A peculiar repetitive sequence in the rat genome

Hiroshi Nojima and Hirofumi Sokabe

Department of Pharmacology, Jichi Medical School, Minamikawachi-machi, Tochigi-ken 329-04, Japan

Received 14 July 1986

We report here a new type of peculiar repetitive sequence, $A_{15}T(TC)_9T_{12}$, which was detected at 750 base pairs (bp) upstream of a rat calmodulin processed pseudogene by DNA sequencing of cloned DNA fragments. This sequence element could possibly form a cruciform structure with a 12-AT-pair stem, exposing $(CT)_9$ sequences as a loop. S_1 nuclease protection experiments failed to identify this element as a cruciform structure but instead detected an alternating purine pyrimidine tract at 50 bp downstream of this element.

Total genomic Southern blotting showed that the rat genome contains only a few of these elements.

Calmodulin gene *Cruciform structure* *Repetitive sequence* *Nuclease S_1* *Southern blotting*

1. INTRODUCTION

It has been suggested from renaturation experiments that a large proportion of most eukaryotic genome is comprised of repetitive DNA sequences [1]. Repetitive DNA sequences are classified into two broad categories such as satellite DNA and interspersed repetitive sequences [2]. Interspersed repetitive DNAs with low sequence complexity such as $(AT)_n$ or $(TG)_n$ have been reported in some eukaryotes [3]. These kinds of alternating purine-pyrimidine sequences are suggested to change the DNA helical conformation from the right-handed B to the left-handed Z form under certain conditions [4–7] and that such conformational change may offer a landmark for some DNA-binding proteins regulating gene expressions [8]. These DNA elements were shown to be widely distributed and conserved throughout eukaryotic evolution by hybridization experiments [9–11].

Here, we report an unusual repetitive sequence element observed at the 5′ upstream of a rat calmodulin pseudogene. This DNA element falls in the category of 'simple interspersed repetitive DNAs' but is not widely distributed in the rat genome.

2. MATERIALS AND METHODS

2.1. *Cloning and sequencing genomic DNA fragments*

The cloning and subcloning of rat calmodulin pseudogene (λSC9) from a rat genomic library have been described [13]. DNA sequences were determined by the dideoxy nucleotide chain-termination method [14,15].

2.2. *S_1 nuclease digestion*

The *Ava*I fragment of calmodulin pseudogene (λSC9) was ^{32}P-labeled at the 5′-end with [γ-^{32}P]ATP (4000 Ci/mmol, ICN) and T_4 polynucleotide kinase (Takara Shuzo, Japan), then digested with *Pst*I. The resultant approx. 270 bp fragment was digested with S_1 nuclease (PL Biochemicals) as described [16]. The digested samples were electrophoresed in 16% polyacrylamide gel in the presence of 7 M urea.

2.3. *Total genomic Southern blot*

High-M_r rat liver DNA (20 μg each) was digested by restriction enzymes, electrophoresed in a 0.7% agarose gel, and transferred to a nylon membrane filter (Biodyne; PALL) [17]. The filter was hybridized with a nick-translated [18] *Pst*I-*Ava*I

