Nucleotide sequence of a reiterated rat DNA fragment

Ramesh C. Gupta

Department of Pharmacology, Baylor College of Medicine, Houston, TX 77030, USA

Received 10 October 1983

A reiterated component of rat DNA was isolated by restriction with *Hin*dIII endonuclease and polyacrylamide gel electrophoresis. Sequence analysis revealed that the fragment was 179 nucleotides long. Unlike the known 370N reiterated rat DNA fragment which contained a high m^5C content (2.7 mole%), this repetitive element contained a rather low m^5C content (0.5 mole%). The present rat repetitive sequence appeared to be of α-type as shown by its significant homologies with α DNA sequenes of African green monkey and human. The possibility of sequence heterogeneity is discussed.

Rat DNA *Repetitive fragment* *α-Type sequence* *Sequence analysis*

1. INTRODUCTION

Highly reiterated sequences are found in nearly all eukaryotic genomes [1]. Reassociation kinetics has been used to determine the multiplicity of the repeated sequences that ranges from a few to millions. Some families of these are dispersed throughout the genome, while others are clustered in tandem repeats at the centromere and telomere regions of the chromosomes. Certain of the sequences (termed satellites) can be isolated from total genomic DNA by virtue of their unique buoyant density or by virtue of their rapid reannealing characteristics after denaturation of sheared total DNA. Some repetitive components have been isolated by digestion of total DNA with restriction endonucleases and gel electrophoresis. Little is known about the function of these sequences, although putative functions are many, including involvement in chromosome pairing, control of gene expression, processing of messenger RNA precursors and participation in DNA replication [2].

The rat genome has been shown to contain repetitive sequences which constitute up to 8–10% of its weight as determined by renaturation kinetics (e.g. [3]). Digestion of rat DNA with *Eco*R1 and *Hin*dIII restrictive endonucleases and gel electrophoresis has shown a series of sharp bands, indicating the presence of repetitive DNA components [4–6]. The authors in [7] reported the sequence of one such component (370N), which represents 1–3% of the genome weight (corresponding to $1.6–4.9 \times 10^5$ copies/haploid genome) [6,7]. We describe here the isolation and sequence analysis of another reiterated rat DNA component and discuss its unique features, and sequence homologies with other mammalian DNA fragments.

2. MATERIALS AND METHODS

2.1. *Materials*

*Hin*dIII restriction endonuclease and calf intestinal alkaline phosphatase (2500 units/mg) were from Boehringer-Mannheim. T_4 polynucleotide kinase was from P-L Biochemicals. [γ-^{32}P]ATP was synthesized in the laboratory as described [8]. Ultrapure acrylamide, bisacrylamide and silver stain kit were from Bio-Rad. All other chemicals were analytical grade.

2.2. *Isolation of DNA and the reiterated fragment*

DNA was isolated by a rapid procedure (to be published) adapted primarily from methods in [9,10]. Briefly, frozen liver from a male Sprague-Dawley rat was homogenized in sodium dodecyl

Published by Elsevier Science Publishers B.V.
00145793/83/$3.00

sulfate and the homogenate incubated with proteinase K. After 3 rapid solvent extractions, DNA was lumped out of the aqueous phase with the addition of ethanol. Residual RNA was removed by a brief RNase treatment. A typical yield was 2.2 mg/g of liver. DNA was restricted with *Hin*dIII endonuclease as in [11]. The digest (~3.5 mg) was electrophoresed on two 5% polyacrylamide slab gels (0.3 × 20 × 30 cm) and the fragments were detected by ethidium bromide staining and eluted from the gel [11].

