Evolutionarily conserved structural features in the ITS2 of mammalian pre-rRNAs and potential interactions with the snoRNA U8 detected by comparative analysis of new mouse sequences

B. Michot*, N. Joseph, S. Mazan+ and J. P. Bachellerie

Laboratoire de Biologie Moléculaire Eucaryote du C.N.R.S., Université Paul Sabatier, 118 route de Narbonne, 31062 Toulouse Cedex, France

Received March 3, 1999; Revised and Accepted April 19, 1999 DDBJ/EMBL/GenBank accession nos AJ132552–AJ132555

ABSTRACT

Mechanisms of ITS2 excision from pre-rRNA remain largely elusive. In mammals, at least two endonucleolytic cleavages are involved, which result in the transient accumulation of precursors to 5.8S rRNA termed 8S and 12S RNAs. We have sequenced ITS2 in four new species of the Mus genus and investigated its secondary structure using thermodynamic prediction and comparative approach. Phylogenetic evidence supports an ITS2 folding organized in four domains of secondary structure extending from a preserved structural core. This folding is also largely conserved for the previously available mammalian ITS2 sequences, rat and human, despite their extensive sequence divergence relative to the Mus species. Conserved structural features include the structural core, containing the 3′ end of 8S pre-rRNA within a single-stranded sequence, and a stem containing the 3′ end of the 12S pre-rRNA species. A putative, phylogenetically preserved pseudoknot has been detected 1 nt downstream from the 12S 3′ end. Two long complementarities have also been identified, in sequences conserved among vertebrates, between the pre-rRNA 32S and the snoRNA (small nucleolar RNA) U8 which is required for the excision of Xenopus ITS2. The first complementarity involves the 5.8S-ITS2 junction and 13 nt at the 5′ end of U8, whereas the other one occurs between a mature 28S rRNA segment known to be required for ITS2 excision and positions 15–25 of snoRNA U8. These two potential interactions, in combination with ITS2 folding, could organize a functional pocket containing three cleavage sites and key elements for pre-rRNA processing, suggesting a chaperone role for the snoRNA U8.

INTRODUCTION

In most eukaryotic organisms, the mature rRNA sequences (SSU rRNA, 5.8S and LSU rRNA) are transcribed in the nucleolus in the form of a single large precursor in which the different mature rRNA sequences are separated by spacer regions, 5′ETS, ITS1, ITS2 and 3′ETS. Nucleolar processing of the large pre-rRNA transcript involves an elaborate pathway of endo- and exonucleolytic cleavages taking place on a precursor RNA assembled with ribosomal proteins into an RNP particle of very complex structure. Maturation of pre-rRNA also includes the formation of a complex pattern of nucleoside modifications on the mature sequences of the precursor, most of them corresponding to 2′-O ribose methylations and pseudouridylation. Throughout ribosome biogenesis, the pre-ribosomal RNP particle transiently associates with scores of snoRNAs (small nucleolar RNAs), only a minority of which are required for definite pre-rRNA cleavages (1). Most snoRNAs guide the two major types of rRNA nucleoside modifications through transient base-pairing with pre-rRNA at the modification site (2–4). While pre-rRNA processing has been extensively studied in various organisms, including Tetrahymena, Xenopus, mouse, rat and human, the major advances on the question have been achieved in yeast Saccharomyces cerevisiae. Through the development of elegant in vivo systems involving a powerful combination of biochemical and genetic approaches, a wealth of detailed information is now available in yeast, both on the cis-acting elements and the trans-acting factors required at definite stages of pre-rRNA processing (5). Although the biological role of the ITS spacers is not well understood, the utilization of the yeast model has definitely shown their importance for a faithful production of mature rRNAs. It has also revealed that processings of the different spacers are tightly coupled. Thus, deletions in the 3′ETS affect the processing of both the ITS2 and 5.8S rRNAs (6,7), while the integrity of a stem–loop within this spacer is required for internal cleavages taking place within ITS1 (8). Moreover, a

*To whom correspondence should be addressed. Tel: +33 5 61 33 58 65; Fax: +33 5 61 33 58 86; Email: bmichot@ibcg.biotoul.fr

