The spread of sequence variants in Rattus satellite DNAs

David A. Epstein¹, Frank R. Witney² and Anthony V. Furano

Section on Genomic Structure and Function, Laboratory of Biochemical Pharmacology, National Institute of Arthritis, Diabetes, and Digestive and Kidney Diseases, National Institutes of Health, Building 4, Room 116, Bethesda, MD 20205, USA

Received 24 October 1983; Revised and Accepted 1 December 1983

ABSTRACT
The genus Rattus has two related families of satellite DNA: Satellite I consists of tandem arrays of a 370 base pair repeat unit which is a dimer of two 185 base pair portions (a, b) which are about 60% homologous. Satellite I' consists of tandem arrays of a 185 base pair repeat unit (a') which is about 85% homologous to a and 60% homologous to b. R. norvegicus contains only satellite I but R. rattus contains both satellites I and I'. We examined certain aspects of satellite DNA evolution by comparing the spacing at which variant repeat units of each satellite have spread among non-variant repeat units in these two species. With but one exception, in R. rattus, 15 different variant repeat units have spread among nonvariant repeat units of satellite I, with a spacing equal to the length of the (a,b) dimer. Similarly, fourteen different variant repeat units of the monomeric satellite I' have mixed among non-variant repeat units with a spacing equal to the length of the (a') monomer. These results suggest that a mechanism involving homologous interaction among satellite sequences could account for the spread of variant family members. We also found that a sequence variant present in certain portions of the dimeric repeat unit of satellite I is more efficiently amplified (or less efficiently corrected) than variants occurring in other regions. This was not true for the monomeric repeat unit of satellite I'.

INTRODUCTION
Metazoan genomes are characterized by families of repeated DNA elements. Some families contain 100,000 or more near-identical members (i.e., repeat units) that are organized in tandem arrays which are usually located at telomeric and centromeric regions of chromosomes. These DNA families are called satellite DNA [reviewed in (1, 2)], and whatever their function, which may be limited to meiotic events (3), structural studies of these multicopy families have revealed several characteristics pertinent to the biological mechanisms that may determine their evolution.

For example, the high degree of sequence similarity among thousands of family members clustered in various tandem arrays suggests that unequal recombination and/or gene conversion between tandem arrays occurs quite frequently to "homogenize" the members of satellite families (4-10). On the other hand, certain portions or "segments" (11, 12) of satellite families
contain a mixture of variant and non-variant members. Although a particular segment might represent less than 1% of a satellite family, it is readily identified by restriction enzyme analysis if the variant member differs from the non-variant members by a new restriction enzyme site [(11) and see Results]. These segments could be produced by repeated rounds of recombination/gene conversion between a variant member that had arisen by a random base change and nearby non-variant members [see (10)], although other mechanisms have been suggested (11).

Additionally, studies of mouse (13, 14), primate (15-18), and rat satellites (19-21) indicate that many satellite DNAs were formed by recurring rounds of the following steps: amplification of a DNA element, divergence of some of these elements, and subsequent amplification of a divergent pair of elements (13, 14). As Fig. 1 shows, this scheme is exemplified by the 370 base pair (bp) repeat unit of rat satellite I first described in R. norvegicus (19, 20), which is a divergent dimer of two 185 bp elements, called a and b (21). Overall, a and b are about 60% homologous, but they share a pair of 20 bp sequences that are nearly identical to each other. As Fig. 1 shows, this pair of nearly identical sequences is also found in the 185 bp a' repeat unit of the satellite I' family in R. rattus. This repeated homology in the 185 bp rat elements suggests they were produced by an earlier dimerization of a shorter (92/93 bp) ancestral sequence [see Fig. 1 and (19)].

