Volume 8 Number 17 1980 Nucleic Acids Research

The organisation of the long range periodicity calf satellite DNA I variants as revealed by restriction enzyme analysis

Gérard Roizes, Michel Pages and Christian Lecou

Equipe de Recherche de Biophysique du CNRS, USTL, 34060 Montpellier, France

Received 9 June 1980

ABSTRACT : The analysis of a large number of restriction sites within the long range periodicity calf satellite DNA I does not reveal a superimposable shorter repeat. Although some restriction sites are present in almost all the 100,000 tandemly arranged copies of the 1460 bp repetition unit, other sites such as Atu CI occur at much lower frequencies. When present they are distributed randomly along the satellite DNA molecules. The missing sites appear to result from random and presumably single base alterations. Digestion with the enzymes Hha I and Kpn I showed another type of variant to exist within the calf satellite DNA I. Unlike Atu CI the distributions of the variants detected by these enzymes are not random and organised on long stretches of satellite DNA. The possible functional significance and evolutionary implication of these results are discussed.

INTRODUCTION

The use of restriction endonucleases in the study of satellite DNAs has shown these highly repetitive sequences to posses a greater complexity than originally expected.

The restriction analysis of a number of satellite DNAs has shown that both the length and sequence of the repetition units differ considerably (1-4).These tandem repeated DNAs are of two major types : those consisting of very short nucleotide sequences between 2 and 10 base pairs in length such as found in crabs or Drosophila virilis (5-7) and those that have longer and more complex repeat units (1-3). However, there is no clear cut distinction between these two types of distribution. For instance, the mouse satellite DNA has a short repeat which occurs in a longer tandemly organised repeat of 242 base pairs (8). The relation between these two repeats is not yet understood (24).

It has also been shown that restriction enzymes often generate multimers as well as the monomers and this presumably

results from an alteration of the considered restriction site in the DNA sequence. A more important observation which has been interpreted as being the result of unequal crossing over, is the not infrequent occurence of DNA fragments which are fraction -in length- of the monomer (e.g. 1/2 or 1 and 1/2 times the monomer)(2,9).

It has been shown that the heterogeneities which were known to exist in the repeats of calf satellite DNA I are confined to specific regions of the sequence (10). Two types of sites were described, those which are present in all the repetition units(e.g. Eco RI or Mbo I) and those which are present in only a fraction of them (e.g. Hha I, Mbo II, Sac I or Hinf I). These results led to the suggestion that calf satellite DNA I could be organised in several subsatellite fractions perhaps localised on different chromosomes.

The existence of different sorts of heterogeneities in satellite DNAs has now been widely confirmed (9,11,12). When present in almost all of the repetition units, the sites have been called type A while those present in only a small portion of them have been called type B (9).

Satellite DNA variants have been shown to be present on specific Drosophila (13) and human (14-16) chromosomes. Brown and Dover (17) showed that the satellite DNA sequences of the X chromosome of Mus musculus have a specific organisation.

Recently, molecular cloning and sequencing of satellite DNAshas been ofgreat help in determining their fine structures. Of particular interest is the work of Hsieh and Brutlag (3) which shows that the repetition unit of D.melanogaster 1.688 satellite DNA contains a few single base changes at certain hotspots along the 359 bp long repetition unit. It has also been shown that a second tandem repeat of 254 bp has the same original sequence as the 359 bp repeat and is located in a different region of the 1.688 satellite DNA molecules (18). They point out that the variations in the sequence affect only a limited number of sites, presumably for functional reasons and recall their finding of a protein which preferentially binds to 1.688 satellite DNA in a particular region of the repetition unit (19).

In this paper, we describe new restriction analysis data for calf satellite DNA I which confirm and extend our previous results suggesting that the organisation of this satellite DNA is made up of subsatellite fractions. These variants are shown to be organised on long DNA stretches. The organisation of this satellite DNA and its functional implications is discussed.

MATHERIAL AND METHODS :

DNAs : Calf satellite DNA I (1.715 g/cm^3) was prepared from purified calf thymus DNA as already described (10).

