Nucleic Acids Research, 1999, Vol. 27, No. 23 **4533–4540**

Ribosomal internal transcribed spacer 2 (ITS2) exhibits a common core of secondary structure in vertebrates and yeast

N. Joseph, E. Krauskopf[1], M.I. Vera[1] and B. Michot*

Laboratoire de Biologie Moléculaire Eucaryote du CNRS, Université Paul Sabatier, 118 route de Narbonne, 31062 Toulouse Cedex, France and [1]Universidad Nacional Andrés Bello, Avenida República 237 Piso 2, Santiago, Chile

Received September 17, 1999; Revised and Accepted October 7, 1999

ABSTRACT

Molecular mechanisms of ITS2 processing, a eukaryotic insertion between the 5.8S and LSU rRNA, remain largely elusive even in yeast. To delineate ITS2 structural and functional features which could be common to eukaryotes, we first produced phylogenetically supported folding models in the vertebrate lineage, then tested them in deeper branchings and, more particularly, among yeasts. ITS2 comparisons between four Teleostei, a Chondrichthyes specimen and two jawless organisms have revealed a common folding architecture in four to five domains of secondary structure emerging from a preserved structural core. This folding, largely reminiscent of ITS2 architecture in mammals, is also preserved in amphibia and in chicken, despite dramatic sequence variations. Preferential conservation is located around a central loop and at the apex of a long stem in the ITS2 3′-half. Interestingly, these two independent structural features contain, respectively, the 3′-ends of the two transient rRNA precursors 8S and 12S RNA identified in mammals, suggesting a preservation of these intermediates of processing over the entire vertebrate group. Surprising similarities between the vertebrate ITS2 folding shape and that of invertebrates as well as protista have made intriguing the significant differences from the yeast model. A detailed comparative analysis including four relatively close species and *Schizosaccharomyces pombe*, a deep yeast branching, has revealed an alternative phylogenetically supported four-domain folding presenting strong similarities to the vertebrate model. Remarkably, the two best conserved regions of vertebrates have unambiguously preserved counterparts which are also sites for internal processing in yeast. Therefore, molecular mechanisms involved in ITS2 excision in vertebrates and yeast might be more closely related than currently believed and might require a very similar *trans*-acting machinery.

INTRODUCTION

In most eukaryotes, the mature rRNA sequences (SSU rRNA, 5.8S RNA and LSU rRNA) are produced in the nucleolus from a single large precursor (pre-rRNA) containing spacer regions that are sequentially removed through a series of rapid and intricate processing steps. These events take place only when a fully transcribed precursor molecule is assembled into a huge RNP particle. The mechanisms leading to mature rRNA have revealed a tight coupling between the activity of this processing machinery and ribosome assembly. In addition, yeast *in vivo* analyses have demonstrated complex interdependencies in the processing of the spacers that are indeed cleaved in a concerted fashion (1–6). This reflects the need for higher order structure in the pre-ribosomal particle that may be necessary to organize the processing sites in close spatial proximity. Yeast experiments have identified several *cis*-acting elements which are required for spacer excision, while numerous essential *trans*-acting factors have now been characterized in both yeast and vertebrates, including endo- and exoribonucleases (reviewed in 7) and several snoRNAs (8). Nevertheless, many features of ribosome biogenesis remain poorly understood. Not only is little known about the dynamics of pre-rRNA folding, but even structural features of the spacers essential for rRNA processing and ribosome biogenesis remain to be identified, especially in vertebrates. Furthermore, the role of spacer regions is largely speculative, despite data from yeast mutation/deletion experiments pointing out the crucial significance of the integrity of several structural features, both for correct and efficient spacer removal and for biogenesis of active ribosomal subunits (9). Thus, a complete and accurate description of the secondary structure folding of each spacer is an essential step for an effective integration of the very diverse information concerning ribosome biogenesis in order to gain better insight into the molecular mechanisms underlying processing events.

During the pre-rRNA cleavage pathway, the excision of ITS2, a eukaryote-specific insertion, remains the most elusive processing step, even in yeast. Nevertheless, several critical *cis*-acting processing elements were identified in *Saccharomyces cerevisiae* in those regions of ITS2 that show the highest degree of sequence conservation with other yeast species (10). A recent comparative analysis including ITS2 sequences of human, rat and five species of mice (11) has demonstrated that

*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

despite dramatic sequence divergence (12,13), all mammalian organisms share a similar folding pattern, supported by phylogenetic evidence. In these species, ITS2 is organized in four main domains of secondary structure emerging from a preserved central core. Remarkably, two internal cleavages identified in rodents and leading to an 8S (14) and a 12S (15,16) intermediate of maturation are located in the best conserved structural features, a single-stranded segment of the central core and the apical part of a stem in the 3′-half of ITS2, respectively. Definite secondary structure constraints have also been detected in trematodes (invertebrates), a deep metazoan branching (17,18), and in a wide set of plants, including green algae and flowering plants (19,20). Surprisingly, despite considerable size variations between mammals (~1000 nt), trematodes (~350 nt) and plants (mean size 250 nt) and the complete lack of sequence homology between these three groups of species, their ITS2 present very similar folding patterns.

However, the entire eukaryotic kingdom does not share these common characteristics. In fact, yeast ITS2 folding does not exhibit any obvious similarity to the organization described above. In *S.cerevisiae*, ITS2 secondary structure has been predicted from a conjunction of phylogenetic analysis, minimum energy modelling, chemical and enzymatic probing and genetic experiments (21,22). The models produced propose folding into a cross-like structure showing a long irregular duplex between the 3′-end and a sequence close to the 5′-end (10,21), which is absent in mammals, plants and trematodes. Therefore, it is likely that two different solutions were retained during evolution to fold this spacer. This diversification in the structural constraints acting on ITS2 in yeast as compared with mammals, trematodes and plants was suggested by substantial differences in the nucleolytic mechanisms involved for its correct and efficient removal from the pre-rRNA (11). Thus, only one internal cleavage was identified *in vivo* in yeast leading to a 7S intermediate (23), which structurally is closely related to the 12S site of mammals (11). In the same way, no yeast homologue is known for the U8 snoRNA, which is required for cleavages at the 5′-end of 5.8S and 3′-end of 28S RNA in *Xenopus* (24).

