Eukaryotic DNA Diverges at a Long and Complex Pyrimidine-Purine Tract That Can Adopt Altered Conformations*

Richard F. Fowler and Dorothy M. Skinner†
From the University of Tennessee-Oak Ridge Graduate School of Biomedical Sciences, East Tennessee State University, and the Biology Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831

A domain exhibiting major sequence divergences among three cloned repeat units of a complex satellite DNA of the Bermuda land crab contains a repetitive poly-pyrimidine-poly-purine segment consisting of a long C-G tract embedded between runs of CCT*AGG and CGC*AC*GTGGC and their variations. The domain adopts at least two types of altered conformations that are markedly affected by pH and negative superhelical density; only one is sensitive to ionic strength. Supercoil-dependent distortions in helical structure are most pronounced at points of interruption in compositional bias in this domain and a similar although less extensive divergent domain nearby. Since the domain is the site of major sequence divergences among individual satellite repeat units, the altered conformations may be involved in site-specific recombination between repeat units, either those arranged in tandem or those scattered throughout the genome.

Segments of DNA with marked strand compositional bias toward either pyrimidines or purines (Py-Pu) exhibit unusual, and as yet uncharacterized, helical structures (Lee et al., 1979, 1984; Dybvig et al., 1983; Mace et al., 1983; Nickol and Felsenfeld, 1983; Schon et al., 1983; Shen, 1983; Cantor and Efstratiadis, 1984; Drew and Travers, 1984; Evans et al., 1984; Fowler et al., 1985; Pulleyblank et al., 1985). Py-Pu tracts are dispersed throughout the genomes of higher organisms (Birnboim et al., 1979); they are especially abundant in moderately repeated DNAs. Interest in Py-Pu tracts has increased since they have been found upstream from a number of eukaryotic genes (Cantor and Efstratiadis, 1984; Evans et al., 1984; McKeon et al., 1984). Nucleosome sensitivity in Py-Pu tracts has been correlated with transcription (Weintraub, 1985); left-handed (but non-Z) conformations are among the structures proposed to account for the nuclease sensitivity (Cantor and Efstratiadis, 1984). Our data suggest that triple-stranded helical structures (Lee et al., 1979, 1984) with hemiprotonated cytosine base pairs (Gray et al., 1984) might also be involved.

We have shown previously that domains containing repetitive Py-Pu tracts are the sites of some of the major sequence differences in an otherwise closely related family of molecules that comprise a complex satellite DNA of the Bermuda land crab Geocarcinus lateralis and that such domains adopt altered conformations under appropriate conditions (Fowler et al., 1985). This study examines the unusual conformations of one of these domains in greater detail.

In order to study the altered conformations of the divergent domain that contains the longest repetitive Py-Pu segments in the satellite without interference from other domains of unusual sequences, we subcloned a fragment that contains it (Domain III) and a similar but less extensive divergent domain (Domain IV; Fowler et al., 1985; Stringfellow et al., 1985). We describe here analyses of the subcloned fragment with S1, mung bean, and P1 nucleases under varying conditions of superhelical stress, ionic strength, and pH. Restriction sites at the boundaries of the satellite insert or outside it were used to locate regions of nuclease sensitivity.

EXPERIMENTAL PROCEDURES

Plasmids—A 463-bp TaqI fragment containing Domains III and IV of one cloned repeat unit of the satellite (RU, 2089 bp) was subcloned into the ClaI site of pBR322 (Stringfellow et al., 1985). Plasmids were recovered with the insert oriented with either the pyrimidine-biased strand (pT463-I) or purine-biased strand (pT463-II) clockwise with respect to pBR322. Plasmids were purified as described (Fowler et al., 1985).

Generation of Topoisomers—Supercoiled preparations of pT463-I and pT463-II were analyzed because of their different topological properties (see below), although their behavior was identical in the nuclease experiments described here; data from both are presented. Plasmids were relaxed with wheat germ topoisomerase I (Promega Biotec) in either Buffer I (50 mM Tris.HCl (pH 7.6), 10 mM MgCl, 0.1 mM dithiothreitol, 2 mM spermidine (Panayotatos and Wells, 1981; Singleton and Wells, 1982)) at a DNA concentration of 40 μg/ml or Buffer II (50 mM NaCl, 50 mM Tris-HCl (pH 7.9), 1 mM dithiothreitol, 1 mM EDTA, 20% glycerol) at a DNA concentration of 20 μg/ml, in the presence of up to 20 μg/ml ethidium bromide (EB). Incubations in Buffer I at 15, 29, or 37 °C with low concentrations of enzyme (0.01 unit/μg of DNA) led to complete relaxation of pT463-II, but not of pT463-I for which a 7-fold higher topoisomerase I concentration was required. Incubations in Buffer II at 37 °C with 1 unit/μg of DNA completely relaxed both. Topoisomerase activity was stopped by heating for 5 min at 65 °C. Samples were placed on ice, NaAc was added to a final concentration of 0.3 M, and the DNA was EtOH-p precipitated to remove EB. After reprecipitation from 10 mM Tris-HCl (pH 8.0), 1 mM EDTA, 0.3 M NaAc, with 3 volume of EtOH, the products were rinsed with 1 ml of 80% EtOH and dried. Similarly treated samples, but without topoisomerase, have the same −α as native (untreated) DNA, indicating complete removal of EB. Topoisomers were dissolved in 1 mM Tris-HCl (pH 8.0), 0.1 mM EDTA, and analyzed on 1.5% agarose gels containing up to 400 μM chloroguanidine phosphate (Shure et al., 1977). Native plasmids had close

*This research was supported by grants to D. M. S. from the National Science Foundation, National Institutes of Health, and the Exploratory Studies Program, Oak Ridge National Laboratory, and by the Office of Health and Environmental Research, United States Department of Energy, under Contract DE-AC05-84 OR21400 with Martin Marietta Energy Systems. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

†To whom correspondence and reprint requests should be addressed.

