

A three-dimensional working model for a guide RNA from *Trypanosoma brucei*

Thomas Hermann^{1,+}, Beate Schmid[§], Hermann Heumann¹ and H. Ulrich Göringer^{*}

Laboratorium für Molekulare Biologie, Genzentrum der Ludwig Maximilians Universität München, 82152 Martinsried, Germany and ¹Max-Planck-Institut für Biochemie, 82152 Martinsried, Germany

Received March 19, 1997; Revised and Accepted April 23, 1997

ABSTRACT

RNA editing in protozoan parasites is a mitochondrial RNA processing reaction in which exclusively uridylate residues are inserted into, and less frequently deleted from, pre-mRNAs. Molecules central to the process are so-called guide RNAs (gRNAs) which function as templates in the reaction. For a detailed molecular understanding of the mechanism of the editing process knowledge of structural features of gRNAs will be essential. Here we report on a computer-assisted molecular modelling approach to construct the first three-dimensional gRNA model for gND7-506, a ND7-specific gRNA from *Trypanosoma brucei*. The modelling process relied on chemical modification and enzymatic probing data and was validated by *in vitro* mutagenesis experiments. The model predicts a reasonably compact structure, where two stem/loop secondary structure elements are brought into close proximity by a triple A tertiary interaction, forming a core element within the centre of the molecule. The model further suggests that the surface of the gRNA is primarily made up of the sugar-phosphate backbone. On the basis of the model, footprinting experiments of gND7-506 in a complex with the gRNA binding protein gBP21 could successfully be interpreted and provide a first picture for the assembly of gRNAs within a ribonucleoprotein complex.

INTRODUCTION

Guide RNAs (gRNAs) are small, metabolically stable RNA molecules only identified within the mitochondria of kinetoplastid protozoan parasites (1). The molecules are *trans*-acting molecular components and play a central role during the maturation of mitochondrial pre-mRNAs, a process known as kinetoplastid (k)RNA editing (reviewed in 2). During kRNA editing, uridine nucleotides are site-specifically inserted and deleted into otherwise non-translatable mRNA molecules. The primary sequences of gRNAs contain the information for the processing reaction and this information is transferred to the

pre-edited mRNAs in a base pairing interaction. The process is likely catalysed by a series of enzymatic reactions, presumably acting within a large ribonucleoprotein (RNP) complex (3–5).

gRNAs have an average length of 50–70 nt with a strong A/U nucleotide bias. On average 15 uridines are post-transcriptionally added to their 3'-ends. Our current understanding of the reaction steps of an editing cycle allows the assignment of three basic functions for gRNAs. First, by base pairing to the pre-edited mRNAs proximal to an editing site, they define the endonucleolytic cleavage sites within the mRNA molecules (6). Second, as mentioned above, they function as templates for U insertion and deletion. Third, they potentially prevent the diffusion and loss of the upstream mRNA portion by base pairing of the 3' oligo(U) tail following endonucleolytic cleavage (5).

The guiding capacity of gRNAs is determined by the overall length of the molecules and, as a consequence, multiple gRNAs are generally required for complete editing of a specific pre-mRNA (7,8). These gRNAs, though differing in primary sequence, function in the same biochemical reaction and are assembled into an identical RNP complex (3–5, reviewed in 9). As a working hypothesis we propose that, due to the lack of similarity within the primary structure of gRNAs, it is higher order folding rather than a common sequence motif that is essential for recruiting gRNAs into the editing apparatus. This notion is supported on two accounts. First, different gRNAs have been shown to become cross-linked to the same set of mitochondrial proteins, which indicated common binding sites for the polypeptides (10–12). Second, a detailed structure probing analysis of four *Trypanosoma brucei* gRNA molecules demonstrated that these RNAs have similar secondary structures despite their variable primary sequences (13). The foldings are characterized by two imperfect stem/loop elements with both the 5'- and 3'-ends in a single-stranded conformation. Aside from these data, no other structural information exists for gRNAs, although >200 potential gRNA sequences from eight different kinetoplastid organisms have been published (14).

Here we describe the computer-assisted molecular modelling of the three-dimensional architecture of gRNA gND7-506 from *T. brucei*. As primary input data we used published information from accessibility probing experiments where base-specific modification reagents and structure-specific enzymatic probes

* To whom correspondence should be addressed. Tel: +49 89 8578 2475; Fax: +49 89 8578 3810; Email: goeringer@biochem.mpg.de

Present addresses: ⁺Institut de Biologie Moléculaire et Cellulaire, 67084 Strasbourg, France and [§]Memorial Sloan-Kettering Cancer Center, New York, NY 10021, USA

were used (13). The reliability of the model was verified by examining two gND7-506 mutant RNAs where base interactions predicted from the model were either disrupted or altered. The proposed three-dimensional model is supported by footprinting data of gND7-506 in a complex with gBP21, a gRNA binding protein from *T.brucei* (15). The model is intended as a basis for further work on the structure/function correlation of gRNAs during the editing reaction.

MATERIALS AND METHODS

General modelling procedures

Graphical model building was performed using the SYBYL molecular modelling software package (Tripos, St Louis, MO) on a Silicon Graphics Indy workstation. Geometries of standard A-form ribonucleotides were taken from the monomer library implemented in SYBYL, which uses the coordinates from Arnott and Hukins (16). Bond geometries in graphically constructed models were corrected by local minimization under the AMBER forcefield (17), ignoring electrostatic contributions but taking into account hydrogen bonding. Coordinates of the model can be obtained from the authors upon request.

Biochemicals

Diethylpyrocarbonate (DEPC) was purchased from Serva and dimethylsulfate (DMS) from Merck. Ribonucleases T1 and T2, T4 polynucleotide kinase and T4 RNA ligase were from Gibco BRL. Cobra venom RNase V1 was purchased from USB and AMV reverse transcriptase from Stratagene. [³²P]pCp (3000 Ci/mmol), [α -³²P]ATP (3000 Ci/mmol) and [γ -³²P]ATP (5000 Ci/mmol) were purchased from NEN or Amersham. Oligodeoxynucleotides were synthesized by solid support chemistry using *O*-cyanoethyl-*N,N*-diisopropylphosphoramidites.

