J Mol Evol (1995) 41:189–195

JOURNAL OF MOLECULAR EVOLUTION

Molecular Cloning and Complete Nucleotide Sequence of the Repeated Unit and Flanking Gene of the Scallop *Pecten maximus* Mitochondrial DNA: Putative Replication Origin Features

Ali Rigaa,[1,2] Monique Monnerot,[2] Daniel Sellos[1]

[1] Laboratoire de Biologie Marine du Collège de France, BP 225, 29182, Concarneau, France
[2] Centre de Génétique Moléculaire, CNRS, 91190, Gif sur Yvette, France

Received: 20 July 1994 / Accepted: 21 November 1994

Abstract. In the bivalve mollusc *Pecten maximus,* the size of the mitochondrial DNA molecules ranges from 20 to 25.8 kbp. This variability is mainly correlated with the occurrence of a variable domain composed with two to five 1.6-kbp repeated units tandemly arrayed in the genome. DNA fragments spanning the 1,586-base-pair-long repeated element and the nearest flanking gene have been cloned and sequenced. This sequence was analyzed regarding its base composition and potential secondary structures. The repeated unit domain was positioned and oriented with regard to the known flanking gene. It ends 2 base pairs upstream relative to the beginning of the $tRNA^{gly}$ gene. The peculiar properties of the repeated unit were compared with those of the 1,442-bp repeated element found in the mitochondrial genome of the deep sea scallop *Placopecten magellanicus.* This comparison provided evidence for the absence of nucleotide conservation, except for a small sequence engaged in a secondary structure, but argued for a strong pressure maintaining domains with specific nucleotide content. A possible role for the conserved sequence is discussed.

Key words: Mollusc — mtDNA — Repeated unit — Replication origin — $tRNA^{gly}$ gene

Correspondence to: A. Rigaa

Introduction

The "conserved genetic economy" of the mitochondrial genome of vertebrates has not been until now seriously altered–22 tRNA genes, 2 rRNA genes (coding for the small and large ribosomal subunit RNAs), and 13 genes coding for proteins implicated in oxidative phosphorylation metabolism in a covalently closed, circular molecule with a length of from 16.5 to 17.5 kb. MtDNA length variations are now well documented in vertebrate as well as in invertebrate mt genomes. Large-scale size variations of mtDNA between related species, within several species, and in some cases among molecules within an individual are mainly attributed to the existence of different copy numbers of repeated sequences, usually arranged in direct tandem within the mtDNA molecule. (For a review: see Rand 1993.) Repeated sequences are mostly found in a unique region of the molecule which is known for vertebrates (Clayton 1982, 1984) and insects (Clary and Wolstenholme 1985) to contain signature elements which control both replication and transcription. In some cases, increase in length corresponds to duplication of large coding regions. This was shown for the crested newt (Wallis 1987), nematodes (Hyman et al. 1988), and lizard (Moritz and Brown 1986).

In invertebrates, the occurrence of repeated sequences sometimes accounts for unusual lengths of mt genomes: 36 kb for weevils (Boyce et al. 1989), 20 to 42 kb for some species of scallop (Gjevatj et al. 1992).

The repeated sequences exhibit the common feature of being folded in stable secondary structures (Cornuet et

al. 1991). Even if no direct evidence for repeat-sequence implication in replication and/or transcription control mechanisms of mtDNA exists, their occurrence, and their maintenance, give rise to an attractive but purely speculative hypothesis concerning selective advantage or disadvantage for size classes of mt genomes (Rand 1993).

In this paper, we present nucleotide sequence determination and analysis of the domain of *Pecten maximus* mitochondrial DNA encompassing the repeated units. We have characterized the 1,586-nucleotide repeated unit with special attention to secondary structure and possible function in comparison with that of *Placopecten magellanicus* (La Roche et al. 1990), and we have localized the beginning of the coding region of *Pecten maximus* mitochondrial DNA.

Materials and Methods

Pecten maximus scallops were fished on the west coast of Brittany (France). Mitochondrial DNA was extracted from fresh individual gonads according to Rigaa et al. (1993).

Southern Blot and Hybridization. Mitochondrial DNA fragments were transferred overnight from gels to Hybond-N^+ membranes (Amersham) passively in alkaline solution. Then the membranes were washed and dried at 80°C for 2 h. Prehybridization was performed for 1 h, at 65°C, in 0.025 M NaH_2PO_4, pH 6.5, 5 × SSC, 1% sodium sarcosyl, 1 × Denhardt's solution, 0.025 mg/ml salmon sperm DNA. Hybridization was performed in the same medium with radioactive probes (10^6 cpm/ml). Membranes were washed in 2 × SSC, 0.2% SDS, and 1 × Denhardt's solution at room temperature for 30 min, and then in 2 × SSC, 0.1% SDS at 65°C over 45 min. Wet membranes were exposed at −70°C to Kodak film overnight.