In order to delineate the structural and functional features that could be preferentially preserved among all eukaryotic ITS2, we have first produced phylogenetically supported folding models for seven fish sequences available. They include four Teleostei, one Chondrichthyes and two jawless organisms covering >530 000 000 years of vertebrate evolution. We then identified structural features conserved among vertebrates through comparative analysis at the secondary structure level in fish, two amphibians, a bird and mammals. Finally, we revisited the yeast model according to the structural features that have been phylogenetically maintained during evolution not only among vertebrates, but also in invertebrates and plants.

MATERIALS AND METHODS

Sequences analysed

Fish sequences used in this work were extracted from the EMBL database: *Cyprinus carpio*, U87963 (25); *Salvelinus namaycush*, S74137 (26); *Rexea solandri*, U88398 (27); *Misgurnus fossilis*, X07623 (28); *Eptatretus stouti*, AF061797; *Petromyzon marinus*, AF061798; *Hydrolagus colliei*, AF061799. The two amphibian ITS2 sequences, *Xenopus laevis* (X59734) and *Xenopus borealis* (X59733), are from Furlong and Maden (29), while the only ITS2 sequence among birds, *Gallus gallus*, was kindly provided by Crouch *et al.* (30). Yeast ITS2 sequences presented in this work are those previously used by Yeh and Lee (10,21,22) and that of *S.pombe* (31). ITS2 foldings were also produced for a broad sampling of eukaryotic sequences including: the Protista specimens *Prorocentrum micans* (32), *Tetrahymena thermophila* (33), and *Dictyostellium discoideum* (34); a nematode, *Caenorhabditis elegans* (35); a sea urchin (Echinodermata), *Lythechinus variegatus* (36); the fungi *Neurospora crassa* (37) and *Thermomyces lanuginosus* (38) and several representative species within the genus *Candida* (39). Foldings not presented here are available upon request from the authors.

Sequence alignment and secondary structure prediction

An iterative method involving both the thermodynamic and comparative approaches was applied to derive the best phylogenetically supported folding models. An alignment initially performed using the program Multalin (40) was progressively optimized according to secondary structure homology (41). Optimal and suboptimal foldings were predicted for each sequence on a thermodynamic basis (42) using the software package from the Genetics Computer Group, University of Wisconsin (43). Proposed foldings were compared and solution maximizing structural homologies were retained and checked for the presence of compensatory changes. Then the alignments were refined and new foldings predicted according to these results. Secondary structure drawings were produced with ESSA (44).

RESULTS

While ITS2 presents a dramatic range of length variation among vertebrates (from 262 to 1155 nt, respectively, in *X.laevis* and human), its size remains relatively homogeneous within each major group of species (Fig. 1). It is ~300 nt in amphibians, ranges from 309 to 486 nt in fish and is considerably expanded in mammals with ~1000 nt, except in the rat, in which it is substantially shorter (765 nt). This high level of sequence variation, which is a common feature of ribosomal spacers, implies that detectable conserved segments can be identified only among closely related species. Therefore the four Teleostei, which have diverged over ~90 000 000 to ~150 000 000 years, might present a rate of variation of ITS2 sequences suitable for comparative analysis designed to evaluate sequence/structural conservation and derive phylogenetically supported secondary structure models. The availability of a Chondrichthyes specimen gave us the opportunity to test the Teleostei model in the whole Gnathostomata phylum, while the lamprey and hagfish sequences have extended it to jawless fish over a vertebrate evolutionary time of >530 000 000 years.

Five dispersed but unambiguously conserved sequence segments (named *a–e*) encompassing about a third of the ITS2 length have been identified in Teleostei. They are interspersed with more variable regions where size variations accumulate (Fig. 2). Surprisingly, *H.colliei* showed a similar level of sequence conservation despite having diverged from Teleostei

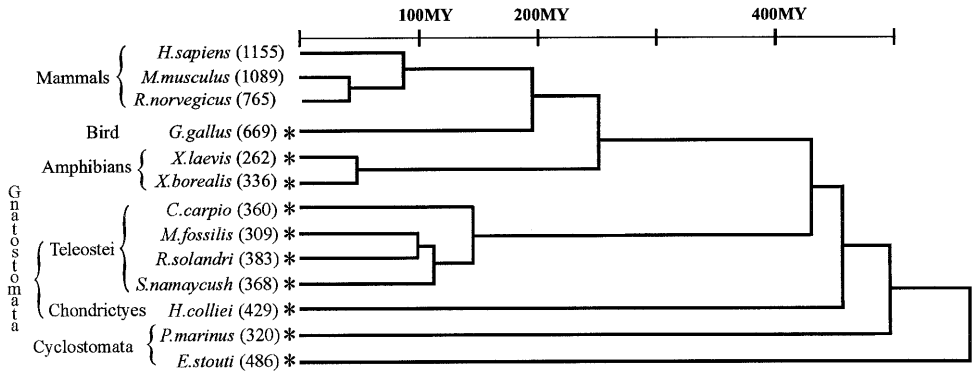

Figure 1. Phylogenetic relationships among vertebrate species. This tree was drawn from Forey and Janvier (45), Janvier (46) and Rasmussen *et al*. (47). ITS2 length is given in parentheses for each species. Stars refer to species used in the present work.