RNA synthesis and construction of mutant gND7-506

gRNAs were transcribed from linearized plasmid DNA templates with 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 MgCl₂, 0.1 mM Na₂EDTA. The RNAs were renatured after purification (95°C, 2 min, followed by cooling the samples to 25°C at a rate of 1°C/min). RNA concentrations were determined by UV absorbency measurements at 260 nm using extinction coefficients (ϵ_{260} l/mol/cm) calculated from tabulated values for the di- and mononucleotides (18).

Synthetic genes for two mutated gND7-506 RNAs [single site mutations at either position 43 from A to C (A₄₃C) or at position 14 from A to U (A₁₄U)] were constructed by self-assembly of overlapping 5'-phosphorylated oligodeoxynucleotides according to Reyes and Abelson (19). The annealed DNA fragments were cloned into plasmid pBS- (Stratagene) and transformed into competent *Escherichia coli* SURE cells (Stratagene). Positive clones were identified by restriction enzyme digestion of isolated plasmid DNA and sequenced to verify the desired base change.

Radioactive end-labelling

Oligodeoxynucleotides and gRNAs were 5'-end-labelled using T4 polynucleotide kinase and [γ -³²P]ATP. gRNAs were 3'-end-

labelled using T4 RNA ligase and [³²P]pCp according to standard procedures. All radiolabelled nucleic acids were purified by denaturing polyacrylamide gel electrophoresis.

RNA structure probing and primer extension analysis

Chemical modification reactions were performed as described by Christiansen *et al.* (20) using DMS and DEPC as modifying reagents. Reactions were performed with 0.2 μ g gRNA and 0.8 μ g bulk yeast tRNA as a carrier in a volume of 10 μ l. Incubation was for 5 and 10 min for DEPC modifications or 2 and 5 min for reactions with DMS at 27°C, which is the optimal growth temperature of insect stage trypanosomes.

Reverse transcription reactions were performed using the oligodeoxyribonucleotide primer AAAATTCACCTATATA, complementary to positions 47–65 of gND7-506. Modified gRNAs (1.3 pmol) and 5'-end-labelled primer (1.0 pmol, ~50 000 c.p.m.) were annealed in 10 mM Tris-HCl, pH 6.9, 40 mM KCl, 5 mM Na₂EDTA at 75°C for 30 s, followed by incubation at room temperature for 30 min. Extension reactions were performed at 37°C for 30 min with 1 U AMV reverse transcriptase per 10 μ l reaction. Complementary DNA products were separated on denaturing 10% (w/v) polyacrylamide gels.

gND7-506/gBP21 complex formation and footprinting experiments

Recombinant (r) gBP21 protein (15) was mixed with 5'- or 3'-end-labelled gND7-506 RNA in 60 μ l 6 mM HEPES, pH 7.5, 50 mM KCl, 2.6 mM MgCl₂, 0.1 mM Na₂EDTA, 0.5 mM DTT, 6% (v/v) glycerol. To ensure complete binding of the gRNA, the protein was used in at least a 10-fold molar excess over RNA. Association of the two molecules was achieved at 27°C for 30 min. Complexes were digested with RNase T1 (125 mU/ μ l), RNase T2 (100 mU/ μ l) and RNase V1 from cobra venom (25 mU/ μ l) at 27°C for 7.5 min in the presence of yeast tRNA (0.25 μ g/ μ l) as a carrier. Control samples did not contain any enzyme but were otherwise treated identically throughout the experiment. Reactions were stopped by the addition of KOAc, pH 6 (final concentration 270 mM), followed by extraction with phenol and ethanol precipitation and were analysed on 10% (w/v) polyacrylamide gels containing 8 M urea. Experiments were performed with three different r-gBP21 isolates that had been tested for their binding activity in nitrocellulose filter binding assays (15,21).

Circular dichroism measurements

CD spectra were measured at 27°C in 6 mM HEPES, pH 7.5, 50 mM KCl, 2.6 mM MgCl₂, 0.1 mM Na₂EDTA, 0.5 mM DTT, 6% (v/v) glycerol. Spectra were recorded from 300 to 195 nm with data acquisition every 0.1 nm. Scans were repeated 10 times and averaged. Mean molar residue ellipticities (θ_m) were calculated per mole of nucleotide monomer.

Temperature-dependent UV spectroscopy

Absorbance versus temperature profiles (melting curves) were recorded at 260 nm in 50 mM potassium cacodylate, pH 7.5, 50 mM KCl, 2.6 mM MgCl₂, 0.1 mM Na₂EDTA, 6% (v/v) glycerol 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. T_m values were determined from derivative plots of absorbance versus temperature as $0.5 \Delta A_{260}$.

RESULTS AND DISCUSSION

gRNA gND7-506

Trypanosoma brucei gRNA gND7-506 directs the editing of a sequence domain near the 5'-end of the mRNA for subunit 7 of NADH dehydrogenase (ND7) (8). Within that sequence stretch 15 uridylyate residues are inserted and six U nucleotides are deleted. gND7-506 has been identified in total RNA preparations from both procyclic and bloodform stage trypanosomes (8). A secondary structure model for gND7-506 was derived from surface probing experiments with base-specific modification reagents and structure-specific RNases in combination with temperature-dependent UV spectroscopy data (13). Figure 1A outlines the proposed structure, which consists of two hairpin elements, termed stem/loop I and stem/loop II, separated by seven single-stranded nucleotides. Both terminal ends of the molecule are likely in a single-stranded conformation, with evidence for a partly helical arrangement of the 3' oligo(U) tail (13). Figure 1B shows a representative autoradiogram of a RNase accessibility experiment which emphasises the various domains of the gND7-506 RNA.

