ANNA SALUDA-GORGUL, JÓZEF JAWORSKI and JANUSZ GREGER *

NUCLEOTIDE SEQUENCE OF SATELLITE I AND II DNA FROM ALPACA \textit{(LAMA PACOS) GENOME}

\textit{Department of Biochemistry 1, Institute of Physiology and Biochemistry, Medical School of Łódź, Lindley’a 6; 90-131 Łódź, Poland}

\textit{Department of General Chemistry, Institute of Physiology and Biochemistry, Medical School of Łódź, Lindley’a, 6; 90-131 Łódź, Poland}

Received 20 December, 1989; Revised 30 March, 1990

Agarose and polyacrylamide gel electrophoresis of a total alpaca \textit{(Lama pacos)} DNA, digested with several restriction enzymes, revealed the presence of two tandemly organized repetitive DNA sequences, named Satellite I and Satellite II. Three Xhol-monomers from Satellite I DNA and two BspRI-monomers from Satellite II DNA were sequenced. As determined by dot hybridization analysis, the total alpaca DNA consists of 3.4\% and 1.3\% Satellites I and II, respectively. Computer search revealed no homology to any primate, rodent or mammalian sequences published in the Gen Bank Release 48.0 (February 1987) and the EMBL Bank Release 10 (December 1986).

A substantial portion of the mammalian genome is composed of highly and moderately repetitive DNA sequences, which can be divided into two classes: tandemly repeated sequences - termed satellites; and reiterated segments, whose copies are interspersed among other genomic DNA sequences. Tandemly repeated DNA can sometimes be isolated as satellite bands in density gradients.

Restriction enzyme analysis of satellite DNAs have shown that their repeated units are from a few base-pairs to hundreds base-pairs in length \cite{1,2}. Analysis of the organization of an eukaryotic genome is a major challenge facing molecular biology today, and determination of the structure and function of satellite DNAs occupies an important position in this field. A satisfactory model of the structural and functional organization of a genome must define the site of highly repetitive DNAs, and of satellite DNA in particular.

The structural peculiarities and functional role of satellite DNAs provide important insights into evolution of the eukaryotes.
Sequence comparisons between different families of repetitive DNA from the same species indicate that this sequence divergence occurs through random mutation and some unknown amplification mechanisms [1, 2].

There is no experimental evidence regarding the function of satellite DNA. However, the association with heterochromatin suggests that it may be important for chromosome structure [1, 2, 3]. Hybridization experiments in situ have shown that only the constitutive heterochromatin is hybridized with the satellite DNAs.

The aim of this work was to analyze the sequence organization of the basic repeated units (monomers) of the Alpaca Satellite I and II DNA and their presence in genomes of different animals.

MATERIALS AND METHODS

**Enzymes.** Restriction endonucleases EcoRI, PstI, BspRI were prepared according to published procedures [4]. Other restriction nucleases were from Pharmacia (Uppsala, Sweden). The T4 DNA ligase was obtained from the Department of Genetics, Warsaw University, Poland. DNA polymerase "Klenow fragment" and polynucleotide kinase were purchased from Pharmacia, Uppsala, Sweden; deoxyribonuclease I was obtained from Boehringer (Mannheim, F.R.G.).

**Restriction endonuclease digestions, agarose and polyacrylamide gel electrophoresis.** Conditions for digestion of DNA with restriction enzymes and agarose (Serva)- or polyacrylamide (Serva)-gel electrophoresis were reported previously [5, 6, 7]. Agarose gels 0.8% to 2% (w/v) and polyacrylamide gels 5-6% (w/v) made up in either Tris/acetate or Tris/borate buffers [8] were used.

**Preparation and fractionation of DNA.** High molecular weight (total) alpaca (Lama pacos) DNA was isolated from the leucocytes as described by Poncz et al. [9]. The alpaca DNA fraction enriched in Satellite II (BspRI) was prepared by centrifugation in the Cs$_2$SO$_4$/Ag$^+$ preparative density gradient as described by Filipski et al. [10].

Plasmid DNA was prepared according to procedures published elsewhere [6, 8].

**Isolation of restriction fragments and cloning.** The total alpaca DNA was digested with nuclease XhoI and separated on 1% (w/v) agarose gel. Three DNA bands of about 270 bp, 540 bp and 810 bp were isolated from this agarose gel by a freeze-thawing procedure [11, 12]. The mixture of these fragments of alpaca DNA was ligated to pBR322-vector cleaved with SalI. Transformation of *Escherichia coli* C-600 SF was performed according to the procedure described by Maniatis *et al.* [8].