Cloning and Sequencing. The 1.6-kb *Ava*I fragment and the adjacent 2.15 kb *Ava*I fragment produced by the digestion of *P. maximus* mtDNA were extracted from a 1% agarose gel in TAE using the Geneclean kit (Bio 101). Then these fragments were blunt-ended with the Klenow fragment of DNA polymerase I and ligated into the *Sma*I restriction site of dephosphorylated pBluescript II KS plus and minus phagemids (Stratagene).

Concerning the 1.6-kb *Ava*I fragment, subclones were made taking advantage of unique *Dra*I and *Eco*RV restriction sites found within it. In addition, to insure the determination of the sequences, deletion clones were obtained by Exonuclease III digestion.

The 2.15-kb *Ava*I cloned fragment was used to generate a subclone by digestion with *Hpa*I (RO). The nucleotide sequences of each end of this insert were obtained by double-stranded DNA sequencing using T7 DNA polymerase.

Computer Analysis. Sequence analysis was accomplished using the GCG program. The "compare" program was run with 21 as a window and 14 for strengency. Stability profiles of duplex DNA were obtained through the B.I.S.A.N.C.E. DNA analysis package (Dessen et al. 1990). The environmental parameter has been set to W = 3.

Results

A previous study on *P. maximus* mtDNA had evidenced the existence of an intra- and interindividual length polymorphism (Rigaa et al. 1993). This length polymorphism was attributed to the presence, in a unique variable domain (VD, Fig. 1), of 1.6-kb-long tandemly repeated units in which *Ava*I cleaves once. After digestion with this enzyme, the occurrence of two copies in the variable domain generates a single 1.6-kb fragment and two other fragments (2.15 and 3.45 kbp) each containing a part of the repeated unit (Fig. 2A). Any additional copy leads to an accumulation of 1.6-kb fragment, as could be seen after hybridization of mt DNA *Ava*I digests from different individuals with the cloned 1.6-kbp insert (Fig. 2B). Moreover, it was found that *Ava*I digestion also generates a 1.6-kb fragment mapped elsewhere in the genome (Fig. 1).

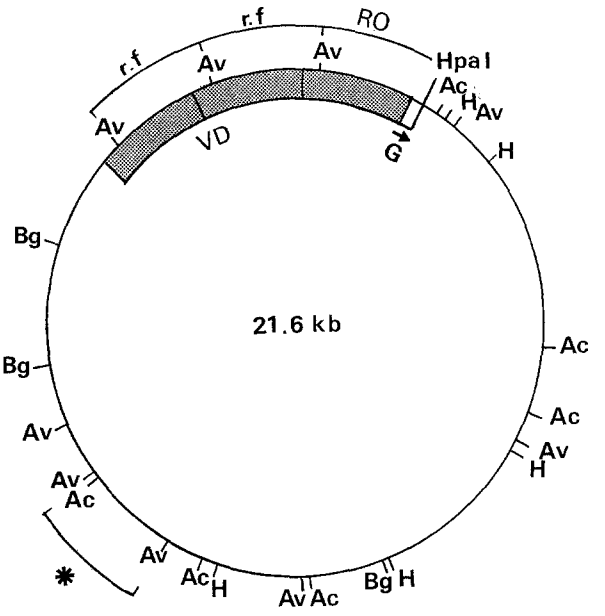

Fig. 1. Restriction map of *Pecten maximus* mtDNA molecule for four endonucleases—*Av: Ava*I, *Ac: Acc*I, *Bg: Bgl*II, and *H: Hin*dIII. This mtDNA molecule size (21.6 kb) is the most common length found among analyzed *P. maximus* mt genomes, with three repeated units; *r.f.* indicates two repeated 1.6-kb *Ava*I fragments. *RO* is the *Ava*I-*Hpa*I fragment. The *shaded arc* in the map indicates the variable domain (*VD*) where the repeated units are tandemly organized. The *asterisk* indicates the other 1.6-kb *Ava*I fragment from the coding part of the genome. *G* and the *arrow* indicate the location of the first coding gene ($tRNA^{gly}$) and its orientation of transcription.

Sequence Determination in the Length-Variable Domain of P. maximus *mtDNA*

The entire nucleotide sequence of the repeated fragment (r.f.) is 1,592 bp long (Fig. 3). The sequences of the two ends of the adjacent *Ava*I–*Hpa*I fragment (RO) were aligned with the r.f. sequence. RO1 is the 125 nucleotides of the left part of the clone, and the RO2 sequence is the 233 nucleotides of the 3′ end of the same clone.

The 125 nucleotides of the RO1 sequence reveal a quite complete similarity with that of the beginning of r.f. However, some dispersed nucleotide substitutions can be noted at positions 8, 22, 47, 70, and 72.