The formation of segments and the production and amplification of divergent dimers are unexplained. Frequent recombination (and/or gene conversion events) between even partially homologous sequences would produce repeated sequences of identical sequence (4). However, out of register recombination (or gene conversion events) initiated at the homologous regions (22) of repeat elements such as those shown at the bottom of Fig. 1 could produce variant repeat elements that in some cases could be longer than existing family members and result in a divergent dimer. Alternatively, if the monomeric units of a tandem array, such as satellite I', interact with each other mainly as "dimers" and not monomers (see bottom of Fig. 1, left), then it would be possible for the right and left hand members of the "dimer" to diverge from each other and evolve into a divergent dimer. For this to take place, the register governing the homologous interactions of the repeat units in tandem arrays would have to be determined by factors other than the length of the repeat element itself (e.g., chromatin structure). Since R. rattus contains both a monomeric (satellite I') and dimeric satellite (satellite I) (21), we can compare, in the same animal, the register which governs the homologous interaction of the
members of these two related satellite DNAs (see Fig. 1).

To do this we determined by restriction enzyme analysis the spacing of variant members among non-variant members in segments of satellite I' and satellite I. Our results on 15 different variants in the satellite I arrays of *R. norvegicus* and *R. rattus* and 14 in the satellite I' arrays of *R. rattus*, showed that, with but one possible exception in *R. rattus* satellite I, the register for placement of sequence variants in the various segments is determined by the length of the basic repeat unit; i.e., *a'* units interact as monomers; *a,b* units as dimers. In addition, we found a bias in the efficiency with which at least one variant was spread and/or corrected in *a,b* arrays, as compared to *a'* arrays.

**MATERIALS AND METHODS**

**DNA Sources**

The preparation of high molecular weight DNA from *R. norvegicus* and *R. rattus*, as well as the construction, isolation and characterization of plasmids *R. n.* 10 (satellite I) and *R. r.* 3 (satellite I') are described in Witney and Furano [(21); also see legend to Fig. 1].

**Restriction Enzyme Digestions**

Restriction enzyme digestions of *R. rattus* or *R. norvegicus* DNAs (isolated from the livers of single animals) were carried out in 25 μl reactions at 37°C for 5 hours using conditions specified by the supplier. Each reaction contained 5-7 μg of DNA and 1-2 U of enzyme/μg of DNA. (One U is that amount of enzyme that will completely digest 1 μg of DNA in 1 hour at 37°C.) In some cases completeness of digestion was tested by adding a second dose of enzyme along with a suitable "marker" DNA, such as lambda phage or SV40 DNA. After blotting (see below), the marker fragments were detected by hybridization with radioactive marker DNA. In all cases we detected the expected number of marker fragments and no partial restriction digestion products.

**Blot Hybridizations**

DNA fragments were separated by electrophoresis on 1% agarose gels in 50 mM Tris-acetate buffer, pH 8.0. The DNA fragments were transferred to nitrocellulose filters (BA-85, Schleicher and Schuell) by blotting (23) as modified by Wahl et al. (24). After incubation for at least 18 hours at 65°C in a solution containing 5X SSC (1X SSC is 0.15 M sodium chloride, 0.015 M sodium citrate, pH 7.2), 0.1% polyvinylpyrrolidone, 0.1% Ficoll, 0.25% sodium dodecyl sulfate, and 100 μg per ml of sonicated salmon sperm DNA, the filters were incubated in the same buffer solution with about 2 x 10⁶ cpm of ³²P-labeled
plasmid DNA made radioactive by nick translation (25). The specific activities of the plasmid DNAs are given in the legend to the figures. The radioactive plasmids were boiled 10 minutes in 10 mM sodium phosphate buffer, pH 7.0, before being added to the filters. Incubation was carried out for 24 hours at 65°C, after which the filters were washed at room temperature with 2 changes (200 ml each) of 2X SSC and 2 changes (200 ml each) of 3 mM Tris base (26). Each wash was carried out for 20 minutes. After drying, the filters were exposed to x-ray film (Kodak XAR-5) at -70°C, using an intensifying screen (DuPont, Cronex Lightning-Plus).