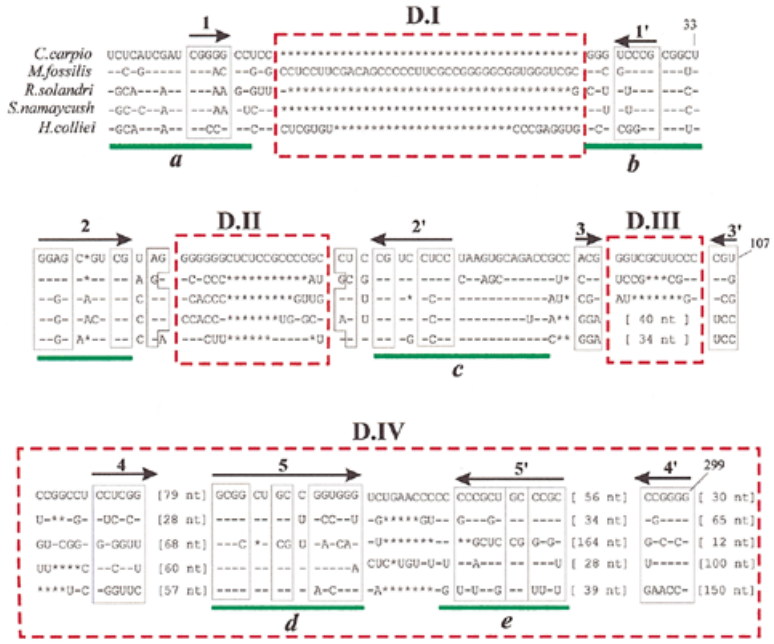

Figure 2. Alignment between Gnathostomata ITS2 regions. Nucleotides identical with *C.carpio*, which serves as a reference, are replaced by hyphens while stars refer to gaps. The best conserved segments are underlined and named *a*–*e*. Nucleotides involved in conserved base pairings are boxed by a thin line with the two strands of each stem delineated by arrows in opposite orientation and identified as 1–5. The dotted boxes (D-I–D-IV) locate four variable domains.

more than 400 000 000 years ago. Thus, the five sequence segments that are preferentially preserved among Teleostei have a clear equivalent in the Chondrichthyes specimen. On the basis of this alignment, fish ITS2 can be divided into two main regions. The first is particularly well preserved and extends over the ~85–110 nt of the 5′-end. It spans the three conserved segments *a*–*c* that are separated by two sites of insertion/deletion, termed D-I and D-II, respectively. Their size variation is relatively limited, except for *M.fossilis*, which shows the longest D-I domain (40 nt), while it displays the shortest ITS2 among fish. The second region is longer and extends downstream from segment *c*, encompassing variable regions D-III and D-IV. Although this region is the most variable, with large size variation, we have detected a 30–40 nt long sequence that contains two short conserved tracts named *d* and *e*. At the secondary structure level, we have identified five stems supported by compensatory changes. Three of them (stems 1, 2 and 5) involve Watson–Crick base pairings between conserved segments and delineate three helical structures. In contrast, stems 3 and 4, which involve more rapidly variable sequences, were identified only through a comparison of folding models.

In spite of important changes in sequence, the Teleostei species share a very similar pattern of secondary structure

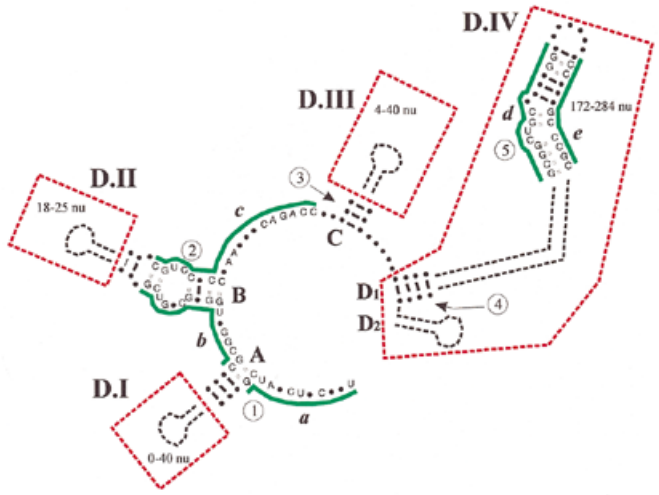

Figure 3. Secondary structure comparison in Gnathostomata. Filled circles denote nucleotides that differ among the five fish sequences (including one nucleotide insertion/deletion). Thick bars point to pairings supported by compensatory changes. The variation range of domains D-I to D-IV (boxed by a broken line and defined as in Fig. 2) is given. Other symbols have the same meaning as in Figure 2.

where homologous sequence segments have homologous locations (Fig. 3 and Supplementary Material, Fig. S1). This folding is organized into four or five independent domains of secondary structure, termed A–D_2. The basal parts of domains A–D_1 correspond, respectively, to the four conserved stems 1–4. The two short domains, A and B, are located within the best conserved 5′-end, domain C encloses a more variable helix immediately downstream of conserved segment *c*, while domain D_1 is considerably longer and presents significant structural variations between the four Teleostei. It encompasses most of the highly variable 3′-half of ITS2 and can fold into a giant stem with lateral branches in the three longest ITS2 (*C.carpio*, *R.solandri* and *S.namaycush*). Its tip part, named stem 5, is remarkably preserved and indeed corresponds to one of the best conserved structural features. An additional short helix (D_2) can be proposed downstream of D_1 in the 3′-end of ITS2 of three of the species (*C.carpio*, *M.fossilis* and *S.namaycush*). *Hydrolagus colliei* ITS2, in line with its common pattern of sequence conservation with Teleostei, can fold into a secondary structure largely reminiscent of these four species (see Supplementary Material, Fig. S1). Thus, a very similar and strong structural constraint has been exerted on the mode of variation and on the folding of ITS2 in fish over an evolutionary time covering ~450 000 000 years. In the five Gnathostomata, the four stems 1–4 organize a common core of secondary structure, which brings the three conserved segments (*a*–*c*) and the two ends of ITS2 into close proximity, while *d* and *e* are paired at the apical part of a long and variable helix. Compensated base changes observed within these stems give direct phylogenetic evidence supporting this secondary structure model, whereas a preferential localization of insertion/deletion at the apex of conserved stems also preserves, although more indirectly, the overall folding shape. This model largely improves the only one previously proposed for *M.fossilis* (28) in which the sole structural feature in common with our model was the terminal hairpin loop of stem 5.