1 The abbreviations used are: Py-Pu, pyrimidine-purine tracts; bp, base pair(s); kb, kilobase pair(s); EB, ethidium bromide; EtOH, ethanol; NaAc, sodium acetate; Pu/Py, alternating purines and pyrimidines; −α, negative superhelical density.

2 S. Spengler, personal communication.
4.5), 1 mM MnCl₂ for 60 min at a DNA concentration of 40 pg/ml, and 1 mM MnCl₂ was replaced by MnCl₂ to increase specificity for the primary S1-sensitive region (Shishido, 1979; see Fig. 2). Plasmids were linearized with S1 nuclease (Pharmacia P-L), in buffers similar to those used for S1, and enzyme levels were adjusted to achieve approximately the same extent of linearization in all samples (see Fig. 5A). As a control on nuclease activity under conditions unfavorable for a particular enzyme, single-stranded M13 DNA was included in preliminary titrations. After treatment with S1, mung bean, or P1 nuclease, samples were placed on ice, made 5 or 12.5 mM in EDTA and rinsed with 80% EtOH.

Restriction Digests and Gel Electrophoresis—1.5% agarose and non-denaturing 7% polyacrylamide gels were as described (Fowler, et al., 1985). Denaturing 8% acrylamide: Bisacrylamide (9:1) gels were used for high resolution mapping, as well as for strand separations and sequence analysis (Maxam and Gilbert, 1980).

Diggests with Single-Strand-Specific Nucleases—S1 digestions were routinely performed at DNA concentrations of 40 μg/ml in 40 mM NaAc (pH 4.5), 1 mM MnCl₂, and 1 mM MnCl₂. In most S1 reactions, ZnSO₄ was replaced by MnCl₂ to increase specificity for the primary S1-sensitive region (Shishido, 1979; see Fig. 2). Plasmids were linearized with S1 nuclease (Pharmacia P-L) at a level of 1.5 units/μg of DNA for 60 min at 37 °C. For high resolution mapping of single-strand nicks, 0.15 units of S1/μg of DNA was used; incubation was at 37 °C for 1–60 min to achieve the desired extent of nicking. Digestions with 30–320 mM NaCl were in 40 mM NaAc (pH 4.5), 1 mM MnCl₂ for 60 min at a DNA concentration of 40 μg/ml, and 1.5 units of S1/μg of DNA.

Mung bean (Pharmacia P-L; Kroeker et al., 1976), P1 (BRL; Fujimoto et al., 1974), and S1 nuclease digestions over a range of pH were carried out in 50 mM NaCl, 50 mM NaAc (pH below 6) or 50 mM Tris-HCl (pH above 6), 1 mM ZnSO₄ for 15–150 min at 37 °C at a DNA concentration of 40 μg/ml. Enzyme levels were: S1, 1.5–290 units/μg; P1, 0.016–3.1 units/μg; mung bean nuclease, 3–200 units/μg. Incubation times and enzyme levels were titrated to achieve approximately the same extent of linearization in all samples (see Fig. 5A). As a control on nuclease activity under conditions unfavorable for a particular enzyme, single-stranded M13 DNA was included in preliminary titrations. After treatment with S1, mung bean, or P1 nuclease, samples were placed on ice, made 5 or 12.5 mM in EDTA and rinsed with 80% EtOH.

High Resolution Mapping of S1 Sites by Labeling of Restriction Sites Outside the Nicked Region—Native pT463-II was treated at 37 °C with 0.15 unit of S1/μg of DNA. Samples (5 μg) were taken at 1, 5, 15, 30, and 60 min; 0.25 μg of each was electrophoresed on agarose to monitor the increase of nicked molecules with time. Samples (2 μg) were then linearized at either the HindIII site 7 bp upstream from the satellite insert, or the EcoRI site 56 bp downstream from the satellite insert in the vector, and labeled at the protruding 5'-OH termini as described (Fowler et al., 1985) with γ-[^32P]ATP (ICN; 4500 Ci/mmol).

