Proposed Secondary Structure of Eukaryote Specific Expansion Segment 15 in 28S rRNA from Mice, Rats, and Rabbits†

Sofia L. Larsson and Odd Nygård*

Natural Science Section, Södertörn University College, S-141 04 Huddinge, Sweden, and Department of Zoological Cell Biology, Arrhenius Laboratories E5, Stockholm University, S-106 91 Stockholm, Sweden

Received September 28, 2000; Revised Manuscript Received November 28, 2000

ABSTRACT: The expansion segments in eukaryotic ribosomal RNAs are additional RNA sequences not found in the RNA core common to both prokaryotes and eukaryotes. These regions show large species-dependent variations in sequence and size. This makes it difficult to create secondary structure models for the expansion segments exclusively based on phylogenetic sequence comparison. Here we have used a combination of experimental data and computational methods to generate secondary structure models for expansion segment 15 in 28S rRNA in mice, rats, and rabbits. The experimental data were collected using the structure sensitive reagents DMS, CMCT, kethoxal, micrococcal nuclease, RNase T_1, RNase CL3, RNase V_1, and lead(II) acetate. ES15 was folded with the computer program RNAStructure 3.5 using modification data and phylogenetic similarities between different ES15 sequences. This program uses energy minimization to find the most stable secondary structure of an RNA sequence. The presented secondary structure models include several common structural motifs, but they also have characteristics unique to each organism. Overall, the secondary structure models showed indications of an energetically stable but dynamic structure, easily accessible from the solution by the modification reagents, suggesting that the expansion segment is located on the ribosomal surface.

Prokaryotic and eukaryotic ribosomes share a common structural RNA core. Ribosomal functions such as peptidyl transferase activity and interaction sites for translation factors and tRNA have been mapped to this phylogenetically conserved structure (*1*−*3*). Mammalian rRNAs are longer than their prokaryotic counterparts. Blocks of nucleotides inserted in the common core account for the main increase in size. The inserts vary greatly in length and nucleotide composition between species, yet they always occur at the same sites in the conserved structure. The inserted sequences are usually termed variable or V regions (*4*, *5*), divergent or D regions (*6*, *7*), or, as used here, expansion segments (ES)[1] (*1*, *8*). The expansion segments are in many cases GC rich and almost always lack the high degree of methylation present in the universal core (*9*). There is also a high frequency of repetitive sequences in many expansion segments, which might have arisen from slippage during replication (*10*), or through unequal crossing over (*4*, *11*).

The question of whether expansion segments serve a function has been widely debated, and there is evidence both for and against any functional role. It seems improbable that the expansion segments persist in eukaryotes without having some role in ribosome assembly or in translation. However, it has been suggested that they simply are remnants of transcribed spacers that are tolerated because they do not disturb the other ribosomal components (*12*). Experimental approaches to investigating the role of the expansion segments have shown that ES19 in domain III of yeast 26S rRNA may be removed, or replaced with the slightly longer but equivalent sequence from mouse, without disrupting ribosomal assembly or polysome formation (*5*). On the other hand, insertion of a 19-nucleotide sequence into ES3 of the small subunit rRNA of yeast interferes with the assembly of 40S subunits (*13*). Similarly, deletion or replacement of ES27 in *Thermus thermophilus* 26S rRNA with an unrelated sequence of similar length is lethal (*7*). However, the expansion segment can be replaced with the homologous sequence from *Saccharomyces cervisiae*, *Dictyostelium discoideum*, or *Caenorhabditis elegans*. Jeeninga et al. (*14*) suggested that part of the equivalent expansion segment in yeast is involved in 60S subunit assembly based on deletion studies. These experiments show that certain expansion segments are essential even though their exact function remains unknown.

The species-dependent variability in length and sequence of the expansion segments makes comparative sequence analysis an insufficient approach for prediction of the secondary structure. The high GC/AT ratio found in most of the expansion segments and the high percentage of repetitive sequences also cause problems when using secondary structure modeling by energy minimization, as these sequence elements allow several secondary structure models with nearly equal free energy. Thus, additional experimental data are needed to identify the most probable secondary structure models.

† This work was supported by Swedish Natural Research Council Grant B 650-1981117/2000.

* To whom correspondence should be addressed: Natural Science Section, Södertörn University College, Box 4101, S-141 04 Huddinge, Sweden. Phone: +46 8 585 88701. Fax: +46 8 585 88510. E-mail: odd.nygard@sh.se.

[1] Abbreviations: ES, expansion segment; CMCT, 1-cyclohexyl-3-(2-morpholinoethyl)carbodiimide metho-*p*-toluenesulfate; DMS, dimethyl sulfate; DOC, sodium deoxycholate; EDTA, ethylenediaminetetraacetic acid; EGTA, ethylene glycol bis(2-aminoethyl ether)-*N*,*N*,*N*′,*N*′-tetraacetic acid; RNase, ribonuclease; RT-PCR, reverse transcription polymerase chain reaction; SDS, sodium dodecyl sulfate.

10.1021/bi002286q CCC: $20.00
Published on Web 02/15/2001

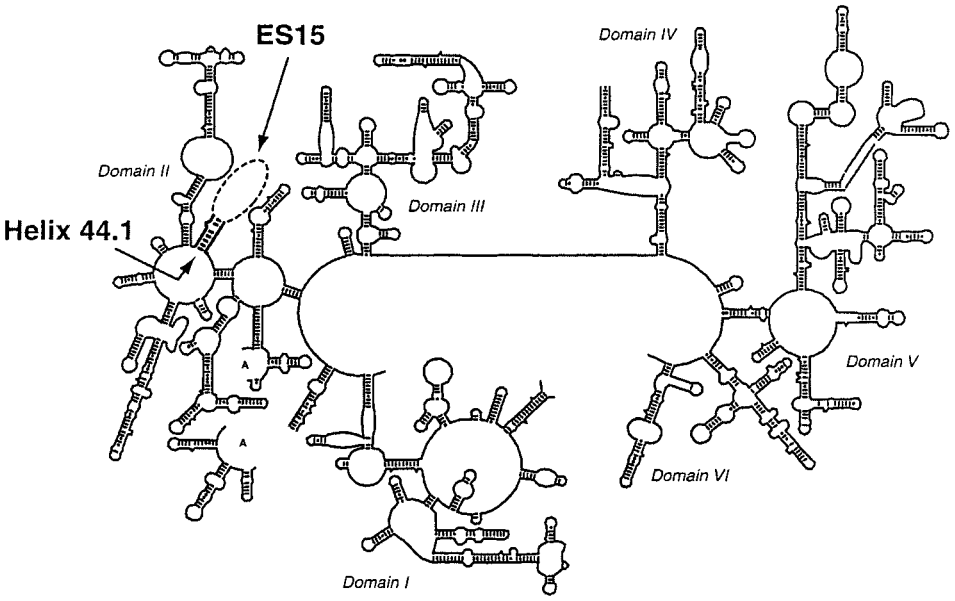

FIGURE 1: Line model of the mouse 28S rRNA according to Gutell et al. (*31*). The expansion segment studied here, ES15, is located in domain II of mouse 28S rRNA, and protrudes from the equivalent helices in rat and rabbit rRNA. Helix 44.1, which is part of the conserved ribosomal core, is included in helix I of our structure models.

Here we have analyzed the secondary structure of expansion segment 15 (ES15) using a wide range of structure specific RNA modifying chemical reagents and enzymes. ES15 protrudes from helix 44.1 in domain II of 28S rRNA (Figure 1) and has so far only been found in mammalian 28S rRNA (*4*, *15*−*17*). The experimental data and available phylogenetic information were used in predicting secondary structure models for mouse, rat, and rabbit ES15 using the energy minimization-based secondary structure program RNAStructure 3.5.

