Volume 15 Number 7 1987 Nucleic Acids Research

Nonrandom distribution of MMTV proviral sequences in the mouse genome

Julio Salinas, Marino Zerial*, Jan Filipski, Michel Crepin[1] and Giorgio Bernardi+

Laboratoire de Génétique Moléculaire, Institut Jacques Monod, 2 Place Jussieu, 75005 Paris and [1]Institut Pasteur, 25 Rue du Docteur Roux, 75015 Paris, France

Received December 1, 1986; Revised and Accepted March 3, 1987

ABSTRACT

Integrated sequences of mouse mammary tumor virus (MMTV) have been localized in the genomes of five inbred mouse strains (Balb/c, C3H, DBA/2, A.TH, 129-SV) and one mammary tumor cell line (GR). Two major classes of MMTV sequences have been detected in mouse DNA fractions as obtained by Cs_2SO_4/BAMD (3,6-bis-(acetatomercurimethyl)dioxane) density gradient centrifugation. The first one corresponds to previously described endogenous sequences (Mtv loci), whereas the second one corresponds to endogenous sequences not previously known, and/or recently acquired; in the case of GR cells exogenous sequences may also be present in this class. The genome distribution is somewhat different for the two classes of sequences, the first one being practically only present in the lightest DNA segments of the mouse genome (GC$\simeq$ 38%); the second one being also represented in heavier segments (GC$\simeq$ 43%). This integration pattern suggests that "ancient" endogenous sequences are practically only localized in genome segments of roughly matching composition, whereas exogenous and recently acquired endogenous MMTV sequences may also be present in heavier fractions.

INTRODUCTION

A widely accepted view is that viral sequences integrate at random in the genome of infected cells. This idea is essentially based on the analysis of the restriction fragments from the host genome which carry the viral sequences, and on the primary structures of the host-virus junction sequences. The underlying, tacit assumption that the genome of the host cell is uniformly receptive for the incoming viral sequences is, however, also responsible for the random integration model.

Our investigations (see ref. 1 for a review) have led to a different view of viral integration, at least as far as

© IRL Press Limited, Oxford, England.

mammalian and avian genomes are concerned. Indeed, we have shown that the nuclear genomes of warm-blooded vertebrates is characterized by a strong compositional compartmentalization, in that they are mosaics of very long (>300 Kb) DNA segments, the isochores. These are fairly homogeneous in base composition and can be pooled into a small number of major classes distinguished by different GC levels (GC being the molar ratio of dG+dC). Such classes can be separated and used to study the genome distribution of any sequence that can be probed. This approach has revealed, among other things, that the GC levels of genes and of specific families of interspersed repeated sequences (which are believed to be mobile in the genome; 2-4) match the GC levels of the isochores in which they are located.

This genome localization approach was applied to integrated viral sequences. In the case of bovine leukemia virus (BLV) most integrated sequences were found in GC-richest DNA molecules; these represented less than 15% of the bovine genome (5). When the composition (54% GC; ref. 6) of the viral DNA became available, it appeared that there was a match in composition between viral sequences and DNA molecules from the host genome in which integration had taken place. Likewise, eight out of nine hepatitis B virus (HBV) sequences were found (7) in an isochore class of a human hepatoma cell line that represented less than 4% of the host genome. Again the GC-rich viral sequences (49% GC; ref.8) were located in the GC-richest DNA molecules from the host genome.

The results obtained with both BLV and HBV indicate that the integration of these viral sequences is not random, but "targeted" towards isochore compartments having a "matching" base composition. Needless to say, the observed "targeting" may be due to the fact that viral sequences are only stable or more stable in host genome segments of matching composition.

Here we have investigated the localization of integrated mouse mammary tumor virus (MMTV) in the genome of five inbred strains of mice (Balb/c, C3H, DBA/2, A.TH, 129/SV; the latter two were studied here for the first time) and of a mammary tumor line from GR mice. The viral integrations analyzed in the

present work are different from those previously studied (BLV and HBV) in both base composition and origin.

First of all, MMTV sequences are less rich in GC than BLV and HBV sequences. Indeed 68% of the MMTV genome (representing both LTRs, the gene env and about half of each pol and gag genes), have a GC level of 44%; moreover, different segments of the genome have the same GC level (9-13). MMTV sequences would, therefore, be expected to integrate in the "light", GC-poor, compartment of the host genome. This compartment comprises isochore classes which are a) the most abundant, since they represent about 2/3 of the mouse genome; b) the most conserved in base composition relative to the genome of cold-blooded vertebrates; and c) the poorest in the doublet CG (1).

