The EMBO Journal vol.13 no.23 pp.5689–5700, 1994

Disruption of RNA editing in *Leishmania tarentolae* by the loss of minicircle-encoded guide RNA genes

Otavio H.Thiemann[1], Dmitri A.Maslov[2] and Larry Simpson[1,2,3]

[1]Department of Biology, [2]Howard Hughes Medical Institute and [3]Department of Medical Microbiology and Immunology, University of California, Los Angeles, CA 90024, USA

[3]Corresponding author

Communicated by E.De Robertis

RNA editing in kinetoplastids appears to be a labile genetic trait that is affected by prolonged cell culture. The transcripts of the G1–G5 cryptogenes are pan-edited in the recently isolated LEM125 strain of *Leishmania tarentolae*, but not in the UC strain which has been in culture for 55 years. At least 32 minicircle-encoded guide RNAs (gRNAs) for the editing of G1–G5 transcripts are present in LEM125 and absent in UC. We hypothesize that specific minicircle sequence classes encoding gRNAs for the editing of these transcripts were lost during the long culture history of the UC strain. The protein products, which include components of complex I of the respiratory chain, are probably not required during the culture stage of the *Leishmania* life cycle.

Key words: cryptogenes/guide RNAs/*Leishmania*/minicircles/RNA editing

Introduction

The mitochondrial genome of the trypanosomatids is composed of two forms of catenated circular DNAs, maxicircles and minicircles (Simpson, 1987). The 30 kb maxicircles of *Leishmania tarentolae* are present in ~50 copies and encode rRNAs and several mitochondrial structural genes. The minicircles are present in a significantly higher copy number, 5000–10 000 molecules per network, and in *L.tarentolae* ~900 bp in size with a single guide RNA (gRNA) (Sturm and Simpson, 1991) gene located in the variable region. The sequences of the transcripts of several maxicircle genes are modified by RNA editing, with the edited sequence information residing in multiple overlapping gRNAs encoded in both maxicircles and minicircles (Simpson *et al.*, 1993; Stuart, 1993; Benne, 1994).

A comparison of the maxicircle sequences of *L.tarentolae* and *Trypanosoma brucei* had previously identified six G-rich intergenic sequences, which are conserved in both species in location and polarity but not in sequence (Simpson *et al.*, 1987). It was speculated that these sequences represent pan-edited cryptogenes. This suggestion was verified in the case of G6, the transcripts of which for both species were shown to be pan-edited yielding mRNAs encoding a ribosomal protein S12 homologue (Maslov *et al.*, 1992; Read *et al.*, 1992). In *T.brucei*, pan-editing of G1 (=CR1), G2 (=CR2), G4 (=CR4) and G5 (=CR5) was also demonstrated (Souza *et al.*, 1992, 1993; Correll *et al.*, 1994; Read *et al.*, 1994). The proteins encoded by the mature edited CR1 and CR2 RNAs proved to be homologues of NADH dehydrogenase subunits 8 (ND8) and 9 (ND9) (Souza *et al.*, 1992, 1993), respectively. The protein encoded by edited CR5 showed a weak similarity to NADH dehydrogenase subunit 3 (ND3) (Read *et al.*, 1994). The lack of editing in ND8 and ND9, in an old laboratory strain of *Crithidia fasiculata*, has also been reported (Sloof *et al.*, 1994). However, a recently isolated strain of *C.fasciculata* has not been examined.

Multiple gRNAs have been identified in *T.brucei* which could mediate portions of these editing events (Corell *et al.*, 1993), and the known genomic complexity of >900 different minicircle-encoded gRNAs in this species could clearly account for the remainder of the editing sequence information. In the case of the UC strain of *L.tarentolae*, however, the genomic gRNA complexity (Table I) is not sufficient to mediate pan-editing of transcripts of G1–G5: A total of eight maxicircle-encoded and 15 minicircle-encoded gRNAs have been identified, all of which were involved with the editing of genes other than G1–G5 (Maslov and Simpson, 1992).

The UC strain of *L.tarentolae* was isolated from a gecko in Algeria in 1939 (Parrot and Foley, 1939) (originally the Parrot TarII strain) and maintained in axenic culture as the promastigote form (which is the form in the insect

Table I. Guide RNA complexity in *L.tarentolae* UC and LEM125 strains

Cryptogenes	Number of gRNAs UC + LEM125 maxicircle DNA	UC + LEM125 minicircle DNA	Total (expected)
COII	1	0	1
COIII	0	2	2
ND7	2	0	2
CyB	2	0	2
MURF2	2	0	2
MURF4 (A6)	0	6	6
RPS12 (G6)	1	7	8
	UC + LEM125	LEM125	
ND8 (G1)	1[a]	9	10 (14)
ND9 (G2)	1	8	9 (17)
G3	2[b]	1	3 (6)
G4	1	9	10 (15)
ND3 (G5)	1	5	6 (9)
Unassigned	1[c]	0	1
Total	13	47	60 (83)

[a]gND8-XIII, a putative gRNA.
[b]gG3-II, a putative gRNA.
[c]gM150, a putative gRNA found in a gRNA/mRNA misguided chimera.

© Oxford University Press

vector) in various laboratories. A comparison of the restriction enzyme digestion profile of the kinetoplast DNA (kDNA) of the UC strain with profiles of digested kDNAs from several more recent isolates of *L.tarentolae* (Wallbanks *et al.*, 1985), indicated that the minicircle complexity of the UC strain was substantially lower than that of the recently isolated strains (Gomez-Eichelmann *et al.*, 1988). Since minicircles encode gRNAs (Sturm and Simpson, 1990), this analysis raised the possibility that the larger minicircle complexity in the latter strains would give rise to a larger gRNA repertoire, which would be sufficient for editing of G1–G5.

Here, we examine the editing of G1–G5 and the corresponding gRNA complexity in a recently isolated strain of *L.tarentolae*, and present evidence that a loss of minicircle sequence classes containing specific families of gRNA genes can occur during prolonged culture. This result has implications not only for the maintenance of trypanosomatids in culture but also for the evolution of RNA editing in general.

Results

The pre-edited cryptogenes G1–G5 are almost identical in sequence in the UC and the LEM125 strains

Maxicircle sequences from LEM125 were PCR-amplified from kDNA as described in Materials and methods, and the sequences of G1–G5 were compared with the corresponding sequences from the UC strain (Simpson *et al.*, 1987). The homologous sequences were almost completely identical between the two strains, the only differences being an A substituted for a G at the 5′ end of G1 (Figure 1A), an A deletion and a G addition in G2 (Figure 2A), and a G and A addition in G4 (Figure 4A), any or all of which could actually represent PCR artifacts rather than true polymorphisms. The G3 and G5 sequences were identical in both strains. This evidence suggests that these represent closely related strains of *L.tarentolae*, in spite of the fact that they were isolated in Algeria and France 46 years apart (Parrot and Foley, 1939; Wallbanks *et al.*, 1985).

The transcripts of the G1–G5 cryptogenes in LEM125 are pan-edited, and the edited mRNAs encode the respiratory complex I components ND8, ND9 and ND3 and two unidentified proteins

Based on the known 3′–5′ progression of editing (Abraham *et al.*, 1988; Stuart *et al.*, 1989; Decker and Sollner-Webb, 1990; Maslov and Simpson, 1992), partially edited RNAs from each G-rich region were obtained by RT-PCR from total kinetoplast RNA (kRNA) from LEM125. An oligo(dT) primer (S-399) to the poly(A) tail was used to synthesize cDNA, and unedited oligonucleotides complementary to genomic sequences in the 5′ domains of G1–G5 were employed as 5′ PCR primers. The consensus sequences derived from the partially edited cDNAs were confirmed and extended by direct primer extension sequencing of edited RNAs. In addition, a consensus sequence for the 5′ region of the editing domain was obtained from each G-rich region by sequencing cloned PCR-amplified 5′ G-tailed cDNA molecules of edited RNAs.

A similar procedure failed to obtain consensus edited sequences for transcripts of G1–G5 from the UC strain except for a limited region at the 3′ end of G5 (Figure 9).

LEM125 G1

A consensus edited sequence (Figure 1A) was determined from the sequences of five partially edited cDNAs and 15 clones containing the 5′ end of the edited RNA (data not shown). The mature edited G1 RNA of LEM125 is 520 nucleotides in length, and is pan-edited by the addition of 215 uridines in 99 sites and the deletion of 41 uridines in 17 sites. An open reading frame (ORF) of 145 amino acid residues which is encoded by the edited transcript is homologous to the ND8 polypeptide encoded by the edited G1 (CR1) RNA in *T.brucei* (Souza *et al.*, 1992). The alignment of the *T.brucei* ND8 sequence with the LEM125 G1 sequence (Figure 1B) showed a Z value of 36 SD units above the mean of aligned randomized sequences (Kanehisa, 1982), which is highly significant. In addition, the characteristic motif CxxCxxCxxxCP identified in the *T.brucei* ND8 (Souza *et al.*, 1992) protein is also observed in the predicted G1 polypeptide (Figure 1B). The AuG methionine initiation codon in G1 RNA is created by the addition of one uridine in site 115, and the uAG termination codon, is created by editing in site 6. The fully edited G1 RNA has an unedited 35 nucleotide 5′ terminal sequence. At the 3′ end the G1 transcript has an overlap of 14 nucleotides with the 3′ end of the G2 transcript, which is transcribed in the opposite direction.

The homology with the *T.brucei* ND8 protein (Souza *et al.*, 1992) provides strong evidence for the authenticity of this edited sequence.

LEM125 G2

A consensus edited sequence was determined from the sequences of nine partially edited cDNAs and 13 clones containing the 5′ end of the edited RNA (data not shown). The mature edited G2 RNA is 652 nucleotides in length with 335 uridine additions in 125 sites and 40 uridine deletions in 15 sites (Figure 2A). The edited RNA encodes an ORF of 196 amino acids which is homologous to the ND9 polypeptide encoded by the edited transcript of the CR2 gene from *T.brucei* (Figure 2B) (Souza *et al.*, 1993). The Z value of the alignment is a highly significant 17.5 SD units (Kanehisa, 1982). The first AuG methionine codon is created by editing at site 139 and coincides with the first methionine codon predicted for the edited CR2 transcript from *T.brucei*. The termination codon (uAG) is created by editing in site 7. A possible in-frame non-canonical translation initiation codon, AUA-isoleucine, is localized in the 5′ unedited sequence of G2, which extends 30 nucleotides 5′ of the last editing site.

