Nucleic Acids Research, 1995, Vol. 23, No. 16 3093–3102

The secondary structure of guide RNA molecules from *Trypanosoma brucei*

Beate Schmid, George R. Riley[1], Kenneth Stuart[1,2] and H. Ulrich Göringer*

Laboratorium für Molekulare Biologie, Genzentrum der LMU München am MPI für Biochemie, 82152 Martinsried, Germany, [1]Seattle Biomedical Research Institute, Seattle, WA 98109-1651, USA and [2]Pathobiology Department, SC38, University of Washington, Seattle, WA 98195, USA

Received July 5, 1995; Revised and Accepted July 21, 1995

ABSTRACT

RNA editing in kinetoplastid organisms is a mitochondrial RNA processing phenomenon that is characterized by the insertion and deletion of uridine nucleotides into incomplete mRNAs. Key molecules in the process are guide RNAs which direct the editing reaction by virtue of their primary sequences in an RNA–RNA interaction with the pre-edited mRNAs. To understand the molecular details of this reaction, especially potential RNA folding and unfolding processes as well as assembly phenomena with mitochondrial proteins, we analyzed the secondary structure of four different guide RNAs from *Trypanosoma brucei* at physiological conditions. By using structure-sensitive chemical and enzymatic probes in combination with spectroscopic techniques we found that the four molecules despite their different primary sequences, fold into similar structures consisting of two imperfect hairpin loops of low thermodynamic stability. The molecules melt in two-state monomolecular transitions with T_ms between 33 and 39°C and transition enthalpies of –32 to –38 kcal/mol. Both terminal ends of the RNAs are single-stranded with the 3′ ends possibly adopting a single-stranded, helical conformation. Thus, it appears that the gRNA structures are fine tuned to minimize stability for an optimal annealing reaction to the pre-mRNAs while at the same time maximizing higher order structural features to permit the assembly with other mitochondrial components into the editing machinery.

INTRODUCTION

Guide RNAs (gRNAs) are small, metabolically stable RNA molecules, identified within the mitochondria of kinetoplastid protozoan parasites such as *Trypanosoma, Leishmania* and *Crithidia*. The molecules are involved in the unusual maturation process of mitochondrial pre-mRNAs termed kinetoplastid (k)RNA editing. During kRNA editing, exclusively uridine nucleotides are site specifically inserted and sometimes deleted from otherwise incomplete mRNA molecules by an unknown mechanism and unknown organellar machinery (reviewed in 1,2). Guide RNAs have been suggested to provide the information during the editing reaction based on the observation that their primary nucleotide sequence has base complementarity to domains of the final mRNA sequence. This gRNA dependency of the editing reaction has recently been confirmed *in vitro* for the deletion of uridines from a synthetic pre-mRNA (3).

A large number of different gRNA molecules have been experimentally identified and various potential gRNA coding sequences have been located on the mitochondrial genomes of the above mentioned organisms (1,2 and references therein). In *Trypanosoma brucei*, gRNA genes have been detected with only one exception (4) on the so-called minicircle mitochondrial DNA. Every minicircle has the capacity to encode three to four gRNA genes which are always oriented on the same DNA strand and are usually flanked by imperfect 18 bp inverted repeat sequences (5–8). Guide RNAs have an average length of 50–70 nucleotides (nt) with a strong A/U nucleotide bias. On average 15 uridines are post-transcriptionally added at their 3′ ends and this U-tail has been suggested to function as the reservoir for the U-addition or deletion reaction or might be involved in a stabilising base-pairing interaction with the pre-mRNA upstream of an editing site (9–11). Guide RNAs can be capped *in vitro* with guanylyltransferase and GTP suggesting that they are primary transcripts (6,9) and in *T.brucei* most of the gRNAs contain a RYAYA sequence motif at their 5′ terminus (12).

The guiding capacity of a given gRNA is limited by the overall length of the molecule and as a consequence multiple gRNAs are generally required for the complete editing of a specific pre-mRNA (13,14). Aside from the limited set of common features (noted above), these gRNAs have different primary sequences yet they function in the same biochemical reaction. This suggests, that similar features might exist on a higher order structural level perhaps within functionally important domains, since molecules engaged in the same cellular process generally fold into identical structures (15). No structural investigation of gRNAs has heretofore been reported and, so far, the only experimental indication for common structural features was deduced from the fact that different gRNAs could be crosslinked to the same mitochondrial proteins, indicating common binding sites for the polypeptides (16–18).

In order to gain insight into the structure/function correlation of gRNAs and to overcome the absence of phylogenetic data, we

* To whom correspondence should be addressed

chose to experimentally verify the secondary structure of four different gRNAs from *T.brucei*. We used a combination of temperature dependent UV spectroscopy, chemical and enzymatic probing techniques and a secondary structure prediction algorithm to describe the solution structure of these molecules. We demonstrate that at 200–300 mM potassium cation concentration the molecules fold into structures with stabilities below the thermodynamically most stable conformations and we propose simple, imperfect, double stem–loop structures as common secondary structure foldings for the four molecules. We interpret these data as a conservation of general structural features that may be important for the recognition and assembly of gRNAs into the editing apparatus.

MATERIALS AND METHODS

Biochemicals

All chemicals were reagent grade or better. Ribonucleases T1 and T2, T4 polynucleotide kinase and T4 RNA ligase were from Gibco-BRL. Cobra venom RNase V1 was purchased from USB, AMV Reverse transcriptase from Stratagene and nuclease S1 from Boehringer. [^{32}P]pCp (3000 Ci/mmol), [α-^{32}P]ATP (3000 Ci/mmol) and [γ-^{32}P]ATP (5000 Ci/mmol) were purchased from NEN or Amersham. Diethyl pyrocarbonate (DEPC) and Kethoxal were from Serva, Dimethyl sulfoxide (DMS) from Merck and 1-cyclohexyl-3(2-morpholinoethyl) carbodiimide metho-*p*-toluene sulfonate (CMCT) was from Sigma. Oligodesoxynucleotides were synthesized by automated solid-support chemistry using *O*-cyanoethyl-*N,N*-diisopropyl-phosphoramidites. Oligonucleotide concentrations were determined by UV absorbency measurements at 260 nm using extinction coefficients (ε_{260}, l mol^{-1} cm^{-1}) calculated from tabulated values for the di- and mononucleotides (19).

Construction of synthetic gRNAs and RNA synthesis

Synthetic genes for the guide RNAs gA6-14 (20,21), gA6-48 (16), gND7-506 and gCyb-558 (8) (for nomenclature see 14) were constructed by self-assembly of overlapping synthetic oligodesoxynucleotides according to Reyes and Abelson, (1989) (22) and cloned into plasmid pBS$^-$ (Stratagene). Positive clones were identified by restriction enzyme digestion of isolated plasmid DNA and sequenced on an automated DNA sequencer using the dye terminator technology. Guide RNAs were transcribed from linearized plasmid DNA templates using T7 polymerase following standard procedures. Transcripts were purified from non-incorporated NTPs by size exclusion chromatography, recovered by ethanol precipitation and dissolved in 50 mM potassium cacodylate pH 7.2, 150 mM KCl, 2.6 mM $MgCl_2$, 0.1 mM Na_2EDTA. The molecules were renatured after transcription by incubation at 95°C for 2 min followed by a slow cooling period (1°C/min) to 23°C before putting the sample on ice. Extinction coefficients (ε_{260}, l mol^{-1} cm^{-1} × 10^{-5}) for the various molecules were calculated from the sum of the nucleotide absorptivity as affected by adjacent bases (23): gA6-14: 7.42, gA6-48: 6.95, gND7-506: 6.55, gCyb-558: 7.42.