2.3. *Terminal labeling, strand separation, and sequence analysis*

For 5′-^{32}P-labeling, restriction fragment (5 μg) was dephosphorylated by incubation at 38°C for 30 min with alkaline phosphatase (0.1 unit/μl) in 200 μl of 50 mM Tris–HCl (pH 8.0), and 0.1 mM EDTA. Phosphatase was removed first by incubation at 50°C for 45 min with 10 mM nitrilotriacetic acid, followed by deproteinization with phenol. Dephosphorylated fragment was recovered by ethanol precipitation and then 5′-labeled with T_4 polynucleotide kinase and [γ-^{32}P]ATP (~0.5 mCi) as in [11]. Strands were separated on a low cross-linked (50:1) 5% polyacrylamide gel and the single-stranded fragments were eluted [11]. The DNA sequence was determined by the base-specific chemical reactions (G, A + G, C, and C + T) as in [11] and thin (0.2 mm thickness) polyacrylamide sequencing gels. Location of m^5C in the sequence was determined by the appearance of a gap or weak band in C and C + T slots [12].

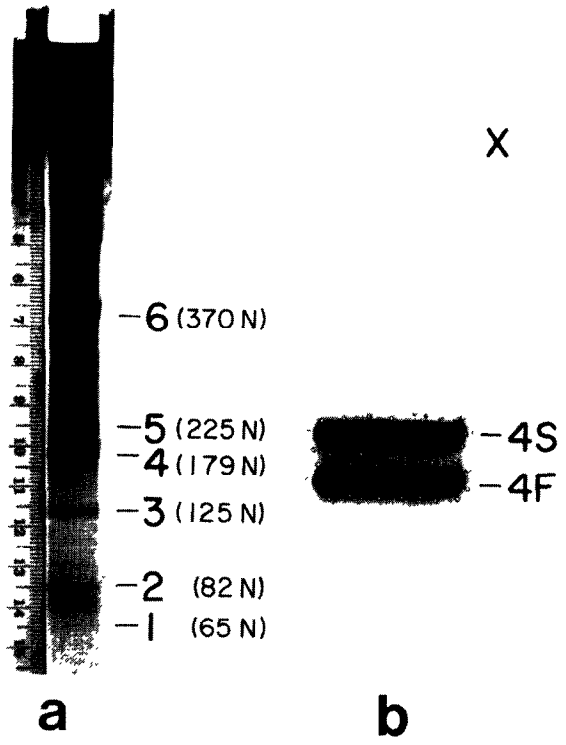

Fig.1. (a) Fractionation of *Hin*dIII restiction digest of rat DNA on a 5% polyacrylamide slab gel (0.2 × 20 × 30 cm). Fragments were detected by silver staining and the sizes determined by using gel markers or sequence analysis. (b) Strand separation of 5′-^{32}P-labeled fragment 4 on a low cross-linked (50:1) 5% polyacrylamide slab gel (0.2 × 20 × 30 cm). Detection by autoradiography. F and S denote the fast and slow migrating strands, respectively; X indicates a minor component.

3. RESULTS AND DISCUSSION

3.1. *Isolation and sequence analysis of the reiterated fragment*

Digestion of total rat DNA with *Hin*dIII restriction endonuclease showed two main bands (not shown) upon polyacrylamide gel electrophoresis and staining of the gel with ethidium bromide. However, after staining with silver stain, a more sensitive detection procedure ([13] also Bio-Rad bulletin 1089), at least 4 more bands were observed, as numbered fragments 1–6 in fig.1a. Gel markers, prepared by *Hin*fI digestion of ϕX174 DNA, were used to estimate the sizes of fragments 1–3 and 5, while the lengths of fragments 4 (present paper) and 6 [7] were determined by their sequence analysis. The latter two fragments refer to the two distinct bands detected after silver staining. Fragment 6 corresponds to the main *Hin*dIII fragment in Novikoff hepatoma DNA reported in [6] and sequenced by authors in [7] from rat liver DNA, and is reported to represent 1–3% of the genomic weight [6,7]. Based on ultraviolet absorption measurements, fragment 4 was found to represent 0.2–0.4% of the genomic weight, while fragment 6 represented 0.8–1% which is consistent with the lower value reported in [7]. The reiteration frequency of fragment 4 was therefore calculated to be 67 000–134 000 copies/haploid from the known fragment length (179N, see below), the fraction of the genome it repre-

sents, and the estimated genome size (6×10^9 base pairs/haploid genome) [14].

