Volume 316, number 3, 247–252 FEBS 12042 February 1993

Conserved secondary structures in the ITS2 of trematode pre-rRNA

Bernard Michot[a], Laurence Despres[b,c], François Bonhomme[b] and Jean-Pierre Bachellerie[a]

[a]*Laboratoire de Biologie Moléculaire Eucaryote, Institut de Biologie Cellulaire et de Génétique du CNRS, Université Paul Sabatier, 118 Route de Narbonne, 31062 Toulouse Cédex, France,* [b]*URA CNRS 1493, Laboratoire Génome et Populations, Univ. Montpellier II, 34095 Montpellier Cédex, France* and [c]*URA CNRS 698, Centre de Biologie et d'Ecologie, Tropicale et Méditerranéenne, Université de Perpignan, 66860 Perpignan Cédex, France*

Received 5 November 1992; revised version received 11 December 1992

Biological functions of transcribed spacer regions in eucaryotic pre-rRNAs remain elusive so far. Utilization of the comparative approach to analyse their secondary structure has been impeded by the extensive sequence divergence observed among most of the specimens available to date. However, we have taken advantage of the recent derivation of a set of largely similar sequences for different schistosome species to look for the presence of constrained secondary structures within the internally transcribed spacer 2 (ITS2). We show that several common features of secondary structure are shared by these species despite sequence variation, with some of them supported by compensatory changes, suggesting a significant role of ITS2 as an RNA domain during ribosome biogenesis.

Pre-rRNA; ITS2; Secondary structure; Trematode; Evolution; rRNA processing

1. INTRODUCTION

Mature rRNAs are produced by the processing of a large precursor from which different transcribed spacer regions are sequencially removed through an elaborate pathway of cleavage steps [1]. In eucaryotes, transcribed spacer regions may represent a very substantial fraction of the length of the primary transcript [2]. They are particularly expanded in mammals, in which they amount to about half the size of the ribosomal RNA transcription unit. Although these transcribed spacer regions are obvious candidates for important roles in the control of ribosome biogenesis, elucidation of their biological function and of the molecular mechanisms involved in their accurate excision still remain a major challenge. Recent functional analyses performed on yeast *S. cerevisiae* ribosomal RNA genes clearly show that the structural integrity of the transcribed spacer regions is an essential prerequisite for correct processing of mature rRNA and biogenesis of active ribosomal subunits [3,4]. The derivation of reliable secondary structure models for each transcribed spacer region would undoubtedly represent a major step towards a detailed understanding of their biological role. The comparative sequence analysis provides the most powerful tool for identifying the biologically relevant folding pattern of an RNA molecule, i.e. its native structure within the cellular context [5]. However, an essential prerequisite for its effective utilization is the availability of a collection of sequences exhibiting a substantial number of nucleotide differences while remaining similar enough for unequivocal sequence alignments. Due to the high rate of sequence variation of transcribed spacers, which may exhibit dramatic size variation and extensive sequence divergence even among moderately distant species [6–8], this condition could not be fulfilled with the species available in the literature for most transcribed spacers. Nevertheless, the presence of phylogenetically conserved secondary structure elements in the 5′ externally transcribed spacer was recently revealed by the comparative analysis of a limited set of vertebrate sequences [9]. In the present study we have focused our attention on another transcribed spacer, ITS2, which interrupts the eucaryotic large subunit rRNA molecule and has no procaryotic equivalent [2]. Even between species like mouse and rat, which have diverged only 9–12 million years ago [10], sequence similarities are restricted to very minor portions of ITS2 [6]. The realization that ITS2 could serve as a sensitive indicator of phylogenetic relationships among very closely related species, owing to its high rate of sequence variation, results in a surge of new sequence determinations [11–13]. In a recent molecular phylogenetic study of schistosomes (platyhelminthes: trematoda), complete ITS2 sequences have been determined for seven different species which are parasites of humans, rodents or rumi-

Correspondence address: J.P. Bachellerie, Laboratoire de Biologie Moléculaire Eucaryote, Institut de Biologie Cellulaire et de Génétique du CNRS, Université Paul Sabatier, 118 Route de Narbonne, 31062 Toulouse Cédex, France. Fax: (33) (61) 33 58 86.

Abbreviations: ITS2, internal transcribed spacer 2.