Radioactive end-labelling

Oligodesoxynucleotides were 5′ end-labelled using T4 polynucleotide kinase and [γ-^{32}P]ATP according to standard procedures. Guide RNAs were 3′ end-labelled using T4 RNA ligase and [^{32}P]pCp as described by Bruce and Uhlenbeck (1978) (24). All radiolabelled nucleic acids were purified by denaturing polyacrylamide gel electrophoresis.

Temperature-dependent UV spectroscopy

Absorbance versus temperature profiles (melting curves) were recorded at 260 nm in 50 mM potassium cacodylate pH 7.2, 2.6 mM $MgCl_2$, 0.1 mM Na_2EDTA using a thermoelectrically controlled Perkin Elmer lambda 16 spectrophotometer. The temperature was scanned at a heating rate of 1 or 2°C/min at temperatures between 10 and 95°C. Melting curves were carried out as a function of gRNA concentration (0.02–0.7 μM) at KCl concentrations between 0 and 1 M. T_m values were determined from derivative plots of absorbance versus temperature as 0.5 ΔA_{260}. Melting enthalpies (ΔH_{vH}) were calculated based on the equation: $\Delta H_{vH} = 4\,RT_m^2\,(d\alpha/dT)_{T=Tm}$ (23, 25). Hyperchromicity values were calculated as described by Hung *et al.* (1994) (26): hyperchromicity (%) = $100 \times A_{260}$ (95°C) – A_{260} (10°C)/A_{260} (10°C). The release of counterions (ΔnK^+) was determined using the equation: $dT_m/d\log [K^+] = -0.9\,(2.303\,RT_m^2/\Delta H_{VH})\,\Delta nK^+$ (27). Standard Gibb's free energies ($\Delta G°$) and standard free entropies ($\Delta S°$) were calculated as described by Marky and Breslauer (1987) (25) for a temperature of 27°C which is the optimal growth temperature for procyclic stage trypanosomes.

Chemical probing

Chemical modification reactions were performed essentially as described by Christiansen *et al.* (1990) (28) using DMS, DEPC, kethoxal and CMCT as modifying agents. Reactions were carried out with 0.2 μg gRNA in the presence of 20 μg/ml bulk yeast tRNA as a carrier in 70 mM HEPES–KOH pH 7.8, 10 mM $MgCl_2$, 270 mM KCl, 1 mM DTT at 27°C for times ranging between 2 and 30 min.

Structure-specific enzymatic probing

Enzymatic probing of gRNAs was performed as described (28) with the following nucleases: Cobra venom RNase V1 (50 and 100 mU), ribonuclease T1 (500 and 750 mU), RNase T2 (3 and 6 U) and nuclease S1 (0.2 and 0.4 U). Reactions were performed in 30 mM Tris–HCl pH 7.8, 20 mM $MgCl_2$, 300 mM KCl, 1 mM DTT with 0.2 μg gRNA and 4 μg bulk yeast tRNA as a carrier except for the nuclease S1 reaction which was performed in 50 mM NaOAc pH 4.6, 50 mM NaCl, 1 mM $ZnCl_2$ and 5% (v/v) glycerol. Incubation was for 7.5 min at 27°C. Reactions were stopped by adding NaOAc pH 6 to a final concentration of 0.3 M followed by extractions with phenol and phenol–chloroform and precipitating the RNA material with ethanol.

Primer extension analysis

Reverse transcription reactions were performed using the following desoxyoligonucleotides primers: gA6-48: AAATTATTCTTTATAC, gA6-14: AAAATAATTATCATATC, gND7-506: AAAATTCACTATATA, gCyb-558: AATTTATTCCCTTTA. Modified gRNAs (1.3 pmol) and 5′ end-labelled primer molecules (1.0 pmol, ~50 000 c.p.m.) were annealed in 10 mM Tris–HCl pH 6.9, 40 mM KCl and 5 mM Na_2EDTA at 75°C for 30 s followed by an incubation at room temperature for 30 min. Extension reactions were performed at 37°C for 30 min using 1 U AMV

reverse transcriptase per 10 µl reaction. Complementary DNA products were separated on denaturing 10% (w/v) polyacrylamide gels.

RNA secondary structure predictions and randomization analysis

RNA secondary structures were calculated using the MFOLD subroutine of the GCG software package (29) based on a free energy minimization algorithm (30,31). Calculations were performed for 27°C, which is the optimal growth temperature for procyclic stage trypanosomes. Experimental data from the probing experiments were implemented using the prevent/force subroutines of the MFOLD program. Randomization of gRNA sequences was performed using the program MacStAn (32). The statistical difference (S) of the experimental gRNA structures to randomised RNAs was calculated using the equation: $S = (\Delta G_{ran} - \Delta G_{exp})/SD$ in analogy to Le and Maizel (33). ΔG_{exp} is the experimentally verified Gibb's free energy value, ΔG_{ran} the mean of 10 random permutations of the real sequence and SD the standard deviation of the randomised sequences. The statistical difference is expressed in SD units. A search for potential pseudoknotted structural motifs was performed using the program STAR (34).

RESULTS

Optical melting studies

For our study we chose four previously identified gRNAs from *T.brucei* (8,14) with primary sequences as shown in Figure 1A. The molecules vary in length between 64 and 79 nt (including 9 or 10 vector derived bases at their 5′ ends) and are predicted to fold into minimal free energy structures (Fig. 1B) with stabilities in the range of –4.4 to –8.2 kcal/mol (at 27°C). Based on the current paradigm for the RNA editing mechanism (36), these molecules direct editing of four individual domains within three different pre-mRNA molecules. Guide RNAs gA6-14 and gA6-48 are specific for two regions within the ATPase (A6) pre-mRNA, whereas gND7-506 and gCyb-558 direct the editing of sequences within the NADH 7 (ND7) and apocytochrome b (Cyb) messenger RNAs. The molecules were transcribed by standard *in vitro* methods and contain 3′ oligo(U) tails either 10 or 15 nt long (see Fig. 1), thus mimicking the posttranscriptional polyuridylylation process within the mitochondrial organelle (9).

To gain first insight into the specific foldings of the four gRNAs in solution, the molecules were renatured after transcription and then thermally unfolded. This process (helix–coil transition) was monitored by the increase in UV absorbancy. At 200 mM monovalent and 2.5 mM divalent cation concentrations all four molecules melted in a broad transition with a midpoint, termed T_m–1, varying between 33 and 39°C depending on the gRNA. Figure 2 shows a representative set of the four normalised melting profiles along with the differentiated curves. Both the broad shapes of the transitions and the low T_m values indicated that none of the molecules contained extended helical domains, which would result in much sharper melting signals (23,37). A small high temperature transition (T_m–2) in the range of 70–77°C could also be detected and the hyperchromicities at 260 nm were determined to vary between 15 and 20%. All four molecules displayed asymmetric peaks at lower temperatures (~20°C, see arrowheads in Fig. 2) suggesting multiple melting processes probably evidence for tertiary interactions or secondary structure changes. Increasing the gRNA concentrations ≤35-fold had no effect on the two T_m values (Fig. 3), indicative of a unimolecular melting of single stranded hairpin structures. Standard thermodynamic parameters for the main transition were therefore calculated according to a two state transition model (23,25) and are listed in Table 1. As anticipated from the transition temperature for T_m–1, the corresponding melting enthalpies (ΔH_{vH}) for all four molecules were low, in the range of –32 to –38 kcal/mol. Consequently, the derived ΔG°_{27} values were calculated to be as low as –0.9 to –1.5 kcal/mol, approximately five to six times less than the theoretical minimal free energy structures (see Fig. 1A).

Table 1. Thermodynamic parameters[a] calculated for the main transition, T_m–1

gRNA	T_m (°C)	ΔH_{vH} (kcal/mol)	ΔG°_{27} (kcal/mol)	ΔS°_{27} (kcal/mol deg)
gA6-48	37 ± 2	–35 ± 3	–1.1 ± 0.3	–1.3 ± 0.1
gA6-14	36 ± 5	–32 ± 6	–0.9 ± 0.6	–1.2 ± 0.2
gND7-506	39 ± 2	–35 ± 6	–1.5 ± 0.3	–1.3 ± 0.2
gGyb-558	33 ± 2	–38 ± 3	–0.9 ± 0.3	–1.4 ± 0.1

[a]Values are given ± SD (T_m, ΔH_{vH}) or ± absolute uncertainties (ΔG°_{27}, ΔS°_{27}).

