INVERTED SEQUENCES IN RAT DNA

S. SZALA, J. MICHALSKA, H. PATERAK, B. BIENIEK and M. CHORĄZY
Department of Tumor Biology, Institute of Oncology, 44-100 Gliwice, Poland

Received 9 March 1977

1. Introduction

Eucaryotic DNA contains 'palindromic' sequences (or 'foldback' DNA) which are defined by their concentration-independent renaturation which is consistent with intramolecular base pairing of inverted sequences [1–9]. There are two kinds of inverted sequences:

(1) Inverted-repeated sequences separated by non-repeated sequences which form after denaturation and reassociation 'hairpins' – structures in which complementary sequences on a single DNA strand form a base-paired 'stem' – with wide distribution of length (averaging about 1000 base pairs) and terminated in 'single-stranded loops', ranging from 400 to thousands of base pairs in length;

(2) Inverted-repeated sequences not separated by other sequences, forming during denaturation and reassociation 'hairpins' with average length about 1100 base pairs without discernible loops.

In previous studies [10,11] we showed that reassociation of 'foldback' type of sequences in rat DNA is 'concentration-independent' to C_0t-value $<< 10^{-4}$. The amount of reassociated 'foldback' DNA is a function of DNA fragment size. We studied also the thermal stability as well S_1 resistance and transcription properties of foldback DNA using RNA polymerase from *Escherichia coli*. We found that rat foldback DNA contains long pyrimidines tracts above 80 nucleotides in length [11]. In this report are presented data on molecular structure of rat foldback DNA revealed by electron microscopy, and preliminary results on sequencing of polypyrimidine tracts isolated from foldback DNA.

2. Materials and methods

The experimental methods used for the isolation of rat foldback DNA and also details on denaturation, reassociation, hydroxyapatite chromatography, S_1 digestion and electron microscopy are described in the previous papers [10–12]. The long pyrimidine tracts were isolated from foldback DNA after treatment with formic acid–diphenylamine according to Straus and Birnboim [13]. Purified oligopyrimidine sequences were transcribed into complementary RNA (cRNA) with *E. coli* RNA-polymerase. The incubation mixture, total volume 32 μl, was composed of: 0.033 M Tris–HCl (pH 7.9), 0.09 M KCl, 6 mM 2-mercaptoethanol, 33 mM $MgCl_2$, 0.5 mM EDTA, 0.4 mM sodium pyrophosphate (pH 7.8) and 0.5 mg/ml bovine serum albumin (Sigma Chem. Co., St. Louis, Mo.), 5 μg RNA-polymerase (Boehringer, Manheim) isolated from *E. coli* MRE 600, 0.1 μg pyrimidine tracts, and 0.165 mM (final conc.) GTP and ATP (Sigma Chem. Co., St. Louis, Mo.). [α-^{32}P]GTP or [α-^{32}P]ATP, spec.act. 2–10 mCi/mmole (The Radiochemical Centre, Amersham) were added, usually at 0.1 mCi/sample. Incubation time was usually 2 h at 37°C. DNAase, 1 μg, (RNAase-free, Worthington, Freehold, NJ) was then added and incubation continued for 15 min. The mixture was diluted with 100 μl 0.5% SDS and extracted 2 times with phenol saturated with 1% SDS, 1 mM EDTA. Water phase was then applied at room temperature to a column of Sephadex G-100 (1 × 20 cm) equilibrated 0.01 M

 North-Holland Publishing Company – Amsterdam

Tris–HCl (pH 7.5), 0.1 M NaCl, 0.005 M EDTA and 0.5% SDS. The cRNA eluted in the void volume was precipitated with ethanol and purified from SDS by redissolving it in 2% sodium acetate (pH 5.0) and precipitated again with ethanol. Approximately 0.5–1 × 10^5 cts/min of polypurine cRNA was digested with RNAase T_1 (Sankyo, Tokyo) for 30 min at 37°C in 5 μl incubation mixture (10 mM Tris–HCl, pH 7.6 and 1 mM EDTA) containing 10 μg *E. coli* tRNA (enzyme to RNA ratio 1:20). Digestion products were dried and separated on two dimentional system: electrophoresis on acetate cellulose (Schleicher-Schüll) at pH 3.5 and thin-layer chromatography on cellulose impregnated with polyethyleneimine developed with a pyridine–formate buffer 1.5 M (pH 3.5). For other details see: Southern and Mitchell [14] and Southern [15], also Szala et al. [11].