Linearized nicked plasmids labeled at their HindIII ends were cleaved with EcoRI; those labeled at their EcoRI ends were cleaved with HindIII. Restriction digestions were loaded onto 7% nondenaturing polyacrylamide gels from which 496-bp fragments containing 463 bp of DNA were transferred by 30 bp of the vector were recovered for each time point. The faster migrating (pyrimidine-biased) and slower migrating (purine-biased) strands of the pT463-II EcoRI/HindIII 496-bp fragment were recovered from denaturing gels and sequenced (Maxam and Gilbert, 1980). Fragments with 5'-[^32P]P)HindIII ends were electrophoresed on denaturing gels along with a sequencing ladder derived from the 5'-[^32P]pyrimidine-biased strand (HindIII end-labeled); fragments with 5'-[^32P]P)EcoRI ends were electrophoresed on denaturing gels along with a sequencing ladder derived from the 5'-[^32P]purine-biased strand (EcoRI end-labeled). 3'-End-labeling reactions with Shishido et al. (1982) were in a buffer described by Shen (1980) with all four α-[^32P]dNTPs (ICN; >3000 Ci/mmol); initial incubations were for 30 min at room temperature.

RESULTS

Characteristics of Domains III and IV—The DNA segment studied here is a 463-bp TaqI fragment (Fig. 1) that contains Domains III and IV of one cloned variant of the crab satellite subeloned into pBR322. The parent G + C-rich satellite (Skinner, 1967) has an average repeat unit of 2.07 kb (Lamarca et al., 1981); there are 16,000 copies per genome. Earlier, we compared the sequences of three cloned satellite variants: RU (2089 bp) close to the average length of most repeat units, TRU (1674 bp) truncated by an extra EcoRI site at its 3' end, and EXT (2639 bp) extended by a 5-fold amplification of a 142-bp segment (Skinner et al., 1982; 1983; Bonnewell et al., 1983) that is found scattered throughout the genome. Although the variants have distinctly different lengths, they are >95% homologous over a number of segments, one as long as 560 bp. Major sequence changes, including deletions from 15–108 bp, localized to four domains (I–IV) containing C-G, tracts where n = 11–23 and one domain (V) containing alternating purines and pyrimidines (Pu/Py). In Domains I, III, and IV, the C-G tracts are embedded in segments of heterogeneous repetitive Py-Pu similar strand bias. In RU, TRU, and EXT, these domains undergo S1 nuclease-sensitive structural alterations under negative superhelical stress. Furthermore, RU, TRU, and EXT appear to assume their own characteristic higher order structures in any given divergent domain (Fowler et al., 1985).

Domain III contains the most extensive repetitive segment (194 bp) found in RU as well as in TRU, where deletions shorten it to 178 bp. In EXT, it is shortened to only 71 bp. To determine the most extensive repetitive segment, RU, TRU, and EXT were scanned using the TTTA tetramer as an end probe (Stringfellow et al., 1985). Further evidence of marked sequence changes here (Fowler et al., 1985). In RU, TRU, and EXT, the domain begins on one strand with a series of CCTs and variations followed by a long C tract and repetitions of CGCAC pentamers and variations; pyrimidine bias is interrupted between the CCT segment and the C tract. In RU, the interruption is included in a TTTA tetramer embedded between a (CCT)₄ and a C₈ tract; TTTA is found nowhere else in RU, nor anywhere in TRU or EXT. Our data suggest that the TTTA possesses unusual properties in this environment. Although less extensive than Domain III, Domain IV shares

certain characteristics in RU, TRU, and EXT. It, too, consists of a C tract embedded between repetitive segments of similar bias ("C$_{\text{C}}$, T") with interruptions in strand bias at the 5' end of the C tract. The major sequence change in Domain IV is an 11-bp insertion in EXT.

**Altered DNA Conformations Detected by S1 Nuclease—**

Unusual DNA conformations were first detected in experiments on native supercoiled pT463-I and -II linearized with high levels of S1 nuclease. The positions of S1 attack were mapped relative to the EcoRI site in pBR322 (Fig. 2). A cluster of nuclease-sensitive sites spanning approximately 100 bp mapped to the extensive repetitive Py-Pu domain. The strongest signal localized to the TTAA region between the (CCT-AGG)$_{15}$ and (C-G)$_{15}$ tracts; weaker signals were observed at the boundaries of the cluster, ~50 bp in both directions. This same banding pattern for Domain III is seen when RU, the full length repeat unit, is treated similarly (Fowler et al., 1985). Substitution of Mn$^{2+}$ for Zn$^{2+}$ suppressed S1 nuclease sensitivity in Domain IV (Fig. 3B; 0.21-kb fragment), boundary of the (CCT-AGG)$_{15}$ tract (0.15-kb fragment), and Domain IV (0.45-kb fragment) are indicated. M, marker DNA, φX174 digested with HinfI (lane 9). B, as in A except $\sigma$ = 0.05 (lane 10); M, as in A (lane 11).

**Fig. 2.** Nondenaturing gels of linear or supercoiled forms of pT463 subclones treated with S1 nuclease in the presence of either Zn$^{2+}$ or Mn$^{2+}$. A, native plasmids were treated with S1 (1.5 units/μg of DNA) before (lanes 3, 4, 8, and 9) or after (lanes 1, 2, 6, 7) digestion with EcoRI. Brackets, S1-specific bands that map to Domain III; since pT463-I and pT463-II have opposite orientations, their banding patterns are inverted. Arrow (lane 3 and 4), signal attributed to divergent Domain IV; analogous fragments in lane 8 and 9 are ~40 bp, too small to be seen in this photograph. Lane 5, marker DNA, φX174 digested with HinfI. B, overloaded gel (20 μg of DNA/lane) of native plasmids treated as in lanes 4 and 9 in A to permit visualization of minor bands. Bold arrows, signals from TTAA region (lane 1, 0.21-kb fragment; lane 2, 0.29-kb fragment); brackets, signals from surrounding tracts in Domain III.

