Nucleic Acids Research, 1994, Vol. 22, No. 8 ***1489–1495***

Editing of *Trypanosoma brucei* maxicircle CR5 mRNA generates variable carboxy terminal predicted protein sequences

Laurie K.Read[1], Kenneth D.Wilson[1,+], Peter J.Myler[1,2] and Kenneth Stuart[1,2,*]
[1]Seattle Biomedical Research Institute, 4 Nickerson Street, Seattle, WA 98109-1651 and [2]Pathobiology Department, SC38, University of Washington, Seattle, WA 98195, USA

Received November 23, 1993; Revised and Accepted March 16, 1994 GenBank accession no. L26251

ABSTRACT

RNA editing post-transcriptionally modifies several mRNAs from the maxicircle of kinetoplastid parasites by addition and removal of uridine residues. We report here that maxicircle CR5 transcripts of *Trypanosoma brucei* are edited in two domains separated by an eight nucleotide sequence that remains unedited. The large 5′ domain is edited to a consensus sequence while the smaller 3′ domain is edited to multiple final sequences. In all, 205–217 Us are inserted and 13–16 encoded uridines are deleted from the CR5 mRNA, producing a mature transcript 75–80% larger than the unedited transcript. The edited RNAs predict small, highly hydrophobic proteins. The carboxy terminal 15–30% of these predicted proteins have multiple different amino acid sequences as a result of the variable edited 3′ mRNA sequence, but these fall into two families of sequence. Limited amino acid sequence and hydrophobicity profile similarities suggest that the protein encoded by edited CR5 mRNA may be a subunit of NADH dehydrogenase.

INTRODUCTION

The mitochondrial (kinetoplast) genome of kinetoplastid parasites is composed of two types of circular molecules termed maxicircles and minicircles. The maxicircle component of the genome encodes rRNAs and constituents of the respiratory apparatus, and is thus analogous to the mitochondrial genomes of other organisms. However, the proteins encoded by many maxicircle genes were not identified by their DNA sequences, since the creation of mature mRNAs requires post-transcriptional editing by insertion and deletion of uridines (for recent reviews, see 1–3). Indeed, editing of some *Trypanosoma brucei* maxicircle transcripts is so extensive that greater than 50% of their mature mRNA sequences are the result of editing. The sequence information for editing resides in small RNAs, termed guide RNAs (gRNAs), which are complementary to the edited mRNA sequences (4–6). In *T.brucei*, the large majority of gRNAs identified to date are encoded in the minicircle component of the mitochondrial genome (5–7).

Genes encoding transcripts that undergo extensive RNA editing are characterized by a pronounced G vs. C strand bias (1). Based on this characteristic, six small regions of the maxicircle, termed CR 1–6 (for cytosine-rich templates 1–6), were postulated to encode extensively edited transcripts in *T.brucei*. Each of these transcripts has now been demonstrated to be extensively edited throughout their length (8–11, R.Corell and K.Stuart, manuscript in preparation). Edited CR1 and CR2 transcripts, now termed ND8 and ND9, respectively, encode subunits 8 and 9 of the NADH dehydrogenase (also called NADH:ubiquinone oxidoreductase), a component of the mitochondrial respiratory complex I (9,10). Edited CR6 transcripts show limited homology to the ribosomal protein S12 (RPS12) (8,12). In this paper, we report the edited sequence of *T.brucei* maxicircle CR5 mRNA. CR5 transcripts have two editing domains, the smaller of which is not edited to a consensus sequence. The protein(s) predicted by edited CR5 mRNA shows limited similarity to NADH dehydrogenase subunit 3 (ND3).

MATERIALS AND METHODS

Cell culture and mitochondrial isolation

Bloodstream and procyclic forms from *T.brucei brucei* IsTaR1 (EATRO 164) and *Leishmania tarentolae* promastigotes (UC strain) were grown as previously described (13,14). Mitochondria were isolated by the method of Harris *et al.* (15) and stored frozen at −70°C until RNA extraction. RNA was isolated from total cells or mitochondria using the guanidinium isothiocyanate–phenol–chloroform method of Chomczynski and Sacchi (16).

Oligonucleotide probes and primers

The oligonucleotides used in this study are shown below with restriction sites incorporated at their 5′ ends underlined. Their position and orientation relative to CR5 sequence are shown in Fig. 1. Oligonucleotides TbCR5-2, TbCR5-8, and TbCR5-12 are

*To whom correspondence should be addressed at: SBRI, 4 Nickerson Street, Seattle, WA 98109-1651, USA

+Present address: Department of Biochemistry, University of Washington, Seattle, WA 98195, USA

the same sense as the mRNA; oligonucleotides TbCR5-7, TbCR5-9, TbCR5-10, and TbCR5-14 are complementary to the mRNA.

TbCR5-2	GCGAATTCGAATCACGGTGTGAACAGGAATGG
TbCR5-7	GCGAATTCAACAAAAAACATCATAAAC
TbCR5-8	GCGGATCCTCGCCTTTTTACTTTAG
TbCR5-9	GCGAATTCAAAAATAAAACAACACTATCAC
TbCR5-10	GCGAATTCAACGAAAACAATCAAAAAAATGC
TbCR5-12	GCGGATCCGTTGTTGTTTGTGGTTTTCG
TbCR5-14	GCGGATCCAAATCTCTTTACCCCCTTCAGTG
Bam-dG_{10}	CCGGATCCGGGGGGGGGG
XSC-dT_{17}	GACTCGAGTCGACATACGATTTTTTTTTTTTTTTT

PCR amplification and cloning

cDNAs were obtained using the polymerase chain reaction (PCR) by previously described strategies (8,9). First strand cDNA was synthesized according to manufacturer's instructions using MMLV reverse transcriptase (Superscript; BRL). PCR was carried out using Taq DNA polymerase (Boehringer Mannheim) in either a Perkin-Elmer Thermocycler or an Ericomp EasyCycler. Six different PCR-generated libraries were used to determine the fully edited CR5 sequence. Conditions for cDNA synthesis and PCR amplification for each of these libraries is described below, followed by a description of the cloning of PCR products which was similar for all libraries.

