rRNA genes from the lower chordate *Herdmania momus*: structural similarity with higher eukaryotes

Bernard M.Degnan, Jun Yan, Clifford J.Hawkins[1] and Martin F.Lavin
Queensland Institute of Medical Research, Bramston Tce., Herston, Queensland 4006 and [1]Chemistry Department, University of Queensland, St Lucia, Queensland 4072, Australia

Received July 17, 1990; Revised and Accepted October 12, 1990 EMBL accession no. X53538

ABSTRACT

Ascidians, primitive chordates that have retained features of the likely progenitors to all vertebrates, are a useful model to study the evolutionary relationship of chordates to other animals. We have selected the well characterized ribosomal RNA (rRNA) genes to investigate this relationship, and we describe here the cloning and characterization of an entire ribosomal DNA (rDNA) tandem repeat unit from a lower chordate, the ascidian *Herdmania momus*. rDNA copy number and considerable sequence differences were observed between two *H. momus* populations. Comparison of rDNA primary sequence and rRNA secondary structures from *H. momus* with those from other well characterized organisms, demonstrated that the ascidians are more closely related to other chordates than invertebrates. The rDNA tandem repeat makes up a larger percentage (7%) of the genome of this animal than in other higher eukaryotes. The total length of the spacer and transcribed region in *H. momus* rDNA is small compared to most higher eukaryotes, being less than 8 kb, and the intergenic spacer region consists of smaller internal repeats. Comparative analysis of rDNA sequences has allowed the construction of secondary structures for the 18S, 5.8S and 26S rRNAs.

INTRODUCTION

The genes that code for the three rRNA molecules found in the ribosome from higher eukaryotes are transcribed as a single unit by RNA polymerase I (1, 2). The transcription unit, consisting of the 18S, 5.8S and 26S/28S rRNA genes, is tandemly repeated many times, with copy number per haploid genome ranging from about 50 to 10,000 in different species (3). Separating each rRNA cistron is a non-coding spacer DNA, termed the non-transcribed spacer (NTS) or intergenic spacer (IGS), which is often made up of smaller internal repeat units (4, 6). Although rRNA genes are conserved throughout evolution, the rDNA length varies considerably among higher eukaryotes, ranging in size from 7.2 kb in the nematode, *Caenorhabditis elegans* (7), to 44 kb in the rat (8). The major source of variation is length differences in the IGS (9).

The pioneering experimental work on the secondary structure of ribosomal RNA was carried out on the 16S and 23S subunits of *E. coli* (reviewed in 10). Although the structures of the core regions are conserved in all organisms, the cytoplasmic rRNAs of eukaryotes are larger than those of prokaryotes. This is the result of an initial insertion of blocks of sequence into the rRNA genes and subsequent extension of helices and loops within these blocks (9, 11). Consequently, these regions have been termed expansion segments (12). The locations of expansion segments are conserved in eukaryotes, although sequence is not. The original higher eukaryotic secondary structure models, the 18S rRNA from *Xenopus laevis* (13) and *Artemia salina* (14), and the 28S rRNA from *Xenopus laevis* (12) and mouse (15), were constructed based on comparative analyses and thermodynamic stability.

The tadpole larva stage of ascidians is considered to be a living representative of the original progenitor to all vertebrates. The relationship of ascidian larvae with chordates is based on the presence of notochord-like cells in the tail, a hollow dorsal nerve cord and a cerebral vesicle (16, 17). Although the ascidian is one of the more studied organisms in developmental biology (18, 19), very little is known about the composition and structure of the genome of these animals. We have cloned and characterized an entire tandem rDNA repeat unit from the ascidian *Herdmania momus* and have constructed secondary structure models for the 18S, 5.8S and 26S rRNAs in order to obtain a better understanding of the phylogenetic relationships between chordates and other eukaryotes. Primary sequence and secondary structure of conserved and divergent regions are compared with known vertebrate and invertebrate rRNA genes.

MATERIALS and METHODS

Isolation of rDNA

Herdmania momus (suborder Stolidobranchia, family Pyuridae) were collected from Heron Reef, Great Barrier Reef and from Middle Reef, Stradbroke Island, Queensland, Australia. *Pyura stolonifera* (suborder Stolidobranchia, family Pyuridae) (20) were collected from the tidepools at Hastings Point, New South Wales. Gonads from the hermaphrodites were dissected, placed in millipore filtered sea water, and teased apart, releasing sperm and oocytes. DNA was extracted from sperm by lysing in NETS (100 mM Tris-HCl, pH 7.6, 200 mM NaCl, 40 mM EDTA, 2% SDS, 250 μg/ml Proteinase K (Boehringer-Mannheim)), followed by phenol/chloroform extraction (21), and purified by CsCl isopycnic gradient centrifugation (22).

High molecular weight DNA from *H. momus* (Heron Reef) was digested with different restriction endonucleases (Promega) and satellite DNA from each digest was gel purified (21). DNA fragments were ligated into the appropriate restriction sites of Bluescript, pSK+ (Stratagene). The recombinant clones were screened using random prime-labelled, *H. momus* XbaI satellite DNA (8kb). Inserts were hybridized to Southern blotted *H. momus* genomic DNA to check their authenticity (21).

Characterization and sequencing of cloned rDNA

Two recombinant plasmids, one containing an 8.0 kb PstI insert, designated pHRP75, and the other a 4.4 kb ClaI insert, designated pHRC44C, were shown to hybridize to the major satellite band of *H. momus* DNA. Each insert was restriction mapped and shown to be homologous with the other by Southern blot hybridization (21). Sequencing of pHRP75 and pHRC44C was performed by the dideoxy chain termination procedure (23) using ssDNA template produced from the pSK+ helper phage system (Stratagene). A strategy using a combination of subcloned fragments and nested unidirectional deletions (24), was used to sequence both pHRP75 and pHRC44C (Fig. 1B). Sequence differences between the two fragments were noted. At least two sequencing reactions were performed for DNA obtained from separate subclones of the same fragment. To ensure no loss of sequence data at the subcloning sites, other clones were constructed to ensure sequence was read through the restriction sites.

Secondary structures for the small subunit were based on the models of Nelles *et al.* (14), Gutell *et al.* (25) and Neefs *et al.* (26), and for the large subunit on the models of Michot *et al.* (15), Gutell and Fox (27) and Gutell *et al.* (28). The *H. momus* sequence was aligned to identify conserved and divergent regions. Expansion regions that could not be constructed using comparative analysis were determined on the basis of thermodynamic stability (29).