3. Results and discussion

In electron micrographs of rat foldback DNA isolated to $C_0t << 10^{-4}$ from DNA with the average molecular weight of 32 000 base pairs, we identified mainly linear double-stranded molecules (90% of molecules) which do not have discernible loops and 10% of molecules containing single-stranded loop ends (fig.1). The number-average and weight-average loop-lengths, calculated from the distribution of the length of looped molecules (fig.2), are 0.41 μm and 1.68 μm (or 1230 and 5040 base pairs) respectively. The number-average and weight-average double-stranded 'stems' for looped molecules are 0.35 μm and 0.74 μm (or 1050 and 2220 base pairs) respectively. For 'unlooped' molecules the number- and weight-average lengths are equal to 0.69 μm and 0.85 μm (or 2070 and 2550 base pairs) respectively.

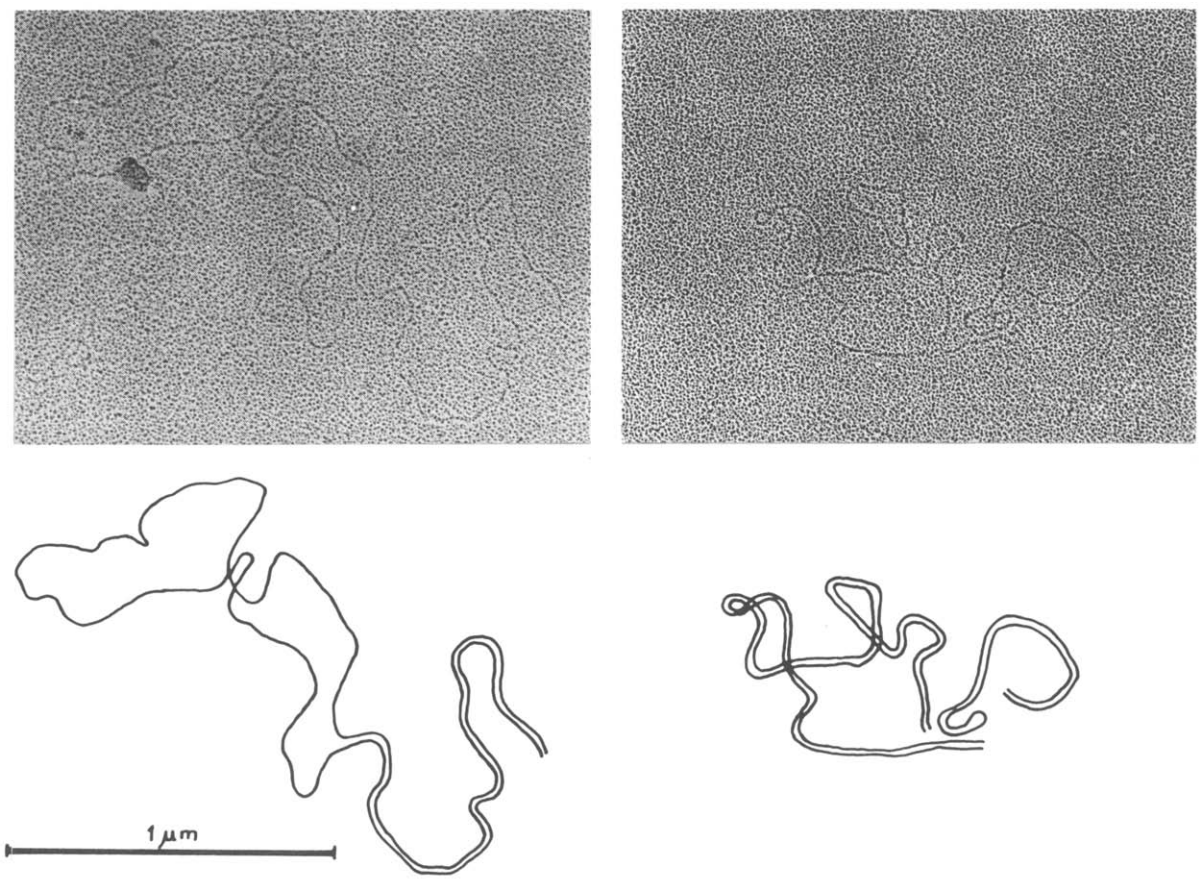

Fig.1. Electron micrographs of looped and unlooped hairpins in rat foldback.DNA.