Remarkably, the ITS2 of *E.stouti* and *P.marinus*, which do not present any significant sequence homology with the Gnathostomata, can nevertheless fold very similarly (see supplementary material, Fig. S2). A comparison at the secondary structure level among fish allowed detection of homologues of stems 1 and 2 and the identification of a set of invariant positions in sequences corresponding to segments *a*–*c*. In these two species, the 3′-terminal region can fold into a hairpin structure that shows remarkable thermodynamic stability with 17 consecutive Watson–Crick base pairs in *P.marinus* and 11 bp interrupted by only one mismatch in *E.stouti*. This stem is the probable counterpart of stem D_2 in the Gnathostomata. In contrast, we failed to identify an unambiguous homologue of stem 5 although the long central region between *c* and D_2 (195 and 270 nt in lamprey and hagfish, respectively) can fold into a long branched helix reminiscent of domain D_1 of the Gnathostomata. In addition, jawless fish are distinguished by the apparent lack of the short and variable domain C, which is a common feature of the five Gnathostomata. The determination of new sequences more closely related to *E.stouti* on the one hand and *P.marinus* on the other should help to refine the D_1 folding for which thermodynamically very close alternative structures can be proposed and to identify a potential homologue of stem 5.

The folding characteristics of fish ITS2 present numerous features in common with the mammalian model (11). These features should correspond to ancestral characters and may therefore be preserved in other vertebrate groups. Indeed, in amphibians too, five conserved sequence tracts, previously identified by sequence alignment between *X.laevis* and *X.borealis* (29), can adopt a secondary structure folding identical to their homologue in fish and mammals (see supplementary material, S3a). Segments *a*–*c* guide the folding of domains A and B in the central core, while *d* and *e* form the tip of a long and thermodynamically stable unbranched helix. Remarkably, *G.gallus* ITS2 can also fold according to the secondary structure shape of other vertebrates. Segments *a*–*c* are the sole sequences unambiguously common for amphibians and mammals, while some similitude between the tip part of the long stem D_1 and stem 5 of mammals suggests a probable conservation of this structural element also in the bird lineage (see supplementary material, Fig. S3b). Interestingly, both amphibians and birds lack domain C, as do jawless fish, as opposed to other fish and mammals, suggesting that insertion/deletion events have occurred independently several times during vertebrate evolution.

A detailed comparison of structural features which are conserved among vertebrate groups emphasizes the spectacular conservation of the 5′-half of the central core (in the ITS2 5′-part) and of stem 5 (Fig. 4a and b, respectively). Phylogenetic conservation of stem 1 is essentially restricted to the four basal consecutive base pairs, while it is more extended on stem 2 and encompasses an internal loop and a mismatch (Fig. 4a and c). Regarding stem 5, the presence of short sequence tracts common to each set points to the ancestral character of this stem, despite some intergroup folding variation (Fig. 4b and c). Strikingly, the 3′-ends of the 8S and 12S pre-rRNA intermediates of maturation identified in mammals are located precisely in these best preserved structural features. They lie in the

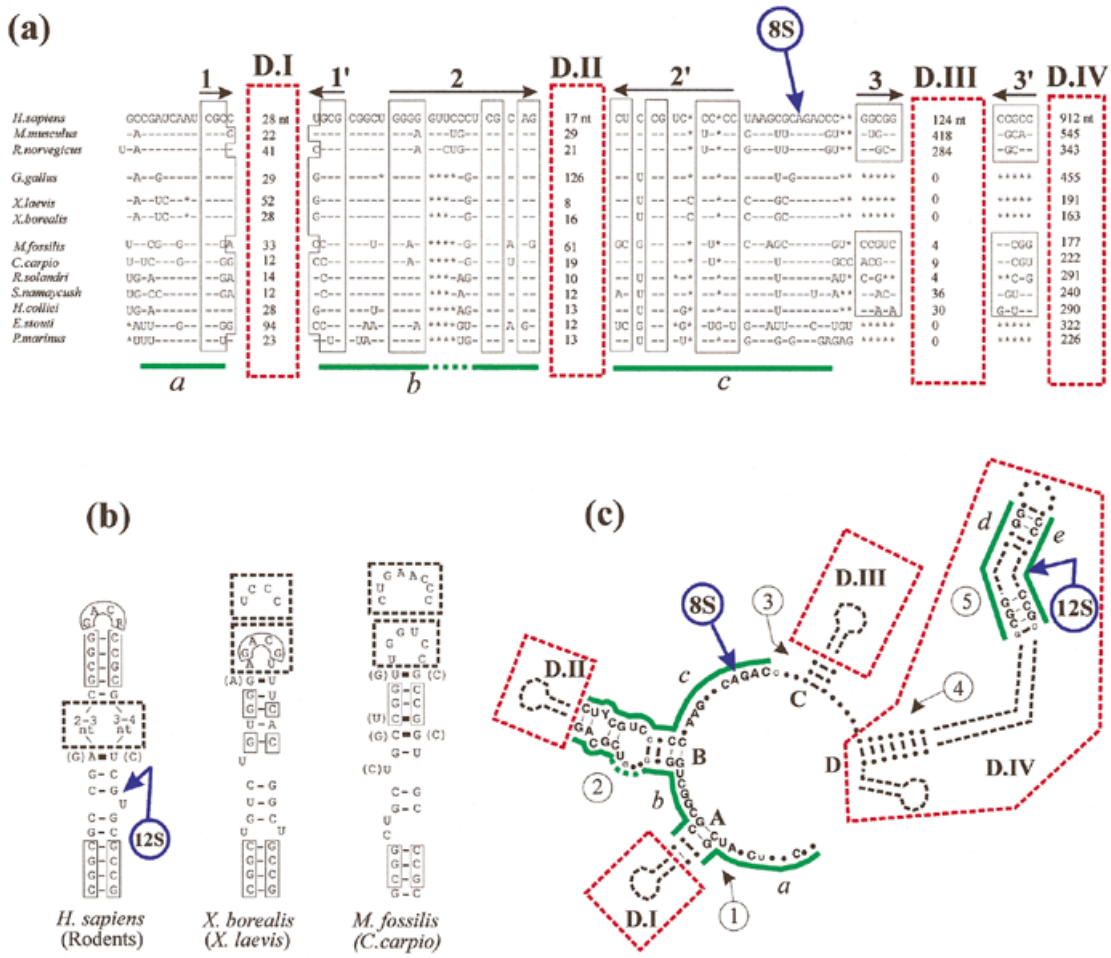

Figure 4. ITS2 conserved structural features over the whole vertebrate group. (**a**) Alignment in the ITS2 5′-region. Hyphens represent nucleotides identical to *Homo sapiens*, while stars indicate gaps. The best conserved segments *a–c* are underlined (green lines). An arrow locates the 8S RNA internal cleavage site previously identified in rodents (14). D-I to D-IV identify four variable domains. (**b**) Comparison of stem 5. In each of the three sets of comparison, *H.sapiens*, *X.borealis* and *M.fossilis*, respectively, serve as reference, changes in species that are compared are given in parentheses and variable parts of the stem are identified by dotted boxes. Thin lined boxes show conserved nucleotides and thick bars point to compensated changes. An arrow indicates the position of the 12S RNA cleavage identified in rodents. (**c**) Secondary structure consensus for vertebrate ITS2.