Genome size and rDNA copy number

The genome size of the two populations of *H. momus* and *P. stolonifera* were determined by 33258 Hoechst dye fluorescence (Hoeffer Fluorometer) of DNA (30) from a known number of sperm. Sperm, from five individuals, were purified from oocytes and other gonad debris by filtering through 60 μm nylon mesh, counted on a haemocytometer, pelleted by centrifugation at 5000 g for 5 min, and resuspended at 2×10^9 sperm/ml with sterile sea water. Sperm were then lysed by the addition of an equal volume of NETS, vortexed vigorously, freeze/thawed, and adjusted to the equivalent of 10^8 sperm/ml with the addition of TEN (10 mM Tris-Cl, pH 7.6, 0.1 mM EDTA, 100 mM NaCl). The fluorescence of the lysed sperm samples was measured and related to a standard curve using ng levels of *H. momus* DNA and 33258 Hoechst dye. The amount of DNA/sperm and hence per haploid genome, was calculated.

rRNA copy number per haploid genome was determined using quantitative filter hybridization with dot-blotted genomic and plasmid DNA using a number of rDNA probes (31). High molecular weight DNA from the ascidians and pHRP75 were freeze/thawed repeatedly in order to shear the DNA and nick the plasmid, and aliquotted, in triplicate, in the range 6 to 50 ng for the ascidian DNA and 0.5 to 4 ng for pHRP75. λDNA was added to each aliquot so that the final amount of DNA was 100 μg. Samples were alkali denatured and dot blotted following standard protocols (21). Two random prime-labelled rDNA probes were hybridised to the dot blotted DNA; an XbaI/SacI fragment (18S) and an XhoI/EcoRI fragment (26S), both from pHRP75. Counts hybridizing to each DNA dot were measured and a standard curve was constructed from the hybridization profile to pHRP75. From this curve, genomic copy numbers were determined and related to the amount of DNA on the filter, and hence number of haploid genome equivalents.

Other analyses

H. momus (Heron) rDNA probes were used in qualitative Southern blot analysis (21) to compare *H. momus* (Heron Reef) to *H. momus* (Middle Reef) and *P. stolonifera.* Probes used were pHRP75 and the ClaI/PvuII IGS fragment from pHRP75. The genomic DNAs were pooled from at least five individuals of each group of animals.

RESULTS

rDNA cloning and sequencing

Digestion of DNA from the ascidian *H. momus* with PstI, and XbaI, resulted in a single restriction satellite band about 8 kb in size, while digestion with a number of other restriction enzymes generated smaller, more numerous satellite bands (Fig. 1A). The major satellite was cloned from PstI (Fig. 1A) and ClaI digested DNA (not shown), checked by hybridizing back onto genomic DNA and restriction mapped. The restriction maps of pHRP75 (PstI fragment) and pHRC44C (ClaI fragment) are shown in Fig. 1B.

The complete sequence of an entire rDNA unit was determined from pHRP75 and pHRC44C (Fig. 2). The arrows below the

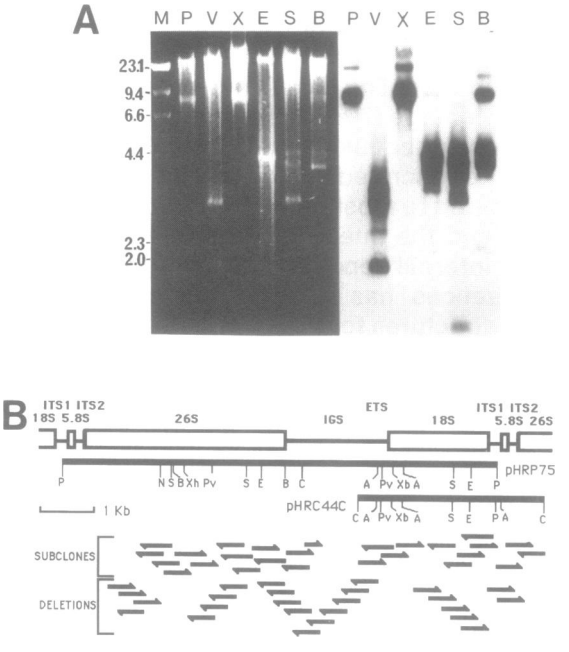


Figure 1. (A) Hybridization of pHRP75 to restriction enzyme digested *H. momus* genomic DNA. The left hand panel shows ethidium bromide stained genomic DNA and the right panel the hybridization pattern. M, markers; P, PstI; V, PvuII; X, XbaI; E, EcoRI, S, SacI; B, BglII. (B) Sequencing scheme of pHRP75 and pHRC44C. Restriction maps of pHRP75 and pHRC44C are shown in relation to the rDNA tandem repeat. Note that an arrow does not represent a single subclone, but is representative of the sequencing of at least two subclones. On average, 400–450 bases were read from ssDNA template with a double loading on a 50 cm gel. A, AccI; B, BglII, C, ClaI; E, EcoRI; N, NcoI; P, PstI; Pv, PvuII; S, SacI; X, XhoI; Xb, XbaI. IGS, intergenic spacer; ETS, external transcribed spacer; ITS 1 and 2, internal transcribed spacers.

restriction map identify the fragments that were sequenced (Fig. 1B). The entire 18S rRNA gene was sequenced in both clones and these were shown to be greater than 99% identical. Fragments from pHRP75 were used exclusively to sequence the 3′ end of the 26S rRNA gene and most of the IGS, and fragments from pHRC44C were used to exclusively sequence the 5.8S, internal transcribed spacer 2 (ITS 2) and the 5′ end of the 26S gene. In the region where minor differences were observed, the sequence presented is that for clone pHRC44C.