Since the four gRNAs obviously fold into structures of very low thermodynamic stability it was important to test whether the determined ΔG values were significant. For this reason we employed a statistical analysis with randomized RNA sequences of identical base composition and length. We generated 10 randomised sequences for each of the four gRNAs and calculated their minimal Gibb's free energies using the energy rules published by Freier *et al.* (30) and Zuker *et al.* (31). The statistical deviation (S) of the experimentally derived ΔG values was determined as the difference between the mean of the ΔG values for the randomised RNAs (ΔG_{ran}) and ΔG_{exp}, the free energy value derived from the melting data divided by the standard deviation (SD) of the random sequences (33). In all four cases, the ΔG_{exp} values varied for more than two standard deviations from the ΔG_{ran} values (≤4.7 SD), indicating a significant difference to the randomised RNAs.

The folding of nucleic acids can be strongly affected by the concentration of monovalent cations (27,38). Therefore, we investigated the melting behaviour of gA6-48 (as a representative for all four gRNAs) at increasing KCl concentrations (0–1 M). The data were analysed in a plot of T_m versus log cKCl (data not shown) and demonstrated that both melting transitions behaved independently of the salt concentration, as great as 1 M KCl. A moderate increase of T_m–2 ($dT_m/d\log$ cKCl = 4.1°C) was calculated to amount to a release of less than one K^+ per gRNA molecule, consistent with the result that only short helices become unfolded (39).

Chemical and enzymatic probing and secondary structure modelling

The higher order structure of the four *in vitro* synthesised gRNA molecules was probed by chemical modification and ribonuclease accessibility experiments. The four individual bases in single stranded, non-stacked conformations were monitored with kethoxal (G, N-1, N-2); CMCT (U, N-3 > G, N-2); DEPC (A,

A

name	sequence
gA6-48	5' $_{9(N)}$ AUAUAUAAUC CAUUAUCAGU UGCGAGAUUG UAGUAUAAAG AAUAAUUUUUU UUUUUU 3'
gA6-14	5' $_{10(N)}$ AUAUACUAUA ACUCCGAUAA CGAAUCAGAU UUUGACAGUG AUAUGAUAAU UAUUUUUUUU UUUUUUUUU 3'
gND7-506	5' $_{9(N)}$ AUAAAUACGA UGUAAAUAAC CUGUAGUAUA GUUAGUGUAU AUAGUGAAUU UUUUUUUUU 3'
gCyb-558	5' $_{9(N)}$ GGGAGAUUAAA AGACAAUGUG AAUUUUUAGG UGAUAAAGGG AAUAAUUUUUU UUUUUUUUUU 3'

name	length (nt)	ΣGC/ΣAU	ΔG°_{27} (kcal/mol)	basepairs (n)
gA6-48	64	0.33	-6.8	13
gA6-14	79	0.34	-5.8	20
gND7-506	67	0.34	-8.2	15
gCyb-558	70	0.37	-4.4	15

B

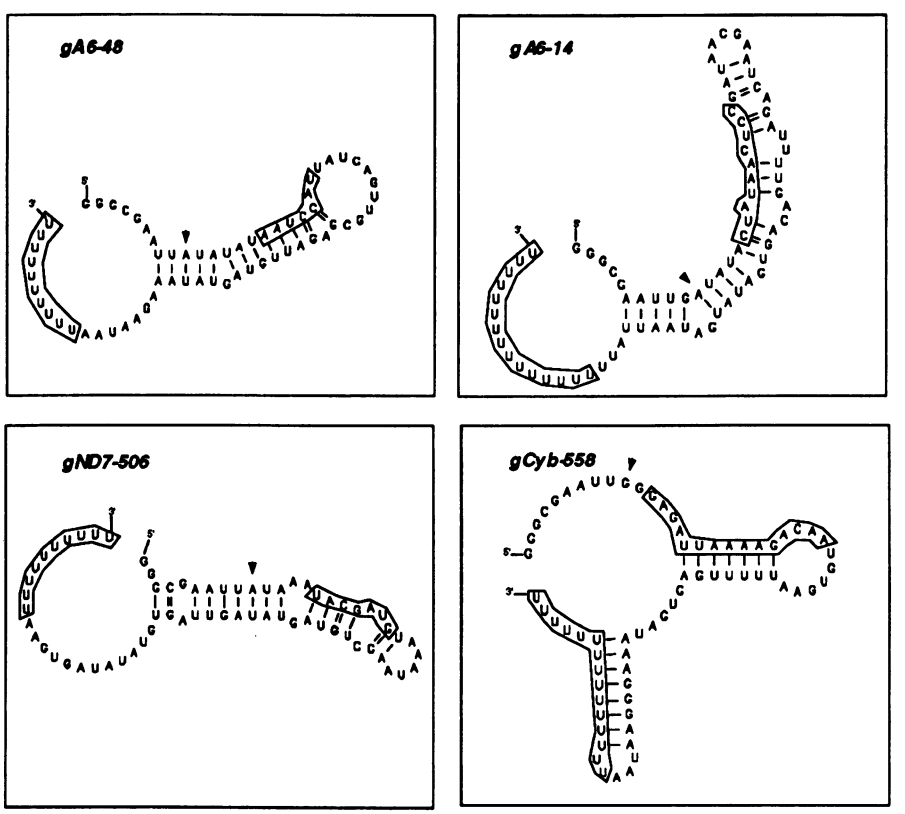

Figure 1. (**A**) Primary sequences of the four different *T.brucei* guide RNA molecules used in this study. The anchor sequences necessary for the initial base-pairing interaction with the pre-edited mRNAs and the 3′ terminal oligo(U) tails are underlined. The length of the molecules as well as the GC versus AU nucleotide bias were calculated including the 9 or 10 vector derived nucleotides at the 5′ ends. Gibb's free energies were calculated for a temperature of 27°C, which is the optimal growth temperature of procyclic stage trypanosomes. The number of basepairs (*n*) were taken from the predicted minimal free energy secondary structures as shown in (**B**). The structures were generated using the MFOLD software (29) and RNA_d2 for the graphical output (59). Anchor and oligo(U) sequences are boxed. Arrowheads annotate the presumed 5′ ends of the molecules *in vivo*.

N-7) and DMS (G, N-7 > A, N-1 > C, N-3) as well as with ribonucleases T1 (G-specific) and T2 or nuclease S1 (both single strand-specific). Nucleotides reactive to cobra venom nuclease V1 were considered either to be involved in base-pairing interactions or in a helical single stranded arrangement (40). The pattern of reactivity was visualized by oligodesoxynucleotide