Construction of the three-dimensional model

Starting from the secondary structure, construction of the three-dimensional model was performed in a stepwise fashion, beginning with independent modelling of the two helices. As initial building blocks we used coordinates from sequence-identical domains of either known RNA crystal structures or NMR-derived RNA solution structures stored in the PDB database. For stem I, coordinates from the PDB files 4TRA (tRNA^{Phe} from *E. coli*; 22), 1SCL (sarcin/ricin loop of rat 28S rRNA; 23) and 1RNA [synthetic poly(UA) helix; 24] were used. Stem II was constructed from sub-structures of PDB files 1ELH (helix I of *E. coli* 5S rRNA; 25) and 1RNK (synthetic pseudoknot; 26), in addition to the aforementioned files 1SCL, 1RNA and 4TRA. Coordinates for the three G/U base pairs within gND7-506 (U₃₁/G₅₅, G₃₂/U₅₄, U₃₃/G₅₃) were derived by superimposition of several G/U pairs obtained from four crystal structures of tRNA^{Phe} (PDB files 2TRA, 3TRA, 4TRA and 5TRA) (22) and from G/U base pairs in 1ELH and 1RNK. The A/A bulge within stem II (positions A₃₄ and A₅₂) was modelled according to a similar motif in the sarcin/ricin loop (PDB file 1SCL). Linking of the individual base pairs was done taking into account standard A-form double strand RNA geometry, resulting in a base stacking pattern in line with the experimental CV probing data (13). For an additional verification, we performed circular dichroism (CD) measurements and confirmed all typical A-form features (27) for gND7-506: a large negative ellipticity at 210 nm, a moderate negative ellipticity at 240 nm and a positive band around 265 nm (data not shown).

The bases of several single-stranded nucleotides immediately flanking stem I were modelled under the assumption of optimal stacking, again, as indicated by the CV probing data (13). G₁₈, the first loop nucleotide of stem/loop I was built to stack on top of base C₁₇ and bases A₂₄, A₂₅ and U₂₆ were arranged to staple onto A₂₃, thus extending the A-type conformation of helix I.

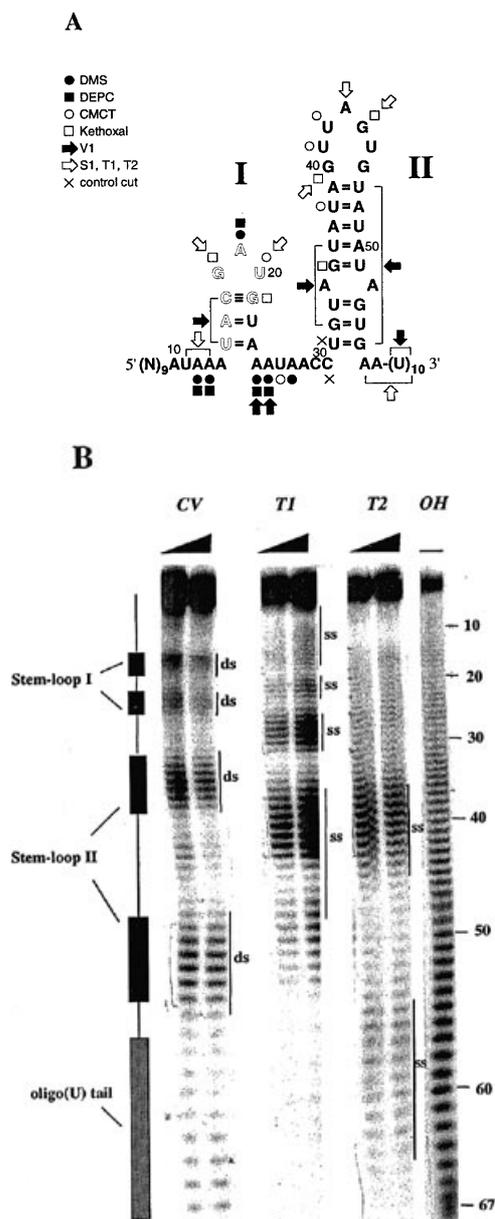


Figure 1. (A) Secondary structure model of *T. brucei* gRNA gND7-506, summarizing the chemical modification and enzyme accessibility data used as input in the modelling procedure (13). DMS, dimethylsulfate; DEPC, diethylpyrocarbonate; CMCT, 1-cyclohexyl-3-(2-morpholinoethyl)carbodiimide; T1, RNase T1 (G-specific); T2, RNase T2 (single strand-specific); S1, nuclease S1 (single strand-specific); CV, cobra venom nuclease V1 (specific for double strands and stacked bases). The two stem/loop elements are marked as I and II. Bases near the 5'-end which are involved in initial base pairing to the pre-mRNAs ('anchor interaction') are given as outline letters. (B) Representative autoradiogram of an enzyme accessibility experiment with cobra venom nuclease V1 (CV), RNases T1 and T2 and ³²P 3'-end-labelled gND7-506 RNA as described (13). The double-stranded (ds) domains of the RNA are shown as black rectangles, single-stranded (ss) regions are given as black bars and the 3' oligo(U) tail as a grey rectangle. OH refers to an alkaline hydrolysis ladder.

Restraining G₁₈ as described reduced the possible conformational space for modelling of the 3 nt loop of stem I (bases G₁₈-A₁₉-U₂₀). The final structure of the loop was assessed using the following data: A₁₉ was strongly modified by DMS and

DEPC and thus was arranged to be fully solvent accessible by pointing outward from the loop. In contrast, U₂₀ was only weakly reactive towards 1-cyclohexyl-3-(2-morpholinoethyl)carbodiimide (CMCT) and, as a consequence, was arranged to point to the inside of the loop.