Based on sequence similarity with other higher eukaryotes, the putative ends of mature 18S, 5.8S and 26S were determined (Fig. 2A). The boundaries are estimations and further analyses are required to map exact sites of cleavage of the precursor rRNA. The total length of the rDNA tandem repeat is 7,967 bp and has

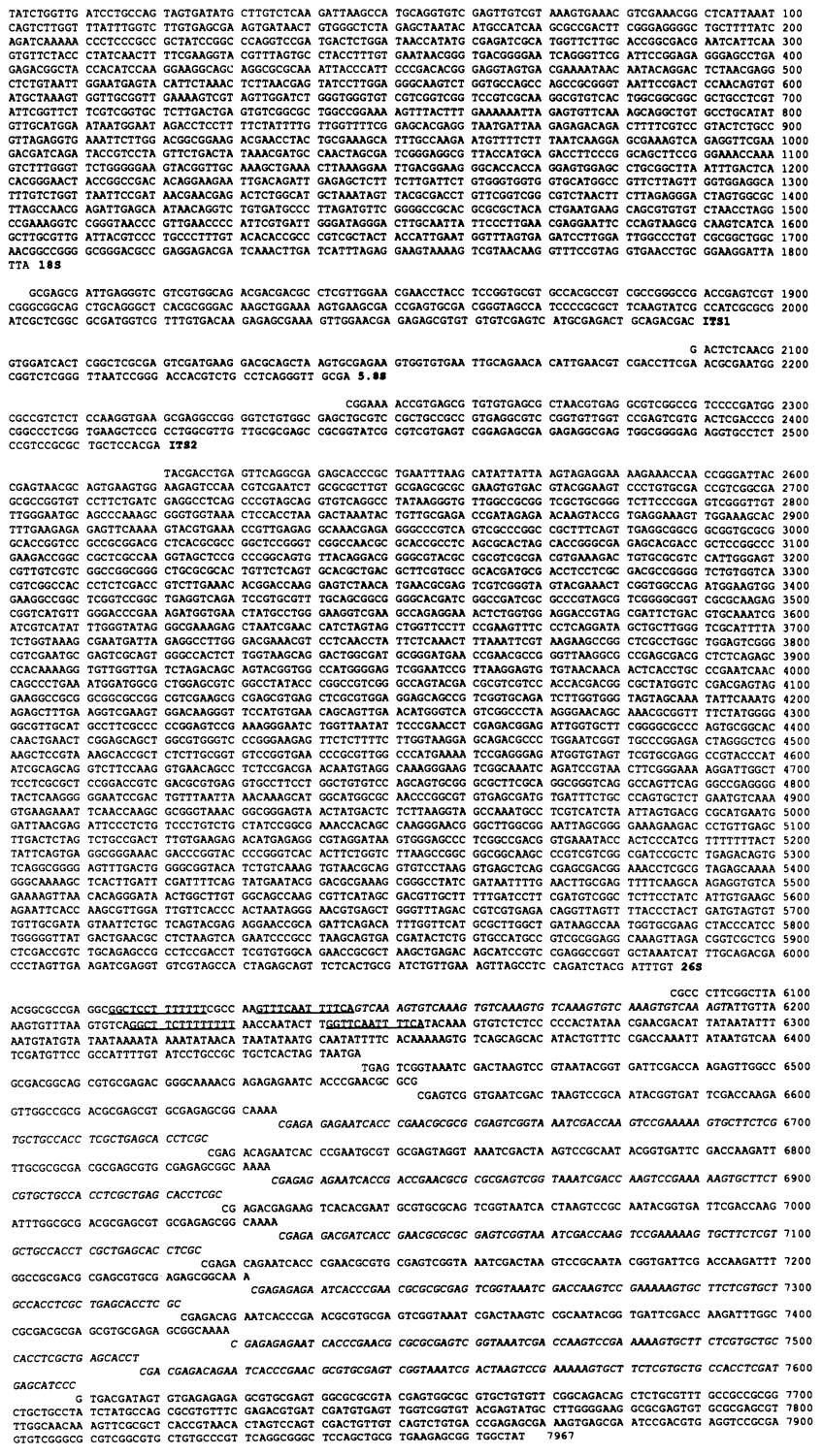


Figure 2. The sequence of an entire rDNA tandem repeat from *H. momus*, beginning at the 18S gene. (A) The 18S, 5.8S and 26S rRNA genes. (B) IGS region. Numbering begins at the 5′-end of the 18S rRNA (Fig. 2a) and finishes at the 3′-end of IGS (Fig. 2b). Each internal repeat within the IGS is depicted by a line change. One set of internal repeats is italicized. Small repeats at the 5′ end of the IGS are underlined and italicized.

a GC content of 56.0%. The coding regions of the mature 18S, 5.8S and 26S are 1,803, 155 and 3,566 bp respectively and all three regions have a high GC content (Table 1). The IGS is comprised of a series of tandem internal repeats approximately 100 bp in length flanked at the 3′ end by unique sequence and at the 5′-end by an AT rich region containing shorter repeats (Fig. 2B) . There are 12 copies of the internal repeat present as 2 sets of 6 repetitive elements. Greater than 90% sequence similarity is observed between repeats within each set, while the sequence is approximately 70% homologous between the two sets. Each category of repeat is arranged in alternate fashion with the other resulting in 5 double repeats of 200 bp each. This block is flanked by a single copy of one repeat at the 5′-end and a copy of the other at the 3′-end (Fig. 2B).

Secondary structure

Secondary structures for the *H. momus* rRNA were constructed by comparing the *H. momus* primary sequence with the mouse, brine shrimp and general models, and drawing a *H. momus* secondary structure that fitted the previously established models (14, 15, 25–28). Divergent regions were analysed by comparison of the secondary structures over a wider range of animals (human, *Xenopus, Drosophila* and *C. elegans)* and by an RNA secondary structure analysis program (29). The secondary structure models for *H. momus* rRNA are presented in five figures: 18S, Figs. 3a, b, and 5.8S − 26S, Figs. 4a, b, c. In general, the *H. momus* models are largely consistent with the 18S model of Nelles *et al.* (14) Gutell *et al.* (25) and Neefs *et al.* (26), and the 5.8S − 28S model of Michot *et al.* (15), Gutell and Fox (27) and Gutell *et al.* (28). Arrows indicate compensatory mutations (changes that occur in the primary sequence without affecting the secondary structure) found in the ascidian rRNAs when compared to the above models and *Xenopus* models (12, 13). Since variable regions D2, D8 and D12 differ significantly from previous published structures in their primary sequence, structural arrangements are based more on energy than comparative analysis and these portions of the model should be considered as tentative. Numbering of helices and divergent regions follow those of Nelles *et al.* (14) and Michot *et al.* (15). Any discrepancies between ascidian and previous secondary structure models are listed below.