+Present address: Equipe ATIPE, UPRESA Développement et Evolution, Bâtiment 444, Université Paris-Sud, 91405 Orsay Cedex, France
deletion within ITS2 not only prevents biogenesis of the large subunit but also affects the maturation of small-subunit rRNA (9). Finally, efficient production of 25S rRNA requires ITS1 sequences (10,11). The processing steps which have been best studied so far are those leading to the production of SSU-rRNA, which requires four essential snoRNAs, U3, U14, snR30 and MRP (12). In contrast, the molecular mechanisms by which ITS2, a eukaryote-specific insertion, is removed and the mature 3′ end of 5.8S and 5′ end of 28S rRNA are generated remain largely elusive. ITS2 sequences exhibit a high rate of size and sequence variation during evolution of eukaryotes, which has generally hampered the derivation of reliable secondary structure models by comparative analysis. Yeast _Saccharomyces_ ITS2 is the only specimen for which the secondary structure has been unambiguously resolved, through a combination of chemical and enzymatic probing, minimum energy modeling, genetic experiments and phylogenetic analyses (13,14).

Among distant yeast species, while its size may vary substantially (66–240 nt) and its conserved sequences are restricted to a few short tracts, ITS2 can fold into a largely similar secondary structure (15). _In vitro_ mutational analysis guided by phylogenetic comparisons has pointed to the crucial importance of some of the most conserved structural elements for proper ITS2 excision and biogenesis of active ribosomal subunits (16). Vertebrate ITS2 do not exhibit any sequence homology with yeast ITS2. They are generally much longer, exhibiting dramatic variations in size, from 262 nt in Xenopus _laevis_ (17) to > 1000 nt in mammals. The rate of mammalian ITS2 sequence variation is so high that even among species as closely related as rat and mouse, only a few short tracts of significant sequence similarities can be detected (18,19).

Our present knowledge about the excision of ITS2 suggests major differences between yeast and vertebrate mechanisms. In both yeast and vertebrates, the mature 3′-terminus of 5.8S rRNA and 5′-terminus of LSU rRNA are formed by endonucleolytic cleavage(s) in ITS2 followed by action of exonucleases. However, the two groups of organisms differ in the number of endonucleolytic cleavages involved. In mammals, two precursor intermediates to 5.8S rRNA have been identified, the 8S and 12S pre-rRNAs (20–22) corresponding to a 5.8S rRNA still linked to an ITS2 extension, of 118 and 581 nt, respectively. In contrast, only one internal cleavage has been detected in yeast, corresponding to the accumulation of a pre-5.8S rRNA processing intermediate, termed 7S, containing an ITS2 extension of ~140 nt in _Saccharomyces_ (23) and ~100 nt in Schizosaccharomyces _pombe_ (24). Moreover, while in _X. laevis_ snoRNA U8 is required for cleavages at the 5′ end of 5.8S and 3′ end of 28S, through mechanisms which remain unclear (25), yeast cells do not seem to contain a U8 snoRNA homolog.

A detailed knowledge of the ITS2 folding should provide the basis for further analyzing ITS2 function in ribosome biogenesis and the molecular mechanisms involved in its excision from pre-rRNA. Since a first attempt based on the sole mouse–rat comparison (18), no ITS2 folding model based on the comparative approach has been proposed for mammals, due to the lack of other related sequences. With this objective in mind, we have sequenced the ITS2 of four new species belonging to the _Mus_ genus, in order to derive a set of sequences sufficiently related to each other, i.e. appropriate for a comparative approach of their secondary structure. By a combination of minimum energy models and phylogenetic comparisons, we have identified a set of structural features, conserved not only among rodents but also in humans, which are likely to contain important functional elements of the ITS2, particularly for its excision. Finally, we have used this novel information on ITS2 folding to test, on a comparative basis among distant vertebrates, the potential for its base-pairing interactions with the snoRNA U8, and have identified a phylogenetically supported base-pairing that could play a key role in the excision of ITS2.