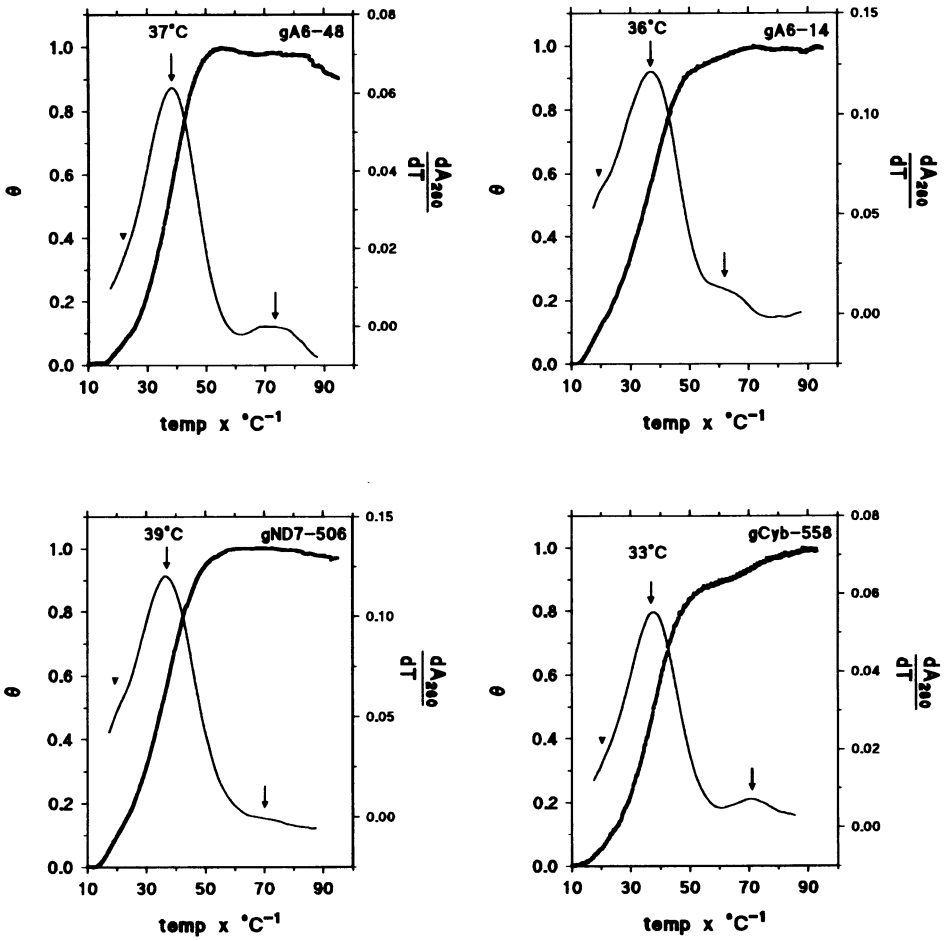

Figure 2. Representative set of melting profiles of the four tested gRNAs. Thermal denaturations were monitored at 260 nm in 50 mM potassium cacodylate pH 7.2, 150 mM KCl, 2.6 mM $MgCl_2$ and 0.1 mM Na_2EDTA. Plotted is the normalized absorbance (θ), as a function of the temperature (T) with $\theta = (A_T - A_i)/(A_f - A_i)$; A_T is the absorbance at temperature T, A_i the absorbance at the initial temperature (10°C), and A_f, the absorbance at the final temperature of 95°C (58). In addition, the first derivatives (dA_{260}/dT) of the melting curves are shown. Arrows annotate the two main melting transitions T_m–1 and T_m–2 and small arrowheads indicate low temperature transitions possibly due to tertiary interactions.

directed reverse transcription followed by gel electrophoresis of the cDNA products. To gain information about the nucleotides at the primer binding site we additionally performed ribonuclease digests with 3′ end-labelled gRNAs.

Representative autoradiographs of the structure probing experiments for all four gRNA molecules are shown in Figure 4. Figure 5 and Table 2 summarize the quantitative analysis of the experiments using guide RNA gND7-506 as an example. The data were used to construct secondary structure models which are summarized in Figure 6. As predicted from the UV melting curves, the four guide RNAs fold into similar structures containing two stem–loop elements which we termed 5′- and 3′ hairpin (see also Fig. 5, panel CV). The single-stranded region between the two helical domains varied from 3 to 11 nt, depending on the gRNA. The 5′-ends of the molecules were also single-stranded, varying in length from 2 to 11 nt. No evidence for a folding of the 5′ vector sequences was found since these nucleotides were always accessible to the single strand-specific chemicals (for example see Fig. 4, panel gCyb-558 and Table 2, nucleotides 1–10). Only in two experiments we detected one or two additional CV sites in gA6-48 and gA6-14 (Fig. 4, top two panels). However, these cuts were not reproducible and therefore not considered significant. Further support for a structural and functional silent behaviour of the vector derived sequences comes from the fact that a gA6-14 transcript containing the exact 5′ sequence was competent to form presumed RNA editing intermediates, so-called chimaeras, *in vitro* (20).

Invariably, the anchor sequences of the four molecules were identified as part of the 5′ hairpin. In three of the four cases, the 5′ helix consists of only 3 bp thus not requiring a large energy input for unwinding. In all cases the 5′ helix was the thermodynamically, less stable stem of the two helical elements.

Loop sizes varied between 3 and 7 nt with the exception of gCyb-558, where a 12 membered loop extends the 3′ stem structure. However, the 3′ part of that loop ends in an $A_{45}AAGG_{49}$ sequence, suggesting a stacking interaction on top of the helix which terminates with nucleotide G_{50}. Similarly, it seems possible that the two adenosines (A_{24} and A_{25}) in gND7-506 might extend the helical sequence $G_{21}UA_{23}$ of the 5′ hairpin of that molecule.

All four gRNA molecules showed evidence for a single-stranded conformation of the oligo(U) tail immediately following

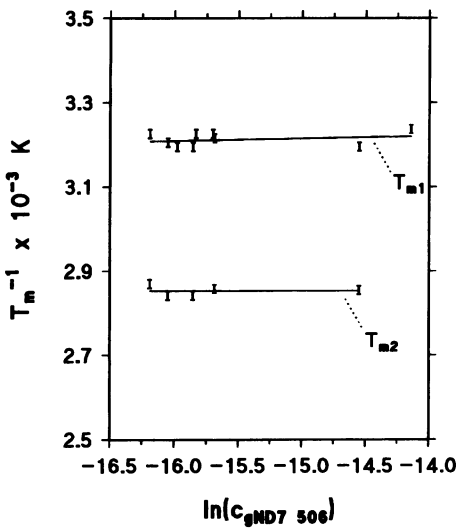

Figure 3. Concentration dependence of the two transition temperatures T_m–1 and T_m–2 for guide RNA gND7-506 (van't Hoff plot). The individual data points were fitted by a linear least squares calculation with error bars representing standard deviations.

the 3′ stem–loop. Adjacent to these signals we found double strand-specific CV endonuclease sensitive sites. In two cases (gA6-14, gND7-506) these sites overlapped the single-stranded signals. Two explanations can account for this result: the oligo(U) sequence might fold back on itself, thus adopting a double helical structure via U/U base-pairing (41,42) or, the U-tail might be in a helical, single-stranded conformation which has been shown to be a substrate for cobra venom endonuclease (40).

The modification data clearly revealed a fraying effect at both ends of the different helical elements. This made it difficult to exactly determine the closing base pair of the various stems and therefore, the most intense signal was used as an indication for the terminal base pair. In the majority of cases A–U interactions terminated the stems but U–Gs, as well as in one case a C–U basepair (gA6-14), (42,43) were also found. Taken together, we interpret these results as an indication for a dynamic breathing of the helical ends (44).

Termination signals in untreated RNA samples occurred predominantly at the junctions of helical domains to single-stranded sequences (see Fig. 5). Guide RNA gA6-48 possibly contains two G–A basepairs in its 3′ helix. The 5′ stem–loop of that molecule consists of three A–U bp interrupted by two looped out C nucleotides, C_{19} and C_{20}. The latter position also gave rise to a cut in the untreated control sample.

DISCUSSION

Guide RNAs play a central role during kRNA editing by functioning as quasi 'templates' in the process (3). The reaction is presumably initiated by the formation of a short antiparallel duplex structure, the so called anchor interaction, formed between a 5′ sequence stretch of gRNAs and complementary nucleotides within the pre-mRNAs located 3′ of an editing site. This anchor duplex becomes extended as the editing reaction proceeds, with the gRNAs primary sequence specifying the sequence of the edited mRNA by a combination of Watson–Crick and G–U base pairing.