18S rRNA. The *H. momus* 18S rRNA model, in general, concurs with the Nelles *et al.* (14), Gutell *et al.* (25) models and the more recent Neefs *et al.* (26) model. We followed the numbering system of Nelles *et al.* (14), which differs somewhat from a more recent model (26). The 5′ end of the 18S rRNA is presented as a pseudoknot (13, 25) (Fig. 3a). Helix 3 follows the structure proposed by Atmadja *et al.* (13), however the first strand of helix 3 can also basepair with the ssRNA strand between helices 16 and 17, resulting in a shortened helix 17 and an elongated helix 15 (25) (Fig. 5a). Helix 5 in Vl is extended at the base so that

Table 1. Structure and composition of a rDNA tandem repeat unit

Region	Length (bp)	%GC	Location*
18S gene	1803	50.4	1–1803
ITS 1	286	64.3	1804–2089
5.8S gene	155	56.8	2090–2244
ITS 2	276	68.9	2245–2520
26S gene	3566	57.1	2521–6086
IGS	1881	54.2	6087–7967
TOTAL	7967	56.0	–

*Location refers to position in nucleotide sequence (Fig. 2A, B).

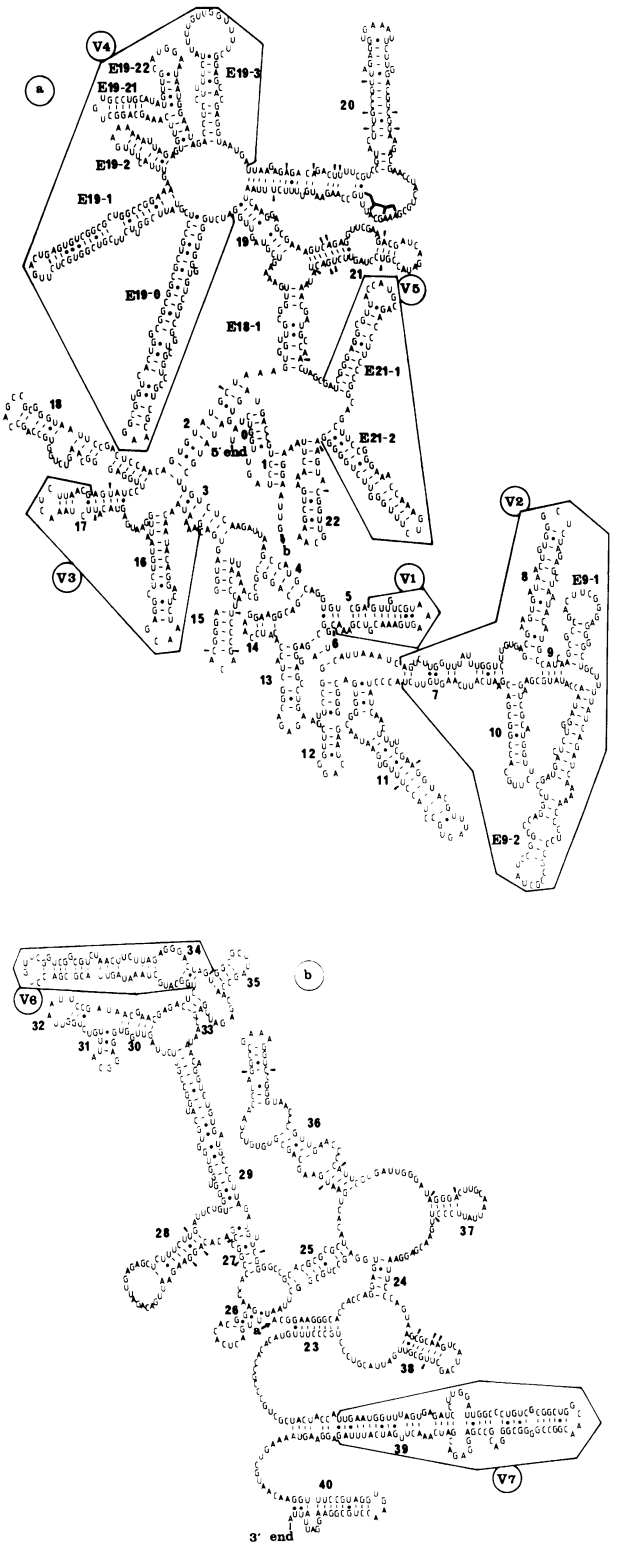
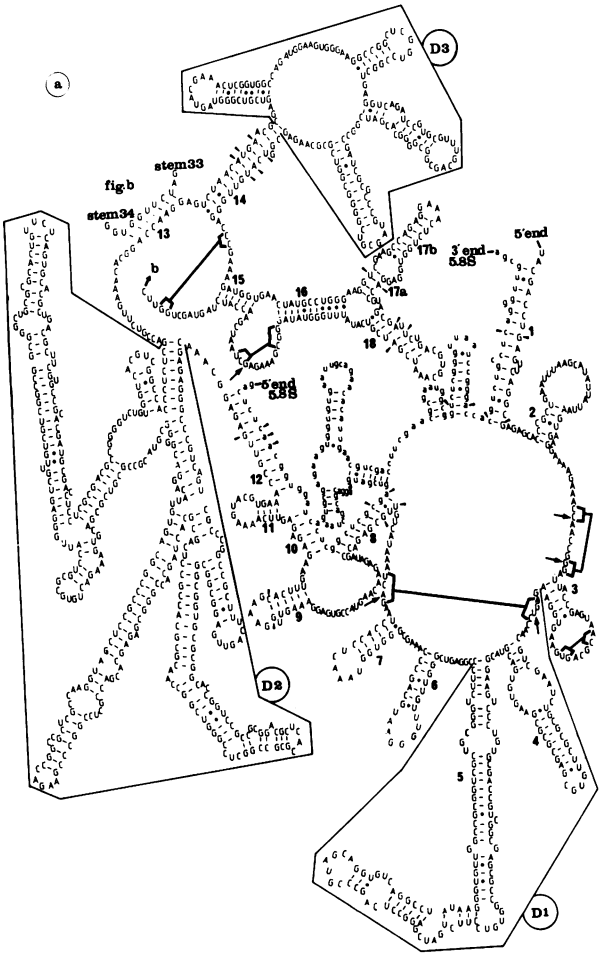


Figure 3a–b. Secondary structure model of the 18S rRNA of *H. momus*. Numbering system follows that of Nelles *et al.* (14). Compensatory mutations in relation to the *Artemia salina* (14) and *Xenopus laevis* (13) models are represented by small arrows. Basepairings proposed by Gutell *et al.* (25) are boxed and connected by a line.

the flanking single stranded loop regions are incorporated into the helix. Both V2 and V3 concur more closely with the Neefs *et al.* (26) model than the Gutell *et al.* (25) model. In V2, the helices are more extensively base-paired and helix 9 closes the terminal loop as in the *Artemia* model (14). Although an alternative basepairing of helix 3 shortens helix 17 (Fig. 5a), other structures that would completely eliminate this helix appear unlikely. Although helix E18−1 follows the Nelles *et al.* (14) format, the GU noncanonical basepairing in the middle of the helix probably does not exist. Such an internal loop would destabilise the entire helix, resulting in a structure similar to that proposed by Gutell *et al.* (25) (Fig. 5c). Helices E19−21, E19−22 and E19−3 of *H. momus* can alternatively be constructed to form a pseudoknot (Fig. 5b), as recently proposed (26). The V5 region, as presented, is similar to that proposed by Nelles *et al.* (14). However, more recent models (25, 26), are structured differently. Although *H. momus* does not possess compensatory mutations to support either model, the conservation of eukaryotic A−U higher-ordered basepairing (32) as shown in Fig. 5c supports the more recent models.