For the relative orientation of the two helices to each other we examined the experimental probing data for inaccessible base positions and identified several adenosine residues which, based on the secondary structure model, showed unexpected reactivities: A₁₄ at the 5' base of stem/loop I was unreactive towards DEPC and any of the single strand- and double strand-specific enzymatic probes. It reacted only very weakly with DMS. Similarly, A₄₃ within the loop on top of stem II was unreactive towards DMS and DEPC, but reactive towards the single strand-specific RNase T2. Finally, the two adenines at positions 24 and 25 were DMS and DEPC reactive but also displayed a sensitivity towards the double strand-specific cobra venom enzyme. An inspection of the possible orientations of stem I and stem II to accommodate the modification data of A₁₄ and A₄₃ revealed a possible triple A interaction involving in addition A₂₄ at the 3'-end of stem I. Precedence for tertiary interactions of three adenylate residues has been reported for tRNA^{Gln} (28) and thus we further investigated this possibility in more detail. Due to their proximity to stem I, the positions of A₁₄ and A₂₄ were more or less fixed and the resulting A/A interaction caused a slight widening of the helix, probably explaining the observed fraying effect in the chemical and enzymatic probing studies (13). For the docking of A₄₃ to A₁₄/A₂₄ at the base of stem I, two mutually exclusive orientations were conceivable. While the two possibilities lead to a different packing of stems I and II, considering the given length of the single-stranded region between the two stems only one orientation remained as a realistic alternative. Experimental evidence for the existence of the triple A interaction in gND7-506 will be shown below. After fixing the orientations of stems I and II, the best arrangement of the stems was obtained by a manual docking procedure which continuously monitored the distances of the RNA fragment termini that had to be connected by single-stranded regions (Distance Monitoring option, SYBYL software).

Lastly, the single-stranded but very likely stacked 3' oligo(U) extension was modelled from the poly(U) half of a poly(A/U) RNA helix and further optimized by forcefield minimization *in vacuo* (17). The tail was joined with stem/loop II and the resulting final model is outlined in Figure 2. It is characterized by a tight parallel packing of the two stem structures whose major grooves face each other. The 3' oligo(U) tail is dangling from that core structure, showing a high degree of flexibility. The majority of the surface of the molecule is made up by the sugar-phosphate backbone of the RNA, since almost all single-stranded bases are either pointing inwards from the loop structures or are buried at the interface of the two helices. Bases protruding freely into the solvent are exceptions, such as A₁₉, which was strongly modified in the chemical probing experiments. The insert in Figure 2 depicts the exact geometry and hydrogen bonding pattern of the triple A interaction.

Comparison of the base accessibilities in the model with experimental probing data

As outlined above, construction of the model relied in part on experimental data of base accessibilities derived from chemical

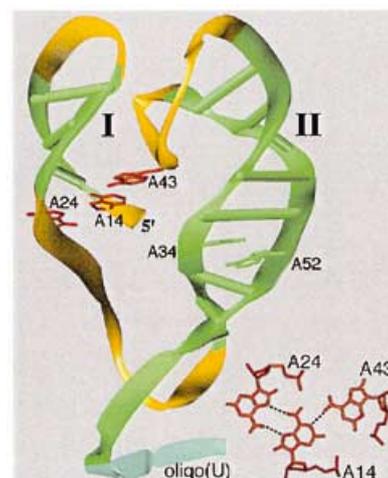


Figure 2. Ribbon model of the helix arrangement of the gND7-506 gRNA. Stems I and II are in green, loop regions are in orange and the oligo(U) tail is in blue. The bases A₁₄, A₂₄ and A₄₃ are annotated in red. Bases A₃₄ and A₅₂, which form a bulge within helix II, are in green. Stems I and II are parallel packed facing their major grooves. The 7 nt loop on top of stem II is folded in between the helical grooves, forming a triple A tertiary interaction with the bases of stem I. The geometry of the triple A interaction of bases A₁₄/A₂₄/A₄₃ is shown as an insert (bottom, right). The drawing was made on a Silicon Graphics Indy workstation using SYBYL 6.1 (Tripos, St Louis).

probing experiments with kethoxal, DEPC, CMCT and DMS. To allow for a quantitative comparison of the model with the probing results, we calculated theoretical accessibilities for relevant base atoms using the GEPOL algorithm (29) and the atomic radii of Rose *et al.* (30). The values were normalized with accessibilities of corresponding atoms in isolated, unpaired model nucleotides and finally scaled to match the 0–4 scale that was used for classification of the experimental data (13). Figure 3 depicts a graphical representation of these results, showing the differences in base accessibilities (model data minus experimental data) for the four chemical probes. From the 25 nt that had been identified to be reactive to the different probes, the model predicted the accessibility of 20 base positions correctly (80%). For 3 nt (12%) the model was in line with the modification for one probe but deviated for a second modification reagent. Only for two nucleotides (U₁₅ and G₃₅) did the calculated accessibilities notably underestimate the experimental values (8%).

Site-directed mutagenesis

To collect experimental evidence for the predicted triple A interaction of A₁₄, A₂₄ and A₄₃ within the core of the gND7-506 structure, two single site mutant gRNAs were constructed. A base change at position 43 from A to C (mutant A₄₃C) was predicted to disrupt the interaction (Fig. 4A), whereas an A→U transversion at position 14 (mutant A₁₄U) was thought to potentially stabilize the tertiary interaction (Fig. 4B), similar to the A₉/A₂₃/U₁₂ triple in *E. coli* tRNA^{Phe} (29).

