Molecular Characterization of a Repeat Element Causing Large-Scale Size Variation in the Mitochondrial DNA of the Sea Scallop *Placopecten magellanicus*

*Julie La Roche,* *Marlene Snyder,* *† Douglas I. Cook,* *Kathleen Fuller,* *and Eleftherios Zouros*

*Department of Biology, Dalhousie University, and †Mount St. Vincent University, Halifax, Nova Scotia; and ‡Department of Biology, the University of Crete*

The scallop *Placopecten magellanicus* has the largest reported animal mitochondrial DNA (average 35 kb) and exhibits large inter- and intraindividual length variation owing to the varying copy number of a repeated element. We have characterized the repeat array by using restriction mapping and sequence analysis. The repeated element consists of 1,442 bp flanked on either side by the sequence ACTTTCC in a direct orientation. The array contains two to eight copies of the repeated element arranged in a direct orientation and in tandem. Only complete copies of the element are present in the array. The repeat element contains three regions with characteristic nucleotide sequences: a 10-bp inverted repeat shown to extrude into a cruciform in a supercoiled DNA plasmid, a 120-bp tract rich in G/C (70%) and adjacent to the inverted repeat, and periodically interspersed homopolymer runs of A and T occurring near the middle of the element which induce DNA curvature in dimeric constructs of the element. The element appears to be unique to *P. magellanicus*. The structural properties of the repeat element and its organization in an array of repeats may be important in explaining the generation and maintenance of large-scale mitochondrial DNA size variation observed in many animal species.

**Introduction**

Widespread variation in the size of the mitochondrial DNA (mtDNA) molecule of invertebrate metazoans and poikilotherm vertebrates has been recently documented. The variation occurs among species, between individuals of the same species, and even among molecules within an individual (see refs. in Moritz et al. 1987; Rand and Harrison 1989). The range of sizes, previously reported to be from 14.2 kb to 19.5 kb (Fauron and Wolstenholme 1976), has been extended to 42 kb (Snyder et al. 1987). In addition to variation in size, variability in the gene arrangement relative to mammalian mtDNA has also been reported (Clary et al. 1982; Clary and Wolstenholme 1985; Cantatore et al. 1987; Himeno et al. 1987; Wolstenholme et al. 1987; Hyman et al. 1988; Jacobs et al. 1988). Taken together, these findings have altered considerably the perception of animal mtDNA, which had, as recently as 1985, been called “an extreme example of genetic economy” (Attardi 1985).

We have recently reported on the atypical mtDNA of the sea scallop, *Placopecten magellanicus*, which is the largest metazoan mtDNA (42 kb) described to date and

---

1. Key words: mitochondrial DNA, scallops, repeated elements.

Address for correspondence and reprints: Dr. Julie La Roche, Department of Applied Science, Brookhaven National Laboratory, Upton, New York 11973.

* Mol. Biol. Evol. 7(1):45–64. 1990. © 1990 by The University of Chicago. All rights reserved.

0737–4038/90/0701–0004$02.00
which also exhibits extremes of variability in size from individual to individual (Snyder et al. 1987). Several types of size variation of animal mtDNA have been recognized (see Moritz and Brown 1987 for review). Among those most common are small variations in the number of nucleotides in short homopolymer runs, variations in the copy number of tandemly repeated sequences, and deletions or tandem duplications of large parts of the genome. The variability observed in the scallop mtDNA is of the second type. Varying copy number of a repeated sequence in animal mtDNA is apparently a widespread phenomenon, having been reported in lizards (Densmore et al. 1985), crickets (Harrison et al. 1985), Drosophila (Solignac et al. 1986), nematodes (Powers et al. 1986), newts (Wallis 1987), and fish (Bentzen et al. 1988). A comparison of the characteristics of the repeated arrays in animal mtDNA may be important in explaining the generation and maintenance of this source of genetic diversity.

In the present paper we present both the sequence of the 1.45-kb repeat element from scallop mtDNA and the sequences of the unique fragments that flank the array of repeats. The only other complete sequence of an mtDNA repeat element is that of the cricket, Gryllus firmus (Rand and Harrison 1989). We present evidence that the elements are directly and tandemly repeated, and we give the approximate position of the repeat array in relation to other known functional regions of the molecule. The repeat element exhibits several regions of sequence capable of conferring altered structures on the molecule. The molecular information we present here is essential for understanding the nature of mtDNA size variation in P. magellanicus, and it may also help explain the unusually large size of this molecule.

Material and Methods

DNA Cloning and Sequencing

Collection of scallops and the isolation of mtDNA were essentially done according to a method described by Snyder et al. (1987). Two PstI fragments, one encompassing the repeat element and the second encompassing one of the flanking regions, were cloned in pUC18 plasmid by using techniques described by Maniatis et al. (1982, pp. 390–391). The two PstI fragments were subcloned in M13mp19 in both orientations, and the replicative-form (RF) DNA was isolated using an alkaline lysis method (Maniatis et al. 1982, pp. 90–91). Ordered deletions were made by the method of Henikoff (1984). The RF DNA was linearized by treatment with BamHI and SstI, followed by unidirectional hydrolysis with exonuclease III. The DNA was then treated with mung bean nuclease, followed by a fill-in reaction. Under the chosen experimental conditions, the DNA was hydrolyzed at a rate of 400 bp/min. Samples were collected every 30 s. For each strand, the religated recombinants were selected according to size so that the deletions spanned the entire length of the fragment. Single-stranded templates were prepared and sequenced by the dideoxy chain-termination method (Sanger et al. 1977). The EcoRI/PstI clone containing the other flanking region was cloned and partially characterized by comparing its restriction-enzyme profile to that of the 1.45-kb PstI clone. A portion of the EcoRI/PstI clone, encompassing the region from the EcoRI site to the KpnI site, was subcloned into pUC19 and partially sequenced using the double-stranded dideoxy termination procedure described by Meirendorff and Pfeffer (1987).