**MATERIALS AND METHODS**

**Polymerase chain reaction (PCR) and DNA sequencing**

Tissues from the following four mouse species belonging to the _Mus_ genus, _Mus_ _cervicolor_, _Mus_ _caroli_, _Mus_ _Coelomys_ _pahari_ and _Mus_ (_Pyromys_) _playthrix_, were kindly provided by Dr F. Bonhomme (Université Montpellier 2). ITS2 sequences were determined on cloned PCR fragments following amplification of genomic DNA. In the first step, primers spanning conserved segments at the 5′ end of 28S rRNA (LH3_reverse: 3′-TCCCTTTTCCTTTTGGTTGCTCCCTAGGCCC-5′) and 3′ end of 5.8S rRNA (LHI_forward: 3′-TTTGTGTGCTGATCTGTCGCTTTAAGATA-5′) were used. PCR conditions were set up according to the recommendations of the _RecT_ DNA polymerase supplier (Biolabs) using 0.2–0.5 ng genomic DNA (typical thermocycling profile: 30 cycles with 1 min at 98°C, 1 min at 55°C and 2 min at 80°C). PCR products were digested by EcoRI and BamHI before cloning in pTZ18R. Automated sequencing (ALF) was performed on double-strand plasmid DNA, using the Pharmacia PL AutoRead Kit and universal sequencing primers. To complete the sequencing of the central region of the ITS2, three additional primers, selected within portions of the ITS2 sequence conserved between the different _Mus_ species: (8S_forward: 3′-GACCATTGAGCCCTCTGCC-5′; 5.8S+340_forward: 3′-CCAGGGTGCCCGGCGGTGCT-5′ and 12S_reverse: 3′-AGGCCGCGGCCGGCGCGCCCT-5′), were used in sequencing reactions performed in the presence of fluorescein-15-dATP or dUTP (Boehringer Mannheim) according to the supplier’s protocols. Likewise, sequence ambiguities at the 3′ end of ITS2 were solved by utilization of an additional sequencing primer at the 5′ end of 28S (28S-b_reverse: 3′-GCTTGAGCTTAGTCTGACC-5′). Both strands were sequenced at least twice for each species. The sequences have been deposited at the EMBL data library under the accession numbers AJ132552 (_M._ _cervicolor_), AJ132553 (_M._ _caroli_), AJ132554 (_M._ _Coelomys_ _pahari_), AJ132555 (_M._ _Pyromys_ _playthrix_).

**Sequence alignment and secondary structure prediction**

In order to achieve the best phylogenetically supported folding models, we used an iterative method in which an alignment initially performed using the program Multalin (26) was progressively optimized according to secondary structure homologies. Optimal and suboptimal foldings were predicted for each sequence on a thermodynamic basis (27) using the software package of the Genetic Computer Group, University of Wisconsin (28). Proposed foldings were compared and solutions maximizing structural homologies retained and checked for the presence of compensatory changes. The alignments were then refined and new foldings derived from the improved alignments. Finally, we created a database incorporating secondary structure informations in the alignment file (29), to search for potential pseudoknots and derive consensus secondary structures.

**Polymerase chain reaction (PCR) and DNA sequencing**

Tissues from the following four mouse species belonging to the _Mus_ genus, _Mus_ _cervicolor_, _Mus_ _caroli_, _Mus_ _Coelomys_ _pahari_ and _Mus_ (_Pyromys_) _playthrix_, were kindly provided by Dr F. Bonhomme (Université Montpellier 2). ITS2 sequences were determined on cloned PCR fragments following amplification of genomic DNA. In the first step, primers spanning conserved segments at the 5′ end of 28S rRNA (LH3_reverse: 3′-TCCCTTTTCCTTTTGGTTGCTCCCTAGGCCC-5′) and 3′ end of 5.8S rRNA (LHI_forward: 3′-TTTGTGTGCTGATCTGTCGCTTTAAGATA-5′) were used. PCR conditions were set up according to the recommendations of the _RecT_ DNA polymerase supplier (Biolabs) using 0.2–0.5 ng genomic DNA (typical thermocycling profile: 30 cycles with 1 min at 98°C, 1 min at 55°C and 2 min at 80°C). PCR products were digested by EcoRI and BamHI before cloning in pTZ18R. Automated sequencing (ALF) was performed on double-strand plasmid DNA, using the Pharmacia PL AutoRead Kit and universal sequencing primers. To complete the sequencing of the central region of the ITS2, three additional primers, selected within portions of the ITS2 sequence conserved between the different _Mus_ species: (8S_forward: 3′-GACCATTGAGCCCTCTGCC-5′; 5.8S+340_forward: 3′-CCAGGGTGCCCGGCGGTGCT-5′ and 12S_reverse: 3′-AGGCCGCGGCCGGCGCGCCCT-5′), were used in sequencing reactions performed in the presence of fluorescein-15-dATP or dUTP (Boehringer Mannheim) according to the supplier’s protocols. Likewise, sequence ambiguities at the 3′ end of ITS2 were solved by utilization of an additional sequencing primer at the 5′ end of 28S (28S-b_reverse: 3′-GCTTGAGCTTAGTCTGACC-5′). Both strands were sequenced at least twice for each species. The sequences have been deposited at the EMBL data library under the accession numbers AJ132552 (_M._ _cervicolor_), AJ132553 (_M._ _caroli_), AJ132554 (_M._ _Coelomys_ _pahari_), AJ132555 (_M._ _Pyromys_ _playthrix_).
structure drawings were produced using ESSA (30). Complementarities between U8 snoRNA and 32S pre-rRNA were searched with both the GCG package and the ESSA program. Alignments and foldings not presented here are available upon request.