The novel *Fasciola hepatica* sequence is available under the Genbank accession number L07844. The previously reported schistosome sequences are also available, for *S. mansoni* (L03658) *S. haematobium* (L03656), *S. intercalatum*, a sequence identical to *S. bovis* (L03657), and *S. japonicum* (L03660).

nants [13]. Their high level of similarity prompted us to perform a comparative analysis in terms of RNA folding pattern. In order to better evaluate the significance of the results, the ITS2 sequence of a species belonging to another genus of trematodes, *Fasciola hepatica*, a digenetic species, was also determined and included in the comparisons.

2. MATERIALS AND METHODS

F. hepatica ITS2 sequence was determined after PCR amplification of ITS2 using a pair of primers chosen in phylogenetically conserved sequences at the 5′ end of 28 S rRNA and in 5.8 S rRNA, respectively, as reported elsewhere [13]. The comparative analysis was performed according to a previously reported approach [14]. Optimal and suboptimal RNA foldings predicted on a thermodynamical basis [15,16] were compared and solutions which maximized structural homologies among the different species were examined for the potential presence of comparative support. Computations were performed using the software package of the Genetic Computer Group, University of Wisconsin [17].

3. RESULTS AND DISCUSSION

Although sequence determinations had been performed on seven different schistosome species [13], only four ITS2 sequences turned out to be sufficiently dissimilar from each other to be useful for the comparative approach. Phylogenetic trees derived from molecular markers suggest that the two most closely related species (*S. intercalatum* and *S. haematobium*) may have diverged 1–4 million years ago while the separation of *S. japonicum* from the branch leading to the 3 other species could have occurred 24–70 million years ago [13]. As shown in Fig. 1, these specimens can be unambiguously aligned over the entire length of ITS2. As for the digenetic trematode *F. hepatica*, its sequence diverges so extensively from the group of schistosomes that, except for the 34 5′-terminal nucleotides, no sequence tract can be aligned, unless secondary structure considerations are taken into account (see below).