Screening of recombinant plasmids was carried out as described by Birnboim & Doly [13]. Clones containing alpaca Satellite I sequences were designated pGK-1, pGK-2, pGK-7.
For the purpose of sequence analysis DNA of alpaca from recombinant plasmid pGK-1 was transferred to pBR322-vector cleaved with ClaI and the part of alpaca DNA from plasmid pGK-7 was transformed to pUC19-vector cleaved endonuclease Sall.

The DNA fraction enriched in alpaca Satellite II DNA was digested with BspRI and the digestion products were separated on 6% (w/v) polyacrylamide gel. The band of about 250 bp was isolated from the preparative gels and ligated to pUC19-vector cleaved with HincII [8]. The vectors were placed into *E. coli* JM 101 and screened for inserts by standard methods [8, 14]. Clones containing alpaca Satellite II sequences were designated pA12, pA19.

Radiochemicals, DNA sequencing and analysis. [γ-32P]ATP was prepared from [32P]orthophosphate (Amersham or Świerk), as described by Maxam & Gilbert [15], and [α-32P]dATP, [α-32P]dCTP, [α-32P]dGTP (1 000–3 000 Ci/mmmol) were prepared in this laboratory, according to Reeve & Huang’s protocol [16].

Determinations of nucleotide sequences were carried out according to the Maxam & Gilbert procedure [15].

5'-Recessed ends of restriction fragments were dephosphorylated with calf intestinal phosphatase (Boeringer) and labelled with [γ-32P]ATP and polynucleotide kinase. Fragments with 3'-recessed ends were labelled with [α-32P]dTTP and the “Klenow fragment” of *E. coli* DNA polymerase I [8]. The entire nucleotide sequences of all the investigated monomers of alpaca Satellite I and II DNA were determined in both strands.

Sequence analysis was performed on an Apple II computer using programs of Larson & Messing [17], and an Amstrad PC 1512, using software developed in this laboratory [18]. The probability that the observed homology between different regions of sequences occurs by chance was computed according to the double-matching probability of McLachlan [19] and was corrected for sequence length [20].

DNA labelling and blot-hybridization analysis. The DNA probes used for hybridization were labelled by a nick-translation method [8] or the suitable restriction inserts were labelled on the ends. After digestion with appropriate restriction nuclease and separation on 1.2–2% agarose gels, the DNA fragments were transferred onto nitrocellulose filters (BA85; Schleicher-Schuell) as described by Southern [21].

Filters with DNA were baked for 2 h at 80°C under vacuum or damp filters were UV-irradiated for 10 min at a distance of 12 cm [22]. The conditions for hybridization and washing of the filters were maintained according to published protocols [23]. After washing, the filters were exposed to Roentgen XM X-ray film (Foton XM).

Copy number determination. The frequency of the Satellite I DNA and the Satellite II DNA in alpaca genome was determined using a dot-blot technique [24]. Known amounts of the total alpaca DNA preparation, the recombinant
plasmid DNA (pGK1-25, or pA12) and also suitable vector DNA were fixed on cellulose filters. Each plasmid contained a single copy (monomer) of Satellite I DNA (or the Satellite II DNA). Hybridization with the labelled satellite monomer from plasmid pGK1-25 (or pA12) was carried out in one sealed plastic bag at 65°C for 24 h in a solution of 0.1% bovine serum albumin, 0.1% Ficoll, 0.1% poly(vinylpyrrolidone), 0.5% sodium dodecyl sulphate, 0.6 M NaCl, 0.06 M sodium citrate, 150 mg/ml denatured *E. coli* DNA as a carrier. The filters were washed in 0.15 M NaCl, 0.015 M sodium citrate and 0.1% sodium dodecyl sulphate at 52°C.

Comparison of the radioactivity bound within the linear range of response was used to determine the copy number of Satellite I or Satellite II in the alpaca genome.

![Electrophoretic patterns of the total alpaca DNA obtained on 6% polyacrylamide gel, after digestion with various restriction nucleases: lane 1, HindIII; lane 2, EcoRI; lane 3, BstI; lane 4, Aval; lane 5, XhoI; lane 7, BspRI; lane 8, PstI; lane 9, SacI; lane 6, complete BspRI digest of the pBR322 DNA as a DNA length marker](image)

Fig. 1. Electrophoretic patterns of the total alpaca DNA obtained on 6% polyacrylamide gel, after digestion with various restriction nucleases: lane 1, HindIII; lane 2, EcoRI; lane 3, BstI; lane 4, Aval; lane 5, XhoI; lane 7, BspRI; lane 8, PstI; lane 9, SacI; lane 6, complete BspRI digest of the pBR322 DNA as a DNA length marker.
RESULTS AND DISCUSSION

Isolation of the Alpaca Satellite I (XhoI) clones

The total alpaca DNA was cleaved with various restriction endonucleases and the digestion products were analyzed on 6% (w/v) polyacrylamide gels (Fig. 1). Digestion with three enzymes (XhoI, AvaI and SacI) resulted in the presence of a distinct ethidium bromide-stainable band, that migrated to a position in the gel corresponding to about 270 bp. In addition, distinct bands at position corresponding to about 540 bp and 810 bp were seen with these enzymes. This progression of bands in multiples of 270 bp immediately suggested that they originated from a tandem array and were undoubtedly satellite sequences.

