Proc. Natl. Acad. Sci. USA
Vol. 90, pp. 9242–9246, October 1993
Biochemistry

The boundaries of partially edited transcripts are not conserved in kinetoplastids: Implications for the guide RNA model of editing

(RNA editing/trypanosomes/cytochrome-*c* oxidase/polymerase chain reaction/molecular evolution)

LAURA F. LANDWEBER, ALEXANDER G. FIKS, AND WALTER GILBERT

Department of Cellular and Developmental Biology, Harvard University Biological Laboratories, 16 Divinity Avenue, Cambridge, MA 02138

Contributed by Walter Gilbert, June 21, 1993

ABSTRACT **We have studied partially edited molecules for the cytochrome-*c* oxidase subunit III (COIII) transcript from two species of the insect trypanosome *Herpetomonas*. We found unexpected patterns of editing, in which editing does not proceed strictly 3′ to 5′, in 24 of 61 partially edited clones. A comparison of the partially edited molecules between the two kinetoplastid species revealed an 8- to 10-nt shift in precisely defined editing boundaries, sites at which editing pauses before binding of the next guide RNA after formation of a stable duplex between a guide RNA and mRNA. This suggests that the region of base pairing between individual guide RNAs and the COIII transcript is not strictly conserved in kinetoplastids, implying gradual evolution of the editing process.**

Kinetoplastid RNA editing, the addition or deletion of uridines from trypanosomatid mitochondrial transcripts, creates >90% of the 288 amino acid codons in the cytochrome-*c* oxidase subunit III (COIII) transcript in both *Trypanosoma brucei* (1) and *Herpetomonas* (2). The mechanism of RNA editing is thought to involve several guide RNA (gRNA) molecules, small maxicircle or minicircle transcripts that mediate editing by base pairing with specific regions of the edited transcript, allowing some G·U base pairs. Complete editing proceeds 3′ to 5′ and requires a set of overlapping gRNAs. Editing by each gRNA creates an anchor sequence for binding the next gRNA (3, 4).

The progressive realignment of gRNA with mRNA (5) produces a series of intermediates as uridines are added to or deleted from active editing sites. These molecules typically contain 3′ contiguously edited sequences and 5′ contiguously unedited sequences, separated by a junction region which contains both correctly edited and incorrectly edited sequences. Partially edited molecules have been found with sites that are not edited precisely 3′ to 5′, particularly in the junction region (5–8). These unexpected patterns of editing result from incomplete editing, the addition or deletion of fewer than the correct number of uridines at an editing site, and misediting, either the editing of sites which are not edited in the mature transcript or the excessive editing of normal editing sites. Pairing of an incorrect gRNA with the mRNA may also lead to misediting (9). Presumably, most of these unexpected patterns are eliminated by reediting of the misedited sites, producing a maturely edited transcript (5).

Editing often pauses at the same position, producing an excess of partially edited molecules which are correctly edited up to a common boundary between the 5′ end of the contiguously edited region and the 3′ end of the junction region. These boundaries reflect gRNA utilization and define the 5′ boundaries of some gRNA/mRNA anchor duplex regions for cytochrome *b* in *Leishmania tarentolae* (6) and ATPase 6 and COIII in *T. brucei* (5, 8). Presumably the pauses occur when one gRNA has finished editing a region and has formed a stable duplex with the mRNA, which must be displaced by the anchor region of the next gRNA (4–6). This suggests that one can use a reverse approach to infer the locations of potential gRNAs within the COIII sequence and to compare these locations between species by sampling the pool of steady-state intermediates from each species. In the absence of information about the gRNAs for *Herpetomonas*, we sought to map the boundaries of partially edited transcripts by this approach, in order to compare the location of potential gRNA/mRNA anchor regions between two species of *Herpetomonas*, *H. megaseliae* and *H. mariadeanei*, which we chose because they have substantially diverged in both the unedited and the edited sequences but are more closely related to each other than either is to *T. brucei* (2).