The two mutant gRNAs were transcribed *in vitro* from DNA templates which had been verified for the presence of the desired base changes. As an initial test of their structures we examined the UV melting profiles of the two mutant RNAs. Wild-type gND7-506 RNA, as published previously (13), is characterized by a main melting transition with a T_m of 39°C. This transition

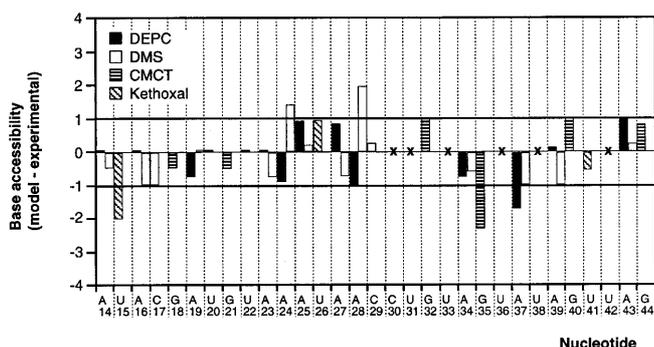


Figure 3. Quantitative comparison of the base accessibilities in the model with the experimental probing data (13). The differences of base accessibilities towards DMS, DEPC, kethoxal and CMCT (model minus experimental) are shown using a relative reactivity scale ranging from 0 (not accessible) to 4 (very accessible). Adenosines are listed as two values, accounting for modification of the N7 position by DEPC or the N1 position by DMS. The values for guanosines were averaged to account for modification of both the N1 and N2 nitrogens by kethoxal. Hydrolysis positions in untreated control samples are annotated with an x and were not considered in the calculation. Considering an error margin of ± 1 , the model is in line with 80% of the experimental data.

was interpreted to reflect primarily the melting of stem/loop II with the melting of stem/loop I superimposed on it. In addition, a small low temperature transition at $\sim 20^\circ\text{C}$ and a minor high temperature transition at 70°C have been described. They were assigned to either temperature-induced changes of the secondary structure or unidentified higher order transitions. Interestingly, both mutant gRNAs exhibited melting profiles indistinguishable from wild-type gND7-506 RNA (data not shown). This was not too surprising, since a disruption of the predicted H bonding pattern of the triple A interaction, which is certainly of low thermodynamic stability (see insert in Fig. 2), need not necessarily create a distinct change in the melting profile. Therefore, we decided to rely on chemical modification experiments with base-specific reagents (DMS and DEPC) to test the structures of the two gND7-506 mutant RNAs. Figures 4C and D summarize the results of the chemical probing experiments. For the $A_{43}C$

mutant RNA an enhanced accessibility predominantly at nucleotide positions at the base of stem I was found (Fig. 4C). In contrast, the $A_{14}U$ mutant RNA showed a small but significant reduction in the accessibilities of bases in that region (Fig. 4D). These data can be interpreted in the first instance as a disruption of the triple A interaction as a consequence of the $A_{43}C$ mutation and in the second case as an enhanced core interaction by the formation of a $U_{14}/A_{24}/A_{43}$ triple.

gRNA-protein interaction

Finally, support for the model was gained from footprinting experiments of gND7-506 in a complex with gBP21, a gRNA binding protein from *T. brucei* (15). The arginine-rich polypeptide has been shown to bind to gRNA molecules with an equilibrium dissociation constant (K_d) in the nanomolar range. Protein motifs known to mediate RNA binding (32) are absent in gBP21. Based on the sensitivity of the gRNA/gBP21 complexes to elevated monovalent cation concentrations, the current model for the interaction of the two molecules involves, at least in part, ionic contacts. The association with gBP21 stabilizes the gRNA structure as judged from hyperchromicity experiments of the RNP complex in comparison with naked gRNA (15).

The footprinting experiments were performed with the single strand-specific RNases T1 and T2 and the double strand-specific cobra venom nuclease V1. Protected nucleotides were found exclusively within and at the base of stem/loop II (Fig. 5A and B). Nucleotide A_{37} , loop nucleotide G_{44} and the sequence stretch G_{53} – A_{56} at the 3'-end of helix II were strongly protected. Several other positions were shielded to a weaker or only to a very weak extent (see colour code in Fig. 6A). One nucleotide position (U_{42}) exhibited an enhanced reactivity in the RNP complex when compared with naked gRNA. U_{42} is located in the apical loop of hairpin II. Its sugar-phosphate residue is exposed on the surface of the model and thus it is very accessible, even in free gRNA. A graphical representation of these data on the three-dimensional model revealed a clustering of protected nucleotides on the solvent-accessible side of helix II (Fig. 6B) across its minor groove, which is indicative of a defined gRNA binding site for gBP21.

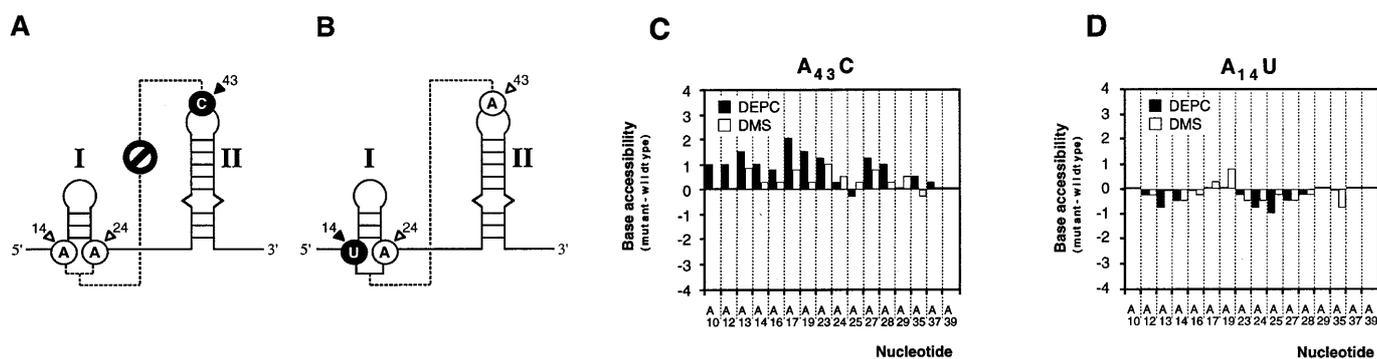


Figure 4. Evaluation of single site mutant gND7-506 gRNAs to test the presence of the triple A tertiary interaction within the core of the gRNA structure (dotted line connecting bases A_{14} , A_{24} and A_{43}). (A) Secondary structure of gND7-506 emphasizing the $A_{43} \rightarrow C$ ($A_{43}C$) base change, which was predicted to disrupt the tertiary interaction. (B) Mutant $A_{14} \rightarrow U$ ($A_{14}U$) was anticipated to stabilize stem I by the formation of a U_{14}/A_{24} base pair and not interfere with the tertiary interaction. (C and D) Results of the chemical probing experiments for both gND7-506 mutant RNAs. The two graphs list the accessibility differences (mutant minus wild-type) of bases 10–39 for DMS and DEPC. (C) mutant $A_{43}C$; (D) mutant $A_{14}U$. The data are derived from densitometer scans of non-saturated autoradiograms scaling the accessibility values from 0 to 4 (see legend to Fig. 3). Positive values correspond to an increase in the accessibility and negative values convey a shielding effect when compared with the wild-type.