MATERIALS AND METHODS

Chemicals. Dimethyl sulfate (DMS) and 1-cyclohexyl-3-(2-morpholinoethyl)carbodiimide metho-*p*-toluenesulfate (CMCT) were from Aldrich-Sigma. 2-Keto-3-ethoxybutyraldehyde (kethoxal) was from ICN Pharmaceuticals. Lead(II) acetate was from Merck. The PCR kit, micrococcal nuclease from *Staphylococcus aureus*, ribonuclease CL3, and deoxynucleotides were from Boehringer Mannheim. Ribonucleases V_1 and T_1 were from Amersham-Pharmacia Biotech. Superscript reverse transcriptase was from Life Technologies, Inc. The DNA sequencing kit was from Perkin-Elmer, and the fluorescent cDNA probe was from Oswell. Unlabeled cDNA probes were from Scandinavian Gene Synthesis. The rRNA sequences used for primer annealing during PCR, sequencing, and primer extension were G1835−U1849 and C2025−A2039 (mouse numbering).

Buffers. The ribosome buffer consisted of 30 mM Hepes-KOH (pH 7.6), 70 mM KCl, 0.25 M sucrose, 2 mM $MgCl_2$, and 5 mM mercaptoethanol. The hybridization buffer consisted of 50 mM Hepes-KOH (pH 7.6) and 100 mM KCl.

Preparation of Mouse and Rat Ribosomes. Mouse (NMRI strain) and rat (Sprague-Dawley strain) liver homogenates were prepared as described by Morley and Jackson (*18*). Runoff ribosomes were prepared by incubating the homogenates in the presence of 0.1 mg/mL creatine phosphokinase, each of the 20 amino acids (each at 125 μM), 0.7 mM $Mg(CH_3COO)_2$, 85 mM KCl, 25 mM ATP, 50 μM GTP, and 30 mM creatine phosphate (*18*). The mixtures were incubated for 10 min at 30 °C. The samples were treated with DOC and Triton X-100 to final concentrations of 1.9 and 0.9% (v/v), respectively, and centrifuged to remove larger aggregates. The supernatants were treated with NH_4Cl and $Mg(CH_3COO)_2$ at final concentrations of 0.5 M and 5 mM, respectively, to remove ribosome-bound translational factors. The samples were layered onto a 15 to 28% (w/v) continuous sucrose gradient in 20 mM Tris-HCl (pH 7.6), 5 mM $MgCl_2$, 5 mM 2-mercaptoethanol, and 70 mM KCl. The runoff ribosomes were pelleted by centrifugation for 6 h at $200000g_{av}$ at 5 °C and dissolved in ribosome buffer.

Preparation of Rabbit Ribosomes. Rabbit (New Zealand red strain) reticulocyte polysomes were prepared according to the method of Nygård and Nilsson (*19*). Rabbit reticulocyte lysates were layered onto discontinuous sucrose gradients consisting of 1 mL of 35% (w/v) sucrose in 20 mM Tris-HCl (pH 7.6), 25 mM KCl, and 2 mM $Mg(CH_3COO)_2$ superimposed on 1 mL of 60% (w/v) sucrose in the same buffer. The gradients were centrifuged for 170 min at $257000g_{av}$ at 5 °C, and the pelleted ribosomes were dissolved in ribosome buffer.

Chemical Modification of rRNA in Intact Ribosomes. Ribosomes (30 pmol) in final volumes of 50 μL were modified as follows. DMS modification was carried out for 5 min at 37 °C in the presence of 20 and 90 mM DMS (*20*). CMCT modification was carried out for 15 min at 37 °C in the presence of 20 and 100 mM CMCT (*20*). Kethoxal modification was carried out essentially according to the method of Stern et al. (*21*), except that the kethoxal concentration of the stock solution was 34 mg/mL in 99% (v/v) ethanol. Samples were incubated for 60 min at 4 °C in the presence of 9 and 23 mM kethoxal. Cleavage with lead(II) acetate was carried out essentially according to the method of Ciesiolka et al. (*22*). The samples were incubated for 10 min at 35 °C in the presence of 1.4 and 2.1 mM $Pb(CH_3COO)_2$. The reaction was stopped by addition of EDTA. Control samples were identically treated but without the modifying reagent.

Table 1: Chemical and Enzymatic Reagents Used in This Study

reagent	molecular mass (Da)	specificity[a]	position of modification or cleavage	ref
DMS	126	ss A and ss C	N1-A, N3-C (N7-G[b])	*32*
CMCT	424	ss G and ss U	N3-U, N1-G	*32*
kethoxal	148	ss G	N1-G, N2-G	*32*
micrococcal nuclease	17000	ss N	p↓N	*59*
ribonuclease T_1	11000	ss G	Gp↓	*60*
ribonuclease CL3	16800	ss C > ss A, ss G	Cp↓	*61*
ribonuclease V_1	15900	ds N or stacked N	N↓p	*45, 62*
lead(II) acetate	261	ss or dynamic ds regions	Np↓	*22, 37*

[a] ss, single-stranded; ds, double-stranded; N, any base. [b] Cannot be detected by primer extension.

The ribosomes were precipitated with 2.5 volumes of ethanol and 0.1 volume of 3 M $NaCH_3COO$ (pH 5.2) and pelleted by centrifugation. The pellets were dissolved in 0.1 M Tris-HCl (pH 7.6) and 0.5% (w/v) SDS, and the RNA was extracted with phenol as described below.

Enzymatic Modification of rRNA in Intact Ribosomes. Ribosomes (30 pmol) in final volumes of 50 μL were incubated with modifying enzymes as follows. Micrococcal nuclease cleavage was performed by incubating the ribosomes for 15 min at room temperature in the presence of 1 mM $CaCl_2$ and 3 or 12 units of micrococcal nuclease (*20*). The reaction was terminated by addition of EGTA to a final concentration of 2 mM, and the rRNA was immediately extracted with phenol as described below. Cleavage with RNases was carried out as follows. For RNase T_1, cleavage of the ribosomes was incubated for 15 min at 50 °C with 2 and 10 units of nuclease. RNase CL3 cleavage was carried out for 15 min at 50 °C in the presence of 0.02 and 0.10 unit of nuclease. Cleavage with RNase V_1 was carried out for 5 min at 37 °C with 0.4 and 2.0 units of nuclease. Control samples were identically treated but without the nuclease. The reaction mixtures were diluted with an equal volume of 0.1 M Tris-HCl (pH 7.6) and 1% (w/v) SDS and subjected to phenol extraction as described below.

Mild Deproteination. In a series of experiments, the ribosomes were subjected to a mild deproteination before modification (*23*). Briefly, the ribosomes were incubated for 1 h at 37 °C in the presence of 1 mg/mL proteinase K and 0.5% (w/v) SDS, final concentrations, followed by mild phenol extraction.

Phenol Extraction of RNA. Ribosomal RNA was extracted from the ribosomes with a 1/1 phenol/chloroform mixture saturated with 0.1 M Tris-HCl (pH 7.6). The extracted rRNA was precipitated with 2.5 volumes of ethanol and 0.1 volume of 3 M $NaCH_3COO$ (pH 5.2), dissolved in H_2O, and stored in aliquots at −80 °C.

Primer Extension. Primer extension was as previously described (*20*). Briefly, 0.5 pmol of RNA was hybridized to 0.25 pmol of fluorescent probe in hybridization buffer. The samples were incubated for 3 min at 90 °C and allowed to cool to room temperature for 10 min. All four deoxynucleotides, at final concentrations of 0.5 mM, and 37.5 units of reverse transcriptase were added and the samples incubated for 30 min at 45 °C. The resulting cDNA was analyzed on 4.75% (w/v) acrylamide sequencing gels in an Applied Biosystems 377 DNA sequencer.

Sequencing of ES15 by RT-PCR. rRNA was used in reverse transcription and the resulting cDNA subjected to PCR with probes flanking the expansion segment. The generated DNA fragment was sequenced using an Applied Biosystems 377 DNA sequencer.

Modeling the Secondary Structure. Secondary structure models of the expansion segments were generated using the RNA folding program RNAStructure 3.5 (*24*) based on Mfold (*25*−*27*). The modification and cleavage data were used to control the folding procedure.

RESULTS AND DISCUSSION

Secondary structure models of mammalian 28S rRNA contain a specific extension of conserved helix 44.1 located in domain II of 28S rRNA (Figure 1) (*1*). Here we have combined available phylogenetic information with chemical and enzymatic modification data to construct a secondary structure model for ES15 in mice, rats, and rabbits.