CAGAC motif of the central loop immediately downstream of stem 2 and in stem 5 (Fig. 4a and c), respectively. The preferential conservation of these functional sites suggests their key role in ribosome biogenesis. In contrast, the more versatile constraints acting on the folding of rapidly variable regions rather favor their involvement as structural domains, the role of which could be to preserve the functional 3-dimensional organization of conserved features (Fig. 4c).

In an attempt to test and extend our model to deeper branchings represented by a single organism, we have imposed folding characteristics identified in vertebrates as a prerequisite. ITS2 of the two other metazoan sequences available, *C.elegans* and *L.variegatus*, can exhibit the typical folding shape as well as unambiguous equivalents of segments *a–c* and of stems 1 and 2 with a CAGAC motif at a position homologous to the mammalian 8S RNA site. Similarly, three protista, *P.micans*, *T.thermophila* and *D.discoideum*, also show a four-domain architecture largely reminiscent of the vertebrate model, despite a complete lack of sequence homology (foldings not shown; see Materials and Methods).

The general ITS2 folding model which is now emerging among eukaryotes renders the significant differences from the *S.cerevisiae* model (9,10,21,22) particularly intriguing. This has prompted us to re-examine it using an evolutionary approach including *S.pombe*, a deep yeast branching, in addition to the four relatively close species initially used (10,22). Remarkably, we have identified an alternative phylogenetically supported folding which is largely similar to our general model (Fig. 5). Changes result from replacement of the long-range irregular helix between the two ends of ITS2 (domain III in the Yeh and Lee model) with new stems located on each of its two strands (Fig. 5b). The helix proposed in the 5′-region is the unambiguous homologue of vertebrate domain B, as revealed by an identical location 5 nt downstream of domain A (domain

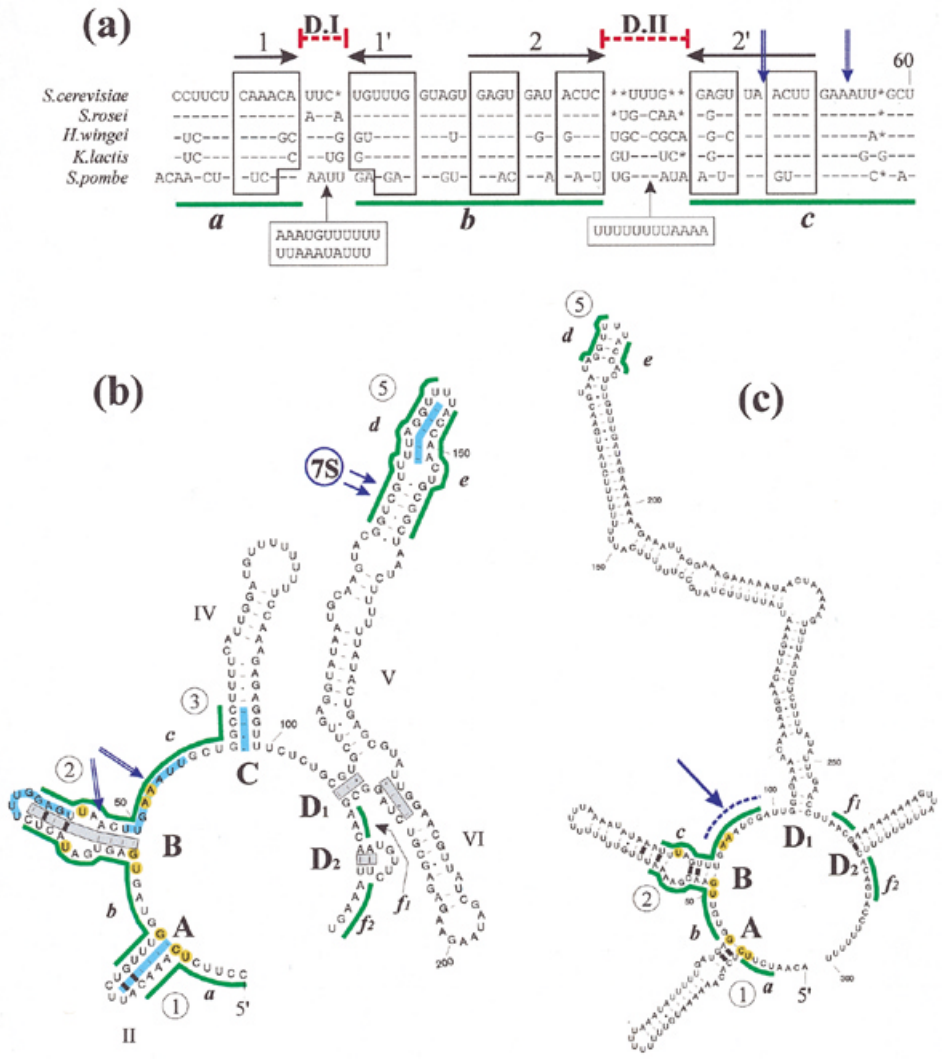

Figure 5. ITS2 conserved structural features among yeast species. (**a**) Sequence alignment over the 5′-region of yeast ITS2. Symbols have the same meaning as in Figure 2. (**b**) ITS2 secondary structure folding in *S.cerevisiae*. The two adjacent single arrows locate an *in vivo* internal cleavage site (7S), whereas the two double arrows point to potential cleavages identified by Chamberlain *et al.* (50). Sequences conserved between *S.cerevisiae*, *Saccharomyces rosei*, *Hansenula wingei* and *Kluyveromyces lactis* are overlined and identified as *a*–f_2. Correspondences with the Yeh and Lee model (21) are given in Roman numerals, while grey stems are new ones predicted in this work. Stems and nucleotides coloured blue show the stems and nucleotides essential for ITS2 processing (10). (**c**) Secondary structure folding of *S.pombe*. Green lines point to sequences which are conserved between the five yeast species compared. The blue arrow and blue dotted line locate a potential cleavage identified in *S.pombe* ~101 bases downstream of the 3′-end of 5.8S rRNA and 210 bases upstream of the 5′-end of 25S rRNA (48). In (b) and (c) thick bars reveal compensated changes either between *S.cerevisiae*, *S.rosei*, *H.wingei* and *K.lactis* in (b) or between these four species and *S.pombe* in (c). Nucleotides conserved with vertebrates are written in yellow circles.