S1 Nuclease Assay

Supercoiled M13 RF DNA containing the 1.45-kb PstI fragment was digested at a concentration of 100 µg/ml in 50 mM ammonium acetate, pH 4.6, 50 mM NaCl,
1 mM ZnCl$_2$ with 400 U nuclease S1 (Pharmacia)/ml at 15°C for 2 h. The reaction was terminated by adding excess Tris(hydroxymethyl)aminomethane HCl/ethylene-diaminetetraacetic acid (EDTA) buffer, pH 8.0, followed by phenol extraction. The S1 nuclease-treated DNA was then digested with restriction enzymes that cut only once in the plasmid.

Gel Electrophoresis and Southern Blot Analysis

DNA was digested to completion with restriction endonucleases according to manufacturer's instructions and run on horizontal 0.8% agarose gels in Tris(hydroxymethyl)aminomethane acetate/EDTA buffer, pH 7.8, and then was transferred to nitrocellulose (Maniatis et al. 1982, pp. 382–386). Probes labeled with $^{32}$P were generated by nick-translating whole plasmids or fragments purified from low-melting-point agarose gels. Single-stranded DNA from M13 phage deletion clones was also used as a template in random-primer labeling reactions, and the products were used as hybridization probes. Polyacrylamide gels (6%) were run in Tris(hydroxymethyl)aminomethane borate/EDTA buffer, pH 8.3.

RNA Isolation and End-labeling

RNA was isolated from purified mitochondria by using the guanidine thiocyanate/CsCl gradient method (Chirgwin et al. 1979). The total purified RNA was end-labeled with $^{32}$P-ATP by using T4 polynucleotide kinase (Maniatis et al. 1982, pp. 122–123). $^{32}$P-labeled RNA was used as a probe on nitrocellulose blots of agarose gels of mtDNA restriction digests.

Results

Sequence of the Repeat Element and Organization of the Array of Repeats

Figure 1a gives the cleavage map of the Placopesten magellanicus mtDNA molecule for five restriction enzymes. Fragments II and III (to which we also refer as PstI fragments 1.45 kb and 0.75 kb, respectively) were cloned and used as probes against restriction digests of total P. magellanicus mtDNA. As shown in figure 1b, the 1.45-kb PstI clone hybridized to fragments I–III and VI of figure 1a and the 0.75-kb PstI clone hybridized to fragments II, III, and VI. A subclone of fragment III that does not have sequence identity with fragment II (i.e., does not hybridize to fragment II under high-stringency conditions) also hybridizes to fragment VI. This demonstrates that fragment VI contains sequence identity with both the repeat element and the unique sequence of fragment III which flanks the array of repeats.

Total RNA extracted from mitochondria isolated from adductor muscles of P. magellanicus was used as a probe against a Southern transfer of P. magellanicus mtDNA. The RNA hybridized only to a BamHI/BamHI 3.6-kb fragment (data not shown) that contains fragment V (fig. 1a). Evidence for the location of the large rRNA gene in fragment V was obtained from the hybridization of P. magellanicus mtDNA with an EcoRI/PstI fragment of cloned Paracentrotus lividus (sea urchin) mtDNA containing the large (16S) rRNA gene (fig. 1c). The cloned rRNA gene hybridized to an HpaI/HpaI 1.3-kb fragment only (fragment V in fig. 1A). A second fragment of the sea urchin mtDNA containing the small (12S) rRNA gene did not hybridize to any sequence of the P. magellanicus mtDNA molecule.

The DNA sequence of the 1.45-kb PstI repeat unit (fragment II) is given in figure 2A. To determine the extent of sequence overlap between the repeat unit and its
FIG. 1.—a, Restriction map of the *Placopesten magellanicus* mtDNA molecule with four repeated elements. This represents the most common type of molecule in natural populations. The *EcoRI* site at 12 o'clock is used as the reference origin. Fragments of interest are identified by roman numerals on the outside of the circular map. Thick portions on the map identify those parts of the genome whose DNA sequences appear in fig. 2. The hybridization patterns between fragments I–VI are indicated on the inside of the circular map. Fragment IV has sequence similarity with clone pdyHB from *Drosophila yakuba* (from Dr. D. Wolstenholme of the University of Utah), encoding a subunit of the respiratory chain NADH dehydrogenase (ND2, partial), cytochrome oxidase subunits I and II, ATPase subunits 8 and 6, and cytochrome oxidase subunit III (partial) (Clary and Wolstenholme 1983). Fragment V contains the gene for the 16S rRNA. B = *BamHI*; E = *EcoRI*; H = *HpaI*; S = *SphI*; P = *PstI*. b, Autoradiograms of hybridizations of the 0.75-kb *PstI* fragment (fragment III) and of the 1.45-kb *PstI* fragment (fragment II) to restriction digests of mtDNA from *P. magellanicus*. The DNA was electrophoresed on a 0.8% agarose gel, and a double transfer was obtained by placing the gel between two nitrocellulose membranes (Maniatis et al. 1982).
flanking regions that is implied by the hybridization data of figure 1b, we sequenced the 0.75-kb PstI fragment (fragment III) and characterized the 2-kb EcoRI/PstI fragment (fragment I) by restriction and sequence analysis. The DNA sequence of fragment