As in the case of G1, the significant homology with the *T.brucei* ND9 protein sequence provides strong support for this consensus edited sequence.

LEM125 G3

The genomic G3 sequence is the shortest cryptogene of *L.tarentolae* and shows some unusual editing features. A consensus edited sequence was determined from the sequences of five partially edited RNAs (data not shown). The transcript is edited by 35 uridine additions and 14 uridine deletions, yielding an RNA molecule 205

A

```
                                    gND8-XIII (Mc)                                   gND8-XII(mc)
                  ..uuuuuU-UUuUaagCaagUaUUaaaaaUaAAAA-5'   .aUaaaUggagaUgUaCgUgagagaUaCaaCaCaaU-GagaUCa-5'
            ↓
5'-TAAACATATATAATGTATTAGATTAAATTAAATAGTAA G   G   A GA    G TTTTC     G   G     A A G G     A G  G G  ATC   GG
5'-UAAACAUAUAUAAUGUAUUAGAUUAAAUUAAAUAG*AAuGuuuGuuuAuGAuuuuuGuUUUCuuuuuuuGuuuGuuuuuAuAuGuGuuuuuuAuGuuGuGuuA*CuuuGGuuuuA
                                       116 115 114 113      111          109    108    107    105   103    101  99 98  97   96
                                         M  F  V  Y  D  F  C  F  S  F  F  V  C  F  Y  M  C  F  L  C  C  V  T   L  V  L

                   gND8-X (mc)
   ..aCaaUUaUagaCaC-AAGUaCCaUUGGUagaagaCGCAAAUAU-AAC-5'
                                                                gND8-IX (mc)
                                   ..UagaaaU-AGCU—ACgUUgaaUCUUGUaaUaUagCGAaCGGCaaaUaCAC-5'
CCA  GGAG  GACCA  G  AG A   G GTTTCG GG AACCA      GCG     ATTTGATTTG GG   AGAACG  G A  GCC GCCG   A GTGA    A ATTTTGCCC
CCAuuGGAGuuGACCAuuGuuAGuAuuuGuG*UUCGuGGuAACCAuuuuuuGCGuuuuuA*UUGA**UGuGGuuuAGAACGuuGuAuuGCCuGCCGuuuAuGUGAuuuuAuA***UGCCC
  95   94    93 92 91 90   88  87 86        85     84  83  82  81  80      79   77  76    75      73    71
P  L  E  L  T  I  V  S  I  C  V  R  G  N  H  F  L  R  F  Y  W   C  G  L  E  R  C  I  A  C  R  L  C  D  F  I    C  P

                                            gND8-VII(mc)
                      ..UaU-AUUaaUAGAUaagaGGUUaCaCgagAUaUauUCGAUAGC——AGCAACA-5'
                                                                                  gND8-VI(mc)
                                                     ..uugaUAGU——AGCaaCaUaagUaaCgUCgaAAACaUaCGUaCa
AAG   AGC CTTTTTTAGA G  CG  G G  AGAAG   A GTTGG  ATCGG    CCGA G G   AA A  AG  ATCGTTTTCG  G A   A  G GG  TTTG A GCA G
AAGuuuAGCuC*****UAGAuGuuCGuuGuGuuAGAAGuuuAuG*UGGuuAUCGGuuuuCCGAuGuGuuuAAuAuuAGuuAUCG***UCGuuGuAuuuAuuGuGGuuUUUGuAuGCAuGu
  70   69   68    67 66  65   63     62  60  59      58      57    55    53  52    51   50    48 47   45   44   42 4
 S  L  A  L      D  V  R  C  V  R  S  L  C   G  Y  R  F  S  D  V  F  I  Y  S  Y  R    R  C  Y  Y  C  G  F  C  M  H  V

                           gND8-IV (mc)
              ..UgUaagaaaCaaUagagCgaCaaCaUgaCGGUACAU-5'                                         ..C———UagCaaCaag
                                                                gND8-III (mc)
aaC-5'                                                  ..UaCaUaaaUaU-G-CGUGCAUUUagaCAAGaUaagC———CAACAACA -5'
 G CCAACATTTTGACTTTTTTGCCA  ACTTGCA  CA G      G  A    G  G  G A  GCCA G A   A GTCTGCACCTAAA   GTTT A   GTTTTG  G  G
uuGuCCAACA****GAC******GCCAuuAC**GCAuuCAuGuuuuuuGuuAuuuuGuuGuuGuAuuGCCAuGuAuuuAuG*C*GCACCUAAAuuuGUUUuAuuuG****GuuGuuGuuu
41       39      38     37  36   35      33   32  31  30 29  27   26   24   22 21        20      18    17   16 15  14
C  P  T     D      A  I  T    H  S  C  F  L  L  F  C  C  C  I  A  M  Y  L  C    A  P  K  F  V  L  F  G     C  C  F

  gND8-II (mc)
aUaCgaU——AAACUaagaAUAAACACAAA-5'
                              gND8-I (mc)
        ..gUagaUaCaagaUagaUCgagUCaCaUaG—AGCUCUUCAUAUA -5'
 A G  ATTTTTGA    A   G G    G   AG    AG G A CTTTCGAGAAGTATATTTGATTAATAATTTAATAAATTAATTTTATAGACAGTT
uAuGuuA***UUUGAuuuuuAuuuGuGuuuuGuuuAGuuuAGuGuAuC**UCGAGAAGUAUAUUUGAUUAAUAAUUUAAUAAAUUAAUUUU(A)n
  12   11      10    9 8  7     6      5  4 3 2 1                                   ^^^^^^^^^^^^^^^^^
 M  L     F  D  F  Y  L  C  F  V  Stop
```

B

```
                                               I                                II                                  III
L. t.   1 MFVYDFCFSFFVCFYMCFLCCVTLVLPLELTIVSICVRGNHFLRFYWCGLERCIACRLCDFICPSLALDVRCVRSLCGYRFSDVFIYSYRRCYYCGFCMHVCPT 104
T. b.   1 MFFFDFLFFFFVCFYMCFVCCVTICLPIELTIVSLLVRGNHFLRFYWCGLERCIACRLCDLICPSLALDVRVGWSFGGHRFADWFTLSYRRCIYCGFCMHVCPT 104

L. t. 105 DAITHSCFLLFCCCIAMYLCAPKFVLFGCCFMLFDFYLCFV 145
T. b. 105 DAITHSLFVMCFCCLAMYLLAPKFLLFGCCFMLFDFYLCFV 145
```

Fig. 1. The pan-edited G1 (ND8) cryptogene in the LEM125 strain. (**A**) The genomic and the mature edited mRNA sequences are shown, with the editing sites numbered 3′−5′. Added uridines are shown with lower case letters and deleted uridines as *. The editing domain is underlined. The cognate gRNAs are shown above the mRNA sequence, with canonical base pairs indicated by parallel lines, G−U base pairs by colons and A−C base pairs by dots. Guide nucleotides in the gRNA sequences are indicated as lower case letters a and g. The gRNA anchor regions are indicated by stippling. The methionine (AuG) and the stop (uAG) codons are indicated in bold and double underlining. The predicted amino acid sequence is shown below. (Mc) maxicircle-encoded gRNA, (mc) minicircle-encoded gRNA. The arrow indicates the site of an A to G substitution in the genomic seuqence of LEM125(3C16) as compared with UC. The arrowheads indicate the overlap with the 3′ end of G2 (ND9). (**B**) Alignment of the predicted amino acid sequences of the edited G1 (ND8) mRNAs from *L.tarentolae* LEM125 and *T.brucei* (Souza *et al.*, 1992). The stippled boxes labeled I−III indicate the conserved domains present in *T.brucei* (Souza *et al.*, 1992). The motif 'CxxCxxCxxxCP' is indicated by bold letters. (L. t.) *L.tarentolae* LEM125 sequence, (T. b.) *T.brucei* sequence.

nucleotides in length (Figure 3). The additions occur in the first 15 sites and the deletions in site 16 (one U residue deleted), site 17 (five U residues deleted) and site 18 (eight U residues deleted). In order to sequence the 5′ end of the edited G3 RNA, RT-PCR was performed using a 3′ edited primer covering sites 11–15 and a 5′ oligo(C) primer complementary to a 5′-tailed oligo(G) sequence. All 10 clones yielded the sequence shown in Figure 3.

This pattern of deletions occurring at the 5′ end of an editing domain is highly unusual. In addition, the genomic sequence contained between the last two deletion sites 17 and 18 has the characteristics of a G-rich pre-edited sequence, raising the possibility that the 5′ end of the edited G3 RNA is not functional in translation and could represent aberrant editing. An ORF of 51 amino acids is created by these editing events, with an AUA codon (isoleucine) being a possible in-frame non-canonical translation initiation codon (Figure 3). No significant homology to any sequence in the database was detected.

The 5′ end of the G3 RNA overlaps 39 nucleotides with the 3′ end of the MURF1 transcript, and the 3′ end of the G3 RNA overlaps with the 5′ end of the ND1 transcript by 36 nucleotides (Simpson *et al.*, 1987).