Library 1: First strand cDNA was synthesized from 5 μg procyclic form (PF) mitochondrial RNA using 100 ng XSC-dT_{17} as primer. Ten percent of this product was PCR amplified using 300 nM each XSC-dT_{17} and TbCR5-2 (which corresponds to unedited CR5 sequence beginning 94 nts from the 5′ end). Amplification conditions were 30 cycles of 94°C for 1 min, 45°C for 30 sec, 72°C for 1 min. Five percent of the resulting product was re-amplified under the same conditions. Approximately one-third of the clones generated in this manner were completely unedited. The remaining clones displayed only limited editing at their 3′ ends, and no consensus sequence was evident.

Library 2: Twenty μg total bloodstream form (BF) RNA was C-tailed using poly(A) polymerase as previously described (17) and cDNA synthesized with 100 ng Bam-dG_{10} as primer. PCR amplification of ten percent of the product was performed using 300 nM each Bam-dG_{10} and TbCR5-2 under the conditions described for Library 1. One of the five clones sequenced from this library was a 5′ truncated RNA (18) with substantial edited sequence. Of the remaining four full length clones, all of which displayed different 5′ sequence, three contained identical sequence at editing sites (ESs) 1−10. (ESs are defined as positions between nonuridine nucleotides; 6).

Library 3: Based on the consensus sequence derived from Library 2, a primer corresponding to 3′ edited sequence (TbCR5-7) was used to obtain clones edited further 5′. cDNA was synthesized using 20 μg of BF total RNA with 100 ng XSC-dT_{17} as primer. The resulting cDNA was dC-tailed using terminal deoxynucleotidyl transferase (BRL) according to the manufacturer's instructions. Ten percent of the dC-tailed cDNA was amplified using 300 nM each Bam-dG_{10} and TbCR5-7 for 2 cycles of 94°C for 1 min, 45°C for 1 min, 72°C for 1 min and 28 cycles of 94°C for 1 min, 50°C for 1 min, 72°C for 1 min. Five of the clones from this library had identical edited sequence extending to ES 60, with the exception of ESs 14, 15,16, and 19, and this was assumed to be the fully edited sequence (later confirmed by Library 6; see below). The remaining 34 clones displayed highly variable edited sequences which differed from each other, and several clones were internally primed at runs of Gs.

Library 4: The edited consensus 5′ of ES 60 was determined by PCR amplification of dC-tailed cDNA with Bam-dG_{10} and the TbCR5-9 primer which corresponds to consensus edited sequence derived from Library 3. cDNA was synthesized using 20 μg BF total RNA and 100 ng TbCR5-7 as primer and dC-tailed as described above. Ten percent of the product was PCR amplified for 30 cycles of 94°C for 1 min, 45°C for 30 sec, 72°C for 1 min using 300 nM each Bam-dG_{10} and TbCR5-9. Nine of twenty-one clones sequenced from this library contained identical edited sequence up to ES 125.

Library 5: To determine the extreme 5′ edited sequence, ten percent of the dC-tailed cDNA from Library 4 was PCR amplified using 300 nM each Bam-dG_{10} and TbCR5-10 under conditions described for Library 4. TbCR5-10 corresponds to consensus edited sequence derived from Library 4. Eight of eleven clones sequenced contained identical sequence up to the 5′ end of the mRNA with the exception of ES 148 which contained with one, two, or three Ts with essentially the same frequency.

Library 6: In an attempt to determine a consensus sequence for ESs 14, 15, 16, and 19, cDNA was PCR amplified with primers corresponding to edited 5′ and 3′ sequence. (Although the 5′ TbCR5-12 primer is somewhat downstream of the 5′ end, 95% of all clones from Library 4 edited at these sites were fully edited at their 5′ ends.) cDNA was synthesized from 20 μg BF total RNA using 100 ng XSC-dT_{17} as primer. Ten percent of the product was PCR amplified using 300 nM each TbCR5-7 and TbCR5-12 for 4 cycles of 94°C for 1 min, 45°C for 30 sec, 72°C for 1 min and 26 cycles of 94°C for 1 min, 55°C for 30 sec, and 72°C for 1 min. Three of 11 clones sequenced were partially edited throughout their length. The other 8 clones were primarily fully edited and confirmed the edited sequence derived from Libraries 3 and 4. However, alignment of these clones provided no consensus sequence at ES 12, 14, 15, 16, or 19.

All PCR products were digested with appropriate restriction enzymes, size selected on 1.5% agarose, transferred to NA45 paper (Schleicher and Schuell), and eluted for 1 hour at 65°C in 1 M NaCl, 0.1 mM EDTA, 20 mM Tris (pH 8.0). Size-selected products were ligated into digested pBluescript II SK− (Stratagene) and ligation mixtures used to transform *E.coli* DH5α F′IQ competent cells (Gibco BRL). Recombinants were identified by colony PCR using the vector-specific primers ZL and ZR (8,19). Sequencing was done by the dideoxy chain termination method using Sequenase (USB) and resolution on 6% acrylamide−7M urea gels or by Taq cycle sequencing using an ABI Model 373A DNA sequencer. The resulting sequences were analyzed using DNASTAR (Madison, WI), ESEE (20), CLUSTAL V (21), and the University of Wisconsin Genetics Computer Group software (22).