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/86/$3.50

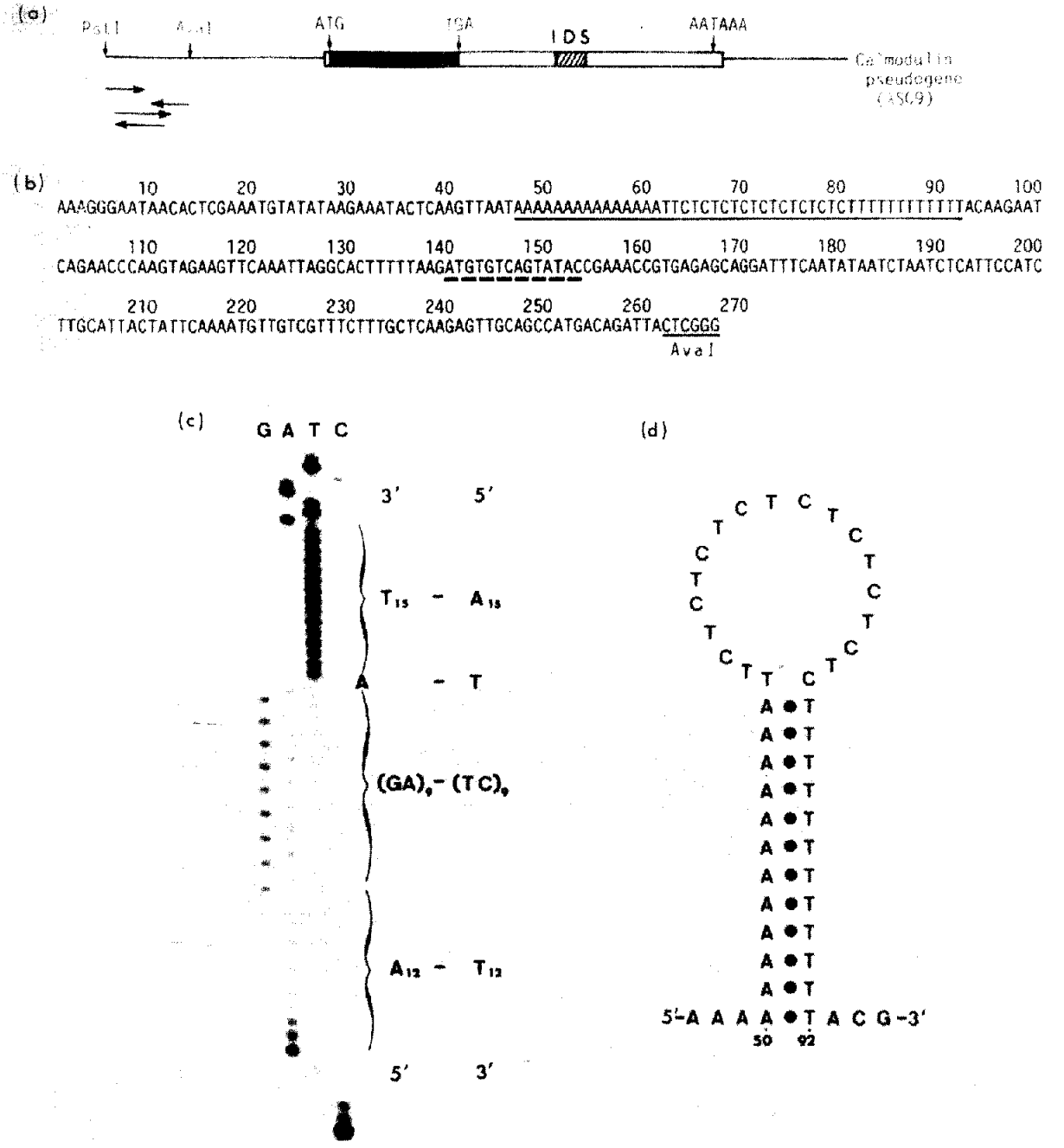

Fig.1. (a) A schematic illustration of the calmodulin pseudogene and the location of analyzed DNA fragments (*Pst*I-*Ava*I). Coding (■) and noncoding (□) regions, and the position of an inserted middle repetitive sequence identifier sequence (IDS) [22] are denoted [13]. Strategies for DNA sequencing are shown by horizontal arrows. (b) Primary structure of a repeated sequence (underlined) and the surrounding regions. A stretch of alternating purine pyrimidine sequences digested by S_1 nuclease (see fig.2) is shown by a broken underline. (c) An autoradiogram of 4% sequencing gel displaying the DNA sequences around the repeated sequence determined by the dideoxy chain-termination method [14,15]. (d) A possible cruciform structure formed around the repetitive sequence.

fragment (fig.1) of λSC9 [13]. The hybridization was conducted at 50°C for 40 h in 3 × SSC (1 × SSC: 0.15 M NaCl, 0.015 M sodium citrate) and 0.2% each of Ficoll, polyvinylpyrrolidone and bovine serum albumin. The filter was washed sequentially in 3 × SSC, 1 × SSC, 0.3 × SSC and 0.1 × SSC at 50°C for 15 min each in the presence of 0.1% SDS before exposure to X-ray films.

