Volume 4 no. 1 January 1977 Nucleic Acids Research

Characterization of foldback sequences in hamster DNA using electron microscopy

Anne J. Bell and Norman Hardman

Department of Biochemistry, University of Aberdeen, Marischal College, Aberdeen AB9 1AS, UK.

Received 26 November 1976

ABSTRACT

Foldback sequences in nuclear DNA from cultured Hamster fibroblasts (BHK-21/C13 cells) have been characterized by electron microscopy. One half of the structures observed when denatured hamster DNA is allowed to anneal in the range $0 < C_o t < 1 \times 10^{-4}$ M sec result from the annealing of inverted sequences forming foldback DNA. The remainder have a probable bimolecular origin. arising from rapidly-annealing sequences of satellite-like complexity. The average length of the inverted sequences in the foldback molecules is about 0.9 kilobases. There is estimated to be about 42,000 such sequences (21,000 pairs) in the hamster genome, approximately 45% of which form looped structures with a mean loop length of 1.74 kilobases.

Contrary to previous reports, binding of the renatured duplex molecules to hydroxyapatite results in a poor recovery of structures containing identifiable foldback sequences, due to preferential enrichment of the bound fraction with duplexes formed by intermolecular annealing.

INTRODUCTION

In all eukaryotic organisms so far studied, isolated nuclear DNA has been shown to contain repetitive and non-repetitive sequences [2]. The studies of Davidson *et al.* [3] and Donelson and Hogness [4] have shown that the organization of the repetitive sequences differs in detail between species. Classes of repeated sequences in eukaryotic DNA may be organized either in direct register, or some may be present in inverted order. Such sequence arrangements can be distinguished experimentally by the physical properties they confer upon the polynucleotide chains which contain them.

Repeated sequences present in direct order can be detected in denatured eukaryotic DNA fragments by a more rapid rate of association than unique sequences, when annealed under appropriate conditions [2]. Some of these repeated sequence

elements, the so-called middle-repetitive sequences, are interspersed with unique sequences [5], which has given rise to the general view that they may function as regulatory elements in the control of gene expression in higher organisms [6,7,8].

Inverted repeat sequences have similarly been found in the DNA of all higher organisms studied [3,9]. In contrast to DNA molecules containing direct-order repeated sequences, chains with inverted sequences have self-complementary copies of the sequence in each strand. Consequently, annealing results in rapid, concentration-independent, intra-chain base pairing. DNA molecules containing such sequences, often called foldback or hairpin DNA, can therefore be separated from the bulk of the DNA molecules by binding the spontaneously-annealing fraction selectively to hydroxyapatite, as described in several studies [3,9,10]. The function of inverted sequences in eukaryotic DNA is presently not known, but similar to some other classes of repeated sequences it has been suggested that they may also be involved in the control of genetic expression [11].

One of our primary interests is to investigate the organization of DNA base sequences surrounding genetically active segments of the eukaryotic genome, in relation to the sequence organization of the primary RNA transcripts and processed messenger RNA molecules. The hamster fibroblast is particularly relevant to this type of study because of its widespread use as an experimental system for the study of tumour virus-induced cell transformation. However, the organization of repetitive and non-repetitive sequences in hamster DNA has hitherto not been studied in detail. The purpose of the present series of experiments was to investigate foldback sequences in nuclear DNA from BHK-21/C13 cells.

We have used the electron microscope to characterize unfractionated hamster foldback DNA, and to examine foldback molecules bound to hydroxyapatite crystals. It is shown that hairpin molecules with shorter duplex stems are lost preferentially from the hydroxyapatite-bound fraction. In addition, a class of long, linear duplex-containing DNA

molecules, probably intermolecular in origin, contaminates the foldback fraction. The contaminating duplexes are selectively enriched by the hydroxyapatite fractionation procedure. These results suggest that in some cases there may be serious difficulties associated with the interpretation of data based on hydroxyapatite binding when this technique is used to 'purify' foldback molecules or to study the properties and organization of foldback sequences in eukaryotic DNA.

MATERIALS & METHODS

Chemicals and Enzymes

All chemicals were AnaLar grade (BDH Ltd., Poole, U.K.). Enzymes were purchased from the Sigma Chemical Co., London. Hydroxyapatite was prepared according to the method of Tiselius et al.[12], as modified by Wilson and Thomas [13], or was purchased from Bio-Rad Laboratories, Bromley, U.K., as Bio-Gel HTP. Both sources gave comparable results.

Preparation of Nuclear DNA from BHK-21/C13 Cells

[^{3}H]-labelled DNA was prepared from nuclei of cells grown exponentially at 37°C for 2.5 days in monolayer culture, in the presence of medium containing [Methyl-^{3}H] thymidine (0.25 μCi/ml). Cells were screened routinely for the absence of mycoplasma contamination. Cells were harvested at 4°C in hypotonic buffer (2 mM $MgCl_2$, 10 mM Tris-HCl, pH 8.0), lysed with 0.3%(v/v) Triton-N101, and the nuclei purified by repeated pelleting by centrifugation for 10 min at 800 x g through a solution containing 0.35 M sucrose, 2 mM $MgCl_2$, 10 mM Tris-HCl, pH 8.0. Nuclei were lysed using 2%(v/v) Sarkosyl NL, and then digested with heat-treated pancreatic RNase (50 μg/ml) at 37°C for 30 min. Proteinase K was added (500 μg/ml) and digestion continued for 4 hr at 37°C. DNA was isolated from the digest by successive isopycnic centrifugation steps, the first in a gradient containing Caesium sulphate, and the second containing Caesium chloride.