The schistosome sequences share an identical folding

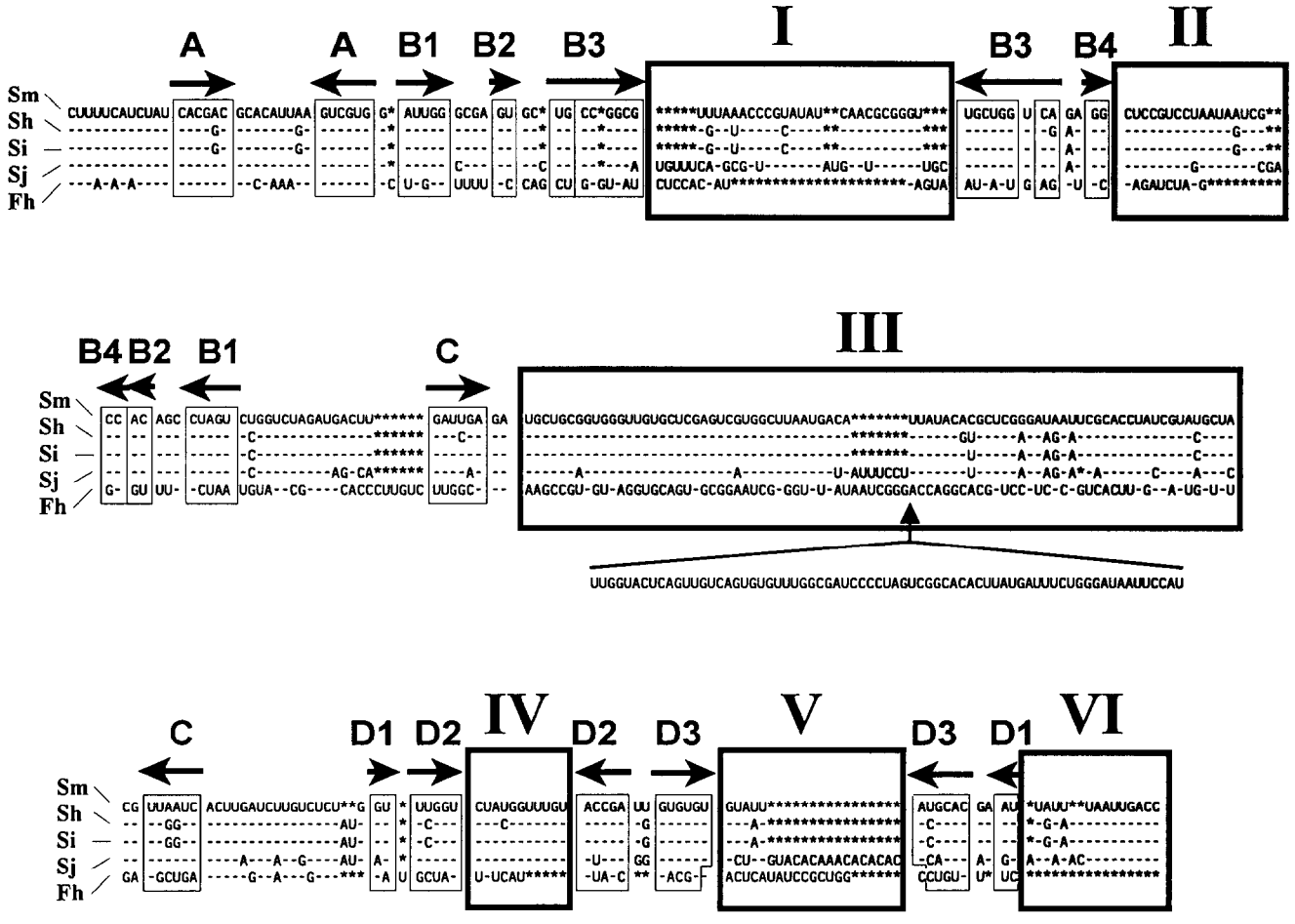

Fig. 1. Sequence alignments maximizing secondary structure conservation for platyhelminth ITS2. The schistosome sequences *S. mansoni* (Sm), *S. haematobium* (Sh), *S. intercalatum* (Si), *S. japonicum* (Sj) are essentially aligned as proposed elsewhere [13], except for the location of an oligonucleotide insertion in *S. japonicum* (or the 5′ end of box I sequence), which was slightly modified on the basis of RNA secondary structure considerations. Stars refer to gaps. Nucleotides involved in base pairings are boxed by a thin line, with the two strands of each stem delineated by arrows in opposite orientation and stems designated by letters, as indicated in Fig. 2. Portions of the sequence which cannot be aligned between schistosomes and *Fasciola hepatica* (Fh) are boxed (thick line).