This fragment was isolated from ethidium bromide-stained gel as the integrity of the fragments recovered from a silver-stained gel has not yet been verified. For the sequence analysis, the fragment was 5′-^{32}P-labeled and electrophoresed on a strand-separation gel (fig.1b). A faint band (X, fig.1B) amounting to ~10% of the total radioactivity in fragment 4 was always found under these conditions, indicating that the duplex fragment was contaminated with a minor component. The entire sequence of both strands was determined unambiguously from the gel patterns displayed on 8 and 20% polyacrylamide–8 M urea gels (not shown). Variation in the distances observed on the sequencing gel between G8, C9 and A10 bands may reflect a localized structural effect. Attempts to overcome this by electrophoresis at higher gel temperature (55–60°C) were unsuccessful. Any possibility for the presence of a m^{5}C residue preceding C9 was ruled out, as only one G residue was found (which would face C9) in the opposite strand. The nucleotides above chain length 167 were identified by restriction of this fragment with *Hin*fI endonuclease, 5′-^{32}P-labeling, strand separation, and sequence analysis of the shorter fragment (not shown). Weak bands (no. 31) obtained in C and C + T slots suggested that C31 was largely methylated. Note that this was the only m^{5}C residue detected in the entire sequence. Unlike the 370N repeated fragment of rat DNA which contains 10 CG dinucleotides all carrying methylated C residues [7], the present reiterated fragment contained only one CG sequence with the C being methylated. Partial methylation was also detected in the opposite strand.

As shown by sequence analysis, there are 5 *Sau*3A sites, one *Hin*fI site, and one *Taq*I site, as indicated in the complete nucleotide sequence presented in fig.2. An interesting result was found when the end-labeled fragment was digested with these enzymes. The *Hin*fI and *Taq*I digests both showed two smaller fragments undigested. On the other hand, *Sau*3A showed 6 bands instead of the two bands expected; each band contained 10–20% of the total radioactivity, in addition to ~15% radioactivity remaining in the parent fragment. When *Hin*fI, *Taq*I, and *Sau*3A digestions were carried out in the presence of ϕX174, pBR322, and

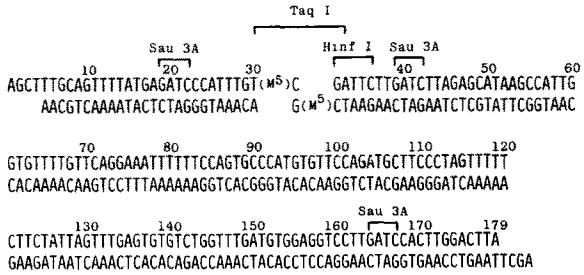

Fig.2. Complete nucleotide sequence of the 179 bp reiterated component of rat DNA. Restriction nuclease cleavage sites are indicated by brackets. Parenthesis denotes that C31 is not fully methylated.

λ DNAs, respectively, as internal standards, the carrier DNAs were found to be digested completely. The lack of complete digestion of the rat fragment may be due to the presence of the m^{5}C residue within the restriction site, as when m^{5}C is present adjacent to the *Hin*fI site this enzyme is inhibited [7]. However, this is unlikely for the *Taq*I site, because Tm^{5}CGA is reported to be cut with this enzyme [15]. There is no explanation for the incomplete digestion with *Sau*3A, as all the 3 sites are located far from the methylated residue. However, any possibility of 10–15% methylated C residue being present in the recognition site which would inhibit the cleavage [15] cannot be ruled out. These results suggest that this fragment is heterogeneous (to the amount of 10–20% or more at each position in the sequence). We are currently exploring this possibility in our laboratory by molecular cloning techniques. The sensitivity of sequencing techniques do not allow the determination of such levels of heterogeneity. It is important, in this regard, to note that significant sequence heterogeneity has been reported in the *Eco*RI monomers of the 370N rat DNA component [16].