Preparation and Fractionation of Foldback DNA (Table 1)

8.8 μg of [^{3}H]-labelled hamster nuclear DNA (specific activity 2.3 x 10^4 cts/min/μg, weight-average single chain size 7 kilobases) in 3.9 ml of 0.12 M sodium phosphate, pH 6.8, was heat-denatured for 10 min at 100°C, cooled to 60°C, and

Table 1

Classification of Structures Observed in Hamster Foldback DNA

DNA	No. Structures Observed	% OF MOLECULES										
		Looped Hairpins type			Unlooped Hairpins type			Multiple Hairpins	Complex	Single Chains	Linear ds/ss	Forks
		a	b	c	a	b	c					
Unfractionated	798	3.0	0.6	0.3	2.4	2.5	0.6	0.9	0.6	88.2	0.9	0.1
Unbound to HAP	309	3.9	0	0	1.6	0	0	0.3	1.0	92.9	0	0.3
HAP-bound Foldback DNA	275	6.3	2.3	1.1	6.9	27.0	15.4	1.8	11.9	10.9	15.4	1.1
Formamide-denat. HAP-bound Foldback DNA	320	0.6	0	0	1.1	1.7	1.1	0	1.1	90.0	3.4	0.6

Structures are listed according to their structural appearance. The theory behind hairpin classification is described in Figure 1. Multiple hairpins are those structures which could be traced completely and contain two or more discernible hairpins. Complex molecules are structures with chain overlaps precluding an unambiguous interpretation. The linear double stranded/single stranded (Linear ds/ss) category contain a duplex segment with single chain terminals at each end. Forks are duplexes with at least three single chain terminals, as described previously[14]. Representative structures are shown in Plate 1. Experimental details are described in the text.

annealed to $C_o t = 1 \times 10^{-4}$ M sec. DNA samples (0.06 µg) were spread immediately for electron microscopy using a 55% formamide hyperphase, as described below. An identical sample of denatured and annealed DNA was passed through a 0.3 g column of hydroxyapatite maintained in a water-jacketed column at 60°C. Unbound single chains were washed through the column with 30 ml of 0.12 M sodium phosphate, and observed by electron miscoscopy. The duplex fraction was eluted with 12 ml of 0.4 M sodium phosphate, pH 6.8, then diluted with water to give 0.12 M sodium phosphate and prepared immediately for electron microscopy. Total recovery of DNA from columns was >95%. An average of 10% of the recovered material was found in the duplex fraction. For hydroxyapatite-bound molecules observed using 80% formamide, the duplex fraction was dialysed overnight into 0.1 mM EDTA, 10 mM Tris-HCl, pH 7.4, and concentrated by desiccation to 0.2 ml. 50 µl of this material was added to 0.8 ml of formamide and 0.1 ml of 0.1 M EDTA, 1.0 M Tris-HCl, pH 8.5, and was then denatured at 50°C for 10 min. After cooling quickly to room temperature, 50 µl of cytochrome C solution (1mg/ml) was added. Samples were spread immediately for observation in the electron microscope using a hypophase containing formamide (50%,v/v) and 10 mM Tris-HCl, pH 8.5.

Hydroxyapatite Co-fractionation of Foldback DNA and denatured Escherichia coli DNA

4.6 μg of [^{3}H]-labelled hamster nuclear DNA (specific activity 2.3 x 10^{4} cts/min/μg) and 18 μg of [^{14}C]-labelled E.coli DNA (specific activity 3.0 x 10^{2} cts/min/μg) were co-sheared to 8 kilobases, alkali-denatured in 0.5 M NaOH, neutralized with 0.05 M sodium phosphate, pH 6.8 and annealed to $C_o t$ = 1 x 10^{-3} M sec. The sample was passed through a hydroxyapatite column at 60°C, as described above. Single chains were removed from the column with 30 ml of 0.12 M sodium phosphate. The duplex fraction was then eluted with 12 ml of 0.4 M sodium phosphate. 91% of the [^{3}H] and 100% of the [^{14}C] radioactivity were recovered. Of the recovered material, 12.9% of the [^{3}H] counts and $<$0.01% of the [^{14}C] counts were recovered in the duplex fraction.

Electron Microscopy

Preparations were made as described by Hardman[14], using the protein monolayer technique of Kleinschmidt et al.[15], incorporating the modifications of Davis and Hyman[16] for use with formamide solutions. Hyperphase solutions contained DNA, 0.03 M sodium phosphate, pH 6.8, 50 μg cytochrome C/ml, 55%(v/v) formamide and 10 mM EDTA, 100 mM Tris-HCl, pH 8.5. The iso-denaturing temperature for DNA spreading under these conditions was calculated to be 63.4°C using the method described in the text for a 80% formamide hyperphase. Specimens were picked up on to copper grids coated with a Parlodian support film, and stained for 30 sec in a fresh solution of 50 μM uranyl acetate and 50 μM HCl in 90%(v/v) ethanol. Samples were then immersed for 10 sec in 2-methylbutane and air-dried.

Grids were rotary shadowed with Pt/Pd (4:1, w/w) and viewed in an AEI-EM6 electron microscope. Length measurements were made using a Depose-HC map measurer. The nucleotide equivalent per unit length of duplex DNA was found to be 3.05 kilobases/μm by spreading bacteriophage Lambda DNA duplexes under the above conditions using a 55% formamide hyperphase.

RESULTS & DISCUSSION

Electron Microscopy of Unfractionated Hamster Foldback DNA

Purified, sheared BHK-cell nuclear DNA was heat-denatured, annealed to $C_o t = 1 \times 10^{-4}$ M sec, and samples prepared for observation in the electron microscope, using a 55% formamide hyperphase. From electron micrographs, which were taken at random locations on grids, DNA structures were classified according to their structural appearanne. The results are shown in Table 1. The rationale behind the classification of hairpin structures is illustrated in Figure 1, and is similar to the theories presented previously[9,10].

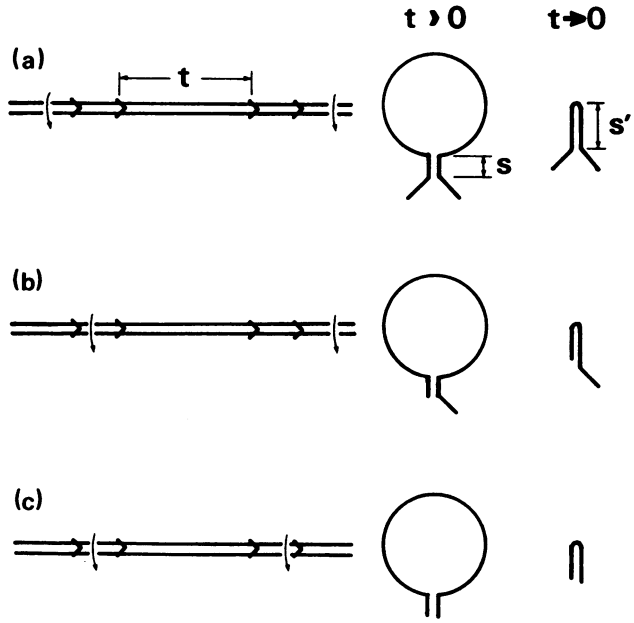


Figure 1
A duplex DNA segment is illustrated containing two homologous sequences arranged in inverted order, the polarity being indicated by ⇉⇇. Arrows show alternative sites of shear cleavage. The complementary inverted sequences in denatured chains undergo intramolecular annealing, forming hairpin structures with a loop corresponding to the length of the intervening non-homologous sequence, t . If t is sufficiently large to be observed in the electron microscope a looped structure results ($t > 0$), otherwise unlooped hairpins are observed ($t \to 0$). Type-a hairpins are those shown in (a), where measurements of the duplex stems, s and s', can be used to determine the length of the homologous inverted sequence. (b) and (c) illustrate the formation of hairpins of type-b and type-c respectively, which result if duplex chain scission occurs within the inverted sequence.