Restriction endonucleases : All the restriction endonucleases used in this study were prepared in the laboratory with the exception of Taq I and Kpn I which were gifts of E.Southern and C.A. Reynaud respectively. The others (Alu I, Atu BI, Atu BVI, Atu CI, Ava I, Ava II, Ava III, Bal I, Bam HI, Bgl I, Bgl II, Eco RI, Eco RII, Ecl I, Hae III, Hga I, Hha I, Hind II, Hind III, Hinf I, Hpa I, Hpa II, Mbo I, Mbo II, Pst I, Pvu I, Pvu II, Sac I, Sal I, Sfa NI, Tgl I, Xba I, Xho I, Xho II, Xma I, Xma III) were prepared using procedures described in the literature. It is interesting to note that the first enzyme (Hind III) was prepared in 1972 and still works after being kept more than 8 years at -20°C.

DNA hydrolysis and gel electrophoresis : The DNA samples were digested in the buffer conditions corresponding to theconsidered restriction endonuclease. When labelled satellite DNA was used for the analysis, cold marker DNA samples (λCI, SV 40, pBR 322 or Ad 2) were added as internal controls to check the degree of digestion. Gel electrophoresis was performed either in 1.4 % agarose or 6 % acrylamide vertical gels. The DNA bands were revealed by ethidium bromide under UV light or by autoradiography.

5' end DNA labelling : Calf satellite DNA I was cut by Eco RI to completion. ^{32}P from hot ^{32}P ATP (Amersham, 3000 Ci/mMole) was transfered to the 5' DNA ends by the exchange reaction (20). After labelling the DNA was passed through Sephadex G 50. The DNA fractions were dialysed and then precipitated with alcohol and redissolved in the proper buffer.

DNA fragment lengths estimation :In most experiments, SV 40 DNA (generous gift of P.Nardeux,Villejuif) samples were digested with

Atu BI (isoschizomer of Eco RII) or with Hae III and were used as markers. In 1.4% gels, the DNA fragments from about 200 up to 1500 bp can be estimated with good approximation whilst the 6% actylamide gels were used for estimating lengths smaller than 300 bp.

RESULTS:

When cut by Eco RI, calf satellite DNA I gives a single fragment of 1460 bp in length. Using larger quantities of DNA, Gautier et al (21) revealed minor DNA fragments present in equimolar amounts. Their lengths, when added together come to approximately twice that of the repetition unit. It is not yet possible to say how these two minor fragments are related to the bulk of calf satellite DNA I. There are, however, indications that they could have a different origin or have evolved independently. For instance, Hind III site is present in the 1750 bp fragment but not in the 1050 bp fragment when the reverse is true for Pst I. No Hha I and Kpn I sites are present in either of these fragments although they are both present in the bulk of the satellite DNA as we show further on. These fragments, as well as the dimer and the trimer of the 1460 bp repeat, can be clearly detected on autoradiographs after labelling the Eco RI satellite fragments at the two 5' ends of the DNA and electrophoresis in an agarose gel (Fig.1). These minor bands are due to the alteration of a few of the Eco RI sites in the satellite DNA. Although hard to quantitate precisely, the Eco RI sites are missing in only a very few of the 100,000 copies of the repeat found in the calf haploid genome. The Mbo I site is present like Eco RI in nearly all the repeats (Fig.2). The Pst I site is present twice in each repeat (Fig.1b) and the fact that the fragments are visible under UV indicates that there is a higher proportion of missing Pst I sites than the Eco RI sites. The localisation of the two Pst I sites within the repeat (Fig.2) leads to the generation of two monomer fragments which comigrate in gels. The different Pst I sites have not been mutated to the same extent since the two labelled fragments, resulting from an alteration of one or other of the sites do not occur in the same proportions. The