RESULTS

The length of the ITS2 sequence exhibits significant variation among the four additional species of the Mus genus analyzed in this study, ranging from 1080 to 1219 nt ([M. cervicolor, 1080 nt; M. (C) pahari, 1094 nt; M. caroli, 1102 nt; M. (P) platythrix, 1219 nt]. However, the values remain collectively relatively close to the size of the Mus musculus ITS2, 1089 nt (18), which is substantially larger than the 765 nt rat homolog (31). All the novel Mus ITS2 share the same highly biased base content noticed for the Musmusculus sequence (18), i.e. very high in GC (~75%) and very low in A (~6.5%). As shown in Figure 1, two of the novel ITS2 sequences, M. cervicolor and M. caroli, appear more particularly related to Musmusculus: the three sequences can be unambiguously aligned over most of their length, the limited size differences (~26 nt) resulting from a few short insertions/deletions essentially located in the 3′ half of the ITS2. The two other ITS2 sequences, for M. (C) pahari and M. (P) platythrix, are also strongly similar to the three other Mus sequences over most of their length but they exhibit a substantially larger number of segmental insertions/deletions. On the basis of the sequence alignments, the Mus ITS2 can be divided into three main regions (Fig. 1). The first one extends over the 120 5′-terminal nucleotides, spanning the strongly conserved segments a–c. This region contains only two sites of very short segmental insertions/deletions, termed D.I and D.II. Located immediately downstream from the first one, a second region of ~550 nt contains nine conserved sequence tracts, d–l, which amount to about half its size. Conserved segments d–l are separated by more variable sequence stretches which frequently exhibit short to moderate size variations (1–33 nt). The third region, corresponding roughly to the ITS2 3′ half, extends downstream from conserved segment l. Although this appears as the most variable region, with several large segmental insertions/deletions (among them, a 109 nt additional sequence in M. (P) platythrix), it nevertheless contains two conserved sequence tracts, m and n, of which n is the longest at 80 nt, over the whole ITS2.