MATERIALS AND METHODS

Cell Culture. *H. megaseliae* (30209) and *H. mariadeanei* (30708) obtained from the American Type Culture Collection were grown in BHI medium (Difco) and LIT medium (15), respectively, with 10% fetal bovine serum at 25°C.

cDNA Synthesis, Amplification, Cloning, and Sequencing. Oligonucleotides COIII and 4ED were used to amplify partially edited molecules from *H. mariadeanei* first-strand cDNA synthesized with random hexamers and murine reverse transcriptase (Pharmacia), and oligonucleotides COIII-2 and a species-specific primer complementary to the 3′ never-edited region (3′NE) were used to amplify *H. megaseliae* and *H. mariadeanei* first-strand cDNA synthesized with the 3′-NE primer and Superscript reverse transcriptase (BRL) from total RNA (10), using 45 cycles of 0.75 min at 94°C, 1.5 min at 50°C or 58°C, and 1.5 min at 72°C (11). PCR products were spin-purified (Centricon-30, Amicon), cloned (T-vector, Novagen, and TA Cloning, Invitrogen), and 11 clones obtained by the first method were screened by colony PCR and sequenced on both strands with Sequenase (United States Biochemical). Clones obtained by the second method were screened by hybridization to a mixture of oligonucleotides (4ED, 8ED, and 11ED) complementary to the 3′ end of the *T. brucei* edited COIII transcript. Fifty positive clones were sequenced on both strands with *Taq* DNA polymerase (Promega; *fmol* sequencing system). Clone 1.36, obtained by the second procedure, was diluted $<10^{-8}$, reamplified, and cloned exactly as above, and 12 clones were sequenced to quantify PCR error.

Oligonucleotides. Oligodeoxynucleotide sequences were as follows: COIII, 5′-GAAGGAGAGGGGAGGTTTCG-3′; COIII-2, 5′-CCA(A/G)GGAGAAGCAGGGAC(C/A)GA-3′; *H. mariadeanei* 3′NE, 5′-GTATTGTTGTTATAACTACT-

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: COIII, cytochrome-*c* oxidase subunit III; gRNA, guide RNA.

*The sequences reported in this paper have been deposited in the GenBank data base (accession nos. U00597–U00618, U00622–U00658, and U00660–U00661).

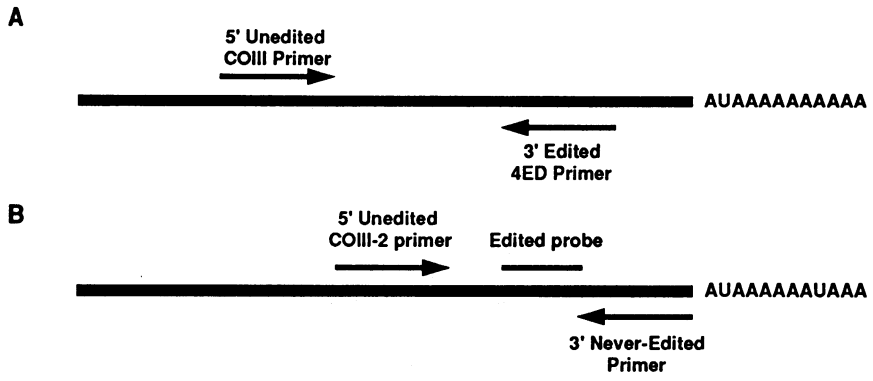

FIG. 1. PCR amplification schemes. With strategy A, 11 clones were obtained from *H. mariadeanei*; with B, 14 clones were obtained from *H. mariadeanei*, and 36 from *H. megaseliae*. Combination A allowed us to study a 481-nt region of the edited transcript. Combination B allowed us to study just the last 292 nt. Clones obtained by strategy B were probed with a mixture of 3′ edited oligonucleotides to ensure that they were partially edited. Primers are not drawn to scale.