The 270 bp XhoI-band and two closest neighbouring bands (540 bp and 810 bp) were excised from the agarose gel. The alpaca DNA from this gel was isolated and the mixture of these DNA bands was ligated into the vector

Fig. 2 A. Electrophoresis on 1.2% agarose-gel of the total alpaca DNA digested with various restriction enzyme. Lane 1, TaqI; lane 2, HpaII; lane 3, XhoI; lane 4, BstNI; lane 5, BspRI; lane 6, EcoRI. B. Southern-blot hybridization analysis of digestion products shown in A. Nick-translated plasmid pGK-7 was used as a labelled probe
pBR322. Next, C-600SF transformed cells, containing recombinant plasmids with repetitive inserts from alpaca DNA, were identified by restriction enzyme analysis (not shown).

Three plasmids containing repetitive fragments of alpaca DNA were isolated and designated pGK-1, pGK-2, pGK-7. 270 bp inserts (monomers) were contained in pGK-1 and pGK-2. The pGK-7 plasmid consisted of two monomers of about 270 bp and 810 bp fragment DNA, all ligated in vitro during the transformation. The nature of these monomer inserts was studied by digestion with several endonucleases and gel electrophoresis.

The 3' ends isolated TaqI-TaqI fragments, containing alpaca DNA, were labelled by "Klenow fragment" DNA polymerase I and used in turn to probe a Southern-blot, containing the total alpaca DNA digested with various restriction enzymes. The nick-translated pGK-7 plasmid was used as a probe to this filter with total alpaca DNA digested. The result was the same (Fig. 2 A, B).

All enzymes gave very similar hybridization patterns with the probes (labelled TaqI-fragments from pGK-1 and pGK-2, and nick-translated plasmid pGK-7), suggesting that these enzymes cleaved the alpaca DNA at different points in a tandem array, whose monomer length was about 270 bp.

This sequence of monomers was called the Alpaca Satellite I DNA.

**Sequence analysis of monomers of the Alpaca Satellite I DNA**

For the sequencing, TaqI-TaqI insert DNA of alpaca from plasmid DNA pGK-1 was recloned in pBR322-vector cleaved with endonuclease C1aI (plasmid pGK1-25), and the alpaca monomer XhoI-XhoI from plasmid DNA pGK-7 was recloned in pUC19-vector, cleaved with restriction enzyme SstI (pUGK-19).

All three monomers present in plasmid DNAs (pGK1-25, pGK2, pUGK-19) were sequenced by the Maxam & Gilbert [15] procedure. The

---

Fig. 3 I-III. I. Comparison of nucleotide sequences of several monomers of the Alpaca Satellite I DNA-(XhoI): a, 267 bp long DNA monomer inserted into pBR322-vector (plasmid pGK1-25); b, 266 bp long DNA monomer from plasmid pGK-2; c, 267 bp long DNA monomer derived from pGK-7 and cloned in pUC-19 vector; d, consensus sequence constructed on the basis of sequences a, b and c. The consensus represents the most frequent base at every residue. The dot at any position indicates the same nucleotide as in the monomer from pGK1-25. The letters indicate that the nucleotide differs from that in the monomer sequence from pGK1-25. Hyphens have been omitted for clarity. II. Arrows (A-G) indicate localization of regions of significant homology on the monomer of the Alpaca Satellite I DNA, shown in Fig. 3 I-III. III. Homology between several regions of the monomer sequence of the Alpaca Satellite I DNA (A-G). All sequences are matched to obtain maximum homology. Identical nucleotides at particular positions are marked with asterisks. Data concerning positions are shown above and under the compared sequences. On the right hand of the compared sequences, is shown the percentage of their homology [number of identical nucleotides (bp)/length of the region (bp)] × 100.
entire sequences of monomers of the Alpaca Satellite I DNA are shown in Fig. 3-I a, b, c. Two of the monomers were 267 bp long, and one - 266 bp long. All monomers of the Alpaca Satellite I DNA were about 57% G-C rich.