At the secondary structure level, we have identified 11 stems, termed I–IV and 1–7, that are conserved among the five Mus species (Fig. 1). Nine of them involve Watson–Crick base pairings within or between the above-mentioned conserved segments. The four stems I–IV delineate four independent domains of secondary structure identified A–D (Fig. 2). Domains A and B are both located within the conserved 5′ terminal region, while domains C and D roughly span the central region and the more variable 3′ half of ITS2, respectively. Sequences of stems I–III are strongly conserved. While stem II cannot be tested by the comparative approach, stems I and III are phylogenetically supported with, in addition to a G:C–C:G change in stem III, several semi-conservative G:U–G:C changes in both stems. As for stem IV, although the delineation of the two sequence tracts involved in its formation in each of the five Mus sequences is less straightforward due to the relatively high divergence of the 3′ half of the ITS2, it is also supported on a comparative basis. The formation of stems I–IV bring into close proximity five of the most conserved segments of the ITS2, a, b, c, k and l, providing a structural core for the folding of the entire ITS2 into four independent domains of secondary structure linked by relatively short single-strand tracts of conserved sequence. It seems noteworthy that one of these linker segments, between domains B and C, contains one of the two sites of internal ITS2 processing, corresponding to the 3′ end of the 8S precursor to 5.8S RNA. Intriguingly, the core of secondary structure constituted by stems I–IV and single-strand linkers seems to be under a compositional constraint significantly different from the rest of the ITS2. It is markedly enriched in conserved adenines, with 11.6% of its nucleotide positions invariant between the five Mus species being adenines (versus only 4.2% for the other conserved tracts outside the conserved central core). Moreover, most of these conserved adenines (12 out of 18) are unpaired. Finally, it seems noteworthy that the single-strand region immediately downstream from stem IV at the 3′ end of the ITS2, although highly variable in size and nucleotide sequence, exhibits a common feature in all the Mus species, with the systematic presence of a short CG-rich motif flanked by two 8–13 nt pyrimidine-rich sequences.

Despite the extensive variation of its sequence, domain C can adopt a very similar folding in the five Mus species (Fig. 3). In all cases it folds into a single very long helix (accommodating ~120 bp with a substantial amount of helix irregularities) which exhibits a Y-shaped apical part and very few branched lateral stems. This folding pattern is largely determined by pairings between two sets of conserved tracts, d–j and e–i, respectively, which are phylogenetically supported, as depicted in Figure 3. The folding of the conserved apical Y-shaped structure is also supported on a comparative basis, with the occurrence of compensatory changes on five of its base pairings.

Domain D exhibits a more complex secondary structure (Fig. 4). Although its overall folding backbone looks very similar for the five species, several stems differ extensively in line with the high rate of variation of this region of the ITS2 sequence. The 3′ part of segment m is involved in a long-distance interaction with two oligonucleotide tracts which are very short, i.e. which are not included in the set of conserved segments a–n. Interestingly, segment n is located at the tip of a long unbranched helix (stem 6) which contains the 3′ end of the 12S pre-rRNA intermediate. Stems 5 and 7, both involving conserved sequence motifs shorter than 14 nt, are supported by phylogenetic evidence. The 109 nt additional segment present in M. (P) platythrix as compared to other Mus folds into a single, long unbranched stem dramatically extending stem 5 at its apex.

Only a subset of the conserved tracts identified among the different Mus sequences can be identified within the ITS2 of the two other, more distant mammalian sequences, rat and human, which are 764 and 1155 nt, respectively (Fig. 5). Nevertheless, the overall secondary structure predicted for rat and human resembles the mouse model, with the presence of the same four basic independent domains, A–D, folded into a very similar pattern. The sequence of the 120 nt at the 5′ end of the ITS2 (conserved tracts a–c of the Mus sequences) is also well preserved not only in rat but in humans too, definitely appearing as the best conserved portion of the mammalian ITS2 sequence. Our folding proposal is further supported by the occurrence of a compensatory change in stem II in a human–rodent comparison (Fig. 5b). In rat and human domain C also folds in an unbranched helix. In human, however, this helix is considerably shortened and lacks the Y-shaped terminal structure. Over domain C, the sequence...
Figure 1. (Opposite and above) Alignment of the ITS2 sequences for the five species of the *Mus* genus. Nucleotides identical with *M. musculus*, which serves as a reference, are shown by hyphens (stars refer to gaps). Conserved segments (longer than 13 nt and exhibiting at least 80% homology as compared to the *M. musculus* sequence) are delineated by a green underline and termed α–n. Nucleotides involved in conserved base pairings are boxed by a thin line with the two strands of each stem delineated by a pair of numbered arrows in opposite orientation. The four stems identified by yellow boxes, and numbered I–IV, form the conserved basis, as depicted in Figure 2, of the four independent domains of ITS2 secondary structure which extend into more variable regions D.I–D.IV (denoted by red dotted boxes). Conserved stems numbered 1–7 are located within the more variable regions, as depicted in Figures 3 and 4.