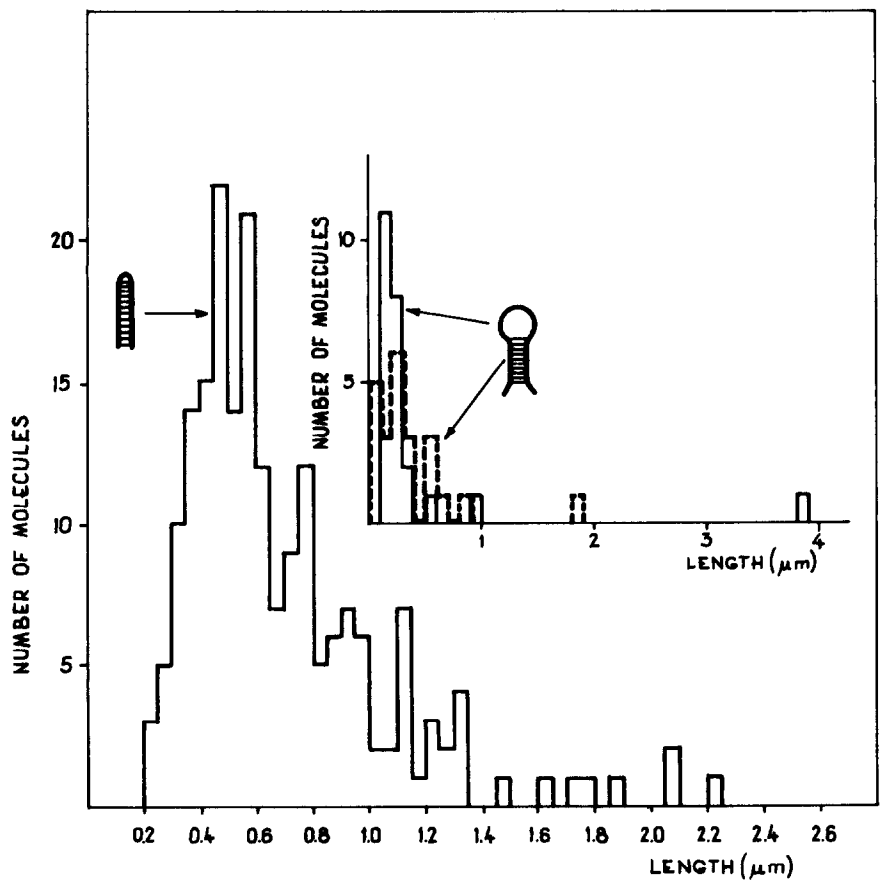

Fig.2. Length distribution of hairpins without discernible loops. Insert: Observed distribution of duplex-length and loop-length of looped hairpins.

The amount of both types of inverted-repeated sequences in rat genome was roughly calculated. Taking the haploid amount of rat DNA equal to 3×10^9 base pairs and assuming that 4% of rat DNA form double-stranded 'stems' [10], we obtain for the 'looped sequences': 6000 of sequences

$$\left[\frac{3 \times 10^9 \times 0.04 \times 0.1}{2 \times 1000}\right]$$

and for 'unlooped sequences': 27 000 of sequences

$$\left[\frac{3 \times 10^9 \times 0.04 \times 0.9}{2 \times 2000}\right].$$

The total amount of palindromic sequences/haploid rat genome is then 33 000 and is compatible with the amount of hairpin forming sequences present in the main band portion of the mouse haploid genome (40 000 obtained by Cech and Hearst [7], and for hamster (42 000) [23] but different from data for human DNA (about 120 000 palindromic sequences [9], or for *Drosophila* (2000–4000 inverted-repeated sequences) [6]. We do not know if these differences are species specific or are due to different methods of isolation of the foldback DNA. While in the mouse foldback DNA, looped molecules amount to about 60%, in rat DNA we found only small amount (10%) of looped molecules. On the other hand it is very difficult to exclude the possibility that during isolation of foldback DNA by the 'denaturation and

reassociation cycle' some of the very long loops are not destroyed.