The nucleotide sequence (fig.2) indicated that this repetitive element (40.2%) contains a somewhat higher G + C content than reported for the 370N repeat (38.1% [7]). An inspection of the sequence (fig.2) showed that about half of the total nucleotide sequence comprises interspersed direct and inverted repeats but most of these are tetra- and pentamers. The only significantly large internal repeat is CTTGATC (36–42, 161–167). No palindromic or regulating sequences are apparent. A comparison of the sequences of the present re-

petitive element and the type 2 Alu equivalent repetitive elements [17] showed no homology.

A comparison of the rat DNA sequence with α DNA sequence of African green monkey [18] and the related human DNA sequence [19], as described in fig. 3, shows (1) that the length of the rat repeat is 9 nucleotides longer than the monkey repeat, or 8 nucleotides longer than the human repeat, and (2) that 37% of the base sequence arrangement is the same between rat and monkey repeats, and 41% between rat and human repeats, and the homology is distributed randomly. Although the monkey and human repeats exhibit much larger homology (virtually same chain length and 65% same base order) [19], a significant homology found between rat and the other mammalian repeats suggests that the rat 179N fragment presumably represents the α-type sequence. There is an uneven distribution of A's and T's (~42%) in each strand of the rat primate sequence.

To determine if the 179N DNA fragment was repeated tandemly and/or interdispersed, rat DNA was restricted with *Hin*dIII endonuclease and the fragments were resolved on agarose gel, transferred to nitrocellulose paper, and hybridized to this nick-translated element as in [20]. The results of this experiment are presented in fig.4a,b. The discrete radioactive bands of approximate chain lengths 179, 358, 716, 1253, etc. (panel b) indicated the presence of multimers of the 179N component, but several intermediate multimers, i.e. tri, penta, hexa, etc. were either missing or present only in trace amounts. Furthermore, a significantly large amount of the radioactivity in the upper region of the gel was found in the limit *Hin*dIII digest. These results suggest that the 179N element is largely interdispersed in the related but diverged sequences, and only a small fraction may be present in short tandem repeats.