The mean value of sequence homology between the different monomers studied is equal to 93.7% and for consensus sequence - 94.6% (Fig. 3-I d). Only point mutations and 1-3 point deletions are observed along the monomers.

Sequence organization within the 267 bp repeated unit of the Alpaca Satellite I DNA is of interest because of the existence of palindromic sequence organization of high inner homology amounting to 63 bp (Fig. 3-II, Fig. 3-III A). The second part of monomer sequence is significantly rich in longer and shorter sequences mutually homologous (Fig. 3-II and Fig. 3-III). The copy number of the Alpaca Satellite I DNA in the alpaca genome was established to be equal to $7.6 \times 10^5$, on the assumption that the size of the alpaca genome is equal to $6 \times 10^9$ nucleotides (Fig. 4.A).

**Sequence monomers of the Alpaca Satellite II DNA**

A small alpaca DNA fraction “B” was prepared as a result of centrifugation of the total alpaca DNA in the Cs$_2$SO$_4$-Ag$^+$ preparative density gradient. The DNA of fraction “B” was digested with endonuclease BspRI and separated on 6% (w/v) polyacrylamide gels. A band of about 250 bp (Fig. 5) was isolated from preparative gels. In addition, the next bands were at positions corresponding to multiples of 250 bp. This picture of bands on the gel suggested that they originated from tandem array sequences, and were probably satellite DNAs. At the same time the DNA of fraction “B” was not digested with XhoI or AvaI enzymes (not shown).

The 250 bp BspRI-BspRI fragment of alpaca DNA was excised from the gel and a portion was cloned into the vector pUC-19. The second portion of this DNA was labelled by nick-translation and used to identify pUC-transformed JM 101 cells, containing repetitive DNA inserts. Four plasmid DNAs, containing these repetitive DNA fragments of alpaca, were isolated and two of them, namely pA12 and pA19, were sequenced by the Maxam & Gilbert [15] procedure. The complete sequences of both fragments of alpaca DNA are shown in Fig. 6-I.

The character of the alpaca DNA sequence in plasmids pA12 and pA19 was determined by a blot-hybridization experiment. The total alpaca DNA was digested with various restriction endonucleases, and the digestion products were separated on 2% (w/v) agarose gel and transferred onto nitrocellulose paper. The clone pA12 was used as a probe to a Southern-blot-hybridization. The result of this experiment (Fig. 7A, B) confirmed that fraction B of DNA belonged also to a satellite DNA of alpaca. This fraction was called the Alpaca Satellite II DNA.
Fig. 4. A. Estimation of copy number in the alpaca genome of monomers belonging to the Alpaca Satellite I DNA. The total alpaca DNA, pGK1-25 DNA and pBR322 DNA were hybridized in one vessel with labelled monomer DNA, isolated from pGK1-25 plasmid DNA. The sets of dots contain 12.5, 25, 50, 100 ng of total alpaca DNA (●) or represent the difference of radioactivity between the recombinant pBR-plasmid DNA (pGK1-25) and pBR322 DNA (■). Radioactivity in the dots was counted on scintillation counter. To compare the results from three independent experiments, all counts of a given experiment corresponding to different amounts of DNA [N (a)] were divided by the results obtained for the dot containing 100 ng of total alpaca DNA [N (100)]. The values N(a)/N(100) are plotted against amounts of DNA present in the dots. Assuming that 267 bp DNA (the monomer of Alpaca Satellite I DNA) comprises 5.77% of the recombinant pBR-plasmid we find the 1 μg of total alpaca DNA contains (a1/a2) × 0.0577 μg of monomers of the Alpaca Satellite I. This value corresponds to 7.6 × 10^5 copies per alpaca genome containing 6 × 10^9 nucleotides.

B. Estimation of copy number in the alpaca genome of the monomers belonging to the Alpaca Satellite II DNA. The total alpaca DNA, pA12 DNA and pUC19 DNA were hybridized in one vessel with labelled insert of the alpaca DNA from pA12, as described in Materials and Methods. The sets of dots contain 6.3, 12.5, 25, 50, 100 ng of the total alpaca DNA (●) or represent the difference of radioactivity between the recombinant pUC-plasmid DNA (pA12) and pUC19 DNA (▲). The applied methods were as in Fig. 4A. The values N(b)/N(100) are plotted against amounts of DNA present in the dots. Assuming that 249 bp DNA (the monomer of the Alpaca Satellite II DNA) comprises 8.48% of the recombinant pUC-plasmid (pA12), we find that 1 μg of total alpaca DNA contains (b1/b2) × 0.0848 μg of monomers of the Alpaca Satellite II. This value corresponds to 3.17 × 10^5 copies per alpaca genome containing 6 × 10^9 nucleotides.
Fig. 5. Electrophoretic pattern of DNA preparations obtained on 6% polyacrylamide gel after digestion with BspRI restriction nuclease. Lane 1, complete BspRI digest of pBR322 plasmid DNA as a DNA length marker; lane 2, fraction “B” of DNA obtained after centrifugation of the total alpaca DNA in the Cs$_2$SO$_4$/Ag$^+$ preparative density gradient.
General characteristic of the Alpaca Satellite II sequence