Gel electrophoresis and hybridizations

Ten μg total RNA per lane was electrophoresed on 1.5% agarose gels containing 0.66 M formaldehyde and 20 mM MOPS, pH 7.0 (23), and RNA transferred to Nytran by capillary transfer. An *in vitro* transcribed probe incorporating α[^{32}P]-UTP and complementary to edited CR5 mRNA was synthesized using a clone from Library 4 which corresponds to the 5′ 296 nucleotides of fully edited CR5 mRNA sequence as a template. Filters were prehybridized for 6 hours at 60°C in 5 × SSPE (1 × SSPE=90 mM NaCl, 10 mM NaH_2PO_4, 1 mM EDTA), 1% SDS, 1 ×

Denhardt's solution, 50% formamide, and 150 μg/ml herring sperm DNA, and hybridization was performed overnight in the same solution containing 2 × 10^6 dpm/ml probe at 60°C. Blots were then washed twice in 2 × SSPE, 0.1% SDS for 15 minutes at 60°C and twice in 0.1 × SSPE, 0.1% SDS for 15 minutes at 60°C. A probe corresponding to nearly the entire unedited CR5 sequence was synthesized by incorporation of α[^{32}P]-dATP in a PCR reaction as described (8) using a plasmid containing a maxicircle fragment spanning the CR5 gene (pTKHR38; 24) and the TbCR5-8 and TbCR5-14 primers. Filters were probed and washed as previously described (8).

RESULTS AND DISCUSSION

Edited CR5 sequence

The fully edited CR5 sequence (Fig. 1) was derived from alignment of sequence from 104 cDNA clones. Our approach to determining the sequence entailed the generation of six cDNA

Table I. Numerical representation of number of Ts in the variable 3′ region of clones generated by PCR amplification with 5′ and 3′ primers corresponding to edited CR5 mRNA sequence

DNA:	A	C	TG	A	A	G	G	G	G	G	TA	A	A	G	
ES:	23	22	21	20	19	18	17	16	15	14	13	12	11		total Ts
Clone															
G6	0	0	1	1	4	2	2	4	4	0	0	2	4		24
G7	0	0	1	1	3	1	1	0	0	1	0	4	4		16
G8	0	0	1	1	1	1	1	1	0	1	0	2	4		13
G9	0	0	1	1	3	1	1	1	0	1	0	2	4		15
G11	0	0	1	1	3	1	1	1	0	1	0	4	4		17
G12, G1	0	0	1	1	3	1	1	0	1	1	0	4	4		17
G14	0	0	1	1	1	1	1	0	1	0	0	2	4		12

Clones were generated as described in Materials and Methods (Library 6). The number of Ts in each editing site (ES) and the total number of Ts in the region presented is given. CR5 DNA sequence in this region is shown above the numerical representation.