A

```
                                                                              ..aaUaUagUaUUUaUU————————C
                                                                                ||:|:|::||:||::         |
5'- CGTTATATAATTTAATATCTATTAAGTATAA A G    A    TGTTTA  A GTTTCG  G G G   G G G G A      G   G G  G AGA GGTTTTTTTG
5'- CGUUAUAUAAUUUAAUAUCUAUUAAGUAUAAuAuGuuuuuAuuuuUG***AuuAuGUUUCGuuGuGuGuuuGuGuuGuuGuuAuuuuuuuGuuuGuGuuGuAGAuGG*******G
                                140  138   137   136 135          133  131        128    126   125   124 123   121  120    119
       I  S  I  K  Y  N  M  F  L  F  L    I  M  F  R  C  V  F  V  L  L  L  F  F  C  L  C  C  R  W        V

gND9-XIV(Mc)                                                                                        gND9-XII (mc)
gaaaagaUaUaaaaUaaCUaACaUC—AAA-5'                                ..agaUaagaaAUUggaUaCUagaAAAAAUaaCaUaaCaaaU
  A G    G  GA   G AGTTTTTTG       A   G   G G G      A       GTTTGA G    A    TAA   A GA    TTTTTA  G A G   A
uuuuuuuAuGuuuuGuuGAuuGuAG**UUUUGuuuuuuuuuAuuuGuuuGuGuGuuuuuuuuuAuuuuuuG**UGAuGuuuuuAuuuuuUAAuuuAuGAuuuUUUUUAuuGuAuuGuuuA
  118      116 115 114   112            111    110 109   107      106      105  104    102     101      100  99  98       97   95   94
 F  L  C  F  V  D  C  S    F  V  F  F  Y  L  F  V  C  F  F  L  F  F  V    M  F  L  F  F  N  L  W  F  F  L  L  Y  C  L

CUaaaU-5'
||||||
GA   A    G A AGA    GCGG      G     GTTTTTTCG    A     GATA A GTTTTTTG    G   A  A  A G    CGAGTTTTTAAG    G A  GA   G
GAuuuAuuuuGuAuAGAuuuuuGCGGuuuuuGuuuuG****UUCGuuuuAuuuuGAUAuAuG**UUUUGuuuuGuuuAuuAuuAuGuuuuCGAG****UAAGuuuuGuAuuGAuuuGuuu
  93   92    90    89        88  87    86          85   84       82            80    79 78 77     75          74        73  72  71  70
D  L  F  C  I  D  F  C  G  F  C  F  V      R  F  I  L  I  Y    V  L  F  C  L  L  L  C  F  R  V     S  F  V  L  I  C  F

                                                   gND9-VIII(mc)
                              ..UagUaUaUaaaAGCUUGAgCU———UCggaaaCUAAACAAACCAAA-5'     ..uuuuUaUUgaUaagUGU
                                                                              gND9-VII(mc)
                                                            ..uuuUgaaCUaaaaUaagagaCaCCaUUACUAACAAAUG -5'
  G     GG    GG      AG    A          G AG   A GCG   A G A   TTGAACG  GATTTAG      GATTTG    GG     G      G GG AA GA   G    ACA
uuuuuuGuuuuuuGGuuuGGuuuuuAGuuuAuuuuuuuuGuAGuuAuGCGuuAuGuAuuuUUGAACGuGA***AGuuuuuGAUUUGuuuGGuuuuGuuuuuuGuGGuAAuGAuuGuuuACA
 69         68      67      66     65  64       63  62  61   60    58  57          56   55     54          53     52     51  50  49     47  46
 F  L  F  F  G  L  V  F  S  L  F  F  C  S  Y  A  L  C  I  F  E  R  E    V  F  D  L  F  G  F  V  F  C  G  N  D  C  L  H

gND9-VI (mc)
-AGUggagaUaCagCUAACUgagaaACAACCAAAA-5'
                                          gND9-V(mc)                                        gND9-III (mc)
                  ..aUagCUaaaggaagUgCaUUUCaaaUaGGU—AACaaaCCaaaCa-5'          ..gCaUaUgaUaUgUUUCUGUaaCaaUaUU
TTCG     A G  GATTGA      G  GG       G G AAAG    A CCATTTTG    GG    G     ATTTTG   GTATTTA       G A G   A G GAAGATA   G  G GA
*UCGuuuuuAuGuuGAUUGAuuuuuuGuuGGuuuuuuuuuuGuGuAAAGuuuAuCCA**UUGuuuGGuuuGuuuA***UGuuG*A***AuuuuuGuAuGuuAuGuGAAGAUAuuGuuGuGA
  44    43 42        41     40        39     38        36  35   34       33     32   31   30    29       27       26  25     23    21   20  19
 R  F  Y  V  D  W  F  F  V  G  F  F  L  C  K  V  Y  P    L  F  G  L  F  M     L   N     F  C  M  L  C  E  D  I  V  V  I

          gND9-II (mc)
  ..U———UaAUaaggCaCagUACaaaaAGGUUaGAACGUUaUaU-5'
aaCGGUAgU———AC-ACAAUAUAC -5'
 GCCAC ATTTTG TG    G G  GTG     TCCAA  TTTGCAA  A AG       A    TATTTA    TGTTA  GAA  CTGTCTATAAAATTAATTTATTAAATTATTAATCAA
uuGCCACuA****GuUGuuuuGuGuuGUGuuuuUCCAAuUUUGCAAuAuAGuuuuuAuuuUA**UAuuUGUUAuGAAuCUGUCUAUAAAAUUAAUUUAUU(A)n
 17        16  15    14    13   12 11      10         9         8  7      6       5    4     3          2      1
 A   T   S      C   F   V   L   C   F   S   N   F   A   I  Stop
```

B

```
L. t.   1   ISIKYNMFLFLIMFRCVFVLLLFFCLCCRWVFLCFVDCSFVFFYLFVCFFLFFVMFLFFNLWFFLLYCLDLFCIDFCGFCFVRFILIYVLFCLLLCFRVSFVLI  103
T. b.   1   ......MCIFLCLFRLCFCLILFYCLCCRWCFVCFVDCSFLFFYCFVSFFLFYCMFLFFNLWFLFLYCCDLLLIDFCGFCFCRFMLLYILFCLFLCVRLCFVLC  97

L. t. 104   CFFLFFGLVFSLFFCSYALCIFEREVFDLFGFVFCGNDCLHRFYVDWFFVGFFLCKVYPLFGLFMLNFCMLCEDIVVIATSCFVLCFSNFAI....  196
T. b.  98   CLFVFFGLCFSFSCFCYAFLLLERECFDLFGFYFVGNDILHRLFVDWFFVGFFLLKCYPLFGLFVLLFCVLVEEIVCTFTMLFLLLHTNFYLHYFI  194
```

Fig. 2. The pan-edited G1 (ND9) cryptogene in the LEM125 strain. (**A**) See Figure 1A for symbols and nomenclature. The two possible initiation codons (AUA and AuG) are indicated by double underlining. The arrowheads indicate the overlap with the 3′ end of G1 (ND8). (**B**) Alignment of the predicted amino acid sequences of the edited G2 (ND9) mRNAs from *L.tarentolae* LEM125 and *T.brucei* (Souza *et al.*, 1993). (L. t.) *L.tarentolae* LEM125 sequence, (T. b.) *T.brucei* sequence. The two possible initiation codons are indicated in the *L.tarentolae* sequence by underlined bold case.

```
                                                                                                   gG3-III (mc)
                                                                                     ..AACAUUUaaUaaaaUaaggaCaaUaCCga
5'- AAATAAAATTATATGAATTAAGTATAAGAACAAGGATTTTTTTTGAAAGGAGAGGACAGTCTTAAGGATTTTTGAGGTAGAGA TTTGAGGG  G    A     A  A GG
5'- AAAUAAAAUUAUAUGAAUUAAGUAUAAGAACAAGGA********GAAAGGAGAGGACAGUCUUAAGGA*****GAGG*AGAGAuUUUGAGGGuuGuuuuAuuuuuAuuAuGGuu
                                                     18                                      17       16       15          14   13      12    11     9
    I   K   L   Y   E   L   S   I   R   T   R   R        K   E   R   T   V   L   R   R      G   E   I   L   R   V   V   L   F   L   L   W   L

     gG3-II (Mc)
 ..UUUaaUaCUaaaaUaAUAAAUAAAUUAA-5'
                                   gG3-I (Mc)
CCC-5                     ..uaaaaUaaUaUaagUUCA-5'
GGGGGACCAGA  A GA    A CATTTA     A  A  A    GAGTTTTCGATATTAAAATAACATAGCTATTATTTTTGTTGTGAAAAGTCAAATAAACAGCT
GGGGGACCAGAuuAuGAuuuuAuCAUUUAuuuuAuuAuAuuuGAGUUUCGAUAUUAAAAUAACAUAGCUAUUAUUUUUGUUGUGAAAAGU(A)n
                8      7     6    5           4    3   2    1
  G   D   Q   I   M   I   L   S   F   I   L   L   Y   L   S   F   R   Y  Stop
```

Fig. 3. The partially pan-edited G3 cryptogene in the LEM125 strain. The genomic sequence and the apparently mature edited sequence are shown, with the editing sites numbered sequentially. See Figure 1A for symbols and nomenclature. The putative mis-edited sequence of the G3 transcript is indicated by double underlining. The open box indicates an unedited possible domain connection sequence. The three identified gRNAs are shown above the mRNA sequence. The arrowheads indicate the sequence overlap with ND1 at the 3′ end and MURF1 at the 5′ end.

A

[Figure 4A: sequence alignment of the G4 cryptogene with guide RNAs gG4-I (mc) to gG4-XIV(mc)]

B

```
L.t.   1 MYNCKSVVVFGFFMYYFVCCIFYFCFFFCLFDLCVMVYFIYIAFVCLFVLVCYVVILLLCFDCIVFFLFTVIIIFFFLFIFWVFIFLLLFWFVFGLFFF 100
T.b.   1 .............MYSLWYILLLFCFLFLLFVCVGMFYLFCYSCLFFFVVLGCDFLLVF.WLYSLFFLWRYNFVYFFFLFCFVFFVLLFLFGLFGFFLYF 86

L.t. 101 FYVCVLIFYFEFLFMLFFVFCGFLLFVMFILFFISFFVLIVLVFCWLLFIFVFRFFCMTFCILILIGLF 169
T.b.  87 LL..CFVLFFDLFFMLFFVLGGFFVFVFFFCLCLFLFVVVVILLVWLLLLFVYRFIYMRFLF....... 146
```

Fig. 4. The pan-edited G4 cryptogene in the LEM125 strain. (**A**) See Figure 1A for symbols and nomenclature. The N-terminal amino acid sequence not present in *T.brucei* is underlined, and the methionine initiation codon at site 107 which is also present in *T.brucei* is indicated by M.
(**B**) Alignment of the predicted amino acid sequences of the edited G4 mRNAs from *L.tarentolae* LEM125 and *T.brucei* (Corell *et al.*, 1994). The N-terminal extended amino acid sequence of LEM125 is indicated by bold type and underlining.