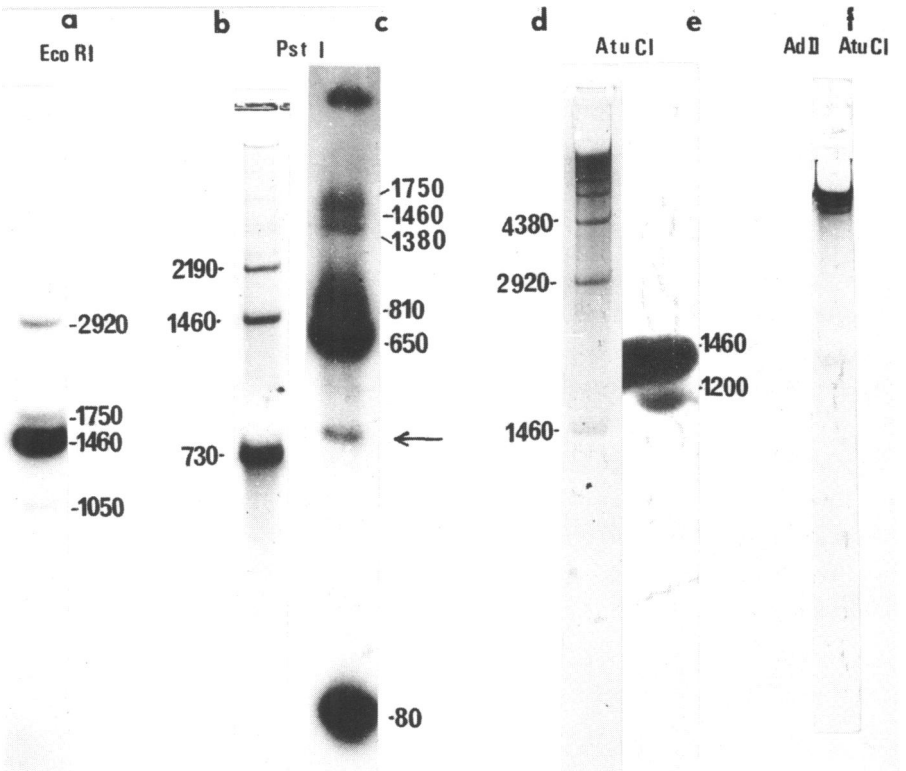

Fig. 1 :Variations in the rates of alteration of several restriction sites in calf satellite DNA I: Calf satellite DNA I was cut by Eco RI and labelled with ^{32}P as described in Material and Methods. It was electrophoresed as such in 1.4% agarose gel (a) or after complete digestion by Pst I (c) or Atu CI (e). Lane (f) shows the pattern of Ad2 DNA hydrolysed by Atu CI in the same assay as (e); it shows that the digest is complete. The cold satellite DNA after complete digestion by Pst I (b) or Atu CI (d) was electrophoresed similarly. The distances migrated by the DNA fragments cannot be compared directly as the gels were run separately. The electrophoresis in lane (c) was performed in a 6% polyacrylamide gel. One of the fragments obtained when the minor 1050 bp fragment in lane (a) is cut by Pst I is indicated by an arrow (lane c). The 260 bp fragment which should be present in lane (e) is not shown as it had run off the gel.

810 bp fragment is found much more frequently than the 1380 bp (Fig. 1c and Fig.2).

Atu CI (TGATCA) cuts calf satellite DNA I at one of the three Mbo I (GATC) sites in the repetition unit (Fig.1). Although Mbo I is present in nearly all the repetition units, Atu CI is

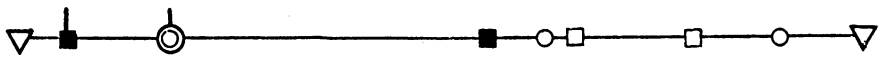

Fig 2 : The physical map of calf satellite DNA I : Map showing the location of the restriction sites : Eco RI (▽), Mbo I (○), Hind II (□), Pst I(■) and Atu CI (◯). ▌ indicates the sites which are frequently altered and presumably subject to one base change (hotspots).

missing much more frequently (Fig.1). As the Atu CI sites present are randomly distributed along the satellite DNA molecules we conclude that only two of the six nucleotides of the Atu CI recognition site are susceptible to random alteration.It is very unlikely that this alteration corresponds to base methylation, as the only modified base known to exist in calf DNA is C in the dinucleotide CpG. Therefore, this shows that at least some of the heterogeneities which are encountered in calf satellite DNA I are restricted to single base changes, as in 1.688 D. melanogaster satellite DNA.