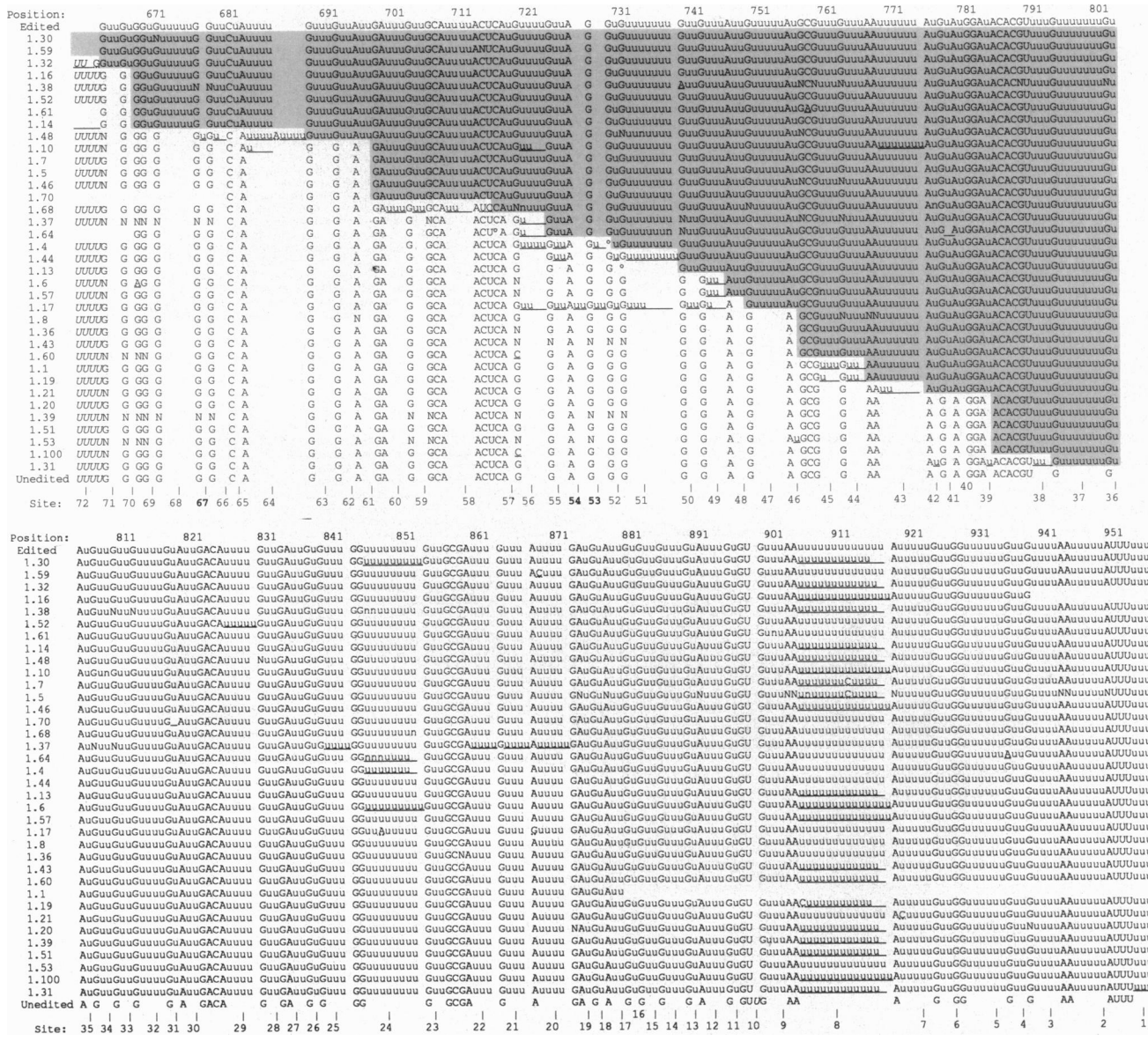

FIG. 2. Alignment of 36 partially edited *H. megaseliae* clones. Positions are numbered according to the completely edited sequence (2). Editing sites are numbered from the 3′ end of the transcript; new editing sites are in boldface. Lowercase u, uridine added by editing; italicized *U*, uridine deleted by editing in the mature transcript. Circles indicate nonuridine deletions. The contiguously edited region is shaded. Partially edited or misedited sites, sites not consistent with editing 3′ to 5′, and errors are underlined. Editing at site 9 creates the UAA stop codon. The GenBank accession nos. for these sequences are U00597–U00618, U00622–U00634, and U00661.