5.8S − 26S rRNA. The *H. momus* 5.8S − 26S rRNA model adhered closely to the Michot *et al.* (15) and Gutell and Fox (27) models. We followed the numbering system of Michot *et al.* Gutell and Fox (27) proposed a small helix between helices 2 and 3 (Fig. 4a). The ascidian sequence contains a compensatory mutation in this region, when compared to both vertebrate and invertebrate sequences, and this would support such a model (7, 12, 15, 33) (Fig. 4a). The basepairing of four nucleotides of



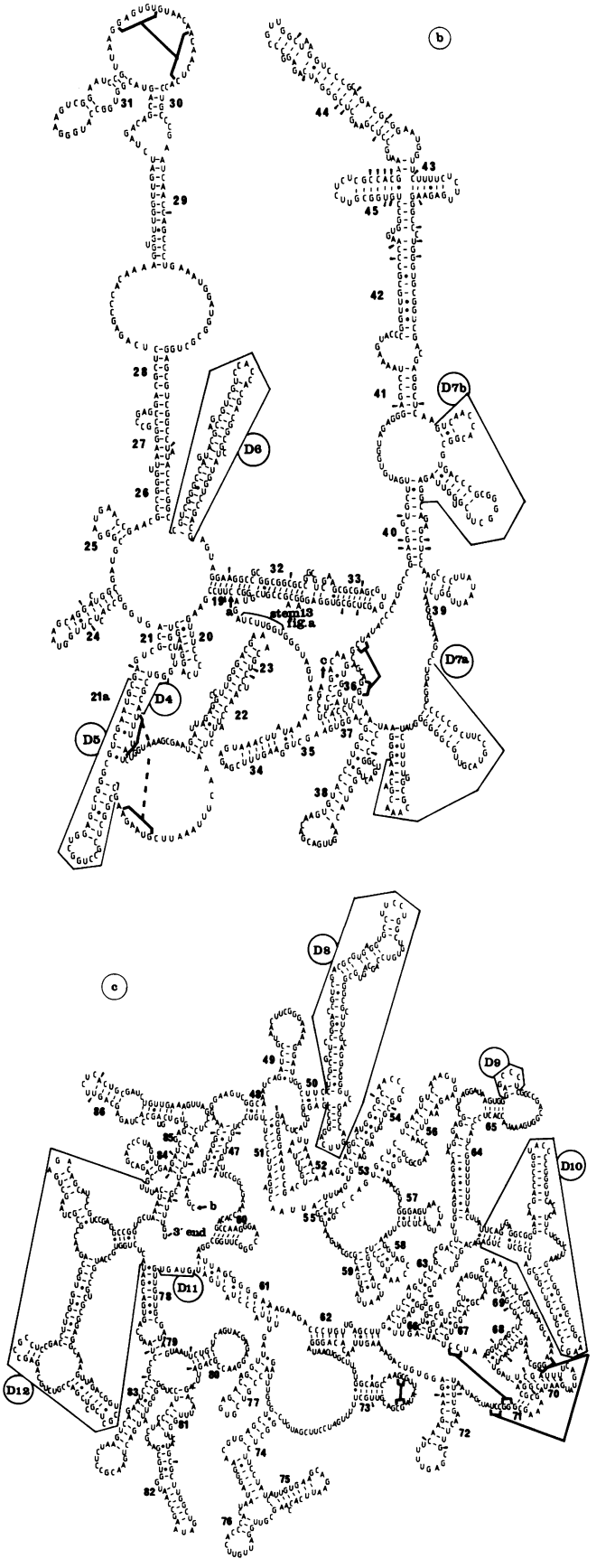


Figure 4a−c. Secondary structure model of the 5.8S−26S rRNA of *H. momus*. Numbering system follows Michot *et al.* (15). Compensatory mutations in relation to the mouse (15) and *Xenopus laevis* (12) models are represented by small arrows. Basepairings proposed by Gutell and Fox (27) are boxed and connected by a line.

ssRNA located between helices 3 and 4 with four nucleotides between helices 7 and 8 (27) contains a compensatory mutation in relation to other animals. Expansion segment Dl is similar to the Michot *et al.* (15) model with the exception of the absence of a branch at the end of helix 5 (Fig. 4a). D2 in the ascidian has been constructed to contain one small and two large secondary helices, with the first helix branched into three tertiary helices. Like the *Drosphila* model (33), the *H. momus* model does not adhere closely to the proposed three subdomains of the D2 segment (34). The proposed B helix in the D3 segment (27, 35), has a base change in *H. momus* that results in the helix being completely devoid of Watson-Crick basepairing, and is therefore not included in the model. The helix located between helices 15 and 16, as proposed by Gutell and Fox (27) is elongated in *H. momus* by a compensatory base change in the hairpin loop (Fig. 4a). Helix 20 in Fig. 4b is arranged in the extended form present in the *Xenopus* model (helix 28 in *Xenopus*) (12). Helices 21–33, including D4 and D5, appear to fit the *Xenopus* model more so than the other models. Helix 21 is shortened by 4 noncompensatory base changes and helix 21a which corresponds to helix 30 in Clark *et al.* (12) has been extended (Fig. 4b). Because of the presence of helix 21a, helices 22 and 23 are shortened. The more conventional structure (15, 27) is also shown in the D5 region. The helix proposed in the hairpin loop at the end of helix 30 (27) is extended in *H. momus*. Expansion segment D6 in *H. momus* is a simple helix similar to that of *Xenopus* (Fig. 4b). A compensatory mutation in helix 38 extends the helix and supports the *Xenopus* model (helix 44). However, the *H. momus* 26S rRNA sequence can be arranged to fit the Gutell and Fox (27) model in this region (Fig. 5d). The ascidian sequence contains three compensatory mutations when structured according to Gutell and Fox (27), supporting that model (Fig. 5d). D7a arrangement consists of two helices instead of one helix. Noncompensatory base changes prevent the formation of the short helix 46 which is found in the mouse model (15). Noncompensatory base changes at the end of helix 49 result in a larger terminal loop than in the mouse or general models, resembling the structure found in *C. elegans* (7) (Fig. 4c). The expansion segment D8 is similar in structure to other eukaryotes (Y-shaped), however helix 50 is much shorter than all other eukaryote structures previously studied (34). Helix 52 is lengthened as in *Xenopus* (helix 61). Helices 55 and 56 can be arranged as in Gutell and Fox (27), resulting in an extended helix 56 and only one nucleotide of ssRNA flanking helix 55. D9 is almost completely lost in *H. momus*, being only 8 bases in length. The primary sequence of D10 in *H. momus* can be arranged most favourably using the hammerhead configuration proposed for *Xenopus* (12). D12 is similar in structure to that in the mouse model, but smaller,