Editing can be extensive and sometimes requires the consecutive action of multiple different gRNAs on a single pre-mRNA

Table 2. Summary of the chemical modification and enzyme accessibility data for gND7-506

nucleotide	modification / strength	enzyme accessibility
G1	Ke/1	none
G2	Ke/1, CMCT/0.5	T1/1
G3	Ke/1.5, sometimes control cut	T1/2.5
C4	weak control cut	weak control cut
G5	Ke/1.5, CMCT/0.5	T1/2
A6	DMS/1, DEP/3	T2/1
A7	DMS/1, DEP/3	T2/1
U8	CMCT/3	T2/0.5
U9	CMCT/3	T2/0.5
A10	DEP/ 3, DMS/ 0.5-1	none
U11	CMCT/ 4	none
A12	DEP/ 3, DMS/ 0.5-1	none
A13	DEP/ 3, DMS/ 1.5-2	none
A14	DEP/ non-reactive, DMS/ ≤ 0.5	none
U15	CMCT/ 2	CV/ 1-1.5
A16	DMS 1, DEP/ non-reactive	CV/ 2.5-3
C17	DMS/ 1	CV/ 2
G18	Ke/ 2	CV/ 1-1.5, T1/ 0-0.5
A19	DEP/ 4, DMS/ 3.5	CV/0-0.5
U20	CMCT/ 0.5	T2/≤ 0.5
G21	Ke/ 1.5	CV/0.5, T1/≤1.0
U22	none, sometimes control cut	CV/ 1-1.5
A23	DEP/ non-reactive, DMS/ 0.5-1	CV/ 1.5-2
A24	DMS/ 1, DEP/ 1	CV/ 2, T1*/ 0.5
A25	DMS/ 1, DEP/ 1.5-2	CV/ 1.5-2, T1*/1.0
U26	CMCT/ 2.5-3	CV/ 0.5, T1*/ 1.5, T2/ 0.5
A27	DMS/ 1, DEP/ 1.5	T1*/ 1.5, T2/ 0.5
A28	DMS/ 1, DEP/ 1	T1*/ 1.5, T2/ 0.5
C29	DMS/ 0.5-1	T1*/ 1.0, T2/ 0.5
C30	control cut	T1*/ 0.5
U31	control cut	CV/ 1.0
G32	none, sometimes control cut (very weak)	CV/ 2.0, (not T1)
U33	control cut	CV/ 2.5
A34	DMS / 0.5-1, DEP/ 1.5	CV/ 3.5
G 35	Ke/ 4	CV/ 4, T1/1.0, T2/0.5
U36	control cut	CV/ 3.5, T2/ 1.5
A37	DMS/ 1, DEP/ 3 sometimes control cut	CV/ 2.5, T2/ 2
U38	control cut	CV/ 1, T2/ 2.5
A39	DEP/ 1.5, DMS/ 1	CV/ ≤ 0.5, T2/ 4, T1/ 4
G40	Ke/ 1.5	T2/ 4, T1/ 4
U41	CMCT/ ≤ 1	T2/ 4, T1/ 4
U42	control cut	T2/ 4
A43	DEP and DMS non-reactive	T2/ 2.5
G44	Ke/ 1-1.5	T2/ 1.5-2, T1/ ≤0.5
U45	nd	T2/ 1.5
G46	nd	T2/ 0.5, T1/ 0.5
U47	nd	CV/ 0.5
A48	nd	CV/ 0.5-1, T1*/ ≤ 1.0
U49	nd	CV/ 1.5, T1*/ ≤ 1.0
A50	nd	CV/ 2.5
U51	nd	CV/ 3.5
A52	nd	CV/ 3.5-4
G53	nd	CV/ 3, (not T1), T2/ 0.5
U54	nd	CV/ 2, T2/ 0.5
G55	nd	CV/ 1, T2/ 0.5-1, (not T1)
A56	nd	T2/ 1, CV/ 0.5
A57	nd	T2/ 1
U58	nd	T2/ 0.5-1
U59	nd	T2/ 0.5-1
U60	nd	T2/ 0.5-1
U61	nd	T2/ 0.5-1, CV/ 0.5
U62	nd	T2/ 0.5-1, CV/ 0.5-1
U63	nd	T2/ 0.5-1, CV/ 0.5-1
U64	nd	T2/ 0.5, CV/ 1
U65	nd	CV/ 1
U66	nd	CV/ 1
U67	nd	CV/ 2

The data are averaged from five independent experiments. Relative reactivities were estimated from densitometer scans of the autoradiographs with a scale ranging from 0.5 to 4.0 equivalent to marginally reactive to hyperreactive. T1* refers to single-stranded non-G nucleotides sensitive to RNase T1 at high concentrations and nd annotates not determined. Double-stranded domains are shown as black rectangles, the 3′ oligo(U) tail as a hatched rectangle and the 5′ vector derived nucleotides as an open rectangle.

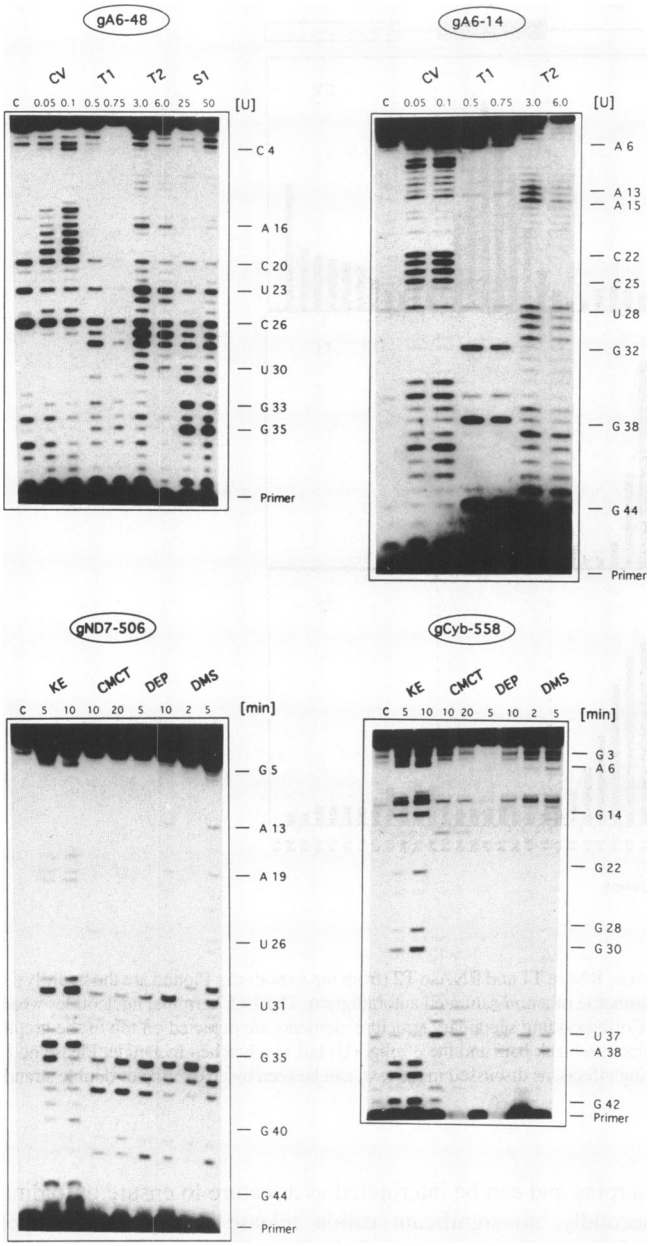

Figure 4. Structure mapping of *in vitro* transcribed *T.brucei* gRNAs. A representative set of autoradiographs for the four individual gRNAs is shown. The molecules were treated with limiting amounts of single strand specific reagents [KE = kethoxal, DEP = diethyl pyrocarbonate, DMS = dimethyl sulfate, CMCT = 1-cyclohexyl-3(2-morpholinoethyl) carbodiimide metho-*p*-toluene sulfonate] for incubation times at 27°C as indicated on the top of the autoradiographs. Digestions with RNases T1 and T2 or nuclease S1 were performed with different enzyme activities as indicated. CV annotates treatment with the double strand specific RNase V1 from cobra venom. Cleavage/modification sites were detected using primer extension analysis and are indicated on the right. Lane C annotates a control sample with no enzyme or modification reagent added but treated identically throughout the experiment.