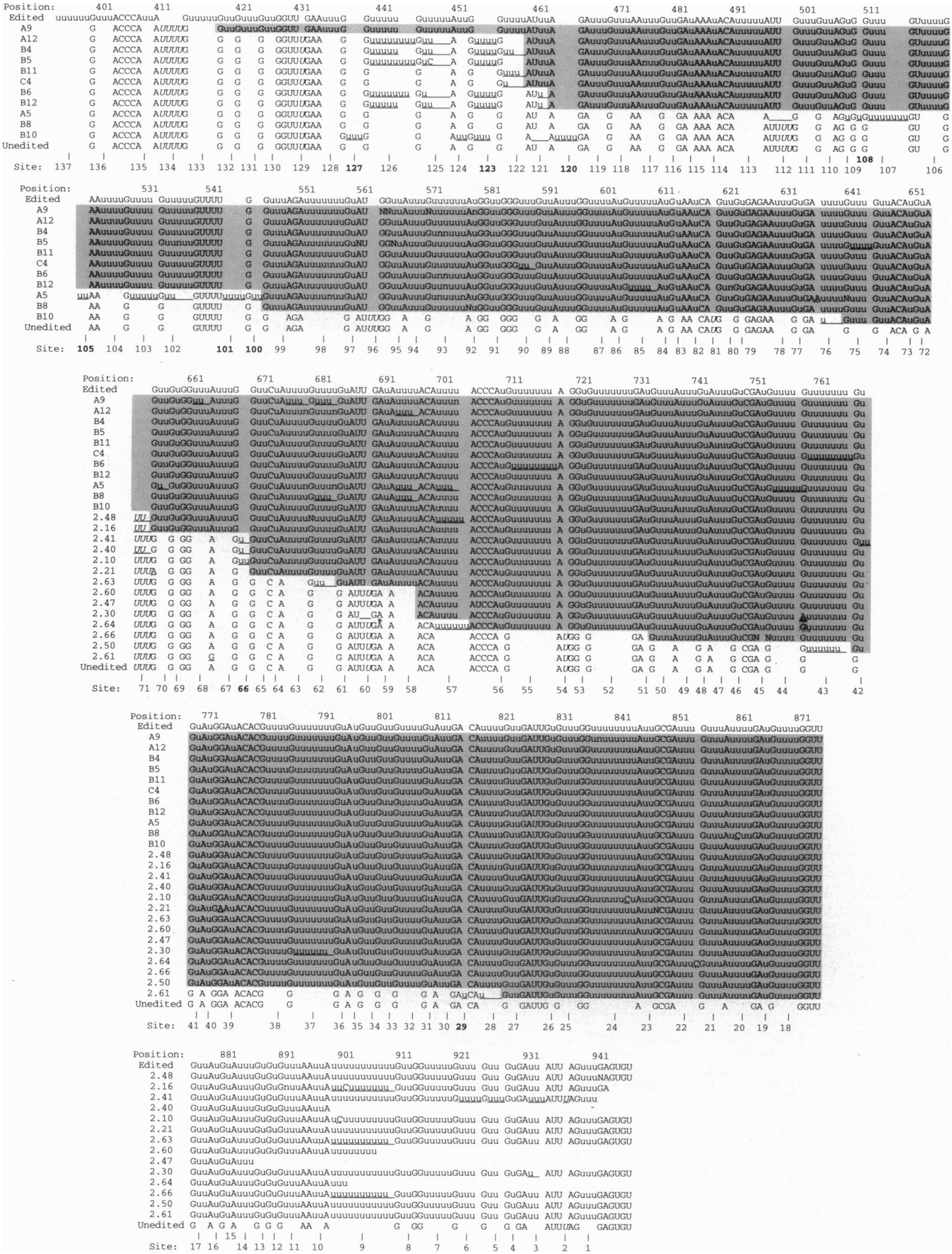

FIG. 3. Alignment of 25 partially edited *H. mariadeanei* clones. Symbols and numbering are as in Fig. 2 legend. Editing at site 11 creates the UAA stop codon. The GenBank accession nos. for these sequences are U00635–U00658 and U00660.