Chemical and Enzymatic Probing of the Secondary Structure. Secondary structure data for ES15 in mice, rats, and rabbits was collected using the structure-sensitive chemical and enzymatic probes listed in Table 1. The structure probing experiments were performed a minimum of three times for each reagent and organism. Each round of modification was then analyzed two to three times by primer extension using an ES15 specific probe. Typical examples of primary modification data are shown in Figures 4−7 (see below for discussion). The experimental data are summarized in Figure 3.

For the collection of secondary structure data, we used intact 80S ribosomes. Consequently, it was essential to exclude interference of ribosomal proteins during structure probing. Therefore, we also performed chemical and enzymatic modification of rRNA that had been prepared from the ribosomal particles with proteinase K and a mild phenol extraction (*23*). The modification data obtained from the deproteinized rRNA were the same as those found in intact ribosomal particles (not illustrated), indicating that the structure analysis was not influenced to any larger extent by ribosomal proteins.

Alignment of Sequences. During the modification experiments, we discovered discrepancies between the published sequences in GenBank and our modification patterns. ES15 from mice, rats, and rabbits was therefore subjected to RT-PCR and direct sequencing. The resulting mouse and rabbit sequences contained heterogeneities, whereas the rat sequence was completely homogeneous. The latter sequence differed slightly from the published sequence (GenBank accession number V01270) (*17*) (Figure 2).

A comparison of the obtained mouse sequence with the previously published genomic sequence (GenBank accession number X00525) (*16*) showed that the population of rRNA

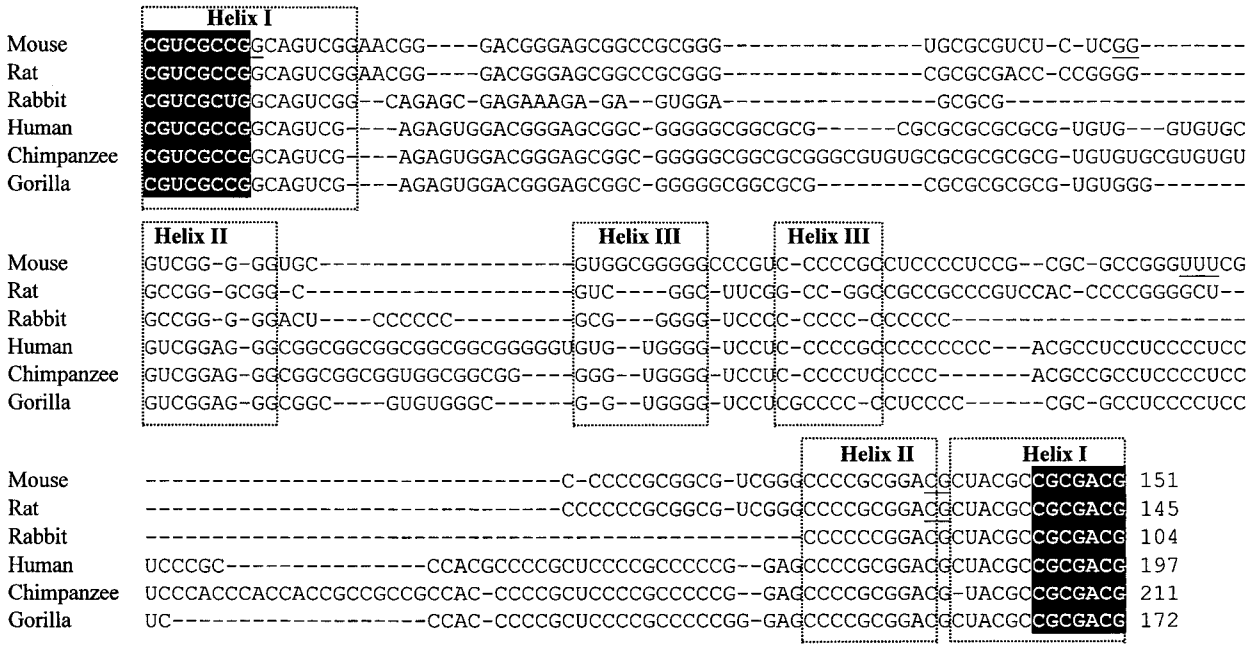

FIGURE 2: Alignment of ES15 in mouse, rat, rabbit, human, chimpanzee, and gorilla 28S rRNA. The alignment was performed in part with the ClustalW alignment program (*28*, *63*) and complemented by a manual search for common motifs (dotted boxes; see the text). Helix 44.1 (shown in Figure 1) is denoted with black boxes. Mouse, rat, and rabbit sequences were determined by RT-PCR and direct sequencing (see the text). The sites of insertions, deletions, and substitutions in mice and rats compared to the published sequences are underlined. Human, chimpanzee, and gorilla 28S rRNA sequences are from Kuo et al. (*15*). Numbers at the end of each sequence indicate the total number of nucleotides for each sequence, including helix 44.1.

isolated from mouse ribosomes was a mixture of mainly two sequences, one being almost identical to the published sequence, and the other differing by a deletion of five nucleotides (Figure 2). The latter was the dominating sequence.

In the case of rabbit ES15, a similar explanation seems plausible. However, in this case we were not able to locate specific positions for deletions and/or insertions. The dominating rabbit ES15 sequence is shown in Figure 2. It should be noted that the base specific structure probing data are in agreement with the mouse, rat, and rabbit sequences listed in Figure 2.

The obtained sequences were aligned with human, chimpanzee, and gorilla ES15 sequences (*15*) using ClustalW (*28*). As seen in Figure 2, mouse, rat, and rabbit ES15 contained a high percentage of repeats of three to five guanines or cytosines. Similarly, primate ES15 also contains highly repetitive sequences, which is particularly evident in humans (*15*).

Because of the highly repetitive nature of the sequences, several different alignments were possible. Therefore, we identified probable helical motifs, common to all ES15 sequences, using the modification data. These common motifs were the extension of the initial helix, helix I in Figure 2, an internal helix (helix II), and a helix−loop motif (helix III).

Prediction of Secondary Structure. The secondary structure of ES15 was modeled using RNAStructure 3.5 (*24*), a PC version of Mfold (*25*−*27*). The RNAStructure software uses energy minimization to construct secondary structure models of a given RNA sequence. In our case, without any restrictions, the software generated a large number of different secondary structures with very similar free energies, due to the repetitive nature of the sequences. To minimize the number of possible structures, we used the modification data, including the previously identified helices I−III as restrictions during folding. A comparison of secondary structure predictions of, among others, 5S rRNA, made with and without structure probing data, shows that accuracy of prediction is improved from 26 to 87% when modification data are included (*24*).

The suggested secondary structures for ES15 in mice, rats, and rabbits are shown in panels A−C of Figure 3, respectively. The calculated free energies of the structures are −38.7, −39.0, and −29.4 kcal/mol for mice, rats, and rabbits, respectively. However, one should note that the RNAStructure software does not include algorithms for estimating the free energy of noncanonical base pairs, such as C·C, G·A, and U·C. These types of base pairs have recently been identified in 23S rRNA as well as in many other types of RNA structures (*29*, *30*), and we suggest that they occur in ES15 (see below). In calculating the free energy of our structure models (Figure 3), we could therefore not include the stability contributed by the noncanonical base pairs.

The three models contain an initial helix, helix I, which includes the conserved helix 44.1 (Figure 1) that belongs to the structural core (*31*). At the end of helix I, there is a short single-stranded branch point, leading into hairpin A and helix II (Figure 3). Helix II branches off into helix III and hairpin B. The only exception to this general structure model is rabbit ES15, which only has one helical arm on the far side of helix II (Figure 3C).