were done at high stringency. Lane A, EcoRI; lane B, BamHI; lane C, PstI; lane D, EcoRI/BamHI; lane E, EcoRI/PstI; lane F, BamHI/PstI; lane G, EcoRI/BamHI/PstI. The top arrow corresponds to a 3.6-kb BamHI/PstI fragment that encompasses fragment VI of fig. 1a. The second, third, and fourth arrows correspond to fragments I, II, and III of fig. 1a, respectively. Numbers in the far-right-hand side are molecular sizes in kilobases. c, Localization of the 16S rRNA gene in the Placopecten magellanicus mtDNA molecule. Digests of P. magellanicus mtDNA were run on agarose gels, transferred to nitrocellulose membranes, and probed with a clone from the Paracentrotus lividus mtDNA containing the 16S rRNA gene (Dr. P. Cantatore, Universita di Bari). Shown bands correspond to fragments in fig. 1a. The left panel is an ethidium bromide-stained gel. Lane 1, 1-kb ladder marker (BRL); lane 2, PstI/SphI digest; lane 3, BamHI/PstI digest; lane 4, HpaI digest. The right panel is an autoradiogram of the same gel after being probed with the P. lividus mtDNA probe (medium-stringency hybridization). Lane 2, Hybridization to an SphI 8.6-kb fragment; lane 3, hybridization to a BamHI 4-kb fragment; lane 4, hybridization to an HpaI 1.3-kb fragment. Both the SphI 8.6-kb and the BamHI 4-kb fragments contain the HpaI 1.3-kb fragment which is shown in fig. 1 as fragment V. The arrow marks the position of the HpaI 1.3-kb fragment in both the ethidium bromide-stained gel and in the autoradiograph. Hybridization in lane 1 (right panel) is due to sequence similarity between the cloning vector and the 1.6-kb marker (BRL product profile).
III, shown in figure 2B, revealed that this fragment is 746 bp long and that there is complete identity in the sequence of the first 513 bp of fragments II and III. The identity stops abruptly at bp 514.

The lack of hybridization between fragments I and III implies no strong sequence similarity between fragment I and the first 513 nucleotide base pairs of the repeat unit that are shared by fragments II and III. Sequence similarity between fragments I and II must, therefore, be confined to the last two-thirds of the length of the repeat unit. Indeed, both the relative positioning of cleavage sites in fragment I for the enzymes StuI, SmaI, StyI, and KpnI and a site sensitive to S1 nuclease treatment (fig. 2C) corresponded to those in the repeat element between pb 658 and 1449 (fig. 2A). These observations establish that fragment I contains the sequence of fragment II from the end (position 1449) up to the KpnI site at position 657 (fig. 2A). To determine how much farther the repeated sequence extends into fragment I, we obtained the sequence of 185 nucleotide pairs from the KpnI site toward the EcoRI end of the fragment. The sequence, given in figure 2C, shows a nearly complete identity between positions 1–
mtDNA Size Variation in Scallops

157 and positions 507–663 in figure 2A. Alignment of the two sequences revealed that the nucleotide base at position 522 and the nucleotide bases at positions 564 and 565 of the sequence in figure 2A are missing in the sequence in figure 2C (positions 16 and 58–59, respectively). The sequence identity between fragments I and II ended abruptly at position 1 of figure 2C.

The data in figures 2A–2C allow the construction of the array containing the repeat elements, as shown in figure 3. The nucleotide sequence ACTTTCC is present in all three sequences. In figure 2C, it marks the beginning of the identity between fragment I and II, and in figure 2B it marks the end of the identity between fragments II and III. The sequence analysis from both flanking regions (fragments I and III) indicates that the repeat unit starts at bp 514, continues through positions 1449, then goes to bp 1, and ends at bp 506. Both the array as a whole and each individual sequence of 1,442 bp is flanked on either side by the direct repeat ACTTTCC (positions 507–513 in fig. 2A). For an array containing the 1,442-bp sequence n times, there will be n + 1 copies of the oligonucleotide. In our survey of natural populations n ranged from two to eight.

Further evidence that the elements are directly repeated within the array was initially obtained by treating six scallops’ mtDNA with Smal, which cuts ~500 bp away from one of the physical ends of the repeat element (position 507 in fig. 2A). The presence of inverted repeat elements within the array would generate fragments larger and smaller than 1.45 kb. In all animals, we observed only one band, which corresponded to the 1.45-kb fragment and thus indicated that the elements were arranged as direct tandem repeats. A large survey of mtDNA from 300 individual animals treated with Stul (representing a collection of ~1,200 elements) and from 30 animals treated with PstI produced the same result. Only one exception was observed, from an Stul digestion which produced a 2.8-kb fragment; this is consistent with a loss of the site in one internal element.