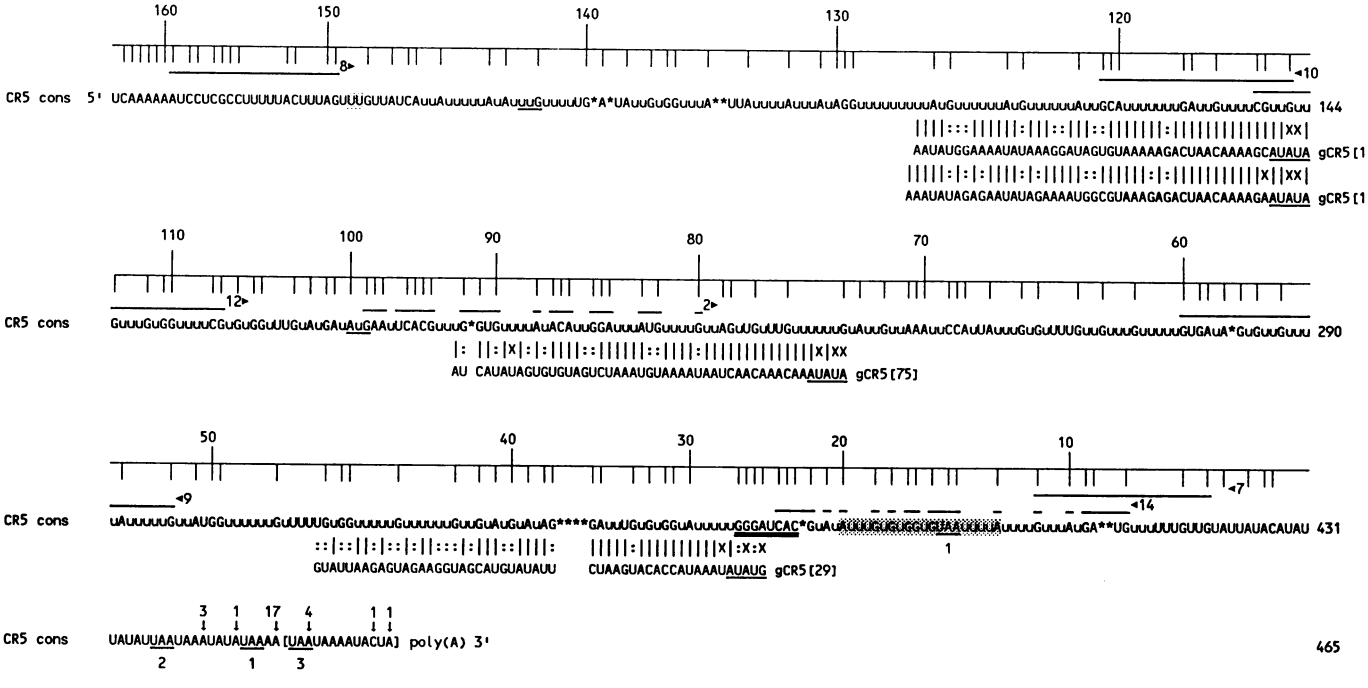

Figure 1. *T.brucei* CR5 edited mRNA sequence, potential gRNA sequences, oligonucleotides, and polyadenylation sites. The consensus edited sequence derived from CR5 cDNAs is shown with potential editing sites (*i.e.*, sites between non-U nts) numbered proceeding 3′ to 5′ beginning at the furthest 3′ non-U, non-A residue. The nt numbers shown at the right begin at the most common 5′ end. Uppercase letters represent encoded nts, lower case u's represent uridines added by editing, and asterisks represent encoded uridines deleted by editing. Shaded regions indicate sequences which are not edited to a final consensus sequence (see text). The 5′-most UUG present in all clones, the 5′-most AUG, and potential termination codons are underlined; termination codons are numbered according to their reading frame. The eight nucleotide sequence which remains unedited in the final sequence and which defines the boundary between the 5′ and 3′ editing domains is double underlined. Sequence at the 3′ terminus of a minority of clones is bracketed. Polyadenylation sites from cDNA clones in Libraries 1 and 2 are indicated by arrows with the numbers indicating their frequency (in a total of 27 clones). Potential gRNA sequences are shown 3′ to 5′ below the mRNA beginning with an RYAYA sequence (underlined), with Watson–Crick basepairs indicated by (|), G:U basepairs by (:), and mismatches between mRNA and gRNAs by (X). gCR5[29] and gCR5[75] are encoded on the TbmiT13 and TbmiT4 minicircles, respectively (R.Corell, *et al.*, manuscript in preparation), gCR5[114] is encoded on a *T.equiperdum* minicircle (43), and gCR5[116] was identified by PCR amplification of minicircle cassettes as previously described (6). Sequences corresponding to oligonucleotides complimentary to RNA (◀) or the same sense as RNA (◀) are overlined and designated by abbreviated names relative to those shown in Materials and Methods.

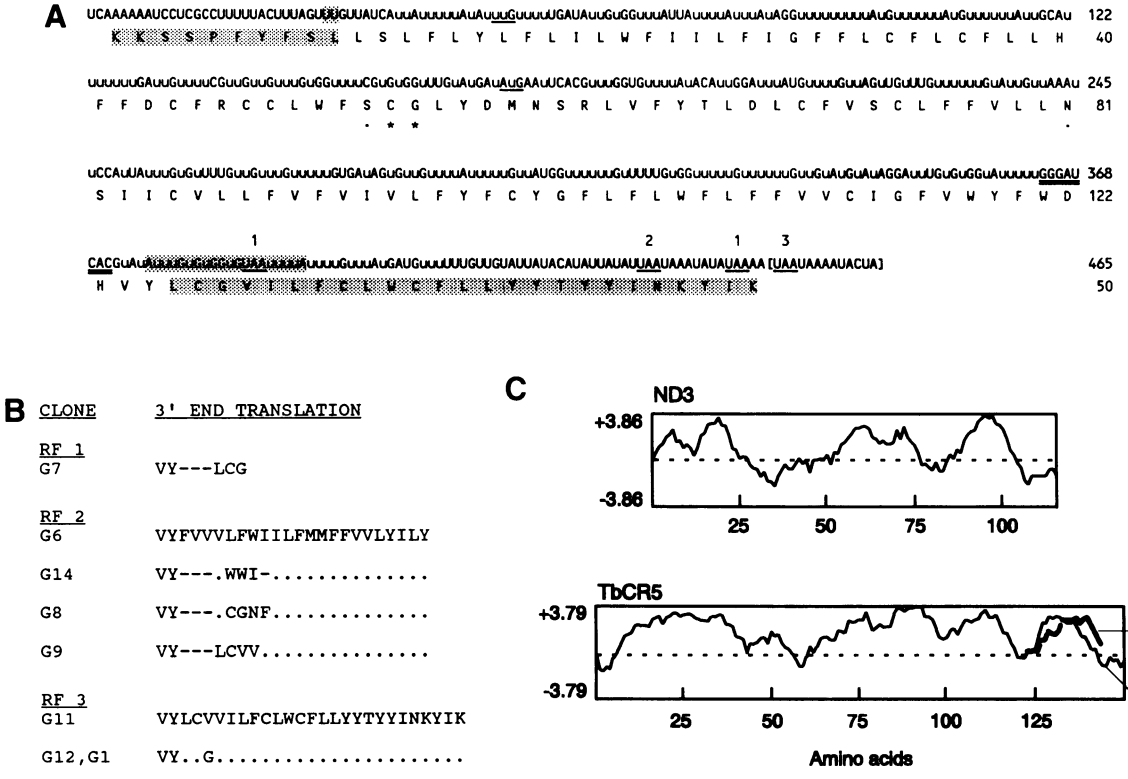

Figure 2. Amino acid sequence(s) predicted from edited CR5 mRNA. **A.** Translation of consensus edited CR5 RNA sequence assuming three Us at ES 148 and using the 3′ end sequence of clone G12. Potential initiation and terminaiton codons, ambiguously edited regions, and 8 nt internal unedited sequence are indicated as in Fig. 1. Variable predicted amino acid sequence resulting from ambiguously edited sequence is shaded. Asterisks indicate amino acids present in the CR5 protein which are conserved in all mitochondrial ND3 proteins and periods indicate neutral (S) and conservative (N) replacements for amino acids conserved in all ND3 proteins. Other symbols as in Fig. 1. **B.** Variation in C-terminal amino acid sequence predicted from different cDNA clones (Library 6). Periods indicate residues identical to those shown on the top line within each reading frame and dashes indicate gaps introduced to facilitate alignment. **C.** Kyte−Doolittle hydrophobicity plots of predicted CR5 proteins (TbCR5) and ND3 from *Bos taurus* (ND3). The CR5 sequence is numbered from the beginning of the ORF shown in Fig. 2A. The different C-terminal sequences predicted from clones G12 (thin line) and G8 (thick line) are shown.

libraries, and took advantage of the general 3′ to 5′ progression of editing (25) which results in cDNAs being edited at their 3′ ends and unedited at their 5′ ends. This approach has been successful in determining the fully edited sequences of a number of edited RNAs (8,9,26,27).

