NUCLEOTIDE CLUSTERS IN DEOXYRIBONUCLEIC ACIDS. X. SEQUENCES OF THE PYRIMIDINE OLIGONUCLEOTIDES OF MOUSE L-CELL SATELLITE DNA.

Klaus Harbers, Barbara Harbers and John H. Spencer

Department of Biochemistry, McGill University
Montreal, Quebec, Canada H3G 1Y6.

Received April 24, 1974

SUMMARY

All the major pyrimidine oligonucleotides from the light and heavy strands of mouse L-cell satellite DNA have been sequenced. The sequences within each strand show a close homology with each other. In the heavy strand thymines predominate at the 5'-ends and cytosines at the 3'-ends and the longest pyrimidine tracts are the two octamers T-T-T-T-C-C-T-C and C-T-T-T-T-T-T-C. The hexamers T-T-T-T-C-C, T-T-T-C-T-C and T-T-T-T-T-C occur most frequently and their sequence homology with each other and the rest of the pyrimidine oligonucleotides of the heavy strand suggests that the basic repeating unit contained a pyrimidine hexamer. In the light strand sequence homologies and frequency of occurrence of C-T-T suggest the basic repeating unit also contained a pyrimidine trimer. These conclusions support the theory that mouse satellite DNA consists of a short repeating unit which has undergone several base changes during evolution.

INTRODUCTION

Mouse satellite DNA comprises about 10% of the total cellular DNA (1) and its high rate of renaturation has led to the suggestion that it is composed of a highly repetitive base sequence (2,3). Sequence analyses of satellite DNA's from guinea pig (4) and the kangaroo rat Dipodomys ordii (5) have resulted in descriptions of short basic base sequences and slight variants repeated many times per genome. As a first step in the sequence determination of the basic repeating unit of mouse satellite DNA we recently published the complete pyrimidine catalogue of this DNA (6). In this communication we report the base sequence of all the major pyrimidine oligonucleotides present in mouse satellite DNA.

MATERIALS AND METHODS

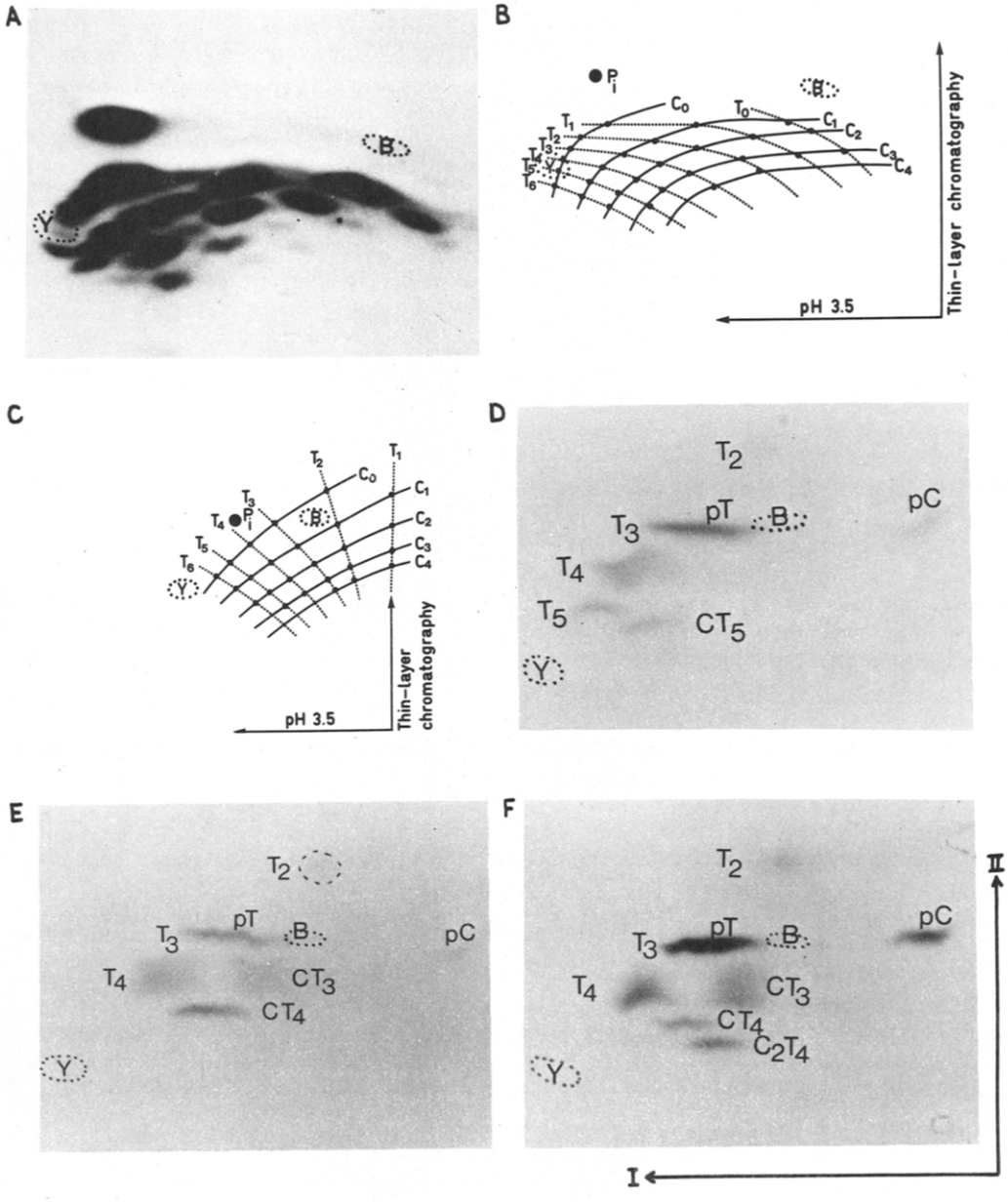
[^{32}P] labeled mouse L-cell DNA was isolated, purified, the satellite DNA isolated and the light and heavy strands of the satellite DNA separated as previously described (6).