**Isolation of Radiolabeled 185 bp EcoRI Fragments from R. norvegicus Satellite I**

Ten µg of R. norvegicus DNA was digested with EcoRI for 4 hours at 37°C in a solution (40 µl) containing 50 mM NaCl, 100 mM Tris-HCl (pH 7.5), 10 mM MgCl₂, and 40 units of enzyme. The reaction was then adjusted to 6 mM β-mercapto-ethanol, 40 µM each of dGTP, dCTP, TTP; 200 µCi of [alpha-³²P]dATP (3,000 mCi/µmol); 1 unit of the Klenow fragment of DNA polymerase (27) was added to the reaction, and synthesis was carried out for 20 minutes at 20°C. The reaction mixture was then electrophoresed on an 8% polyacrylamide gel, and the 185 bp EcoRI fragments were isolated from the gel and used as the substrate for the restriction digestion experiments described in Results.

**RESULTS**

We previously showed that the cloned R. rattus satellite I sequences in plasmid R. r. 3 (see legend to Fig. 1) do not hybridize under our conditions to satellite I DNA of either R. norvegicus or R. rattus (21). This may not be surprising because even though the R. r. 3 clone contains a 93 bp portion of satellite I' (a') that is overall 85% homologous to the corresponding sequence in satellite I (a), the longest stretch of complete homology is only 20 bp. Furthermore, the last 20 bp of the a' sequence in clone R. r. 3 is not homologous to the b portion of satellite I which alternates with the a sequence in this satellite and interrupts the remaining homology between the cloned a' sequence and its homolog in satellite I [see Fig. 1, and Fig. 3 of (21)]. The R. norvegicus satellite I sequences in plasmid R. n. 10 (see legend to Fig. 1) do not hybridize to satellite I' DNA of R. rattus, but hybridize well to the satellite I present in this genome. Therefore, restriction fragments containing either satellite I' or satellite I can be unambiguously distinguished with the clones described in the legend to Fig. 1. The extent of hybridization by plasmid R. n. 10 with R. rattus DNA is about 20% that with R. norvegicus DNA. This is due to the fact that R. rattus contains about one-fifth of the amount...
FIGURE 1. Diagram of an evolutionary scheme for generating the structure of Rattus species satellites I and I'. The open box represents the ancestral 92/93 bp unit of the satellite that was amplified into the long tandem arrays characteristic of satellite DNA. Over time, divergence occurs in some of the repeat elements of the tandem array (stippled box). This divergence may create a novel restriction enzyme site allowing detection of the variant element by digestion with the restriction endonuclease. If the divergent element and a neighboring element are amplified as a single unit, the basic repeating unit of the satellite would be a 185 bp unit composed of related 92/93 bp elements. This is the current evolutionary unit of the R. rattus satellite I' (21). If the 185 bp repeat unit was amplified in tandem with yet another related 185 bp unit, the basic repeat of that satellite would be a 370 bp unit composed of a divergent dimer of elements. This is the evolutionary unit of the R. norvegicus and R. rattus satellite I. R. rattus contains only about one-fifth as much satellite I as R. norvegicus; the remainder of the satellite sequences in R. rattus is satellite I'. R. norvegicus does not contain satellite I' DNA.

Sequence Variants in the R. norvegicus Satellite I Tandem Arrays

When satellite I DNA is digested with an enzyme that has a site in the canon-
FIGURE 2. Hybridization of *R. norvegicus* satellite I DNA to DNA from *R. rattus* and *R. norvegicus*. Seven µg of *R. norvegicus* (A, B) or *R. rattus* (C) DNA was digested with various restriction enzymes and electrophoresed on a 1% agarose gel. The DNA fragments were then transferred to nitrocellulose filters and
hybridized with $^{32}$P-labeled R. n. 10 plasmid as described in Materials and Methods. R. n. 10 contains a 185 bp segment of R. norvegicus satellite I (see Fig. 1). The plasmid was labeled to a specific activity of about $3 \times 10^7$ cpm/ug. 2 x $10^6$ cpm of radioactivity was added to each hybridization reaction. The hybrids in (B) and (C) were detected by autoradiography. An ethidium bromide-stained gel identical to that used for the blot used in (B) is shown in (A). The sizes of the fragments are given in kilobase pairs in (A) and in base pairs in (B) and (C). The numbers at the top of each gel denote the enzyme used to digest the rat DNA; m, HindIII digested lambda DNA. The sizes of the 370 bp HaeIII or HindIII satellite I bands or the 185 bp satellite I' band were derived from DNA sequence analysis (19, 21). The sizes assigned to higher multiples of these bands was based on their comigration with DNA fragments of known molecular weight after gel electrophoresis on agarose or polyacrylamide gels of various concentrations (data not shown).