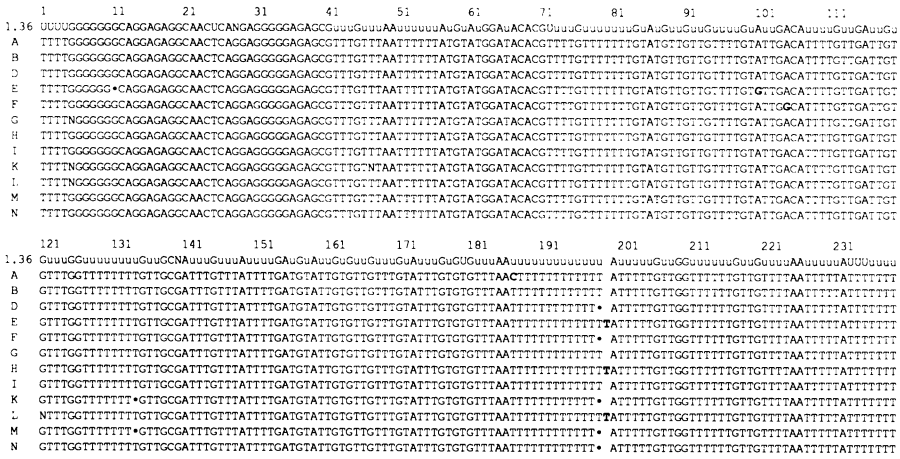

FIG. 4. Reamplified clones from partially edited clone 1.36. Substitutions are in boldface; deletions are indicated by filled circles.

CACT-3′; *H. megaseliae* 3′NE, 5′-CTATATTCTACACA-CACTT-3′; 4ED, 5′-AAATTACACACACAAATACAT-AAC-3′; 8ED, 5′-CAAACTAAATCAACAAAATGTCAA-3′; 11ED, 5′-ACAAAACGTGTATCCATACACAAA-3′. Oligonucleotides 4ED, 8ED, and 11ED were a gift from V. Volloch (Boston Biomedical Research Institute).

Statistical Analysis. The program POISTFIT by R. C. Lewontin was used to analyze each of the three data sets obtained by separate PCRs, to test whether the number of clones containing RNA editing boundaries at each of the possible editing sites can be fit to a Poisson distribution (the null hypothesis). Each class represents the number of editing sites sampled at which 0, 1, 2, 3, 4, or 5 clones share a common boundary. We conservatively estimated the zero class (number of editing sites which were not the boundaries of any clone) by the inclusive number of editing sites covered in the survey (sites of uridine addition or deletion in the correctly edited mRNA located between the first and last partially edited sites) less the number of editing sites which were boundaries in one or more of the clones.

RESULTS AND DISCUSSION

Amplification of Partially Edited Transcripts. We used two PCR strategies based on the 3′-to-5′ progression of editing to isolate partially edited COIII transcripts, by a combination of either 5′ unedited and 3′ edited primers (7) or a 5′ unedited primer and a downstream primer complementary to a never-edited sequence in the 3′ untranslated region, just upstream of the poly(A) tail (Fig. 1). Clones obtained by the second method were screened with a 3′ edited probe to identify ones which were partially edited. We sequenced 11 *H. mariadeanei* clones obtained by the first method and 14 additional *H. mariadeanei* clones and 36 *H. megaseliae* clones obtained by the second method. The sequences of all 61 clones are shown in Figs. 2 and 3. We found 23 different patterns of partial editing in 25 *H. mariadeanei* clones, and 22 different patterns in 36 *H. megaseliae* clones.