With the possible exceptions of Hpa II, Hha I (10) and Ava I (results not shown) none of the restriction enzymes used in this study give apparent patterns of the B type like those generated in Mus musculus and guinea pig satellite DNAs (9,11). The extended ladders of DNA fragments obtained with Hpa II, Hha I and Ava I can be at least partly explained by the presence of methylated Cs in the CpG part of their recognition sites (10,20, 21).

Fig.2 shows the localisation of the Eco RI, Mbo,I,Hind II, Pst I and Atu CI sites on the physical map of DNA. The rates of random alteration of this basic pattern varies from almost zero (Eco RI, Mbo I...) to more than 50% (Atu CI) and are presumably

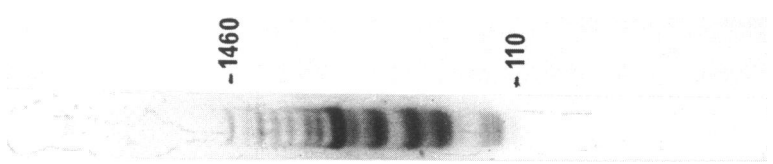

Fig. 3 :Digestion of calf satellite DNA I with Xho I : The shortest and largest fragments are indicated by their lengths in base pairs. Electrophoresis was carried out in 1.4% agarose.

caused by base changes. All these sites can therefore be considered as type A. Therefore, like 1.688 D. melanogaster satellite DNA, calf satellite DNA I contains sequences which do not seem to be affected by base changes, whilst single base changes occur more or less frequently at certain hotspots (3).

A quite different type of organisation is revealed when calf satellite DNA I is digested with a number of other enzymes. In this paper a few characteristic examples derived from the digestion of this satellite DNA with Xho I, Hha I and Kpn I are shown. The full description of these results will be published elsewhere (Pagès,Lecou and Roizès, in preparation).

Fig.3 shows the complete digest of calf satellite DNA I with Xho I. Not all the DNA fragments present in the multiband pattern observed occur in equimolar yields but the sum of the lengths of all the bands when added together comes to a multiple of 1460 bp.

An alternating distribution of altered and unaltered Xho I sites could generate a multiband pattern of this sort. However, the following examples make us think that this is not the case and that there is probably, for Xho I as well as for other restriction sites present in this satellite DNA, another explanation for this kind of pattern.

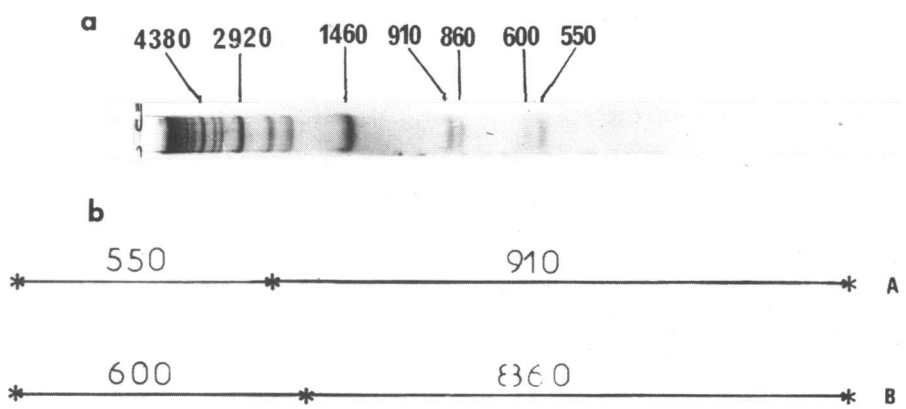

Fig. 4 : Digestion of calf satellite DNA I with Hha I : Digest of the satellite DNA with Hha I after gel electrophoresis on 1.4% agarose (a). The two possible distributions of the two Hha I sites present per repetition unit (b). Hha I (*).