Physical Features of the Repeat Element

Search using the yeast, maize, and mammalian mtDNA codes identified several small open reading frames (of which the largest was 369 bp long) in the sequence of the repeated element. Neither these sequences nor the sequence of any other segment of the repeated element (and the sequences that flank it) showed significant similarity with known nucleic acid or protein sequences. However, the repeat element contains several unusual structural features: (1) There is a large inverted repeat of 10 nucleotides occurring at positions 697–706 and 712–721 (fig. 2A) that may allow for the extrusion of a cruciform with a stem of 10 bp and a loop of five nucleotides. (2) The region immediately adjacent to the inverted repeat is 70% G/C rich over 110 bp, in contrast with the 40% G/C content of the whole PstI 1.45-kb element. (3) The two ends of the PstI 1.45-kb fragment (positions 1–200 and 1100–1449) contain several short homopolymer runs of A’s and T’s, occurring at a periodicity of ~10 bp, that may induce physical bending of this DNA fragment.

Formation of a Cruciform

The extrusion of an inverted repeat to form a cruciform in vivo depends primarily on its stability in supercoiled DNA molecules. To investigate the possibility that the cruciform is extruded in vivo (fig. 4a), we carried out S1 nuclease mapping experiments on RF M13mp19 DNA that contained one copy of the 1.45-kb PstI fragment. The restriction map of this construct is shown in figure 4b. The change in mobility of the
<table>
<thead>
<tr>
<th>10</th>
<th>20</th>
<th>30</th>
<th>40</th>
<th>50</th>
<th>60</th>
<th>70</th>
</tr>
</thead>
<tbody>
<tr>
<td>CTCAGGCTT</td>
<td>TTTTTTTT</td>
<td>CCCGGTTTTA</td>
<td>TTACCGCCCTA</td>
<td>ATAAAAATGCG</td>
<td>ATAAAAAGCCG</td>
<td>CTCTGGAAAA</td>
</tr>
<tr>
<td>80</td>
<td>90</td>
<td>100</td>
<td>110</td>
<td>120</td>
<td>130</td>
<td>140</td>
</tr>
<tr>
<td>ATAAAGAAAA</td>
<td>ACGGCCAACG</td>
<td>GCAATTTTGG</td>
<td>TTTAAAAAGG</td>
<td>AGCTTTTTTT</td>
<td>CGAAMAAAAA</td>
<td>AAATAAAAAC</td>
</tr>
<tr>
<td>150</td>
<td>160</td>
<td>170</td>
<td>180</td>
<td>190</td>
<td>200</td>
<td>210</td>
</tr>
<tr>
<td>TTAACGCTCGC</td>
<td>CGAACAAAAA</td>
<td>AATTCTTTTA</td>
<td>TTTTGTTCGA</td>
<td>AAATTTTTGA</td>
<td>ATGAACTTGT</td>
<td>CTGAAACGAA</td>
</tr>
<tr>
<td>220</td>
<td>230</td>
<td>240</td>
<td>250</td>
<td>260</td>
<td>270</td>
<td>280</td>
</tr>
<tr>
<td>GCACGTTTTTA</td>
<td>TAAACATATA</td>
<td>CCAAGGCACGA</td>
<td>CTTAGGGGAA</td>
<td>AAAAAGCACTT</td>
<td>ATTTTTTTGCG</td>
<td>CACTTTTTCG</td>
</tr>
<tr>
<td>290</td>
<td>300</td>
<td>310</td>
<td>320</td>
<td>330</td>
<td>340</td>
<td>350</td>
</tr>
<tr>
<td>AAAAAAGCCTGTCAGAATGCGCTTGG</td>
<td>CTGCCGCTCAGAATGG</td>
<td>AGCTTTAGTGCGTTTTT</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>360</td>
<td>370</td>
<td>380</td>
<td>390</td>
<td>400</td>
<td>410</td>
<td>420</td>
</tr>
<tr>
<td>ACTCGGAGG</td>
<td>TAAAGTTAAA</td>
<td>CTCCGCCCAC</td>
<td>TCCCGTTTAC</td>
<td>CACGGAGGG</td>
<td>GAAATAATCC</td>
<td>CACCTTTGCAG</td>
</tr>
<tr>
<td>430</td>
<td>440</td>
<td>450</td>
<td>460</td>
<td>470</td>
<td>480</td>
<td>490</td>
</tr>
<tr>
<td>GCTTAAAGCC</td>
<td>TAAACCTTTT</td>
<td>ACTACACTAC</td>
<td>CTCGACCCCC</td>
<td>CATTCCTACT</td>
<td>GCCCAGCCCC</td>
<td>TTCTCTCAC</td>
</tr>
<tr>
<td>500</td>
<td>510</td>
<td>520</td>
<td>530</td>
<td>540</td>
<td>550</td>
<td>560</td>
</tr>
<tr>
<td>AAGGCTAGCA</td>
<td>CGGGCACTTTG</td>
<td>TCAGCAATTTT</td>
<td>TCACGCTTTGG</td>
<td>CTGACCTCAG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>570</td>
<td>580</td>
<td>590</td>
<td>600</td>
<td>610</td>
<td>620</td>
<td>630</td>
</tr>
<tr>
<td>CCGGCAAAGG</td>
<td>AAAATCTATT</td>
<td>GCCGGCTCAGC</td>
<td>AAGACTACAG</td>
<td>TCGAGTAACTA</td>
<td>CTATAAATAC</td>
<td>CAAAAGAAC</td>
</tr>
<tr>
<td>640</td>
<td>650</td>
<td>660</td>
<td>670</td>
<td>680</td>
<td>690</td>
<td>700</td>
</tr>
<tr>
<td>TCTACCAGC</td>
<td>CCGTCTCACC</td>
<td>ACGAGCTGTT</td>
<td>AGCAAGCTT</td>
<td>TCTTAAACAC</td>
<td>CTTCTTTCG</td>
<td>CTATAACCG</td>
</tr>
<tr>
<td>710</td>
<td>720</td>
<td>730</td>
<td>740</td>
<td>750</td>
<td>760</td>
<td>770</td>
</tr>
<tr>
<td>TCGGCTCGCC</td>
<td>TCGGCTCGCC</td>
<td>CTCTGTTTTC</td>
<td>ACCGCTTATCA</td>
<td>ACCGCTTAT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>780</td>
<td>790</td>
<td>800</td>
<td>810</td>
<td>820</td>
<td>830</td>
<td>840</td>
</tr>
<tr>
<td>CTTTCCTTTC</td>
<td>CCGTCTTACG</td>
<td>CCGGCTTACG</td>
<td>CCTAACTTGG</td>
<td>TGGTCTTCCC</td>
<td>CCTGCTTCCG</td>
<td>CCTGCTTGG</td>
</tr>
<tr>
<td>850</td>
<td>860</td>
<td>870</td>
<td>880</td>
<td>890</td>
<td>900</td>
<td>910</td>
</tr>
<tr>