A large fraction (88%) of the DNA structures observed were found to be single chains, with no discernible duplex region. Over 85% (10.3/11.9) of the duplex-containing structures were hairpin-like, consistent with their being formed by intramolecular base pairing. Only 8.5% (1.0/11.9) of the duplex structures (linear ds/ss, and forks) were of a configuration corresponding to the major bimolecular rapidly-annealing components of Hamster DNA, studied by Bell and Hardman[17]. Some examples of the various classes of observed structure are shown in Plate 1.

Hydroxyapatite Fractionation

In an attempt to eliminate contaminating single chains, therby enriching the duplex foldback fraction, denatured hamster nuclear DNA, annealed to C_0t = 1 x 10^{-4} M sec, was passed through hydroxyapatite crystals at 60°C in 0.12 M sodium phosphate. This procedure effectively removes all unbound single chains in the 0.12 M sodium phosphate wash, as shown by the control experiments shown in Figure 2, and as previously demonstrated by Wilson and Thomas[9].

In similar experiments, the hamster foldback fraction was obtained for 7 kb[18] DNA by heat-denaturation, annealing to C_0t = 1 x 10^{-4} M sec, and selective binding to HAP[19]. The unbound fraction, washed through the column with 0.12 M sodium phosphate, and the duplex-containing fraction obtained by elution with 0.4 M sodium phosphate, were examined by electron microscopy as above. The DNA structures observed were classified and compared with unfractionated DNA. The results are presented in Table 1.

The results show that all DNA structural types observed to contain duplex segments are enriched by selective binding to HAP, and the proportion of single chain species is diminished, in accord with the known fractionation properties of HAP. Not all duplex species were retained by HAP with equal efficiency. Type-a looped hairpins and multiple hairpins are enriched approximately two-fold (6.3/3.0 and 1.8/0.9, respectively) whereas the enrichment of type-b and type-c looped hairpins is about four-fold (2.3/0.6 and 1.1/0.3, respectively). About three-fold enrichment is observed for

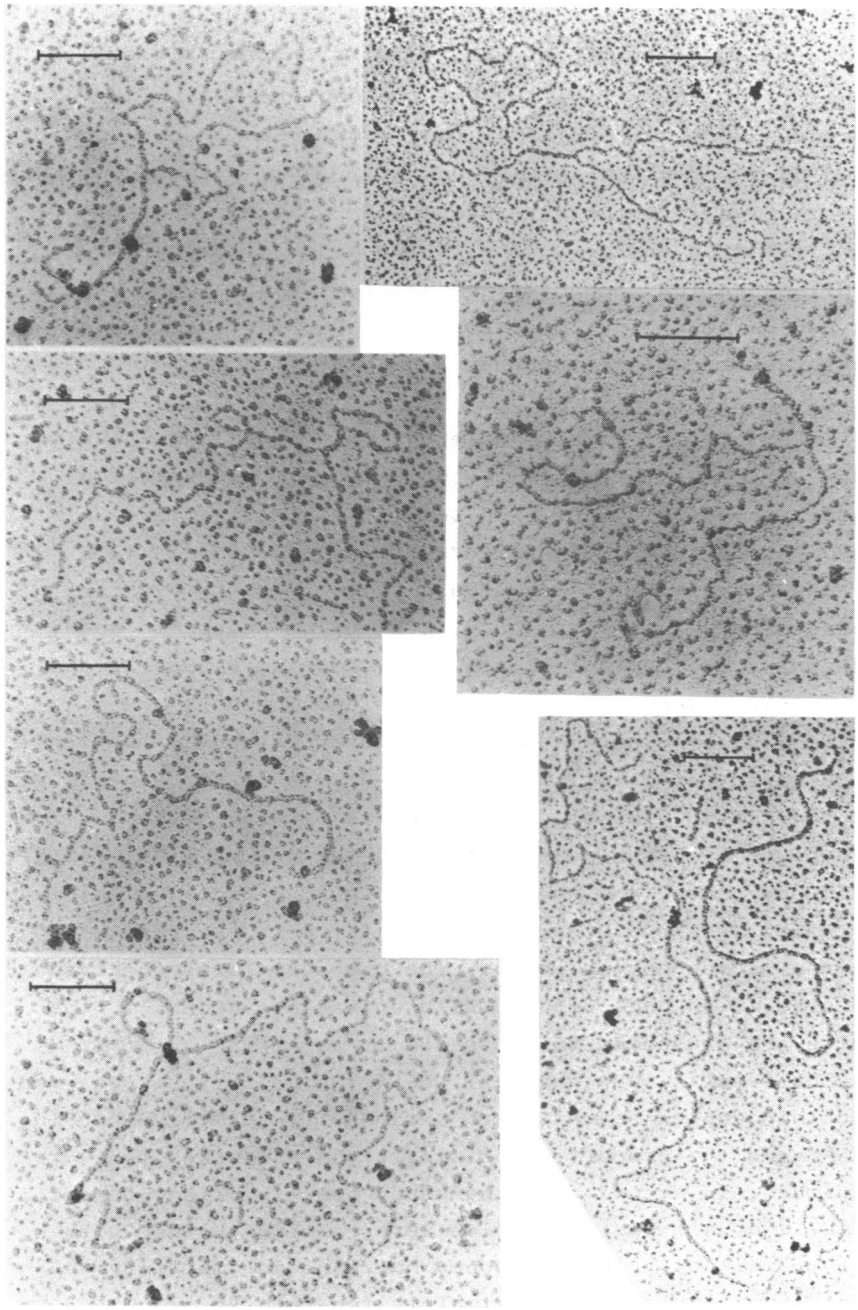
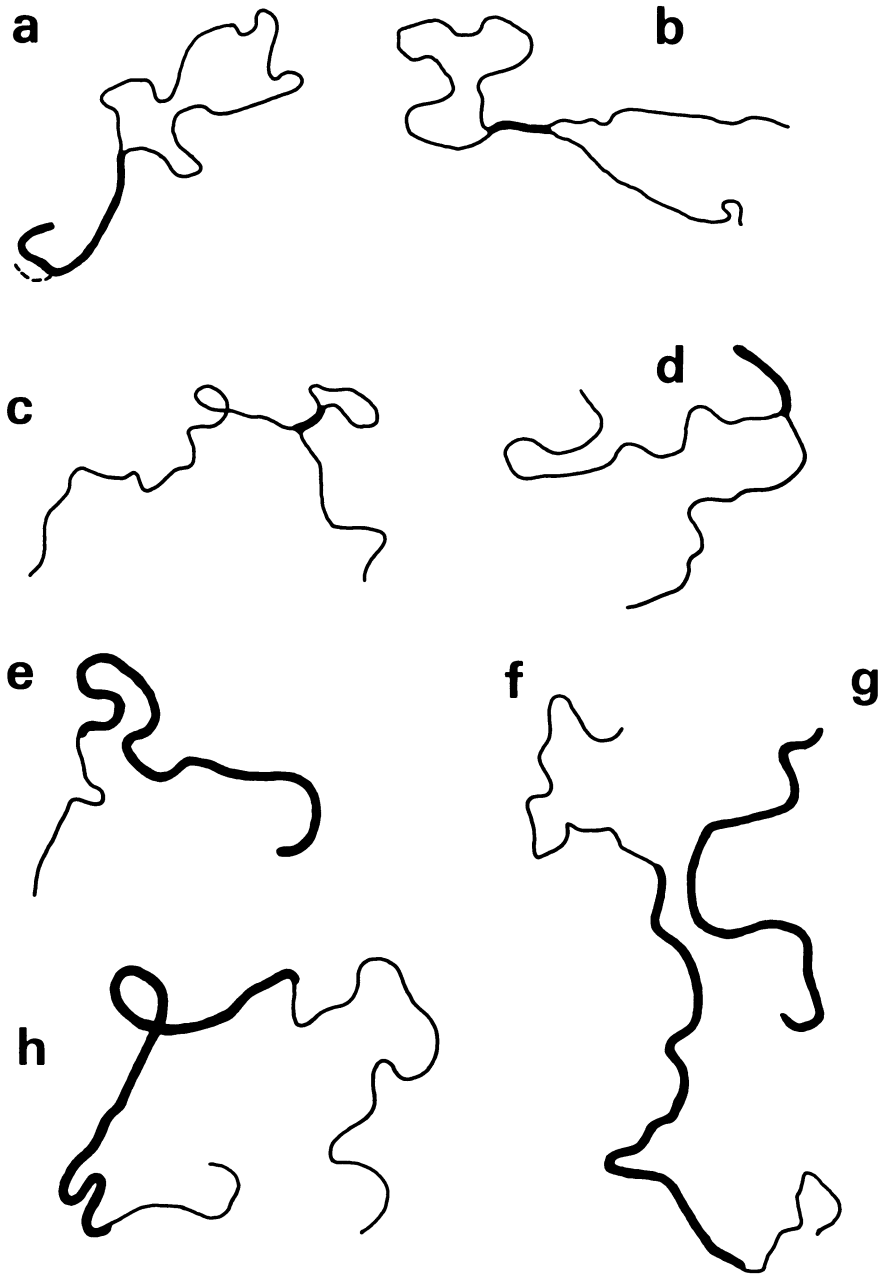