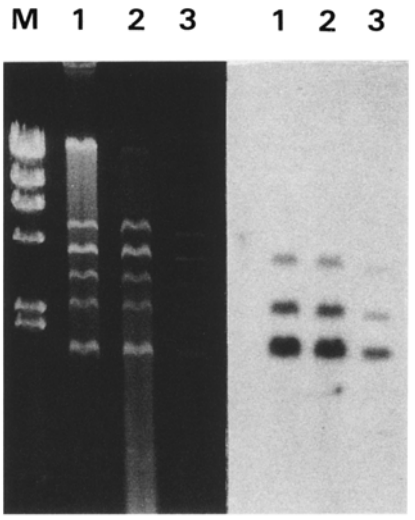

Fig. 2. **A** Electrophoresis of mt DNA *Ava*I digests of three individuals of *P. maximus*. The mt genome lengths are 24.8, 23.2, and 21.6 kb, for lanes *1*, *2*, and *3*, respectively. *M* = lambda *Hin*dIII fragments. **B** Hybridization of A, transferred to Hybond N^+ (Amersham) membrane with the cloned 1.6-kb *Ava*I repeated fragment. The autoradiography displays intense bands for the following *Ava*I fragments—3.45, 2.15, and 1.6 kb—showing homology of the *Ava*I 1.6-kb repeat with the adjacent flanking fragments.

The first 121 nucleotides of RO2 are almost identical to nucleotides 1242–1362 of r.f. Two minor differences have been noted: a transition at position 1330 for r.f. and two deletions/additions at positions 1339 and 1360 for r.f.

In the absence of sequence for two contiguous *Ava*I fragments and according to the above data, the best image of the 1,586-bp repeated-unit sequence of *P. maximus* mtDNA is as follows: Starting at nucleotide 1363 of r.f. it continues through the *Ava*I site at position 1592; then we can assume it goes to nucleotide 1 and ends at nucleotide 1362. The variable domain of *P. maximus* mtDNA is organized with direct, tandemly linked, repeated units. The right part of the variable domain, still under analysis, shows that the coding region of the mitochondrial genome starts immediately, two nucleotides downstream of the repeated unit, with the first-identified gene of *P. maximus* mtDNA—that is, the $tRNA^{gly}$ gene (Fig. 3).

Organization of the Repeated Unit

Using the *Drosophila* mtDNA genetic code, no open reading frames larger than 150 nucleotides were detected on either strand, and we have been unable to find any homology for polypeptides predicted from such open reading frames in protein data libraries.

With dot-plot analyses, one can recognize two peculiar domains: Stretches (four to 11 nucleotides) of A's and T's are very frequent in the middle part of the repeated unit, while G-stretches (three to six nucleotides), accompanied with some stretches of C's, are abundant in the last quarter of the unit.

A surprising feature arose from dot-plot analyses of complementary strands: two 84-bp-long sequences, from nucleotide 201 to nucleotide 284 and from position 845 to 928, show a high level of complementarity (94%), with only five mismatches. These sequences are susceptible to form a secondary structure, exhibiting 75% of A+T over a length of 728 nucleotides and containing most of the homopolymers of A's and T's present in the whole repeated unit. This A+T–rich domain is located exactly in the middle part of the repeated unit. However, the right part of the repeat unit can be subdivided into subdomains: the most interesting one, from nucleotide 1115 to 1288 (Fig. 3), contains 54% G (60% of G+C) and will be referred to as the G-rich domain.

Other potential secondary structures have been deduced from the same analysis. Two sequences (H1 and H2) can be folded in hairpin structures as shown in Fig. 4A. However, in this figure, for H1, only a stem of 13 bp (without the two TATA blocks involved in the stem) is shown to keep a maximum of structural homology with the inverted repeat sequence found in *P. magellanicus*. Stem sequences of both H1 and H2 show a remarkable conservation of the nature of nucleotides and could be aligned with the putative hairpin sequence found in *P. magellanicus* (Fig. 4B). H1 lies between nucleotides 1289 and 1328 in the right domain of the repeated unit next to the G-rich part and 33 nucleotides upstream of the end of the repeated sequence (Fig. 5B). It has also been found in the 2.15-kb adjacent *Ava*I fragment without any modification (positions 48–88). H2 represents a shorter sequence with another putative hairpin structure lying at position 1487–1504 on r.f. (that is, in the left domain of the repeat unit) in the 5′ domain of the repeated unit, 128 nucleotides from the beginning. Thus the two sequences, H1 and H2, are localized separately at each end of the repeat unit (Fig. 5A).

Discussion

Variation in copy number of tandemly repeated sequences is a widespread phenomenon in animal mtDNA and has been reported for five other scallop species out of the seven studied by Gjetvaj et al. (1992). The size of the unit ranges from 400 bp (*Aequipecten opercularis*) to 1,442 bp (*P. magellanicus*).