Surprisingly, we were not able to determine a consensus sequence for ESs 12, 14, 15, 16, or 19 (shaded in Fig. 1) using this strategy, even when PCR amplification was performed with primers corresponding to edited 5′ and 3′ sequence (Library 6, Materials and Methods). In clones from Library 6 which are otherwise nearly fully edited, the total number of Ts in the ambiguous ESs is highly variable and does not always increase by multiples of three (Table 1). Thus, the clones are shifted out of frame relative to each other at several positions in the ES 12−19 region, resulting in different protein sequences at the carboxy terminus of the predicted protein as discussed below. We have previously observed similar, but less extreme, sequence variation within open reading frames (ORFs) of other extensively edited *T.brucei* maxicircle genes (10,27). One possible explanation of such variation is that several gRNAs may direct editing of this region of the mRNA, each specifying a different sequence. Alternatively, the variability in the 3′ domain of *T.brucei* CR5 mRNA might be due to loss of the minicircle(s) encoding gRNA(s) specific for editing of this region, similar to the absence of gRNAs for G1−5 (=CR1−5) mRNAs in the UC strain of *L.tarentolae* (28). Such a loss might allow any gRNA which could form a minimal anchor in this region to direct its editing. This scenario seems unlikely, however, since ESs 20−22 are edited to a consensus sequence (Fig. 1 and Table I), thus suggesting their editing is directed by a specific family of gRNAs anchoring downstream of these ESs.

The variable sequence at some ESs in the 3′ end of CR5 mRNA does not prevent editing of the rest of the message to a consensus sequence (Fig. 1). This may be due to the presence of an 8 nt region (doubly underlined in Fig. 1) just 5′ of the ambiguous ESs which remains unedited in the final consensus sequence. The 8 nucleotide sequence may define the boundary between 3′ and 5′ editing domains similar to those previously described in the *T.brucei* ND7 (27) and *L.tarentolae* RPS12 mRNAs (12). This sequence could presumably serve as an anchor region for a gRNA which could direct editing of the 5′ domain. As observed in other extensively edited transcripts, short regions at the 5′ and 3′ ends of fully edited CR5 transcripts also remain unedited in the fully edited mRNA. In addition, one ES in the 5′ end of CR5 mRNA lacks a consensus sequence; ES 148 contains one, two, or three Ts with essentially equal frequency. However, since ES 148 is probably outside the ORF (see below), the number of Us at this site would not be expected to affect the function of the mature RNA. Thus, fully edited CR5 mRNA is generated by the addition of 205−217 uridines and the deletion of 13−16 encoded uridines

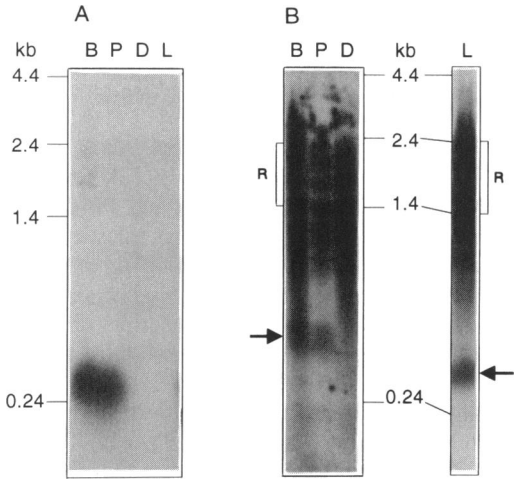

Figure 3. Northern blot analysis of CR5 RNAs. Blots containing 10 μg of total RNA from *T.brucei* bloodstream forms (B), procyclic forms (P), or dyskinetoplasic mutants (D), and *L.tarentolae* promastigotes (L) were probed with **(A)** an unedited PCR-synthesized CR5 probe corresponding to nearly the entire CR5 gene or **(B)** a radiolabelled *in vitro* transcript complimentary the 5′ two-thirds of the edited CR5 mRNA sequence. Arrows indicate CR5 signal. Shadowing due to rRNA is indicated by R. Size markers (BRL) are shown in kb.

to give a mature transcript of between 445−459 nucleotides assuming the most common 3′ end (see below). The resulting mature CR5 mRNA is 75−80% larger than the unedited transcript.

The 5′ end of the CR5 mRNA was determined from 43 dC-tailed cDNAs, 40 of which had identical 5′ ends (Fig 1). Three clones had 5′ ends 1, 2, or 3 nucleotides 3′ of this site. The most common 5′ end maps 50 nucleotides into the 5′ end of the adjacent CR6 gene (8) which is transcribed from the opposite strand. The 3′ polyadenylation sites (derived by PCR amplification of RNA with XSC-dT_{17} or C-tailed RNA with Bam-dG_{10}) were more variable than the 5′ ends (Fig. 1). The most common poly(A) addition site in the CR5 transcript is within the four As of a UAUAUAAAA sequence. Given the most common 3′ end and including encoded As, 3′ end of the CR5 gene overlaps at least 22 nucleotides of the adjacent ND4 gene, which is transcribed from the opposite strand. This is based on the position of the ND4 termination codon since the precise ND4 3′ end is not known; thus, the overlap between the two genes is likely greater than 22 nucleotides. Three of 4 BF clones (all edited) had Us in their poly(A) tails, as did 3 of 23 PF clones (2 edited; 1 unedited). In addition, one BF clone also had a C residue in the poly(A) tail.

