Internal organization of long repetitive DNA sequences in sea urchin genomes

(repeated sequences/sequence organization/DNA reassociation)

NIRUPA CHAUDHARI AND SYDNEY P. CRAIG*

Department of Biology, University of South Carolina, Columbia, South Carolina 29008

Communicated by Philip Leder, August 22, 1979

ABSTRACT In keeping with earlier reports, we have found that reassociated long repeat DNA from sea urchin is thermostable, indicating the absence of evolutionarily diverged families of repeated sequences. However, we found that when fragments of radiolabeled long repeat DNA were denatured and reassociated with intact long repeat driver DNA, then sheared to 350 basepairs and assayed for thermal stability, the level of mismatch found in the duplexes varied inversely with the length of the starting fragments. This effect was shown to be due directly to the physical size of the molecules involved in reassociation. These results are consistent with, and support a model for, long repeat DNA in which short units of repetition are arranged in precise arrays. The significance of this arrangement of sequence units within long repeat DNA is discussed.

Repeated sequences, composing 20–60% of the nuclear DNA of a large variety of animals, have been shown to fall into two distinct size classes, which are generally referred to as the "long" and "short" repetitive sequences (1). Long and short repeats have been demonstrated within theDNAs of animals in several phyla (2–6). These studies have also revealed that reassociated long repeat sequences are more thermostable than short repeats. Thus, long repeat sequences have been considered to belong to families of identical repeats, whereas short repeat sequences belong to families with evolutionarily diverged copies. Although detailed studies have been performed on the sequence complexity, copy number, and electron microscopic appearance of both long and short repeat DNA (3, 5–7), until recently, little was understood of the internal organization of the majority of long repeat sequences.

During the course of our earlier investigations, we discovered that when long repeat DNA was sheared, denatured, and then reassociated with itself, mismatched duplexes were produced (8, 9). This indicated that long repeat DNA might in fact contain divergent families of repeats. Therefore, we postulated a model in which (i) long repeat sequences are composed of tandem short units of repetition, 150–300 basepairs (bp) long, (ii) some of these units are members of evolutionarily diverged families of more or less related sequences, and (iii) the units of repetition are precisely ordered, with a certain version of a sequence always adjacent to specific other sequences (8, 9). This precise arrangement would favor the formation, in intact arrays, of long and perfectly matched duplexes over short, mismatched ones (see ref. 9).

The present paper is aimed at testing some aspects of this model. We have investigated the relationships between fragment length and criterion of reassociation on the one hand and the stability of duplexes formed during the reassociations of long repeat DNA on the other. The results reported here strongly support our model for the presence, and precise tandem arrangement, of short units of repetition within long repeat DNA.

MATERIALS AND METHODS

Preparation of Long Repeat DNA. 3H-Labeled and unlabeled DNA was extracted as described from embryos and sperm of the sea urchins Lytechinus variegatus, L. pictus, and Sphaerechinus granularis (8). The specific radioactivity of labeled preparations was approximately 10⁶ cpm/μg of DNA. Long repeat sequences were prepared from radiolabeled or unlabeled DNA, after single strands, 4000–6000 nucleotides long, had been reassociated until the product of DNA concentration (moles of nucleotides per liter) and incubation time (sec), the Cq, was 20. Single strand-specific (S1) nuclease from Sigma was used to remove unassociated single strands. The length of enzymatic digestion was calculated as described (7, 9), and, unless otherwise noted, a digestion estimate of 0.95 was utilized. Long and short repeated sequences were resolved on a column of Bio-Gel A-50m in 0.12 M sodium phosphate buffer, pH 6.8 (10). The same column was also used for analytical molecular sizing in some experiments.

DNA Reassociations. DNA that was to be subsequently digested with S1 nuclease was denatured by heating to 100°C for 5 min and was reassociated in 0.5 M NaCl/0.01 M, 1,4-piperazinediethanesulfonic acid (Pipes), pH 6.8, at 65°C. Other samples were reassociated either in 0.12 M phosphate buffer at 68°C or in 0.4 M phosphate buffer at 68°C. Cq values were adjusted to the standard rate in 0.12 M phosphate buffer (10).

