Volume 9 Number 22 1981 Nucleic Acids Research

The primary structure of bovine satellite 1.715

Claire Gaillard, Janine Doly, Jordi Cortadas and Giorgio Bernardi

Laboratoire de Génétique Moléculaire, Institut de Recherche en Biologie Moléculaire, 75005 Paris, France

Received 21 September 1981

ABSTRACT: The primary structure of the 1402 bp repeat unit of bovine satellite 1.715 has been determined using a dimer inserted at the SalI site of plasmid pBR322 and cloned in E. coli. In contrast with bovine satellites 1.706, 1.720b and 1.711a, the 1.715 satellite has a complex sequence with no obvious internal short prototype repeat. The sequence consists however of repeats ranging in length from 6 to 13 nucleotides. In addition, the hexanucleotide, AGATGA, present in the prototype sequences of satellites 1.706, 1.720b and 1.711a, is found in satellite 1.715 in repeats as long as, or longer than, 8 nucleotides, establishing a homology link among those satellites on one hand and satellite 1.715 (and the related satellite 1.711b) on the other. In turn, this suggests a common evolutionary origin. A comparison of the maps for 16 restriction enzymes of cloned and uncloned satellite indicates very little sequence divergence among the repeat units of the latter, most of the differences being due to methylation.

INTRODUCTION

Previous investigations from our laboratory, involving density gradient centrifugation in the presence of DNA ligands (1-3), have shown that 23% of the bovine genome is formed by 8 satellite DNAs (3, 4). This allowed us to study their long-range periodicities (5), extending previous results (6-10), and to show that the bovine satellites share identical repeat lengths, a fact suggesting a common evolutionary origin.

The problems of the internal periodicities and of the evolutionary relationships among bovine satellite DNAs have been studied more recently at the nucleotide sequence level in a series of investigations. These showed that the 1.706 satellite, which is organized in an unusual structure of superimposed long- and short-range repeats (11), consists of different variants of a basic 23 nucleotide sequence, which is itself composed of a dodecanucleotide and a related undecanucleotide (12). The basic prototype repeat of the 1.706 satellite exhibits a very high homology to the 46 nucleotide repeat of the 1.720b satellite, which is in turn composed of two related 23 nucleotide sequences (13). On the other hand, the 1.711a satellite also largely consists

of variants of a basic 23 nucleotide repeat closely related to the prototype sequence of the 1.706 satellite; the rest of the repeat unit of the 1.711a satellite consists of an unrelated 611 nucleotide sequence, which is not internally repetitive. A similar insert exists in the repeat unit of the 1.711b satellite, which is otherwise very close to the repeat unit of the 1.715 satellite (14). While the 1.706, 1.720b and 1.711a satellites have a common origin well established by the homologies found at the sequence levels, it is not yet known whether the two related satellites 1.711b and 1.715 also share a sequence homology with them. In the present work we have determined the primary structure of the 1.715 satellite after cloning in E. coli, and we have compared the restriction map of the cloned repeat unit with those of non-cloned satellites. The aims of these investigations have been the following. First of all, we wanted to elucidate the point first mentioned concerning the evolutionary relationship of the 1.715 (and of the related 1.711b) satellite with those previously studied. Second, we considered of interest to obtain sequence information in the case of a satellite in which restriction analysis indicated a complex structure with no simple short internal repeat; in fact, the available sequence data essentially concern satellites formed by very short internal repeats. Third, we wanted to check the existence of sequence heterogenities which have been suggested for several satellites and in particular for the 1.715 satellite (10, 15).

MATERIALS AND METHODS

DNAs. The 1.715 calf satellite DNA was prepared as previously described (3, 4). Plasmid pBR322 DNA was prepared according to Clewell and Helinski (16).

Digestion with restriction enzymes and gel electrophoresis. Restriction enzymes were obtained from BioLabs (Beverly, MA) or Bethesda Research Laboratory (Rockville, MD) and were used as recommended by the suppliers. Agarose gel electrophoresis was done as described (17). Polyacrylamide gels (0.12 x 20 x 40 cm) were run in Peacock's buffer (18). Ethidium bromide staining and photography of the gel were done according to Prunell et al. (19).

Construction of recombinant plasmids. Ligation of the SalI satellite digest to pBR322 was done with T4 DNA ligase from PL-Biochemicals (Milwaukee, WI) as recommended by the suppliers. Transformation of competent cells of E. coli strain HB101 with the recombinant plasmids was done according to Bolivar et al. (20). Transformants were plated on ampicillin containing media, selected for their tetracycline sensitivity and purified by streaking on ampicillin plates.

