Rat repetitive sequence: consensus sequence of *Taq* I-298 base pairs fragmentHideaki Kikuchi, Takao Sekiya¹, Susumu Nishimura² and Minro Watanabe

Department of Cancer Chemotherapy and Prevention, Research Institute for Tuberculosis and Cancer, Tohoku University, Sendai, Miyagi 980, ¹Oncogene Division and ²Biology Division, National Cancer Center Research Institute, Chuo-ku, Tokyo 104, Japan
Submitted September 8, 1987

Accession no.Y00450

The DNA of Buffalo rats was digested with the restriction endonuclease *Taq* I. The *Taq* I digested DNA was electrophoresed on an 8% polyacrylamide gel, and the 298 base pairs band was isolated. The recovered DNA fragment was ligated to plasmid pBR322 previously cleaved with *Cla* I. Out of 82 clones, four positive clones (ID-4, IIA-7, IIB-9, IIC-10), were sequenced.

The homology of sequence among them is about 97%. Homologous sequence of ID-4 was searched using the data base of Gen Bank. The sequences of high homology score were found in rat gamma casein gene intron 1 (91.7%) (1), rat (Sprague-Dawley) cytochrome C pseudogene (91.6%) (2) and rat repetitive sequence, clone 3B5 (89.6%) (3).

The results of hybridization in aqueous solution (data not shown) indicate at least 12,500 copies in the rat genome. The pattern of Southern hybridization (data not shown) suggests the presence of two classes of organization. One is tandem repetition so that 298 base pairs band appeared by the digestion with *Taq* I. Another organization of this sequence is an interspersion in rat genome, such as the cases described above (1,2). For each position in the figure, an asterisk (*) represents the same residue found in the ID-4 clone, and a dash (-) signifies a space inserted to allow the best possible match with the ID-4 sequence. The bottom line is the consensus sequence.

```

                                     50
ID-4      CGAGTGTAT TCCCTTTCCC GGTATCCGGG CAAACAATCC CCTTCCCCCT
IIA-7      ***** A**T***A** ***** **C*****C
IIB-9      *****T ***** *****--C-A A*CC*****
IIC-10     ***** **T**T*** *****_*** *TC*****

```

```

Consensus CGAGTGTAT TCCCTTTCCC GGTATTCCGGG CAAACAATCC CCCTCCCCCT

```

```

                                     100
CCCTTTCCTT ATGGGTGTTC CCCTCCCAAC CCTCCCCCA TTGCCGCCCT
*---***** ***** ***** *****
***** ***** *****A*****
***** ***** *****T*****

CCCTTTCCTT ATGGGTGTTC CCCTCCCAAC CCTCCCCCA TTGCCGCCCT

```

```

                                                    150
CCCCCACCA GTCTAGTTCA CTGGGGGTTT AGTCTTAGCA GGACCCAGGG
*****T_* *****
*****T** *****
*****A** *****

CCCCCATCA GTCTAGTTCA CTGGGGGTTT AGTCTTAGCA GGACCCAGGG

                                                    200
CTTCCCCTTC CACTGGTGCT CTTACTAGGA TATTCATGTC TACCTATGAG
*****
*****T *****
***** *****

CTTCCCCTTC CACTGGTGCT CTTACTAGGA TATTCATTGC TACCTATGAG

                                                    250
GTCAGAGTCC AGGGTCAGTC CATGTATATA TAGTCTTTGG GTAGTGGCTT
A-*****
*****
*****
*****C--- -*****A*

GTCAGAGTCC AGGGTCAGTC CATGTAT--- -AGTCTTTAG GTAGTGGCTT

                                                    298
AGTCCCTGGA AGCTCTGGTT GCTTGGCATT GTTGACATA TGGGGACT
*****
*****A* -*****
*****

AGTCCCTGGA AGCTCTGGTT GCTTGGCATT GTTGACATA TGGGGACT

```

REFERENCES

1. Yu-Lee, L. and Rosen, J.M. (1983) J.Biol.Chem. 258, 10794-10804.
2. Scarpulla, R.C. (1985) Nucl.Acids Res. 13, 763-775.
3. Witney, F.R. and Furano, A.V. (1984) J.Biol.Chem. 259, 10481-10429.