LEM125 G4

A consensus sequence was obtained from the sequences of six partially edited cDNAs and 16 clones containing the 5′ end of the edited RNA (data not shown). The mature edited LEM125 G4 RNA sequence is 573 nucleotides in length and is edited by the addition of 326 uridines in 110 sites and the deletion of five uridines in two sites (Figure 4A). Two in-frame AUG methionine putative translation initiation codons were observed. One is located in unedited 5′ sequence and the other is created by editing in site 107. The latter aligns with the observed AUG codon in *T.brucei* (Corell *et al.*, 1994). A predicted stop codon (UAA) is present in the 3′ unedited sequence. The predicted ORF from mature edited G4 RNA, assuming that the first AUG is the initiator, is 169 amino acids long (Figure 4B) and shows a limited similarity with the ORF derived from the fully edited CR4 RNA from bloodstream *T.brucei* (Corell *et al.*, 1994); the alignment has a Z value of 4.1 SD units, which is at the limit of statistical significance (Kanehisa, 1982). No significant homology with any other database sequence was observed.

The fully edited LEM125 G4 RNA has a 38 nucleotide 5′ unedited sequence. Neither the 5′ end nor the 3′ end of G4 RNA overlap with the transcription units of the adjacent ND4 and COI genes (Simpson *et al.*, 1987).

LEM125 G5

A consensus edited sequence was determined from the sequences of 11 partially edited cDNAs and 17 clones containing the 5′ end of the edited RNA (data not shown). The LEM125 mature edited G5 sequence is 411 nucleotides long with 67 Us added in 71 sites and five Us deleted in three sites (Figure 5A). An ORF of 115 amino acids is created, but the first AuG methionine codon is located at editing site 58, which is 26 amino acid residues downstream from the first in-frame amino acid (Figure 5B).

A

```
                                                                                                   gND3-IX (mc)
                                                              ..uuuuuuuuuaCaaUAUaaagagaCUagCgagaGUaaCaaCaaaUaCUaaaaGC-5'
     vvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvvv       :::|||||:||||:.||||:|:||||:|:|:|||||||||||||||||||||
5' -TCACTCAAAAAATTCACGCATTTATAATTTATATAGGTCGATAGGTATTAGTTTGGGAGGAGGGAAAAAG G  GCA       GA  G    CG  G  G   A GA    CG
5' -UCACUCAAAAAAUUCAAGCAUUUAUAAUUUAUAUAGGUCGAUAGGUAUUAGUUUGGGAGGAGGGAAAAAGuGuuGCAuuuuuuuGAuuGuuuuCGuuGuuGuuuAuGAuuuuCGu
                                                                       73 72     71      70  69   68 67 66      64 63
                                                             F  G  R  R  E  K  V  L  H  F  F  D  C  F  R  C  C  L  W  F  S  C

                                                                                                  ..aUagCUaAAUUaaaaUaggagUUAU
                                                                          gND3-VII(mc)            |||:|:|||||||||||||::|:||||
                                                          ..uugagaUaaUgCgagaCaaUgUAAAAUaaGG——UaAUAACCAGAUcu-5'
                                                            |::|:|||||:|:|:||||||::||||||||||    |||||||||:||.
G GG   G A GACA GAA  CACGA   G    ATTG   C A AGA   A G TTCGCAG    A  A G    G  A GTTTTA  CCTTTA  A  GG TTAA    A     AATA
GuGGuuuGuAuGACAuGAAuuCACGAuuuGuuuA*UGuuuCuAuAGAuuuAuGuUUCGCAGuuuuAuuAuGuuuuGuuAuGUUUUAuuCC***AuuAuuGGuUUAAuuuuAuuuuuAAUA
62   61  59   58    57       56  55 54   53 52      50   48          47  46 45  44   43 42        41  40  39 38  37     36     35
 G   L  Y  D  M  N  S  R F  V  Y   V  S  I  D  L  C  F  A   V  L  L  C  F  V  M  F  Y  S    I  I  G  L  I  L  F  L  I

                                                                                          gND3-II (mc)
                                                                         ..CACaaaUUgUaggCUggGGCaCgUGACAAACAAA-5'
  gND3-V (mc)                                      gND3-III(mc)               ||:|||||:||::||::|:|||:||||||||||
CgaCaCCaCaaCaUauagUuCaC-5'          ..UaUaUUaaggUgaaaCgUagagaCaaUaaaaCCUCaU-A-CAUAGAUcu-5'            ..aUaaaUaaagaUaaaggaa
|:|||||||||||||| ||| |||              :|:|::||::::||||:||:|:||||||||||||||| | ||||:||.                  |:|||:|||:|||||::||
G  G GG G  G A    A G G AAA  G        G G GG    G     G A     G  A     GGAG ATTTGTA   AA A   GA  CTG G AC G    G     A
GuuGuGGuGuuGuAuuuuAuGuGuAAAuuGuuuuuuGuGuGGuuuuGuuuuGuAuuuuuGuuAuuuuGGAGuA*UUGUAuuuAAuAuuuGAuuCUGuGuACuGuuuGuuuuuAuuuuuuu
34  33 32  31    29  28  26   25   24   23      21    20   19   18   17   16     15 14     13     11    10   9 8 7  6   5     4
V  V  V  L  Y  F  M  C  K  L  F  F  V  W  F  C  F  V  F  L  L  F  W  S  I   V  F  N  I  W  F  C  V  L  F  V  F  I  F  F

  gND3-I (Mc)
aaaUgaUUgaaaaUUaUUUAUCACAU-5'
||||:|||:||||||||||||||||
  A  AA    AA AAATAGTGTATTCTCTTATTTAATACTAAAATAATAAATAAATTAAATAAAAAAA
uuuAuuAAuuuuuAAuAAAUAGUGUAUUCUCUUAUUUAAUACUAAAAUAAUAAAUAAAUU(A)n
4   3    2    1
 L  L  I  L  I  N  S  V  F  S  Y  L  I  L  K  Stop
```

B

```
L. t.    1 ..............................FGRREKVLHFFDCFRCCLWFSCGLYDMNSRFVYVSIDLCFAVLLCFVMFYSIIGLILFLIVVVLYFMCK 69
                                         :.    :|||||||||||||||||||||||||:|: .:||||.   | ||::  |||.::||::|:||:::|
T. b.    1 KKSSPFYFSLLSLFLYLFLILWFIILFIGFFLCFLCFLLHFFDCFRCCLWFSCGLYDMNSRLVFYTLDLCFVSCLFFVLLNSIICVLLFVFVIVLFYFCY 100


L. t.   70 LFFVWFCFVFLLFWSIVF....NIWFCVLFVFIFFLLILINSVFSYLILK 115
            |:.:: :.|::  ::|:    ::::|.:::|.:: ::|. . :   .:|
T. b.  101 GFLFLWFLFFVVCIGFVWYFWDHVYLCGVILFCLWCFLLYYTYYINKYIK 150
```

Fig. 5. The partially pan-edited G5 (ND3) cryptogene in the LEM125 strain. (**A**) See Figure 1A for symbols and nomenclature. The putative unedited 5′ sequence of G5 is indicated by double underlining and bold type. The initial methionine of the predicted amino acid sequence is indicated by bold type and underlining. The 5′ end of the pan-edited G6 (RPS12) cryptogene (Maslov *et al.*, 1992) is indicated by an arrow. The arrowheads indicate the sequence overlap with G6 (RPS12) at the 5′ end. (**B**) The alignment of the predicted amino acid sequences of the edited G5 (ND3) mRNAs of *L.tarentolae* LEM125 and *T.brucei* (Read *et al.*, 1994) is shown. The sequence in LEM125 which corresponds to the potentially unedited sequence is indicated by double underlining. The longer amino acid sequence reported for the edited G5 (CR5-ND3) mRNA in *T.brucei* (Read *et al.*, 1994) is indicated by bold type and underlining. The first methionine in the LEM125 sequence is indicated in bold type.

The ORF from the edited G5 sequence has significant similarity to the ORF derived from the CR5 sequence of *T.brucei* (Read *et al.*, 1994); the alignment has 41% amino acid matches and a Z value of 10.6 SD units (Kanehisa, 1982). Homology searches against the databases yielded limited similarity of the G5 amino acid sequence with NADH dehydrogenase subunit 3 (ND3) sequences, as was reported previously for the CR5 amino acid sequence from *T.brucei* (Read *et al.*, 1994).

The 5′ end of the G5 transcript overlaps 51 nucleotides with the 5′ end of the RPS12 transcript (Maslov *et al.*, 1992), which is transcribed off the opposite strand.

Partially edited G5 RNAs from the UC strain showed a limited consensus sequence identical to the 3′ end of the editing domain in edited LEM125 G5 RNA (Figure 9). A maxicircle-encoded gRNA (gND3-I) that could guide the editing events in this block was detected by a computer search of the known maxicircle sequence, and is present in both LEM125 and UC strains (Figure 7). The presence of this maxicircle-encoded gRNA in the UC strain could account for the correct editing of block I. We show below that the adjacent upstream gRNA in LEM125 is minicircle-encoded and is absent in the UC strain cells due to a lack of this minicircle class, thereby accounting for the lack of correct upstream editing in partially edited G5 RNAs from the UC strain (Figure 9).

Edited transcripts of G1–G5 are detected in LEM125 cells by Northern hybridization analysis and are not detected in UC cells

The presence of mature edited transcripts for G1–G5 in the UC and LEM125 strains was tested by Northern blot hybridization analysis using as probes oligonucleotides antisense to the edited sequences (Figure 6). Edited transcripts can be detected in kRNA from LEM125 cells and not in kRNA from UC cells. The weak signal in the G4 lane migrating ahead of the edited RNA position probably represents non-specific hybridization. As a control, edited RPS12 RNAs were detected in kRNA from both strains.