Helix I. The bases in the proposed helix I were essentially protected from single-strand specific modification in ribosomes and in deproteinized RNA (not illustrated). The only exceptions were U13 and G15 in mouse (Figure 3A) and C14 in rat (Figure 3B) that were weakly accessible for modification. In the proposed secondary structure models, U13 and G15 are involved in U-A and G·U base pairs positioned at the termini of short helices (Figure 3A). The

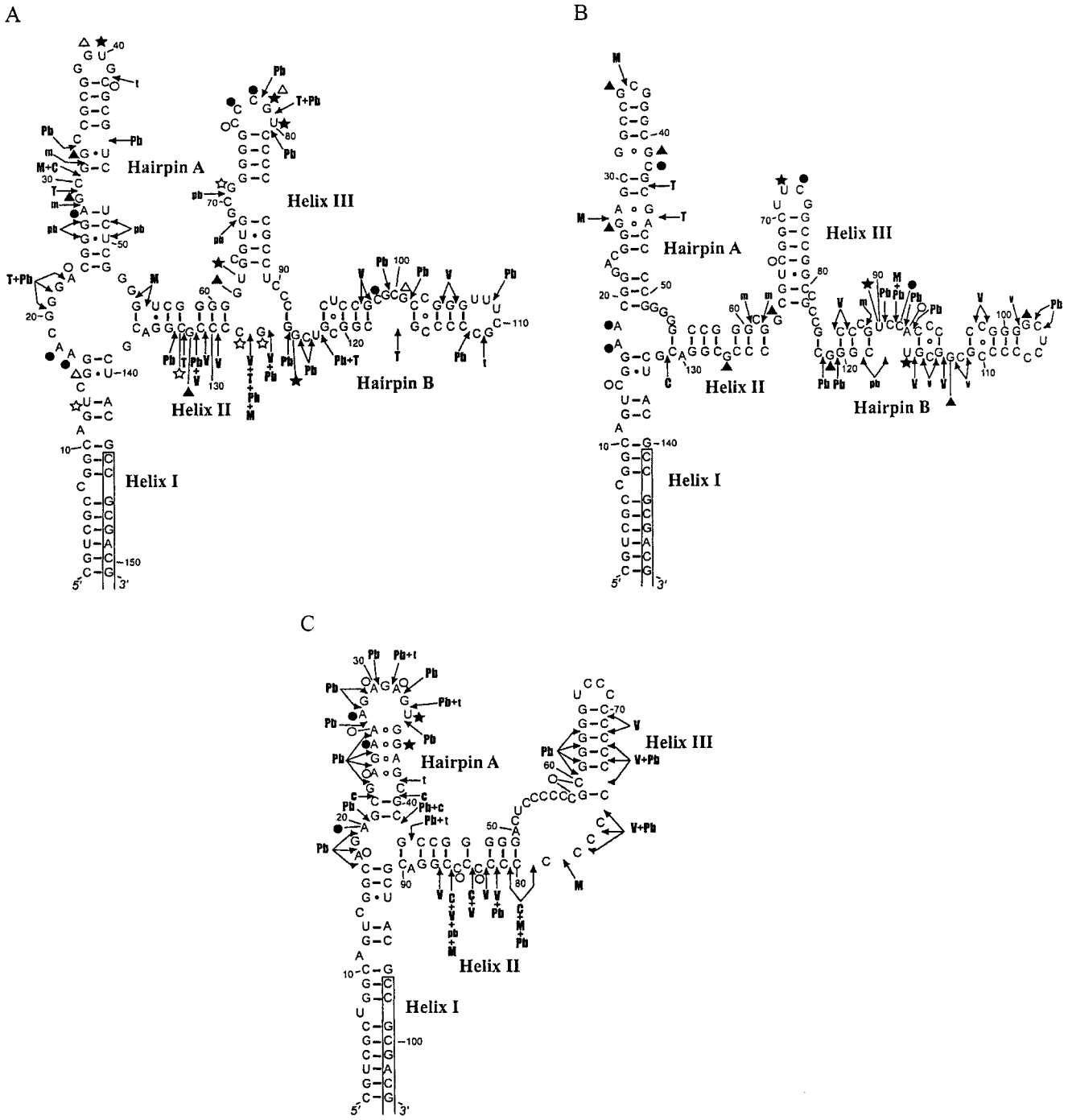

FIGURE 3: Proposed secondary structure of ES15. The first nucleotide from the 5′-end in the initial helix is numbered as position 1. Sites available for modification by DMS (○ and ●), CMCT (☆ and ★), and kethoxal (△ and ▲). Sites available for cleavage by nuclease and Pb(II) are denoted by arrows and an identifying letter: micrococcal nuclease (m and M), RNase CL3 (c and C), RNase T_1 (t and T), RNase V_1 (v and V), and Pb(II) (pb and Pb). Black symbols and capital letters show strong modifications or cuts; white symbols and small letters indicate weaker modifications or cuts. Helices identified in the alignment (Figure 2) are denoted by Roman numerals. The remaining helices are named A and B to distinguish them from the conserved helices. Boxed nucleotides indicate overlap of the hybridization site of the probe used in primer extension. (A) Proposed secondary structure of mouse ES15. The free energy (ΔG) is −38.7 kcal/mol (see the text for discussion). (B) Proposed secondary structure of rat ES15. The free energy (ΔG) is −39.0 kcal/mol. (C) Proposed secondary structure of rabbit ES15. The free energy (ΔG) is −29.4 kcal/mol.

kethoxal reactivity of the N2 position of G15 is compatible with the suggested involvement in a G·U base pair (*32*). However, the N3 position of U13 should not be accessible for CMCT modification when U13 is paired with an adenine. Previous studies of rRNA secondary structures have indicated that bases in base pairs at helix termini have a higher probability of being accessible for single-strand specific chemical modification (*33*, *34*) presumably due to an increased dynamic instability of base pairs next to loops, hinges, and bulged nucleotides (*33*).

The initial seven base pairs in helix I belong to helix 44.1 (*31*). The site for hybridization of the probe used in primer extension (seen boxed in Figure 3) overlaps this conserved helix. This part of the 28S rRNA has previously been found to be inaccessible to single-strand specific chemical modification, supporting its participation in a helix (*34*).

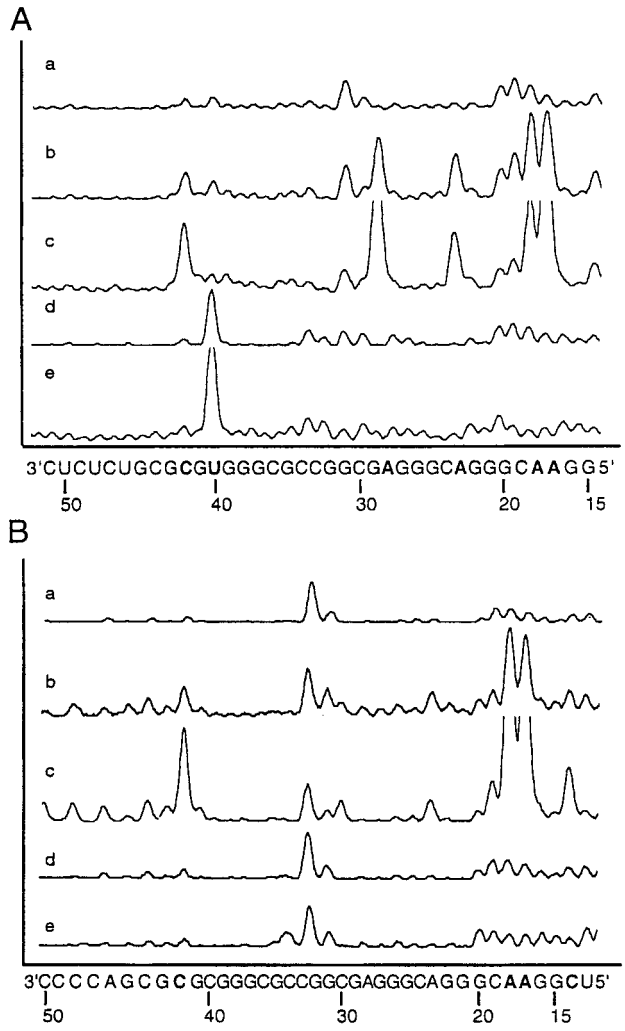

FIGURE 4: Selection of the secondary structure probing data for the nucleotides proposed to form hairpin A in mice and rats: (A) mouse, (B) rat, (a) control sample, (b) 20 mM DMS, (c) 90 mM DMS, (d) 20 mM CMCT, and (e) 100 mM CMCT. The numbering of nucleotides is according to panels A and B of Figure 3. Modified bases are in bold. The sequence has been shifted by one nucleotide to compensate for the fact that primer extension stops one nucleotide before the modification site.