Comparison of the Repeated Units of P. maximus *and* P. magellanicus

Based on primary sequence, the comparison of *P. maximus* and *P. magellanicus* repeat units (1,586 and 1,442 bp, respectively) exhibits no clear similarity whatever the

```
r.t.    1 CCCGGGGTGT TTAATAAGTG GGTTTATAGG TAGGTCAGTG GAGCCTCTTT AGAGCTAGCG
RO1     1 -------G-- ---------- -A-------- ---------- ------T--- ----------

r.f    61 AAGTCGTAAC TCGTAGTTGG GTTTGGGGGA AGAGCATGGC TTCAAAATTC GTAAGTCAGT
RO1    61 ---------T -T-------- ---------- ---------- ---------- ----------

r.f   121 GCGTGTTTTT CAAGAGGAAA AATGGGGGTG CCACAGGAAT TTACTCACTA AGTACAAAAT
RO1   121 -----

r.f   181 ATATTTTTTG TAAAAAAAGT TTTTTTTTTC CAAAAAAGTC GGGTTTTTTT TGAAAATTTC

      241 TCTTATTTTG TGCAAAATGG GACGTACAAA AAAGGCAAAA AAAACGCGAA AAAAATGTTG

      301 AAAAAACGTT GAAAAAAGGG GTGTTTTTGG AACCTGTAAG TCAAAAGTCT TTTTTAGGAG

      361 TAAAATCGCT TTTTAACAGG AATTTACGTA GTAAAATGGG GGGGTTTTTT TCCAAAAAAG

      421 TTGCAAAAAA TCGTTTTTTT TGATGGTTTT TATAAAAATT AAGGAAAATA AGTGTTTTTT

      481 TTATAAAATT TTATAAAATT TTACTTAGTG CGGGGTTTTG TTAGGGTTTA GTTCTGTTTG

      541 TTTTTTAAAA AAATAAGAAA TTTTTGTAAA AAAAAGTGTC ATTTTTTTAG GAGTTTTTCG

      601 AGGTCTTTCA AGAGGTTCAG GGGAACTATC TTTCTAGCTT AAAATTTTTT ATTTTTTTTT

      661 GTAAAAAAAT TTTTTTTTTT AAGGGGGAGT AAATTTATTG CATTGTAAGT GGGGCACAAA

      721 AACTTGGTTT TAGGGGGTAT AAAAAATCTA AGAATTGGGC AATTTTGCCT ATTAAGAAGT

      781 TTTGTGGAAT ATCTTGAGAA ACAGGGTTTT TTGTTGTTTT TTTTGCCAAA AAAAAAAATT

      841 TTTTTTTTTT TGGCTTTTTT TGTACGTGTC ATTTTGCACA AAATAAGAGA AATTTTCAAA

      901 AAAAACCCTA CTTTTCTAGG AAAAAAAATC TCGTGGGAGA AAGGGGACAT TTACCGAAAA

      961 TAATTTTCGG TGAATGGCCC CTTTGTTGGG GGATTTTTGG GGCATGGCCG AAAATAGCTC

     1021 CGACTAGCGA CACTGTAAAG GGTAAACCCC CTCCCCCAAA AAAATAAATT CCCCTATAAT

     1081 CCTTCCTTTA TCTCCCCAGG AATTTTATTC CAGAGGGAGG GAGAGGGAGG TAGGGGCTCC

     1141 AGTTCTAGGA AGGGGGAGGG GGATAGGGGG TGGGGGAAGA AGAAGTGGTT CGGGGTGTGG

r.f  1201 AGGGAGAGGG AGGGTTTGTA AAAAGAGGTG GGGGGGTTCT AGGGGGAGGA GAATTTTCTC
RO2     1                                                  --------- ----------
                                                             H1
r.f  1261 ATAGCCCCCT AGGATGGGGT GAGGGGGTTA TAGGGGGAAG GGGATATCCA TCCCCTGCTT
RO2    20 ---------- ---------- ---------- ---------- ---------- ----------

r.f  1321 CCCCCTATAC GACTGGTG*A ACATGAACGA TTAGTGGTTT AGTAGAAATT GGCAAGTTGG
RO2    80 ---------T --------G- ---------- ---------* --TTTGTATC TCTGGTGTAA
                                                                [----------

r.f  1381 CAGATGGATG TGTTTGACTT AGGATCAAAA GAGGGGGTTA TAGTAGTGCC CCACTTGTTT
RO2   140 ATAGCATGCG CGCCTTCCAA GCGTGAGGTC TTCTTAGTTG AAGGGGGTGT ATAGGGTAAG
          ---------- tRNA^Gly ---------->
                                                                  H2
r.f  1441 GGGCAGGGTA GTATATGTGT AGTACGTTGG GTTCATGCCC CGCTAGTGGG AAGATGTTTC
RO2   200 AGGCGTTAGT CTATTAAGGG CATCCCGCTG TTAAC

r.f  1501 TTCCCTCCTG CCTAGGTTGT CTAGTAGTTT AGAGATTGAA GTCTGGAGTG TAAAAAACGT

r.f  1561 CCATAGATGT ACTTCAAGGG TTGTTCACCC GGG
```

Fig. 3. Complete nucleotide sequence of the repeated fragment (*rf*) and alignments with *RO1* and *RO2* sequences. *Underlined positions* 201–284 and positions 928–845 correspond to complementary nucleotide sequences, susceptible of forming a long stem of 84 bp. *Bold* sequences (from 1289 to 1329 and from 1487 to 1504) are, respectively, the H1 and H2 putative secondary structures, which can be folded in a hair-pin. *RO1* is the sequence of the first 125 nt of the left part of the *RO* fragment. *RO2* is the 233 nt of the right part of the same clone. Only nucleotides that are different are indicated. The *arrow* indicates the sequence and the transcriptional direction of the tRNA gene which has been identified by both standard cloverleaf structure and derived anticodon (*bold* type).