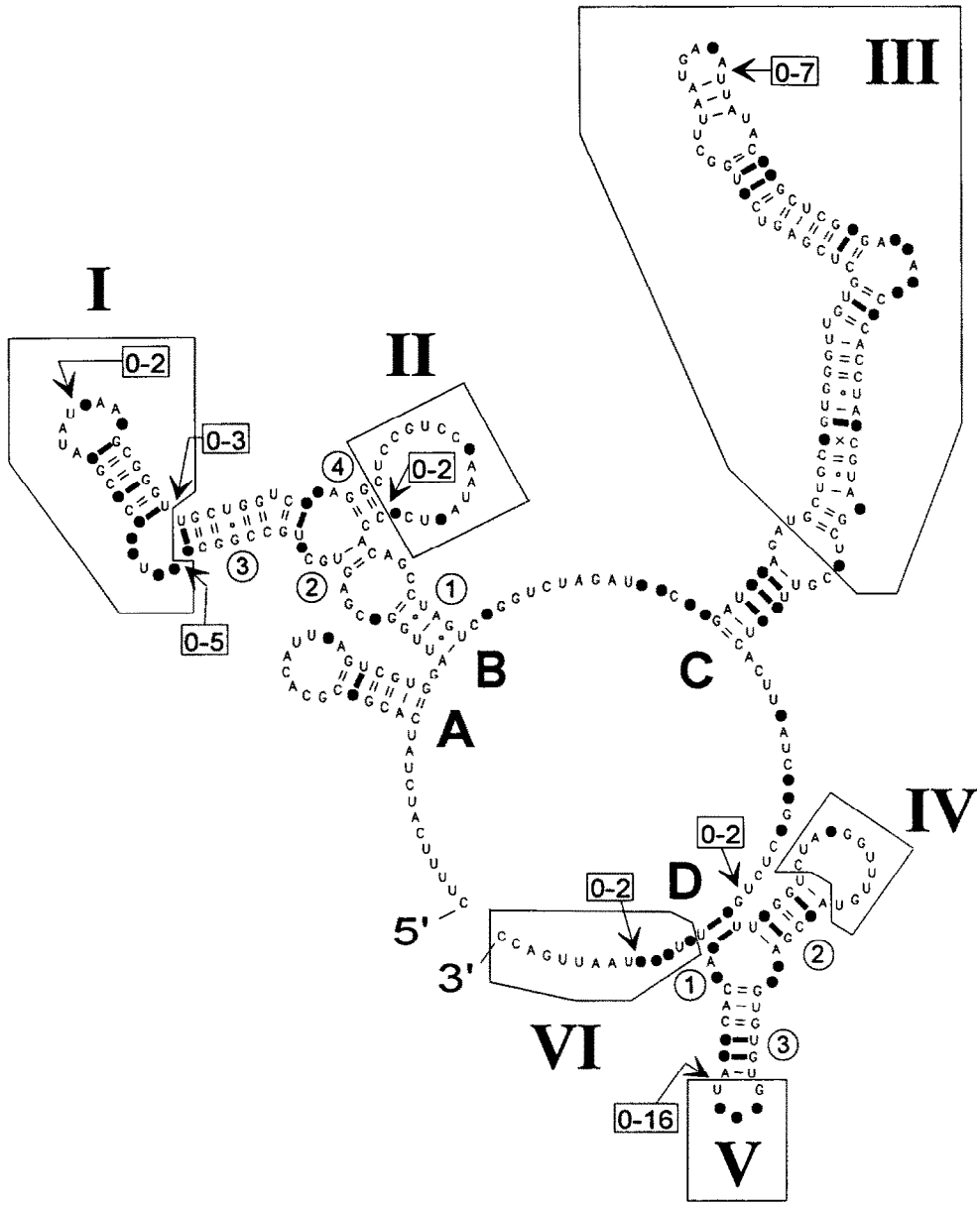

Fig. 2. Common folding of schistosome ITS2 sequences. Nucleotide positions which differ among schistosomes (including by point insertion/deletion) are denoted by a filled circle. Sites of oligonucleotide insertions are shown by arrows, with indication of the range of nucleotide variations. Pairings supported by compensatory changes are denoted by a thick bar, with the single non-compensatory change denoted by an ×. GU pairings are symbolized by a small open circle. Stems are identified by letters, with portions of composite stems further denoted by numbers, as defined in Fig. 1. Portions of schistosome ITS2 which cannot be aligned with the *Fasciola* sequence are boxed and numbered as in Fig. 1.

pattern (Fig. 2) in which four independent domains of secondary structure, termed A–D, can be identified. A number of nucleotide differences among these sequences are observed over proposed base-paired positions. Remarkably, all preserve the base-pairings. In fact, most of these changes concern only one of the two paired positions, with the presence of G·U appositions which, although less stable than the Watson–Crick complementarities, retain the RNA helical structure. Obviously, the occurrence of either a A:U/G·U or a G:C/G·U change is of weaker comparative value, on a statistical basis, than a truly compensatory change. However, their systematic accumulation, together with the presence of a couple of *bona fide* compensatory changes, definitely favors the Fig. 2 model. Moreover, this schistosome consensus folding finds additional support (see below) when the comparison is extended to another trematode, *Fasciola hepatica*, which does not belong to schistosomes. The comparative support among schistosomes appears strong for the elongated stem C, with eight compensatory changes vs. a single non-compensatory one. Stems B and C are also signifi-