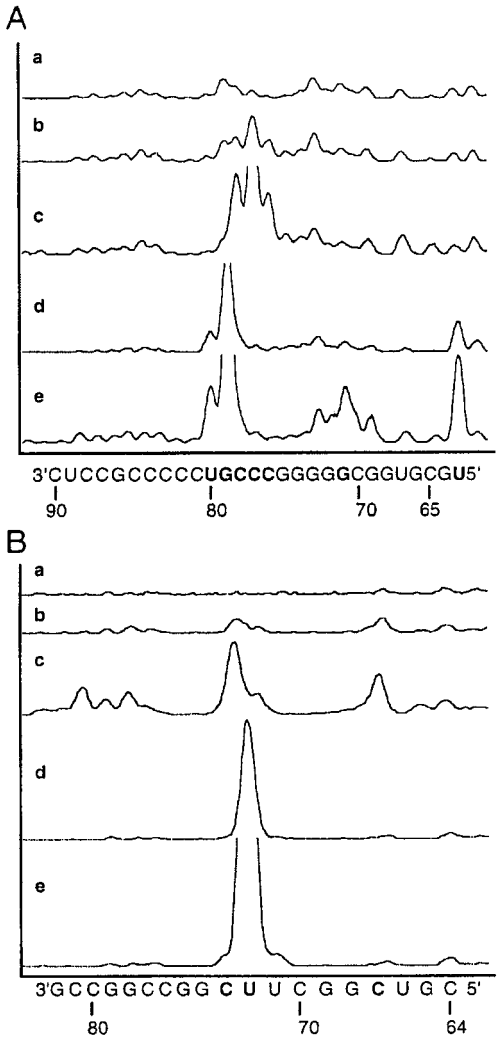

FIGURE 5: Selection of the secondary structure probing data for the nucleotides proposed to form helix III in mice and rats: (A) mouse, (B) rat, (a) control sample, (b) 20 mM DMS, (c) 90 mM DMS, (d) 20 mM CMCT, and (e) 100 mM CMCT. Modified bases are in bold. For the numbering of nucleotides, see the legend of Figure 4.

Helix I was essentially identical in all three species, but as the rabbit sequence lacked the AA motif following helix I in mice and rats, an additional C-G base pair could be formed by pairing C17 with G91 (Figure 3C).

Hairpin A. The sequences proposed to form hairpin A in Figure 3 exhibited considerable sequence variations. In rabbit ES15 (Figure 3C), the sequence contains purines almost exclusively, with seven of the 21 bases being adenines. As rabbit ES15 contains no uracil rich sequence complementary to the bases in the putative loop, the adenines could not form canonical Watson−Crick base pairs with any other region in the expansion segment. In the model of rabbit ES15, hairpin A consists of a six-base pair helix and an apical heptaloop. The loop is accessible for lead(II) cleavage, and four of the bases are also accessible for DMS or CMCT modification. The helical part of the putative hairpin A contains a stack of four A·G/G·A base pairs similar to that observed in 23S rRNA and other crystallized RNAs (*29*, *35*, *36*). Note that the contribution of these base pairs to the stability of the structure was not considered by the folding software (see above). The N1 position of the adenines in the three A·G base pairs was available for modification by DMS. This type of modification is possible if these base pairs are in the sheared conformation. The sheared conformation

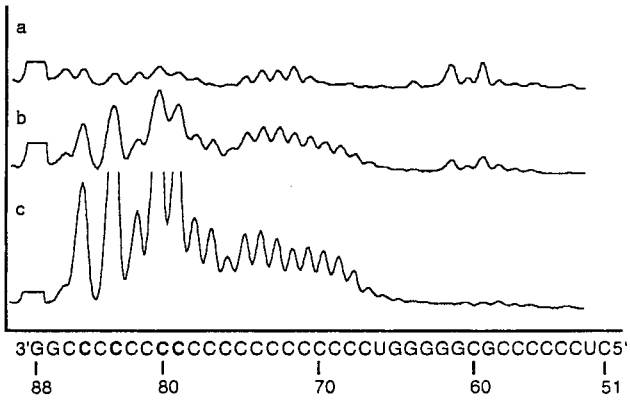

FIGURE 6: Selection of the secondary structure probing data for the nucleotides proposed to form helix III in rabbits: (a) control sample, (b) 0.02 unit of RNase CL3, and (c) 0.10 unit of RNase CL3. The numbering of nucleotides is according to Figure 3C. Modified bases are in bold. The sequence has been shifted as described in the legend of Figure 4.

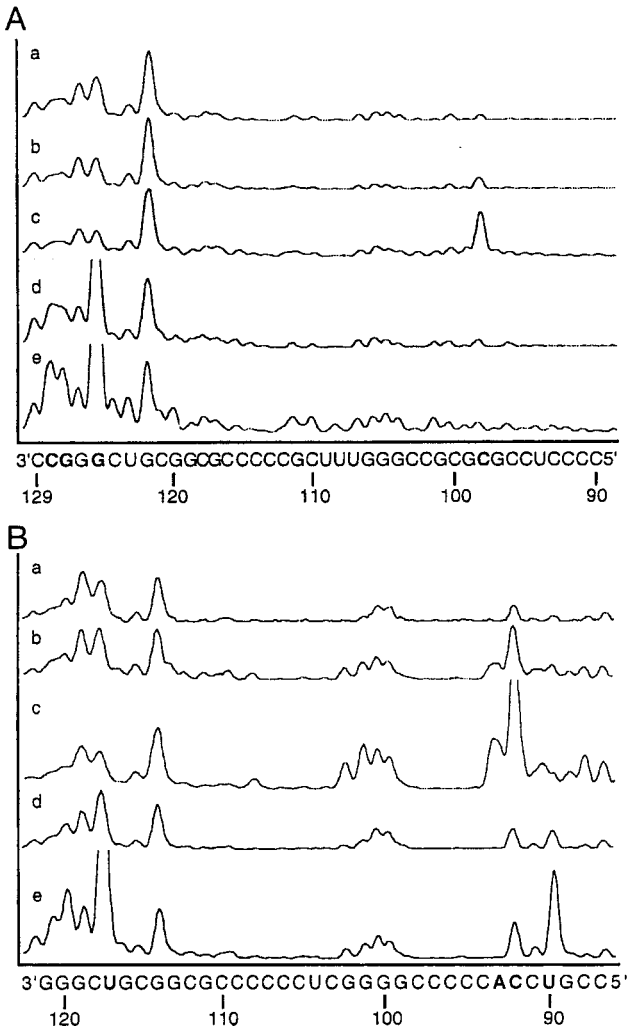

FIGURE 7: Selection of the secondary structure probing data for the nucleotides proposed to form hairpin B in mice and rats: (A) mouse, (B) rat, (a) control sample, (b) 20 mM DMS, (c) 90 mM DMS, (d) 20 mM CMCT, and (e) 100 mM CMCT. The numbering of nucleotides is according to panels A and B of Figure 3. Modified bases are in bold. The sequence has been shifted as described in the legend of Figure 4.

is also compatible with the strong CMCT reactivity of the N1 position of G36. The helical part of hairpin A contained two nucleotides (G23 and C39) that were positioned as bulged nucleotides in the model in Figure 3C due to the RNase CL3 accessibility. However, as the bulged bases are inaccessible for CMCT and DMS modification, it is possible that the bases form a G-C base pair and that the nuclease cleavage is due to structural distortions caused by the adjacent stack of proposed noncanonical base pairs.

As mentioned above, the sequence forming the apical part of the putative hairpin A was very accessible for modification by lead(II) acetate (Figure 3C). Experiments on synthetic model oligonucleotides and rRNAs indicate that all single-stranded regions and double-stranded regions with increased flexibility are accessible for lead(II) cleavage (*22*, *37*−*40*). Thus, the accessibility of the apical region of hairpin A for lead(II) acetate cleavage supports the conclusion that the bases were not involved in Watson−Crick base pairing. The possibility of modeling the loop as a GAGA tetraloop, with A28·G33 as a closing base pair, was rejected, as previous observations show that GAGA tetraloops are inaccessible to lead(II) cleavage (*40*).