3. RESULTS AND DISCUSSION

The primary structure of the *Pst*I-*Ava*I fragment is given in fig.1. We observed an unusual repetitive sequence, $A_{15}T(CT)_9T_{12}$, in this region (underlined in fig.1b). It is located at about 750 bp upstream of a rat calmodulin processed pseudogene (λSC9) [13]. λSC9 is a typical processed pseudogene probably introduced into the genome through an mRNA-mediated process of insertion. One of the interesting features of this pseudogene is that it harbors multiple mutations in a small region of the genome such as insertions of one nucleotide (C) and two 17 bp direct repeats in the coding region and an integration of a middle repetitive sequence 'identifier sequence (IDS)' [19] in the 3′-noncoding region. The former two insertions cause frameshift mutations preventing it from encoding a functional protein. Although it bears a TATA box (TATAA), CAT box (CCAAT) and GC box (CCGCCC) in the 5′ upstream region, S_1 nuclease mapping suggested that λSC9 is not transcribed, at least in rat brain [13]. An autoradiogram of the sequencing gel around an unusual repetitive sequence is given in fig.1c. It possibly forms a cruciform structure with a stem of 12 AT pairs leaving a tract of $(CT)_9$ as a loop (fig.1d).

To examine whether this cruciform structure actually exists, a *Pst*I-*Ava*I fragment radiolabeled at the 5′-end of the *Ava*I site was digested with S_1 nuclease (fig.2). It is expected to display a band of 180 nucleotide (nt), however, it shows a band at 115 nt. This corresponds to an alternating purine-pyrimidine tract at 140 nt (broken line in fig.1b), ATGTGTCAGTATAC, which possibly forms an S_1-sensitive structure such as left-handed DNA of Z form [4–7]. Thus, the $A_{15}T(TC)_9T_{12}$ tract may not be in the cruciform structure at least under our experimental conditions. It is reported that a cruciform structure often displays anomalous electrophoretic migration [20]. The *Pst*I-*Ava*I fragment migrates at an expected size in both agarose (not shown) and polyacrylamide gel electrophoresis (fig.2), also indicating the absence of a cruciform structure.

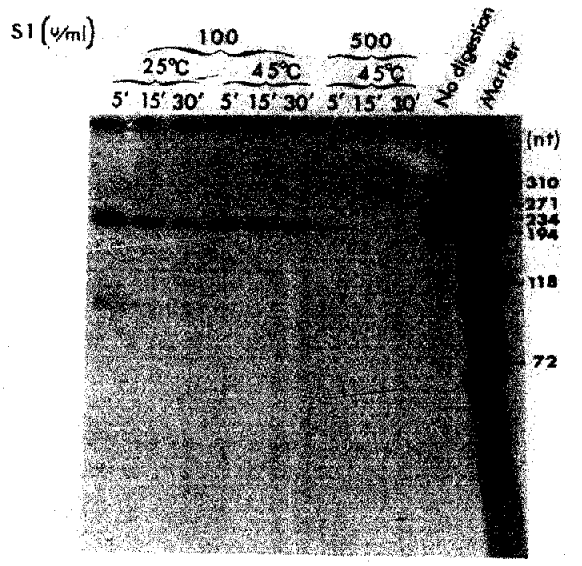

Fig.2. The DNA fragment radiolabeled at the 5′-end of the *Ava*I site was digested by S_1 nuclease (100 or 500 U/ml) at 25 or 45°C for 5, 15 and 30 min. The protected fragments were electrophoresed in 16% polyacrylamide gel in the presence of 7 M urea. Radiolabeled *Hae*III fragments of φX174 DNA were comigrated as size markers.

To determine the distribution of the $A_{15}T(TC)_9T_{12}$ tract in the rat genome, total genomic Southern blotting was performed with a nick-translated [18] *Pst*I-*Ava*I fragment as a probe. As shown in fig.3, only a few bands were observed for digestions of several restriction enzymes. The result indicates that this DNA element is not widely distributed in the rat genome.

Alternating polypurine or polypyrimidine tracts have been previously reported in other eukaryotic genomes such as $(CT)_{25}$ in sea urchin histone spacer [20] and $(CT)_{30}$ and $(GA)_{28}$ in mouse I_gC_μ [21]. A $(CT)_9$ tract sandwiched with poly(A) and poly(T) observed around a rat calmodulin pseudogene therefore is a new type of simple repetitive sequence.