Plate 1
Hamster nuclear DNA was denatured and annealed to a $C_o t$ value of 1 x 10^{-4} M sec, and molecules spread for observation in the electron microscope using a 55% formamide hyperphase, as described for Table 1. (a) A probable type-c looped hairpin



(a doubtful chain is indicated by a broken line), (b) type-a looped hairpin, (c) type-a looped hairpin, (d) type-a unlooped hairpin, (e) type-b unlooped hairpin, (f) linear double/single chain molecule, (g) type-c unlooped hairpin, and (h) linear double/single chain molecule. The bars represent 0.2 μm.

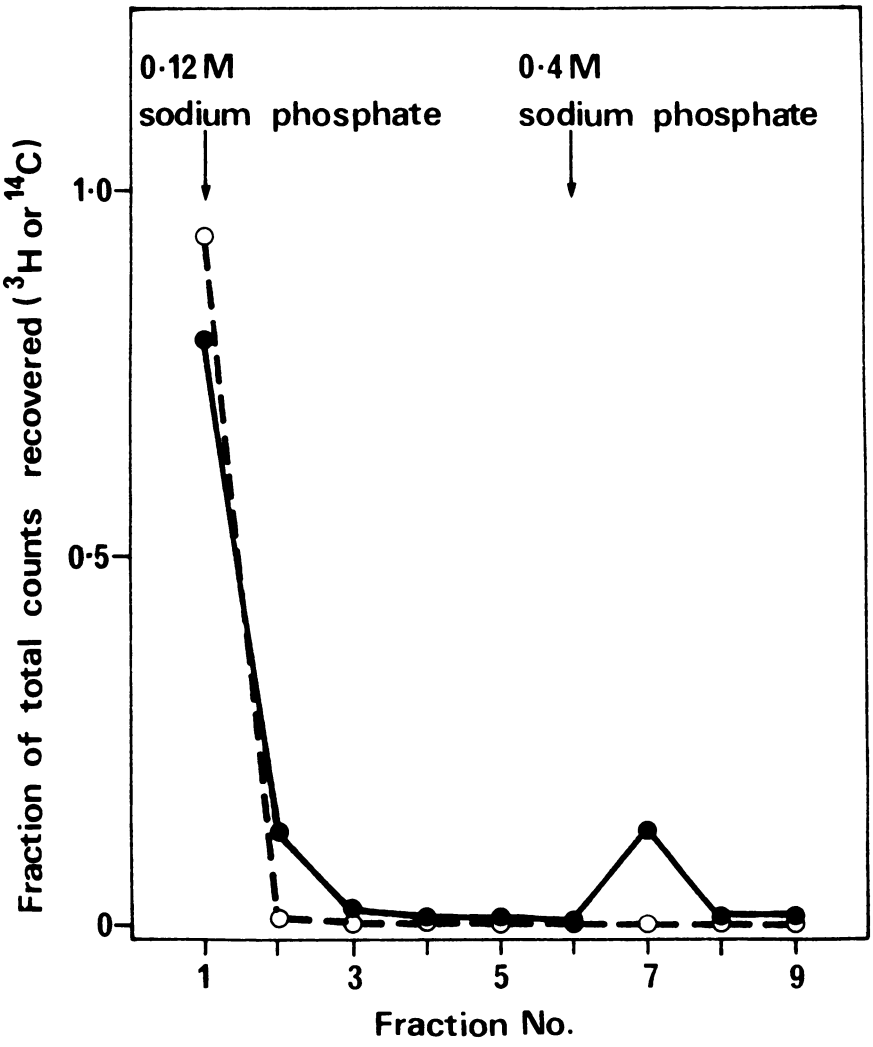


Figure 2
Foldback DNA was isolated using hydroxyapatite charomatography. Experimental details are given in the text. O—O [^{14}C] radioactivity (*E. coli* DNA), ●—● [^{3}H] radioactivity (Hamster DNA).