Four potential gRNAs which could direct editing of the CR5 transcript were identified by computer search of minicircle and maxicircle sequences (6) (Fig 1). The length of complementarity (allowing G:U basepairing) between the putative gRNAs and the edited CR5 sequence ranges from 43 to 46 nt. All four potential gRNAs are encoded in minicircles between 18 bp inverted repeats (5,29), and all have a 5′-RYAYA-3′ sequence at their 5′ ends which has been postulated to be involved in transcription initiation (5). Interestingly, gCR5[114] and gCR5[116] could potentially direct editing of almost exactly the same region of the mRNA to the same final sequence, although their sequences differ greatly due to the flexibility allowed by G:U as well as A:U basepairing. However, the presence of the putative CR5 gRNAs was not examined by Northern analysis, and since gCR5[114] is encoded on a *T.equiperdum* minicircle it is possible that it is not present in our *T.brucei* stock. No potential gRNAs were found which could direct any of the editing patterns at the variable 3′ editing sites.

Predicted protein sequence(s)

The longest ORF in the edited CR5 sequence begins either two or 25 nts from the 5′ end depending on the sequence at ES 148 (see Fig. 2A). The 5′-most AUG in this reading frame is created by uridine insertion. The 175 nucleotide 5′ untranslated region (UTR) defined by this AUG codon is significantly longer than those of other edited mRNAs which range from 28 to 91 nts (30,11). Thus, translation of edited CR5 mRNA may use an alternate initiation codon, as has been suggested for ND1 and MURF1 transcripts of *T.brucei* (2), ND1 and ND7 transcripts of *Crithidia fasciculata* (31) and ND7 transcripts of *L.tarentolae* (32) which do not contain 5′ AUGs in their mature sequences. AUN and GUG codons are believed to act as initiation codons in many mammalian mitochondrial transcripts (33), and GUG and UUG have been proposed as initiation codons in prokaryotes (34,35). Utilization of the first UUG codon which is present in all CR5 clones (underlined in Fig. 2A) would extend the protein sequence 41 amino acids beyond that predicted with the AUG initiation codon. Two AUA and two AUU codons present between this UUG and the first AUG are also candidates for initiation codons. Alternatively, the lack of an AUG codon near the 5′ end of *T.brucei* CR5 mRNA could indicate that the CR5 mRNA is non-functional, again parallel to the UC strain of *L.tarentolae* (28).

Genomically encoded UAA termination codons are present in all three reading frames near the 3′ end of the CR5 transcript (Figs. 1 and 2A). Thus, it is possible that all edited RNAs, regardless of 3′ domain sequence, will possess termination codons. However, the termination codon for reading frame three is 3′ of the most common polyadenylation site (Figs. 1 and 2A). Only 6 of 27 clones examined extended far enough 3′ to include this termination codon (Fig. 1). An additional possibility is that U addition in the poly(A) tail could generate an in-frame stop codon in a transcript otherwise lacking one.

The longest ORF in the edited CR5 mRNA sequence predicts proteins ranging from 71−93 amino acids and having molecular masses from 8.7−11.5 kDa depending on 3′ sequence and assuming an AUG start codon (Fig. 2). If the first UUG codon is utilized for translation initiation, proteins ranging from 112−134 amino acids and having molecular masses between 13.7 and 16.6 kDa are predicted. The variation in edited CR5 3′ mRNA sequence results in different predicted amino acid sequences in the carboxy terminal 15−30% of the protein (Fig. 2). Two reading frames predominate (with minor variations between clones), one of which contains a carboxy terminal extension high in charged and polar amino acids (reading frame 3) (Figs. 2B and 2C). All predicted proteins are highly hydrophobic with phenylalanine, leucine, and valine contents averaging 22, 19, and 13%, respectively. The Kyte−Doolittle hydrophobicity profile (36) of the predicted CR5 protein product predicts three (or possibly four) membrane spanning domains (Fig. 2C).

Searches of protein and translated nucleic acid databases with the CR5 mRNA (using 3′ variations in both clones G12 and G8)

showed no significant homology to any other protein, although the analysis was complicated by the hydrophobic nature of the sequence. However, PROFILESEARCH analysis (37) revealed 17 of the top 30 matches to NADH dehydrogenase subunit 3 (ND3). Multiple alignment of these sequences showed low levels of amino acid homology between the CR5 and ND3 sequences (11.5−16.8% identity; 39.7−53.4% similarity) which were not significant upon RDF2 shuffling analysis (38) (z_0=0−3 standard deviations above the mean). However, ND3 protein sequences are not highly conserved between species; scores of z_0=0 are obtained between ND3 proteins of some species in RDF2 analyses. Only six amino acids are absolutely conserved between mitochondrially encoded ND3 sequences (39). The CR5 predicted protein contains two of these (a CG dipeptide) and has a conservative replacement for a third (N for D) and a neutral replacement for a fourth (S for E) (see Fig 2A). In addition, several other matches or conservative replacements to the consensus sequence were evident throughout the central portion of the protein, which is the most highly conserved region (39). The predicted size of the *T.brucei* protein is consistent with that of other ND3 proteins which range from approximately 13−16.5 kD (39). Moreover, there is similarity between the hydrophobicity profiles of the *T.brucei* protein and other ND3 proteins (Fig. 2C) (39), the primary characteristic of which is three potential membrane spanning domains. While these results are not conclusive, they suggest that edited CR5 mRNA may encode ND3. Furthermore, small size and hydrophobic nature are properties which are shared by a number of subunits of NADH dehydrogenase (40). Thus, if CR5 does not encode ND3, it may encode another subunit of NADH dehydrogenase.