**Fig. 3.** Coarse mapping of S1-sensitive sites in the RU-TaqI 463-bp fragment at increasing $\sigma$. A, nondenaturing gel displaying topoisomers of pT463-I (σ = 0.001–0.054), treated with S1 nuclease (1.5 units/μg of DNA, 60 min) in the presence of Mn$^{2+}$, followed by EcoRI digestion (lanes 1–8). Signals from TTAA region (0.21-kb fragment), boundary of the (CCT-AGG)$_{15}$ tract (0.15-kb fragment), and Domain IV (0.45-kb fragment) are indicated. M, marker DNA, φX174 digested with HinfI (lane 9). B, as in A except $\sigma$ = 0.001–0.146. N, native pT463-I, $\sigma$ = 0.05 (lane 10); M, as in A (lane 11).

the TTAA region as the most sensitive supercoiled-dependent S1 site in the RU variant and is consistent with observations that this domain competes favorably for S1 nuclease with other potential supercoil-driven structures such as Z-DNA and cruciforms (Fowler et al., 1985). A very similar Py-Pu site located 180 bp 5' to the chicken pro-a2(I) collagen gene requires $\sigma$ = 0.024 to induce S1-sensitive structural alterations (Finer et al., 1984), in excellent agreement with our observations. Although Domain IV (Fig. 3, 0.45-kb fragment) of the satellite and regions of pBR322 become more nuclease-sensitive with increasing $\sigma$, digests of pT463 topoisomers with $\sigma$ up to 0.14–0.15 detect no additional altered structures once the signal from the TTAA region appears (Fig. 3B; 0.21-kb fragment).

An **Altered Conformation Detected by S1 Exists at NaCl Concentrations up to 320 mM**—Asionic strength increases, duplex stability becomes greater and base mismatches due to out-of-register strand slippage in repetitive DNA sequences are suppressed. Because S1 sensitivity in a sea urchin histone gene spacer region that mapped to a tract of repetitive TC-AG was abolished at 120 mM and higher NaCl (Hentschel, 1982), we examined the effect of increased NaCl concentrations on nuclease sensitivity in the satellite fragment.

Digestions were performed in S1 buffer with 50–320 mM NaCl and enzyme at a level of 1.5 units/μg of DNA (Fig. 4). Samples were linearized at all salt concentrations, although cutting decreased somewhat as salt concentration increased, presumably due to slight inhibition of S1 (Vogt, 1973; Hentschel, 1982). Restriction enzyme mapping of the S1-linearized samples revealed the characteristic signals from the TTAA region (Fig. 4A; 0.21-kb fragment) and surrounding repetitive structures at NaCl concentrations as high as 320 mM. As ionic strength increased, signals from the TTAA region and the boundary of the (CCT-AGG)$_{15}$ tract (0.15 kb) persisted, but...
the signal from the 3' boundary of the (CGCAC-GTGGCG)$_3$ tract (0.24 kb) diminished (Fig. 4B). This suggests that different structural alterations occur on either side of the TTAA region and that strand slippage does not account for the high nuclease sensitivity of the TTAA region or the boundary of the (CCT.AGG)$_5$ tract. At higher salt concentrations, an additional fragment slightly shorter than that observed at pH 4.3-7.3 (lanes 2-5) appeared in Domain IV yielding a second primary cleavage site for a given nuclease; *, significant sensitivity; **, variations on sequences shown; $\Theta$, tracts of C; $\Psi$, 100% pyrimidine-biased domain but not simple repeating blocks, regions of RU conserved in TRU and EXT. Sequences of divergent Domains III and IV and a Py-Pu domain which has not undergone divergence are represented in abbreviated form above their respective positions; "", variations on sequences shown; $\Theta$, tracts of C; $\Psi$, 100% pyrimidine-biased domain but not simple repeating 

**TABLE I**

<table>
<thead>
<tr>
<th>Nuclease</th>
<th>pH 4.3</th>
<th>pH 5.2</th>
<th>pH 6.5</th>
<th>pH 7.3</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td>+</td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mung bean</td>
<td>+</td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P1</td>
<td>+</td>
<td></td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>

**FIG. 4.** Coarse mapping of SI-sensitive sites in the RU-TaqI 463-bp fragment at increasing salt concentrations. A, nondenaturing gel displaying native pT463-I ($\gamma = 0.05$) treated with S1 nuclease at increasing NaCl in the presence of Mn$^{2+}$, followed by EcoRI (lanes 1-6). Signals from the TTAA region (0.21-kb fragment), the boundary of the (CCT.AGG)$_5$ tract (0.15-kb fragment), and from Domain IV (fragments centered at 0.45 kb) are indicated. B, plot of data from a Zenieh soft laser microdensitometer scan of the gel in A showing signal intensities versus NaCl. Intensities are expressed as per cent combined signal strengths detectable within the size range of the fragment. $\bullet$, 0.15-kb fragment; $\Theta$, 0.21-kb fragment; $\Theta$, 0.24-kb fragment.