type-a unlooped hairpins (6.9/2.4). In contrast, retention of other duplex-containing molecules by HAP is considerably greater, about ten to twenty five-fold. The results for unbound foldback DNA confirms these observations, showing the presence of significant numbers of those duplex-containing structures which are less efficiently retained by HAP. In order to find a possible explanation for the apparently

anomolous binding properties of HAP, measurements were taken from electron micrographs to determine the overall chain size and the length of the duplex segment of each of the major categories of DNA structure observed. The results are summarized in Table 2.

Table 2

Stem Lengths and Total Single Chain Lengths of Unfractionated and HAP-Fractionated Hamster Foldback DNA

Type of Structure	LENGTH (μm) Unfractionated DNA	HAP-bound Foldback DNA	Unbound Fraction
Single chains	1.47(100)	1.32(27)	1.29(200)
Unlooped Hairpins			
(i) stem			
a	0.32(31)	0.29(21)	0.10(13)
b + c	1.50(22)	1.20(44)	ND
(ii) total length			
a	2.19(27)	1.29(27)	1.51(13)
b + c	3.53(12)	3.10(43)	ND
Looped Hairpins			
(i) stem			
a	0.30(33)	0.29(14)	0.10(14)
b + c	0.49(7)	ND	ND
(ii) total length			
a	2.64(33)	1.76(7)	1.55(4)
b + c	2.51(7)	ND	ND
(iii) loop length			
a	0.58(33)	0.71(14)	0.40(4)
b + c	0.50(7)	ND	ND
Linear ds/ss			
(i) duplex segment	0.93(8)	1.20(26)	ND
(ii) total length	3.90(8)	4.20(26)	ND

Lengths s, t, and s' were measured from hairpin structures and some other categories, observed and classified as described in Table 1 and Figure 1. Total chain lengths were determined in all cases by the summation of the single chain lengths plus twice the duplex length of individual molecules. Figures in brackets refer to the number of observations. Those structural types not detected in the sample are indicated by ND.

The general conclusion from the analysis of the data presented in Tables 1 and 2 is that the structures with smaller duplex regions are those retained least efficiently by HAP. Moreover, in the case of the type-a looped and unlooped hairpin classes, shorter hairpins (about 0.1 μm or 0.3 kb) appear in the unbound fraction. However, it is known from previous studies[13] that under similar conditions HAP crystals can retain DNA duplexes much smaller than 0.3 kb in length very efficiently. Possible explanations for the present results are that the smaller hairpins are selectively prevented from binding to HAP either because they are

covalently attached to single chain terminals or loops, or because of sequence mismatching. Wilson and Thomas[9], in a different context, have described what may be a similar effect with HeLa-cell foldback DNA. In this previous study, however, the melting temperature of the hairpin fraction seems to rule out the possibility of extensive mismatching. Since the HAP binding step clearly leads to selection of hairpins containing larger duplex stems, we have confined the following detailed analysis of hairpin stem lengths to unfractionated DNA molecules, and used this data to provide estimates for the length of hamster DNA foldback sequences.

Analysis of Hairpin Structures

Of the three possible types of hairpin structure (Figure 1) stem lengths of type-a hairpins are the only measurements which give an unambiguous estimate of the length of the inverted sequences. Cech and Hearst[10] determined that the total single chain molecular weight of hairpin-containing strands in mouse main-band DNA was not significantly different (13.4 kb) from the bulk single chain molecular size of 12.6 kb, used in their experiments, consistent with the interpretation that the structures are formed on one strand by intramolecular annealing. In the present study, the average single chain size of type-a looped and unlooped hairpins, compared with the single chain category, gives a different result, but one which we believe is nevertheless consistent with these classes of structures forming as a result of intramolecular annealing. From the length measurements presented in Table 2 for unfractionated DNA, it can be seen that the average single chain length of type-a looped and unlooped hairpins is longer than the single chain category. For unlooped hairpins, we explain this marginally larger size (2.19μm compared with 1.47 μm) as being due to the requirements for two 0.32 μm copies of complete tandem inverted sequences each terminated by a single chain tail in order to be detected in the type-a category, a criterion which would be met only by the larger molecules in a population with a relatively short overall chain length. A similar augument can explain the larger

single chain size of type-a looped hairpins (2.64 μm), since these molecules carry an additional 0.58 μm single chain loop (Table 2). These considerations were probably not necessary in the previous work described above[10], because the contribution of the average duplex stem (about 1.0 kb) was a much smaller contribution to the total DNA chain size.

We have used measurements of stem and loop lengths of type-a hairpins to determine the length of the inverted sequences in hamster DNA (Figures 3 and 4). For unlooped hairpins, the duplex stem lengths, s', were in the range 0.1-3.0 kb, with a number-average size of 0.98 kb (Figure 3a). Stem lengths of type-a looped hairpins gave a similar distribution, but with a slightly smaller number-average size (0.92 kb, Figure 3b). Loop sizes, t , ranged from 0.3-9.0 kb, with a mean value of 1.8 kb (Figure 4).

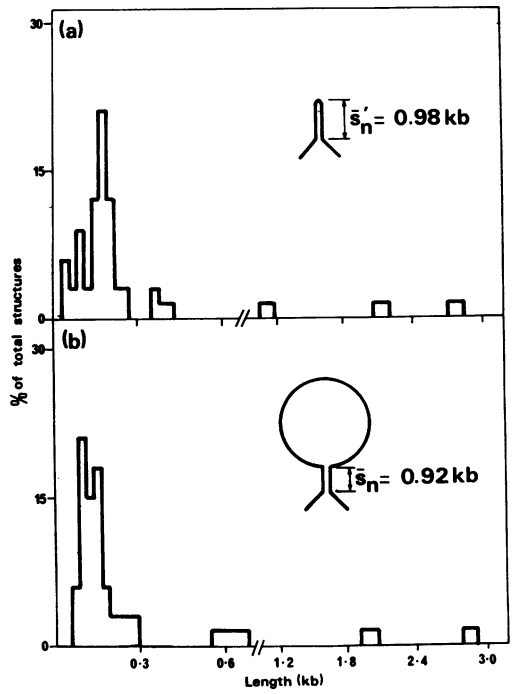


Figure 3
Hairpin stem lengths, s' and s, were measured from 31 type-a unlooped hairpins, shown in (a), and 33 type-a looped hairpins, shown in (b), in unfractionated hamster foldback DNA, observed as described for table 1. The number-average stem lengths ($\bar{s}'_n$ and $\bar{s}_n$) are shown in each case.