(7,13,14). Based on the paradigm that different RNA molecules with an identical biochemical function fold into analogous higher order structures (15,45), we were interested whether gRNA molecules adopt a common secondary structure. To test this hypothesis, we used a set of chemical and enzymatic structure probes combined with thermal denaturation studies to experimentally verify the secondary structures of four different *T.brucei* guide RNAs.

Our results show that the RNAs in solution, despite their different length and primary sequences, fold into similar secondary structures with Gibb's free energies below the predicted most stable conformation. A statistical analysis indicated that these structures do not occur by chance in randomized RNAs of identical length and base composition. The molecules fold into two consecutive hairpin elements separated by a single-stranded region of variable length. The temperature transition profiles can therefore be interpreted as primarily the melting of the 3′ stem–loop structures with the melting of the 5′ hairpins as a superimposition which can not be dissected by thermodynamic methods. Although the structures are not identical, their similarity is high enough to rationalize the previous finding that different gRNAs bind to the same mitochondrial proteins, which had indicated similar protein binding motifs on the various RNAs (16–18).

The 5′-ends of the four structures are in a single-stranded conformation followed by a small hairpin which contains the anchor sequence presumed necessary for the initial interaction with the pre-mRNAs. This strongly suggests an unwinding of that stem structure to initiate the editing reaction. Within the 3′ U-tails of the RNAs we noticed an overlap of single strand-specific enzyme cuts with double strand-specific sensitivities (see Fig. 5). As mentioned above, this can be explained either by a stacked, single-stranded, helical conformation of the U-tail or, alternatively, by folding back of the U-tail on itself. Although U–U basepairing has been demonstrated in the central part of a synthetic RNA double helix (42) as well as for the homopolymer poly(U) at low temperatures (~5°C) (41), we favour the helical single strand explanation for the following reasons. First, our probing experiments were performed at 27°C, much above the stable formation of isolated U–U base pairs. Secondly, a high degree of flexibility is expected for the oligo(U) tail. Guide RNAs have been shown to be covalently linked, via their U-tails, to multiple editing sites in chimaeric gRNA/pre-mRNA molecules (10,20) and some of these linkage points can be several editing sites apart, despite the fact that the gRNA is always anchored at the same position (10,46). Precedence for a helical conformation of non-basepaired nucleotides can be found in the 3′ ends of tRNA molecules, where the CCA sequence extends the helical geometry of the acceptor stem (47) as well as in the homopolymers poly(C) and poly(A) (48). Since the gRNA U-tail has been suggested to function as the reservoir for the U-addition/deletion reaction (10,11), the single-strandedness of that part of the molecule has an interesting consequence for the refolding and potential reuse of gRNAs. After the release of the molecules from the mRNA, gRNAs should be able to refold rapidly into the above described 2D-structures since a truncated or extended U-tail should have no consequence for a proper refolding reaction.

We have not identified structural differences for gRNAs acting at a 3′ editing domain (gA6-14, gND7-506, gCyb-558) versus late acting gRNAs (gA6-48). However, we can not rule out that tertiary interaction might distinguish between the two types of gRNAs. An indication for such 'higher order interactions' may be deduced from the asymmetry of the differentiated melting curves at low temperatures. Additionally, the fact that the 5′ haipins could be identified at 27°C can be interpreted as circumstantial

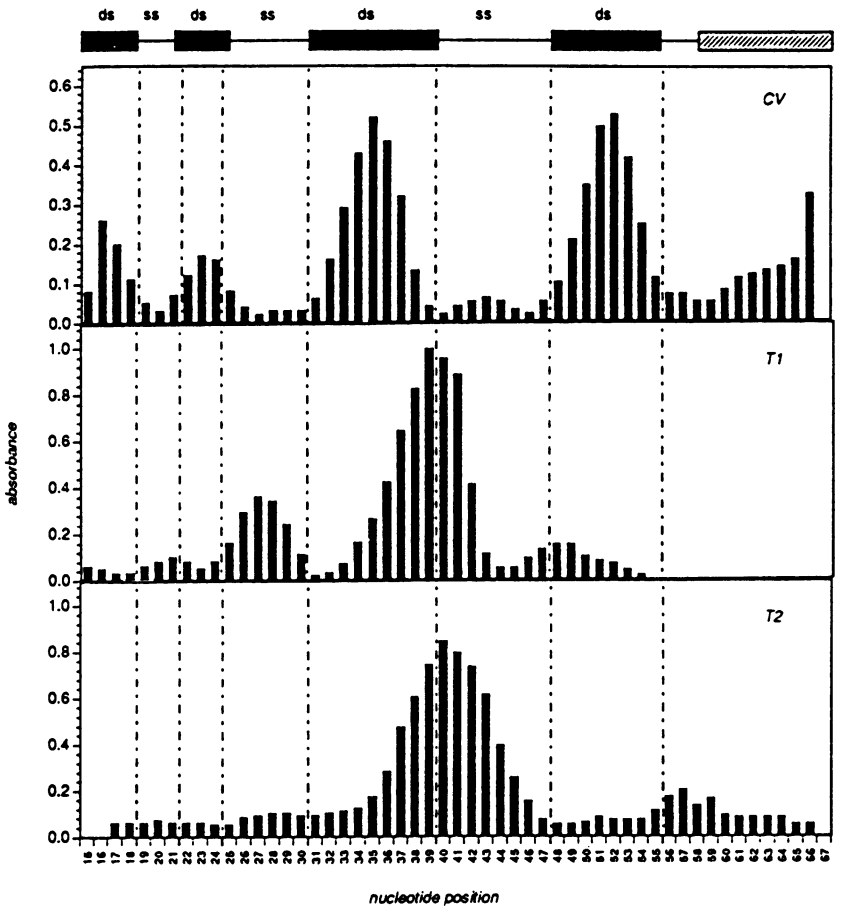

Figure 5. Hydrolysis profiles of 3′ end-labelled gND7-506 digested with CV endonuclease, RNase T1 and RNAse T2 (from top to bottom). Plotted are the hydrolysis intensities at the individual nucleotide positions in absorbance units generated by densitometrie of a non-saturated autoradigram. The 14 5′ terminal nucleotides were not properly resolved on the various gels and hence were not quantitatively assessed. Corresponding secondary structure elements are depicted on top of the graph with double stranded (ds) domains shown as black rectangles, single-stranded (ss) regions as black bars and the 3′ oligo(U) tail as a hatched rectangle. Please note the four helical domains, equivalent to two stem structures, in the CV experiment. Fraying effects, as discussed in the text, can be seen by an overlap of double strand and single strand specific intensities, most notably at positions 36–39.

evidence for tertiary interaction which might stabilise RNA domains of otherwise low thermodynamic stability. A computer simulation to search for pseudoknot structures (34) did not reveal any potential sequences. We have no satisfactory explanation for the transition at T_m–2. High temperature melting processes have been reported for other small size RNAs (49) and were interpreted as an aggregation phenomenon. This cannot be used as an explanation in our system, since T_m–2 was concentration independent (see Fig. 3).