In rat foldback DNA we have found long polypyrimidine sequences above 80 nucleotides in length (see fig.5 in [11]). In order to find out whether these polypyrimidine tracts contain internal, highly repeated sequences, like for instance polypyrimidine sequences in *Drosophila* genome [16,17] we analysed the distribution of oligonucleotides obtained by RNAase T_1 digestion of complementary, polypurine RNA obtained from transcription of polypyrimidine tracts isolated from foldback DNA (fig.3). Polypurine RNA labelled with [α-^{32}P]ATP, contains nine oligonucleotides: G, (52.4% total radioactivity), AG (22.6%), A_2G (13.7%), A_3G (6.9%), A_4G (2.1%), A_5G (1.1%), A_6G (0.5%), A_7G (0.3%) and A_8G (0.25%). Similar kind of analysis of cRNA obtained from polypyrimidine tracts of DNA of *Drosophila melanogaster* indicates the presence only of the two oligonucleotides AG and A_2G, suggesting that polypyrimidine tracts from Drosophila include highly repeated sequences TC and TTC [17]. Our analysis of cRNA transcribed from polypyrimidine tracts of rat foldback DNA argue against the internal repetitivity of sequences composed mainly of pyrimidines and indicate only that all possible combination of C and T are present in the polypyrimidine sequences (nevertheless G + AG + A_2G = 89% total radioactivity). Whether polypyrimidine sequences are present in looped molecules or in unlooped molecules we still do not know, but from the appearance of polypyrimidine tracts in S_1 digested molecules of rat foldback DNA it is unquestionable that these sequences are present in double-stranded stems, but not in the loops.

If we assume that pyrimidine tracts of about 80 nucleotides in size present in inverted-repeated sequences reach 3×10^{-4}%, long pyrimidine tracts in inverted-repeated sequences amount to:

$$\frac{3 \times 10^9 \times 3 \times 10^{-4}}{80} = 11\,000/\text{haploid genome.}$$

This means that not all of the inverted-repeated sequences contain long pyrimidine tracts because total amount of inverted-repeated sequences is equal to 33 000. Some of the inverted-repeated sequences contain most probably only one long (about 80 nucleotides) polypyrimidine sequence.

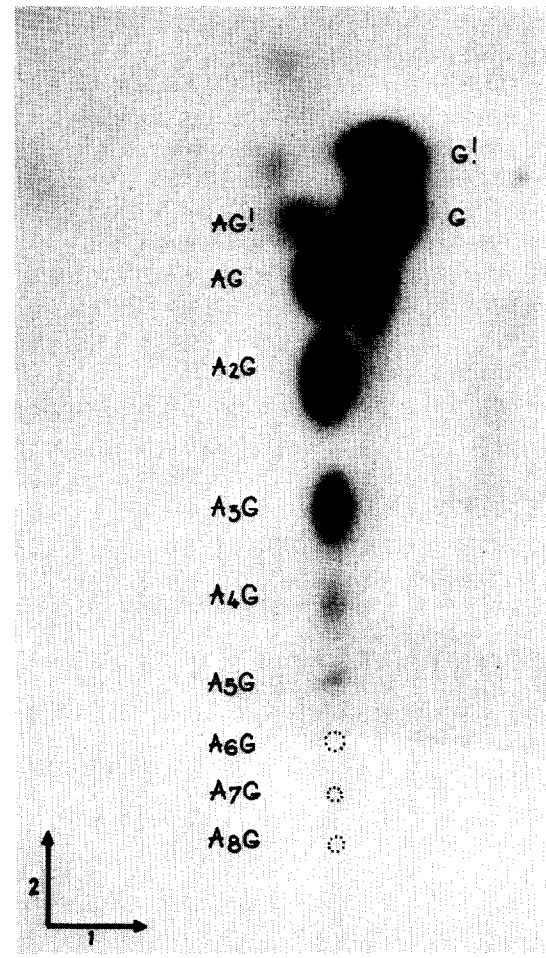

Fig.3. Fingerprint of RNAase T_1 digest of cRNA (labelled with [α-^{32}P]ATP) synthesized from polypyrimidine tracts isolated from rat foldback DNA. The digest was fractionated in two dimensions: (1) high voltage electrophoresis at pH 3.5 on cellulose acetate strip, then (2) on thin-layer plate of cellulose impregrated with polyethyleneimine developed with pyridine–formate buffer, 1.5 M (pH 3.5). (!) = cyclic oligonucleotides.