II in Lee model) and strong structural similarities. Its basal part, which is largely supported by compensatory base changes (Fig. 5a), can indeed fold into a stem–loop structure identical to vertebrate stem 2. A prediction based on the thermodynamic approach shows that *S.pombe* can adopt a very close folding pattern (Fig. 5c; 48). A comparison with other yeast species gives new phylogenetic support for stems 1 and 2, while an equivalent of stem 5 can be proposed on the basis of a short sequence conservation. In contrast, domain C, the basal stem of which (stem 3) was demonstrated as essential in *S.cerevisiae* for 26S RNA production (10), was absent. Nevertheless, alternative folding solutions suggest a possible equivalent for domain C in *S.pombe* too that could be tested with additional related species. A comparison at the secondary structure level between the vertebrate and yeast ITS2 folding models has pointed to nine conserved nucleotides with identical positions in the vicinity of stems 1 and 2 (Fig. 5b and c). With respect to the 3′-end of ITS2, a stem boxed by two short conserved sequence tracts (*f1* and *f2*) could correspond to vertebrate domain D_2. It should be noted that the new pairings proposed here are fully compatible with chemical and enzymatic probing performed by Yeh and Lee (21). Finally, additional comparative proofs supporting our model were obtained by folding the ITS2 of several other fungi (not shown; see Materials and Methods).

DISCUSSION

The present study demonstrates an unexpected preservation of ITS2 folding architecture not only among distant vertebrates, but more surprisingly between yeast and vertebrates. Moreover, our detailed comparative analysis reveals a preferential conservation of the two regions containing internal processing sites. These observations might reflect a common function(s) in ribosome biogenesis. In particular, they point to maturation pathways and molecular mechanisms of ITS2 excision, which may be more largely similar than previously believed. Thus, the two transient pre-rRNA species, corresponding on the one hand to the 8S species and on the other hand to the 12S species of mammals, could be essential intermediates in the production of functional mature 5.8S and LSU rRNA in both vertebrates and yeast. However, only one internal processing site was identified *in vivo* in *S.cerevisiae* and *S.pombe* (7,49), two yeast species with the best known processing pathways. Surprisingly, our ITS2 folding models predict that the internal cleavage occurring in each of these two species is not homologous, in contrast to what is currently predicted. In *S.cerevisiae* the 7S pre-rRNA 3′-end occurs in stem 5 and corresponds to the 12S pre-rRNA of mammals (11), while in *S.pombe* the sole internal cut identified lies ~100 bases downstream of the mature 3′-terminus of 5.8S rRNA (49). Its location immediately downstream of stem 2 indicates that it corresponds to the 8S pre-rRNA species of mammals (Fig. 5c). Therefore, if only one cut is required in yeast with the probable essential function of providing an entry for exonucleases, our observation would suggest that either of these two potential processing sites could be a functional target for endonuclease depending on the species. The presence of two internal cleavage sites in mammals could increase the processing efficiency by excising a long internal segment and thus reducing the length of the ITS2 sequence, which must be degraded by a trimming reaction to produce the 3′- and 5′-ends of 5.8S and 28S rRNA, respectively. Nevertheless, there is some indication of a possible mammal-like 8S pre-rRNA processing intermediate in *S.cerevisiae*. RNase P indeed cuts *S.cerevisiae* pre-rRNA *in vitro* precisely in the single-stranded region where the mammalian 8S RNA 3′-end lies (50) and the *in vivo* transient terminus of *S.pombe* (Figs 4c and 5c). Conformational rearrangements induced by this cleavage *in vivo* could make the 3′-end of this *S.cerevisiae* pre-rRNA species immediately accessible to exonucleolytic activity leading to rapid production of the mature 3′-end of 5.8S rRNA. A similar hypothesis could also explain the lack of observable *in vivo* cleavage corresponding to an equivalent of the mammalian 12S cut in stem 5 of *S.pombe* ITS2.

However that may be, the strong conservation at the two potentially cleaved internal processing sites may be related to the positioning of the enzymatic complex within the ITS2 conserved core through direct interactions. Recently, *S.pombe* stem 5 has been demonstrated to be a site for specific interactions with soluble factors (48). Stem 2, identified in this work, seems also to be a crucial *cis*-acting element required for production of the 5.8S RNA 3′-end. Two sets of clustered substitutions located in stem 2 and in the single-stranded segment between stems 2 and 3 (Fig. 5) severely decrease the level of mature 26S rRNA, but do not affect production of the 5′-terminus of 26S RNA (10). Interestingly, additional strong phylogenetic support for a key function of stem 2 and adjacent conserved sequences is given by three species of the genus *Candida*. In these organisms, which exhibit dramatically shortened ITS2 sequences (66–89 nt), stem 2 and the 3′-adjacent GAAAU motif seem to be the sole conserved structural features (not shown).