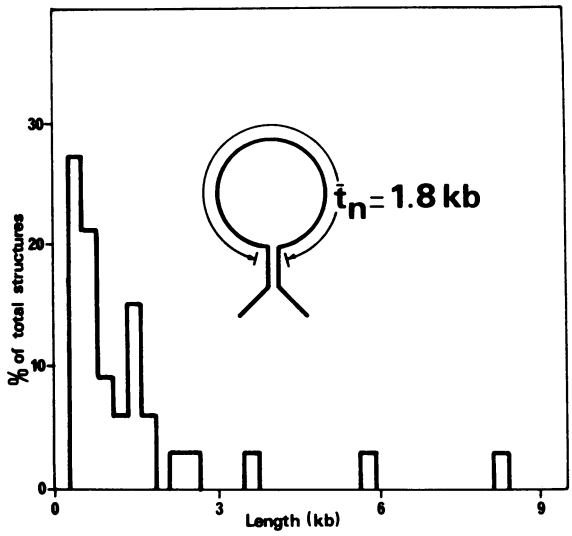


Figure 4
Loop lengths, t , of 33 looped hairpins (type-a) are shown from unfractionated foldback DNA, observed as described for Table 1. The number-average loop length, $\bar{t}_n$, is indicated.

Assuming that random cleavage occurs in regions of hamster DNA containing hairpin sequences, the fraction, F , of hairpins which would be expected to possess two recognizable single chain termini is given by equation 4 of Cech and Hearst[10]:

$$F = \frac{L - t - 2s - 2b_o}{L - t - 2s_o}$$

where L is the single chain fragment length, t the loop size, s the hairpin stem length, s_o the minimum observable stem length, and b_o the minimum length of single chain terminus which can be recognized. We have previously estimated the values of s_o and b_o to be about 0.03 μm (Hardman, N. and Jack, P.L., unpublished results). For the average looped hairpin (L = 2.64 μm, t = 0.58 μm, s = 0.3 μm) we calculate F to be 0.7. We therefore expect that about 3/4 of the looped hairpins in unfractionated DNA to have two single chain termini (type-a, Figure 1) which is in reasonable accord with observation (F_{obs} = 3.0/3.9 = 0.77, Table 1). On the other hand, for unlooped hairpins, where L = 2.19 μm, t = 0 and s = 0.32 μm, F becomes 0.7, which is far greater than the

observed value of 0.31 (2.4/5.5, Table 1). From this, we conclude that a second group of molecules is present in the type-b and type-c hairpin categories which either contain much longer hairpin sequences (approx. 1.5 μm, Table 2), or alternatively are molecules which have been wrongly classified as hairpins. In the section which follows, we deduce that these large, linear duplex-containing molecules in the type-b and type-c hairpin classes are contaminants which result from intermolecular annealing.

Origin of Large Hairpin-Like Molecules

Two lines of evidence suggest that a large proportion of the molecular structures scored as type-b and type-c unlooped hairpins are products of intermolecular annealing. First, the average duplex stem length is far greater than might be expected if all three classes of hairpins are derived from the same population of molecules (see above). Second, the total chain length of type-b and type-c unlooped hairpins (3.53 μm, Table 2) is similar to the total chain length of the linear ds/ss category (3.9 μm, Table 2), a component which is known to be characteristic of one of the rapidly-annealing bimolecular components in hamster DNA, described by Hardman and Bell[20]. The total chain length of both these components is over twice the average single chain molecular size of 1.47 μm, in accord with an intermolecular origin.

To investigate this point further, hydroxyapatite-bound foldback DNA, which preferentially enriched with these long, duplex-containing molecules (Table 1), was denatured at 50°C for 10 min in the presence of 84% formamide and observed by electron microscopy using a hyperphase containing 80% formamide. Taking a value of 88°C for the t_m of the HAP-bound foldback fraction in 0.1 M sodium phosphate (Figure 5), and using the empirical relationship derived by Thomas and Dancis[21] between melting temperature and molarity of monovalent cation, the Δt_m in 0.02 M Na^+, used for DNA spreading using 80% formamide, is -14.5°C. A formamide concentration of 80% produces a denaturing effect equivalent to raising the temperature by 57.6°C, as shown by McConaughy, Laird and McCarthy[22]. Thus, the iso-denaturing temperature equivalent

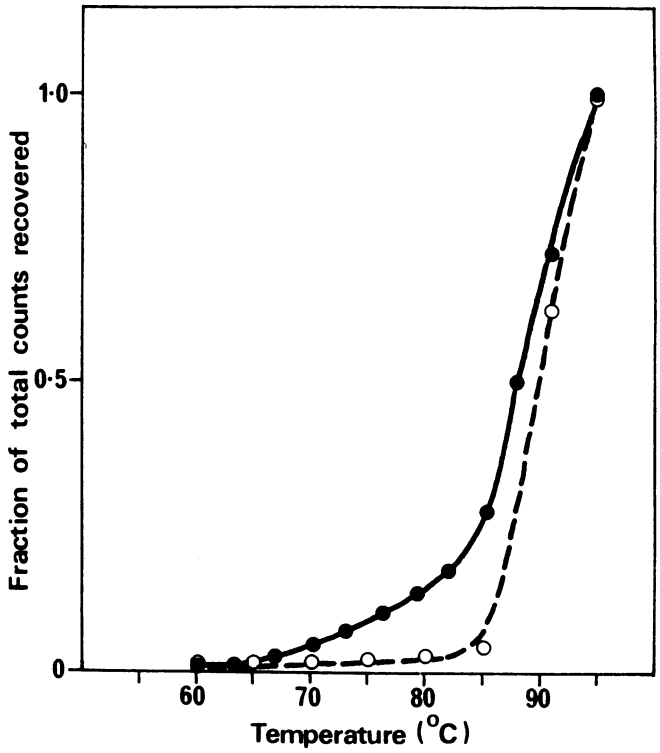


Figure 5
8 μg of heat-denatured, [^{3}H]-labelled hamster nuclear DNA was annealed to $C_o t$ = 1 x 10^{-4} M sec, and bound to a 0.3 g HAP column, as described in the experimental details to Table 1. After elution of single chains with 30 ml of 0.12 M sodium phosphate, the column was equilibrated with 10 ml of 0.1 M sodium phosphate at 60°C. The temperature was raised in 3°C increments, eluting the denatured material with 4 ml of 0.1 M sodium phosphate at each temperature. Duplex material still bound to the column at 95°C was eluted using 4 ml of 0.4 M sodium phosphate. 102% of the [^{3}H] label was recovered. In a separate experiment, 2 μg of native [^{3}H]-labelled hamster nuclear DNA was bound to HAP, and thermally eluted using 5°C increments, similar to the above procedure. 102% of the radioactive label was recovered. ●—● HAP-bound foldback DNA, O—O native hamster DNA.