The DNA had a specific activity of 5×10^4 cts/min/μg (Cerenkov radiation). The depurination products (7) were fractionated by the ionophoresis-homochromatography thin-layer system of Brownlee and Sanger (8) as modified for pyrimidine oligonucleotides by Ling (9). The separated oligonucleotides were visualized by radioautography and eluted from the thin-layer plate with triethylammonium bicarbonate, pH 8.0.

Spleen phosphodiesterase (EC.3.1.4.1.) snake venom phosphodiesterase (EC.3.1.4.1.) and bacterial alkaline phosphomonoesterase (EC.3.1.3.1.) were obtained from Worthington Biochemical Corp. (Freehold, N.J., U.S.A.). Dephosphorylation of the pyrimidine oligonucleotides and partial digestion either with spleen or snake venom phosphodiesterase were performed as described by Ling (9). Fractionation of the partial digestion products was by ionophoresis-homochromatography as described above (8,9).

RESULTS

Figure 1A shows a radioautogram of the [^{32}P]-labeled pyrimidine oligonucleotides present in a diphenylamine-formic acid hydrolysate of satellite DNA separated by ionophoresis-homochromatography. Studies by Ling (9) and in our laboratory (10) have shown that the pyrimidine oligonucleotides separate according to base composition in the electrophoretic step, and according to chain length in the chromatography step, thus the relative position of the separated oligonucleotides to each other is composition dependent. This allows the accurate prediction of the base composition of a pyrimidine oligonucleotide by its position on the chromatogram. A base composition grid derived from the pyrimidine oligonucleotide separation pattern of Figure 1 A is shown in Figure 1 B. These results confirm the unusual pyrimidine oligonucleotide distribution of mouse L-cell satellite DNA (6). In the present study the pyrimidine oligonucleotides were dephosphorylated prior to separation and sequence analysis to provide the non phosphorylated 5' and 3' termini required by spleen and snake venom phosphodiesterases respectively. A base composition grid derived from the separation pattern of the dephosphorylated pyrimidine oligonucleotides is shown in Figure 1C.


A
B
Y
B
P i
Co
To
C1
C2
C3
C4
T1
T2
T3
T4
T5
T6
Y
pH 3.5
Thin-layer chromatography
C
D
T2
pT
pC
T3
T4
T5
CT5
E
F
CT3
CT4
C2T4
II
I

In a previous study (6) we showed that the majority of pyrimidine pentanucleotides and longer occur in the heavy strand of the satellite DNA and this observation allowed their isolation directly from double-stranded DNA without the necessity for strand separation. Di-, tri- and tetranucleotides were isolated from separated strand preparations.