Figure 2. A conserved core of secondary structure delineates four regions. Filled circles denote nucleotide positions that vary among the five mouse sequences. Pairings supported by compensatory changes are depicted in green. Domains D.I–D.IV, schematized by a red box (broken line) with indication of the range of size variation, and stems I–IV are as defined in Figure 1. The four regions delimited by stems I–IV are termed A–D, respectively. The 3′ end of the 8S precursor to 5.8S rRNA identified in rat (19) is shown by a blue arrow. Other symbols as in Figure 1.

Homology between rat and mouse is restricted to conserved segment e, which spans the site of preferential cleavage in vitro of rat ITS2 by protein B23 endoribonuclease, and to a short portion of conserved segment f which is also conserved in human. In contrast, conserved segment e of rodents has no obvious homolog in human, but a similar pyrimidine-rich tract is found at the same location of the ITS2 secondary structure. Domain D folds into a complex branched structure, in which the lack of...
Figure 3. Folding of domain D.III. The portion of the folded ITS2 shown here for the different Mus species corresponds precisely to the part of the aligned sequences in a red dotted box termed D.III in Figure 1. Conserved segments (as defined in Fig. 1 legend) are delineated by a thick overline. Pairings supported by compensatory changes (when comparing the corresponding sequence to the M. musculus reference) are depicted in green. Conserved stems are numbered 1–4, as in Figure 1.

Homology between the Mus, rat and human sequences makes it difficult to propose homologous structural features. This region, site of the largest size variations among mammalian ITS2, is considerably expanded in human and shortened in rat, as compared to the Mus species. There is only one tract of significant sequence homology between the Mus, rat and human sequences in this ITS2 region, which corresponds to the longest of the segments conserved among the Mus species, n. Accordingly, stem 6, which contains the 3' end of rodent 12S pre-rRNA—this is supported by the occurrence of two compensatory base changes...
Figure 4. Folding of domain D.IV. The portion of the folded ITS2 shown here for the different Mus species corresponds precisely to the part of the aligned sequences in a red dotted box termed D.IV in Figure 1. The blue arrow points on a cleavage site initially identified in M.musculus (20,21). Other symbols are as in Figure 3. The dot overlines denote sequence tracts shorter than 14 nt which are conserved among the five Mus sequences.
Figure 5. Secondary structure model for the ITS2 of rat (a) and human (b). The rat and human sequences are respectively from Subrahmanyam et al. (31) and Gonzalez et al. (34). Counterparts of sequence tracts conserved among the five Mus species are termed as in the previous figures and denoted by thick overlines. Although its sequence is not conserved between human and rodents, the human sequence termed e (dotted line) and its positional equivalent in rodents, tract e, are both pyrimidine-rich. Independent domains of secondary structure A–D correspond to regions A–D of the Mus ITS2 delineated in Figure 2. The 3′ ends of the 8S and 12S intermediates in the formation of 5.8S rRNA identified in rat and mouse, respectively, and a site of preferential cleavage of the rat ITS2 by protein B23 endonuclease (49) are also indicated (blue arrows).

in human as compared to rodents (Fig. 5b)—appears as one of the best preserved structural features of the mammalian ITS2. Remarkably, 8–12 nt immediately downstream from the 3′ end of the 12S precursor are complementary to a large portion of conserved segment f in domain C, with seven consecutive pairings more particularly preserved and two compensatory changes observed between humans and rodents (Fig. 6). It is noteworthy that the alternative base pairing interaction proposed for segment f in the secondary structure model depicted in Figure 3 is not phylogenetically supported.

Since snoRNA U8 is required for the production of 5.8S and 28S from vertebrate 32S pre-rRNA, we have systematically looked for evolutionarily conserved complementarities between the snoRNA and pre-rRNA sequences. In addition to the short hallmark motif boxes C and D, the conservation among the four available U8 sequences (Fig. 7) is essentially restricted to the 25
DISCUSSION