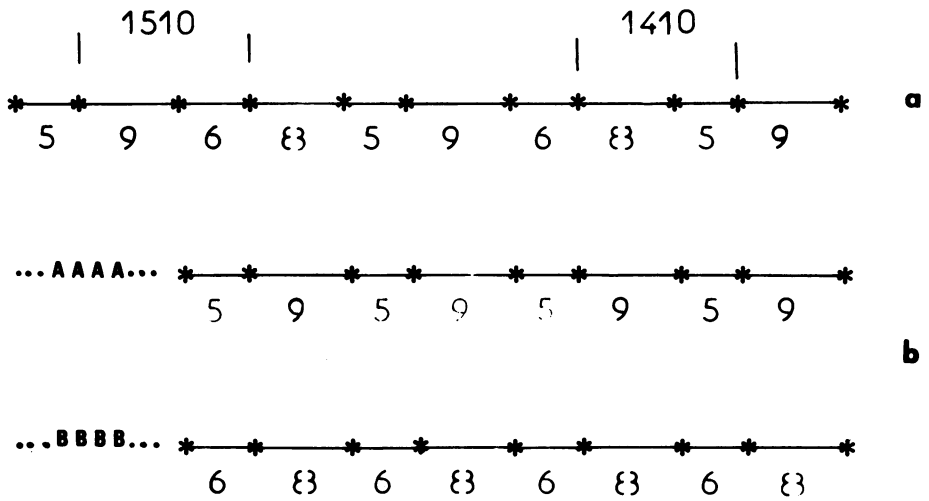

Fig. 5 a and b : Two possible types of arrangement of Hha I sites along the satellite DNA molecules. The numbers 5, 6, 8 and 9 correspond to the 550, 600, 860 and 910 bp fragments shown in Fig. 4 which are generated in the repetition units A and B when digested with Hha I (*).

Fig. 4 illustrates the distribution of the Hha I sites in calf satellite DNA I found earlier (10). The digestion pattern (fig.4a) which is partial because some of the GCGC sites are methylated, shows that none of the fragments (1410 and 1510 bp) which would have been generated if the two Hha I site distributions A and B defined in Fig.4b were mixed up as shown in Fig.5a, are actually found.

This shows that the A and B repetition units are organised in tandem on separate long stretches of satellite DNA as shown in Fig.5b. We consider this situation to arise from the organisation of calf satellite DNA I in subsatellite fractions made of different repetition units (defined by the Hha I restriction

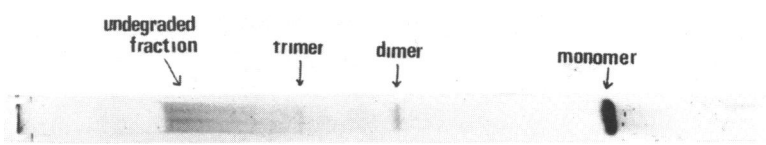

Fig. 6 :Digestion of calf satellite DNA I with Kpn I ; The digestion was shown to be complete using SV40 DNA as an internal standard.

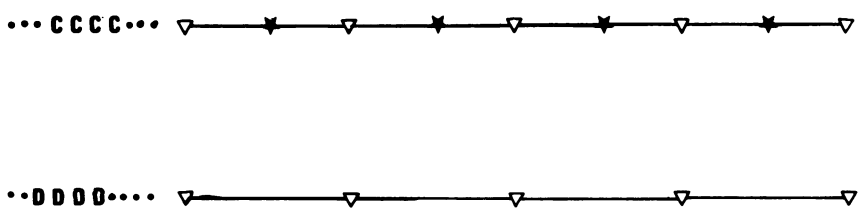

Fig.7 : The two subsatellite fractions revealed by Kpn I : Eco RI (▽), Kpn I (★).

sites) and tandemly arranged.

Another striking example of this situation is shown by the restriction pattern of calf satellite DNA I obtained with Kpn I (Fig.6).