<td>TOGCCCTGCA</td>
<td>CGAACCTGTC</td>
<td>TCTACTCCCC</td>
<td>TCTTTGGCTA</td>
<td>ATCCTGTTCC</td>
<td>ATCCTGTTCC</td>
<td></td>
</tr>
<tr>
<td>920</td>
<td>930</td>
<td>940</td>
<td>950</td>
<td>960</td>
<td>970</td>
<td>980</td>
</tr>
<tr>
<td>TTTTTTTTTTGGG</td>
<td>TCTGCAGGCA</td>
<td>CTCCGCCGGG</td>
<td>GAGACCCAGA</td>
<td>TTTCTTTTGGG</td>
<td>TGGGGAAGCTG</td>
<td></td>
</tr>
<tr>
<td>990</td>
<td>1000</td>
<td>1010</td>
<td>1020</td>
<td>1030</td>
<td>1040</td>
<td>1050</td>
</tr>
<tr>
<td>GCTTTAAATT</td>
<td>TTTACGTTCC</td>
<td>CCGGGGCGGA</td>
<td>CCGGCGATA</td>
<td>TCTTTTTAATC</td>
<td>TAAATTTTTT</td>
<td></td>
</tr>
<tr>
<td>1060</td>
<td>1070</td>
<td>1080</td>
<td>1090</td>
<td>1100</td>
<td>1110</td>
<td>1120</td>
</tr>
<tr>
<td>GCTGCCGTAC</td>
<td>CCAAAGCTTTA</td>
<td>TTGGCTTCCG</td>
<td>CCGGGCTGCA</td>
<td>GCCTTTTCTG</td>
<td>TCCAGAAAAAC</td>
<td>ACGCAGATTTA</td>
</tr>
<tr>
<td>1130</td>
<td>1140</td>
<td>1150</td>
<td>1160</td>
<td>1170</td>
<td>1180</td>
<td>1190</td>
</tr>
<tr>
<td>CTTTTGTTTTTTA</td>
<td>CTTTAATATTG</td>
<td>GCAGAACGGCA</td>
<td>TCCTCTGACG</td>
<td>GCTTTTTTTTAC</td>
<td>GCTTTTTTTTAA</td>
<td>AACTTTTTTA</td>
</tr>
<tr>
<td>1200</td>
<td>1210</td>
<td>1220</td>
<td>1230</td>
<td>1240</td>
<td>1250</td>
<td>1260</td>
</tr>
<tr>
<td>TCGGATAAGTAA</td>
<td>AAGCAGGTTG</td>
<td>TTAACGCTCAG</td>
<td>ATTTTTTTTTT</td>
<td>ATTTTTTTTTT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1270</td>
<td>1280</td>
<td>1290</td>
<td>1300</td>
<td>1310</td>
<td>1320</td>
<td>1330</td>
</tr>
<tr>
<td>TTTGACCCGGG</td>
<td>GTTTTTTTTG</td>
<td>ACTTTTTTTT</td>
<td>TTTTTTTTTT</td>
<td>TTTCTTTTGGG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1340</td>
<td>1350</td>
<td>1360</td>
<td>1370</td>
<td>1380</td>
<td>1390</td>
<td>1400</td>
</tr>
<tr>
<td>AAAAAATACCA</td>
<td>TTTTTTTTTG</td>
<td>TTTTTTTTTT</td>
<td>TTTTTTTTTT</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1410</td>
<td>1420</td>
<td>1430</td>
<td>1440</td>
<td>1450</td>
<td>1460</td>
<td>1470</td>
</tr>
</tbody>
</table>
FIG. 2.—A, Nucleotide sequence of the PstI 1.45-kb fragment (fragment II in fig. 1a). Shown are the restriction sites of selected enzymes, the physical ends of the repeat (in three-sided boxes), the seven-nucleotide flanking sequence (in the enclosed box; positions 507–513), the inverted repeat sequence at positions 697–721 (enclosed in a dashed-line box), the G/C-rich region (bracketed with arrows; positions 723–838), and interspersed homopolymer runs of A and T (underlined) [runs of A and T beyond position 286 or before position 1185 were not underlined on the arbitrary criterion of being separated from the main concentration of runs by >40 bp. The exclusion of these few runs does not affect in any substantial way the ability of runs to induce DNA bending. B, Nucleotide sequence of the PstI 0.75-kb fragment (fragment III in fig. 1a). Notation is as in panel A. The sequence is identical to that in panel A until position 513. C, Composite of restriction-map data and nucleotide sequence of the EcoRI/PstI 2-kb fragment (fragment I in fig. 1a). Part of the fragment starting at position −24 and ending at the KpnI site was sequenced. The part from KpnI to the end of the fragment (PstI site) was mapped by restriction analysis. Each dash from position 158 to the end corresponds to a nucleotide base. Single stars indicate restriction sites. The five stars marking the position of S1 nuclease cleavage correspond to the five nucleotide bases forming the loop of the cruciform in fig. 4a. Numbering of the sequence corresponds to the beginning of identity with the sequence in fig. 2A, at position 507. The restriction map, the S1 nuclease site, and the nucleotide sequence from position −30 to position 158 indicate strong sequence similarity between fragment I, from position 0 to the end, and fragment II from position 507 to position 1449 (fig. 2A). The dashes at positions 16 and 58–59 indicate nucleotide bases that are present in fragment II but that were not found in fragment I.
main band of the plasmid after treatment with S1 nuclease indicated that the supercoiled plasmid is linearized by the enzyme (fig. 4c, first panel). The results from digests of the linearized plasmid with three restriction enzymes (fig. 4c, second panel) indicated that S1 nuclease cleaves at a single site within the 1.45-kb \textit{PstI} insert and at a position corresponding to that of the inverted complementary sequence, as predicted from the restriction map of the plasmid (fig. 4b). These results indicate that the sequence of the 10-bp inverted repeat allows for stable extrusion of a cruciform in a supercoiled \textit{mtDNA} molecule.