Sequence analysis of oligonucleotide CT_5 by partial snake venom phosphodiesterase digestion is shown in Fig. 1D. The separated oligonucleotides shown in the radioautogram result from the progressive removal of nucleotides from the 3' end of CT_5. The base compositions of the oligonucleotides were assigned from the grid shown in Figure 1C. The first base removed was C, $CT_5 \rightarrow T_5$ then T, $T_5 \rightarrow T_4$ etc. giving the sequence T-T-T-T-T-C. If more than one sequence is present in an oligonucleotide of a given base composition, partial digestion with an exonuclease will result in a pattern of spots on the thin-layer plate which contains at least one "branching point". An example of this is shown in Figure 1E, the result of partial digestion of oligonucleotide CT_4 with snake venom phosphodiesterase. Removal of the 3' terminal residues from CT_4 yields two products T_4 and CT_3 which results in a "branch point" on the radioautogram. The total pattern of oligonucleotide spots shows T_3 the only trinucleotide

---

Figure 1 - Two-dimensional fractionation of [^{32}P]-labeled pyrimidine oligonucleotides of mouse L-cell satellite DNA.

Dimension I: electrophoresis on a cellulose acetate strip in pyridine-acetate buffer, pH 3.5;

Dimension II: homochromatography on DEAE-cellulose thin-layer plates at 60°C, eluent 3% partially hydrolyzed yeast RNA containing 7M urea.

The separated oligonucleotides were visualized by radioautography. Y = position of yellow dye marker, B = position of blue dye marker.

(A) Depurination products of [^{32}P]-labeled satellite DNA.
(B) Grid pattern of 3', 5' phosphorylated pyrimidine oligonucleotides. The diagram is based on the oligonucleotide pattern of Figure 1A. The position of each oligonucleotide is represented by a dot (·). Solid lines join oligonucleotides of similar cytosine content, and broken lines of similar thymine content.
(C) Grid pattern of terminally dephosphorylated pyrimidine oligonucleotides.
(D) Products resulting from partial digestion by snake venom phosphodiesterase of hexanucleotide CT_5.
(E) Products resulting from partial digestion by snake venom phosphodiesterase of pentanucleotide CT_4.
(F) Products resulting from partial digestion by snake venom phosphodiesterase of hexanucleotide C_2T_4.