Two fractions are clearly detected. The first contains one Kpn I site every 1460 bp with an alteration rate of about 10% giving rise to the dimer and trimer which are visible under UV. The second, which represents more than 30% of the total satellite DNA molecules, has no Kpn I site in the repetition units. It is highly unlikely that this undegraded DNA fraction is a contaminant of the satellite DNA I. Such a contaminant would have to have exactly the same density in CsCl (the calf satellite DNA I used has a unimodal profile in the analytical ultracen-

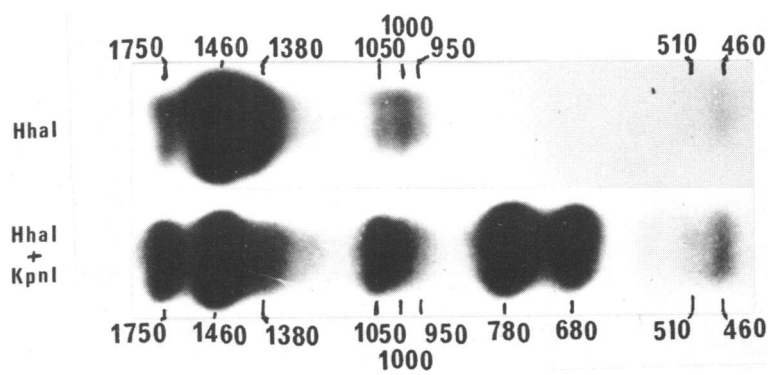

Fig. 8: The calf satellite DNA I was labelled as described in Fig. 1 after being cut by Eco RI. None of the minor DNA fragments (1750 and 1050 bp) generated by Eco RI are affected by Hha I or Kpn I. The 90 bp DNA fragment generated by Hha I which is not affected by Kpn I has run off the agarose gel.

trifuge)(10) and show the same periodicity of 1460 bp when digested with Eco RI or other restriction enzymes. So it is reasonable to suppose that the undegraded DNA seen in Fig.6 has the same basic sequence as the bulk of the satellite DNA I.Unlike Atu CI (Fig.1) where the mutated sites appear to be random, the distribution of the mutated Kpn I sites is clearly not random. This means that two sorts of satellite DNA molecules (defined by the Kpn I sites) are organised independantly as shown in Fig.7.

Two types of molecules are, therefore, detectable with both Hha I and Kpn I.

Analysis, after double digestion with Kpn I and Hha I,of the satellite DNA which had been previously cut by Eco RI and 5' ^{32}P labelled, allowed us to conclude that the A (Hha I) repetition units could either be of the C or D type (Kpn I). The same holds for the B (Hha I) repetition units (Fig. 8).

This conclusion is drawn from the fact that the 1000 and 950 bp fragments (Fig. 9) are affected by the presence of a Kpn I site in similar proportions, since their relative intensities remain approximatively the same as in the digest with Hha I alone. The important decrease in intensity of the 1380 bp fragment in the same comparative experiment leads to the same conclusion

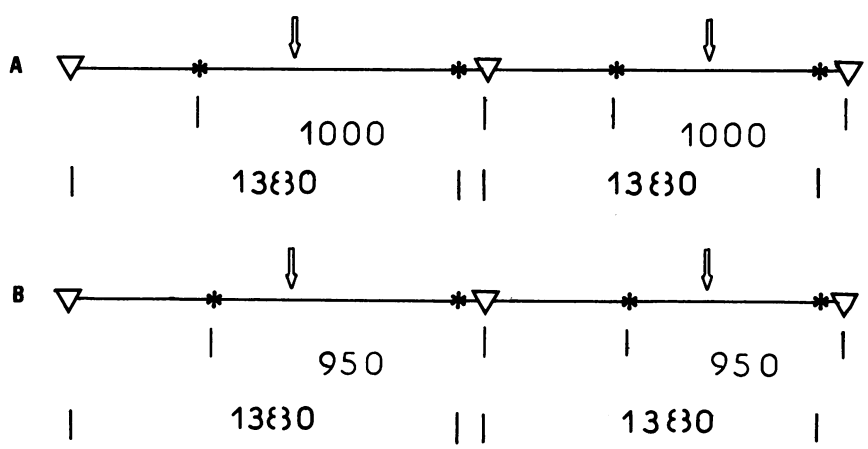

Fig. 9 : A and B represent the two subsatellite fractions revealed by Hha I. |1380| , |1000| and |950| indicate the two labelled DNA fragments which may contain a Kpn I site (⇩). Eco RI (▽), Hha I (*).

