, Abraham JM, Stuart K. 1988. Extensive editing of the cytochrome c oxidase III transcript in *Trypanosoma brucei*. *Cell* 53:413–422.
- Hajduk SL, Adler B, Madison-Antenucci S, McManus M, Sabatini R. 1997. Insertional and deletional RNA editing in trypanosome mitochondria. *Nucleic Acids Symp Ser* 36:15–18.
- Kable ML, Heidmann S, Stuart KD. 1997. RNA editing: Getting U into RNA. *Trends Biochem Sci* 22:162–166.
- Kable ML, Seiwert SD, Heidmann S, Stuart K. 1996. RNA editing: A mechanism for gRNA-specified uridylation insertion into precursor mRNA. *Science* 273:1189–1195.
- Maslov DA, Sturm NR, Niner BM, Gruszynski ES, Peris M, Simpson L. 1992. An intergenic G-rich region in *Leishmania tarentolae* kinetoplast maxicircle DNA is a pan-edited cryptogene encoding ribosomal protein S12. *Mol Cell Biol* 12:56–67.
- Seiwert SD, Heidmann S, Stuart K. 1996. Direct visualization of uridylation deletion in vitro suggests a mechanism for kinetoplastid RNA editing. *Cell* 84:831–841.
- Simpson L, Wang SH, Thiemann OH, Alfonzo JD, Maslov DA, Avila HA. 1998. U-insertion/deletion Edited Sequence Database. *Nucleic Acids Res* 26:170–176.
- Sloof P, Benne R. 1997. RNA editing in kinetoplastid parasites: What to do with U. *Trends Microbiol* 5:189–195.
- Souza AE, Myler PJ, Stuart K. 1992. Maxicircle CR1 transcripts of *Trypanosoma brucei* are edited and developmentally regulated and encode a putative iron-sulfur protein homologous to an NADH dehydrogenase subunit. *Mol Cell Biol* 12:2100–2107.
- Stuart K. 1993. The RNA editing process in *Trypanosoma brucei*. *Semin Cell Biol* 4:251–260.
- Stuart K, Allen TE, Kable ML, Lawson S. 1997. Kinetoplastid RNA editing: Complexes and catalysts. *Curr Opin Chem Biol* 3:340–346.
- Sturm NR, Maslov DA, Blum B, Simpson L. 1992. Generation of unexpected editing patterns in *Leishmania tarentolae* mitochondrial mRNAs: Misediting produced by misguiding. *Cell* 70:469–476.