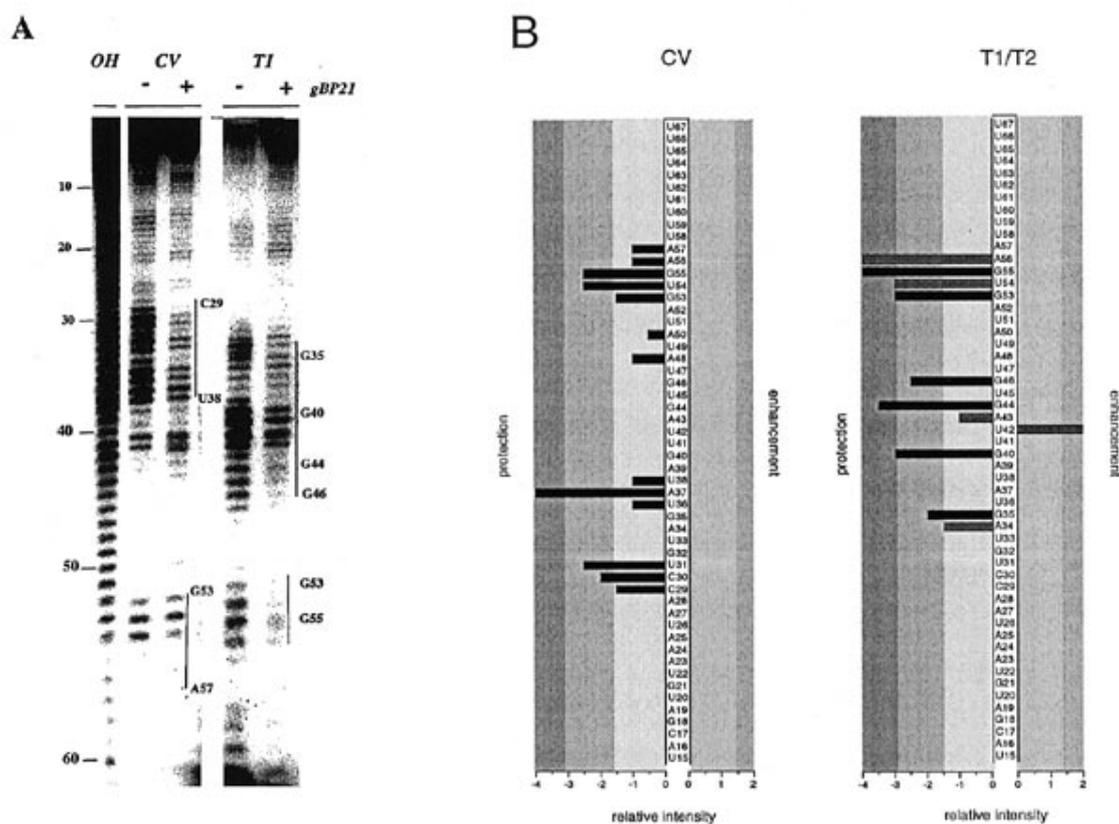


Figure 5. Determination of the binding site of *T. brucei* gBP21 protein on gND7-506 RNA. (A) Representative autoradiogram of a nuclease footprint experiment with ^{32}P 3'-end-labelled gND7-506 (0.6 μM) digested with cobra venom nuclease V1 (CV, 25 mU/ μl) and RNase T1 (T1, 125 mU/ μl) in the presence (+) and absence (-) of recombinant gBP21 (6 μM). Protected areas are marked with a vertical bar. OH refers to an alkaline hydrolysis ladder. (B) Bar graph summarizing the relative accessibility differences (+/- r-gBP21 protein) of nucleotide positions for cobra venom nuclease (CV, left panel), RNase T1 (right panel, black bars) and RNase T2 (right panel, grey bars) derived from three independent experiments. Nucleotide positions without a bar remained unchanged in their accessibility to the three enzymes when in a complex with gBP21. Bars pointing to the left are equivalent to protection of the nucleotide position in the presence of the protein, whereas bars pointing to the right represent enhanced accessibility. The data stem from densitometer scans of non-saturated autoradiographs using a normalized scale of 2 to -4 (0 no, 1 weak, 2 medium, 3 strong and 4 very strong difference in the nucleotide accessibility; + and - designate enhancement and protection respectively).

CONCLUSION

The computer-assisted modelling of RNA structures has proven a valuable tool in providing an understanding of the relationship between structure and function of a variety of RNAs (33-35). Although gRNAs have been identified as key components in the mitochondrial RNA editing reaction, a molecular understanding of their precise function is still lacking. This deficit, at least in part, is a consequence of the scarcity of structural data for these RNAs.

In the present study we have used chemical and enzymatic probing data to model the three-dimensional architecture of gND7-506, an ND7-specific gRNA from *T. brucei*. The main features and conclusions from the model are as follows. The folding of gND7-506 presents a remarkably compact structure. Two hairpins, which are the basic secondary structure elements of gND7-506, are closely packed together and likely connected by a tertiary interaction of three adenosine residues. The triple A interaction is formed between two adenosines at the base of stem I and a third A residue which is part of the apical loop of hairpin II. As a consequence, the two helices are lined up in an almost parallel fashion, facing each other with their major grooves.