(Fig. 8). We therefore, conclude that the following four types of DNA molecules are present in calf satellite DNA I :

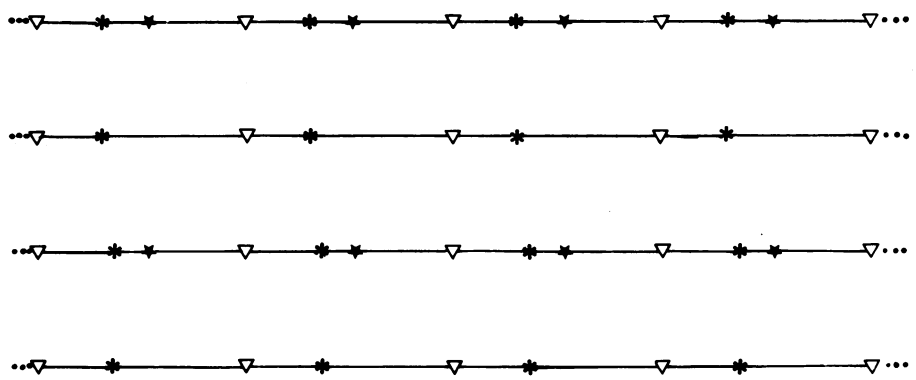

Fig. 10 : The four types of satellite DNA molecules revealed by Hha I and Kpn I : Eco RI (▽), Hha I (*), Kpn I (★). To simplify the diagram we have omitted the Hha I site which occurs in all the repeats 90 bp from each Eco RI site.

Other subsatellite fractions have been detected with other enzymes and further analysis will show if they are organised in a similar way. It is quite possible that the calf satellite DNA I is differentiated into a number of subsatellite fractions.

We do not yet know if the specific sequence differences present on different stretches of satellite DNA have a biological significance.

DISCUSSION AND CONCLUSIONS.

The results reported in this paper show the structure of calf satellite DNA I to be relatively complex.

Basicaly it consists of more than 100,000 tandemly arranged copies of a repetition unit which contain an Eco RI site every 1460 bp. The same holds for several other restriction sites such as Mbo I, Hind II, Sac I... which are distributed with the same periodicity. However, some other sites such as Atu CI and one of the two Pst I sites are present in this repetition unit arrangement with lower frequencies. The missing sites appear to result from random and presumably single base alterations. This is a situation very similar to that describe by Hsieh

and Brutlag (3) for the D.melanogaster 1.688 satellite DNA where only about 25 to 30 bases of the 359 of the repetition unit are altered in large proportions. It is not yet possible to answer the question whether or not certain regions are more stable for functionnal reasons. These results establish the alternating arrangement of constant and variable regions within the repetition units.

The localisation of a large number of restriction sites (not all shown here) within the repetition unit has not revealed a shorter repeat superposed on the 1460 bp. We, therefore, consider calf satellite DNA I as a long range periodicity satellite DNA such as the D.melanogaster (3) or the α green monkey (4) satellite DNAs. It differs, in that respect, from the numerous other satellite DNAs which have very short repeats,(e.g. in D.virilis (7), or in mouse (8)).In the mouse satellite DNA, the shorter repeat could be an ancestor of the long range periodicity repeat of 242 bp (2,8,24).

Unequal crossing over seems to be widely accepted as the most probable mechanism responsible for the appearance of intermediate fragments in the mouse satellite DNA pattern obtained with Eco RII (2,9). The same mechanism can explain the patterns obtained with Hae III with mouse satellite DNA (2,9,17) and Hind II and guinea pig satellite III DNA respectively (11). In these examples, only small proportions (1 to 2%) of the satellite DNAs are susceptible to these enzymes and give rise to ladder patterns (type B pattern) with the same basic distance between the sites as for those present in the whole satellite DNA (type A patterns).

Southern (2) argues that the extent of crossing over is unlikely to be high enough to fix and maintain the rather high degree of homology of the 10^6 copies of the mouse satellite DNA. This is not the conviction of G.Smith (23) who thinks that unequal crossing over can account for the origin and evolution of repetitious DNAs.