**Sequence-induced Curvature in the Repeat Element**

Small homopolymer runs of A's and T's as well as more complex DNA sequences can physically induce curvature in a short DNA sequence (Marini et al. 1982). The A/T-rich region of the repeat element (positions 1-200 and 1100-1449) contains several such homopolymer runs and shows some similarity with sequences of kinoplast DNA from \textit{Leishmania tarentolae}, in which the presence of physically bent DNA has been demonstrated (Wu and Crothers 1984). The possibility that the region of A and T homopolymers contained in the repeat element is bent was examined by constructing a dimer of the 1.45-kb \textit{PstI} fragment in pUC18 (fig. 5). The construct was subsequently digested with \textit{PstI} and \textit{StyI}, two enzymes for which there is a single site in the element. It can be seen from figure 5 that both enzymes generate fragments of equal length containing identical sequences, but in a different order. In the \textit{PstI} digest, the homopolymer runs of A and T are located at the end of the 1.45-kb fragment. However, the same homopolymer runs are located close to the middle of the 1.45-kb fragment generated from the \textit{StyI} digest. Figure 5A shows the migration pattern of the \textit{PstI} and \textit{StyI} fragments in agarose gels, which is consistent with the true size of

![Diagram](image-url)
the fragments. On a 6% acrylamide gel (fig. 5B), the migration of the 1.45-kb \textit{Sty}I fragment is much slower than that of the \textit{Pst}I 1.45-kb fragment, giving an apparent size of \(~2\) kb. These results are consistent with current views that a centrally located sequence-induced bend retards the migration of a DNA fragment in polyacrylamide gels and gives a spuriously large size estimate for the fragment (Wu and Crothers 1984). This experiment, therefore, indicates that a sequence-induced curvature of the DNA is likely to occur in the area of the repeat element containing the A and T homopolymers.
FIG. 4.—a, Base pairing of the inverted repeat sequence (positions 697–721 in fig. 2A) leading to cruciform extrusion in supercoiled DNA. b, Restriction map of the plasmid used for the S1 nuclease mapping experiment presented in panel c. The plasmid contains one copy of the repeated 1.45-kb PstI fragment of *Placopecten magellanicus* mtDNA (shown with a thin line). The map shows the predicted sizes of the restriction fragments (in kilobases) and the position of the S1 nuclease site corresponding to the site of the cruciform within the repeated element. c, S1 nuclease mapping of the extruded cruciform in a supercoiled M13mp19 RF containing one copy of the 1.45-kb PstI fragment (fig. 4b). The left-hand-side panel shows a 0.8% agarose gel of the plasmid DNA before (Control) and after (S1) treatment with S1 nuclease. In the control lane most of the plasmid DNA appears in the supercoiled configuration (S), although a small fraction of the DNA is present as a relaxed circle (R). After treatment with S1 nuclease, the supercoiled DNA disappears and most of the DNA is present as linear DNA (L). The right-hand-side panel shows the supercoiled DNA linearized by digestion with S1 nuclease and then digested with three restriction enzymes which cut only once in the plasmid. Note that each restriction enzyme yields two separate bands. The restriction fragments of the digests are in agreement with the predicted restriction map.