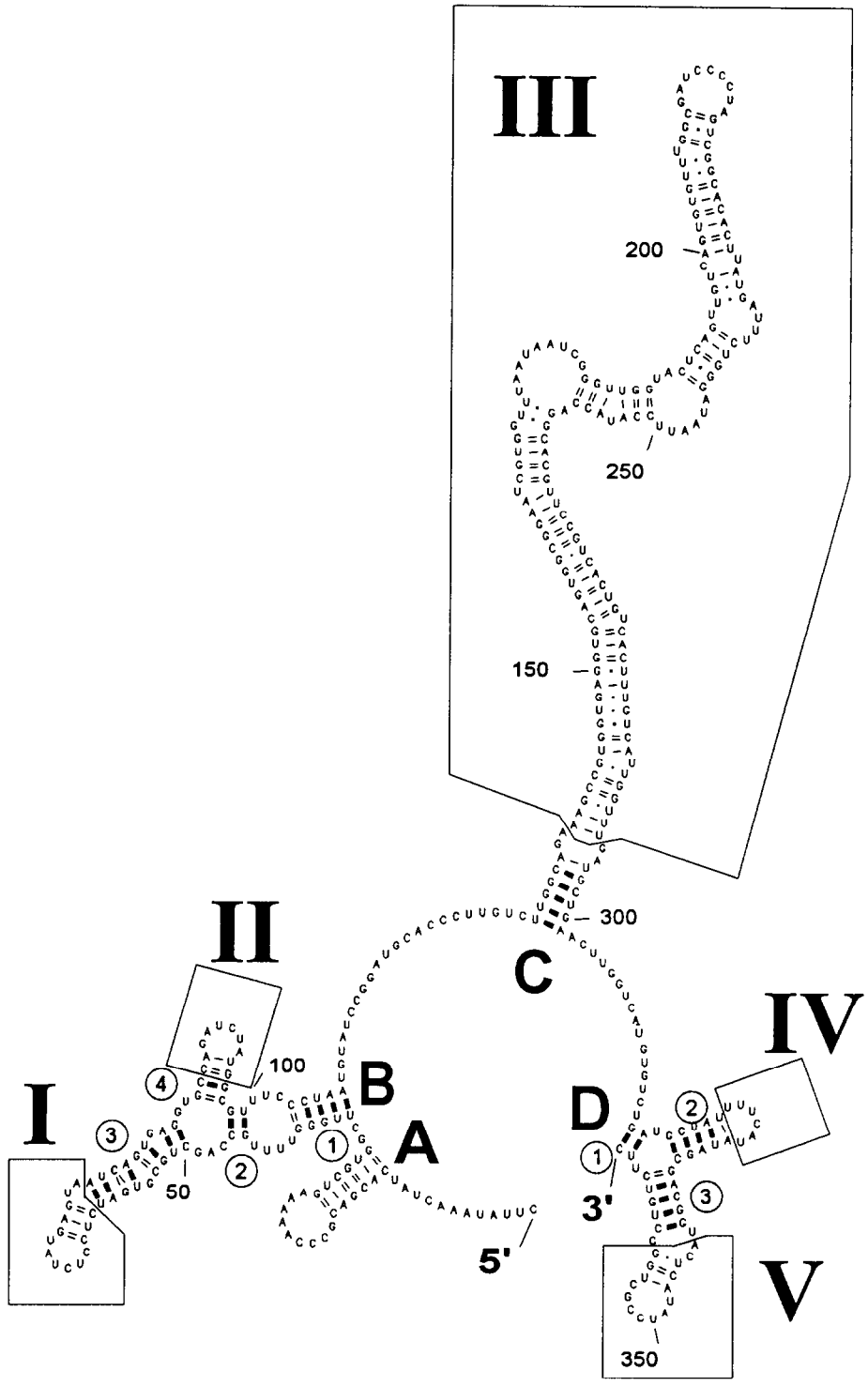

Fig. 3. Folding of *F hepatica* ITS2 sequence. Portions of ITS2 which cannot be aligned with schistosome sequences even when taking into account the secondary structure similarity are boxed (numbers as in Fig. 1). Base pairings supported by comparative changes between *F. hepatica* and schistosome sequences are denoted by thick bars.

cantly supported on a comparative basis. However, the sequence of stem A is not variable enough (a single nucleotide difference which preserves the base-pairing) for a definitive comparative test. It is also noteworthy that the few insertions/deletions of oligonucleotides among the schistosome sequences do not disturb the stem structures.

It is remarkable that the folding pattern of *F. hepat-*

ica, chosen as a close outgroup reference [18], is closely similar to the schistosome model, as shown in Fig. 3, even over the portions of ITS2 which do not exhibit any obvious sequence similarity. In addition to stem A which is precisely maintained, the basal parts of stems B, C and D are also preserved, with differences mostly restricted to the apical regions of these stems (boxed in Fig. 3). Moreover, even if the apical region of stem C differs in the details of its secondary structure between *F. hepatica* and the schistosomes, it folds in both cases into a very long unbranched stem. As for the conserved parts of stems B–D, they are supported by a large amount of comparative evidence, when comparing *F. hepatica* with the schistosomes, as summarized in Fig. 4. In addition several conserved nucleotides are detected over the single-stranded portions which connect the four conserved stems, with particularly an ACUU motif preserved immediately downstream from stem C. However, this core of the secondary structure is not perfectly conserved and some very minor variations are tolerated in the length of the single-stranded linker regions, with a one-nucleotide difference in the A–B linker, a dinucleotide insertion/deletion in the C–D linker and a hexanucleotide insertion/deletion in the B–C linker.