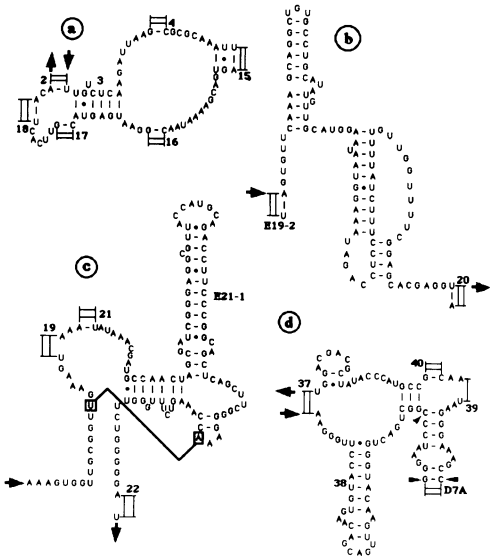


Figure 5. Alternative secondary structure models. (a) 18S rRNA from helix 2 to 15. This model is based on Gutell *et al.* (25). (b) 18S rRNA from helix E19–2 to 20. The pseudoknot is based on the model of Neefs *et al.* (26). (c) 18S rRNA from helix 2 to 22. This model is based on Gutell *et al.* (25). The box nucleotides correspond to the proposed complex interaction (32). (d) 26S rRNA from helix 37 to 40 and is based on the model presented by Gutell and Fox (27).

Ascidian rDNA comparisons

Genome sizes were analysed for two populations of *H. momus* and another Pyuridae, *P. stolonifera* (Table 3). These ranged from 0.27 pg/haploid genome in *H. momus* (Heron Reef) to 0.30 pg/haploid genome in *H. momus* (Middle Reef) and 0.87 pg/haploid genome in *P. stolonifera*. The rDNA copy number/genome varied significantly between *H. momus* populations, with the Heron Reef population containing 2410 copies determined with the 18S probe and 2380 copies with the 26S probe. The Middle Reef population was shown to contain 460 copies using the 18S probe and 350 copies with the 26S probe while *P. stolonifera* contained 1700 rDNA copies with the 18S probe and 1340 copies with the 26S probe (Table 2). The values from the 18S probe are taken to be more accurate since the 26S probe traverses divergent expansion regions, D7b–D10. Hybridization with a random prime-labelled 26S probe will be negatively biased if sequence differences exist within these expansion segments. Qualitative Southern analysis revealed differences in copy number and restriction pattern when comparing the two *H. momus* populations (Fig. 6, unpubl. results). The IGS is known to be the most rapidly evolving region in the rDNA, with large sequence diversity between closely related species (9, 11). The ClaI/PvuII fragment, containing most of the IGS from the cloned rDNA from the Heron Reef population (Fig. 1A), was radiolabelled and hybridized, under low stringency, to Southern blotted ascidian DNA (Fig. 6). Only 0.5 μg of Heron Reef DNA was blotted, while 2 μg of the other ascidian DNAs were blotted. The IGS fragment hybridized to the rDNA of the Heron Reef population (lanes H), but did not to the Middle Reef population (lanes M) or to *P. stolonifera* DNA (lanes P). When this filter was reprobed with pHRP75, under stringent conditions, there was hybridization to all species.

DISCUSSION

Morphological and physiological evidence has classified ascidians as members of the phylum Chordata (16). The combination of

Table 2. Comparison of genome size and rDNA tandem repeat number in ascidians

Species	Genome Size (10^9 bp)	rDNA Copies	% of Genome*
H. momus (Heron Reef)	0.27	2,410	7.1
H. momus (Middle Reef)	0.30	460	1.2
P. stolonifera	0.87	1,700	1.6

*The percentage of genome composed of rDNA repetitive sequence was calculated based on the length of a single repeat being 8 kb.

well characterized developmental features such as mosaic development and a well documented cell lineage (18, 19), small genome size (36, 37), and close phylogenetic relationship with vertebrates (16, 17), suggests that this group of animals is well suited for molecular investigations in development and evolution. Although small lengths of ascidian sequence have been analysed previously (38), the cloning and characterization of an entire rDNA tandem repeat from the ascidian *H. momus* is the first molecular analysis of the well characterized rRNA gene family for this subphylum. The rDNA is an ideal gene family to study in regard to evolutionary associations because the rates of molecular change differ between various domains within the rDNA (11), and therefore can be analysed and applied to a wide range of phylogenetic relationships.

Sequence and *H. momus* population analysis

The rDNA tandem repeat cloned from *H. momus* is 7967 bp in size, the second smallest higher eukaryote characterized, with *C. elegans* being smaller at 7196 bp (7). Unlike the *C. elegans* rDNA, the *H. momus* has an IGS structure similar to vertebrates and *Drosophila*, being made up predominantly of internal repeats. The IGS size differences which are based on the number of repeats, are responsible for much of the interspecific and intraspecific variation in rDNA length in many species (39, 40). Southern analysis of *H. momus* (Heron Reef) genomic DNA (Fig. 1A) demonstrates size variation between rDNA tandem repeats. The enhancer and RNA polymerase I promoter function lie within this repetitive DNA region and the promoter and initiation site map to the repeat closest to the 3′ end of the IGS (41). The IGS undergoes rapid sequence changes by molecular drive, which is the process whereby gene variants are spread through a multigene family and eventually through a sexual population, and by the limited constraints on the RNA polymerase I system to coevolve (11, 42). This rapid coevolution is illustrated by the interspecific incompatibility of rDNA *cis*-elements and RNA polymerase I complexes from different species, and the lack of consensus sequence (43). We have shown that an IGS probe from Heron Reef population does not hybridize to the Middle Reef population under low stringency conditions (Fig. 6), strongly suggesting very poor sequence similarity between IGSs of the two populations. Interestingly, gametes from these two populations can be interchanged to form normal tadpole larvae (unpublished data).