Unexpected Editing Patterns. Thirteen of 25 partially edited COIII clones in *H. mariadeanei* and 11 of 36 clones in *H. megaseliae* show departures from 3′-to-5′ editing in the junction region (Figs. 2 and 3), as in *T. brucei* (8) and *L. tarentolae* (6). We found editing of nine new sites, which are not edited in the mature mRNA, in the region of the *H. mariadeanei* transcript we studied, which contains 553 nt and 128 editing sites in the mature mRNA, and editing of three new sites in *H. megaseliae*, which contains 294 nt and 69 mature editing sites in the region we studied. The number of misedited or incompletely edited sites per clone and the number of new editing sites is much greater in *T. brucei* (8) than in either *Herpetomonas* species, although this may reflect differences in the cloning procedure and the length of the region we used.

The level of misediting also depends on the region studied. For example, in *H. mariadeanei*, six of the eight clones that included sites 122–128 in the junction region were edited at new sites. Either editing by an incorrect gRNA (6, 9) or misediting by the correct gRNA (5) could lead to the longer stretches of incorrect editing in the junction regions of seven *H. mariadeanei* clones (A5, B10, B12, B6, B5, B4, and A12; Fig. 3) and *H. megaseliae* clone 1.17 (Fig. 2).

Although incomplete editing and misediting are much more pronounced in the junction regions (8), we found 60 examples of either unexpected editing or PCR errors (57 additions or

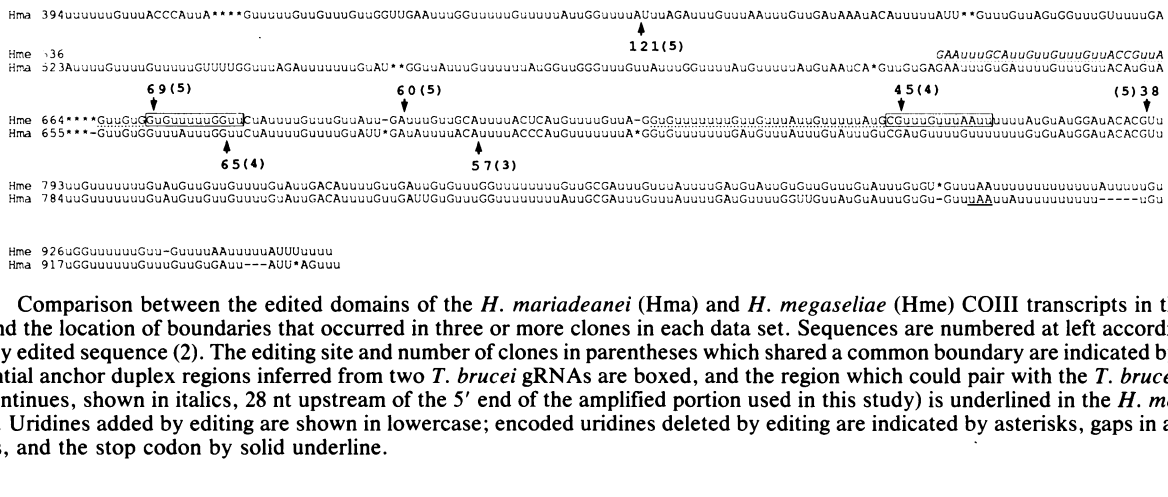

FIG. 5. Comparison between the edited domains of the *H. mariadeanei* (Hma) and *H. megaseliae* (Hme) COIII transcripts in the region studied and the location of boundaries that occurred in three or more clones in each data set. Sequences are numbered at left according to the completely edited sequence (2). The editing site and number of clones in parentheses which shared a common boundary are indicated by arrows. The potential anchor duplex regions inferred from two *T. brucei* gRNAs are boxed, and the region which could pair with the *T. brucei* gRNAs (which continues, shown in italics, 28 nt upstream of the 5′ end of the amplified portion used in this study) is underlined in the *H. megaseliae* sequence. Uridines added by editing are shown in lowercase; encoded uridines deleted by editing are indicated by asterisks, gaps in alignment by dashes, and the stop codon by solid underline.