The most intriguing feature of the four structures is their low thermodynamic stability. The molecules unfold at temperatures as low as 33–39°C with melting enthalpies of –32 to –38 kcal/mol. A possible explanation for this behaviour can be found in the proposed function of the gRNAs. Presumably, complete unfolding has to occur during the editing reaction in order to allow for basepairing with the pre-mRNA molecule (36). A structurally stable gRNA would require a large energy input for unwinding, whereas a molecule of low thermodynamic stability can be unfolded with only a minimal energy requirement. Three additional facts support such a model. First, the observed 'fraying' of the helical ends suggests a dynamic breathing of the hairpins and can be interpreted as a feature to ensure unfolding. Secondly, no significant cation release was measured upon unfolding of the structures, indicating the absence of substantial helical domains which have higher charge densities than single-stranded regions (27,39). Thirdly, none of the four molecules could specifically be cleaved by Pb^{2+} cations (Schmid and Göringer, unpublished) a reaction which requires a specific coordination of the cation for hydrolysis and only tightly folded RNA structures can provide such a structural context (50 and references therein).

The double stem–loop structures might serve as protein recognition sites as well as nucleation points for a complete folding of the RNAs. The individual structures seem to be fine tuned to minimize stability for an optimal unfolding reaction while at the same time maximizing higher order structural features in order to allow a specific association with mitochondrial proteins. Evidence for a protein association to gRNAs has been demonstrated in several experimental systems (51) and simple RNA hairpin structures have been characterized in other cases as protein recognition elements (52,53). Interestingly, one of the gRNA binding proteins, a 90 kDa polypeptide, was shown

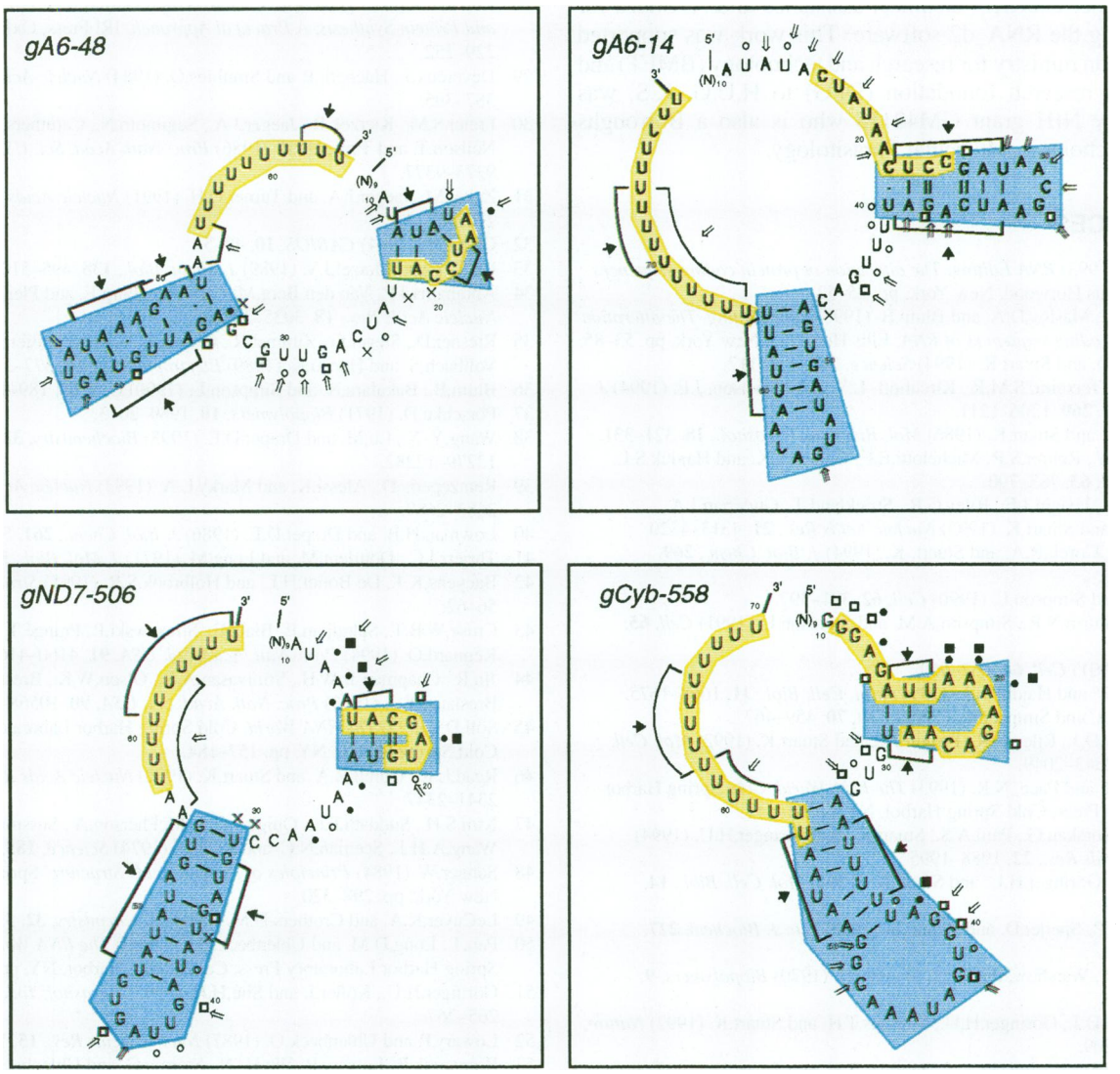

Figure 6. Secondary structure models (59) of the four individual gRNAs summarizing the modification/digestion data. Sensitivities of the RNA molecules to the various reagents and enzymes are indicated by the following symbols: Kethoxal, open square; DMS, filled circle; DEPC, filled square; CMCT, open circle; CV, filled arrow plus bracket, S1, T1; T2, open arrow. X annotates frequent termination sites in untreated control samples. The two stem–loop domains in the four molecules are shaded in blue. Oligo(U) tails and anchor regions are shaded in yellow.

to recognize the 3′ oligo(U) extension of gRNAs (16–18). Based on the data presented here, this might be a phenomenon similar to the RNA–protein interaction of the La protein from HeLa cells which specifically interacts with a short stretch of uridylate residues, 3′ to a stem structure (54,55). Other potential recognition elements for the interacting proteins may be the single strand/helix junctions, the single stranded terminal ends or bulged out nucleotides. In analogy to what has been recently identified for the RNA–RNA interaction of naturally occurring antisense RNAs to their target sequences, such structural features, like the bulged out nucleotides, might be necessary to initiate (and propagate) stable base-pairing between gRNAs and pre-mRNAs or to prevent RNase cleavage (56,57).

In conclusion, we have shown that four different gRNA molecules from *T.brucei* adopt simple, double stem–loop structures with stabilities below the thermodynamic most stable conformation. The depicted secondary structure models can be considered first approximations for potential higher order structures which at present are not known. We interpret the low stability features of the molecules as a direct consequence of the function of gRNAs which presumably completely unfold during the editing reaction. The molecules are not locked into a rigid structure which would require a substantial energy input for unwinding, but rather, fold into flexible structures which can be easily perturbed by external factors. This is consistent with the hypothesis that gRNAs are likely to be functional components of the editing machinery similar to flexible domains in large RNA molecules that mediate functional properties.

ACKNOWLEDGEMENTS

We wish to thank A. Missel, J. Köller, U. Müller and S. Parchmann for helpful discussions and A. Paul, H. H. Shu and A. Souza for critical reading of the manuscript. G. Nörskau is

thanked for her excellent technical assistance and A. Souza for implementing the RNA_d2 software. This work was supported by the German ministry for research and technology (BMFT) and the German research foundation (DFG) to H.U.G. K.S. was supported by NIH grant GM42118 who is also a Burroughs Wellcome Scholar in Molecular Parasitology.