In parallel with this impressive conservation of ITS2 structural features among eukaryotes, a line of experimental evidence suggests a requirement for common *trans*-acting factors in ITS2 processing of yeast and vertebrates (reviewed in 7). Eleven essential proteins constitute a highly conserved eukaryotic RNA processing complex called an exosome containing multiple exoribonuclease activities (51). Stems A and D_2, which are close to the ITS2 ends, could be a part of their stop signals. In contrast, endonucleases required for initial internal cleavages have not yet been identified. Nevertheless, some indications suggest their possible conservation between yeast and mammals. Rpp30 is the human homologue of yeast Rpp1, which is an essential protein subunit of RNase P, possibly required for 7S RNA cleavage in *S.cerevisiae* (52). Moreover, in HeLa cell extracts RNase P is associated with RNase MRP (53), an enzyme that cleaves the yeast pre-rRNA at site A3 within ITS1 (54). Finally, both enzymes localize in the nucleolus in *S.cerevisiae* and mammals (55–57). Therefore, a common basic enzymatic machinery organized around the RNase P–RNase MRP complex, finely tuned to each cleavage site by specific components, could be used in both ITS1 (which requires two internal cleavages) and ITS2 excision and could be largely conserved between yeast and vertebrates. Indirect support was recently given by similarity between structural features of ITS1 and ITS2 in *S.pombe* rRNA (48).

Thus, a complex network of interactions might take place to assemble the pre-rRNA structural features directly involved in processing steps into a probable, relatively compact structure. This complex organization, which allows fine control of the synthesis of SSU and LSU rRNA molecules, also requires structural features belonging to mature rRNA. Directed mutagenesis and expression of mutated pre-rRNAs in yeast have demonstrated an interaction between the 5.8S RNA 3′-end and 28S RNA 5′-end as an essential requirement common to yeast and vertebrate pre-rRNA processing (58). Similarly, stem E20, which closes divergent domain V13 ~2000 nt downstream from the 25S RNA 5′-end, is essential for yeast ITS2 processing (59). In mammals and *Xenopus*, U8 snoRNA is an essential *trans*-acting factor for ITS2 processing. A chaperone role has been proposed for this snoRNA through two long and conserved complementarities involving two adjacent segments of U8 and the ITS2 5′-end and a sequence located immediately upstream of stem E20, respectively (11). In yeast, no equivalent for U8 snoRNA has been identified so far, leading to the conclusion that the processing steps required for ITS2 excision are probably not mediated by a snoRNA. Nevertheless, in yeast, mutations in Nop1p and Nop56p, which are components of the box C-D snoRNAs, surprisingly inhibit the synthesis of 25S rRNA (60,61). In addition, the 9 nt which are located upstream of stem E20 and are involved in the complementarity between U8 and 28S rRNA in vertebrates are perfectly conserved among eukaryotes. These two observations, taken together with the high degree of common structural and enzymatic requirements in yeast and vertebrates, make the apparent lack of an equivalent for U8 snoRNA in yeast particularly stressing. A yeast homologue of U8 snoRNA could be identified in the

S.cerevisiae genome by applying the U8/32S interaction model predicted for vertebrates (11), while the combined use of a yeast and *Xenopus in vivo* analysis system might provide an effective means of addressing the possible common structural features required in yeast and/or vertebrates, respectively.

SUPPLEMENTARY MATERIAL

See Supplementary Material available at NAR Online.

ACKNOWLEDGEMENTS

We are grateful to J. P. Bachellerie for his support, P. Janvier for helpful discussions on fish phylogeny and Y. Henry for critical reading of the manuscript. This research was supported in part by grants to B.M. (ACCSV-07 and -13) from the Ministère de l'Education Nationale de l'Enseignement Supérieur et de la Recherche, and to M.I.V. from the Chilean Commission for Science and Technology (no. 1970633, Fondecyt_Chile).