DNA sequencing. Isolation and purification of cloned DNA fragments, 5' terminal labeling and sequence analysis were performed according to the procedures of Maxam and Gilbert (21 and personal commun.). Sequencing gels were 6%, 8% and 20% in polyacrylamide; 7M or 8.3 M urea was used. Gel dimensions in early experiments were 1 x 200 x 400 mm, in later experiments 0.35 x 200 x 400 mm or 0.35 x 200 x 900 mm for long range sequence analysis. Autoradiography was done at -20°C overnight or at room temperature after previous fixation of the gel with 0.3% CTAB (cethyl trimethyl ammonium bromide) for longer exposures (Kryoukov, personal communication). Kodak NS2T films were used. DNA segments 150 to 250 bp long were sequenced in each experiment. The strategy adopted was such that the restriction sites used for the 5' end labeling were always contained in another sequenced fragment. Both strands were sequenced on 80% of the satellite repeat unit. In several cases this was necessary because of sequence artefacts which appeared in one strand only. These consisted in a non-specific breakage which made sequence reading impossible over 2-8 nucleotides at specific sites, namely at positions 915, 1252 and 1341 on the light strand (the last two positions correspond to a Sau3A and an AluI site, respectively) and positions 295 and 1005 on the heavy strand (corresponding to a Sau3A and a HaeIII site, respectively). These artefacts were not dependent upon the enzyme used in the preparation of the fragment and were always close to tetra-C sequences.

RESULTS

Cloning and sequencing of the 1.715 satellite DNA. The satellite, purified as described in the preceding section, was digested to completion by SalI restriction endonuclease. The restriction pattern of the digest consisted of a series of bands corresponding to a 1,400 bp unit and its oligomers. The satellite digest was ligated to SalI-digested pBR322 plasmid. Twelve clones from E. coli cells transformed by recombinant plasmids were analyzed. Seven of them contained as an insert in their plasmids the 1,400 bp fragment and five the corresponding dimer. One of the latter clones was used for all subsequent experiments; a schematic restriction map of the plasmid is given in Fig. 1.

The EcoRI fragment from the cloned SalI dimer and its flanking segments were sequenced as indicated in Fig. 2. The primary structure of the repeat unit of the EcoRI fragment is given in Fig. 3; the flanking regions (see Figs. 1 and 2) were identical in sequence with the corresponding segments located within the EcoRI fragment.

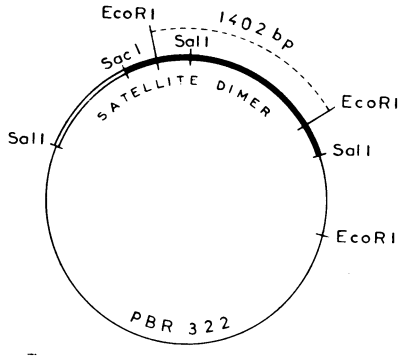

Fig. 1. Schematic restriction map of the hybrid plasmid used for sequencing the 1.715 satellite. The filled-in segment indicates the region of the satellite dimer which was sequenced (see Fig. 3).

Computer analysis of the sequence of the 1.715 satellite DNA. Table 1 presents the base composition and the frequencies of dinucleotides and trinucleotides of the light strand (2) of satellite DNA. Figs. 4 and 5 display the difference histograms of dinucleotides and trinucleotides as obtained from the satellite and as calculated from its base composition. The most remarkable features are on one hand the abundance of GA, AGA and GAG, and, to a lesser extent of AG, TC, TCC, TCG and TCT, and on the other the shortage of TA and of the trinucleotides containing TA.

A computer analysis of the repeated sequences in satellite DNA is presented in Table 2. The numbers and frequencies of all repeated sequences ($\geq$ 6 nucleotides) are given as well as those of repeated sequences identical