Sizing Analysis of DNA. Single-strand lengths of DNAs were estimated as described (8), using alkaline isokinetic sucrose gradients in an SW 41 rotor (Beckman).

Melting Point Determinations. Double-stranded DNAs longer than 400 bp were sheared at 50,000 rpm in a VirTis 60 homogenizer prior to melting, to eliminate the effect of duplex length upon the melting point (10). Radiolabeled DNA was bound to hydroxyapatite in 0.12 M phosphate-buffered saline and thermally eluted in 3.5°C increments as described (8). The t_m was calculated as the temperature, to the nearest 0.5°C, at which 50% of the total radioactivity had eluted from the hydroxyapatite.

RESULTS

A flow diagram of the fractionations and analyses of DNA is shown in Fig. 1. Purified long repeat DNA has been shown to be relatively thermostable (3, 7). In our hands, the t_m of purified long repeat DNA was always within 0.5°C of that for native DNA.

Abbreviations: bp, basepairs; Cq, product of DNA concentration in mol of nucleotides per liter multiplied by time in sec; Pipes, 1,4-piperazinediethanesulfonic acid; t_m, melting point of DNA.

* Present address: Department of Biology, University of Southwestern Louisiana, Lafayette, LA 70504.
DNA (Table 1). In order to test whether this high thermal stability of long repeat DNA is retained after subsequent melts and reassociations, radiolabeled long repeat DNA, sheared to average lengths of 350 bp, was reassociated to Cot 50 with driver quantities of similarly prepared, unlabeled long repeat DNA. The $t_m$s are significantly lowered for the twice-reassociated long repeat DNA of several species of sea urchins (Table 1). We have interpreted this $t_m$ depression to result from the presence, within long repeat DNA, of families of sequences, the members of which have diverged evolutionarily by nucleotide substitution (8, 9).

According to our model for the internal organization of long repeats, the length and ordering of the arrays of units allows the preferential formation of long stable duplexes. Therefore, we predicted that varying the length of the long repeat strands involved in secondary reassociations would result in a parallel change in the thermal stability of the duplexes produced. In order to test this prediction, we generated a broad range of duplex fragment sizes by extensive S1 nuclease digestion of radiolabeled long repeat DNA from *L. variegatus*. At high ratios of enzyme to DNA, S1 nuclease cuts DNA at single-strand breaks and even at single basepair mismatches (11). The digested DNA was then chromatographed on Bio-Gel A-50m (Fig. 2). The average single-stranded lengths of fragments ranged from 1800 to less than 300 nucleotides (Table 2). Aliquots from fractions 11–16 were withdrawn, sheared to 350 bp, bound to hydroxyapatite, and thermally eluted. The $t_m$ of fraction 17 was determined without resharing because this fraction contained molecules of 300-bp average length. The $t_m$s of these fractions (Table 2, column A) are close to one another and to the $t_m$ of *L. variegatus* native DNA (Table 1). We take this as evidence that the fragments were generated by random cleavage of long repeat sequences.

As stated earlier, we predicted that the longer fragments of long repeats should be capable of forming more stable duplexes when reassociated with intact single strands of long repeat DNA. To test this, we denatured and reassociated aliquots of fractions 11–17 (Fig. 2) with excess, unsheared, unlabeled long repeat DNA. These reassociations were taken to Cot 50 before the DNA in each reaction mixture was sheared to 350 bp, and the $t_m$ of the duplexes formed was determined by thermal elution from hydroxyapatite (Fig. 3 and Table 2, column B). These results demonstrate a correlation between the size, during reassociation, of radiolabeled long repeat fragments and the

---

**Table 1. Thermal stability of long repetitive DNA**

<table>
<thead>
<tr>
<th>Species</th>
<th>$t_m$, °C</th>
<th>Native DNA*</th>
<th>Long repeat DNA*</th>
<th>Twice-reassociated long repeat DNA†</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>L. pictus</em></td>
<td>84.0</td>
<td>84.0</td>
<td>76.5</td>
<td></td>
</tr>
<tr>
<td><em>L. variegatus</em></td>
<td>85.5</td>
<td>85.0</td>
<td>75.0</td>
<td></td>
</tr>
<tr>
<td><em>S. granularis</em></td>
<td>83.5</td>
<td>83.5</td>
<td>76.5</td>
<td></td>
</tr>
</tbody>
</table>