**FIG. 5.** pT463-II digested with S1, mung bean, and P1 nucleases over a range of pH. A, 1.5% agarose gel showing 0.5-µg aliquots of DNA from nuclease digests. Nicked (N), linearized (L), and supercoiled (SC) forms are indicated. B, 7% polyacrylamide gel of DNAs (9.5 µg/lane) after digestion with HindIII. Native pT463-II (lanes 1, 6, and 11) was digested with mung bean (lanes 2-5), S1 (lanes 7-10), or P1 (lanes 12-15) nucleases at pH 4.3 (lanes 2, 7, and 12), pH 5.2 (lanes 3, 8, and 13), pH 6.5 (lanes 4, 9, and 14), or pH 7.3 (lanes 5, 10, and 15) under conditions that produce approximately the same extent of linearization. Lanes in B are aligned with corresponding lanes in A. Brackets, divergent Domains III and IV; $\Theta$, TTAA region; $\Theta$, central region of (CCT.AGG)$_3$ tract; $\Theta$, boundaries of repetitive segments; M, marker DNA for B. $\Theta$, X HindIII. C, diagram of pyrimidine-biased strand of RU-TaqI 463-bp fragment showing areas of cleavage by nucleases at pH 5.2. Filled bars, regions of RU conserved in TRU and EXT. Sequences of divergent Domains III and IV and a Py-Pu domain which has not undergone divergence are represented in abbreviated form above their respective positions; **, variations on sequences shown; $\Theta$, tracts of C; $\Psi$, 100% pyrimidine-biased domain but not simple repeating blocks. Vertical lines spaced proportionally, (CCT)$_3$ and (CGCAC) tracts. Brackets below diagram, areas of sensitivity to S1, mung bean (MB), and P1 nucleases; $\bullet$, significant sensitivity; $\Theta$, primary cleavage site for a given nuclease; $\Theta$, minor signal. Positions were mapped relative to either the EcoRI site of pT463-I or HindIII site of pT463-II (both in the vector, to the left of the diagram).

hydrolysis was observed with N. cressa nuclease under a variety of conditions (see also Selleck et al., 1984). S1 and mung bean nucleases produced similar band patterns over the entire pH range. At pH 4.3 or 5.2, cleavages most frequently occurred in Domain III near the TTAA region with secondary sites at the boundaries of the repetitive segments, ~50 bp in either direction. Signals from the boundary of the (CGCAC-GTGGCG)$_3$ tract were less prominent at pH 5.2 than at 4.3. Other sites of nuclease sensitivity were found in Domain IV and at other Py-Pu regions where bias is
Eukaryotic DNA Diverges at a Pyrimidine-Purine Tract

abruptly interrupted. In addition to random cleavages at pH 6.5 or 7.3 with both S1 and mung bean (Table I and Fig. 5B, lanes 2–5, 7–10), very low levels of specific cleavage in the (C-G)$_{12}$ tract and boundary of the (CCT-AGG)$_{15}$ tract were seen with mung bean nuclease. In the absence of cleavages in the satellite insert, or with pBR322 as a control, sites mapped to the pBR322 cruciforms (Lilley, 1980).

P1 nuclease produced a different pH-dependent band pattern in Domain III from that produced by S1 or mung bean nucleases (Table I and Fig. 5B, lanes 12–15). At pH 4.3, cleavages occurred most frequently near the TTA region and the boundary of the (CCT-AGG)$_{15}$ tract, but not at the boundary of the (CGCAC-GTGCG)$_{12}$ tract. By contrast, at pH 5.2, the central region of the (CCT-AGG)$_{15}$ tract was most sensitive; these results may reflect the pH-dependent change in substrate specificity of P1 (Fujimoto et al., 1974). Although strand specificity of early nicking events that produce these fragments has been determined only for S1 (see below), it appears that P1 recognizes only the altered conformation on the pyrimidine-biased strand because cleavages map to the center of the (CCT-AGG) tract (Fig. 5B, lane 13). All three enzymes yield two fragments in Domain IV, however, they produce significantly different amounts of each at different pH values (Fig. 5B). The smaller of the two fragments exhibits increased sensitivity to S1 at high ionic strengths (Fig. 4A). As with S1 and mung bean, at pH values higher than 5.1, P1 cleaves at the pBR322 cruciforms or at random.

These results indicate that the altered conformations in Domains III and IV depend on an acidic environment and suggest the involvement of protonated C residues. Methylation of cytosines allows such structures to form at physiological pH (Lee et al., 1979, 1984; Gray et al., 1984). All three nucleases recognize the boundary of (CCT-AGG)$_{15}$ at pH 4.3 and 5.2, while the boundary of (CGCAC-GTGCG)$_{12}$ is less sensitive at pH 5.2 or at higher ionic strengths and is not recognized by P1 under any of these conditions (Figs. 4 and 5B). Clearly, two types of altered structure exist at the extremes of the repetitive segments.