ical sequence (19) of the repeat unit, essentially all of the high molecular weight tandem arrays are reduced to fragments the size of the repeat unit or some low multiple thereof. This pattern of fragments has been called a type A digestion pattern (11). For example, the sites for the restriction enzymes HaeIII and HindIII occur once in the consensus sequence of the R. norvegicus satellite I repeat unit (19). As can be seen in Fig. 2B (tracks 3 and 19), digestion of R. norvegicus DNA with either enzyme yields satellite I-containing fragments that are 370 bp or some low multiple of 370 bp. These digestions are complete, since adding a second dose of either enzyme [along with a marker DNA to test the efficacy of the second dose (see Materials and Methods)] did not change these patterns (results not shown).

On the other hand, restriction enzymes that have recognition sites not present in the canonical repeat unit of a satellite often digest about 0.5% to 5% of the total satellite arrays, producing a ladder of fragments which range in size from undigested tandem arrays to fragments that in some cases are the length of the repeat unit.

This has been called a type B digestion pattern (11), examples of which are shown in tracks 2, 5, 6, 7, 8, 10, and 16 of Fig. 2B. Although B restriction sites are sufficiently rare so that only 0.5% to 5% of the satellite is digested, they are close enough together to produce fragments as small as low multiples of the repeat unit of the satellite. This indicates that the variant repeat unit that contains the B restriction site is present only in a segment of the satellite family and is not mixed randomly among all members of the family (11).

The "B type" enzymes used in Fig. 2B recognize six distinct sequences. (Enzymes 6 and 7 both recognize GATC.) In all cases the spacing of the B restriction site is 370 bp. In addition to those shown in Fig. 2, we detected 9 other B variants, each of which also occurred at a 370 bp spacing. An
FIGURE 3. Comparison of restriction enzyme sites in satellite I DNA of R. norvegicus (R. n.) and R. rattus (R. r.). DNA was digested with either Hinf I, Dde I, or Eco RI, and the resulting fragments separated on an 8% polyacrylamide gel. The fragments were then transferred to nitrocellulose filters and hybridized with \( \text{^32P}-\text{labelled } R. \text{ n. } 10 \) plasmid as described in Materials and Methods and in the legend to Fig. 2. Hybrids were detected by autoradiography. The sizes of the fragments are given in base pairs. (B) R. n. or R. r. DNA was digested with Eco RI. After transfer of the resulting fragments to nitrocellulose, the filters were hybridized with either R. n. 10 to detect satellite I or R. r. 3 to detect satellite I' sequence. Hybrids were detected by autoradiography.

An explanation for these findings is that a random base change which created a restriction enzyme site not present in the consensus sequence occurred in a single member of the satellite family. The restriction site variant was then propagated (amplified) through a small percentage of the non-variant members by a mechanism involving homologous interaction between variant and non-variant repeat units [see (10)] to produce a satellite segment (11).

Our finding that no spacings other than 370 bp are produced in satellite I arrays [i.e., none of the variant segments show, for example, 1/2-mer (185 bp) or 1 1/2-mer (555 bp) spacing] indicates that if out of register events take place, the products of these reactions have not accumulated in the genome to a detectable extent.
Sequence Variants in the \textit{R. rattus} Satellite I Tandem Arrays

Earlier work \cite{21} showed that \textit{R. rattus} DNA contained a satellite DNA that hybridizes well to the \textit{R. norvegicus} satellite I. As is shown in Fig. 3A, restriction enzyme digestion by enzymes that have multiple sites within the 370 bp consensus sequence of rat satellite I \cite{19} produce nearly identical sets of fragments with both the \textit{R. rattus} and \textit{R. norvegicus} satellite I. Similar results were also found with \textit{HphI} (data not shown). The \textit{HinfI} fragments of sizes other than 92/93 bp, or multiples thereof, are produced because these sites are not equidistant from each other in the a,b tandem arrays of satellite I in both \textit{R. rattus} and \textit{R. norvegicus}. This is apparent from the DNA sequence analysis of \textit{R. norvegicus} satellite I \cite{19, 20}. The results in Fig. 3A indicate that most of the 370 bp satellite I repeat units in \textit{R. rattus} and \textit{R. norvegicus} are quite similar.