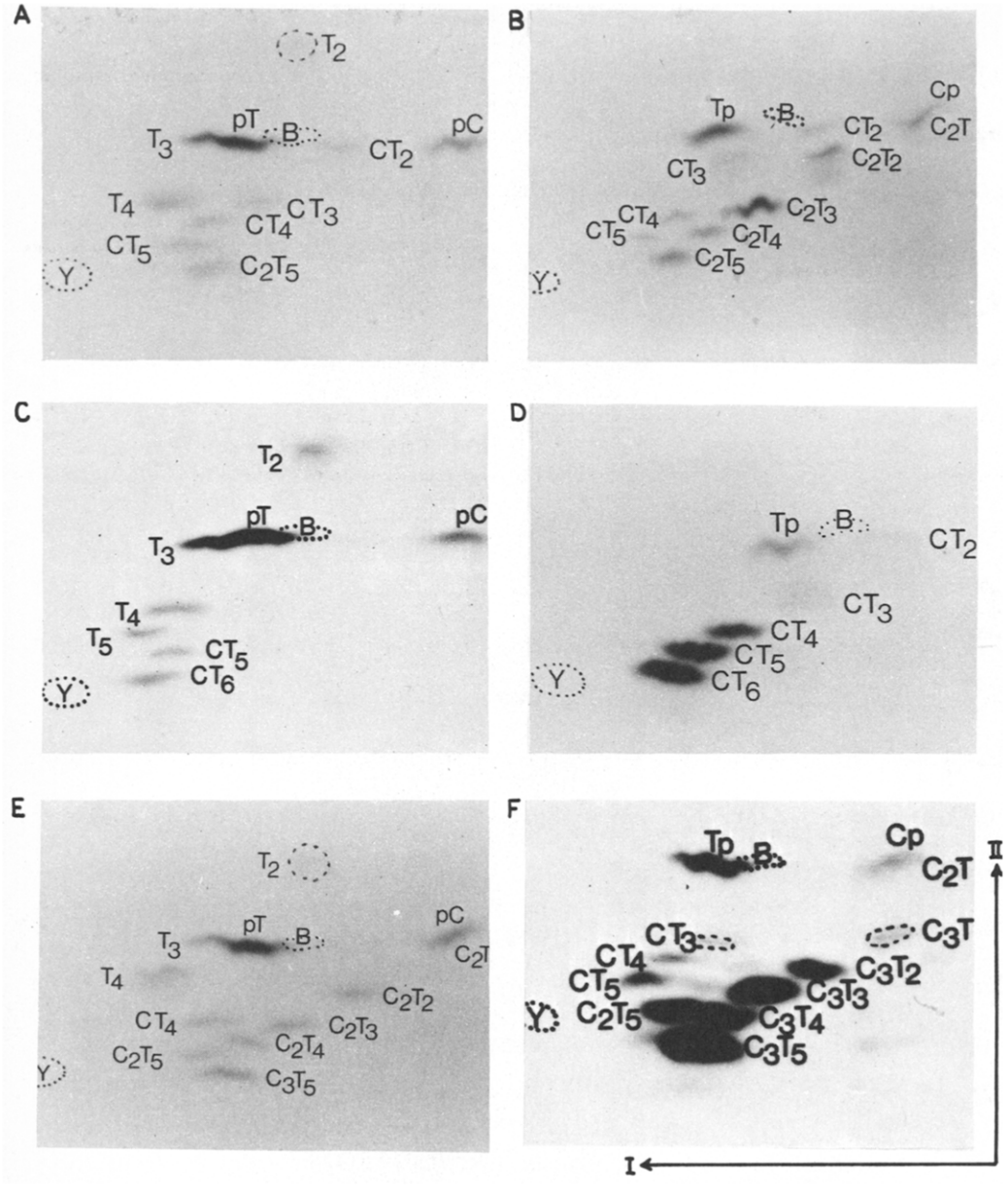


Figure 2 - Two-dimensional fractionation of the products resulting from partial enzyme digestion of [^{32}P]-labeled pyrimidine oligonucleotides from mouse L-cell satellite DNA.

For experimental details see legend to Figure 1.

(A) Partial digestion of heptanucleotide C_2T_5 with snake venom phosphodiesterase.
(B) Partial digestion of heptanucleotide C_2T_5 with spleen phosphodiesterase.
(C) Partial digestion of heptanucleotide CT_6 with snake venom phosphodiesterase.
(D) Partial digestion of heptanucleotide CT_6 with spleen phosphodiesterase.
(E) Partial digestion of octanucleotide C_3T_5 with snake venom phosphodiesterase.
(F) Partial digestion of octanucleotide C_3T_5 with spleen phosphodiesterase.

present and indicates that CT_4 is composed of the two sequences T-T-T-T-C and T-T-T-C-T. Note that pT and TpTpT chromatograph adjacent to one another. Another example of a "branch point" was with oligonucleotide C_2T_4 which comprises two sequences T-T-T-T-C-C and T-T-T-C-T-C (Figure 1F), which occur in about equal amounts. Hepta- and octanucleotides were digested with snake venom and spleen phosphodiesterase in separate experiments to ensure accurate determination of base sequences. As examples of this procedure the radioautograms in Figures 2A and 2B show the heptamer C_2T_5 is a mixture of two sequences, T-T-T-T-C-T-C and C-T-T-T-T-T-C. The other major heptanucleotide, CT_6 has the sequence T-T-T-T-T-C-T (Figures 2C and 2D). The longest oligonucleotide sequenced was the octamer C_3T_5 which consists of two sequences T-T-T-T-C-C-T-C and C-C-T-T-T-T-T-C (Figures 2E and 2F).