* Native and long repetitive DNAs were sheared to 350 bp before thermal elution from hydroxyapatite.
† Radiolabeled long repeat DNA was sheared to 350 bp, then reassociated to Cot 50 with similarly sheared long repeat driver DNA, in 0.4 M phosphate buffer at 68°C. The DNA was then bound to hydroxyapatite and thermally eluted.
thermal stability of the duplexes they produce.

A possible explanation for the spread of \( t_{m5} \) obtained is that the fragments of long repeat DNA (Fig. 2) were not generated by random scission by S1 nuclease but rather represent different populations of sequences with genuinely variant properties. In order to test this possibility, we performed another series of reassociations with all conditions (salt, temperature, tracer-to-driver mass ratios and Cdt) identical to those for column B of Table 2, except that the DNAs from fractions 11–16 were sheared to 350 bp before reassociation. These sheared DNAs were concentrated on hydroxyapatite and were then denatured and reassociated in the presence of unsheared unlabeled long repeat DNA. The duplexes formed in these reactions were assayed for thermal stability and the results are presented in Table 2, column C. All fractions so examined yielded \( t_{m5} \) of 76.5–77°C. Thus, the fractions from Fig. 2 appear not to contain sequences of appreciably different base compositions. The \( t_{m5} \) in column C are close to the \( t_{m5} \) obtained when total long repeat DNA, sheared to 350 bp, was denatured and reassociated with itself (Table 1). The \( t_{m5} \) obtained in these two series (columns B and C) are quite different from each other. Therefore, we conclude that the physical size of the long repeat molecules involved in reassociation determines the level of mismatch introduced into the duplexes formed.

![Fig. 2. Profile of S1 nuclease-generated fragments of long repeat DNA, on Bio-Gel A-50m. Long repeat DNA from L. variegatus was incubated with S1 nuclease to a digestion estimate of 0.999, then chromatographed on Bio-Gel A-50m. Radioactivity in 50-μl aliquots of each 5-ml fraction was determined by scintillation counting.](image)

![Fig. 3. Thermal elution profiles of various fragment lengths, reassociated with unsheared long repeat DNA. Aliquots (6000 cpm) from fractions 11–17 of Fig. 2 were individually concentrated on hydroxyapatite and denatured and reassociated in the presence of 8 μg of long repeat DNA in 0.4 M phosphate buffer at 68°C, to Cdt 50. The reaction products were then diluted and sheared at 50,000 rpm in a Vibra-60 homogenizer. The salt concentration was adjusted to 0.12 M phosphate buffer, and the duplexes (between 45% and 52% of the input radioactivity) were bound to hydroxyapatite at 63°C, then thermally eluted. The melting profiles shown represent, from left to right, fractions 17, 16, 14, 12, and 11 from Fig. 1, and sheared long repeat DNA.](image)

![Fig. 4. Profile of twice-reassociated L. variegatus long repeat DNA on Bio-Gel A-50m. About 150,000 cpm of labeled long repeat DNA was mixed with 18 μg of unlabeled long repeat DNA in 0.3 M NaCl/0.01 M Pipes, pH 6.8. The mixture was divided into two aliquots, which were heat denatured and then incubated to Cdt 10, at either 68°C (○) or 58°C (●). The DNAs were then treated with S1 nuclease to a digestion estimate of 0.95, bound to hydroxyapatite, eluted with 0.4 M phosphate buffer, and chromatographed on Bio-Gel A-50m. Radioactivity in 250-μl aliquots of 5-ml fractions were determined by scintillation counting. Arrows indicate markers as in Fig. 2.](image)