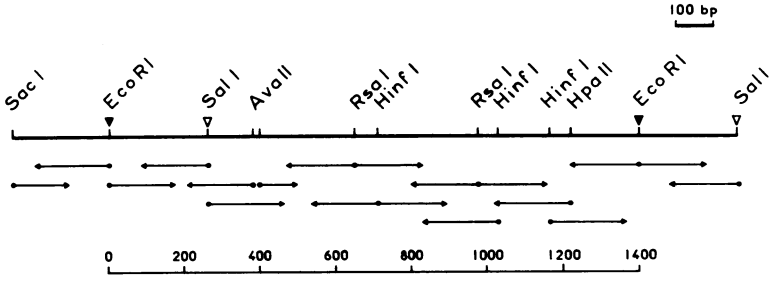

Fig. 2. Sequence strategy used in determining the primary structure of cloned satellite 1.715. Arrows pointing to the right or to the left indicate the starting points and the extent of the sequences determined on the light or on the heavy strand, respectively.

```
   1 GAATTCCCCG TCGTAACTCG AGAATCCCGC CGTAACTCGA GAAAAACCAC GTGGCTCCCC CGTCATCGCA AGATGAAGCC CTTTCCCGCT ACTGCGCCTC
     CTTAAGGGGC AGCATTGAGC TCTTAGGGCG GCATTGAGCT CTTTTTGGTG CACCGAGGGG GCAGTAGCGT TCTACTTCGG GAAAGGGCGA TGACGCGGAG

 101 AGGAGAAGTC CCACGTTAGG AATTGGAGGT CGAAAGGGCC CTTGGCACCC TTGATGCGAC CCACAAAGTT CCCCGAAATC CCGGTCTCCC TCGAGAGGAA
     TCCTCTTCAG GGTGCAATCC TTAACCTCCA GCTTTCCCGG GAACCGTGGG AACTACGCTG GGTGTTTCAA GGGGCTTTAG GGCCAGAGGG AGCTCTCCTT

 201 CACTGAGGTT TTCCGGCACC CCCTCCTCTG AGCCCTTTCT CCCCTCCTGA TCTGGACAGG AGGGTCGACT CCCCTGCTTT GTCTGGAAGG GGTTCCCGAT
     GTGACTCCAA AAGGCCGTGG GGGAGGAGAC TCGGGAAAGA GGGGAGGACT AGACCTGTCC TCCCAGCTGA GGGGACGAAA CAGACCTTCC CCAAGGGCTA

 301 CCTTCCGGTC GCACCTCAGG ATGAGGCCGG TCTCACGAAG ACATTCCAGA CGTGGCCTCG TGGGTGGTTC CACATTCCGT AGGACCCCGA TTTCCCGGTC
     GGAAGGCCAG CGTGGAGTCC TACTCCGGCC AGAGTGCTTC TGTAAGGTCT GCACCGGAGC ACCCACCAAG GTGTAAGGCA TCCTGGGGCT AAAGGGCCAG

 401 CCCTCTTGAT AAGAACCCGA TGCCCGGACA CCTCTCCGAA CTCCAGCCTG TGAATGAAGT CAACACGAAG GGGCAGTTTT TCCGTGCATC GTTCGGAAAA
     GGGAGAACTA TTCTTGGGCT ACGGGCCTGT GGAGAGGCTT GAGGTCGGAC ACTTACTTCA GTTGTGCTTC CCCGTCAAAA AGGCACGTAG CAAGCCTTTT

 501 AACCCCAGGT TCCAAATACA GCTCGACAAG CGGCCTCTCT CCCCGGGGAC ATCTCGAGAG GCAAGCGGAG TTCCATGCCT CAACCCAAGA CGAGGCCTGA
     TTGGGGTCCA AGGTTTATGT CGAGCTGTTC GCCGGAGAGA GGGGCCCCTG TAGAGCTCTC CGTTCGCCTC AAGGTACGGA GTTGGGTTCT GCTCCGGACT

 601 CTCTCCTGTC CCCAGTCTGC AGGGACCCTG CGATCGGAGT CTGAAATCAG AGGTACCCTG CGGTTCCTGC CTCAACTGGA GATGAGGCCC TCTTCCAATG
     GAGAGGACAG GGGTCAGACG TCCCTGGGAC GCTAGCCTCA GACTTTAGTC TCCATGGGAC GCCAAGGACG GAGTTGACCT CTACTCCGGG AGAAGGTTAC

 701 CACCAAGCCC AGTGGAGTCC CGAGAGGCCC CTCCCACCTC CAGTTTCCCT GGCTTCTCAG AGCCACCATG AGAAGCCCCC TGAGGTCACC TGCACAAGTC
     GTGGTTCGGG TCACCTCAGG GCTCTCCGGG GAGGGTGGAG GTCAAAGGGA CCGAAGAGTC TCGGTGGTAC TCTTCGGGGG ACTCCAGTGG ACGTGTTCAG

 801 GAGGGAACCC AGGGTTTCCT GCCTCAACCC GAGAAAGACC TCGAGAGACC TTCTTCAACA CGTCTCGAGG CCACATTCCC CTACCATGGC TCGGGAGCAA
     CTCCCTTGGG TCCCAAAGGA CGGAGTTGGG CTCTTTCTGG AGCTCTCTGG AAGAAGTTGT GCAGAGCTCC GGTGTAAGGG GATGGTACCG AGCCCTCGTT

 901 TGACGCGCTC CCCCTCGCCA CTCGCATGGA GACCCGACTT CCCTGGCGCC CCACGAGAGG CTCACTGACC TCGCCGTCGT ACCTCGTGAG AAACCGCACA
     ACTGCGCGAG GGGGAGCGGT GAGCGTACCT CTGGGCTGAA GGGACCGCGG GGTGCTCTCC GAGTGACTGG AGCGGCAGCA TGGAGCACTC TTTGGCGTGT

1001 CTGGGGCCGC CGCTCGAGAA CAACCCCGAG ATTCCCCCGT CATCGAGAGA TGAGGGCCTT CGTCTCCTGC GTGGCCTAGA GACCAATCTC GCGACCTCTC
     GACCCCGGCG GCGAGCTCTT GTTGGGGCTC TAAGGGGGCA GTAGCTCTCT ACTCCCGGAA GCAGAGGACG CACCGGATCT CTGGTTAGAG CGCTGGAGAG

1101 TCCAAACGCC TCAGGAGGCT TGACTCCCTT GAGTCCACCC AGTGAGCTCC AAGAGATACC CGTCGCGATT CGAGAGCAGA GCGGACGTTC TTTGCTTCCA
     AGGTTTGCGG AGTCCTCCGA ACTGAGGGAA CTCAGGTGGG TCACTCGAGG TTCTCTATGG GCAGCGCTAA GCTCTCGTCT CGCCTGCAAG AAACGAAGGT

1201 CTCGAGATGA ATGCCTGTCT CCCCGGGTGC GTCTGGAATG CAACCCCGAG ATCCCTGTCG CCCCTGGAGA GGAACACTGG CTTCTGGACA CAAGCCTAGA
     GAGCTCTACT TACGGACAGA GGGGCCCACG CAGACCTTAC GTTGGGGCTC TAGGGACAGC GGGGACCTCT CCTTGTGACC GAAGACCTGT GTTCGGATCT

1301 TGAGGTCTAT TGGCCCTGCA GTCACTCGAG AGCAATCCCC AGCTTTCCTT CGCAACTCGA ATGGAAGATT GGACTTGCCT GGGCCAACAC AAGAGGCAGC
     ACTCCAGATA ACCGGGACGT CAGTGAGCTC TCGTTAGGGG TCGAAAGGAA GCGTTGAGCT TACCTTCTAA CCTGAACGGA CCCGGTTGTG TTCTCCGTCG

1401 CT
     GA
```