Preserved structural features in the ITS2 folding of mammals

The rates of divergence of the ITS2 sequence studied in this work appear well-suited for the derivation of secondary structure models by the comparative approach. Although most point mutations correspond to transitions on a single side of the base-pair (semi-conservative changes G:U versus G:C or A:U), a few true compensatory changes are observed, providing strong support for several elementary features of the model, stems III, IV, 1, 5 and 7. Moreover, the overall shape of the ITS2 folding is not altered by relatively frequent insertions–deletions preferentially located at the apex of long stems (for stems II, 5 and 7) or at opposite positions within a duplex, reminiscent of what was previously observed for the rapidly evolving domains of mature eukaryotic rRNAs (32,33), thus providing further indirect evidence in favor of the ITS2 model. Although in the early rodent model (18) the most conserved regions of the ITS2, i.e., domain A, domain B, the upper half of domain C (except for the terminal Y-shaped structure) and stem 6 in domain D, were similarly folded except for a few details, extensive differences are observed over the regions devoid of any sequence homology between rat and mouse, except for stem IV.

The new mouse model also provides a framework for a re-evaluation of ITS2 foldings for rat and human based on a sole thermodynamical basis (34). In rat and human too, the ITS2 can be folded in four regions emerging from a central loop containing three of the five most conserved sequence tracts of the spacer, pointing to the functional importance of this structural core, possibly at some stage of pre-rRNA processing. The high-content of conserved adenines at unpaired positions of this structural core is also in line with this notion, since bulged and looped adenines may represent key elements in RNA–RNA tertiary contacts (35,36) and RNA–ribosomal protein interactions (reviewed in 37–39).

ITS2 exhibit large size differences and extensive sequence variation among distant eukaryotes. However, among groups of more closely related organisms, such as trematodes (40), green algae and flowering plants (41,42) the comparative analysis of different sets of related sequences has resulted in secondary structure models which, although substantially different from each other, nevertheless exhibit several common features with the present rodent model. Thus: (i) ITS2 can be organized around a preserved central core of secondary structure from which emerge four helices and containing a high level of sequence conservation among groups of related ITS2 (18) the most conserved regions of the ITS2, i.e., domain A, B and C. (ii) Within each of these groups, the 5′ end of the ITS2 is highly conserved over ~120 nt and it can fold into two short and strongly constrained stems. (iii) Domain C can fold into a long and generally unbranched stem. (iv) The ITS2 3′ half is the most variable region and it can frequently fold into complex branched structures. However, these common structural features are not found in all eukaryotes. Thus, the phylogenetically supported, experimentally tested yeast ITS2 folds into a cross-like structure delineated by a 5′–3′ terminal stem (13,43) without any obvious similarity to the organization described above, suggesting that the function of ITS2 might substantially differ between fungi and other eukaryotic kingdoms. Analysis of other sets of related ITS2 sequences, such as X.laevis and Xenopus borealis (17), a group

Figure 6. A potential pseudoknot adjacent to the 3′ end of 12S pre-rRNA. (a) Representation of the base pairings involved in the formation of stem 6 and the pseudoknot for mouse, rat and human. The red box delineates a variable portion of stem 6 and green boxes point to compensatory changes. A blue arrow indicates the position of the 3′ end of the 12S pre-rRNA intermediate and the sequence of conserved tract f in bold-face letters (this representation is not intended to imply triple-strand interaction). (b) Schematic representation of the folding of domains C and D in mouse, showing the localization of the potential pseudoknot, with conserved segments depicted in green, and the site of the 109 nt additional sequence in M. P. platythrix shown by a filled triangle.
Potential functional implications

Except for the 3' terminal region, which merely retains a similarly biased base composition, sequences spanning pre-rRNA processing sites appear preferentially conserved among mammalian ITS2, pointing to the biological importance of 8S and 12S pre-rRNAs. Moreover, while conserved segment \( f \) does not span a processing site, it can form with the nucleotides immediately downstream from the 3' end of 12S pre-rRNA a phylogenetically supported pseudoknot which could play a key role in the recognition by the processing endonuclease, given that several RNA-binding proteins, such as rproteins S12, S15 and L23 (36,48), recognize a pseudoknoted structure in their RNA target. Recently, a new endonucleolytic cleavage catalyzed \textit{in vitro} by nucleolar protein B23 (49) has been identified in rat, located within a conserved segment of the rodent ITS2 (Fig. 5a). Protein B23 seems to recognize a three-dimensional structure rather than a ITS2 sequence (49) and the above-mentioned pseudoknot, which brings in close spatial vicinity the 3' end of the 12S RNA sequence (Fig. 6b), could represent a crucial element in the build-up of the three-dimensional structure of the ITS2 competent for pre-rRNA processing. Likewise, the secondary structure folding may also bring into relatively close proximity the 5' and 3' ends of the ITS2 and the 3' processing site of 8S pre-rRNA. Accordingly, the complex chain of nucleolytic cleavages leading to ITS2 excision could take place within a single functional pocket, which could play a major role in the tight temporal linkage of these elementary reactions. Obviously, the identification of tertiary interactions within the ITS2, through the detection of nucleotide covariations among sets of related ITS2 sequences, should provide further insight into this question.