As will be reported elsewhere (Gupta, R.C., in

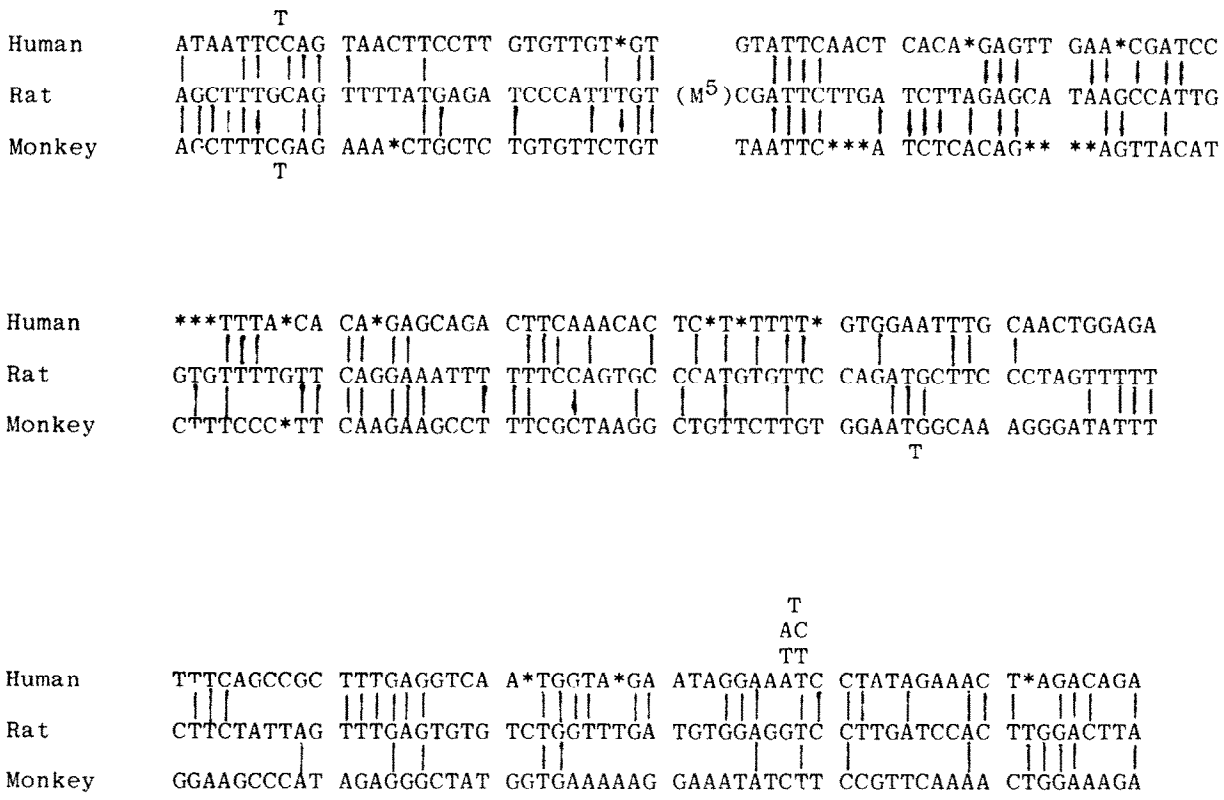

Fig.3. Comparison of the 179N rat DNA sequence with α DNA sequences of African green monkey [18] and human [19]. Certain nucleotides were bulged out in order to accomplish maximum homology. Asterisks denote empty spaces.