A comparison of the mouse and rat sequences forming the putative hairpin A (Figure 3A,B) showed that the sequences were very similar, with nucleotide differences seen at only six positions (Figure 2). Similar secondary structures could therefore be expected. Nonetheless, this part of rat and mouse ES15 exhibited differences in their modification patterns (Figures 3A,B and 4). As the secondary structure models were constructed using the modification data, the differences resulted in slightly dissimilar models. As seen in panels A and B of Figure 3, the main differences between the models are the size of the single-stranded sequence connecting helix I and hairpin A and the subsequent shift of the nucleotides forming the latter structure. In the model of mouse hairpin A, the apical loop is a tetraloop, one of the most common types of apical loops in RNA structures (*41*). In rats, the apical loop is proposed to be a triloop containing the sequence GCG supported by the most common and stable closing base pair, C-G (*42*).

In the model of mouse hairpin A, A28 is suggested to be involved in an A-U base pair at the end of a helix, even though the base was strongly modified with DMS (Figures 3A and 4). As indicated above, this could be due to dynamic breathing at the terminus of the helix. However, DMS reactivity of A28 is also possible if the A28·U48 base pair is in the reverse Hoogsteen conformation, leaving the N1 position accessible for DMS modification (*30*). This type of base pair is typically found at helix termini and can be flanked by a G-C base pair as in the group I intron ribozyme domain (*30*, *43*). In the model in Figure 3A, the structure below the A-U pair is slightly sensitive to lead(II) cleavage. Above the A-U pair is a kethoxal and nuclease sensitive bulge. This bulge is followed by a nonreactive G-C base pair and a G·U base pair in which the N2 position of guanine is available for kethoxal modification (Figure 3A).

Alternatively, A28 is not bound to U48; instead, the kethoxal sensitive G29 pairs with U48, leaving A28 as the bulged nucleotide. The N2 position of G29 would then be available for kethoxal modification. However, this alternative results in a thermodynamically slightly less stable structure than the main model.

In rat hairpin A, there were very few modification sites overall, indicating a relatively stable hairpin structure (Figures 3B and 4). One ambiguous site was the RNase T_1 sensitive G43, which is located in a helix, although just before the final base pair, and should therefore not be accessible. However, as mentioned earlier, this can be attributed to an increased flexibility of base pairs neighboring bulges (*33*). In the model, the kethoxal reactive G41 is positioned in a G·G base pair also involving the nonreactive G31 (Figure 3B). This is possible if the G·G pair is in the GGN7 imino conformation [see Table 1 and Nagaswamy et al. (*30*)].

It has previously been shown that (GA)·(GA) tandem mismatched base pairs, such as those of the G27·A46 and A28·G45 pairs in Figure 3B, constitute a particularly stable mismatch, especially if in the context of neighboring G-C base pairs (*44*). Formation of a G·A pair involving G27 would still allow kethoxal access at the N2 position [see Table 1 and Nagaswamy et al. (*30*)]. The crystal structure of the 50S ribosomal subunit also shows that G·A base pairs are much more common in the rRNA structure than previously anticipated (*29*).

It is possible to shift the nucleotides in the mouse structure of haipin A to create a model more similar to that in rats and vice versa. However, these alternate structures are less compatible with the modification data, since several of the nucleotides that are accessible to single-strand specific reagents would then be modeled as double-stranded. In addition, these structures would be energetically less stable.

Helix II. The sequences proposed to form helix II are highly conserved between the six organisms presented in Figure 2, with only a few nucleotide exchanges between the three species studied here (Figure 3A−C). It was possible to model this helix in several alternate structures, because of the high number of guanines and cytosines. The secondary structure models for helix II presented in Figure 3 were the ones that best suited the experimental data and were most conserved between the three species that were investigated. In the structure models, the bases available for chemical modification have been positioned as bulged nucleotides with one exception, C133 in the mouse sequence (Figure 3). This base was slightly accessible for modification by CMCT but not by DMS (Figure 3A). CMCT-dependent modification of cytosines has previously been seen in other rRNA regions (*34*), but the structural requirements for this type of modification are not known.

In mice and rabbits, helix II was readily cleaved by RNase V_1, indicating a stacked and, most probably, helical structure. According to Lowman and Draper (*45*), RNase V_1 has no absolute requirement for a double-stranded helix; instead, it recognizes certain ordered stacking conformations, the most predominant being that of a phosphate backbone in a helical conformation. This backbone does not necessarily need to be in a double-stranded RNA helix, even though that would naturally be the most probable explanation. In our model, the helical conformation is also supported by the lack of accessibility of single-strand specific modification reagents to helix II.

Helix III. This proposed helix consists mainly of G-C and C-G base pairs (Figure 3). In rats, the bases forming the putative helix were essentially inaccessible for modification by single-strand specific reagents (Figure 5B). One exception was found at C67. This base was slightly accessible for modification by DMS (Figures 3B and 5B), most likely due to the instability of the neighboring U·G wobble base pair (*33*). In mice, the putative three-nucleotide bulge in helix III was supported by CMCT and lead(II) modification data (Figures 3A and 5A).

In rabbits (Figure 3C), the G-C base pairs proposed to form helix III were very accessible to Pb(II), even though the 3′-side of this suggested helix was also highly accessible to RNase V_1. The strong lead acetate cleavage sites indicated a dynamic property of this proposed helix, in contrast to the apparent stability of helix III in mice and rats.

As mentioned above, the recognition element for RNase V_1 is that of a helically stacked phosphate backbone, or a structure similar to it (*45*). Thus, a dynamic structure with helical elements would support the suggested helix III in rabbits. A stacked nature would also explain the suggested single-stranded nucleotides that were highly accessible to RNase V_1, such as nucleotides 76 and 77 in Figure 3C.

During sequencing of rabbit ES15, we noticed that the sequence contained a long stretch of 20 cytosines, C67−C86 (Figure 2). This sequence was extremely sensitive to cleavage by RNase CL3 (Figure 6). At a low RNase CL3 concentration (Figure 6b), cleavage was mainly introduced between C79 and C86. At higher concentrations, the enzyme had access to the phosphate backbone at almost all cytosines in the sequence (Figure 6c). The increased number of cleavage sites is most likely due to a successive destabilization of the RNA upon initial cleavage of the sequence and does probably not reflect the secondary structure of this region. The sequence between C70 and C85 was also hyperreactive to several single- and double-strand specific reagents, suggesting that the sequence forms a very dynamic structure that could result in the complex cleavage pattern summarized in Figure 3C.

The rabbit sequence from G59 to C75 seems phylogenetically most similar to helix III in the other species (Figure 2). On the other hand, this region was extremely accessible to cleavage by lead(II) acetate, while the proposed helices III in mice and rats were essentially inaccessible to cleavage outside of loops and bulges (Figure 3). It is therefore possible that nucleotides G59−C75 in rabbits are more homologous to the proposed hairpin B in mice and rats rather than helix III.

As seen in Figure 3, helix III is closed by a terminal loop. In the model of mouse ES15, the loop is a pentaloop (Figure 3A). The sequence forming the loop was highly reactive toward both single-strand specific reagents and Pb(II) (Figures 3A and 5A). In rats, helix III supports an apical tetraloop (Figure 3B). Tetraloops are the most common type of terminal loops in RNA helices (*46*). The proposed ^{71}UUCG^{74} tetraloop (Figure 3B) was accessible for modification by DMS and CMCT but inert to Pb(II)-induced cleavage (Figures 3B and 5B). The lack of Pb(II) reactivity is consistent with the observations of Ciesiolka et al. (*40*). The UUCG tetraloop has been extensively examined and shown to be one of the most stable tetraloops [see Williams (*47*) and Leulliot (*48*)]. In eukaryotic 18S rRNA, the UUCG tetraloop is most commonly closed by a G-C pair, although a closing C-G pair, such as in our proposed model (Figure 3B), is quite prevalent, and the most stable closing pair for a tetraloop (*47*). Recently, Dale et al. (*49*) showed that a sequence identical to the GGCUUCGGCC sequence present in rat ES15 spontaneously forms a stable hairpin with a UUCG tetraloop, consistent with our model of helix III in rats (Figure 3B).