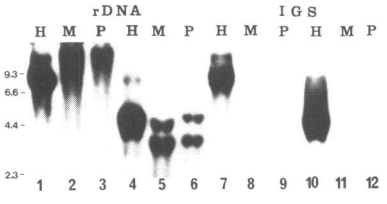


Figure 6. Comparison of ascidian rDNAs by Southern blot analysis. An IGS ClaI/PvuII fragment from pHRP75 was hybridized and washed at low stringency to the filter designated IGS (lanes 7–12). The filter was stripped (21) and reprobed, under stringent conditions with total pHRP75 insert and is designated rDNA (lanes 1–6). Molecular weight standards are shown on the side. H, *H. momus* (Heron Reef), M, *H. momus* (Middle Reef); P, *P. stolonifera*. Lanes 1–3, 7–9, genomic DNA digested with PstI; lanes 4–6, 10–12, genomic DNA digested with EcoRI.

Secondary structure

The *E. coli* rRNA models, based on experimentation (10), have allowed the construction of more secondary models that are based on comparison. These comparisons have reinforced and allowed for the refining of the experimental models. Compensatory base changes in the primary structure, leaving the proposed secondary structure intact, not only validate the proposed helix but suggest a conserved role for this helix (9). A large number of compensatory mutations can be found in the *H. momus* secondary structure when comparing to either brine shrimp or frog 18S models (13, 14) and the mouse and frog 28S models (12, 15). In some cases, the *H. momus* arrangement directly supports a specific model (Figs. 3, 4, 5 and results). Compensatory mutations in the *H. momus* sequence provide phylogenetic evidence for two helices at the 5′ end of 26S rRNA. Both these helices were not included in the mouse model (15), but are proposed as part of a general eukaryotic model by Gutell and Fox (27). The helix situated between helices 2 and 3 contains a U–A basepairing prior to the hairpin loop in all animal models except *H. momus*, which has a C–G basepair (Fig. 4a).

Another feature of the *H. momus* model involves the loss of helix B in expansion segment D3 (35). This highly conserved helix, in order to be maintained in *H. momus*, would require stabilisation without Watson-Crick basepairing. Helix 50, which leads to expansion segment D8 is truncated in *H. momus*. In all eukaryotes helix 50 contains approximately 20 basepairs and an internal bulge (34), but this helix has been reduced to 4 bp in *H. momus*.

The locations of insertion of expansion sequences are conserved in eukaryotes, although sequence is not. The recent construction of secondary structures from invertebrates (7, 14, 33) is of particular interest because of the insight into the conservation of secondary structures with gross changes in primary sequences within the expansion regions (4). Structural conservation among eukaryotes would suggest a functional role for expansion segments in the eukaryote ribosome that has little dependence

Table 3. Comparison of core and expansion segments of eukaryotic 26S/28S rRNA genes

Species	Total Length (bp)	Expansion Length (bp)	Segment Core (% GC)	Expansion Segment (%GC)	% Homology with *H. momus* Core
H. sapiens	5,025	2,625	69.1	82.3	87.0
M. musculus	4,712	2,319	54.0	79.5	86.3
X. laevis	4,110	1,668	45.4	82.4	87.9
H. momus	3,566	1,201	51.3	68.4	–
D. melanogaster	3,945	1,562	45.3	29.7	79.0
C. elegans	3,519	1,153	46.7	53.5	81.5
S. carlsbergensis	3,393	1,014	46.4	50.0	81.1

Homo sapiens (46), *Mus musculus* (15), *Xenopus laevis* (12), *Drosophila melanogaster* (33), *Caenorhabditis elegans* (7) and *Saccharomyces carlsbergensis* (47) sequence data were compiled in Hancock and Dover (44).

on primary sequence. Hancock *et al.* (33) have shown that the general secondary structure of expansion segments proposed by Michot *et al.* (15) is conserved in higher eukaryotes. A eukaryotic model for the expansion segment D2 (34) proposes the presence of a consensus structure containing 3 major secondary helices. Both the *Drosphila* (33) and *H. momus* models deviate from this in that D2 consists of 2 or 3 secondary helices and a variable number of tertiary helices within those helices (33).

Although the degree of structural conservation in expansion segments is not as great as in core segments, the evidence from a number of different taxa suggests a function for these regions and thus a constraint on sequence evolution of these segments (34, 44). Previous claims that expansion elements are functionless and are tolerated by the ribosome because they do not interfere with translation (9) were made without knowledge of structural data from ascidians, fruit fly and nematodes (7, 33). The expansion domains appear to be coevolving in relation to both size and base composition. Michot and Bachellerie (34) compared the size and base composition of D2 to D8 and demonstrated that all eukaryotes possess a D2/D8 ratio of between 1.25 and 1.5. Although *H. momus* D2 and D8 expansion segments are similar in base composition (72.6 and 70.1% GC, respectively), the ratio of D2/D8 differs significantly from other eukaryotes being approximately 3.5, suggesting that the coevolution of these two expansion segments in *H. momus* might not be coupled in the same manner as in other eukaryotes. Whether size variation is related to base composition, with biased slips or rearrangements occurring at equal rates in different expansion segments, or due to functional interactions of expansion segments within the ribosome (33, 42) has yet to be determined. Therefore, specific reasons for the *H. momus* D2/D8 deviation remain obscure.

Comparison of the expansion segment of the ascidian with other eukaryotes (Table 3), shows that *H. momus* possesses a high GC content as found in other chordates. Limited sequence data from a sea urchin, reveals that the ITSs and D1 segment are also both GC rich (45), suggesting that the origin of GC rich expansion segments arose in a primitive deuterostome, before chordates diverged from each other and echinoderms. A recent phylogenetic survey using expansion segment D3, correlated secondary structure of subdomains within D3 with metazoan evolutionary relationships (35). In the two subdomains (II and IV) where vertebrates have a distinctive sequence and folding pattern (35), the *H. momus* primary sequence and secondary structure fits well with these chordate-specific structures. *H. momus* has a smaller proportion of its 26S rRNA composed of expansion segments than other higher eukaryotes, with the exception of *C. elegans*. It is of interest that both ascidian and nematode embryos develop rapidly and in a mosaic manner (18, 19, 36), however, a correlation between the small 26S rRNA gene and rDNA tandem repeat size of both these organisms and embryonic processes seems unlikely.