We do not know the mechanism which gave rise to more than 100,000 homogenous copies of 1460 bp long DNA sequence. Several mechanisms which allow the maintenance of a periodicity of 1460 bp still appear to be operating. These might have a func-

tional basis. Divergence would then occur in less critical regions. However, the difficult point to explain is how the heterogeneities of the sort shown in fig. 10 have arisen. Two explanationscome to mind. The Hha I and Kpn I sites could have reached a high degree of divergence when one or several repeats were amplified and spread along large molecules of satellite DNA. This is compatible with the high proportions of subsatellite fractions found. If this was the case the proportions found would simply reflect the degree of divergence of the corresponding restriction sites occuring at the moment of amplification. Other explanations would necessarily imply constant and very rapid rearrangements or corrections of the repeats throughout the whole genome implying large number of non sister chromatid exchanges. Either type of mechanism could lead to the fixing of the variant satellite DNA molecules in specific regions or even on different chromosomes for functional reasons such as homologous chromosome pairing during meïosis.

A number of points however have, to be cleared up before any of these ideas can be developed further. One would certainly be in better position if the lengths of the subsatellite DNA fractions were known and if one could reject or establish the possible interspersion of different blocks of these variants. It would be of interest to find out if the differences in the subsatellite fractions are restricted to single base changes or if they affect longer sequences as one would expect if they had been fixed for the functional reasons discussed above.

ACKNOWLEDGEMENTS.

The authors wish to thank Dr. B. DOD for helping to correct the English.

Contribution nº 22.

REFERENCES.

1 Botchan,M.R.(1974) Nature,251, 288-292
2 Southern,E.M. (1975) J.Mol.Biol., 94,51-69.
3 Hsieh,T.S. and Brutlag,D.L. (1979) J.Mol.Biol., 135, 465-481
4 Rosenberg,M.,Segal,S.,Kuff,E.L. and Singer,M.L. (1977) Cell, 11, 845-857.
5 Laskowski,M. (1972) Progr.Nucl.Acid.Res.Mol.Biol., 12, 161-188.

6 Gall,J.G. and Atherton,D.D. (1974) J.Mol.Biol., 85, 633-664.
7 Endow,S.A. (1977) J.Mol.Biol., 114, 441-449.
8 Biro,P.A., Carr-Brown,A., Southern,E.M. and Walker,P.M.B. (1975) J.Mol.Biol., 94, 71-86.
9 Hörz,W. and Zachau,H.G. (1977) Eur.J.Biochem., 73, 383-392.
10 Roizès,G.P. (1976) N.A.Res., 3, 2677-2696.
11 Altenburger,W.,Hörz,G. and Zachau,H.G. (1977) Eur.J.Biochem., 73, 393-400
12 Fittler,F. (1977), Eur.J.Biochem., 74, 343-352.
13 Peacock,W.J.,Lohe,A.R.,Gerlach,W.L.,Dunsmuir,P.,Dennis,E.S. and Appels,R. (1978),Cold Spring Harbor Symp. Quant.Biol., 42, 1121-1136.
14 Beauchamp,R.S., Mitchell,A.R.,Buckland,R.A. and Bostock,C.J. (1979) Chromosoma (Berl.), 71,. 153-166.
15 Cooke,H.J. and Hindley,J. (1979) N.A. Res., 6, 3177-3197.
16 Cooke,H.J. (1976) Nature, 262, 182-186.
17 Brown,S.D.M. and Dover,G.A. (1980) N.A.Res., 8, 781-792.
18 Carlson,M. and Brutlag,D.L. (1979) J.Mol.Biol., 135, 127-129.
19 Hsieh,T.S. and Brutlag,D.L. (1979) Proc.Natl.Acad.Sci., USA, 76, 726-730.
20 Berkner,K.L. and Folk,W.R. (1977) J.Biol.Chem., 252, 3176-3184.
21 Gautier,F., Bünemann,H. and Grotjahn,L. (1977) Eur.J.Biochem., 80, 175-183.
22 Taylor,H.J. (1979) Molecular Genetics,Part III, Chromosome structure, Acad.Press, 89-115.
23 Smith,G.P. (1976), Science, 191, 528-535.
24 Shmookler Reis, R.J. and Biro,P.A.(1978) J.Mol.Biol., 121, 357-374.