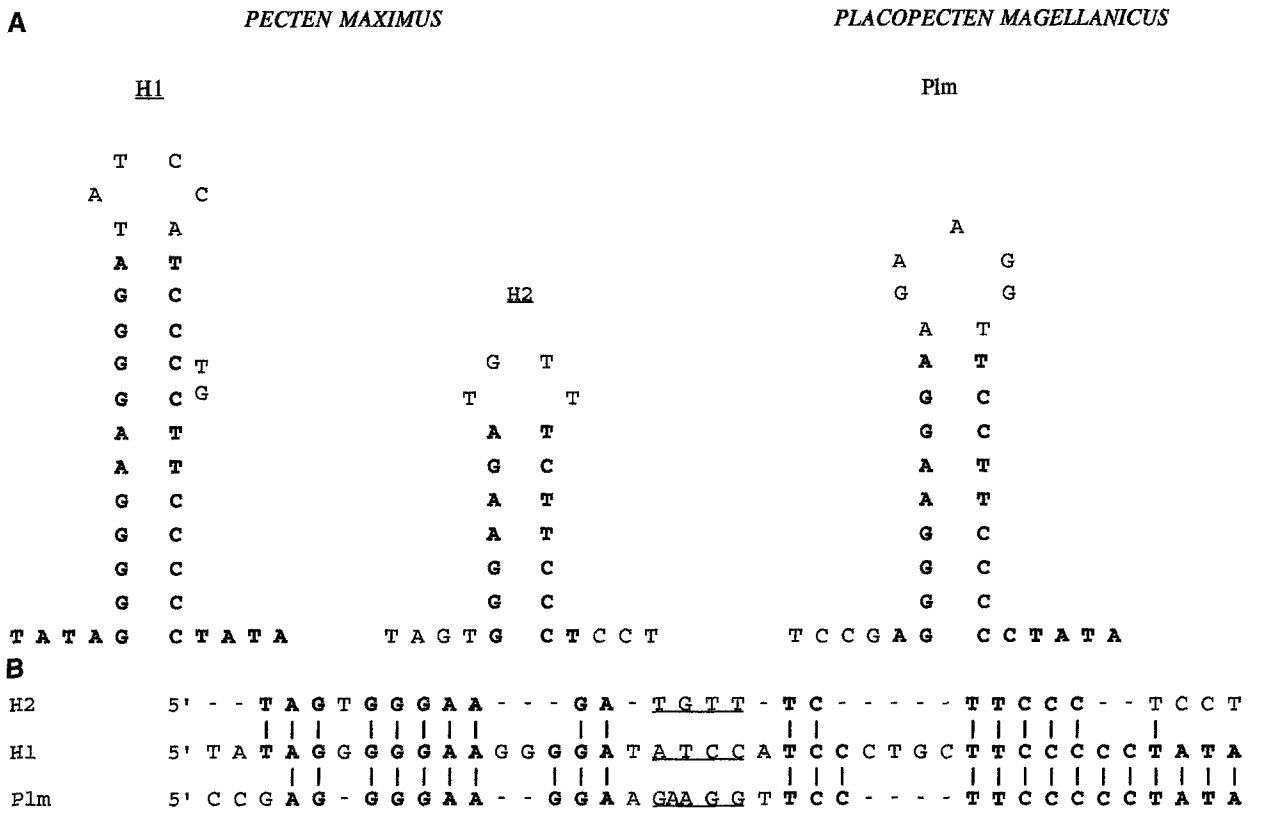

Fig. 4. **A** Putative secondary structures of the H1 and H2 sequences of *P. maximus* that can be folded in a hairpin form. *Plm* represents the complementary strand sequence of the repeated unit of *Placopecten magellanicus* (La Roche et al. 1990). **B** Alignment of the *H1* and *H2* sequences of the repeated unit of *Pecten maximus* with the complementary strand of the repeated unit of *Placopecten magellanicus*. Gaps are introduced for maximal alignment.