ACKNOWLEDGEMENTS

The authors would like to thank the staff of the Heron Island Research Station for all their assistance. This work was supported by an Australian Research Council grant 05-775 to MFL. BMD is supported by a University of Queensland Postgraduate Research Scholarship.

REFERENCES

1. Birnstiel,M.L., Chipchase,M. and Speirs,J. (1971) *Prog. Nucl. Acid Res. Mol. Biol.* **11**, 351-389.
2. Chambon,P. (1975) *Annu. Rev. Biochem.* **44**, 613-638.
3. Long,E.O. and Dawid,I.B. (1980) *Annu. Rev. Biochem.* **49**, 727-764.
4. Fedoroff,N.V. (1979) *Cell* **16**, 697-710.
5. Reeder,R.H. (1984) *Cell* **38**, 349-351.
6. Tautz,D. and Dover,G.A. (1986) *EMBO J.* **5**, 1267-1273.
7. Ellis,R.E., Sulston,J.E. and Coulston,A.R. (1986) *Nucl. Acids Res.* **14**, 2345-2364.
8. Strumph,W.E., Wu,J.R. and Bonner,J. (1979) *Biochemistry* **18**, 2864-2871.
9. Gerbi,S.A. (1985) in R.J. MacIntyre (ed.) *Molecular Evolutionary Genetics*. Plenum, New York, pp. 419-517.
10. Noller,H.F. (1984) *Annu. Rev. Biochem.* **53**, 119-162.
11. Dover,G.A. (1982) *Nature* **299**, 111-117.
12. Clark,C.G., Tague,B.W., Ware,V.C. and Gerbi,S.A. (1984) *Nucl. Acids Res.* **12**, 6197-6220.
13. Atmadja,J., Brimacombe,R. and Maden,B.E.H. (1984) *Nucl. Acids Res.* **12**, 2649-2667.
14. Nelles,L., Fang,B.L., Volkaert,G., Vandenberghe,A. and De Wachter,R. (1984) *Nucl. Acids Res.* **12**, 8749-8768.
15. Michot,B., Hassouna,N. and Bachellerie,J.P. (1984) *Nucl. Acids Res.* **12**, 4254-4279.
16. Berrill,N.J. (1955) *The Origin of Vertebrates*. Oxford Univ. Press, Oxford.
17. Gans,C. and Northcutt,R.G. (1983) *Science* **220**, 268-274.
18. Jeffery,W.R. (1985) *Cell* **41**, 11-12.
19. Whittaker,J.R. (1979) in S. Subtelny and I.R. Konigsberg (eds.), *Determinants of Spatial Organization*. Academic Press, New York, pp. 29-51.
20. Kott,P. (1985) *Mem. Qd. Mus.* **23**, 1-440.
21. Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) *Molecular Cloning. A Laboratory Manual*. Cold Spring Harbor Laboratory, New York.
22. Weeks,D.P., Beerman,N. and Griffith,O.M. (1986) *Anal. Biochem.* **152**, 376-385.
23. Sanger,F., Nicklen,S. and Coulson,A.R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463-5467.
24. Henikoff,S. (1984) *Gene* **28**, 357-359.
25. Gutell, R.R., Weiser, B., Woese, C.R. and Noller, H.F. (1985) *Prog. Nucl. Acids Res. Mol. Biol.* **32**, 156-217.
26. Neefs,J.M., Van de Peer,Y., Hendriks,L. and De Wachter,R. (1990) *Nucl. Acids Res.* **18 Supplement**, 2237-2318
27. Gutell,R.R. and Fox,G.E. (1988) *Nucl. Acids Res.* **16 Supplement**, r175-r269.
28. Gutell,R.R., Schnare,M.N. and Gray,M.W. (1990) *Nucl. Acids Res.* **18 Supplement**, 2319-2335
29. Jaeger,J.A., Turner,D.H. and Zuker,M. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 7706-7710.
30. Cesarone,C.F., Bolognesi,C. and Santi,L. (1979) *Anal. Biochem.* **100**, 188-197.
31. Nisson,P.E., Hickey,R.J., Boshar,M.F. and Crain,W.R. (1988) *Nucleic Acids Res.* **16**, 1431-1452.
32. Gutell,R.R., Noller,H.F. and Woese,C.R. (1986) *EMBO J.* **5**, 1111-1118.
33. Hancock,J.M., Tautz,D. and Dover,G.A. (1988) *Mol. Biol. Evol.* **5**, 393-414.
34. Michot,B. and Bachellerie,J.P. (1987) *Biochimie* **69**, 11-23.
35. Michot,B., Qu,L.H. and Bachellerie,J.P. (1990) *Eur. J. Biochem.* **188**, 219-229.
36. Davidson,E.H. (1986) *Gene Activity in Early Development*. 3rd Edition. Academic Press, New York.
37. John,B. and Miklos,G. (1988) *The Eukaryotic Genome in Development and Evolution*. Allen and Unwin, London.
38. Field,K.G., Olsen,G.J., Lane,D.J., Giovannoni,S.J., Ghiselin,M.T., Raff,E.C., Pace,N.R. and Raff,R.R. (1988) *Science* **239**, 748-753.
39. Wellauer,P.K., Reeder,R.H., Dawid,I.B. and Brown,D.D. (1976) *J. Mol. Biol.* **105**, 487-505.
40. Tautz,D., Tautz,C., Webb,D.A. and Dover,G.A. (1987) *J. Mol. Biol.* **195**, 525-542.
41. Reeder,R.H. (1984) *Cell* **38**, 349-351.
42. Dover,G.A. and Flavell,R.B. (1984) *Cell* **38**, 622-623.
43. Sollner-Webb,B. and Tower,J. (1986) *Annu. Rev. Biochem.* **55**, 801-830.
44. Hancock,J.M. and Dover,G.A. (1988) *Mol. Biol. Evol.* **5**, 377-392.
45. Hindenach,B.R. and Stafford,D.W. (1984) *Nucl. Acids Res.* **12**, 1737-1747.
46. Gorski,J.L., Gonzalez,K.L. and Schmicuez,R.D. (1987) *J. Mol. Evol.* **24**, 236-251.
47. Veldman,G.M., Klootwijk,J., de Regt,V.C.H.F., Planta,R.J., Branlant,C., Krol,A. and Ebel,J.-P. (1981) *Nucl. Acids Res.* **9**, 6937-6952.