Variable carboxy terminal predicted protein sequence due to the absence of a consensus edited sequence in the 3′ domain of *T.brucei* CR5 may reflect non-functionality of the edited mRNA. However, editing of the 5′ domain to a consensus sequence argues against this interpretation since the absence of selective pressure would be expected to result in divergence of the 5′ consensus RNA sequence to a similar extent as the 3′ sequence. It is also possible that the CR5 protein product is essential, but that not all of its domains are required for function. For example, CR5 protein starting at the first methionine may be truncated relative to the wild type protein but still retain function, and sequence flexibility at the carboxy terminus may be tolerated. Additional possibilities are that only a subset of edited mRNAs are selected for translation or that all edited mRNAs are translated but only some CR5 proteins are functional and a certain level of non-functional protein is tolerated.

Translation of edited CR5 mRNAs predicts that about half of the proteins will have hydrophobic carboxy termini, while the remainder will have a hydrophilic extension at their carboxy ends. This suggests that CR5 proteins with different carboxy termini may have slightly different functions thereby providing physiological fine-tuning of parasite respiration. For example, different intermolecular interactions of CR5 proteins in mitochondrial complexes may lead to slightly altered functions. Interestingly, NADH dehydrogenase has been proposed to take part in different energy producing pathways in different forms of BF parasites (41), and it is conceivable that complexes in different pathways have slightly different protein compositions. We have previously pointed out the potential for the production of alternative proteins from some maxicircle genes depending on whether mRNAs are edited or unedited (apocytochrome b and cytochrome oxidase subunit II) or whether they are edited in one or both domains (ND7) (42). Editing of CR5 mRNA to variable sequences in the 3′ domain may represent another mechanism of production of alternative proteins from the same maxicircle gene.

CR5 transcripts

Northern blot analysis of total *T.brucei* RNA reveals unedited and edited CR5 transcripts in both BF and PF parasites (Fig. 3). These transcripts are absent in dyskinetoplastic (DK) mutants which lack kinetoplast DNA, confirming their mitochondrial origin. A probe containing unedited CR5 sequence and spanning nearly the entire CR5 gene detects transcripts of 240−320 nts in PF and 240−370 nts in BF which are slightly more abundant in the latter (Fig. 3A). Unedited CR5 mRNA is 254 nts (without a poly(A) tail); thus, larger transcripts hybridizing with this probe are likely to be partially edited at their 3′ ends. The larger size range of the BF transcripts suggests that there may be more editing of CR5 mRNA in this stage. No transcripts were detected with this probe in the related kinetoplastid, *L.tarentolae* (Fig. 3A). CR5 transcripts ranging from 365−465 nts are detected in both BF and PF *T.brucei* by a radiolabelled *in vitro* transcript complementary to the 5′ 296 nts of edited CR5 sequence and were sometimes observed to be slightly more abundant in BF than in PF parasites (Fig. 3B). The size of these transcripts is consistent with the largest being fully edited CR5 RNA with a poly(A) tail of approximately 10−20 nts. The smear of larger transcripts also detected by this probe probably represents cross-hybridization to nuclear transcripts since it is also observed in DK RNA. The *T.brucei* edited CR5 riboprobe also hybridizes to a transcript of 325−390 nts in *L.tarentolae*, suggesting that a significant amount of *L.tarentolae* CR5 mRNA is edited within the 5′ two-thirds of the molecule. However, the smaller size of the transcripts relative to that in *T.brucei* is consistent with the observation that CR5 (=G5) transcripts in the UC strain of *L.tarentolae* are not fully edited (28). Our inability to detect edited CR5 transcripts with oligonucleotide probes or a shorter riboprobe and the long exposure time necessary to see the signal in Fig. 3B suggests that edited CR5 mRNA is relatively rare in *T.brucei*. This is in contrast to unedited mRNA which was readily detected.

In conclusion, we report here that *T.brucei* CR5 mRNA is extensively edited by uridine insertion and deletion in two domains. Editing of CR5 mRNA spans the entire length of the transcript exclusive of the 5′ and 3′ termini and the eight nucleotide interdomain sequence. Only the large 5′ editing domain is edited to a consensus sequence. The smaller 3′ domain is edited to a variety of sequences which predict two families of carboxy terminal amino acid sequences. Limited similarities of the predicted CR5 protein product suggest that it may encode ND3.

ACKNOWLEDGMENTS

This work was supported in part by National Institutes of Health Grant AI14102 to KS who is a Burroughs Wellcome Scholar in Molecular Parasitology. LKR was supported by National Institutes of Health Postdoctoral Fellowship AI08491. We thank the members of the Stuart lab for helpful discussions, Dr R.Corell for assistance with computer analysis, and Drs S.Anderson, R.Corell, and Y.Skeiky for critical reading of the manuscript. We are also grateful to K.Stankey and J.Bertles for excellent technical assistance.