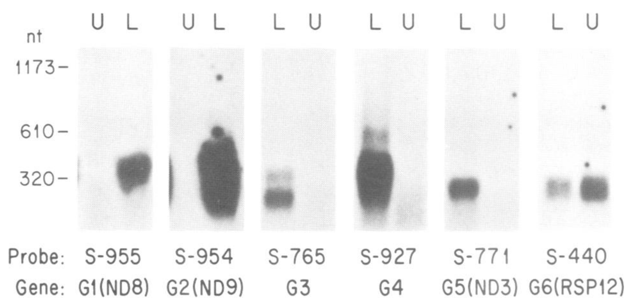

Fig. 6. Northern analysis of edited G1−G6 transcripts from the LEM125 and the UC strains. Total kRNA was electrophoresed in formaldehyde−agarose, blotted and hybridized with labeled oligonucleotide probes for edited mRNAs. L, LEM125 strain; U, UC strain. The RPS12 hybridization was included as a control, since edited transcripts are known to be present in both strains (Maslov *et al.*, 1992).

Identification of 30 new minicircle-encoded gRNAs in LEM125 kRNA that are not present in UC kRNA

A previous analysis of a minicircle DNA library from the UC strain provided evidence for the presence of 16 minicircle-encoded gRNAs (Maslov and Simpson, 1992). In the present study, gRNA libraries were constructed for both the UC and LEM125 strains. The previously identified 16 gRNAs were readily detected in the UC gRNA library, indicating that the two methods are equivalent in sensitivity (data not shown). In this study we performed a gRNA complexity analysis for the LEM125 strain based on the gRNA library and confirmed the presence of the gRNA genes in minicircle DNA by Southern hybridization (data not shown).

As an indication of the completeness of the LEM125 gRNA library, several known UC strain gRNAs (Sturm and Simpson, 1990; Maslov and Simpson, 1992)—gMURF4-I, gMURF4-III, gRPS12-II, gRPS12-III, gRPS12-VII, gRPS12-VIII, gCOIII-I and gCOUIII-II—were identified by random selection and sequencing (data not shown). Oligonucleotide probes for the known gRNAs of the UC strain were then used for negative selection of the LEM125 library. A total of 386 clones which did not hybridize with these probes was sequenced, yielding 32 minicircle-encoded gRNA sequences for G1−G5 editing (from 192 clones) and two maxicircle-encoded gRNA sequences for G1−G5 editing (gG4-IV and gND3-I) (Figures 1A−5A). These results are summarized in Table I. The remainder of the sequenced clones consisted of primer-dimer sequences (117 clones), and sequences of unknown origin (77 clones). Of these 'unknown' sequences, 38 clones were found by computer analysis to represent 17 distinct maxicircle transcripts, from various regions.

Most of these gRNAs were detected in LEM125 kRNA by Northern analysis (Figure 7). The concentration of several gRNAs—gND8-III, gND8-IX, gND9-II, gND9-V, gG3-II, gG4-I, gG4-IV, gND3-II and gND3-V RNAs—was too low to be unambiguously detected by Northern analysis, and their presence was confirmed by primer extension sequencing (data not shown). None of these minicircle-encoded gRNAs was detected in UC kRNA (Figure 7), with the exceptions of gG4-III (= gLt19) and gND3-IX (= gB4).

The transcription origin of these gRNAs was investigated by Southern blot analysis of total *Msp*I-digested kDNA from both strains (data not shown). This analysis showed that, with the exception of gG4-III (gLt19) and gND3-IX (gB4), the 30 new minicircle-encoded gRNAs identified are transcribed from LEM125-specific minicircles which are not present in UC strain kDNA.

Identification of the editing roles of two previously 'unassigned' gRNAs

There are two gRNA transcripts in UC kRNA, gLt19 and gB4 (Sturm and Simpson, 1991), to which we were previously not able to assign editing roles, although the minicircles that encode these transcripts are the highest abundance minicircles in UC kDNA (Kidane *et al.*, 1984; Maslov and Simpson, 1992). The homologue of the gLt19 RNA in LEM125 kRNA is gG4-III, which edits block III of the G4 transcript. In LEM125, the gG4-III RNA has the identical sequence to the gLt19 RNA in the UC strain, although in the latter an 18 nucleotide 3′ extension was present (Figure 8). This 3′ extension in Lt19 would have no guiding function for editing of the G4 mRNA. The gG4-III RNA in LEM125 cells is ~12 times less abundant in steady state kRNA than the non-functional gLt19 RNA in UC cells (Figure 7). The gG4-III minicircle sequence class was also less abundant in the LEM125 kDNA network (113 copies per network of 10 000 minicircles) than the Lt19 minicircle in the UC kDNA network (2500 copies per network) (Maslov and Simpson, 1992).

The homologue of the UC strain gB4 RNA is gND3-IX, which edits the ninth block of the ND3 (G5) transcript. The gND3-IX RNA is approximately four times more abundant in steady state kRNA from LEM125 cells than the gB4 RNA in kRNA from UC cells (Figure 7). The gND3-IX minicircle template in the LEM125 cells is present at ~1750 copies per network as compared with the 2980 copies per network for the B4 minicircle in the UC cells (data not shown).

Analysis of the complete sequences of the LEM125 minicircles encoding the gG4-III and gND3-IX genes showed total identity with the sequences of the Lt19 and B4 minicircles from the UC strain (data not shown). Similar results were obtained for all homologous minicircle sequence classes present in both strains (data not shown). The complete minicircle sequences will be published elsewhere.

New maxicircle-encoded gRNAs

By a computer search of the maxicircle genome with the LEM125 G1−G5 mature edited RNA sequences, five additional putative maxicircle-encoded gRNAs were identified (gND9-XIV, gG3-I, gG3-II, gG4-IV and gND3-I) as shown in Figures 2A−5A. The existence of these gRNA transcripts in both LEM125 kRNA and UC kRNA was confirmed by Northern blot analysis and primer extension sequencing (data not shown).

Several maxicircle-encoded gRNAs (gRPS12-VI, gG4-IV, gND3-I, gCYb-I and gMURF2-II) were detected in the 421 sequenced clones from the LEM125 gRNA library. In one case, a maxicircle sequence encoding a potential gRNA overlapping editing sites 111–116 of ND8 (gND8-XIII) was identified by computer analysis, but no evidence for the existence of this gRNA was obtained by Northern hybridization or primer extension sequencing.

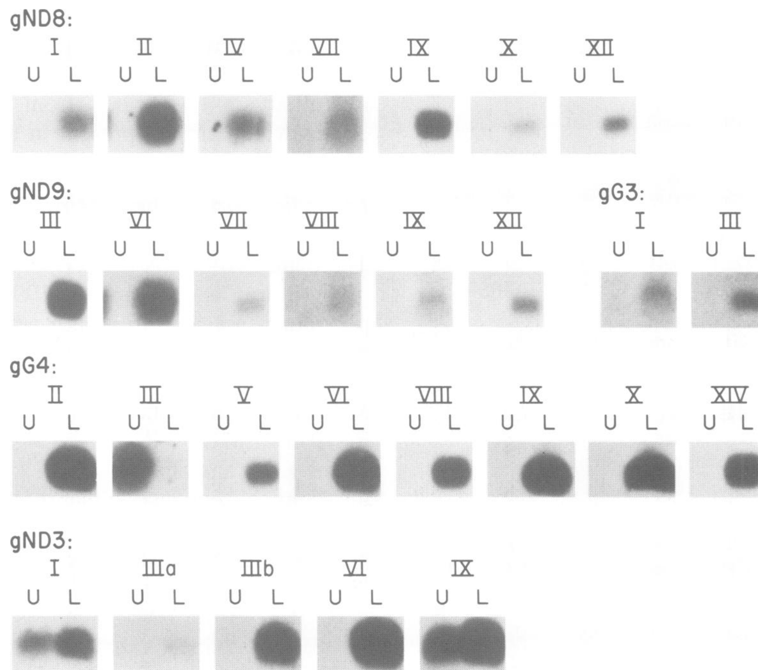

Fig. 7. Northern blot hybridization analysis of the presence of gRNAs for editing of G1–G5 in UC and LEM125 strain kinetoplast RNA. Total kinetoplast RNA was electrophoresed in formaldehyde–agarose, blotted and hybridized with labeled oligonucleotide probes for each gRNA. U, UC strain; L, LEM125 strain.

```
gLt19 (UC)       (5') cuuaaauaacauccagcaaaacgcaaugugguugaagugaaaaauauauaucucuacuguaaauuaaag... (3')
                      ||||||||||||||||||||||||||||||||||||||||||||||||||
gG4-III (LEM125) (5') cuuaaauaacauccagcaaaacgcaaugugguugaagugaaaaauauauau... (3')
```

Fig. 8. The gLt19 RNA from the UC strain has the identical sequence to the gG4-III RNA from the LEM125 strain, but has an extra 18 nucleotides at the 3′ end (underlined).

Mis-editing of G5 transcripts in UC strain

The maxicircle-encoded gRNA transcript involved in editing the first block of ND3 (G5), gND3-I, was detected in both strains (Figure 7). This is consistent with the observed proper editing of block I of G5 RNA in the UC strain, as mentioned above. Since the adjacent upstream gRNA is minicircle-encoded in LEM125 and is absent in the UC strain, the occurrence of incorrect editing in UC G5 RNAs upstream of block I is explainable in terms of the mis-editing hypothesis of Sturm *et al.* (1992), as shown diagrammatically in Figure 9A. Two specific examples of mis-editing by non-cognate gRNAs are shown in Figure 9B for one cDNA sequence. A maxicircle-encoded gRNA which normally mediates editing of block IV of G4 can hybridize to the correctly edited block I sequence, and this gRNA in turn creates an anchor sequence for a gRNA which normally mediates the editing of block IV of MURF4. Additional examples of G5 mis-editing are presented elsewhere (Maslov *et al.*, 1994a).

Redundant gRNAs also occur in Leishmania tarentolae

The presence of extensively overlapping gRNAs is a feature of RNA editing in *T.brucei* (Corell *et al.*, 1993; Riley *et al.*, 1994), presumably due to the extremely large minicircle-encoded gRNA repertoire in that species. Such 'redundant' gRNAs have not previously been observed in an analysis of the complete gRNA repertoire of the UC strain of *L.tarentolae* (Maslov and Simpson, 1992). However, one example of two redundant gRNA sequences occurs in the LEM125 strain: gND3-IIIa and gND3-IIIb (Figure 10). As shown in the Northern analysis in Figure 7, both of the gRNAs are present in LEM125 kRNA and are absent in UC kRNA. The gND3-IIIa RNA has a relatively lower abundance compared with the gND3-IIIb RNA (Figure 7).