High Resolution Mapping with S1 Nuclease—A more detailed examination of Domain III was undertaken using a 10-fold lower nuclease concentration (0.15 unit/µg of DNA) and limited digestion times to minimize linearization and avoid nibbling; conditions were chosen that favored scission of only one strand to produce pT463 populations with a single nick per molecule. Since the data reflect populations of supercoiled plasmids nicked at a given site, it is possible to assign precise locations and determine strand specificity of the most susceptible sites (see “Experimental Procedures”), although relative intensities cannot be compared between strands. After extended digests with higher concentrations of S1 nuclease as seen in Figs. 2–5, linearized molecules accumulated as a result of secondary cleavages opposite the original nick, and nibbling resulted in populations of molecules trimmed to the boundaries of the more sensitive repetitive regions. The pattern observed was reproducible and positions were assigned by comparing to standard size markers and by comparing directly to sequence ladders (Fig. 6).

The TTAA region was most frequently attacked on both strands (Fig. 6); the two A residues are the only purines in a stretch of 80 pyrimidines (Fig. 1, fourth row) and helical structure between the two repetitive pyrimidine segments appeared to be highly perturbed (see “Discussion”). The high resolution nicking pattern suggested that the TTAA segment is in a pivotal position where nicking specificity exchanges between strands of opposing base compositional bias. In addition, the (C-G)$_{12}$ tract exhibited differential sensitivity at either side of its midpoint, and the purine strand of the (CCT-AGG)$_{15}$ tract exhibited sensitivity at certain trimers and at its boundary where bias is interrupted. Sensitivity at the boundary of the (CGCAC-GTGCG)$_{12}$ tract was seen after extensive digestion (Figs. 2–5; additional data with labeled fragments not shown); this could be attributed to strand slippage in unstrained molecules after nicking at a nearby site.

On the pyrimidine-biased strand, a series of nicks began at the interruption in bias and extended through the 5′ half of the C tract but not beyond its midpoint. A weaker signal was observed corresponding to the 5′ boundary of the (CCT-AGG)$_{15}$ tract (Fig. 6, A and C). Other regions of sensitivity on this strand also corresponded to interruptions in pyrimidine bias: (i) a nondivergent Py-Pu domain of 48 bp that is 83% biased and that includes C$_3$ and C$_4$ tracts was sensitive at AGA, a point where bias is once again interrupted (Fig. 1, rows 5 and 6; Fig. 6A, asterisk) and (ii) a divergent domain (IV) that is interrupted by an A residue immediately 5′ to a C tract.

On the purine-biased strand, highest sensitivity was at TTAA, around residue 15 of the G tract (opposite a region of sensitivity on the pyrimidine-biased strand), and at the 3′ end of the (AGG)$_{15}$ tract; relatively high sensitivity was seen between the 6th and 7th and 11th and 12th AGGs, as well as the 5′ half of the G tract (Fig. 6, B and C). Other sites on this strand also corresponded to domains in Domain III where bias is interrupted.

DISCUSSION

We conclude that at least two types of altered conformation within Domain III are recognized by the nucleases. One is associated primarily with the pyrimidine-biased strand and the other with the pyrimidine-biased strand; both depend on relatively low negative superhelical density (−σ = 0.02–0.03) and slightly acidic pH (below pH 6). An asymmetric nicking pattern was observed with nuclease sensitivity on each strand extending downstream in repetitive regions from a point of interruption in base compositional bias; these altered conformations appear to form in a concerted transition. The conformation in the (CCT-AGG)$_{15}$ region differs from the other structure in its sensitivity to all three nucleases and its stability in high salt (Hentschel, 1982) which argues against strand slippage. By contrast, the altered structure in the (CGCAC-GTGCG)$_{12}$ region was suppressed by high NaCl, recognized by only two of the three nucleases, and affected by pH; it did not exhibit sensitivity in high resolution experiments. Domain IV was similar to Domain III in its sensitivity to the nuclease used and in the site nicked in response to changes in pH or ionic strength. The high level of sequence divergence seen in Domain III suggests that the altered conformations might be involved in recombination between repeat units arranged in tandem or scattered throughout the genome.

Hypotheses to Account for the Pattern of Nuclease Sensitivity in Domain III—The nicking pattern seen in Domain III encompassed more than 100 bp including both strands, compared to less than 20 bp for most other Py-Pu nuclease-sensitive sites reported. That the S1 sites were left-handed DNA was concluded from thermodynamic calculations which arrive at 8–18 bp as the length of the most stable structure (Cantor and Efstratiadis, 1984), in good agreement with previously reported values. Furthermore, the low −σ (0.02–0.03) required to induce an altered conformation in the crab satellite site is not sufficient to drive a B→A transition even in very long tracts of GC-GC (Singleton et al., 1982; Rich et al., 1984). It seems reasonable to assume that it might require
FIG. 6. High resolution mapping of strand-specific nicks with S1 nuclease. A, denaturing gel locating S1 sites on the pyrimidine-biased strand. Native pT463-II was treated with low levels of S1 for 1–60 min in the presence of Mn$^{2+}$ (lanes 1–5); HindIII-digested, and 5'' end-labeled (see "Experimental Procedures"); sequencing ladder (lanes 6–9) of the pyrimidine-biased strand of the pT463-II EcoRI/HindIII fragment, 5'' end-labeled at its HindIII site. (C)GCAG$_b$, G$_b$, and (CCT)$_b$ tracts, boxed; untreated fast strand, F, (lane 10); slow strand, S (lane 11); signals from nondivergent Py-Pu domain, *, marker, ϕX174 digested with HindIII and 3'' end-labeled, M (lane 12). B, denaturing gel locating S1 sites on the purine-biased strand. Native pT463 was treated as in (A) except it was EcoRI-digested (lanes 1–5); sequencing ladder (lanes 6–9) of the purine-biased strand of the pT463-II EcoRI/HindIII fragment 5'' end-labeled at its EcoRI site. (AGG)$_b$, G$_b$, and (GTGCG)$_b$ tracts, boxed. C, interpretation of high resolution mapping experiments. Autoradiograms of gels (as in parts A and B) of several time points were scanned with an LKB UltroScan soft laser densitometer interfaced with an Apple Ile computer, data were analyzed with the accompanying LKB software to estimate relative intensities. Signals were normalized to the height of the most intense band in a given lane. Vertical lines reflect intensity of the signal. Positions were assigned by comparison to sequencing ladders such as those in A and B with an accuracy of ±1 bp.