The detection of definite secondary structure constraints on the variation of trematode ITS2 sequences argues strongly for this region of pre-rRNA to have a significant function if its own as a structured RNA domain, probably in the control of preribosome assembly and rRNA processing. In yeast, different deletions of ITS2 have been shown to completely block the formation of mature 26 S rRNA [4]. It is remarkable that the secondary structure proposed for the yeast *S. cerevisiae* [4,9], with the presence of a giant cross-like structure encompassing the entire ITS2, does not exhibit any obvious similarity with the model shown in Fig. 4. Although it was noted [20] that the tobacco and rice ITS2 sequency can be folded into secondary structures which are related to each other and reminiscent of the yeast model, details of these proposals were not supported by comparative proof. In fact, a recent comparative analysis of a collection of higher plant ITS2 of substantial

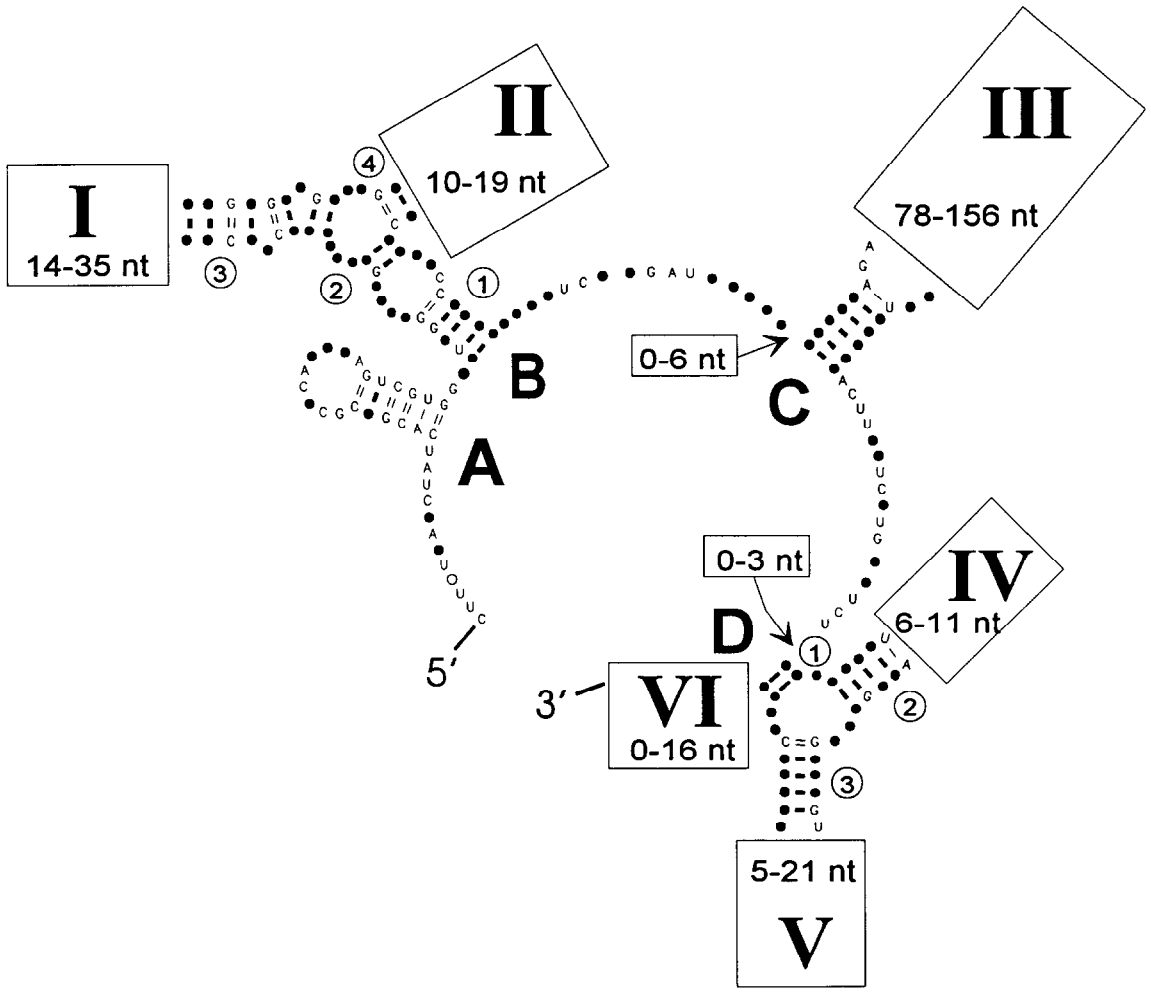

Fig. 4. Common ITS2 folding shared by schistosomes and *F. hepatica*. Symbols as in Fig. 2, with the non-alignable portions schematized by numbered boxes with indication of the range of size variation, in nucleotides.

sequence similarity points to the conservation of alternative secondary structure features, apparently unrelated either to those detected in yeast or to those phylogenetically supported in trematodes (B. Michot and P. Roussea, unpublished results). Such a diversification of structural constraints suggests that the detailed function of ITS2 may have undergone significant changes during the evolution of the major groups of eucaryotes, similar to what has been observed for some of the variable domains of mature rRNAs [14,21]. Our ongoing comparative analysis of other sets of related ITS2 sequences should clarify this issue.