REFERENCES

1. Melekhovets,Y.F., Good,L., Abou Elela,S. and Nazar,R.N. (1994) *J. Mol. Biol.*, **239**, 170–180.
2. van Nues,R.W., Rientjes,J.M.J., van der Sande,C.A.F.M, Zerp,S.F., Sluiter,C., Venema,J., Planta,R.J. and Raué,H.A. (1994) *Nucleic Acids Res.*, **22**, 912–919.
3. Good,L., Intine,R.V.A. and Nazar,R.N. (1997) *J. Mol. Biol.*, **273**, 782–788.
4. Hitchen,J., Ivakine,E., Melekhovets,Y.F., Lalev,A. and Nazar,R.N. (1997) *J. Mol. Biol.*, **274**, 481–490.
5. Liang,W.Q. and Fournier,M.J. (1997) *Proc. Natl Acad. Sci. USA*, **94**, 2864–2868.
6. Allmang,C. and Tollervey,D. (1998) *J. Mol. Biol.*, **278**, 67–78.
7. Venema,J. and Tollervey,D. (1999) *Annu. Rev. Genet.*, **33**, 261–311.
8. Smith,C.M. and Steitz,J.A. (1997) *Cell*, **89**, 669–672.
9. van Nues,R.W., Venema,J., Rientjes,M.J., Dirks-Mulder,A. and Raué,H.A. (1995) *Biochem. Cell Biol.*, **73**, 789–801.
10. van Nues,R.W., Rientjes,J.M.J., Morré,S.A., Mollee,E., Planta,R.J., Venema,J. and Raué,H.A. (1995) *J. Mol. Biol.*, **250**, 24–36.
11. Michot,B., Joseph,N., Mazan,S. and Bachellerie,J.P. (1999) *Nucleic Acids Res.*, **27**, 2271–2282.
12. Michot,B., Bachellerie,J.P. and Raynal,F. (1983) *Nucleic Acids Res.*, **11**, 3375–3391.
13. Goldman,W.E., Goldberg,G., Bowman,L.H., Steinmetz,D. and Schlessinger,D. (1983) *Mol. Cell. Biol.*, **3**, 1488–1500.
14. Reddy,R., Rothblum,L.I., Subrahmanyam,C.S., Liu,M.H., Henning,D., Cassidy,B. and Busch,H. (1983) *J. Biol. Chem.*, **258**, 584–589.
15. Bowman,L.H., Rabin,B. and Schlessinger,D. (1981) *Nucleic Acids Res.*, **9**, 4951–4966.
16. Bowman,L.H., Goldman,W.E., Goldberg,G.I., Hebert,M.B. and Schlessinger,D. (1983) *Mol. Cell. Biol.*, **3**, 1501–1510.
17. Michot,B., Despres,L., Bonhomme,F. and Bachellerie,J.P. (1993) *FEBS Lett.*, **3**, 247–252.
18. Morgan,J.A.T. and Blair,D. (1998) *J. Mol. Evol.*, **47**, 406–419.
19. Hershkovitz,M.A. and Zimmer,E.A. (1996) *Nucleic Acids Res.*, **24**, 2857–2867.
20. Mai,J.C. and Coleman,A.W. (1997) *J. Mol. Evol.*, **44**, 258–271.
21. Yeh,L.C.C. and Lee,J.C. (1990) *J. Mol. Biol.*, **211**, 699–712.
22. van der Sande,C.A.F.M., Kwa,M., van Nues,R.W. van Heerikhuizen,H., Raué,H.A. and Planta,R.J. (1992) *J. Mol. Biol.*, **223**, 899–910.
23. Veldman,G.M., Klootwijk,J., van Heerikhuizen,H. and Planta,R.J. (1981) *Nucleic Acids Res.*, **9**, 4847–4862.
24. Peculis,B.A. and Steitz,J.A. (1993) *Cell*, **73**, 1233–1245.
25. Vera,M.I., Rios,H.M., de la Fuente,E., Figueroa,J. and Krauskopf,M. (1997) *Comp. Biochem. Physiol.*, **118B**, 777–781.
26. Zhuo,L., Sajdak,S.L. and Phillips,R.B. (1994) *Genome*, **37**, 664–671.
27. Colgan,D.J. and Paxton,J.R. (1997) *Mar. Freshwater Res.*, **48**, 103–118.
28. Kupriyanova,N.S. and Timofeeva,M.Y. (1988) *Mol. Biol. Rep.*, **13**, 91–96.
29. Furlong,J.C. and Maden,B.E.H. (1983) *EMBO J.*, **2**, 443–448.
30. Crouch,R.J., Kanaya,S. and Earl,P.L. (1983) *Mol. Biol. Rep.*, **9**, 75–78.
31. Schaak,J., Mao,J.-I. and Söll,D. (1982) *Nucleic Acids Res.*, **10**, 2851–2864.
32. Maroteaux,L., Herzog,M. and Soyer-Gobillard,M.O. (1985) *Biosystems*, **18**, 307–319.
33. Engberg,J. and Nielsen,H. (1990) *Nucleic Acids Res.*, **11**, 6915–6919.
34. Ozaki,T., Hoshikawa,Y., Iida,Y. and Iwabuchi,M. (1984) *Nucleic Acids Res.*, **25**, 4171–4184.
35. Ellis,R.E., Sulston,J.E. and Coulson,A.R. (1986) *Nucleic Acids Res.*, **14**, 2345–2364.
36. Hindenach,B.R. and Stafford,D.W. (1984) *Nucleic Acids Res.*, **12**, 1737–1747.
37. Chambers,C., Dutta,S.K. and Crouch,R.J. (1986) *Gene*, **44**, 159–164.
38. Nazar,R.N., Wong,W.M. and Abrahamson,J.L.A. (1987) *J. Biol. Chem.*, **262**, 7523–7527.
39. Lott,T.J., Burns,B.M., Zancope-Oliviera,R., Elie,C.M. and Reiss,E. (1998) *Curr. Microbiol.*, **36**, 63–69.
40. Corpet,F. (1989) *Nucleic Acids Res.*, **16**, 10881–10890.
41. Corpet,F. and Michot,B. (1994) *CABIOS*, **10**, 389–399.
42. Zuker,M. (1989) *Science*, **244**, 48–52.
43. Devereux,J., Haeberli,P. and Smithies,O. (1984) *Nucleic Acids Res.*, **12**, 387–395.
44. Chetouani,F., Monestié,P., Thébault,P., Gaspin,C. and Michot,B. (1997) *Nucleic Acids Res.*, **25**, 3514–3522.
45. Forey,P. and Janvier,P. (1993) *Nature*, **361**, 129–134.
46. Janvier,P. (1998) *J. Mol. Evol.*, **46**, 375–377.
47. Rasmussen,A.-S., Janke,A. and Arnason,U. (1998) *J. Mol. Evol.*, **46**, 382–388.
48. Lalev,I. and Nazar,R.N. (1999) *Nucleic Acids Res.*, **27**, 3071–3078.
49. Good,L., Intine,R.V.A. and Nazar,R.N. (1997) *Eur. J. Biochem.*, **247**, 314–321.
50. Chamberlain,J.R., Pagán-Ramos,E., Kindelberger,D.W. and Engelke,D.R. (1996) *Nucleic Acids Res.*, **24**, 3158–3166.
51. Mitchell,P., Petfalski,E., Shevchenko,A., Mann,M. and Tollervey,D. (1997) *Cell*, **91**, 457–466.
52. Stolc,V. and Altman,S. (1997) *Genes Dev.*, **11**, 2926–2937.
53. Lee,B., Matera,G., Ward,D.C. and Craft J. (1996) *Proc. Natl Acad. Sci. USA*, **93**, 11471–11476.
54. Lygerou,Z., Allmang,C., Tollervey,D. and Seraphin,B. (1996) *Science*, **272**, 268–270.
55. Jacobson,M.R., Cao,L.G., Wang,Y.L. and Pederson,T. (1995) *J. Cell Biol.*, **131**, 1649–1658.
56. Bertrand,E., Houser-Scott,F., Kendall,A., Singer,R.H. and Engelke,D.R. (1998) *Genes Dev.*, **12**, 2463–2468.
57. Dundr,M. and Olson,M. (1998) *Mol. Biol. Cell*, **9**, 2407–2422.
58. Peculis,B.A. and Greer,C.L. (1998) *RNA*, **4**, 1610–1622.
59. Jeeninga,R.E., Delft,Y.V., de Graaff-Vincent,M., Dirks-Mulder,A., Venema,J. and Raué,H.A. (1997) *RNA*, **3**, 476–488.
60. Tollervey,D., Lehtonen,H., Jansen,R., Kern,H. and Hurt,E. (1993) *Cell*, **72**, 443–457.
61. Gautier,T., Bergès,T., Tollervey,D. and Hurt,E. (1997) *Mol. Cell. Biol.*, **17**, 7088–7098.