The model is supported by a very good fit with the experimental probing data and was further strengthened by analysis of single site mutant gND7-506 RNAs, created by *in vitro* mutagenesis techniques. A base transversion, introduced to disrupt the triple A tertiary interaction, was confirmed to lead to a more open gRNA structure, with an increased accessibility of bases at the interface of the two helices. Conversely, a mutation with the potential to strengthen the tertiary interaction indeed resulted in a gRNA that showed characteristics of a more compact folding.

Lastly, footprinting experiments on gND7-506 in a stable complex with the gRNA binding protein gBP21 were used to support the model. Binding of the polypeptide resulted in the protection of a well-defined set of nucleotide positions, identifying a substantial part of stem/loop II as the interaction site for the protein. Very likely, this binding domain can be defined more precisely if the footprinting experiments are performed with chemical modification reagents instead of the rather bulky enzymatic probes. Interestingly, the gRNA structure remained largely unchanged when in a complex with gBP21. Only one nucleotide (U₄₂) became more accessible to RNase T2. This result is in line with CD measurements of gND7-506 in the presence and absence of gBP21 (15). The data indicated no gross

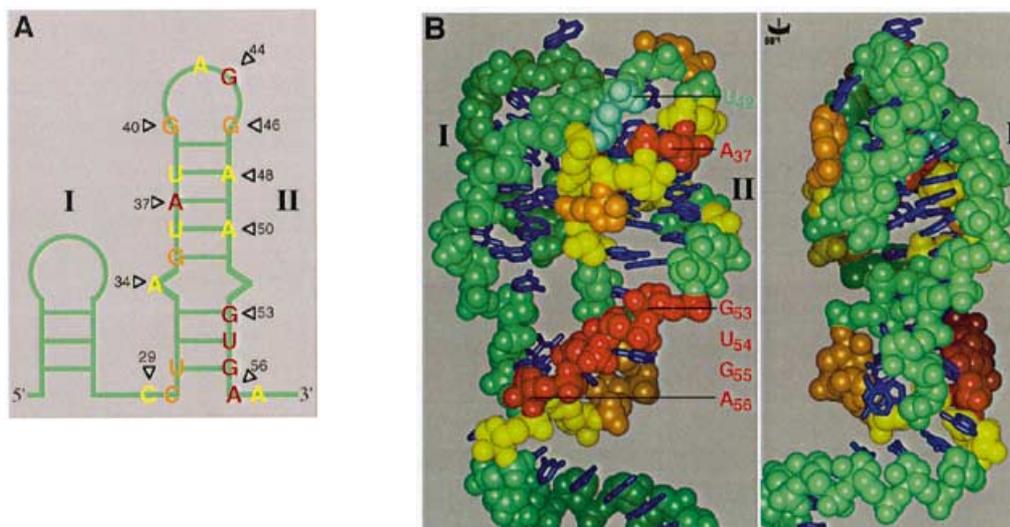


Figure 6. (A) Secondary structure and (B) space filling model of gND7-506 annotating protected backbone positions upon gBP21 binding in red > orange > yellow. Positions with no change in the accessibility pattern are shown in green, bases are given in blue. The sugar-phosphate backbone at nucleotide U₄₂, which exhibited enhanced reactivity in the RNP complex, is coloured light blue. The two orientations of the model differ by a 90° rotation around the vertical axis. The footprinting pattern suggests binding of gBP21 protein to gND7-506 RNA at one side of stem II.

structural rearrangement of the RNA molecule upon protein binding. Since the functional groups of the bases in minor grooves are rather inaccessible and do not allow good discrimination of different base pairs (36), we feel that it is unlikely that the interaction relies on some form of base specificity. The indiscriminate binding behaviour of gBP21 to different gRNA molecules is further support for this assumption (15). As previously suggested (37), it is more likely that the RNA molecule adopts a three-dimensional folding that results in presentation of defined patches of negative charges on the surface of the molecule which correspond to a complementary array of positive charges within the binding domain of the protein. This hypothesis rationalizes experimental data for the cation sensitivity of the gRNA/gBP21 interaction (15). Whether ionic interaction is the general principle for the binding between gRNAs and proteins and whether it also applies for the recruitment of gRNAs into the editing machinery cannot be assessed at the moment. However, the phenomenon might in part be responsible for the observed salt sensitivity of the RNA editing activity *in vitro* (38).

The obvious question that arises from the presented model for gND7-506 is whether all gRNAs can be folded into a similar three-dimensional architecture. For only three additional gRNA molecules has secondary structure information been collected. These RNAs also consist of two imperfect stem/loop elements (13). Thus, similar foldings seem to be conceivable, although they clearly cannot rely on the same triple A tertiary interaction, simply because their primary sequences differ in the relevant sequence domains. However, many other forms of tertiary interactions can easily be envisaged (28). Even for gND7-506, we predict that other interactions contribute to the overall stability of the molecule. Due to the low resolution of the model, only the triple A interaction could be resolved, but additional contacts are required to mediate a stable interaction of the two hairpin elements. Cross-linking studies of gND7-506 as well as of other gRNAs must be performed to ultimately solve this question.

Finally, we would like to emphasize that we view the model only as an initial step towards a molecular understanding of the structure of gRNAs. Hopefully it will stimulate further work on the structure/function correlation of gRNAs to ultimately gain a molecular understanding of the role of gRNAs in the RNA editing process.

ACKNOWLEDGEMENTS

We wish to thank A.Souza, U.Müller and A.S.Paul for critical reading of the manuscript. J.Köller is thanked for r-gBP21 protein preparations and E.Weyher-Stingl and L.Moroder for their help performing the CD measurements. Work in the H.U.G. laboratory is supported by the German Ministry for Education and Research (BMBF) and the German Research Foundation (DFG). T.H. acknowledges support from the Studienstiftung des deutschen Volkes and H.H. from the DFG.