Discussion

We have obtained information on the mechanism of the loss of RNA editing capacity that occurs during prolonged culture of kinetoplastid protozoa. This was accomplished by a comparative analysis of RNA editing in two strains of *L.tarentolae* which differ in the geographical origin and length of time maintained in culture. We have shown that apparently productive pan-editing of the transcripts of the G1–G5 cryptogenes occurs in the recently isolated LEM125 strain (Wallbanks *et al.*, 1985) and not in the UC strain which has been maintained in culture for over 55 years (Parrot and Foley, 1939).

There is no evidence that these strains are isogenic, and in fact they were previously shown to have polymorphisms in the divergent region of the maxicircle and to differ in several chromosomal bands on pulsed field gel analysis (Gomez-Eichelmann *et al.*, 1988). However, the divergent region is known to undergo rapid changes (Muhich *et al.*, 1985) and chromosomal polymorphisms have also been documented to occur in cultures of *L.tarentolae* (Rovai

Fig. 9. Mis-editing of G5 transcripts in UC strain. (**A**) Diagrammatic representation of correct editing of block I in both strains by a maxicircle-encoded gRNA (gND3-I) and incorrect editing of the upstream sequences in the UC strain due to an absence of the cognate minicircle-encoded gRNA(s). (**B**) Example of one partially edited G5 RNA from the UC strain which has correct editing of block I and mis-editing of upstream sequence due to the sequential actions of two non-cognate, misguiding gRNAs. The correctly edited block I is boxed, and the gRNAs are shown above and below the edited RNA sequences. The mature edited LEM125 G5 sequence and the cognate gRNAs are shown for comparison. (LEM) LEM125 edited G5, (UC) UC mis-edited G5.

A

```
                         gND3-IIIb (mc)
3'-..UaUaUUaaggUgaaaCgUagagaCaaUaaaaCCUCaU-A-CAUAGAUcu -5'
3'-..UaUaUUaaagUgaaaCaUaaaagCgaUaaaaCCUCaU-G-CGUAAAUCu -5'
                         gND3-IIIa (mc)
```

B

```
                         gND3-IIIa (mc)
   ..UaUaUUaaagUgaaaCaUaaaagCgaUaaaaCCUCaU-G-CGUAAAUCU -5'
5'...AuuGuuuuuuGuGuGGuuuuGuuuuGuAuuuuuGuuAuuuuGGAGuA*UUGUAuuuAAuAuuuGAuuC...
      25   24   23    21   20   19   18   17   16     15 14    13    11    10
```

Fig. 10. Redundant gRNAs in *L.tarentolae* LEM125. A second gRNA (gND3-IIIa) for the editing of block III of G5 (ND3) is shown. (**A**) Alignment of the two redundant gRNAs, showing the difference in primary sequences. The anchor sequences of both gRNAs are indicated by underlined bold characters. (**B**) Alignment of the gND3-IIIa RNA with the cognate edited mRNA sequence, showing that it contains identical editing information to gND3-IIIb.

et al., 1992). However, the fact that the normally highly polymorphic (Simpson *et al.*, 1987) G-rich maxicircle sequences, G1−G5, are nearly identical in sequence, and the fact that homologous minicircle classes in both strains share perfect sequence identity provide evidence for the close relatedness of these strains.

Editing of the G3 and G5 transcripts in the LEM125 strain may, however, not be complete. In the case of G3, there is an unusual pattern of deletions at the 5′ end of the editing domain encompassing an unedited G-rich sequence. This may represent a situation in which one or more 5′ terminal gRNAs have already been lost during the relatively short culture history of the LEM125 strain, or are not present in this stage of the life cycle. Likewise, the 5′ terminal portion of edited G5 RNA includes an unedited G-rich sequence, and neither an in-frame AUG methionine nor an AUA isoleucine putative translation initiation codon is created by the identified editing events.

The major defect in the UC strain appears to be in the minicircle-encoded gRNA repertoire, as a result of the

loss of several minicircle sequence classes. At least 30 minicircle-encoded gRNA genes for the editing of G1–G5, which are present in the LEM125 strain, are absent in the UC strain (Table I), producing a set of 'pseudo-cryptogenes' (Maslov and Simpson, 1992), the transcripts of which are not productively edited into translatable mRNAs. A complete set of overlapping gRNAs has not been obtained for any of the five cryptogenes in LEM125, such as has been found for the MURF4 and RPS12 cryptogenes in the UC strain (Maslov and Simpson, 1992). But sufficient gRNAs have been identified for each gene to make it likely that a more extensive search could result in the detection of complete sets of overlapping gRNAs, to account for the observed mature edited sequences of G1–G5.

The discovery of two completely overlapping gRNAs mediating the editing of block III of G5 in the LEM125 strain indicates that gRNA redundancy is not limited to the salivarian trypanosomes (Corell *et al.*, 1993; Riley *et al.*, 1994). The extent of this phenomenon, however, is greater in the trypanosomes. In fact, extensive gRNA redundancy in *L.tarentolae* would have limited or even prevented the disruption in editing caused by the loss of specific gRNAs, as apparently has occurred in the UC strain.

Maxicircle-encoded gRNA genes are present and transcribed in both strains, leading to the situation observed for G5 shown in Figure 9, in which the first editing block is correctly edited in both strains due to the presence of a maxicircle-encoded gRNA (gND3-I). The subsequent upstream blocks are edited correctly in the LEM125 strain but are mis-edited in the UC strain due to the lack of the adjacent minicircle-encoded gRNAs (Figure 9, and Maslov *et al.*, 1994a). We speculate that the missing minicircle sequence classes in the UC strain, which encode G1–G5 gRNAs, were rendered functionless in editing and were lost by mis-segregation at kinetoplast division. This would provide a mechanism for the loss of minicircles for the editing of G1–G5 that seems to have occurred in the UC strain.

However, two minicircle sequence classes containing genes for such non-functional gRNAs are still present in the UC strain: Lt19 and B4 (Maslov and Simpson, 1992). It is interesting that these minicircles represent the most abundant sequence classes in the UC strain, whereas the Lt19 homologue (the gG4-III minicircle) in LEM125 is 22-fold less abundant and the B4 homologue (the gND3-IX minicircle) is approximately half as abundant as in the UC strain. Furthermore, the gLt19 RNA from the UC strain has an additional transcribed 18 nucleotides at the 3′ end (Figure 8) which could not mediate correct editing, suggesting that 3′ end processing has been affected in this gRNA, possibly as a result of the lack of an editing function.

The striking differences in relative abundance of specific sequence classes in the two strains may be a result of the apparently stochastic nature of the segregation of daughter minicircles in dividing cells. Selective amplification of specific sequence classes, as is suggested to occur when cells are subjected to stress ('transkinetoplastidy') (Lee *et al.*, 1992, 1994), may also give rise to differences in minicircle copy numbers.

The hypothesis for the loss of minicircle-encoded gRNAs during prolonged culture implies that the *L.tarentolae* cells in culture, which presumably represent the insect stage of the life cycle, do not require functional G1–G5 protein products. The three identified protein products—ND8 (G1), ND9 (G2) and ND3 (G5) (Souza *et al.*, 1992, 1993; Read *et al.*, 1994)—represent components of complex I of the respiratory chain. An absence of complex I involvement in the respiratory chain has been reported for procyclic *T.brucei* (Hill and Cross, 1973), and for culture forms of *Trypanosoma cruzi* (Denicola-Seoane *et al.*, 1992) and *C.fasciculata* (Sloof *et al.*, 1994). In addition, the complete editing of ND7, ND8 and ND9 maxicircle transcripts in *T.brucei* (Souza *et al.*, 1992, 1993; Read *et al.*, 1994) is limited to bloodstream forms which lack complex IV respiration. We have preliminary evidence that the respiration of UC strain *L.tarentolae* cells is completely insensitive to the complex I inhibitor, rotenone, and that the respiration of LEM125 cells is somewhat sensitive to this inhibitor (data not shown), suggesting a partial involvement of complex I respiration in the latter. It is entirely possible that editing of the complex I subunits is also regulated in *Leishmania*, and that cells lacking the ability to edit these mRNAs due to loss of gRNA genes will be incapable of completing the life cycle in the animal host. In order to study this question in the case of *L.tarentolae*, it will be necessary to reproduce the entire life cycle in the laboratory, which has not yet been achieved.

In conclusion, we have shown that RNA editing in *L.tarentolae* is a labile genetic trait which can be readily disrupted by prolonged culture. This finding has implications for the loss of the capacity to go through the entire parasitic life cycle that frequently occurs with cultured trypanosomatids. It may also shed light on the process by which pan-edited cryptogenes are thought to have been substituted with partially edited or fully edited mRNAs several times in the evolution of the kinetoplastids (Landweber, 1992; Maslov *et al.*, 1994; Simpson and Maslov, 1994).

Materials and methods

Cell cultivation, isolation of mitochondria and nucleic acid isolation

LEM125 represents one of a series of stocks of *L.tarentolae* isolated in 1985 from geckos and in 1982 from a sandfly in southern France by Rioux (Wallbanks *et al.*, 1985). We obtained frozen stabilates of these stocks from Dr G.Holtz in 1987. Five of these stocks (LEM87, LEM115, LEM124, LEM125 and LEM306) were analysed and shown to have similar kinetoplast DNA restriction digest profiles in acrylamide gels, and were therefore classified as comprising schizodeme B of *L.tarentolae*, as compared with schizodeme A, which contains several derivatives (UC, T and K strains) of the *L.tarentolae* TarII strain, a 1939 Algerian gecko isolate of Parrot (Parrot and Foley, 1939; Gomez-Eichelmann *et al.*, 1988). The schizodeme A and schizodeme B strains also differ in having different nutritional requirements, several polymorphisms in the divergent region of the maxicircle and a few polymorphisms in chromosome profiles on pulsed field gels, but are basically very similar (Gomez-Eichelmann *et al.*, 1988).

To obtain a clonal line from the LEM125 stock and avoid the possible problem of population heterogeneity (Gomez-Eichelmann *et al.*, 1988), cells were plated onto 0.8% agar plates containing BHI (Difco), 10 μg/ml hemin, 5% heat inactivated calf serum and 1× BME amino acids supplement (BRL). Visible colonies appeared after several days and several were selected for re-cloning. One re-cloned line was selected for this project. Cells were maintained frozen in 10% glycerol and used to initiate cultures on an approximately bimonthly basis.