In summary, rat genome contains about 33 000 inverted-repeated sequences. Some of them are separated by a non-repeated sequence ranging in length from 600–11 400 base pairs. The length of inverted-repeated sequences are in the range from 150–6600 base pairs. At least some (11 000) of the inverted sequences contain heterogenous polypyrimidine tracts (about 80 nucleotides in length). Long palindromic

sequences contain presumably several types of polypyrimidine sequences, e.g. $(Py)_n$, $(Py)_x$, $(Py)_y$ etc.

The biological meaning of the palindromic sequences in eucaryotic DNA are still disputable. It is proposed that palindromes form pairs of hairpin loops in the DNA which could be important for recombination [20]. Cavalier-Smith [19] and Bateman [20] suggested a mechanism for the replication of the telomers of chromosomes which requires terminal palindromes. Lastly, Karrer and Gall [21] and Engberg et al. [22] gave some evidence that rDNA in *Tetrahymena* exists as a linear DNA molecules forming perfect palindromes over 20 000 pairs long.

References

[1] Walker, P. M. B. and McLaren, A. (1969) Molec. Gen. Genet. 104, 104–106.
[2] Britten, R. J. and Smith, J. (1970) Yearb. Carnegie Instn. 68, 378–386.
[3] Davidson, E. H., Hough, B. R., Amenson, C. S. and Britten, R. J. (1973) J. Mol. Biol. 77, 1–23.
[4] Cech, T. R., Rosenfeld, A. and Hearst, J. E. (1973) J. Mol. Biol. 81, 299–325.
[5] Wilson, D. A. and Thomas, C. A., Jr. (1974) J. Mol. Biol. 84, 115–138.
[6] Schmid, C. W., Manning, J. E. and Davidson, N. (1975) Cell 5, 159–172.
[7] Cech, T. R. and Hearst, J. E. (1975) Cell 5, 429–446.
[8] Perlman, S., Philips, C. and Bishop, J. O. (1976) Cell 8, 33–42.
[9] Deininger, P. L. and Schmid, C. W. (1976) J. Mol. Biol. 106, 773–790.
[10] Szala, S., Grabowska, M. and Chorąża, M. (1972) Acta Biochim. Polon. 19, 353–358.
[11] Szala, S., Bieniek, B., Michalska, J. and Chorąży, M. (1976) Biochim. Biophys. Acta 432, 129–144.
[12] Szala, S., Chorąży, M. and Kilarski, W. (1971) FEBS Lett. 17, 96–99.
[13] Straus, N. A. and Birnboim, H. C. (1974) Proc. Natl. Acad. Sci. USA 71, 2992–2995.
[14] Southern, E. M. and Mitchell, A. R. (1971) Biochem. J. 123, 613–617.
[15] Southern, E. M. (1974) Anal. Biochem. 62, 317–318.
[16] Birnboim, H. C. and Sederoff, R. (1975) Cell 5, 173–181.
[17] Sederoff, R., Lowenstein, L. and Birnboim, H. C. (1975) Cell 5, 183–194.
[18] Wagner, R. E. and Radman, M. (1975) Proc. Natl. Acad. Sci. USA 72, 3619–3622.
[19] Cavalier-Smith, T. (1974) Nature 250, 467–470.
[20] Bateman, A. J. (1974) Nature 253, 379.
[21] Karrer, K. M. and Gall, J. G. (1976) J. Mol. Biol. 104, 421–453.
[22] Engberg, J., Andersson, P., Leick, V. and Collins, J. (1976) J. Mol. Biol. 104, 455–470.
[23] Bell, A. J. and Hardman, N. (1976) Nucl. Acid Res. 4, 247–268.