to the spreading conditions in 80% formamide at room temperature (20°C) is 20 + 57.6 + 14.5 = 92.1°C, which is about 4°C above the melting temperature of the HAP-bound DNA, as shown in Figure 5. Hence, the majority of molecules denatured at 50°C in formamide would be expected to remain denatured when observed by electron microscopy. The results shown in Table 1 show that the majority of species observed

are denatured, and are present as single chains under the conditions employed. The major proportion of these single chains are derived from the duplex-containing structures observed in undenatured HAP-bound foldback DNA. The chain length distribution of these molecules is presented in Figure 6, showing the mean length of the chains to be 1.1 μm, close to the mean size of those single chain structures observed in undenatured HAP-bound foldback DNA (1.32 μm), and in unfractionated DNA (1.47 μm) as indicated in Table 2. Moreover, a large proportion of these molecules, when annealed to $C_o t = 1 \times 10^{-4}$ M sec, and prepared for observation by electron microscopy using 50% formamide, form linear ds/ss structures with an overall chain length of 2.5 μm. Thus, although it is possible that some single chain scission has occurred during DNA manipulation, and this is to be expected, a significant fraction of the duplex structures bound to HAP are probably bimolecular in origin. Since the majority of the duplex species bound to HAP are type-b and type-c hairpins and the linear ds/ss category (Table 1), we conclude that the type-b and type-c 'hairpins' do not have a foldback structure, but, like ds/ss linear molecules, result from intermolecular rather than intramolecular annealing. No chains are observed in denatured HAP-bound foldback DNA which correspond in length to the size of these structures, a result which would be expected if the molecules had a true hairpin configuration (3.10 μm and 4.2 μm, Table 2). The misclassification of the structures in the type-b and type-c hairpin categories may be due to the presence of short ($<$ 50-100 bases) single chain termini, too small to be observed using the electron microscope. How these structures might have arisen is not known. Some may arise by preferential shear breakage adjacent to the junctions of the duplex and single chain segments of linear ds/ss molecules on binding to HAP. Some chain scission is known to occur when DNA molecules are bound to HAP using similar procedures[10, 23].

Number of Hairpin Sequences in Hamster DNA

If the interpretation of the above data is correct, then making some assumptions, we can arrive at a crude estimate

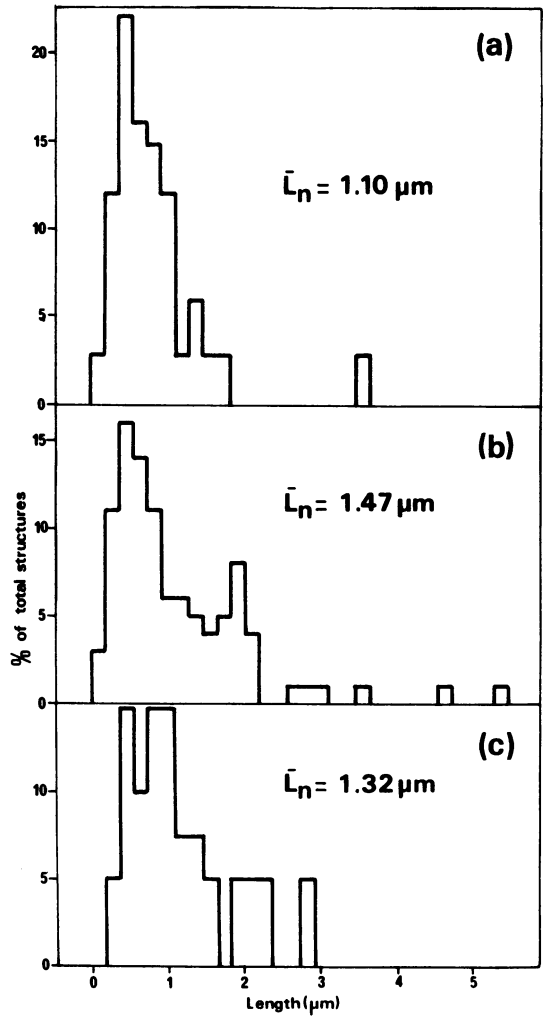


Figure 6
(a) shows the length distribution of 34 single chain category molecules from HAP-bound foldback DNA, denatured at 50°C in 84% formamide and observed by electron microscopy using an 80% formamide hyperphase, as described in the text. The number-average chain length, $\bar{L}_n$, is indicated. The data is compared with (b) 100 single chain category molecules in unfractionated foldback DNA, and (c) 27 single chain category molecules in undenatured HAP-bound foldback DNA.

for the number of hairpin sequences in hamster DNA. For this calculation, we need to know the double chain yield of hairpin sequences, and the extent of contamination, by weight, of these sequences by intermolecular products, namely the

linear ds/ss category and most of the molecules scored as type-b and type-c hairpins. This can be estimated from the electron microscope data from the yield and single chain size of the major structure categories (Table 1 and Table 2). For this analysis, we have neglected the contribution from the minor categories, such as forks, multiple hairpins and uninterpretable (complex) structures, and assume that the fraction of recognizable hairpins of type-a represent 0.7 of the total hairpin sequences present (see above). Thus, we estimate that approximately 42% by weight of the total duplex yield of foldback DNA consists of true hairpin sequences. We have previously determined the total duplex yield of hamster foldback DNA at $C_o t = 1 \times 10^{-4}$ M sec to be 3% of the haploid nuclear DNA content, which corresponds to 9.1×10^4 kilobase pairs, by treatment with S1 nuclease[17]. Since 42% of this material is true foldback sequence (equivalent to 3.8×10^4 kilobase pairs), and each foldback sequence is about 0.9 kb in length, we deduce that there are probably about 21,000 pairs of hairpin sequences in the hamster genome.