The UC strain cells were cultivated in brain heart infusion medium (BHI, Difco) supplemented with 10 μg/ml hemin (Braly *et al.*, 1974). The LEM125 cells were cultivated in BHI-hemin supplemented with 5% fetal calf serum (Gemini Bioproducts, inactivated by heating for 40 min at 54°C) and 1× BME amino acids (BRL).

Mitochondrial fractions were prepared from mid- to late-log phase cells by flotation in Renografin density gradients as described (Braly *et al.*, 1974; Simpson, 1979). kRNA was extracted from purified mitochondria (Simpson and Simpson, 1978), and kinetoplast DNA networks were purified from stationary cell cultures by sedimentation through a caesium chloride step gradient (Simpson, 1979).

Oligonucleotides

Oligonucleotide primers for PCR amplification, hybridization and primer extension assays were synthesized by standard phosphoramidite methods and purified by thin layer chromatography. The following oligonucleotides were utilized in this study.

G-rich region-specific, unedited primers. The corresponding positions in the *L.tarentolae* maxicircle sequence (GenBank entry LEIKPMAX) are given in parentheses. Underlined sequences represent the added 5′ restriction sites.

S-662: GTTTTCGGAAGGAGGGATCGGACC (ND8, 2353–2377);
S-360: GCCGAATTCTAATCAAATATACTTC (ND8, 2640–2624) *Eco*RI;
S-365: GGCGAATTCCTTTTGCAAATTGGAC (ND9, 2684–2700) *Eco*RI;
S-297: TAAGGGATCCGATGTTTAAGTTTCGG (ND9, 2960–2985) *Bam*HI;
S-663: GAAAGGAGAGGACAGTCTTAAGG (G3, 8459–8481);
S-361: GCCGAATTCTTTTATATGTGTGTAC (G3, 8614–8598) *Eco*RI;
S-364: CGGGAATTCTTTCCGTCTCGCCTT (G4, 12888–12903) *Eco*RI;
S-664: TTTTAATTTAATCTTATTTTTATACTTTG (G4, 13131–13102);
S-665: ATTTGTATTTTATAACAAGCATAATAG (G5, 14401–14428);
S-314: TATGGATCCACGCATTTATAATTTAT (G5, 14683–14665) *Bam*HI.

G-rich region-specific, edited primers.

S-726: CTAAACAAAACACAAATAAAAATC ND8
S-764: GGCAATACAACAACAAAATAACAAAAAAC ND8
S-910: AATATACTTCTCGTGATACAC ND8
S-919: TGCGTAATGGCGTCTGTTGG ND8
S-928: CATGGCAATACAACAACAAAATAAC ND8
S-955: GCATACAAAAACCACAATAAATCC ND8
S-969: CdUACdUACdUACdUATGCGTAATGGCGTCTGTTGG ND8
S-767: AAAAACTATATTGCAAAATTGGAAAAACACAA ND9
S-768: CAACAATATCTTCACATAACATACAAAAATTC ND9
S-908: CTTTTTTTTTCCAACAAAAAATCAATC ND9
S-911: TTAATTTTATAGACAGATTCAT ND9
S-918: CAAAACCAAACAAATCAAAAAC ND9
S-953: ATTAATTTTATAGACAGATTCATAAC ND9
S-954: CGCATAACTACAAAAAAATAAACTAAAAC ND9
S-989: ATTAAAAAATAAAAACATCAC ND9
S-992: CdUACdUACdUACdUACAAAAAATAAAAAAAAACAC ND9
S-765: CTCAAATATAATAAAATAAATGATAAAATC G3
S-766: TAAAATCATAATCTGGTCCCCCAACCATAA G3
S-909: TTTTAATATCGAAAACTCAAAT G3
S-920: TCATAATCTGGTCCCCCAACCAT G3
S-993: CdUACdUACdUACdUAGATAAAATCATAATCTGGT G3
S-769: AAATTAAAATACAAAACGTCATACAAAAA G4
S-770: GAAACACAAATATAAATAACATCCAAC G4
S-887: TAAAAAATAAAATAAACATCACAAACAAC G4
S-903: CAAAAAACAACATAAACAAAAATTC G4
S-927: CCACAAAAAACAAAAAACAACATAAAC G4
S-952: TTTTAAAATAATCCAATTAAAATTAAAAT G4
S-990: AAACACTATACAATCAAAAC G4
S-1022: AACACACAAATCAAATAAAC G4
S-1042: CACTATACAATCAAAACACAAC G4
S-1043: CdUACdUACdUACdUACAACAAAATCACAACATAAC G4
S-771: AAAAAAAAAATAAAAACAAACAGTACACAG ND3
S-772: TCCAAAATAACAAAAATACAAAACAAAACC ND3
S-881: CTATTTATTAAAAATTAATAAAAAAAAAATAAAAAC ND3
S-882: CdUACdUACdUACdUAAACAGTACACAGAATC ND3
S-886: CACCACAACTATTAAAAATAAAATTAAACC ND3
S-900: AAGAGAATACACTATTTATTAA ND3
S-907: GGAATAAAACATAACAAAAC ND3
S-914: TAAACAAATCGTGAATTCATGTC ND3
S-926: CAACGAAAACAATCAAAAAAATGC. ND3

Guide RNA library primers and oligo(dT) primer.

S-399: CGCGGATCCTTTTTTTTTTTTTTTTTTTTTTTTT *Bam*HI
S-499: AAAAAAAAAAAAAAT
S-502: AAAAAAAAAAAAAAC
S-503: AAAAAAAAAAAAAAG
S-581: GACTCGAGTCGACATCG$(A)_{18}$
S-584: TCGCGAGCTAGCACTAGT$(C)_{12}$
S-708: CAdUCAdUCAdUCAdUTCGCGAGCTAGCACTAGT
S-709: CdUACdUACdUACdUAGACTCGAGTCGACATCGA

Guide RNA-specific primers.
(Mc: maxicircle encoded gRNA; mc: minicircle encoded gRNA.)

S-1089: CATCTATGTTCTATCTAGCT gND8-I (mc)
S-1090: AGATCGTTGTTCTATGCTAT gND8-II (mc)
S-1146: ATGTATTTATACGCACGTAA gND8-III (mc)
S-1091: AGATGACATTCTTTGTTATC gND8-IV (mc)
S-1147: CCACTGTGGTCTTTGTATG gND8-VI (mc)
S-1148: ATTCTCCAATGTGCTC gND8-VII (mc)
S-1092: ATCTTTATCGATGCAACTTA gND8-IX (mc)
S-1093: ATGTTAATATCTGTGTTCAT gND8-X (mc)
S-1094: TATTTACCTCTACATGCACT gND8-XII (mc)
S-1209: ATTCGTTCATAATTTTTATT gND8-XIII (mc)
S-1149: AATTATTCCGTGTCATGTTT gND9-II (mc)
S-1095: ACGTATACTATACAAGACA gND9-III (mc)
S-1150: TATCGATTTCCCTTCACGTA gND9-V (mc)
S-1096: ATAACTATTCACATCACCTC gND9-VI (mc)
S-1097: ACTTGATTTTATTCTCTGTG gND9-VII (mc)
S-1152: ATCATATATTTTCGAACTCG gND9-VIII (mc)
S-1153: TCTAGTAATTATGCGT gND9-IX (mc)
S-1098: TCTATTCTTTAACCTATGAT gND9-XII (mc)
S-1130: TTATATCATAAATAAGCTTT gND9-XIV (Mc)
S-1325: AAATTTTATTATATTC gG3-I (Mc)
S-1326: AAATTAGATTTTATTATTT gG3-II (Mc)
S-1296: TTGTAAATTATTTTATTCCT gG3-III (mc)
S-1154: ACATTTTGCACTCTAATTTT gG4-I (mc)
S-1155: TATCACTTATATCCATGTTT gG4-II (mc)
S-303: TATTTTTCACTTCAACCACA gG4-III/gLt19 (mc)
S-906: TTATTATTATTTATAATAC gG4-IV (Mc)
S-1099: TATACTATCTCTTATCCTCT gG4-V (mc)
S-1156: TTCAACCCTCTACTTTAAAT gG4-VI (mc)
S-1100: TAAATGTTTACCTCTTTGCC gG4-VIII (mc)
S-1157: CTATCTTTTTCCTCTCATTT gG4-IX (mc)
S-1211: AATGCTCTCTCTATTTACCG gG4-X (mc)
S-1101: TATTCTATTCTTGCTTCTTC gG4-XIV (mc)
S-880: TATTTATTTCTATTTCCTTT gND3-I (Mc)
S-1117: CAGCCTACAATTTGTGTAGA gND3-II (mc)
S-1159: ATATAATTTCACTTTGTATTG gND3-III (mc)
S-1102: TATCGATTTAATTTTATCC gND3-V (mc)
S-1160: ACTCTATTACGCTCTGTTAC gND3-VI (mc)
S-313: AATGTTTGTTATATTTCTCT gND3-IX/gB4 (mc).

PCR amplification of partially edited and 5′ edited G1–G5 transcripts

The partially edited and 5′ edited RNAs were PCR-amplified from cDNAs synthesized with purified LEM125 kRNA using the appropriate 3′ primers, as described previously (Maslov *et al.*, 1992). To obtain the 5′ edited segments of the G1–G5 transcripts, cDNA molecules were synthesized with edited primers S-928 (G1-ND8), S-989 (G3-ND9), S-765 (G3), S-1042 (G4) and S-881 (G5-ND3), dG-tailed with TdT as described in the 'Guide RNA library' section, and PCR-amplified with the primers S-584/S-708 and S-969 (G1-ND8), S-992 (G2-ND9), S-993 (G3), S-1043 (G4) and S-882 (G5-ND3). The RT-PCR products were gel-purified and cloned using the CloneAmp system (BRL) and DH5α Library Efficiency Cells (BRL).