The Repeat Element Is Not Present in the mtDNA of Other Scallop Species

The cloned *PstI* 1.45-kb fragment was hybridized, under medium-stringency conditions, against digested mtDNA from four scallop species: *Argopecten irradians*, *Chlamys islandica*, *C. opercularis*, and *Pecten maximus*. These were chosen as representatives of the closest living relatives of *P. magellanicus*, which is the only extant species in the genus *Placopecten*. The last three species exhibit discrete size variation
due to repeated sequences (B. Gjetvaj and E. Zouros, unpublished results). No hybridization was observed (data not shown), suggesting that the *P. magellanicus* repeat element is unique to this species.

**Discussion**

In the present report we have (1) presented both the sequence of an mtDNA repeat element in the scallop *Placopecten magellanicus* and the sequences of the unique regions that flank the array of repeats, (2) determined the organization of the array of repeats in the molecule, (3) determined the array’s position relative to the 16S rDNA, and (4) examined the potential for altered DNA conformations induced by the presence of unusual sequences present in the repeat element.

The repeat element is 1.45 kb, within the reported size range of animal mtDNA repeated sequences [64 bp in lizards (Densmore et al. 1985) to 3 kb in a nematode (Hyman et al. 1988)] and occurs in direct tandem orientation, as is the case for all reported cases of animal mtDNA repeat elements. In crickets and the nematode the array consists of two or more “complete” repeats flanked (at one or both ends) with “incomplete” runs of the repeat. Our sequence data suggest a different model for the organization of the array of repeats: that of only long, complete repeats flanked on
either side by a 7-bp oligonucleotide. Such an organization has not been observed previously.

Both scallop and nematode mtDNA contain sequences outside the array of repeats that hybridize with the repeat element. In the nematode these fragments contain part of the element’s sequence in inverted orientation (Hyman et al. 1988). In the scallop, a 1.6-kb fragment hybridizing to the repeat element exists across from the array of tandem repeats (fragment VI in fig. 1a) in a region where no size variation has been observed. Preliminary evidence (data not shown) suggests that this fragment has similarity with sequences unique to one of the flanking regions (fragment III in fig. 1a) while it is missing part of the sequence from the repeat element (fragment II in fig. 1a).

In lizards a short 64-bp tandemly repeated sequence is shared by several related species (Densmore et al. 1985). In crickets the repeated sequence is very similar in two closely related species (Rand and Harrison 1989; D. M. Rand, personal communication), and restriction analysis indicates that the same is true in the melanogaster
subgroup of Drosophila (Solignac et al. 1986). As noted, the mtDNA of P. magellanicus is unusual among metazoan mtDNAs because of both its exceptionally large size and the range over which this size varies among individuals. The mtDNA of the bay scallop, Argopecten irradians, is 16.2 kb long and lacks the repeat element of the P. magellanicus mtDNA. As well, three other tested species of scallops lack the repeat element. An ancient duplication of an mtDNA of standard size followed by divergence of the duplicated sequences may explain the size of the P. magellanicus mtDNA (Snyder et al. 1987), but the hybridization patterns of A. irradians and P. magellanicus mtDNA molecules neither exclude nor substantiate this hypothesis. What can be said at present is that the events that are responsible for both the enlargement of the P. magellanicus mtDNA molecule and the appearance of its repeat element occurred after the divergence of extant pectinid species from their common ancestor. That the repeat element of the P. magellanicus mtDNA appears to be unique to this species need not be construed as evidence for a recent origin of the element, as scallop species are separated by very long periods of time [Placopecten and Argopecten can both be traced to the Miocene; other species are much older (Hertlein 1969)].

The sequence data have revealed several structural features of the repeat unit, in addition to the high degree of regularity in the repeat array. Inverted complementary sequences have been reported in the control region of several organisms (Clayton 1982; Wong et al. 1983) and in the origin of light strand replication in the mtDNA of two species [mouse and human (Clayton 1982)]. We have observed a short inverted complementary sequence within the mtDNA repeat element, and a similar sequence has been reported in the only other known animal mtDNA repeat sequence, that of the cricket (Rand and Harrison 1989). In the scallop the inverted sequence is 10 bp long, separated by 5 bp, but in the cricket it is present as a dyad of 14 bp. We have shown that the inverted sequence can be extruded as a cruciform in supercoiled DNA molecules and that it thus may provide a landmark involved in the generation of mtDNA length variation. S1 nuclease cleavage does not by itself establish the existence of a cruciform structure, but the mapping of the cleavage site at the loop position predicted from the sequence is strong evidence for its presence in the expected position.