strands used, except the sequence of the stem defined for H1. However, some homology could be detected when considering the nucleotide content of different domains (Fig. 5B). Indeed, La Roche et al. (1990) described, in a *P. magellanicus* repeat unit, an A+T–rich region with, upstream, a C-rich region preceded by a potential hairpin structure. Assuming that the orientation of this unit is inverted in comparison with that of *P. maximus,* the structure of the two units, judged on the nucleotide content and on the relative organization of the different domains, is very similar (Fig. 5B). The comparable features are: (1) a large, more or less central, A+T–rich domain (75% over 748 nucleotides for *P. maximus* and 721 nucleotides for *P. magellanicus),* (2) a G-rich domain (54% over 174 nucleotides and 62% over 150 nucleotides for *P. maximus* and *P. magellanicus,* respectively), and (3) sequences, called H1 and Plm for *P. maximus* and *P. magellanicus,* respectively, next to the G-rich domain, that could be folded in hairpin secondary structures and followed by the same "TATA" motif. For *P. maximus,* this TATA sequence is located 33 nucleotides upstream of the beginning of the tRNA gene for glycine. No data were presented concerning the mapping of this set of nucleotides on *P. magellanicus* mtDNA in regard to flanking genes, although the coding region is supposed to be close to the repeated domain (Fuller and Zouros 1993).

This similarity in organization and nucleotide content is fully confirmed by the analysis and comparison of stability profiles for the two units (data not shown). Both profiles present a central domain of low stability (the A+T–rich domain) flanked by two more stable ones.

The absence of sequence similarity must be responsible for the negative cross-hybridization experiments performed by Gjetvaj et al. (1992) between the *P. magellanicus* repeat unit and different scallop species mtDNA. These authors concluded that the *P. magellanicus* unit was unique among the *Pecten* species they analyzed, which included *P.maximus,* although *P. maximus* and *P. magellanicus* repeated units clearly share features.

The contrast between the moderate divergence (around 3%) between two repeated sequences observed for an individual *P. maximus* mtDNA and the high divergence between *P. maximus* and *P. magellanicus* units is probably relevant to a concerted-evolution process. This situation has to be compared with that of *Drosophila* species (*melanogaster* subgroup): the repeat units, which account for the variability of mtDNA size, through variability in the A+T–rich region, have been shown to share more homology within a genome than between genomes (Solignac et al. 1986).

It is also of interest to compare the organization of *P. maximus* repeated units with the 1.2-kb noncoding region