DNA sequencing

Plasmid DNA was extracted by a minilysate boiling method (Sambrook *et al.*, 1989) from a 1.5 ml bacterial culture in LB medium supplemented with 100 μg/ml ampicillin. Plasmid DNA was sequenced with the Sequenase version 2.0 kit (USB), using [α-^{32}P]dATP, following the manufacturer's instructions.

5′ mapping of the gRNAs

5′ end-labeled primer extension oligonucleotide (1.5 pmol) and 5 μg kRNA were denatured at 65°C for 5 min and annealed at 40–50°C (according to the T_m of the primer), for 5 min. Elongation–termination reactions were performed at the appropriate temperature for 30 min, using a 3:1 molar ratio of dideoxynucleotides and deoxynucleotides and AMV reverse transcriptase (20 U) (Promega). The extension products were analysed by electrophoresis on 8% polyacrylamide–8 M urea.

Northern and Southern blot analysis

kRNA (5 μg) was fractionated by electrophoresis in formaldehyde–agarose (1.5%) and blotted onto nylon filters (S&S Nytran, 0.2 μm) (Shaw *et al.*, 1988; Blum *et al.*, 1990). The filters were hybridized with 5′ end-labelled oligonucleotides (sp. act. ~10^8 c.p.m./μg).

kDNA (5 μg) digested with *Msp*I was fractionated in 1% agarose and blotted onto a nylon membrane (Magna NT, MSI). The blots were hybridized with 5′ end-labelled oligonucleotides (sp. act. ~10^8 c.p.m./μg).

Guide RNA library

The method of cloning gRNA transcripts from both the UC and LEM125 strains involved cDNA synthesis, addition of a homopolymer tail and PCR amplification. To prime cDNA synthesis on the oligo(U)-tailed gRNA, 250 pmol of the oligonucleotides, S-499, S-502 and S-503, were annealed to gel-isolated gRNA for 5–10 min at 65°C and 15 min at 4°C. cDNA synthesis was performed with SuperScript II RNase H^- reverse transcriptase (BRL) (1000 U) for 30 min at 12°C and 1 h at 37°C. The cDNA products were purified by electrophoresis on 15% polyacrylamide–8 M urea gels, phenol–chloroform-extracted and ethanol-precipitated. The terminal deoxynucleotidyl transferase (TdT) reaction was performed with 25 U of TdT (Boehringer), 1 μM dGTP for 30 min at 37°C. The T-tailed DNA was phenol–chloroform-extracted, and purified through BioGel P-4 (Bio-Rad) spin columns in 10 mM Tris–HCl, pH 8.0, 0.1 mM EDTA. The PCR reaction mixture for the UC cDNA contained 2 pmol of primers S-581 and S-584, 200 pmol of primers S-708 and S-709, 20 mM Tris–HCl, pH 8.3, 1.5 mM $MgCl_2$, 25 mM KCl, 0.05% Tween 20, 100 μg/ml bovine serum albumin, 50 μM of each dNTP and 5 U of AmpliTaq DNA polymerase. The PCR conditions were 5 min at 95°C, followed by 30 cycles at 95°C for 1 min, 50°C for 1 min and 72°C for 1 min.

The gRNA library from LEM125 was PCR-amplified using the 'Hot-Start' PCR technique. The lower mixture contained 5 μl 10× PCR buffer (50 mM KCl, 10 mM Tris–HCl, pH 8.3, 5 mM $MgCl_2$), 10 μl 10× dNTP mixture (2.5 mM each dNTP) and the primers as described above, in a final volume of 50 μl. The upper mixture contained: 5 μl 10× PCR buffer, 1 μl AmpliTaq DNA polymerase (Boehringer, 5 U/ml) and the dG-tailed cDNA, in a final volume of 50 μl. The amplification regime was as follows: 94°C, 5 min; five cycles of 94°C, 30 s, 25°C, 1 min and 65°C, 2 min; and 30 cycles of 94°C, 30 s, 50°C, 1 min and 72°C, 1 min. Gel-purified PCR products were cloned using the CloneAmp system (BRL) and DH5α Library Efficiency Cells (BRL).

In a control experiment, the gRNA library prepared from UC kRNA was screened with a mixture of oligonucleotides specific to 16 gRNA species already identified in that strain: eight gRNAs for RPS12 mRNA, six for MURF4 and two for COIII (Maslov and Simpson, 1992). Fifty positive clones were selected at random and sequenced. Fourteen of these gRNA species were identified in this selection, represented by one to 10 clones each, suggesting that this gRNA library was complete.

Acknowledgements

We would like to thank all members of the Simpson laboratory for advice and discussion, Betina Niner for experimental assistance in the early stages of this project, and Anthea J.Ramos for technical assistance. This research was supported in part by grant AI-09102 from the National Institutes of Health to L.S. O.H.T. was supported by a predoctoral fellowship from the Conselho Nacional de Desenvolvimento Cientifco e Tecnologico (CNPq, Brasil). D.A.M. is on leave from Moscow State University, Russia.

References

Abraham,J., Feagin,J. and Stuart,K. (1988) *Cell*, **55**, 267–272.
Benne,R. (1994) *Eur. J. Biochem.*, **221**, 9–23.
Blum,B., Bakalara,N. and Simpson,L. (1990) *Cell*, **60**, 189–198.
Braly,P., Simpson,L. and Kretzer,F. (1974) *J. Protozool.*, **21**, 782–790.
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.
Corell,R.A., Myler,P. and Stuart,K. (1994) *Mol. Biochem. Parasitol.*, **64**, 65–74.
Decker,C.J. and Sollner-Webb,B. (1990) *Cell*, **61**, 1001–1011.
Denicola-Seoane,A., Rubbo,H., Prodanov,E. and Turrens,J.F. (1992) *Mol. Biochem. Parasitol.*, **54**, 43–50.
Gomez-Eichelmann,M.C., Holz,G.,Jr, Beach,D., Simpson,A.M. and Simpson,L. (1988) *Mol. Biochem. Parasitol.*, **27**, 143–158.
Hill,G. and Cross,G. (1973) *Biochim. Biophys. Acta*, **305**, 590–596.
Kanehisa,M.I. (1982) *Nucleic Acids Res.*, **10**, 183–196.
Kidane,G., Hughes,D. and Simpson,L. (1984) *Gene*, **27**, 265–277.
Landweber,L.F. (1992) *BioSystems*, **28**, 41–45.
Lee,S.-Y., Lee,S.-T. and Chang,K.-P. (1992) *J. Protozool*, **39**, 190–196.
Lee,S.T., Liu,H.Y., Lee,S.P. and Tarn,C. (1994) *Mol. Cell. Biol.*, **14**, 587–596.
Maslov,D.A. and Simpson,L. (1992) *Cell*, **70**, 459–467.
Maslov,D.A., Sturm,N.R., Niner,B.M., Gruszynski,E.S., Peris,M. and Simpson,L. (1992) *Mol. Cell. Biol.*, **12**, 56–67.
Maslov,D.A., Thiemann,O. and Simpson,L. (1994a) *Mol. Biochem. Parasitol.*, in press.
Maslov,D.A., Avila,H.A., Lake,J.A. and Simpson,L. (1994b) *Nature*, **365**, 345–348.
Muhich,M., Neckelmann,N. and Simpson,L. (1985) *Nucleic Acids Res.*, **13**, 3241–3260.
Parrot,L. and Foley,H. (1939) *Arch. Inst., Pasteur d'Algerie*, **17**, 231–232.
Read,L.K., Myler,P.J. and Stuart,K. (1992) *J. Biol. Chem.*, **267**, 1123–1128.
Read,L.K., Wilson,K.D., Myler,P.J. and Stuart,K. (1994) *Nucleic Acids Res.*, **22**, 1489–1495.
Riley,G.R., Corell,R.A. and Stuart,K. (1994) *J. Biol. Chem.*, **269**, 6101–6108.
Rovai,L., Tripp,C., Stuart,K. and Simpson,L. (1992) *Mol. Biochem. Parasitol.*, **50**, 115–126.
Sambrook,J., Fritsch,E.F. and Maniatis,T. (1989) *Molecular Cloning. A Laboratory Manual.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
Shaw,J., Feagin,J.E., Stuart,K. and Simpson,L. (1988) *Cell*, **53**, 401–411.
Simpson,L. (1979) *Proc. Natl Acad. Sci. USA*, **76**, 1585–1588.
Simpson,L. (1987) *Ann. Rev. Microbiol.*, **41**, 363–382.
Simpson,L. and Simpson,A. (1978) *Cell*, **14**, 169–178.
Simpson,L., Neckelmann,N., de la Cruz,V., Simpson,A., Feagin,J., Jasmer,D. and Stuart,K. (1987) *J. Biol. Chem.*, **262**, 6182–6196.
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, pp. 53–85.
Simpson,L. and Maslov,D.A. (1994) *Science*, **264**, 1870–1871.
Sloof,P., Arts,G.J., V.d.Burg,J., V.d.Spek,H. and Benne,R. (1994) *J. Bioenerg. Biomembr.*, **26**, 193–204.
Souza,A.E., Myler,P.J. and Stuart,K. (1992) *Mol. Cell. Biol.*, **12**, 2100–2107.
Souza,A.E., Shu,H.-H., Read,L.K., Myler,P.J. and Stuart,K.D. (1993) *Mol. Cell. Biol.*, **13**, 6832–6840.
Stuart,K. (1993) In Benne,R. (ed.), *RNA Editing. The Alteration of Protein Coding Sequences of RNA.* Ellis Horwood, New York, pp. 25–52.
Stuart,K., Feagin,J. and Abraham,J. (1989) *Gene*, **82**, 155–160.
Sturm,N.R. and Simpson,L. (1990) *Cell*, **61**, 879–884.
Sturm,N.R. and Simpson,L. (1991) *Nucleic Acids Res.*, **19**, 6277–6281.
Sturm,N.R., Maslov,D.A., Blum,B. and Simpson,L. (1992) *Cell*, **70**, 469–476.
Wallbanks,K., Maazoun,R., Canning,E. and Rioux,J. (1985) *Parasitology*, **90**, 67–78.

Received on August 5, 1994

Note added in proof

02e EMBL nucleotide sequence databank accession numbers are: ND8 (G1), Z37535; ND9 (G2), Z37536; G3, Z37537; G4, Z37538; ND3 (G5), Z37539.