comparably elevated values of $-\sigma$ to drive the formation of a type of left-handed helix other than Z-DNA (Gupta et al., 1980). Nevertheless, polyclonal anti-Z-DNA antibodies bind to Py·Pu regions under conditions known to promote structural transitions in those regions (Pulleyblank et al., 1985).

Our observation of such a long nicked segment and another report of nicks that span 30 bp in a similar pattern (Finer et al., 1984) implicate other altered structures such as segments of underwound right-handed helices. If such tracts assume conformations with lowered helical twist, the transition should require relatively low torsional stress since no gross rearrangements in helical structure are necessary (Calladine and Drew, 1984). Long regions of underwound right-handed DNA would be effective sinks of torsional free energy. If helical twist is 16°/bp (as is the lowest extreme for right-handed A-DNA; Dickerson, 1983b), rather than 36° as in "ideal" B-DNA, we calculate that the resulting S1-sensitive structure in the ~100-bp region in the crab satellite would be underwound by (36° – 16°) \times 100 bp)/360 = 5.5 helical turns and would relax to a covalently closed plasmid by an equal number of superhelical turns. Since C·G tracts favor the A conformation (Calladine and Drew, 1984), it is reasonable to expect A-DNA in this segment.

Py·Pu DNAs with repeating sequences are believed to form triple-stranded complexes at and below pH 6 (Lee et al., 1979). The formation of these complexes involves the dismutation of two homologous double helices to form a triple-stranded structure in which the protonated cytosine-rich strand from one duplex winds down the major groove of the other; it is thus protected from attack by S1 nuclease (Lee et al., 1979, 1984). The free purine-rich strand assumes an independent conformation (Lee et al., 1979). That such triplexes may be biologically significant is supported by the recent demonstration (Lee et al., 1984) that they exist at physiological pH (below 8) if the cytosines are methylated. In eukaryotic DNAs, Py·Pu tracts are abundant (Birnboim et al., 1979) and the occurrence of 5 methylcytosines closely parallels that of pyrimidine tracts (Ringer et al., 1985).

Domain III should be amenable to the adoption of an intermolecular triplex under the conditions of nuclease digestions (pH 4.3 and 5.2). Triplex formation is consistent with results obtained for the (CCT·AGG)$_{16}$ tract: (i) the sequence is a repetitive Py·Pu and the purine strand is selectively cut by S1 nuclease; (ii) S1, mung bean, and P1 nuclease sensitivities are pH-dependent; (iii) elevated salt concentrations do not abolish nuclease sensitivity; (iv) boundaries of nuclease sensitivity delimit this tract; (v) the purine-biased strand has a tendency to aggregate when isolated, as if adopting an independent conformation.

Although triplex formation between two (CCT·AGG)$_{16}$ tracts is one mechanism that could account for interactions between homologous molecules, it would require that the two pyrimidine-biased strands be antiparallel. If so, the homologous residues in the (C·G)$_{22}$ and (CGCAC·GTGCG)$_{16}$ tracts would be held in opposite orientation, forbidding direct interactions in those regions. One possible structure for the odd purine strand in the putative intermolecular triplex could be a hairpin-like conformation in which the strand is folded back
upon itself and stabilized by alternate base-pairing schemes between purines involving the N-7 position (Cantor and Schimmel, 1980; Fowler et al., 1985; Pulleyblank et al., 1985).

Under topological constraint, this could lead to misalignment with the homologous residues on the pyrimidine-biased strand which could be relieved by strand slippage, a model consistent with suppression of nuclease sensitivity in the pentamers at increased NaCl concentrations (Fig. 4).

Although this model cannot be ruled out, preliminary experiments designed to demonstrate an interaction between two individual pT463 molecules through triplex formation yielded no evidence for a bimolecular aggregate: (i) lowering the DNA concentration from 40 µg/ml to 5 µg/ml in S1 nuclease digests had no effect, (ii) electron microscopy of pT463 and other plasmids spread at pH 4–5 in S1 nuclease buffer revealed no evidence for intermolecular interactions, (iii) attempts to demonstrate an interaction by cross-linking with (dl)-diepoxybutane (Castleman et al., 1983) failed because the reagent nicked the supercoiled DNAs as it does linear DNA (Kang et al., 1985).