Moreover, most integrated MMTV sequences studied here have an endogenous origin (see ref. 14), whereas both BLV and HBV sequences have an exogenous origin. In other words, MMTV sequences a) behave as stable Mendelian genes; b) correspond to integration events which, although evolutionarily recent (the integration in the germ-line having followed mouse speciation), are ancient relative to exogenous events; c) are characterized by copy numbers and chromosomal locations that are well-known for several strains, and that may vary from strain to strain, and d) are closely related in sequence with their exogenous counterparts, from which they differ because they show mutations and often are transcriptionally silent, owing to methylation (14) or other reasons.

An additional point of interest is the following: since the MMTV genome does not carry an oncogene, it most likely transforms cells by insertional mutagenesis. The genomic localization of MMTV sequences is relevant insofar as such mutagenesis can only concern host sequences located in the same isochore class as the integrated viral sequence. In this connection, it should be recalled that the integration of exogenous proviral DNA in particular genomic sites seems to be important for tumor formation. Two such sites have been identified, int-1 (15) and int-2 (16) in strains C3H and BR6, respectively. Other similar sites have been described (17-19).

MATERIAL AND METHODS

Mouse strains and cell culture. Balb/c, C3H, DBA/2, 129-SV and ATH mouse strains are maintained in our Institute. GR cells from primary explants of mammary tumor (from the Institut Pasteur, Paris, France) were grown in Dulbecco modified Eagle medium (Flow Laboratories, Puteaux, France) supplemented with 10% (v/v) fetal calf serum (Flow Laboratories S.A. Puteaux, France) in an atmosphere of 5% CO_2 saturated with water.

DNA preparation and fractionation. DNA was obtained from the livers of 6-week old female mice (Balb/c, C3H, DBA/2, 129-SV, ATH) or from 10 flasks (175 cm^3, Falcon Labware, Becton Dickinson, France) of GR cells, after treatment with trypsin/EDTA (Flow Laboratories), essentially by using the detergent method of Kay et al. (20). The average sizes of DNA molecules in these preparations were 50-70Kb as determined by electrophoretic mobility. Fractionation of DNA by preparative centrifugation in Cs_2SO_4 density gradient in the presence of BAMD, and analytical centrifugation of the preparative fractions in CsCl density gradient were carried out as previously described (21, 22).

Restriction enzyme digestion and hybridization. These were carried out as already described (21, 22). The probe used was a 0.9Kb Pst I fragment of MMTV LTR cloned in plasmid pBR 322.

RESULTS

A procedure developed in our laboratory (see ref. 1 for a review, and refs. 21 and 22) was used in order to characterize the distribution of integrated MMTV sequences in the genomes of the mouse strains and of the cell line under investigation. The procedure is based on DNA fractionation in a preparative Cs_2SO_4/BAMD density gradient. Fig. 1 shows one such experiment. The results obtained in the other experiments were essentially identical, the CsCl buoyant density of preparative fractions being the same within 0.5 mg/cm^3. Upon analytical CsCl density gradient centrifugation, the Cs_2SO_4/BAMD fractions showed buoyant densities in the expected range, 1.698-1.708 g/cm^3 (21), corresponding to 37-48% GC (23).