REFERENCES

1. Stuart,K. (1991) Trends Biochem. Sci., **16**, 68–72.
2. Stuart,K. (1991) Annu. Rev. Microbiol., **45**, 327–344.
3. Hajduk,S.L., Harris,M.E. and Pollard,V.W. (1993) FASEB J., **7**, 54–63.
4. Blum,B., Bakalara,N. and Simpson,L. (1990) Cell, **60**, 189–198.
5. Pollard,V.W., Rohrer,S.P., Michelotti,E.F., Hancock,K. and Hajduk,S.L. (1990) Cell, **63**, 783–790.
6. Koslowsky,D.J., Riley,G.R., Feagin,J.E. and Stuart,K. (1992) Mol. Cell. Biol., **12**, 2043–2049.
7. Corell,R.A., Feagin,J.E., Riley,G.R., Strickland,T., Guderian,J.A., Myler,P.J. and Stuart,K. (1993) Nucl. Acid. Res., **21**, 4313–4320.
8. Read,L.K., Myler,P.J. and Stuart,K. (1992) J. Biol. Chem., **267**, 1123–1128.
9. Souza,A.E., Myler,P.J. and Stuart,K. (1992) Mol. Cell. Biol., **12**, 2100–2107.
10. Souza,A.E., Shu,H.-H., Read,L.K., Myler,P.J. and Stuart,K. (1993) Mol. Cell. Biol., **13**, 6832–6840.
11. Corell,R.A., Myler,P.J. and Stuart,K.D. (1993) Mol. Biochem. Parasitol. In Press.
12. Maslov,D.A., Sturm,N.R., Niner,B.M., Gruszynski,E.S., Peris,M. and Simpson,L. (1992) Mol. Cell. Biol., **12**, 56–67.
13. Stuart,K., Gobright,E., Jenni,L., Milhausen,M., Thomashow,L. and Agabian,N. (1984) J. Parasitol., **70**, 747–754.
14. Simpson,L. and Braly,P. (1970) J. Protozool., **17**, 511–517.
15. Harris,M.E., Moore,D.R. and Hajduk,S.L. (1990) J. Biol. Chem., **265**, 11368-11376.
16. Chomczynski,P. and Sacchi,N. (1987) Anal. Biochem., **162**, 156–159.
17. Bhat,G.J., Myler,PJ. and Stuart,K. (1991) Mol. Biochem. Parasitol., **48**, 139–150.
18. Read,L.K., Corell,R.A. and Stuart,K. (1992) Nucl. Acids Res., **20**, 2341–2347.
19. Read,L.K., Fish,W., Muthiani,A.M. and Stuart,K.D. (1993) Nucl. Acids Res., **21**, 4073–4078.
20. Cabot,E.L. and Beckenbach,A.T. (1989) Comput. Appl. Biosci., **5**, 233–234.
21. Higgins,D.G., Bleasby,A.J. and Fuchs,R. (1992) Comput. Appl. Biosci., **8**, 189–191.
22. Devereux,J., Haeberli,P. and Smithies,O. (1984) Nucl. Acids Res., **12**, 387–395.
23. Fourney,R.M., Miyakoshi,J., Day,R.S.,III and Paterson,M.C. (1988) Focus, **10**, 5–7.
24. Stuart,K.D. and Gelvin,S.B. (1982) Mol. Cell. Biol., **2**, 845–852.
25. Abraham,J.M., Feagin,J.E. and Stuart,K. (1988) Cell, **55**, 267–272.
26. Bhat,G.J., Koslowsky,D.J., Feagin,J.E., Smiley,B.L. and Stuart,K. (1990) Cell, **61**, 885–894.
27. Koslowsky,D.J., Bhat,G.J., Perrollaz,A.L., Feagin,J.E. and Stuart,K. (1990) Cell, **62**, 901–911.
28. Maslov,D.A. and Simpson,L. (1992) Cell, **70**, 459–467.
29. Jasmer,D.P. and Stuart,K. (1986) Mol. Biochem. Parasitol., **18**, 321–331.
30. Feagin,J.E. and Stuart,K. (1988) Mol. Cell. Biol., **8**, 1259–1265.
31. Shaw,J.M., Feagin,J.E., Stuart,K. and Simpson,L. (1988) Cell, **53**, 401–411.
32. Benne,R., van den Burg,J., Brakenhoff,J., De Vries,B., Nederlof,P., Sloof,P. and Voogd,A. (1985) In Quagliariello,E., Slater,E.C., Palmierei,F., Saccone,C. and Kroon,A.M. (eds.), Achievements and Perspectives of Mitochondrial Research. Elsevier, Amsterdam, pp. 325–336.
33. Gadaleta,G., Pepe,G., De Candia,G., Quagliariello,C., Sbisa,E. and Saccone,C. (1989) J. Mol. Evol., **28**, 497–516.
34. Young,I.G., Rogers,B.L., Campbell,H.D., Jaworowski,A. and Shaw,D.C. (1981) Eur. J. Biochem., **116**, 165–170.
35. Stormo,G.D., Schneider,T.D. and Gold,L.M. (1982) Nucl. Acids Res., **10**, 2971–2996.
36. Kyte,J. and Doolittle,R.F. (1982) J. Mol. Biol., **157**, 105–132.
37. Gribskov,M., McLachlan,A.D. and Eisenberg,D. (1987) Proc. Natl. Acad. Sci. USA, **84**, 4355–4358.
38. Pearson,W.R. and Lipman,D.J. (1988) Proc. Natl. Acad. Sci. USA, **85**, 2444–2448.
39. Fearnley,I.M. and Walker,J.E. (1992) Biochim. Biophys. Acta, **1140**, 105–134.
40. Ragan,C.I. (1987) Curr. Top. Bioenerg., **15**, 1–36.
41. Bienen,E.J., Saric,M., Pollakis,G., Grady,R.W. and Clarkson,.A.B.,Jr. (1991) Mol. Biochem. Parasitol., **45**, 185–192.
42. Stuart,K. and Feagin,J.E. (1992) Int. Rev. Cytol., **141**, 65–88.
43. Barrois,M., Riou,G. and Galibert,F. (1981) Proc. Natl. Acad. Sci. USA, **78**, 3323–3327.