### Table 2. Thermal stability of L. variegatus long repeat DNA fragments reassociated with unsheared long repeat DNA

<table>
<thead>
<tr>
<th>Fraction</th>
<th>Fragment length, in nucleotides</th>
<th>A. Sheared</th>
<th>B. After reassociation as long fragments</th>
<th>C. After reassociation as short fragments</th>
</tr>
</thead>
<tbody>
<tr>
<td>11</td>
<td>1800</td>
<td>86.0</td>
<td>82.5</td>
<td>77.0</td>
</tr>
<tr>
<td>12</td>
<td>1020</td>
<td>85.0</td>
<td>81.5</td>
<td>77.0</td>
</tr>
<tr>
<td>13</td>
<td>880</td>
<td>85.5</td>
<td>80.0</td>
<td>76.5</td>
</tr>
<tr>
<td>14</td>
<td>670</td>
<td>85.0</td>
<td>79.5</td>
<td>77.0</td>
</tr>
<tr>
<td>15</td>
<td>600</td>
<td>84.5</td>
<td>79.0</td>
<td>76.5</td>
</tr>
<tr>
<td>16</td>
<td>460</td>
<td>85.0</td>
<td>78.5</td>
<td>76.5</td>
</tr>
<tr>
<td>17</td>
<td>300</td>
<td>84.0</td>
<td>76.0</td>
<td>—</td>
</tr>
</tbody>
</table>

Variously sized long repeat DNA fragments from fractions 11–17 from Fig. 2 were analyzed as follows: Column A, 5000-cpm aliquots of each fraction were sheared and then the \( t_{m5} \) was determined on hydroxyapatite. Column B, 6000-cpm aliquots of each fraction were concentrated and reassociated with intact long repeat DNA in 0.4 M phosphate buffer at 68°C to Cdt 50. The reaction mixture was sheared to 350 bp, and the \( t_{m5} \) of the duplexes were determined. Column C, 6000-cpm aliquots were sheared, concentrated, and reassociated with unsheared long repeat DNA in 0.4 M phosphate buffer at 68°C to Cdt 50. The \( t_{m5} \) of resultant duplexes were determined on hydroxyapatite.
Table 3. Thermal stability of variously sized duplexes produced during reassociation of long repeat DNA from *L. variegatus*

<table>
<thead>
<tr>
<th>Fraction</th>
<th>DNA reassociated (°C)</th>
<th>DNA reassociated (°C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>11</td>
<td>82.5</td>
<td>82.5</td>
</tr>
<tr>
<td>12</td>
<td>81.0</td>
<td>81.5</td>
</tr>
<tr>
<td>14</td>
<td>79.5</td>
<td>80.0</td>
</tr>
<tr>
<td>16</td>
<td>76.0</td>
<td>76.0</td>
</tr>
</tbody>
</table>

Variously sized duplexes formed by reassociating long repeat DNA were resolved on Bio-Gel A-50m (Fig. 4). Representative fractions from these runs were sheared to 350 bp. The DNAs were bound to hydroxypatite in 0.12 M phosphate buffer at either 56°C (for the 58°C reassociation) or at 68°C (for the 69°C reassociation) and thermally eluted.

When intact long repeat sequences are denatured and reassociated among themselves, the duplexes formed cover a wide range of sizes, from 1200 to 150 bp (unpublished observations). From the foregoing account, the shorter duplexes produced during this second reassociation might be expected to contain a greater proportion of mismatched basepairs than would the longer ones. Also, we wished to establish whether the stringency of criterion during this second reassociation could influence the stability of the duplexes produced. Therefore, 3H-labeled unsheared long repeat DNA was mixed with a 100-fold mass excess of unsealed long repeat DNA in NaCl/Pipes buffer. The batch was then divided into two, and both mixtures were denatured and incubated to Cgt 10, one at 58°C and the other at 68°C. Subsequently, the reaction mixtures were digested with S1 nuclease to remove unreacted strands, and resistant duplexes were bound to hydroxypatite, eluted with 0.4 M phosphate buffer, and chromatographed on Bio-Gel A-50m. The elution profiles of the two sets of DNA on the Bio-Gel column appear nearly identical with respect to the relative amounts of DNA of various sizes (Fig. 4). Selected fractions from Fig. 4 were individually sheared to 350 bp and their TmS were measured by thermal elution from hydroxypatite (Table 3). The TmS of comparably sized duplexes produced at the two reassociation criteria were identical. In addition, they were close to the TmS of duplexes produced by similarly sized molecules of S1-digested long repeat DNA after a second reassociation (Table 2).