Digestion of \textit{R. rattus} DNA with the set of enzymes listed in Fig. 2 produced 8 cases of B patterns in the \textit{R. rattus} satellite I, and with the exception of \textit{HinfII} (lane 5), each with a 370 bp spacing (Fig. 2C). B patterns with a 370 bp spacing were also produced by 7 other enzymes with \textit{R. rattus} satellite I (data not shown) for a total of 15 distinct variant sites. Nine of the 15 variant sites present in the satellite I of each rat species were the same. However, this does not mean that these "segmental variants" \cite{12} have been conserved between the two species (see Discussion).

Sequence Variants in the \textit{R. rattus} Satellite I'

The pattern of fragments generated from the \textit{R. rattus} satellite I' by the restriction enzymes used in Fig. 2 is shown in Fig. 4C. As was the case with the \textit{R. norvegicus} satellite I, an enzyme such as \textit{HaeIII} that has a site in the consensus repeat unit of the satellite (see Fig. 1) generates a type A restriction pattern. As Fig. 4C, track 3, shows, clone \textit{R. r. 3} hybridized to a 185 bp \textit{HaeIII} fragment and a short ladder of fragments that are multiples of 185 bp.

Digestion with enzymes whose sites were not present in the consensus sequence produced B patterns, and in addition to the six B patterns shown in Fig. 4C we detected seven other B variants. In every case the periodicity of the B pattern was 185 bp, the same as the length of the repeat unit of this satellite. These results mean that the repeat units of satellite I' interact as monomers. If reactions as dimers or 1/2-mers (see Fig. 1) occurred, the products of these reactions have not accumulated to detectable amounts.

Nonrandom Alteration of EcoRI sites in the Satellite I Tandem Arrays

In the course of carrying out the experiments described above, we observed a difference in the extent to which the EcoRI sites are lost from satellite I as
FIGURE 4. Hybridization of rat satellite I and I' DNA to R. rattus DNA. Seven µg of R. rattus DNA was digested with various enzymes and electrophoresed on a 1% agarose gel. The fragments were transferred to a nitrocellulose filter and hybridized with ³²P-labeled plasmids as described in Materials and Methods and
in the legend to Fig. 2. (A) shows the ethidium bromide-stained gel used for the blot in (B) and (C): (B) shows the same autoradiogram of the hybrids between the satellite I probe, R. n. 10, and R. rattus DNA shown in Fig. 2C. For (C), the blot used in (B) was reclaimed by incubating it for 10 minutes at room temperature in 0.1 N NaOH. The blot was neutralized in 0.5 M Tris-Cl, pH 7.5, 3.0 M NaCl, blotted damp-dry, and hybridized with the satellite I' probe as described in the legend to Fig. 2. The numbers at the top of the gel denote the enzymes used to digest the rat DNA and are the same as those used in Fig.2.

compared to satellite I'. As Fig. 1 shows, both the a and b units of the satellite I and the a' unit of satellite I' have two EcoRI sites which in tandem arrays are separated by 92 or 93 bp of DNA. Fig. 3B shows that EcoRI digestion of either R. rattus or R. norvegicus satellite I DNA produces a ladder of fragments that are multiples of 92 or 93 bp. The higher multiples result from a loss of the EcoRI site(s) in these tandem arrays. We estimate from these results (and others not shown) that about 20% of the EcoRI sites have been lost in both R. norvegicus and R. rattus. This estimate agrees with that of Pech et al. (19) from their studies with rat satellite I. By contrast, digestion of R. rattus DNA with EcoRI shows that the satellite I' tandem arrays have lost far fewer EcoRI sites (Fig. 3B).