Hairpin B. The sequence forming hairpin B in the secondary structure models was absent in rabbit ES15 (Figure 3C). Even in mice and rats, the sequences showed little similarity except for their high content of guanines and cytosines (Figure 3A,B). This made it virtually impossible to single out one secondary structure model without the secondary structure data.

The proposed secondary structures for hairpin B contain short helical segments, separated by single-strand stretches of variable length, and an apical loop containing five (rat) or six (mouse) nucleotides (Figure 3A,B). The helical nature of the sequences suggested to be involved in the short helices was supported by RNase V_1 cleavage. The apical penta- and hexaloops of the proposed hairpin B showed no sequence similarities (Figure 3A,B). In both species, the loop was accessible to several single-strand specific reagents.

In the models, the apical helices contain a noncanonical C-C base pair (Figure 3A,B). According to Kierzek et al.,

C·C mismatches such as those suggested for C103·C116 in the mouse sequence and C95·C115 and C98·C110 in the rat sequence, destabilizes the helix, but would have a relatively small effect since the flanking G-C pairs stabilize the structure (*50*).

According to the model in Figure 3B, the apical helix in rat hairpin B is connected to a short three-base pair helix via two single-strand bulged bases, C112 and G113. The latter base was accessible for modification by kethoxal, although the whole sequence (111GCGGCG116) was sensitive to RNase V_1 cleavage. As mentioned earlier, this is possible if the bulged nucleotides are stacked in an ordered structure (*45*).

In rats, the initial helical stem of hairpin B is suggested to contain two bulged nucleotides, C88 and G121 (Figure 3B), based on their accessibility to micrococcal nuclease and kethoxal. The helical stem and the apical hairpin are joined together via a bulged loop that was sensitive to single-strand specific modification (Figure 3B). A93 and U117, in the bulge loop, could not be joined in a Watson−Crick type of base pair as the N1 position of A93 and the N3 position of U117 were available for modification by DMS and CMCT, respectively (Figure 7B). The strong reactivity of U117 also excluded formation of a reverse Hoogsteen base pair.

In the model of mouse ES15, the helical stem of hairpin B contained a trinucleotide loop, sensitive to RNase T_1, kethoxal, and lead(II), and a U·C base pair (Figure 3A). U·C base pairs are present in crystals of synthetic RNA double helices and in 23S rRNA (*29*, *51*). The bases in a U·C base pair should not be sensitive to DMS and CMCT modification, as was also the case with the U94·C121 putative pair (Figure 3A and 7A). In the model (Figure 3A), the initial helical stem was connected to the apical part of the hairpin via a tetranucleotide bulge displaying accessibility for DMS, kethoxal, and Pb(II).

General Structural Observations. Overall, ES15 from the three organisms studied was relatively accessible for chemical and enzymatic modification. Both the small reagents DMS, kethoxal, and CMCT and the much larger nucleases induced a large number of strong modifications and cleavages. The available data in the literature suggest that the three-dimensional structure of the rRNA core is identical in ribosomes from eubacteria and eukaryotes and that the extra sequences found in eukaryotic rRNA lie outside this core (*52*). Previous observations show that these extra sequences are significantly more susceptible to nuclease cleavage than the rRNA core, indicating that they are near the surface of the ribosome (*34*, *53*). This is confirmed by recent data. The bacterial counterpart of helix 44.1 is located in the middle of the back of the large ribosomal subunit, where it is exposed on the surface (*29*). It should be noted that archaebacteria, such as *Haloarcula marismortui*, for which the crystal structure of the large ribosomal subunit has been determined (*29*), also contain expansion segments. However, these are fewer in number and shorter than in eukaryotes (*6*, *54*), and there is no counterpart of ES15 in archaea. In a comparison of the three-dimensional structures of rabbit and yeast ribosomes, i.e., ribosomes containing 23S-like rRNA with and without ES15, Morgan et al. (*55*) identified a protruding finger adjacent to the L7/L12 stalk in rabbit 60S ribosomal subunits as ES15. These studies show that ES15 is situated on the surface of the ribosome, where it could be accessible for modification. This is consistent with its suggested role in ribosome degradation in apoptosis (*56*, *57*).

The structural models presented here were based on chemical and enzymatic modification data, as well as phylogenetic comparisons made between the three species that were studied. ES15 sequences are known for an additional three organisms, humans, gorillas, and chimpanzees (Figure 2) (*15*). These sequences are longer, largely due to the higher number of short repeats. A sequence comparison with mouse, rat, and rabbit ES15 allows some predictions concerning the possible secondary structure, although this prediction is hampered by the lack of experimental data. The sequences suggested to form helix I are highly similar between rodents and primates, except that the helix can be extended with a few base pairs in primates. Thus, the primate sequences could be folded into a structure similar to helix I (Figure 3). The sequences that form helix II in mice are present in primates, but contain an additional adenosine. There is also a possibility of forming a homologue to helix III, although the additional cytosines and guanines allow formation of a substantially longer GC rich helical stem terminated by an apical UCCU tetraloop (see Figure 2). Likewise, the sequence between helices I and II in primates allows construction of a hairpin similar to hairpin A, although the additional nucleotides in primate ES15 would result in a more extended helical stem. This was most obvious in chimpanzees, where the additional sequences consist of a large number of short repetitive elements. Construction of a primate structure similar to hairpin B was more difficult due to the large proportion of cytosines found in primate ES15, most notable in the chimpanzee sequence. A folding of these sequences tends to generate more than one hairpin, thereby creating a bulkier structure. As mentioned above, hairpin B was the most divergent between the species that were studied, and was even absent in the short rabbit ES15 sequence (Figure 3C). It is therefore possible that this part of the ES15 structure allows considerable flexibility in both size and structure of the inserted sequences. If so, the sequences forming hairpin B could be expected to have a more exposed position on the ribosomal surface than the remaining ES15. This would be in agreement with hairpin B being the most nuclease sensitive region in ES15.

The secondary structure prediction of *Haloarcula marismortui* 23 S rRNA, based on phylogenetic analysis (*31*), has been shown to be in agreement with that found in the 50S subunit (*29*), with the exception of several additional, mainly noncanonical, base pairs. Thus, comparative analysis generates a surprisingly accurate secondary structure model of ribosomal RNAs. One can therefore assume that the predicted secondary structure of the eukaryotic rRNA core will be largely correct. However, the lack of useful phylogenetic information has made it impossible to generate secondary structure models for the expansion segments of the eukaryotic rRNAs. Secondary structure predictions for these expansion segments will therefore have to rely on other methods while awaiting a crystallographic determination of the eukaryotic ribosomal structure. Here we show that combining available phylogenetic and experimental structure probing data (*24*, *58*) can be useful in generating secondary structure models for these parts of the eukaryotic rRNAs.

ACKNOWLEDGMENT

The technical assistance of Mrs. B. Lundberg is gratefully acknowledged.