Fig. 3. Primary structure of the EcoRI fragment of cloned satellite DNA (see Fig. 2). The upper strand is the light strand in CsCl (2).

TABLE I

Base composition, and frequencies of di- and tri-nucleotides of the light strand of 1.715 satellite DNA

	Base composition			
	A	C	G	T
	21.11	34.52	25.04	19.33
	Frequencies of dinucleotides			
	A	C	G	T
A	5.1	5.7	7.2	3.1
C	6.1	14.0	6.4	8.0
G	8.9	5.8	6.3	4.1
T	1.0	9.1	5.1	4.1
	Frequencies of trinucleotides			
	A	C	G	T
AA	1.07	1.50	1.50	1.00
AC	1.43	2.21	0.79	1.29
AG	2.79	1.21	2.21	1.00
AT	0.21	0.93	1.29	0.71
CA	1.79	2.07	1.50	0.79
CC	2.21	5.71	2.21	3.86
CG	2.43	1.36	1.07	1.50
CT	0.36	3.86	2.29	1.50
GA	2.00	1.71	3.93	1.29
GC	1.36	2.50	0.93	1.00
GG	2.00	1.79	1.29	1.21
GT	0.36	2.00	0.79	0.93
TA	0.21	0.43	0.29	0.07
TC	1.14	3.57	2.43	1.93
TG	1.64	1.43	1.71	0.36
TT	0.07	2.29	0.79	0.93

to parts of longer repeats. Table 2 also gives the fraction of satellite DNA sequences covered by the repeats and the amount of nucleotides present in all repeats neglecting the overlaps. A large number of inverted repeats was also found.