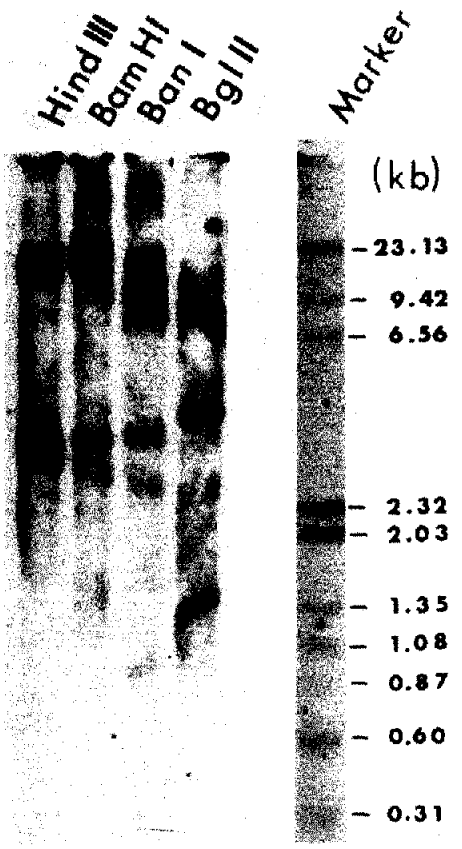

Fig.3. Number and distribution of DNA sequences homologous to the *Pst*I-*Ava*I fragment. See section 2 for details. Radiolabeled *Hin*dIII fragments of λ DNA and *Hae*III fragments of φX174 DNA were comigrated in the same gel to use as size markers and to assess the efficiency of transfer to a nylon membrane, which was the same for fragments of 2–20 kb.

REFERENCES

[1] Davidson, E.H., Hough, B.R., Amenson, C.S. and Britten, R.J. (1973) J. Mol. Biol. 77, 1–23.
[2] Jelinek, W.R. and Schmid, C.W. (1982) Annu. Rev. Biochem. 5, 813–844.
[3] Graves, D.R. and Patient, R.K. (1985) EMBO J. 4, 2617–2626.
[4] Wang, A.H.J., Quingley, G.J., Kolpak, F.J., Crawford, J.L., Van Boon, J.H., Van der Marel, G. and Rich, A. (1979) Nature 282, 680–686.
[5] Singleton, C.K., Klysik, J., Stirdivant, S.M. and Wells, R.D. (1982) Nature 299, 312–316.
[6] Nordheim, A. and Rich, A. (1983) Proc. Natl. Acad. Sci. USA 80, 1821–1825.
[7] Hanniford, D.B. and Pulleybank, D.E. (1983) Nature 302, 632–635.
[8] Nordheim, A. and Rich, A. (1983) Nature 303, 674–679.
[9] Miesfield, R., Krystal, M. and Aranheim, N. (1981) Nucleic Acids Res. 9, 5931–5947.
[10] Hamada, H., Pertino, M.G. and Kakunaga, T. (1982) Proc. Natl. Acad. Sci. USA 79, 6465–6469.
[11] Hamada, H. and Kakunaga, T. (1982) Nature 298, 396–398.
[12] Tauz, D. and Renz, M. (1984) J. Mol. Biol. 172, 229–235.
[13] Nojima, H. and Sokabe, H. (1986) J. Mol. Biol., 190, 391–400.
[14] Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463–5468.
[15] Messing, J. (1982) Methods Enzymol. 101, 20–78.
[16] Berk, A.J. and Sharp, P.A. (1977) Cell 12, 721–732.
[17] Southern, E.M. (1975) J. Mol. Biol. 98, 503–517.
[18] Rigby, P.W.J., Dieckman, M., Rhodes, C. and Berg, P. (1977) J. Mol. Biol. 113, 237–251.
[19] Gough, G.W. and Lilley, M.J. (1985) Nature 313, 154–156.
[20] Sures, I., Lowry, J. and Kedes, L.H. (1978) Cell 15, 1033–1044.
[21] Richards, J.E., Gilliam, A.C., Shen, A., Tucker, P.W. and Blattner, F.R. (1983) Nature 306, 483–487.
[22] Sutcliffe, J.G., Milner, R.J., Bloom, F.E. and Lerner, R. (1982) Proc. Natl. Acad. Sci. USA 79, 4942–4946.