REFERENCES

- Blum,B., Bakalara,N. and Simpson,L. (1990) *Cell*, **60**, 189–198.
- Hajduk,S.L. and Sabatini,R.S. (1996) In Smith,D.F. and Parsons,M. (eds), *Molecular Biology of Parasitic Protozoa*. IRL Press, Oxford, UK pp.134–158.
- Pollard,V.W., Harris,M.E. and Hajduk,S. (1992) *EMBO J.*, **11**, 4429–4438.
- Corell,R.A., Read,L.K., Riley,G.R., Nellissery,J.K., Allen,T.E., Kable,M.L., Wachal,M.D., Seiwert,S.D., Myler,P.J. and Stuart,K.D. (1996) *Mol. Cell. Biol.*, **16**, 1410–1418.
- Seiwert,D.S., Heidmann,S. and Stuart,K. (1996) *Cell*, **84**, 831–841.
- Piller,K.J., Rusché,L.N., Cruz-Reyes,J. and Sollner-Webb,B. (1997) *RNA*, **3**, 279–290.
- Maslov,D.A. and Simpson,L. (1992) *Cell*, **70**, 459–467.
- Koslowsky,D.J., Riley,G.R., Feagin,J.E. and Stuart,K. (1992) *Mol. Cell. Biol.*, **12**, 2043–2049.
- Göringer,H.U., Köller,J. and Shu,H.H. (1995) *Parasitol. Today*, **11**, 265–267.
- Köller,J., Nörskau,G., Paul,A.S., Stuart,K. and Göringer,H.U. (1994) *Nucleic Acids Res.*, **22**, 1988–1995.

- 11 Leegwater,P., Speijer,D. and Benne,R. (1995) *Eur. J. Biochem.*, **227**, 780–786.
- 12 Read,L.K., Göringer,H.U. and Stuart,K. (1994) *Mol. Cell. Biol.*, **14**, 2629–2639.
- 13 Schmid,B., Riley,G.R., Stuart,K. and Göringer,H.U. (1995) *Nucleic Acids Res.*, **23**, 3093–3102.
- 14 Souza,A.E., Hermann,T. and Göringer,H.U. (1997) *Nucleic Acids Res.*, **25**, 104–106.
- 15 Köller,J., Müller,U., Schmid,B., Missel,A., Kruft,V., Stuart,K. and Göringer,H.U. (1997) *J. Biol. Chem.*, **272**, 3749–3757.
- 16 Arnott,S., Hukins,D.W.L. and Dover,S.D. (1972) *Biochem. Biophys. Res. Commun.*, **48**, 1392–1399.
- 17 Weiner,S.J., Kollman,P.A., Case,D.A., Singh,U.C., Ghio,C., Alagona,G., Profeta,S. and Weiner,P. (1984) *J. Am. Chem. Soc.*, **106**, 765–784.
- 18 Puglisi,J. and Tinoco,I. (1989) *Methods Enzymol.*, **180**, 304–325.
- 19 Reyes,V.M. and Abelson,J.N. (1989) *Methods Enzymol.*, **180**, 63–69.
- 20 Christiansen,J., Egebjerg,J., Larsen,N. and Garrett,R.A. (1990) *Ribosomes and Protein Synthesis. A Practical Approach*. IRL Press, Oxford, UK, pp. 229–252.
- 21 Witherell,G.W. and Uhlenbeck,O.C. (1989) *Biochemistry*, **28**, 71–76.
- 22 Westhof,E., Dumas,P. and Moras,D. (1988) *Acta Crystallogr.*, **A44**, 112–123.
- 23 Szweczak,A.A. and Moore,P.B. (1995) *J. Mol. Biol.*, **247**, 81–98.
- 24 Dock-Bregeon,A.C., Bregeon,B., Chevrier,B., Podjarny,A. and Moras,D. (1988) *Nature*, **335**, 375–378.
- 25 White,S.A., Nilges,M., Huang,A., Brünger,A.T. and Moore,P.B. (1992) *Biochemistry*, **31**, 1610–1621.
- 26 Shen,L.X. and Tinoco,I. (1995) *J. Mol. Biol.*, **247**, 963–978.
- 27 Gray,D.M., Liu,J.-J., Ratliff,R.L. and Allen,F.S. (1981) *Biopolymers*, **20**, 1337–1382.
- 28 Gautheret,D., Damberger,S.H. and Gutell,R.R. (1995) *J. Mol. Biol.*, **248**, 27–43.
- 29 Silla,E. and Villar,F. (1990) *J. Mol. Graphics*, **8**, 168–172.
- 30 Rose,G.D., Geselowitz,A.R., Lesser,G.J., Lee,R.H. and Zehfus,M.H. (1985) *Science*, **229**, 834–838.
- 31 Quigley,G.J. and Rich,A. (1976) *Science*, **194**, 796–806.
- 32 Burd,C.G. and Dreyfuss,G. (1994) *Science*, **265**, 615–620.
- 33 Westhof,E., Romby,P., Ehresmann,C. and Ehresmann,B. (1990) In Beveridge,D.L. and Lavery,R. (eds), *Theoretical Biochemistry and Molecular Biophysics*. Adenine Press, New York, NY, pp. 399–409.
- 34 Gautheret,D. and Cedergren,R. (1993) *FASEB J.*, **7**, 97–105.
- 35 Westhof,E. (1993) *J. Mol. Struct. Theochem.*, **286**, 203–211.
- 36 Weeks,K.M. and Crothers,D.M. (1993) *Science*, **261**, 1574–1577.
- 37 Göringer,H.U., Koslowsky,D.J., Morales,T.H. and Stuart,K. (1994) *Proc. Natl. Acad. Sci. USA*, **91**, 1776–1780.
- 38 Seiwert,D.S. and Stuart,K. (1994) *Science*, **266**, 114–117.