Table 3 presents the repeated sequences 8 nucleotides long or longer and their positions on the satellite DNA unit. The repeats are arranged in such a way as to show the sequence homology they share. The results in-

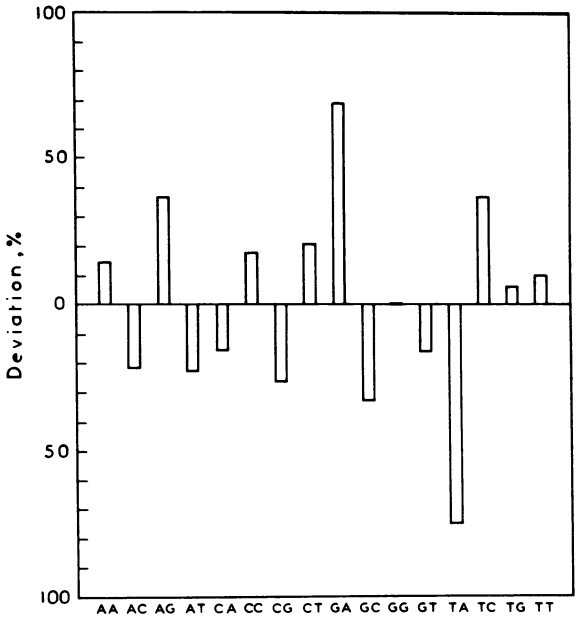

Fig. 4. Difference histogram between the frequency of dinucleotides as found in satellite DNA (light strand) and as calculated from its base composition (see Table I).

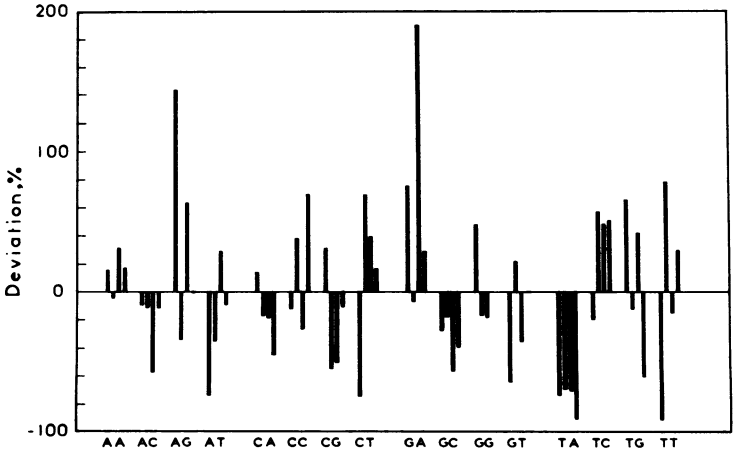

Fig. 5. Difference histogram between the frequency of trinucleotides as found in satellite DNA (light strand) and as calculated from its base composition (see Table I). Third positions in trinucleotide sets are in the order A,C,G,T.

TABLE II

Repeated sequences of 1.715 satellite DNA

Length of repeated sequences	13	12	11	10	9	8	7	6
Total number of repeated sequences (a)	4	1	2	1	11	18	54	131
Number of repeated sequences (b)	4	1	2	1	11	15 1	47 1	100 7
Number of sequences identical to part of longer repeated sequences (c)						2	5 1	20 4
Cumulative percentage of the satellite DNA covered by repeated sequences	7.4	9.1	11.6	13.0	24.3	37.2	63.9	92.4
Cumulative percentage of nucleotides present in all repeated sequences neglecting overlaps	7.4	9.1	12.3	13.7	27.8	47.8	99.8	206.3

The computer search was done for precise repeats (without mismatch and deletions/additions). Repeated sequences as long as or longer than 8 nucleotides are listed in Table III along with their positions in the 1402 bp repeat unit.

(a) No repeated sequences longer than 13 nucleotides were found.

(b) The first line concerns sequences repeated twice, the second one sequences repeated three times. No sequences were found to be repeated more than three times.

(c) The first line concerns sequences present only once, the second one sequences present twice. The other copies of these sequences are present in the longer repeated sequences. No sequences were found to be repeated more than two times neglecting the copies found in the longer repeats.

dicate that the octanucleotide CTCGAGAG is highly represented in the satellite DNA unit. This sequence is present (with at most a 2 nucleotide mismatch) 31 times in the satellite repeat unit. On the other hand, besides repeats made according to the pattern just described and comprising the first four groups of sequences of Table 3, other repeats were found which appear to have a different pattern characterized by high levels of C. As shown in Fig. 6, the repeat unit begins with a tandem of a 21 nucleotide sequence duplication (with a 15% mismatch) containing both types of sequences.