Table 1. Poisson analysis of the distribution of editing boundaries among classes 1–5

Clone type	No. of clones						*P*, %
	0	1	2	3	4	5	
H. megaseliae	16	7	5	0	1	3	0.002
H. mariadeanei A	46	2	2	0	0	1	0.002
H. mariadeanei B	35	5	1	1	1	0	0.2

The data were fit to a Poisson distribution by using the program POISFIT by R. C. Lewontin. The data are significantly different from a Poisson distribution, due to the number of boundaries shared by five clones in *H. megaseliae* and an excess in the zero class, which we conservatively underestimated (see *Materials and Methods*). A and B refer to the cloning strategy in Fig. 1.

deletions of single uridines and three deletions of two uridines) in the contiguously edited regions of all the clones studied. All of the new editing sites, however, and all editing differences greater than two nucleotides occur in the junction regions.

PCR Error. Since we found so many errors in the contiguously edited region, we reamplified and sequenced 12 copies from a single clone to measure the true error due to PCR (Fig. 4). We found 10 uridine additions or deletions (6.0×10^{-3} per uridine) in the reamplified data set; therefore, the rate of uridine additions or deletions in the contiguously edited region (6.2×10^{-3} per uridine) can be attributed to PCR error. There were also three transitions (1.0×10^{-3}), one non-uridine deletion (3.5×10^{-4}), and no transversions ($<3.5 \times 10^{-4}$) in the reamplified data set of 2880 bp, which accounts for the rate of transitions (8.3×10^{-4}), transversions (1.3×10^{-4}), and non-uridine additions or deletions (1.9×10^{-4}) in the contiguously edited region (15,668 bp). As cytosines are rare or nearly absent, the occasional purine deletions found in *L. tarentolae* (6) can also be explained by PCR error. Site 8 in the 3′ untranslated region of *H. megaseliae* is unusually prone to either slippage or misediting (Fig. 2), since 21 of the 36 clones differ from the consensus 13 uridines added in this position; however, the PCR mutation rate was 75% at this site (Fig. 4). Therefore the mutation rate at this site in the partially edited clones is presumably due to slippage of the *Taq* polymerase, rather than misediting or incomplete editing.

Identification of Editing Boundaries. The boundary, or the site at which the maturely edited sequence ends and the junction region begins, typically reflects gRNA utilization and gRNA/mRNA anchor regions. We have analyzed published data sets of partially edited molecules from *L. tarentolae* for which a complete set of overlapping gRNAs is known (4, 12) and have found that boundaries which occur in 10% or more of the clones usually correspond to the 5′-most nucleotide which can base pair with an individual gRNA and can form the anchor duplex with the upstream gRNA. We identified four boundaries in *H. megaseliae* and three boundaries in *H. mariadeanei* that are shared by three or more clones (>10%; Fig. 5). Three pairs of boundaries (two in *H. megaseliae* and one in *H. mariadeanei*) are evenly spaced 28–34 nt apart, precisely the length of the region templated by each of the overlapping gRNAs in *L. tarentolae* (4). This suggests that editing of this region of the transcript in *Herpetomonas* may proceed in blocks of 28–34 nt mediated by a single gRNA, as in *L. tarentolae*. In addition, two separate pairs of *H. megaseliae* boundaries are 59 nt apart, twice the length between a single pair, which suggests that another boundary lies between them, although we did not detect it in three or more clones in this study.

A comparison of these common boundaries between the two species (Fig. 5) reveals that the boundaries in the two *Herpetomonas* species have shifted relative to each other so that each of the boundaries in *H. megaseliae* is located 8–10 nt upstream of the corresponding boundary in *H. mariadeanei*. A statistical analysis of the individual data sets (Table 1) reveals that they are significantly different from a Poisson distribution. Thus we can reject the null hypothesis that the distribution of boundaries in the partially edited molecules is guided by a Poisson process and interpret the shifts between them to be significant.