We have employed the circular permutation test (Wu and Crothers 1984) to examine the possibility that the presence of A and T homopolymer runs in the interior of the element may induce DNA curvature. The results demonstrate that DNA curvature occurs in the expected region of the element. The accumulating reports on sequence-induced DNA curvature indicate that this phenomenon is common. Physically bent DNA has been reported in kinetoplast DNA (Marini et al. 1982), in the autonomously replicating yeast sequences (Williams et al. 1988), and in the lambda origin of replication (Zahn and Blattner 1985). The secondary structures that curved sequences confer on the DNA may play diverse roles (Dean et al. 1987). Laundon and Griffiths (1988) have shown that curved DNA segments are preferential ends of supercoiled DNA molecules and that they thus uniquely orient interwound supertwisted DNA. A curved kinetoplast DNA sequence has been shown to specify a unique binding site for a nicking enzyme from Crithidia fasciculata (Linial and Shlomai 1988).

Length variation may result from errors of replication, imprecise recombination, or transposition. Errors of replication through slippage and mismatching of single strands (Streisinger et al. 1966) have been invoked as a possible mechanism for large duplications and deletions in lizard mtDNA (Moritz and Brown 1987). The presence of a small tandem repeat flanking each element is suggestive of a transposable element, but the lack both of open reading frames of substantial length and of short inverted
FIG. 5.—Left page, Restriction map of the plasmid construct containing two copies of the 1.45-kb PstI fragment in direct tandem repeat (shown with a thin line). Digestion with StyI yields a 1.45-kb fragment which has the homopolymer runs of A and T in the middle of the fragment. Digestion with PstI also yields a 1.45-kb fragment containing DNA sequences identical to the 1.45-kb StyI fragment but with the homopolymer runs of A and T permuted to the ends of the fragment. Right page, A, Normal migration pattern of the 1.45-kb fragment generated by the digestion of the dimer plasmid with either StyI (lane 1) or PstI (lane 2). The arrow shows that in agarose gels the 1.45-kb fragments migrate according to their true size. Right page, B, Migration of the same 1.45-kb fragments in polyacrylamide gels. The 1.45-kb fragment cut with StyI (lane 1), which has the A and T homopolymer runs at the middle, migrates much more slowly than would be expected on the basis of its size. The PstI 1.45-kb fragment, which has the A and T runs at both ends, migrates in accordance with its size (lane 2). Numbers in the right margin indicate lengths in kilobases.
repeats at the ends of the element does not support the hypothesis of tranposition (Calos and Miller 1980). Recombination is believed to be unlikely in the animal mtDNA (Brown 1983), but several aspects of the repeated sequences suggest that it may indeed occur. Unequal crossing-over is a well-documented mechanism for generating variation in the number of tandemly repeated sequences (Parma and Snyder 1973) and for sequence homogenization. Both of these features are present and need to be accounted for in the scallop mtDNA. Regions rich in G/C are common in yeast mitochondrial genomes and have been associated with areas of high recombination incidence (Zamaroczy and Bernardi 1986). Rand and Harrison (1989) have recently discussed how intra- and intermolecular recombination in mtDNA may generate arrays with both varying copy number of a repeat and size heteroplasmy, and Joyce et al. (1988) have suggested that interspersed direct motifs, of the size of the motif flanking the P. magellanicus repeated sequence, were the probable sites for recombination events that have led to rearrangements in wheat mtDNA.

Regardless of the exact mechanism involved, it appears that in P. magellanicus alteration of genome length must be site specific and must occur frequently in order to generate a rich polymorphism of precise repeat-length variants (in ~50% of the
animals surveyed, the copy number differed from the most common type, four; K. Fuller and E. Zouros, unpublished results). This suggestion is supported by the observation that size heteroplasmy is much more frequent than restriction-site heteroplasmy, a fact which, in turn, implies higher mutation rates for size change than for nucleotide substitution (Harrison 1989). Similar support comes from the frequent occurrence of multiple heteroplasmy, i.e., the presence of more than two types of molecules in the same individual [e.g., in crickets (Rand and Harrison 1989), weevils (T. Boyce, M. Zwick and C. Aquadro, personal communication), and scallops (B. Gjetvaj, K. Fuller, and E. Zouros, unpublished results)]. Given that stochastic assortment of mtDNA types leads to the decay of heteroplasmy in the descendents of heteroplasmic individuals (Solignac et al. 1984; Rand and Harrison 1986), these states of multiple heteroplasmy cannot be very old but must owe their existence to recurrent appearance of size variants. A better understanding of the relative importance of the structure and organization of the repeat array in generating this variation awaits the analysis and comparison of sequence data from other animal species which also exhibit size variation due to tandemly repeated elements.

Acknowledgments

We thank Dr. J. Wright for helpful suggestions during the early stage of this work, Drs. D. Wolstenholme and P. Cantatore for providing mtDNA clones, Drs. M. Coulthart, F. Doolittle, and M. Gray for many useful comments on the manuscript, and B. Gjetvaj for help with the drawing of the figures. This work was funded by an NSERC grant to E.Z.

LITERATURE CITED


MASATOSHI NEI, reviewing editor

Received May 11, 1989; revision received August 2 1989

Accepted September 11, 1989