Replacement experiments showed that functional conservation of ITS2 extends only to closely related yeast species, pointing to a high degree of concerted evolution between ITS2 and its processing machinery (16). Particularly, the presence of only one internal cleavage site within yeast ITS2 could be directly related to the above-mentioned peculiarities of yeast ITS2 folding. At the secondary structure level, the environment of the 3' end of the 7S sequence within \textit{S.cerevisiae} ITS2 appears more reminiscent of that of the mammalian 12S sequence than of the 8S intermediate (Fig. 7 and 50), contrary to what was previously suggested (19). Both of them share a similar location at the apex of a conserved stem and contain a GU at the cleavage site, suggesting they represent structural equivalents possibly recognized by a homologous enzymatic complex. Recently, Rpp1, an essential protein subunit of RNase P from \textit{S.cerevisiae}, has been shown to be required for this ITS2 cleavage in yeast (51). Taken together with these observations, the detection of Rpp30, a human homolog of Rpp1 co-purifying with human RNase P , supports the notion that this intermediate step in ITS2 processing has been conserved between fungi and metazoans.
U8: a chaperone for 32S pre-rRNA folding in vertebrates?

Accurate processing of the 3′ end of 5.8S rRNA and 5′ end of 28S rRNA in mammals has been proposed to be mediated by two 5 nt complementarities between the 5′ end of U8 and the 5′ end of 28S rRNA (24). However, no comparative support is available for the two proposed interactions (Fig. 7a), while results of their functional dissection remain ambiguous (52). It is noteworthy that as the result of the two new interactions proposed in this work, stem E20, that closes divergent domain V13 in the mature large subunit rRNA ribose methylated. Accordingly, this complementarity, like possibly the above-mentioned one, might instead mediate a chaperone function of U8 in pre-rRNA folding. Thus, the pair of sequence tracts involved in the two inter-molecular duplexes, adjacent on U8 but far apart on the pre-rRNA, could act cooperatively, as shown for U14 and U3, both required for the production of the SSU-rRNA (61,62). Like U3 and U14, U13 also contains a pair of separate, sizeable complementarities to pre-rRNA (63). However, among the rRNA complementarities in U3, U13 and U14, only one of them, in U14, functions in guiding an rRNA ribose methylation (61). Interestingly, the pairs of rRNA complementarities in U3 (62) and U8 (this study) have a similar location, at the 5′ end of the snoRNA, and both are involved in pseudoknotted structures. Moreover, while a few bona fide methylation guide snoRNAs contain two separate guide sequences, several methylation guide snoRNAs do contain another sizeable rRNA complementarity of unknown function unrelated to any rRNA methylation site (2). Accordingly, the presence of two separate rRNA complementarities appears as a widespread feature of snoRNAs involved in rRNA biogenesis, not only of box C/D snoRNAs but also of box H/ACA snoRNAs which guide the pseudouridylation of rRNAs, providing further support to the notion that they have a common function as RNA chaperones, in addition to more diversified roles in the modifications and/or cleavages of the pre-rRNA (2,58,59,64,65).

ACKNOWLEDGEMENTS

We thank Yves Henry for helpful discussions and critical reading of the manuscript. This work was financially supported by general fundings from the C.N.R.S. and Université Paul Sabatier, Toulouse, and by specific grants from the Ministère de l’Éducation Nationale, de l’Enseignement Supérieur et de la Recherche (ACC SV 07 and 13 to B.M. and ACC SV 01 to J.P.B.), from the C.N.R.S. (Action PCV to J.P.B.) and from Région Midi-Pyrénées.

REFERENCES