Sequences of the other major oligonucleotides that occur in mouse satellite DNA which have been determined by these procedures are given in Table 1. The three different sequences given for tetranucleotide C_2T_2 are not final because of difficulties in interpretation encountered with partial digests of multiple isomeric oligonucleotides. The number of tracts assigned for each sequence in Table 1, is based on a calculation described previously (6) and on the quantitative determination of each oligonucleotide at the "branching point" in partial digests of isomeric oligonucleotides. The calculation of tract numbers allows an easy comparison of the relative frequencies of the different pyrimidine sequences in a satellite DNA.

DISCUSSION

The pyrimidine oligonucleotide sequences described in this communication comprise, together with the mono- and dinucleotide isostichs approximately 85% of the total pyrimidine oligonucleotides of mouse satellite DNA. Sequences of some of the shorter pyrimidine oligonucleotides have been listed previously in a review article (11) without any supporting data. Our results confirm those sequences and in addition provide the sequences of the other penta-, hexa- and octanucleotides.

TABLE 1

Sequences of The Major Pyrimidine Oligonucleotides in Mouse L-Cell Satellite DNA

Component	Heavy Strand	Number of tracts/DNA molecule* (DNA Mol.Wt. 5×10^6 daltons: 15,000 bases)
C_2T_2	T-T-C-C T-C-T-C T-C-C-T	230
CT_3	T-T-T-C T-T-C-T	100 45
CT_4	T-T-T-T-C T-T-T-C-T	85 85
C_2T_4	T-T-T-T-C-C T-T-T-C-T-C	145 145
CT_5	T-T-T-T-T-C	140
C_2T_5	T-T-T-T-C-T-C C-T-T-T-T-T-C	30 20
CT_6	T-T-T-T-T-C-T	50
C_3T_5	T-T-T-T-C-C-T-C C-C-T-T-T-T-T-C	40 40
	Light Strand	
C_2T	T-C-C C-C-T	80 40
CT_2	C-T-T	120
CT_3	C-T-T-T	60

* For details of the calculations of numbers of tracts see text and (6).

The sequences listed in Table 1 are related, having several obvious features in common. In the heavy strand, with two exceptions, thymines occur at the 5' termini. Cytosines predominate at the 3' end. The converse is true in the light strand. The three hexamers T-T-T-T-C-C, T-T-T-C-T-C and T-T-T-T-T-C, which occur in about equal amounts, are

the most common pyrimidine sequences in the heavy strand. C-T-T is the most common sequence in the light strand. Within each strand there is extensive sequence homology centered on the most common sequences. Thus a hexamer in the heavy strand and a trimer in the light strand may be a part of the basic repeating unit of mouse satellite DNA. The results support the view that mouse satellite DNA is composed of a short repetitive base sequence and a number of minor variants of this sequence which differ in only one or two base positions. Currently, we are investigating an enzymatic degradative method to provide oligonucleotides which overlap the pyrimidine sequences.

ACKNOWLEDGEMENTS

The authors thank Drs. A.F. Graham and S. Millward for providing mouse L-cells, and Mr. A. Kabassakalian for technical assistance. This work was supported by grant MT 1453 from the Medical Research Council of Canada. K. Harbers is the holder of a fellowship from Deutsche Forschungsgemeinschaft.

REFERENCES

1. Kit, S. (1962) Nature 193, 274-275.
2. Waring, M. and Britten, R. (1966) Science 154, 791-793.
3. Flamm, W.G., Bond, H.E., Burr, H.E., and Bond, S.B. (1966) Biochim. Biophys. Acta 123, 652-654.
4. Southern, E.M. (1970) Nature 227, 794-798.
5. Fry, K., Poon, R., Whitcome, P., Idriss, J., Salser, W., Mazrimas, J., and Hatch, F. (1973) Proc. Nat. Acad. Sci. U.S.A. 70, 2642-2646.
6. Harbers, K., and Spencer, J.H. (1974) Biochemistry 13, 1094-1101.
7. Burton, K (1968) Methods in Enzymol. 12A, 222-224.
8. Brownlee, G.G.,and Sanger, F. (1969) Europ. J. Biochem. 11, 395-399.
9. Ling, V. (1972) J. Mol. Biol. 64, 87-102.
10. Harbers, B., Delaney, A.D., Harbers, K., and Spencer, J.H., manuscript in preparation.
11. Walker, P.M.B. (1971) Progr. Biophys. Mol. Biol. 23, 145-190.