Conclusions

Under the conditions presently used, we have shown that less than half of the hamster DNA sequences which anneal below $C_o t = 1 \times 10^{-4}$ M sec are probably hairpin sequences. Originally, the reason for the choice of this $C_o t$ value was to allow looped hairpin sequences to anneal which might have been excluded from the analysis at lower $C_o t$ values due to their supposed slower rate of annealing[10,24]. This is also the conventional $C_o t$ value used in several studies of foldback DNA[3,9,23]. Although annealing under these conditions prevents significant intermolecular annealing involving DNA fragments with relatively high sequence complexity, such as E.coli DNA fragments in the present work (Figure 2) and T7 DNA fragments in the experiments performed by Wilson and Thomas[9], it does not eliminate the possibility of contamination by intermolecular components with a much lower, satellite-like, complexity within the hamster genome. Such components have been identified recently in rodent main-band DNA. In the mouse, a component with satellite-like properties is thought to comprise at least

0.5% of the total sequences[25]. In hamster nuclear DNA a similar minor component is present, containing about 1-2% of the nuclear DNA sequences[20].

In those cases where the possibility of intermolecular annealing appears to be eliminated, such as in the study of HeLa-cell foldback DNA by Wilson and Thomas[9], HAP fractionation can be validly used to enrich foldback sequences and study their genome distribution. However, in other instances where the possibility of some intermolecular annealing remains, such as in the case of studies of mouse foldback DNA[9,26], then HAP fractionation may present some special, hitherto ignored, problems. In the present study we find that a significant fraction of the duplex foldback molecules are lost in the unbound HAP fraction and the relative proportion of contaminating intermolecular duplexes increases about three or four-fold in the HAP-bound fraction. We have characterized the HAP-bound sequences, by determination of their melting temperature (Figure 5) and by measurement of HAP yield as a function of DNA fragment size (Bell, A.J. and Hardmen, N., unpublished results), but we do not feel justified in assuming that the results of such experiments reflect meaningfully the properties of the hairpin sequences until the presence of contaminating components is eliminated. Such considerations may not have been taken into account in previous studies of HAP-bound foldback DNA[3,9,23,26].

In spite of these difficulties, the character and abundance of identifiable hairpin sequences in hamster DNA have been investigated in the present study for unfractionated foldback DNA. These properties seem quite similar to those of the corresponding hairpins in mouse main-band DNA[10]. We suggest that the class of large, linear duplex molecules of unknown origin in mouse foldback DNA reported in this previous study[10] are possibly intermolecular in origin. We cannot exclude the possibility that due to the choice of relatively short input DNA single chains in the present work we have eliminated from the analysis particularly long hairpin sequences, or potential hairpins with long loops. Studies are presently in progress to investigate whether such

components are present in hamster foldback DNA, using larger single chains.

It could be argued that contaminants of the foldback DNA fraction might arise from sources other than intermolecular annealing. Residual undenatured DNA, or perhaps the presence of interchain cross-links would contribute to the sources of contamination of the low $C_o t$, hairpin fraction. We consider the first of these possibilities unlikely, since the HAP-bound duplexes can be denatured using a combination of formamide and moderate temperature (Table 2). Considering the second possibility, cross-linked DNA has been detected in Chinese hamster ovary DNA[27], but the amounts are far too small to account for but a minor fraction of hamster DNA found in the foldback fraction in the present study.

Acknowledgements

We are grateful to the Medical Research Council for financial support.

REFERENCES

1 $C_o t$ is the product of the concentration of nucleotide in molar and the time of annealing in sec.
2 Britten, R.J. and Kohne, D.E.(1967) Carnegie Inst. Wash. Yearbook 66, 73-108
3 Davidson, E.H., Hough, B.R., Amenson, C.S. and Britten, R.J. (1973) J. Mol. Biol. 77, 1-23
4 Donelson, J.E. and Hogness, D.S. (1974) Cell 3, 315-325
5 Manning, J.E., Schmid, C.W. and Davidson, N. (1976) Cell 4, 144-155
6 Jelinek, W., Molloy, G., Solditt, M., Wall, R., Sheiness, D. and Darnell, J.E. Jr. (1973) Cold Spring Harbor Symp. Quant. Biol. 38, 891-898
7 Davidson, E.H., Hough, B.R., Klein, W.H. and Britten, R.J. (1975) Cell 4, 77-93
8 Lewin, B. (1975) Cell 4, 11-20
9 Wilson, D.A. and Thomas, C.A. Jr. (1974) J. Mol. Biol. 84, 115-138
10 Cech, T.R. and Hearst, J.E. (1975) Cell 5, 429-446
11 Ryskov, A., Saunders, G., Farashyan, V. and Georgiev, G.P. (1973) Biochim. Biophys. Acta 312, 152-164
12 Tiselius, A., Hjerten, S. and Levin, O. (1956) Arch. Biochem. Biophys. 65, 132-155
13 Wilson, D.A. and Thomas, C.A. Jr. (1973) Biochim. Biophys. Acta 331, 333-340
14 Hardman, N. (1974) Biochem. J. 143, 521-534
15 Kleinschmidt, A.K., Lang, D. and Zahn, R.K. (1950) Z. Naturforsch. 14, 770-779
16 Davis, R.W. and Hyman, R.W. (1971) J. Mol. Biol. 62, 287-301

17 Bell, A.J. and Hardman, N. (1976) 10th Int. Congress Biochem. Abst., p. 21
18 kb = kilobase
19 HAP = Hydroxyapatite
20 Hardman, N. and Bell, A.J. (1976) 10th Int. Congress Biochem. Abst., p. 19
21 Thomas, C.A. Jr. and Dancis, B. (1973) J. Mol. Biol. 77, 43-55
22 McConaughy, B.L., Laird, C.D. and McCarthy, B.J. (1969) Biochemistry 8, 3289-3293
23 Perlman, S., Phillips, C. and Bishop, J.O. (1976) Cell 8, 33-42
24 Schmid, C., Manning, J. and Davidson, N. (1975) Cell 5, 159-172
25 Cech, T.R. and Hearst, J.E. (1975) J. Mol. Biol. 100, 227-256
26 Cech, T.R., Rosenfeld, A. and Hearst, J.E. (1973) J. Mol. Biol. 81, 299-325
27 Rommelaere, J. and Miller-Faures, A. (1975) J. Mol. Biol. 98, 195-218