It should also be noted that two long open reading frames were found in the cloned unit of the 1.715 satellite, extending from position 321 to 641 and from positions 698 to 1120, respectively. Nevertheless, no transcription or translation signals were found in contrast with the case of the insertion of the 1.711a satellite (14).

Comparison between the restriction patterns of cloned and uncloned satellite DNAs. The restriction patterns for 16 enzymes of the cloned and sequenced repetition unit were compared with those of uncloned satellite. Fig. 7 shows the restriction patterns for enzymes which cleave at the same sites in

TABLE III

Repeated sequences (≥ 8) of the 1.715 satellite DNA and their positions (see fig. 3)

CGTAACTCGAGAA	12	31	
CTCGAGAA	36	1013	
CACTCGAGA	1199	1323	
ACTCGAGA	35	1200	
CCTCGAGAG	189	839	
CTCGAGAG	553	1325	840
CTCGAGAGG	190	553	
TCGAGAGA	841	1043	
TCGAGAGCA	1170	1326	
CGAGAGGC	555	721	954
GAGAGGAACACTG	193	1267	
GAGATGAGG	679	1047	
AGATGAGG	1048	1298	
GATGAGGCC	320	681	
CGCCTCAGGAG	95	1107	
CCTCAGGA	97	314	1109
GTCTGGAA	281	1231	
TCTGGACA	251	1283	
TCCCGGTC	179	393	
CCGGTCTC	181	327	
CCTCTCTCC	534	1095	
TCTCCCCGGG	538	1218	
GCTCCCCC	54	907	
TCCCCCGTCATCG	56	1033	
CCGTCGTA	8	974	
TTCCTGCCTCAAC	664	816	
TGCCTCAACCC	576	820	
TCAACACG	460	855	
TTCCCTGGC	745	939	
CTGGCTTCT	749	1277	
CGTGGCCT	351	1070	
CAACCCCGAGAT	1021	1241	
CCACATTCC	370	871	
AGCCCTTTC	77	231	
ACCCTGCG	625	655	
CCCCTCCT	220	241	

GAATTCCCCGTCGTAACTCGA

GAA TCCC GCCGTAACTCGA

GAA

Fig. 6. Sequence of the first 46 nucleotides of the cloned repeat unit of Fig. 3.

both cases. While for some enzymes (AvaII, EcoRII, HincII, MboII and RsaI) no difference was found, for other ones, some of the sites were not cleaved, because modified or absent in some of the units of uncloned satellite. For example, the AluI site at position 1341 and the KpnI and SacI sites were not cleaved in about 20% of the units; similarly, 10% of the PstI sites were resistant. The HhaI and SalI sites showed high resistance to cleavage probably due to the methylation of the CpG dinucleotide present in them. This could also be the case for a few Taq I sites giving rise to faint bands.

On the other hand, in the cloned material some restriction sites were not completely cleaved by some of the enzymes, although the sequencing data showed unambiguously that the sites were present. This was the case for the AluI, HincII and TaqI sites at positions 1341, 459 and 265, (see asterisks in Fig. 6). In the case of AluI, 80% of the cloned material showed cleavage at this site, while TaqI cleavage occurred in 45% of the sites. Cleavage of the HincII site at position 259 occurred in 60% of the cases.

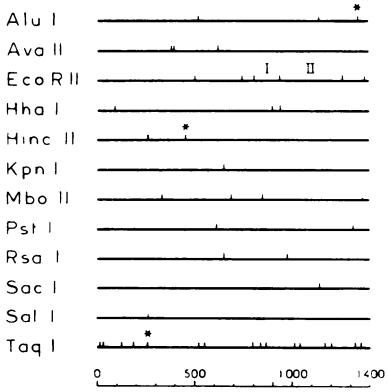

Fig. 7. Restriction maps of the EcoRI repeat unit of cloned and uncloned satellite. These maps concern enzymes whose sites have the same location in the two cases. Asterisks indicate sites showing a partial resistance in cloned satellite (see Text).

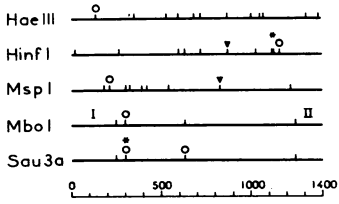

Fig. 8. Restriction maps of EcoRI repeat unit of cloned satellite. Circles and triangles indicate sites absent or only present, respectively, in the uncloned satellite. Asterisks indicate sites showing a partial resistance in cloned satellite (see Text).