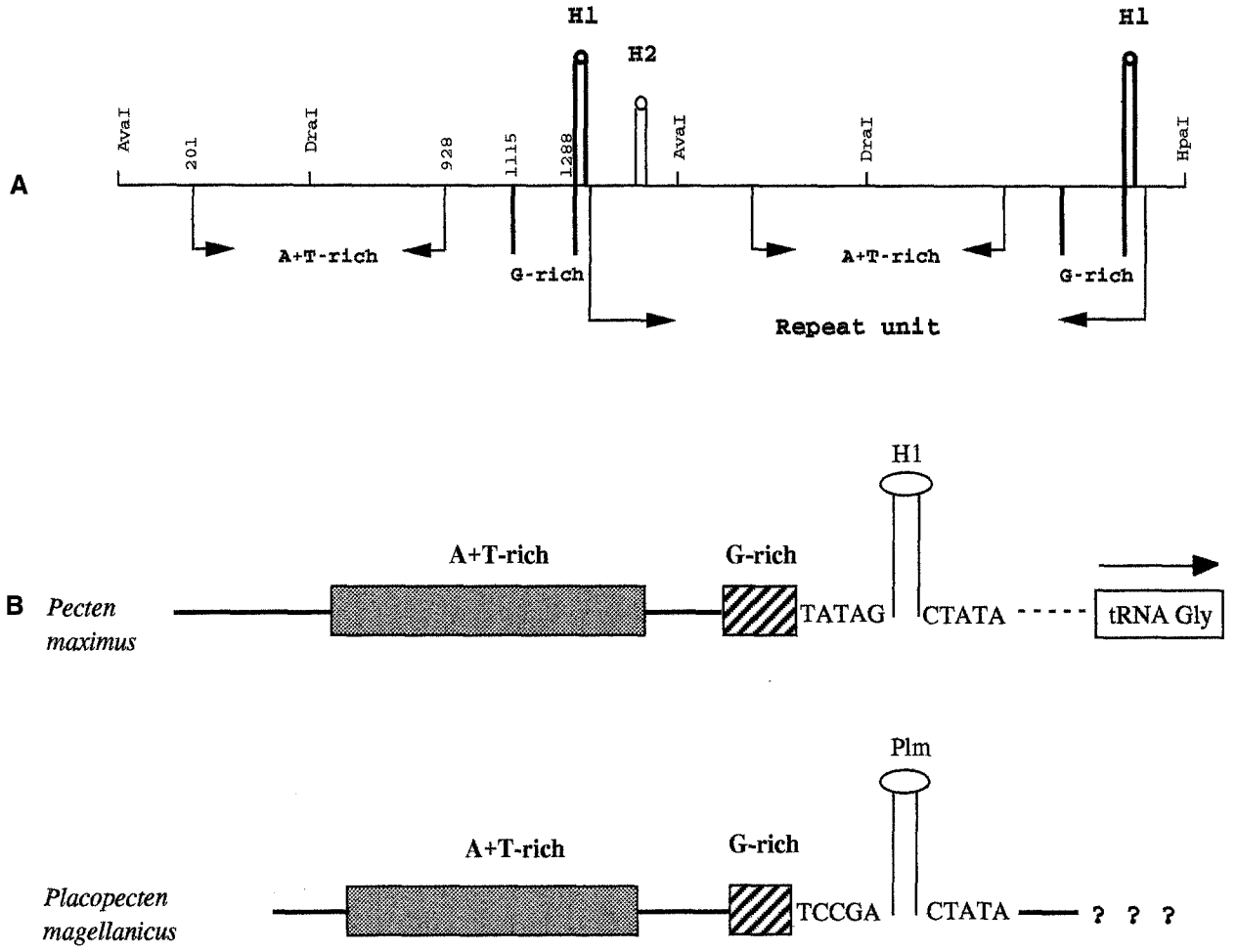

Fig. 5. **A** Restriction map of the 1.6 kg (r.f.) and *AvaI-HpaI* (RO) fragments joined together. *H1* and *H2* identify potential stems and loops forming sequences. **B** Schematic organization of the 1.6-kb and 1.4-kb repeated units of *P. maximus* and *P. magellanicus*. The *gray box* indicates the A+T–rich domain. The *hatched box* indicates the G-rich domain. The *dotted lines* represent the 35 nucleotides upstream to the first gene (tRNA glycine) of *P. maximus*. The *arrow* indicates direction of the transcription. The *question marks* indicate the unknown sequence flanking the repeated unit of the *Placopecten magellanicus* mtDNA (La Roche et al. 1990).

of *Mytilus edulis* mtDNA (Hoffman et al. 1992). When nucleotide content and stability profile are considered, four main domains can also be described, the last one having the highest G/C ratio. As described here, in *P. maximus* and in *P. magellanicus* an A+T–rich domain is flanked by two regions enriched in G+C. This homology of structure between repeat units from two Pectinidae and the noncoding region of *Mytilus edulis* can lead to the hypothesis that length-variable domains of *Pecten* species originate from the amplification of the noncoding region of mtDNA of their common ancestor. This hypothesis does not imply that every species undergoes repeats; indeed, they are absent in some *Pecten* species (Gjetvaj et al. 1992) as well as in *Mytilus edulis*. It does not require, either, a similarity in nucleotide sequence, and it does not preclude other rearrangements such as the ones responsible for present situation in *P. magellanicus:* multiple sites of the variable domain hybridized to the repeat unit, but they seem to contain only part of the repeat unit sequence (Fuller and Zouros 1993).

The inverted repeat (84 bp long) delimiting the 728 bp long A+T–rich domain has never been described in any of the other mitochondrial DNA repeat sequences reported so far (Rand 1993). The presence of this inverted repeat could be an indication of the insertion of a foreign sequence as evidenced in heterochromatin (Miklos et al. 1988, Vaury et al. 1989).

Secondary Structure and Origin of Replication

The lack of genes, the presence of repeats, and the highly stable stem-and-loop structures are consistent with the interpretation of the length-variable domain of *P. maximus* seen as the control region of the molecule. The occurrence of a CSB-2-like sequence (CCCCCTCCCCC at position 1047–1057 Fig. 3) could argue for this (Walberg and Clayton 1981; Zamaroczy et al. 1984). The similarity in the repeated unit organization in *P. maximus* and *P. magellanicus* and the extended identity for the H1 and Plm stems suggest that structural constraints exist on

the repeated unit and also that this domain could have a regulatory role in the two examined mitochondrial genomes.

By carrying out S1 nuclease mapping, La Roche et al. (1990) have demonstrated the extrusion of the inverted repeat of *P. magellanicus* to generate in vitro a cruciform structure. The same was noted by Monforte et al. (1993) for *Drosophila ambigua* mtDNA. These authors concluded that the secondary structure is implied in the origin of replication in accordance with the electron-microscopical observations of the replication origin of *D. yakuba* and *D. virilis* mtDNA (Goddard and Wolstenholme 1980).

A cloverleaf structure is usually associated with hairpins involved in the replication process (Cornuet et al. 1991). Since the sequence lying between H1 and the end of the repeat unit does not exhibit this property, one may think that the $tRNA^{gly}$ which has been located 35 nucleotides downstream of the H1 sequence of the last repeat unit may play a corresponding role.

The "TATA" motif at the 3′ basis of the stem is located precisely 33 nucleotides upstream of the nucleotide that initiates the first coding gene ($tRNA^{gly}$ gene). Moreover, all the other identified genes were found on the same strand of *Pecten maximus* mtDNA (Rigaa et al., personal results). This characteristic reminds the prokaryotic and eukaryotic gene promoters with a "TATA box" located between 10 and 35 nucleotides from the transcriptional initiation site of nuclear genes. A TATA sequence has also been noted near the transcription initiation sites of mtDNA of both vertebrates (Chang and Clayton 1986) and invertebrates (Rand and Harrison 1989; Hoffmann et al. 1992), but its role in the transcription of the mtDNA molecule has still to be proven.