Although the loss of 20% of the EcoRI sites in the rat satellite I could be due to the accumulation of about 3% random base changes in these satellite arrays (13, 19), we found that this is not so; most of the mutations affected only 2 of the 4 EcoRI sites. We observed this when we examined the restriction products of isolated 185 bp EcoRI fragments of R. norvegicus DNA (see Materials and Methods). As diagrammed in Fig. 5, the 185 bp fragments could result from the alteration of a single EcoRI site between any two EcoRI sites in a tandem array. Fig. 5 also shows that subsequent digestion of the 185 bp EcoRI fragments with either HaeIII or HindIII would generate two possible sets of restriction fragments surrounding the altered EcoRI site. If each of the EcoRI sites were altered randomly, one would expect that the amount of both pairs of HaeIII/EcoRI or HindIII/EcoRI double digestion products to be present in the same amount. However, when the 185 bp EcoRI fragments were digested with either HaeIII or HindIII and the resulting fragments analyzed by gel electrophoresis, it was clear that for both enzymes one set of fragments was present in higher concentration than the other (Fig. 6). Digestion of the 185 bp EcoRI fragment with HindIII produced mostly the 129 bp/56 bp fragments, but only small amounts of the 149 bp/36 bp fragments. A nonequal ratio of double digestion products was also produced by HaeIII digestion of the 185 bp EcoRI fragments (Fig. 6). These results suggest that of the four EcoRI sites repeated in the satellite I tandem arrays, two are preferentially conserved (or
FIGURE 5. Representation of *R. norvegicus* satellite I tandem arrays showing the location of EcoRI sites and mutated EcoRI sites that would generate 185 bp fragments. The a (shaded boxes) and b (open boxes) elements of the tandem arrays are also shown (see Fig. 1 and text for explanation). The four possible unique 185 bp EcoRI fragments and the digestion products of these fragments with either HaeIII or HindIII are also shown. The numbers below the boxes are base pairs lengths of the various fragments. The symbols representing the restriction sites are shown in the figure.

lost) compared to the other two. As the diagram in Fig. 5 shows, the EcoRI sites furthest from the junctions between the a and b portions of the a,b tandem arrays were the EcoRI sites that were most commonly lost. Interestingly, these EcoRI sites reside in the most conserved portion of the satellite arrays (Fig. 1 and Introduction).

**DISCUSSION**

Either unequal recombination or gene conversion between different members of the tandem arrays of satellite DNA families can keep the 100,000 or more members of these families very similar in sequence (see Introduction). These mechanisms could also account for the spread of variant members into satellite families (and other multi-gene families) [reviewed in (10)], and although these processes have only been directly demonstrated in fungi and yeast (10, 22), data presented in this paper are consistent with this idea. In *R. rattus*, which contains two closely related satellite families that have basic repeat units of different lengths, restriction enzyme site variants, with one exception (discussed below), have become mixed among non-variant members to produce segments with a periodicity that is identical to that of the basic repeat unit of the satellite. This strongly suggests that segments of variant and non-variant members arose by a mechanism involving homologous interactions among
Digest of 185 bp Eco RI Fragments of R.n. Satellite I
Hind III Hae III

FIGURE 6. Digestion of 185 bp EcoRI fragments of R. norvegicus satellite I. Restriction digests of 185 bp EcoRI fragments from R. norvegicus satellite I DNA with either HindIII or HaeIII. 4 x 10^5 cpm of [32P] 185 bp EcoRI fragments prepared as described in Materials and Methods was digested with HindIII or HaeIII and the products analyzed by electrophoresis on an 8% polyacrylamide gel. The fragments were detected by autoradiography and the sizes of the fragments given in base pairs. The 22 bp fragment generated by HaeIII digestion of the 185 bp EcoRI fragment was electrophoresed off this gel.