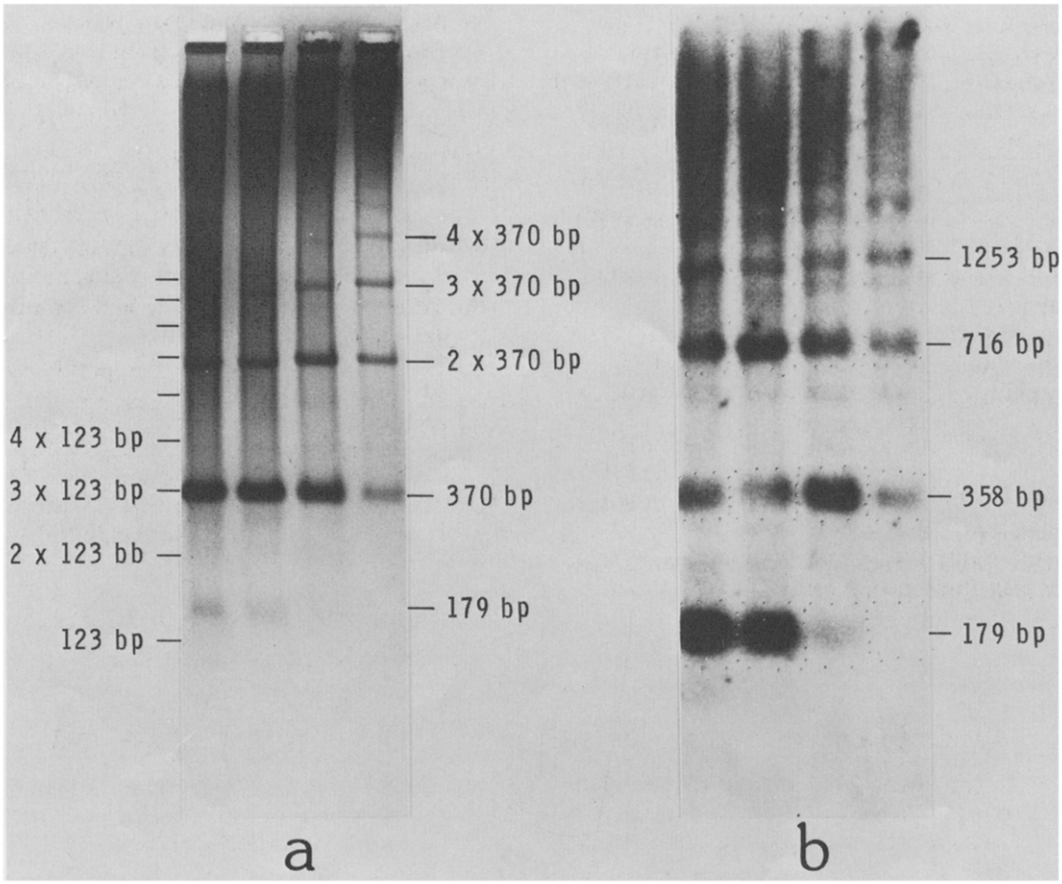

Fig.4. Hybridization of the nick-translated 179N repetitive element with *Hin*dIII restriction digests of rat DNA. The DNA digests (20 μg) were fractionated on a 2% agarose slab gel and the fragments were detected by ethidium bromide staining (a). The hybridization pattern was obtained after autoradiographic exposure at −80°C for 18 h (b). From left to right in each panel, the digestion times are 6, 15, 50 and 140 min. The positions of the size markers, 123 bp ladder (Bethesda Research Lab.), are indicated in panel a.

preparation), the various rat repeated fragments (see fig.1a), particularly the 179N repeat, were found to be the preferential binding sites in vivo for the carcinogens 2-acetylaminofluorene and 2-acetylaminophenanthrene. Further experiments are underway to determine whether this reflects lack of repair in the repeated sequences, as has been reported recently for 2-acetylaminofluorene lesions in α DNA sequences of African green monkey cells [21]. Such studies may contribute to the understanding of the organization of α type sequences, or repeated sequences in general, in the chromatin structure.

ACKNOWLEDGEMENTS

The author wishes to thank Drs L. Rothblum and K. Randerath for helpful discussions. The author also thanks G.K. Gambhir and N.R. Dighe for excellent technical assistance and P. Dodson for expert help with the preparation of the manuscript. This work was supported by USPHS grant CA 30606 and CA 13591.

REFERENCES

[1] Jelinek, W.R. and Schmid, C.W. (1982) Annu. Rev. Biochem. 51, 813–844.

[2] Lewin, R. (1982) Science 217, 621–623.
[3] Bonner, J., Garrard, W.T., Gattesfeld, J., Holmes, D.S., Sevall, J.S. and Wilkes, M. (1973) Cold Spring Harbor Symp. Quant. Biol. 38, 303–310.
[4] Philippsen, P., Streeck, R.E. and Zachau, H.G. (1974) Eur. J. Biochem. 45, 479–488.
[5] Maio, J.J., Brown, F.L. and Musich, P.R. (1977) J. Mol. Biol. 117, 637–655.
[6] Fuke, M. and Busch (1979) FEBS Lett. 99, 136–140.
[7] Pech, M., Igo-Kemens, T. and Zachau, H.G. (1979) Nucleic Acids Res. 7, 417–432.
[8] Gupta, R.C., Reddy, M.V. and Randerath, K. (1982) Carcinogenesis 3, 1081–1092.
[9] Kirby, K.S. (1957) Biochem. J. 66, 495–504.
[10] Marmur, J. (1961) J. Mol. Biol. 3, 208–218.
[11] Maxam, A.M. and Gilbert, W. (1980) Methods Enzymol. 65, 499–560.
[12] Ohmori, H., Tomizawa, J.-I. and Maxam, A.M. (1978) Nucl. Acids Res. 5, 1479–1485.
[13] Merril, C.R., Goldman, D., Sedman, S.A. and Ebert, M.H. (1981) Science 211, 1437–1438.
[14] Singer, D.S. (1979) J. Biol. Chem. 254, 5506–5514.
[15] McClelland, M. (1983) Nucl. Acids Res. 11, r169–r173.
[16] Sealy, L., Hartley, J., Donelson, J., Chalkley, R., Hutchison, N. and Hamkalo, B. (1981) J. Mol. Biol. 145, 291–318.
[17] Page, G.S., Smith, S. and Goodman, H.M. (1981) Nucleic. Acids Res. 9, 2087–2104.
[18] Rosenberg, H., Singer, M. and Rosenberg, M. (1978) Science 200, 394–402.
[19] Manuelidis, L. and Wu, J.C. (1978) Nature 276, 92–94.
[20] Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, New York.
[21] Zolan, M.E., Cortopassi, G.A., Smith C.A. and Hanawalt, P.C. (1982) Cell 28, 613–619.