REFERENCES

1 Stuart,K. (1993) *RNA Editing. The alteration of protein coding sequences of RNA*. Ellis Horwood, New York, pp. 25–52.
2 Simpson,L., Maslov,D.A. and Blum,B. (1993) *RNA Editing. The alteration of protein coding sequences of RNA*. Ellis Horwood, New York, pp. 53–85.
3 Seiwert,S.D. and Stuart K. (1994) *Science*, **266**, 114–117.
4 Kim, K.S., Teixeira, S.M.R., Kirchhoff, L.V. and Donelson, J.E. (1994) *J. Biol. Chem.* **269**, 1206–1211.
5 Jasmer,D.P. and Stuart,K. (1986) *Mol. Biochem. Parasitol.*, **18**, 321–331.
6 Pollard,V.W., Rohrer,S.P., Michelotti,E.F., Hancock,K. and Hajduk,S.L. (1990) *Cell*, **63**, 783–790.
7 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.
8 Riley,G.R., Corell,R.A. and Stuart, K. (1994) *J. Biol. Chem.*, **269**, 6101–6108.
9 Blum,B. and Simpson,L. (1990) *Cell*, **62**, 391–397.
10 Blum,B., Sturm,N.R., Simpson,A.M. and Simpson,L. (1991) *Cell*, **65**, 543–550.
11 Cech,T. (1991) *Cell*, **64**, 667–669.
12 Pollard,V.W. and Hajduk,S.L. (1991) *Mol. Cell. Biol.*, **11**, 1668–1675.
13 Maslov,D.A. and Simpson,L. (1992) *Cell*, **70**, 459–467.
14 Koslowsky,D.J., Riley,G.R., Feagin,J.E. and Stuart,K. (1992) *Mol. Cell. Biol.*, **12**, 2043–2049.
15 Woese,C.R. and Pace, N.R. (1993) *The RNA World*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 91–117.
16 Köller,J., Nörskau,G., Paul,A.S., Stuart,K. and Göringer,H.U. (1994) *Nucleic Acids Res.*, **22**, 1988–1995.
17 Read,L.K., Göringer,H.U. and Stuart,K. (1994) *Mol. Cell. Biol.*, **14**, 2629–2639.
18 Leegwater,P., Speijer,D. and Benne,R. (1995) *Eur. J. Biochem.* **227**, 780–786.
19 Cantor,C.R., Warshow,M.M. and Shapiro,H. (1970) *Biopolymers*, **9**, 1059–1077.
20 Koslowsky,D.J., Göringer,H.U., Morales,T.H. and Stuart,K. (1992) *Nature*, **356**, 807–809.
21 Göringer,H.U., Koslowsky,D.J., Morales,T.H. and Stuart,K. (1994) *Proc. Natl. Acad. Sci. USA*, **91**, 1776–1780.
22 Reyes,V.M. and Abelson,J.N. (1989) *Methods Enzymol.* Vol. 180, pp. 63–69.
23 Puglisi,J. and Tinoco,I. (1989) *Methods Enzymol.* Vol. 180, pp. 304–325.
24 Bruce,A.G. and Uhlenbeck,O.C. (1978) *Nucleic Acids Res.*, **5**, 3665–3676.
25 Marky,L.A. and Breslauer,K.J. (1987) *Biopolymers*, **26**, 1601–1620.
26 Hung,S.-H., Yu,Q., Gray,D.M. and Ratcliff,R.L. (1994) *Nucleic Acids Res.*, **22**, 4326–4334.
27 Rentzeperis,D., Rippe,K., Jovin,T.M. and Marky,L.A. (1992) *J. Am. Chem. Soc.*, **114**, 5926–5928.
28 Christiansen,J., Egebjerg,J., Larsen, N. and Garrett,R.A. (1990) *Ribosomes and Protein Synthesis. A Practical Approach*. IRl Press, Oxford, pp. 229–252.
29 Devereux,J., Haeberli,P. and Smithies,O. (1984) *Nucleic Acids Res.*, **12**, 387–395.
30 Freier,S.M., Kierzek,R., Jaeger,J.A., Sugimoto,N., Caruthers,M.H., Neilson,T. and Turner,D.H. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 9373–9377.
31 Zuker,M., Jaeger,J.A. and Turner,D,H. (1991) *Nucleic Acids Res.*, **19**, 2702–2714.
32 Gast,F.U. (1994) *CABIOS*, **10**, 49–51.
33 Le,S.-Y and Maizel,J.V. (1989) *J. Theor. Biol.*, **138**, 495–510.
34 Abrahams,J.P., Van den Berg,M., Van Batenburg,E. and Pleij,C. (1990) *Nucleic Acids Res.*, **18**, 3035–3044.
35 Riesner,D., Steger,G., Zimmat,R., Owens,R.A., Wagenhöfer,M., Hillen,W., Vollbach,S. and Henco,K. (1989) *Electrophoresis*, **10**, 377–389.
36 Blum,B., Bakalara,N. and Simpson,L. (1990) *Cell*, **60**, 189–198.
37 Pörschke,D. (1971) *Biopolymers*, **10**, 1989–2013.
38 Wang,Y.-X., Lu,M. and Draper,D.E. (1993) *Biochemistry*, **32**, 12279–12282.
39 Rentzeperis,D., Alessi,K. and Marky,L.A. (1993) *Nucleic Acids Res.* **21**, 2683–2689.
40 Lowman,H.B. and Draper,D.E. (1986) *J. Biol. Chem.*, **261**, 5396–5403.
41 Thrierr,J.C., Dourlent,M. and Leng,M. (1971) *J. Mol. Biol.*, **58**, 815–830.
42 Baeyens,K.J., De Bondt,H.L. and Holbrook,S.R. (1995) *Struct. Biol.* **2**, 56–62.
43 Cruse,W.B.T., Saludjian,P., Biala,E., Strazewski,P., Prangé,T. and Kennard,O. (1994) *Proc. Natl. Acad. Sci. USA*, **91**, 4160–4164.
44 Jin,R., Chapman Jr,W.H., Srinivasan,A.R., Olson,W.K., Breslow,R. and Breslauer,K.J. (1993) *Proc. Natl. Acad. Sci. USA*, **90**, 10568–10572.
45 Söll,D. (1993) *The RNA World*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 157–184.
46 Read,L.K., Corell,R.A. and Stuart,K. (1992) *Nucleic Acids Res.* **20**, 2341–2347.
47 Kim,S.H., Suddath,F.L., Quigley,G.J., McPherson,A., Sussman,J.L., Wang,A.H.J., Seeman,N.C. and Rich,A. (1974) *Science*, **185**, 435–440.
48 Sänger,W. (1984) *Principles of Nucleic Acids Structure*. Springer Verlag, New York, pp. 298–320.
49 LeCuyer,K.A. and Crothers,D.M. (1993) *Biochemistry*, **32**, 5301–5311.
50 Pan,T., Long,D.M. and Uhlenbeck,O.C. (1993) *The RNA World*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 271–302.
51 Göringer,H.U., Köller,J. and Shu,H.H. (1995) *Parasitol. Today* **11**, 265–267.
52 Lowary,P. and Uhlenbeck,O. (1987) *Nucleic Acids Res.*, **15**, 10483–10439.
53 Romaniuk,P., Lowary,P., Wu,H.-N., Storma,G. and Uhlenbeck,O. (1987) *Biochemistry*, **26**, 1563–1568.
54 Stefano,J.E. (1984) *Cell*, **36**, 145–154.
55 Kenan,D.J., Query,C.C. and Keene,J.D. (1991) *Trends Biol. Sci.*, **16**, 214–220.
56 Hjalt,T.Å.H. and Wagner,E.G.H. (1995) *Nucleic Acids Res.*, **23**, 580–587.
57 Hjalt,T.Å.H. and Wagner,E.G.H. (1995) *Nucleic Acids Res.*, **23**, 571–579.
58 Sheardy,R.D., Levine,N., Marotta,S., Suh,D. and Chaires,J.B. (1994) *Biochemistry*, **33**, 1385–1391.
59 Perochon-Dorisse,J., Chetouani,F., Aurel,S., Iscolo,N. and Michot,B. (1995) *CABIOS*, **11**, 101–109.