The foregoing results all show that the melting properties of duplexes, produced by purified long repeat sequences, are directly a function of the length of the molecules in question, during reassociation.

**DISCUSSION**

We have shown that sequences in much of long repeat DNA belong to families of similar but nonidentical members. Thus, very short fragments of long repeat sequences seem to be able to reassociate with evolutionarily diverged copies of the sequence, yielding poorly matched duplexes. On the other hand, progressively larger fragments seem to show increasing likelihood of strands finding and pairing with perfect homologs. These results support the existence of an internal organization of long repeats, which favors the formation of stable duplexes and which is broken down during physical fragmentation. Previously, we proposed that short units of repetition may be tandemly ordered so that certain “versions” of given sequences are likely to be adjacent to specific “versions” of other sequences (9). The existence of such organization is supported by the demonstration of an inverse relationship between the size of long repeat fragments and the amount of mismatch found when these sequences are reassociated a second time. Two criteria of reassociation were shown not to affect this relationship. Because all duplexes were sheared to 350 bp before Tm determinations, duplex length could not have contributed to the differences in thermal stability determined. Also, we considered whether fragment length might affect the reassociation criterion, allowing more unstable hybrids to form from shorter fragments. In this regard, bacterial DNA is known to be capable of forming slightly mismatched duplexes when the reassociation criteria are relaxed (12). However, bacterial DNA has been shown to form duplexes of almost identical thermal stabilities after reassociation as either long or short fragments (13). Therefore, we believe that our observations result directly from unusual characteristics of long repeat DNA, and support a novel arrangement of repeat units within these sequences.

The present explanation for the internal organization of long repeat sequences is in keeping with observations from other laboratories. For instance, the TmS reported by others, for purified long repeat DNA, have been within 2–2.5°C of those of native DNA (2, 3, 5, 7) but seldom as close as the 0–0.5°C that we have observed (8, 9). This could be due to the fact that other investigators have started with native DNA fragments, averaging only 2000 bp, as compared to our 4000- to 6000-bp length. Their preparation artifacts might therefore have contained a greater proportion of incomplete long repeats, which would be expected, according to our results, to contain a greater degree of mismatched basepairs.

Eden et al. (3) isolated a highly thermostable repetitive sequence fraction, making up less than 1% of the sea urchin genome, and showed that all copies of these sequences are nearly identical (3). We have shown that a subset of long repeat DNA (roughly 10%), making up about 1% of the total genome of sea urchins, is faithfully conserved in sequence, through extremely long evolutionary periods (8). It is possible that this fraction overlaps that purified by Eden et al. and, likely, contains the genes for rRNA, tRNA, SS RNA and the histones, all of which would behave as nondivergent families of repeats. Our present results do not preclude the presence of some perfectly reiterated sequence families that melt at high temperatures only.

Wu et al. (6) suggested that in the rat, at least some of the long repeat sequences might be composed as “tandem arrays of short repeats.” Several models for the regulation of gene expression have proposed that batteries of redundant sequences may function in the regulation of gene expression (14–16). As discussed in a recent review (16), evidence for such batteries is lacking. Also, this review suggested that sequences within the batteries of repeats would be complementary to the sequences in interspersed short repeats (16). Our own investigations have shown extensive overlap of sequences between long and short repeats (unpublished). Thus, our evidence provides a structural basis in DNA for such models for the regulation of gene expression.

We gratefully acknowledge Dr. Michael Vodkin and Andrew Lesniak for assistance in the preparation of this manuscript. This work was supported in part by North Atlantic Treaty Organization Research Grant 1271. Portions of this work will be submitted by N.C. in partial fulfillment of the requirements for the Doctor of Philosophy degree at the University of South Carolina.

Biochemistry: Chaudhari and Craig