A cloned synthetic insert of 45-bp TC-AG undergoes a structural transition at mildly acidic pH and physiological να, which renders only the pyrimidine strand of the insert nuclease-sensitive and protects from alkylolation N-7 positions of guanines on the purine strand. These results are attributed to Hoogsteen G-C base pairs in alternation with Watson-Crick A-T base pairs (Pulleyblank et al., 1985). If this model applies to the crab (C-G) and (CGCAC GTGCG)5 tracts, residues of the C9 tract would be Hoogsteen-paired with those of the C9 tract, and Hoogsteen G-C pairs would alternate with Watson-Crick A-T pairs in the pentamer region.

Alternatively, the pentamer region might be capable of bending through periodic spacing of base steps as in kinetoplast DNA (Wu and Crothers, 1984). In the pentamer region of RU, there are eight CpC (GpG) base steps spaced at 5-bp intervals, one at each of eight half-turns of B-DNA. A slight kink at each of these steps would result in significant curvature of the helical axis; although seen by S1 and mung bean nuclease and naked supercoiled DNA correlate with regions of RU, there are eight CpC (GpG) base steps spaced at 5-bp intervals, one at each of eight half-turns of B-DNA. A slight kink at each of these steps would result in significant curvature of the helical axis; although seen by S1 and mung bean nuclease, these kinks might not have been detected by P1.

Since stacking interactions of purines are a significant stabilizing force (Calladine and Drew, 1984), the purine-biased strand may be in a relatively nondistorted conformation, hence not recognized by the nuclease except in the putative triplex region and in the G tract. Another possibility is that the pyrimidine-biased strand could assume an independent double-helical conformation stabilized by hemiprotonated cytosine base pairs (Gray et al., 1984).

Calculations of helical twist, base pair roll, and main chain torsion angles (Calladine, 1982; Dickerson, 1983a) can predict structural parameters for a variety of sequences of biological importance such as promoters of prokaryotic (Nussinov, 1984) and eukaryotic genes (Nussinov et al., 1984). Further, they suggest that such deviations are essential for site-specific recognition by proteins. These calculations in the TTAA region (CTTAAC<sub>T</sub>) match the consensus sequence, GGPYCAATCT, for the CAT box (Nussinov et al., 1984). In fact, downstream from and including the 2 A residues, the correlation is exact. Calculations for the repetitive CGCAC pentamers also predict extreme deviations from B-helical geometry. External features of the double helix dictated by the nucleotide sequence are recognized by DNaase I, S1, and micrococcal nuclease (Drew, 1984), as well as RNA polymerase (Nussinov, 1984). These enzymatic probes detect regions of structural perturbation which distinguish their substrates in the DNA. Furthermore, nuclease-sensitive sites in active chromatin and naked supercoiled DNA correlate with regions of large deviations from B-helical geometry and these features may be recognized by regulatory factors (Weintraub, 1983). The altered conformations in the crab satellite may also be specific recognition sites.

Other Observations Suggesting an Altered Conformation in Domain III—Restriction fragments that contain Domain III exhibit anomalous behavior (Fowler et al., 1985). When high concentrations of such fragments were loaded on preparative denaturing gels, they migrated as several distinct bands, suggesting the occurrence of intermolecular interactions. The purine-biased strand aggregated when isolated, remained at the origin of denaturing gels, and was recovered only after prolonged treatment with NaOH. The pyrimidine-biased strand of fragments generated by digestion at the FnuDII site in the second interrupted pentamer (CGCGC) had heterogeneous termini, while the purine-biased strand appeared normal. Fragments of the pyrimidine-biased strand upstream from this site consisted of five bands, one band of the expected size accompanied by four less intense bands apparently decreasing in length by a single nucleotide. In downstream fragments, the expected band was surrounded by two additional bands, one from either side of the cleavage site; this pattern was also observed when CGCGC sequences at other locations in the satellite were cleaved by FnuDII. The conformation of the DNA within the recognition sequence of a restriction enzyme may determine the exact site of hydrolysis (Nishimura and Tsuibo, 1984).

Biological implications for the altered DNA conformation in Domain III are evident in the high level of sequence divergence found there. Domain III is the most highly divergent region in the family of satellite repeat units. Recombination in specific genes has been claimed to occur at short direct repeats through strand slippage, predicted from sequence analyses (Esfratidias et al., 1980; Slightom et al., 1980; Nikaido et al., 1981; Lebo et al., 1983; Hasson et al., 1984) or detected by S1 sensitivity at points of sequence differences (Hentschel, 1982). Internal amplifications of segments of the G + C-rich land crab satellite (Bonnewell et al., 1983) or complete copies of simple sequence satellites (Fowler and Skinner, 1985) are found between inverted repeats. Some of the amplified segments of the G + C-rich satellite, possibly entire repeat units, are scattered throughout the genome. Our demonstration of nuclease sensitivity at points of sequence differences among RU, TRU, and EXT suggests that other altered DNA conformations in addition to strand slippage may be sites of recombination. Although we have not demonstrated interactions between repeat units, if recombination between specific domains of the satellite occurs, it could take place not only at tandemly arranged repeat units (i.e. satellites), but at other loci as well.

Acknowledgments—We thank Dave Allison for EM studies and Frank Kenney for use of the LKB densitometer.

REFERENCES


Eukaryotic DNA Diverges at a Pyrimidine–Purine Tract