The finding of two *T. brucei* COIII gRNAs (K. Stuart, personal communication) which can form perfect anchor duplexes with the regions immediately downstream of two editing boundaries in *H. megaseliae* significantly strengthens the view that the boundaries indicate potential gRNA/mRNA anchor duplex regions. Tb:gCOIII[110] can pair with positions 636–681 of the *H. megaseliae* COIII transcript, requiring only three changes in the gRNA and forming a 12-bp Watson–Crick anchor at positions 670–681, and Tb:gCOIII[78] can pair with positions 730–771, also requiring only three changes in the gRNA and forming a 13-bp anchor at positions 759–771. In both regions templated by these gRNAs, the *T. brucei* mRNA sequence differs from the *H. mariadeanei* sequence by 11 nt but differs from the *H. megaseliae* sequence by only 6 nt (3 of which are A-to-G substitutions which can pair with a corresponding U in the gRNA) and 3 nt, respectively. This suggests that *H. mariadeanei* has diverged in both the mRNA and the gRNA sequences from *T. brucei* and *H. megaseliae*, which are more closely related in this region.

The shift in precisely defined editing boundaries is a surprising result, because we expected that the requirement for multiple changes in the gRNAs would have restricted the drift in the location of the boundaries. Instead, our results suggest that the region of base pairing between individual gRNAs and the COIII transcript is not conserved in kinetoplastids, implying gradual movement over evolutionary time of the anchor region between the gRNA and the mRNA. The shift may have resulted from the extension in *H. mariadeanei* of the region of base pairing between the gRNAs and the mRNA at the 3′ end, with the concomitant loss of base pairing at the 5′ end, conserving the overall distance between editing boundaries. The finding of several overlapping and redundant gRNAs in *T. brucei* (13, 14), which is ancestral to *Herpetomonas* (L.F.L. and W.G., unpublished work), could also explain the presence of multiple editing boundaries. The shift we observed could possibly result from the fixation of different overlapping sets of gRNAs in the two species of *Herpetomonas*. The redundancy in the gRNAs would also enhance their mutation rate, which is thought to be high, because the COIII transcript contains many fixed changes in editing (2).

We thank R. C. Lewontin for discussion, K. Stuart for providing unpublished information, and V. Volloch for oligonucleotides. This work was supported by the National Institutes of Health. L.F.L. is a Howard Hughes Medical Institute predoctoral fellow.

1. Feagin, J. E., Abraham, J. M. & Stuart, K. (1988) *Cell* **53**, 413–422.
2. Landweber, L. F. & Gilbert, W. (1993) *Nature (London)* **363**, 179–182.
3. Blum, B., Bakalara, N. & Simpson, L. (1990) *Cell* **60**, 189–198.
4. Maslov, D. A. & Simpson, L. (1992) *Cell* **70**, 459–467.
5. Koslowsky, D. J., Bhat, G. J., Reed, L. K. & Stuart, K. (1991) *Cell* **67**, 1–20.
6. Sturm, N. R. & Simpson, L. (1990) *Cell* **61**, 871–878.
7. Abraham, J. M., Feagin, J. E. & Stuart, K. (1988) *Cell* **55**, 267–272.
8. Decker, C. J. & Sollner-Webb, B. (1990) *Cell* **61**, 1001–1011.
9. Sturm, N. R., Maslov, D. A., Blum, B. & Simpson, L. (1992) *Cell* **70**, 469–476.
10. Landweber, L. F. & Kreitman, M. (1993) *Methods Enzymol.* **218**, 17–26.
11. Chomczynski, P. & Sacchi, N. (1987) *Anal. Biochem.* **162**, 156–159.
12. Maslov, D. A., Sturm, N. R., Niner, B. M., Gruszynski, E. S., Peris, M. & Simpson, L. (1992) *Mol. Cell. Biol.* **12**, 56–67.
13. Pollard, V. W., Rohrer, S. P., Michelotti, E. F., Hancock, K. & Hajduk, S. L. (1990) *Cell* **63**, 783–790.
14. Koslowsky, D. J., Riley, G. R., Feagin, J. E. & Stuart, K. (1992) *Mol. Cell. Biol.* **12**, 2043–2049.
15. Yoshida, N., Freymüller, E. & Wallace, F. G. (1978) *J. Protozool.* **25**, 421–425.