Acknowledgments. The authors wish to thank J.-C. Mounolou for valuable discussions, N. Dennebouy for her excellent technical assistance, and a particularly insightful anonymous reviewer for comments on an earlier version of the manuscript. This work was partially supported by grants from the "Conseil Régional de Bretagne" and from Ifremer (URM 14).

References

Boyce TM, Zwick ME, Aquadro CF (1989) Mitochondrial DNA in the bark weevils: size, structure and heteroplasmy. Genetics 123:825–836

Chang DD, Clayton DA (1986) Identification of primary transcriptional strat sites of mouse mitochondrial DNA: accurate in vitro initiation of both heavy- and light-strand transcripts. Mol Cell Biol 6:1446–1453

Clary DO, Wolstenholme DR (1985) The mitochondrial DNA molecule of *Drosophila yakuba:* nucleotide sequence, gene organization, and genetic code. J Mol Evol 22:252–271

Clayton DA (1982) Replication of animal mitochondrial DNA. Cell 28:693–705

Clayton DA (1984) Transcription of the mammalian mitochondrial genome. Annu Rev Biochem 53:573–594

Cornuet JM, Garnery L, Solignac M (1991) Putative origin and function of the intergenic region between COI and COII of Apis mellifera L. mitochondrial DNA. Genetics 1128:393–403

Dessen P, Fondrat C, Valencien C, Mugnier C (1990) Bisance: French service for access to biomolecular sequence data-bases. Comput Appl Biosci 6:355–356

Fuller KM, Zouros E (1993) Dispersed discrete length polymorphism of mitochondrial DNA in the scallop *Placopecten magellanicus* (Gmelin). Curr Genet 23:365–369

Gjetvaj B, Cook DI, Zouros E (1992) Repeated sequences and large-scale size variation of mitochondrial DNA: a common feature among scallops (Bivalvia: Pectinidae). Mol Biol Evol 9:106–124

Goddard JM, Wolstenholme DR (1980) Origin and direction of replication in mitochondrial DNA molecules from the genus *Drosophila.* Nucleic Acids Res 8:741–757

Hoffmann RJ, Boore JL, Brown WM (1992) A novel mitochondrial genome organization for the bleu mussel *Mytulis edulis.* Genetics 131:397–412

Hyman BC, Beck JL, Weiss KC (1988) Sequence amplification and gene rearrangement in parasitic nematode mitochondrial DNA. Genetic 120:707–712

La Roche J, Snyder M, Cook DI, Fuller K, Zouros E (1990) Molecular characterization of a repeat element causing large-scale size variation in the mitochondrial DNA of the sea scallop *Placopecten magellanicus.* Mol Biol Evol 7:45–64

Miklos GLG, Yamamoto MT, Davies J, Pirrotta V (1988) Microcloning revels a high frequency of repetitive sequences characteristic of chromosome 4 and the b heterochromatin of Drosophila melanogaster. Proc Natl Acad Sci USA 85:2015–2055

Monforte A, Barrio E, Latorre A (1993) Characterization of the length polymorphism in the A+T–rich region of the Drosophila obscura group species. J Mol Evol 36:214–223

Moritz C, Brown WM (1986) Tandem duplications of D-loop and ribosomal RNA sequences in lizard mitochondrial DNA. Science 233:1425–1427

Rand DM (1993) Endotherms, ectotherms, and mitochondrial genome-size variation. J Mol Evol 37:281–295

Rand DM, Harrison RG (1989) Molecular population genetics of mtDNA size variation in crickets. Genetics 212:551–569

Rigaa A, Monnerot M, Le Gal Y, Sellos D (1993) Mapping and repeated sequence organization of mitochondrial DNA in scallop, *Pecten maximus.* Mol Marine Biol Biot 2:218–224

Solignac M, Monnerot M, Mounolou J-C (1986) Concerted evolution of sequence repeats in *Drosophila* mitochondrial DNA. J Mol Evol 24:53–60

Vaury C, Bucheton A, Pelison A (1989) The b heterochromatic sequences flanking the I elements are themselves defective transposable elements. Chromosoma 98:215–24

Walberg MW, Clayton DA (1981) Sequence and properties of the human KB cell and mouse L cell D-loop regions of mitochondrial DNA. Nucleic Acids Res 9:5411–5421

Wallis GP (1987) Mitochondrial DNA insertion polymorphism and germ line heteroplasmy in the *Trutus cristatus* species complex. Heredity 58:229–238

Zamarocsy M, Faugeron-Fonty G, Baldacci G, Goursot R, Bernardi G (1984) The ori sequences of the mitochondrial genome of a wild-type yeast strain: number, location, orientation and structure. Gene 1152:439–457