Fig. 8 displays the restriction pattern for the enzymes which showed differences in the cleavage or the presence of some of their sites in cloned and uncloned satellite DNA. The following points deserve mention. (a) The HaeIII main band patterns are the same in cloned and uncloned DNA, except that the site at position 137 is not cut in uncloned DNA. Some very faint bands are present in the uncloned material, some of which are due to the resistance of some sites. (b) The HinfI restriction pattern shows two main differences. 1. The site at position 1167 is not cut in uncloned satellite, probably because of the methylation of the CpG dinucleotide formed by the site (GANTC) and the following G. 2. An extra site appears at position 874 in the uncloned material where the sequence is CATTC in the cloned unit; this site could have arisen by a transversion of the 5'C. It should also be noted that the 1122 site is resistant in about 40% of the cases in both cloned and uncloned satellite, despite of the existence of the HinfI site in the cloned unit. (c) An extra MspI site at position 828 is present in most of the uncloned repeat units. Some other faint bands are present in uncloned satellite deriving either from sites which do not exist in cloned satellite or from sites present in some units only. The isoschizomer HpaII cuts the cloned unit at exactly the same sites as MspI, whereas uncloned satellite DNA is not cleaved by HpaII revealing that these sites is methylated. (d) MboI does not cut the cloned unit due to methylation of the site although the sites are present at the positions indicated in Fig. 7. The only difference found is that the site at position 298 is not cut or is absent. Sau3 (an isoschizomer of MboI) does not cut uncloned satellite at positions 298 and 632. It is interesting to note that although the site at position 632 is present in uncloned units (as revealed by the MboI digestion) methylation does not allow the Sau3A enzyme to cut. The site at position 298 in

the cloned unit shows some resistance to degradation. (This corresponds to the sequence artefact in one of the strand mentioned in the Materials and Methods Section).

Finally, it should be pointed out that the restriction maps of Figs. 7 and 8 differ from previously published maps (10,15). A comparison of our results with the previous ones shows, however, that all differences can be easily explained. In the case of AvaII, sites at positions 382 and 397 were taken as a single site. Similarly EcoRII sites at positions 748 and 809 were taken as a single site; in addition the next two fragments were interchanged (fragments I and II of Fig. 7). In the case of MboI the extreme fragments were interchanged (fragments I and II of Fig. 8). Finally, a new AluI site at position 1341 was found.

DISCUSSION

The comparison between the restriction maps for the uncloned 1.715 satellite and its cloned repeat unit show that the sequence of the latter is representative of the native uncloned material. In fact, the restriction maps obtained with 16 enzymes showed that only 7 out of 84 restriction sites were different in the cloned satellite unit relative to the uncloned satellite DNA. If these changes were due to random point mutations, this would correspond to a sequence divergence of 2% between the cloned unit and the uncloned satellite. The actual sequence divergence, however, is certainly much smaller because most of the differences in the restriction maps are due to post-synthetic modifications (methylation) of nucleotides present in restriction sites. In particular, all of the HpaII sites and the Sau3A site at position 632 are methylated in the uncloned satellite DNA. Similarly, methylation at HhaI and SalI sites also must be frequent, the latter one accounting for the high dimer/monomer ratio (5:7) found for the cloned material. Interestingly, the sequenced regions flanking the EcoRI repeat (covering the first and last 260 nucleotides) were found to be identical to the corresponding sequences located within the EcoRI repeat. Our findings favor the view of a small degree of sequence divergence among the repeat units, where it is accompanied by a quantitatively higher level of differences in the methylation pattern.

The analysis of the sequence of Fig. 3 has revealed that, in contrast with the other bovine satellite 1.706, 1.711a and 1.720b (11-14), the 1.715 satellite is not organized in short tandem repeats. Its sequence is made up, however, of a very large number of repetitive elements, 92% of it being formed by repeats as long as, or longer than, 6 nucleotides (Table II). The data of