REFERENCES

1. Gerbi, S. A. (1996) in *Ribosomal RNA: structure, evolution, processing, and function in protein biosynthesis* (Zimmermann, R. A., and Dahlberg, A. E., Eds.) pp 71−87, CRC Press, Boca Raton, FL.
2. Green, R., and Noller, H. F. (1997) *Annu. Rev. Biochem. 66*, 679−716.
3. Noller, H. F., Green, R., Heilek, G., Hoffarth, V., Huttenhofer, A., Joseph, S., Lee, I., Lieberman, K., Mankin, A., Merryman, C., Powers, T., Puglisi, E. V., Samaha, R. R., and Weiser, B. (1995) *Biochem. Cell Biol. 73*, 997−1009.
4. Gorski, J. L., Gonzalez, I. L., and Schmickel, R. D. (1987) *J. Mol. Evol. 24*, 236−251.
5. Musters, W., Concalves, P. M., Boon, K., Raue, H. A., van Heerikhuizen, H., and Planta, R. J. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 1469−1473.
6. Michot, B., and Bachellerie, J. P. (1987) *Biochimie 69*, 11−23.
7. Sweeney, R., Chen, L., and Yao, M. C. (1994) *Mol. Cell. Biol. 14*, 4203−4215.
8. Clark, C. G., Tague, B. W., Ware, V. C., and Gerbi, S. A. (1984) *Nucleic Acids Res. 12*, 6197−6220.
9. Wakeman, J. A., and Maden, B. E. (1989) *Biochem. J. 258*, 49−56.
10. Hancock, J. M., and Dover, G. A. (1990) *Nucleic Acids Res. 18*, 5949−5954.
11. Gonzalez, I. L., Gorski, J. L., Campen, T. J., Dorney, D. J., Erickson, J. M., Sylvester, J. E., and Schmickel, R. D. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 7666−7670.
12. Clark, C. G. (1987) *J. Mol. Evol. 25*, 343−350.
13. Musters, W., Boon, K., van der Sande, C. A., van Heerikhuizen, H., and Planta, R. J. (1990) *EMBO J. 9*, 3989−3996.
14. Jeeninga, R. E., Van Delft, Y., De Graaff-Vincent, M., Dirks-Mulder, A., Venema, J., and Raue, H. A. (1997) *RNA 3*, 476−488.
15. Kuo, B. A., Gonzalez, I. L., Gillespie, D. A., and Sylvester, J. E. (1996) *Nucleic Acids Res. 24*, 4817−4824.
16. Hassouna, N., Michot, B., and Bachellerie, J. P. (1984) *Nucleic Acids Res. 12*, 3563−3583.
17. Chan, Y.-L., Olvera, J., and Wool, I. G. (1983) *Nucleic Acids Res. 11*, 7819−7831.
18. Morley, S. J., and Jackson, R. J. (1985) *Biochim. Biophys. Acta 825*, 45−56.
19. Nygård, O., and Nilsson, L. (1984) *Eur. J. Biochem. 145*, 345−350.
20. Holmberg, L., Melander, Y., and Nygård, O. (1992) *J. Biol. Chem. 267*, 21906−21910.
21. Stern, S., Wilson, R. C., and Noller, H. F. (1986) *J. Mol. Biol. 192*, 101−110.
22. Ciesiolka, J., Lorenz, S., and Erdmann, V. A. (1992) *Eur. J. Biochem. 204*, 583−589.
23. Noller, H. F., Hoffarth, V., and Zimniak, L. (1992) *Science 256*, 1416−1419.
24. Mathews, D. H., Sabina, J., Zuker, M., and Turner, D. H. (1999) *J. Mol. Biol. 288*, 911−940.
25. Zuker, M. (1989) *Science 244*, 48−52.
26. Jaeger, J. A., Turner, D. H., and Zuker, M. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 7706−7710.
27. Jaeger, J. A., Turner, D. H., and Zuker, M. (1990) *Methods Enzymol. 183*, 281−306.
28. Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994) *Nucleic Acids Res. 22*, 4673−4680.
29. Ban, N., Nissen, P., Hansen, J., Moore, P. B., and Steitz, T. A. (2000) *Science 289*, 905−920.
30. Nagaswamy, U., Voss, N., Zhang, Z. D., and Fox, G. E. (2000) *Nucleic Acids Res. 28*, 375−376.
31. Gutell, R. R., Gray, M. W., and Schnare, M. N. (1993) *Nucleic Acids Res. 21*, 3055−3074.
32. Ehresmann, C., Baudin, F., Mougel, M., Romby, P., Ebel, J. P., and Ehresmann, B. (1987) *Nucleic Acids Res. 15*, 9109−9128.
33. Stebbins-Boaz, B., and Gerbi, S. A. (1991) *J. Mol. Biol. 217*, 93−112.
34. Holmberg, L., Melander, Y., and Nygård, O. (1994) *Nucleic Acids Res. 22*, 1374−1382.
35. Scott, W. G., Finch, J. T., and Klug, A. (1995) *Cell 81*, 991−1002.
36. Jovine, L., Hainzl, T., Oubridge, C., Scott, W. G., Li, J., Sixma, T. K., Wonacott, A., Skarzynski, T., and Nagai, K. (2000) *Structure 8*, 527−540.
37. Ciesiolka, J., Lorenz, S., and Erdmann, V. A. (1992) *Eur. J. Biochem. 204*, 575−581.
38. Gornicki, P., Baudin, F., Romby, P., Wiewiorowski, M., Kryzosiak, W., Ebel, J. P., Ehresmann, C., and Ehresmann, B. (1989) *J. Biomol. Struct. Dyn. 6*, 971−984.
39. Winter, D., Polacek, N., Halama, I., Streicher, B., and Barta, A. (1997) *Nucleic Acids Res. 25*, 1817−1824.
40. Ciesiolka, J., Michalowski, D., Wrzesinski, J., Krajewski, J., and Krzyzosiak, W. J. (1998) *J. Mol. Biol. 275*, 211−220.
41. Woese, C. R., Winker, S., and Gutell, R. R. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 8467−8471.
42. Shu, Z., and Bevilacqua, P. C. (1999) *Biochemistry 38*, 15369−15379.
43. Cate, J. H., Gooding, A. R., Podell, E., Zhou, K., Golden, B. L., Kundrot, C. E., Cech, T. R., and Doudna, J. A. (1996) *Science 273*, 1678−1685.
44. Wu, M., McDowell, J. A., and Turner, D. H. (1995) *Biochemistry 34*, 3204−3211.
45. Lowman, H. B., and Draper, D. E. (1986) *J. Biol. Chem. 261*, 5396−5403.
46. Konings, D. A., and Gutell, R. R. (1995) *RNA 1*, 559−574.
47. Williams, D. J., and Hall, K. B. (2000) *J. Mol. Biol. 297*, 1045−1061.
48. Leulliot, N., Baumruk, V., Abdelkafi, M., Turpin, P. Y., Namane, A., Gouyette, C., Huynh-Dinh, T., and Ghomi, M. (1999) *Nucleic Acids Res. 27*, 1398−1404.
49. Dale, T., Smith, R., and Serra, M. J. (2000) *RNA 6*, 608−615.
50. Kierzek, R., Burkard, M. E., and Turner, D. H. (1999) *Biochemistry 38*, 14214−14223.
51. Holbrook, S. R., Cheong, C., Tinoco, I., Jr., and Kim, S. H. (1991) *Nature 353*, 579−581.
52. Dube, P., Bacher, G., Stark, H., Mueller, F., Zemlin, F., Van Heel, M., and Brimacombe, R. (1998) *J. Mol. Biol. 279*, 403−421.
53. Holmberg, L., and Nygård, O. (1997) *Eur. J. Biochem. 247*, 160−168.
54. Michot, B., Qu, L. H., and Bachellerie, J. P. (1990) *Eur. J. Biochem. 188*, 219−229.
55. Morgan, D. G., Ménétret, J.-L., Radermacher, M., Neuhof, A., Akey, I. V., Rapoport, T. A., and Akey, C. W. (2000) *J. Mol. Biol. 301*, 301−321.
56. Houge, G., Robaye, B., Eikhom, T. S., Golstein, J., Mellgren, G., Gjertsen, B. T., Lanotte, M., and Doskeland, S. O. (1995) *Mol. Cell. Biol. 15*, 2051−2062.
57. Houge, G., and Doskeland, S. O. (1996) *Experientia 52*, 963−967.
58. Juan, V., and Wilson, C. (1999) *J. Mol. Biol. 289*, 935−947.
59. Cuatrecasas, P., Fuchs, S., and Anfinsen, C. B. (1967) *J. Biol. Chem. 242*, 1541−1547.
60. Takahashi, K., and Moore, S. (1982) *Enzymes 15*, 435−468.
61. Boguski, M. S., Hieter, P. A., and Levy, C. C. (1980) *J. Biol. Chem. 255*, 2160−2163.
62. Favorova, O. O., Fasiolo, F., Keith, G., Vassilenko, S. K., and Ebel, J. P. (1981) *Biochemistry 20*, 1006−1011.
63. Higgins, D. G., Thompson, J. D., and Gibson, T. J. (1996) *Methods Enzymol. 266*, 383−402.

BI002286Q