Acknowledgements: This work was supported by grants from the Ministère de l'Education Nationale (action: Evolution) and from Région Midi-Pyrénées to J.P.B., and by a fellowship (Lavoisier Program, Ministère des Affaires Etrangères) to L.D.

REFERENCES

[1] Perry, R.P. (1976) Annu. Rev. Biochem. 45, 605–629.
[2] Crouch, R.J. and Bachellerie, J.P., in: DNA Systematics, Vol. I: Evolution (S.K. Dutta, ed.), CRC Press, Boca Raton, Florida, 1986 pp. 47–80.
[3] Musters, W., Boon, K., Van der Sande, C.A.F.M., Van Heerikhuizen, H. and Planta, R.J. (1990) EMBO J. 9, 3989–3996.
[4] 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.
[5] Michot, B., Hassouna, N. and Bachellerie, J.P. (1984) Nucleic Acids Res. 12, 4259–4279.
[6] Michot, B., Bachellerie, J.P. and Raynal, F. (1983) Nucleic Acids Res. 11, 3375–3391.
[7] Furlong, J.C. and Maden, B.E.H. (1983) EMBO J. 2, 443–448.
[8] Ajuh, P.M., Heeney, P.A. and Maden, B.E.H. (1991) Proc. R. Soc. Lond. B. 245, 65–71.
[9] Michot, B. and Bachellerie, J.P. (1991) Eur. J. Biochem. 195, 601–609.
[10] Jaeger, J.J., Tong, H. and Denys, C. (1986) C.R. Acad. Sci. 302, 917–922.
[11] Baldwin, B.G. (1992) Mol. Phylog. Evol. 1, 3–16.
[12] Lee, S.B. and Taylor, J.W. (1992) Mol. Biol. Evol. 9, 636–653.
[13] Despres, L., Imbert-Establet, D., Combes, C. and Bonhomme, F., Mol. Phylog. Evol., accepted.
[14] Michot, B. and Bachellerie, J.P. (1987) Biochimie 69, 11–23.
[15] Zuker, M. and Stiegler, P. (1981) Nucleic Acids Res. 9, 133–148.
[16] Freier, S.M., Kierzed, R., Jaeger, J.A., Sugimoto, N., Caruthers, M.H., Neilson, T. and Turner, D.H. (1986) Proc. Natl. Acad. Sci. USA 83, 9373–9377.
[17] Devereux, J., Haeberli, P. and Smithies, O. (1984) Nucleic Acids Res. 12, 387–395.
[18] Despres, L. (1991) Ph. D. thesis. Université de Montpellier II.
[19] Yeh, L.C.C. and Lee, J.C. (1990) J. Mol. Biol. 211, 699–712.
[20] Venkateswarlu, K. and Nazar, R. (1991) Plant Mol. Biol. 17, 189–194.
[21] Michot, B., Qu, L.H. and Bachellerie, J.P. (1990) Eur. J. Biochem. 188, 219–229.