Both analyzed monomers of the Alpaca Satellite II DNA are 249 bp long and are 51.4% A-T rich. The sequences differed in 8 positions of bases only, and the homology between them was high: 96.7%.

Computer analysis of the sequence organization points to the presence of a few different homologous regions in the monomer of the Alpaca Satellite II DNA (Fig. 6-II, Fig. 6-III). Homology among individual regions is high and amounts from 65% for a long region (positions 88 to 133 — Fig. 6-III C) to 87% for a shorter region (positions 245 to 9 — Fig. 6-III E). All homologous regions of the Satellite II monomer are shown in Fig. 6-II and Fig. 6-III. The copy number of the Alpaca Satellite II DNA in the alpaca genome, determined similarly as for Satellite I, was established to be equal to $3.17 \times 10^5$ copies, on the assumption that the size of the alpaca genome is equal to $6 \times 10^9$ nucleotides (Fig. 4.B).

Several features of the *Lama pacos* repetitive elements described in this paper resemble those of the well defined group of satellite DNA sequences found in animal genomes.

The monomer sequences of two various satellite DNAs of the alpaca genome were isolated and characterized. These sequences were named the Alpaca Satellite I DNA and the Alpaca Satellite II DNA. These alpaca satellites varied in length, content of G-C pairs and with regard to organization of their monomer sequences. Computer analysis of these sequences did not show the presence of short simple repeats along the units studied, characteristic for many satellite DNAs [2, 25-29]. However, it was observed that shorter and longer regions were also homologous. They were located mainly in an Alpaca Satellite I DNA sequence (positions 71-226) which can exist as either a 155 bases long linear fragment or as a cruciform with two 63 base pair long hair-pins (63% homology — Fig. 3-II A).

The internal organization of the Alpaca Satellite II DNA was similar except that it contained three hair-pins differing in length (Fig. 6-II A). These types of internal organization in units of satellite DNAs were observed previously, for example in delphinids family genomes [30] and in Carpine Satellite I DNA [31].

Comparison of these Alpaca Satellite DNAs (I and II) with the suitable part (PRIMATE sequences, RODENT sequences, other MAMMALIAN sequences) of the Gen Bank Release 48.0 (February 1987) and the EMBL Bank Release 10 (December 1986) did not prove a homology with the previously known satellite DNAs of other animals.
Fig. 6 I-III. I. Nucleotide sequences of the 249 bp monomers of the Alpaca Satellite II DNA-(BspRI). a, 249 bp long DNA monomer derived from plasmid DNA pA12; b, 249 bp long DNA monomer derived from plasmid DNA pA19. The dot at any position indicates the same nucleotide as in monomer from pA12. The letter indicates that the nucleotide differs from that in the monomer sequence from pA12. Hyphens have been omitted for clarity. II. Arrows (A-E) indicate localization of the same regions of significant homology, which are shown in Fig. 6-III, along the 249 bp monomer of the Alpaca Satellite II DNA. III. Homology between several regions of the monomer sequence of the Alpaca Satellite II DNA (A-E). All sequences are matched to obtain maximum homology. Identical nucleotides at particular positions are marked with asterisks. Data concerning positions are shown above and under the compared sequences. On the right hand of the compared sequences is shown the percentage of their homology [number of identical nucleotides (bp)/length of the region (bp)] × 100
Fig. 7. Hybridization analysis of the total alpaca DNA. A. Electrophoresis on 2% agarose-gel of the *Lama pacos* DNA digested with various restriction endonucleases: lane 1, not digested the total alpaca DNA; lane 2, EcoRI; lane 3, BspRI; lane 4, BstI; lane 5, BstNI; lane 7, XhoI; lane 8, TaqI; lane 9, HpaII; lane 6, DNA length markers. B. DNAs shown in panel A were transferred onto nitrocellulose filters and used for hybridization experiments with nick-translation pA12 as a probe.
The authors thank Dr. J. Klysik for helpful suggestions and discussions of the experiments and Dr. A. Plucienniczak for his many valuable comments.

REFERENCES