The only exception to the above results was that obtained with HincII digestion of R. rattus satellite I DNA. As lane 5 of Fig. 4B shows, faint bands intermediate in size to the integer multiples of the 370 bp a,b repeat unit were detected at the 555, 925, and 1,300 bp positions. The results with
HindII, as well as the lack of such intermediate bands in the digests produced by MboI/Sau3A (lanes 6 and 7), AceI (lane 16), and 6 other enzymes that recognize distinct sequences and produce B patterns with R. rattus satellite I DNA, were reproduced in another experiment (results not shown). The results with HindII could have several explanations, and among them is that some out of register interaction between the HindII variant and non-variant a,b repeat units has taken place. Another explanation would be that a HindII variant which arose in the a portion of the a,b repeat unit could have spread into the satellite I' tandem arrays of a' repeat units by homologous interactions that would be in register for these tandem arrays. Recent results on gene conversion in fungi (22) suggest that a and a' sequences share sufficient homology for this to occur. Whatever the explanation for the HindII results, the products (i.e., segments) of either out of register interaction among a,b repeat units or intersatellite interaction among a and a' sequences are rare in R. norvegicus and R. rattus.

The finding that the interaction among satellite family members is almost always determined by the length of the repeated unit suggests that divergent dimers which have arisen repeatedly during the evolution of many satellite DNAs are not the products of either out of register interactions or of monomeric arrays interacting as dimers (see Introduction). Since several studies have shown that significantly divergent satellite family members are present where satellite and nonsatellite DNAs join, presumably due to inefficient correction of the satellite members at these junctions (2, 28, 29), it is possible that divergent dimers originate at these regions. If most of an existing satellite array were deleted by recombination between members near but not at the ends of the array, the remnants could include a pair of nondivergent and divergent repeat units. If amplified by a saltatory event, the divergent pair would replace the deleted "monomeric" array. Although this scheme is speculative, it suggests that under the right conditions sequences near satellite/nonsatellite junctions are subject to amplification, a suggestion that might be experimentally testable.

We found a large difference between the loss of the EcoRI sites within the 185 bp units of rat satellite I and I' (Fig. 3); more EcoRI sites are lost from rat satellite I than from satellite I'. Furthermore, the loss of EcoRI sites in satellite I is not random. Two of the four sites were altered more often than the other two (see Figs. 5 and 6), and the two more readily lost EcoRI sites have the same topological relationship to the junction between the divergent portions of the dimer that comprises the satellite I array (see Fig. 5).
In spite of the differences between the loss of the EcoRI sites in satellite I and I', the number of restriction site variants is about the same in both satellite arrays. Fourteen of 24 restriction enzymes, each of which recognize a distinct sequence that is not present in the canonical sequence of satellite I', produced a B pattern, whereas 15 of 21 such enzymes produce B variants in R. norvegicus satellite I. This suggests that satellite I may not be more prone to accumulating sequence variants than satellite I', despite its greater tendency to lose EcoRI sites. Furthermore, by using a computer program (30) to locate sites where a single base change in the canonical sequence would produce the recognition site for the B pattern enzyme, we could find no correlation between the positions of these potential B sites, the strength of the B pattern, and the location of the most often lost EcoRI sites (data not shown). Since the altered EcoRI sites do not reside in mutational "hot spots," then the difference between the loss of EcoRI sites by satellites I and I' could be due to either inefficient correction or biased gene conversion, which, in turn, might be related to the length of the repeat unit or the juxtaposition of the diverged elements in the satellite I arrays.

One final note: using the computer program mentioned above, we were able to correlate the number of potential restriction sites for each of the enzymes tested and the production of a B pattern by any particular enzyme. Not surprisingly there was a good correlation between the number of potential sites and the detection of a B pattern; B patterns were most often found for those enzymes for which four or more potential sites within the repeat unit existed. In several cases nine or more potential sites exist, whereas two or fewer potential sites exist for each of the enzymes that give little or no digestion. The satellite segments that contain a mixture of non-variant members and a given B site variant have also been referred to as "segmental variants" (12), and the finding of B pattern production by a given enzyme in satellites of related species has been taken as evidence for the conservation of segmental variants (12). However, from the above it is clear that nucleotide substitutions at one of any number of sites can produce the "same" segmental variants. Therefore, conclusions about conservation of segmental variants based on restriction enzyme analysis alone may not always be appropriate.

1Present address: c/o Dr. David A. Epstein, Department of Industrial Engineering, Israel Institute of Technology, The Technion, Technion City, Haifa, Israel
2Present address: Bio-Rad Laboratories, 2200 Wright Avenue, Richmond, CA 94804, USA
REFERENCES