Table III suggest that these repeats are largely related to two basic sequences. The main one is represented by the octanucleotide CTCGAGAG, while the other one is characterized by high levels of C. The abundance of repeats related to these sequences is also shown by the large excess of AG, GA, TC, AGA, and GAG over statistical expectations. In addition, both types of basic sequences are represented in the first 46 nucleotides of the sequence of Fig. 3, which are a tandem duplication of a 21 nucleotide sequence. It is of interest that the sequence AGATGA, which is present 5 times in the 1.715 satellite is also found in the prototype sequence of the 1.720b satellite (13) and that the related sequence GAGATGA of the 1.715 satellite is present in the prototype sequence of the 1.706 satellite (13), because this establishes a homology link between the 1.706, 1.711a and 1.720b satellites, which are highly homologous with each other (11-14), on one hand, and the 1.715 satellite and its 1.711b relative, on the other. The homology link indicates, in turn, a common evolutionary origin for all the satellites just mentioned. The strong internal sequence divergence of the repeat units of the 1.715 satellite, which has led to the almost complete disappearance of recognizable shorter internal repeats, suggests that this satellite (as well as the related 1.711b) are evolutionarily older than the 1.706, 1.711a and 1.720b satellites. In this connection, it may be of interest that while the 1.715 satellite (and, in all likelihood, the 1.711b satellite) is strongly localized at the centromeres of all autosomes, the 1.706 satellite, though also confined to the centromere regions of autosomes, is absent from four of them (22). Since the localization of the 1.706 satellite is likely to concern also those of the 1.711a and 1.720b satellites (because of their very high sequence homology), this suggests that these satellites are only present on some of the acrocentric autosomes of the bovine caryotype, which are known to undergo centric fission and fusion. In turn, this may suggest that the latter, more recent satellites have arisen from the ancestral form of the older one(s) in amplification events occurring independently on different chromosomes.

ACKNOWLEDGEMENTS

We thank J.P. Dumas and J. Ninio from this Institute who provided us with the computer programs used to study the satellite sequence and P. Breton for the art work.

REFERENCES

1. Corneo, G., Ginelli, E., Soave, C. and Bernardi, G. (1968) Biochemistry,

7, 4373-4379.
2. Filipski, J., Thiery, J-P. and Bernardi, G. (1973) J. Mol. Biol. 80, 177-197.
3. Cortadas, J., Macaya, G. and Bernardi, G. (1977) Eur. J. Biochem. 76, 13-19.
4. Macaya, G., Cortadas, J. and Bernardi, G. (1978) Eur. J. Biochem. 84, 179-188.
5. Kopecka, H., Macaya, G., Cortadas, J., Thiery, J-P. and Bernardi, G. (1978) Eur. J. Biochem. 84, 189-195.
6. Botchan, M.R. (1974) Nature 251, 288-292.
7. Philippsen, P., Streeck, R.E. and Zachau, H.G. (1974) Eur. J. Biochem. 45, 479-488.
8. Mowbray, S.L., Gerbi, S. and Landy, A. (1975) Nature 253, 367-370.
9. Philippsen, P., Streeck, R.E., Zachau, H.G. and Miller, W. (1975) Eur. J. Biochem. 57, 55-68.
10. Roizes, G. (1976) Nuc. Acids Res. 3, 2677-2696.
11. Streeck, R.E. and Zachau, H.G. (1978) Eur. J. Biochem. 89, 267-279.
12. Pech, M., Streeck, R.E. and Zachau, H.G. (1979) Cell 18, 883-893.
13. Pöschl, E and Streeck, R.E. (1980) J. Mol. Biol. 143, 147-183.
14. Streeck, R.E. (1981) Science, in press.
15. Roizes, G., Pages, M. and Lecou C. (1980) Nuc. Acids Res. 8, 3779-3792.
16. Clewell, D.B. and Helinski, D.R. (1970) Biochemistry 9, 4428-4440.
17. Meunier-Rotival, M., Cortadas, J., Macaya, G. and Bernardi, G. (1979) Nucleic Acids Res. 6, 2109-2123.
18. Peacock, A.C. and Dingman, C.W. (1967) Biochemistry 6, 1818-1827.
19. Prunell, A., Kopecka, H., Strauss, F. and Bernardi, G. (1977) J. Mol. Biol. 110, 17-52.
20. Bolivar, F., Rodriguez, R.C., Greene, P.J., Betlach, M.C., Heyneker, H.L., Boyer, H.W., Crosa, J.H. and Falkow, S. (1977) Gene 2, 95-113.
21. Maxam, A. and Gilbert, W. (1977) Proc. Natl. Acad. Sci. USA 74, 560-564.
22. Kurnit, D.M., Shafit, B.R. and Maio, J.J. (1973) J. Mol. Biol. 81, 273-284.