DNA fractions were then digested with Eco RI, a restriction

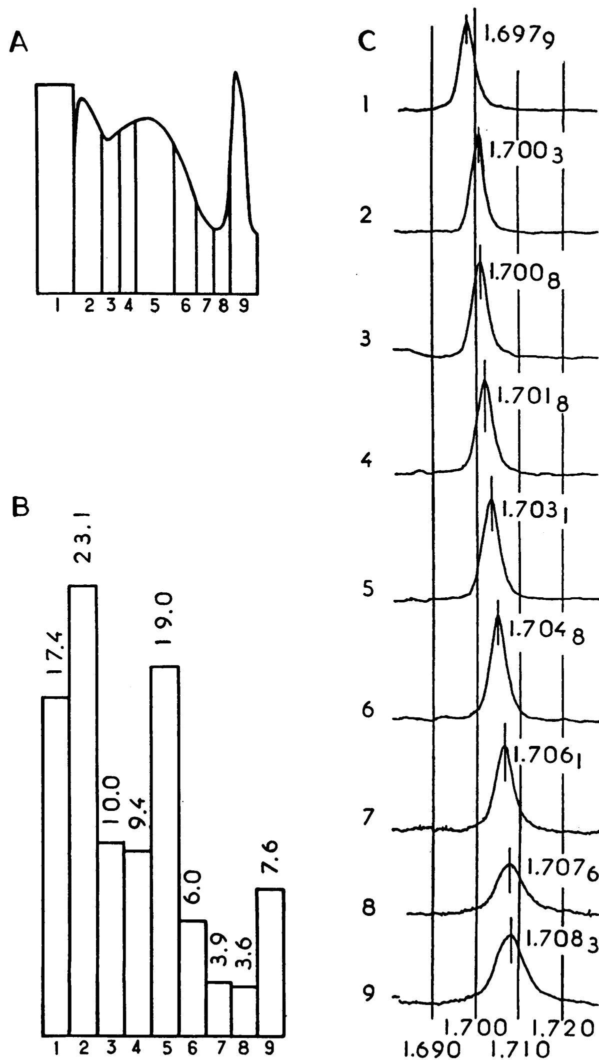

Fig. 1. Fractionation of mouse DNA by Cs_2SO_4/BAMD density gradient centrifugation. 0.5 mg of BALB/c mouse DNA was centrifuged at 30°C in 0.4 M Na_2SO_4, 20 mM $Na_2B_4O_7$, pH= 9.4, 1.58 M Cs_2SO_4 for 68 hours with a BAMD to DNA molar ratio, rf= 0.14, using a 50 Ti rotor at 36,000 rpm and a Beckman L5-50B preparative ultracentrifuge. A. Transmission profile of fractionated DNA as recorded at 253.7 nm. B. Histogram of relative amounts of DNA in the fractions. C. Analytical CsCl profiles of the fractions. The apparent absence of satellite DNA is in all likelihood due to its irreversible binding of BAMD (see ref. 1 for a discussion on this point).

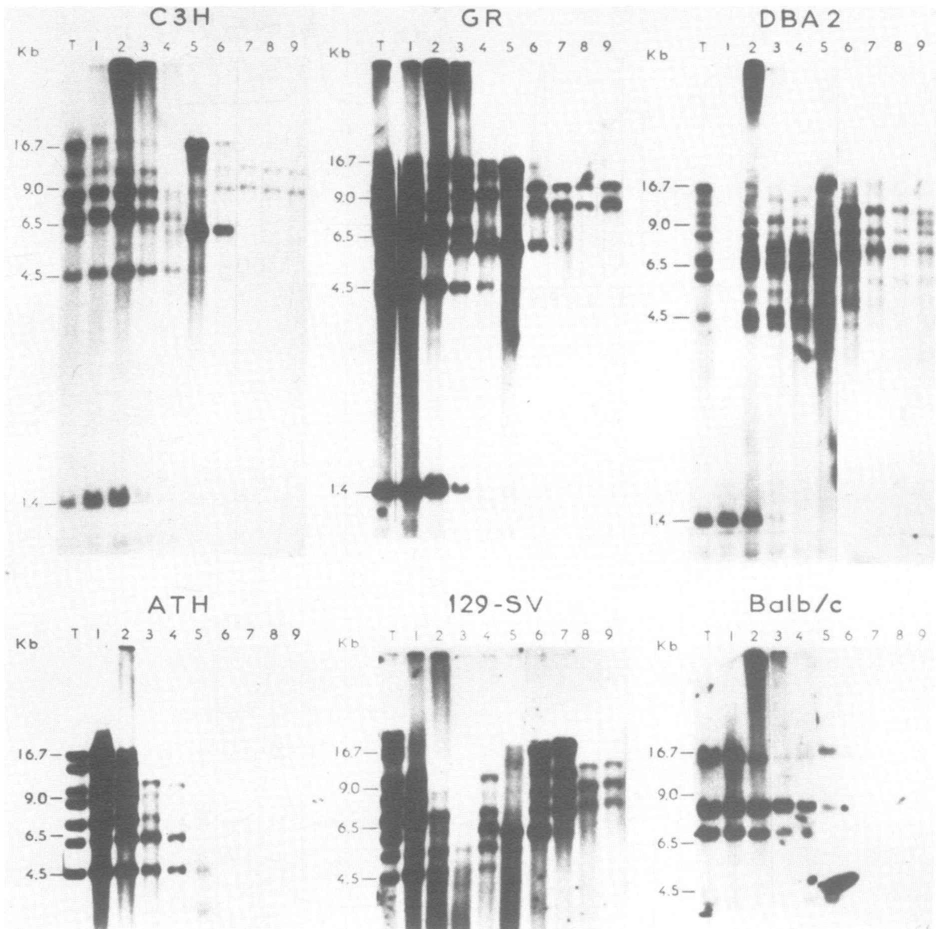

Fig. 2. Location of MMTV LTR sequences in the genomes of the mouse strains investigated in the present work. 10 g of total liver DNA and DNA from Cs_2SO_4/BAMD fractions in amounts corresponding to 100 g of total DNA were digested with EcoRI, separated by electrophoresis on 0.8 % agarose gels and transferred to nitrocellulose filters. Filters were hybridized with the ^{32}P labelled probe and subsequently washed in 3XSSC, 0.1 % SDS at 65°C. In the case of DBA/2, the lack of bands in fraction 1 was due to a poor transfer.

enzyme that cleaves only once most non-defective MMTV proviral genomes and generates, therefore, two unique cell-virus junction fragments; these vary in size according to the location of Eco RI sites in the flanking mouse sequences (24, 25). These

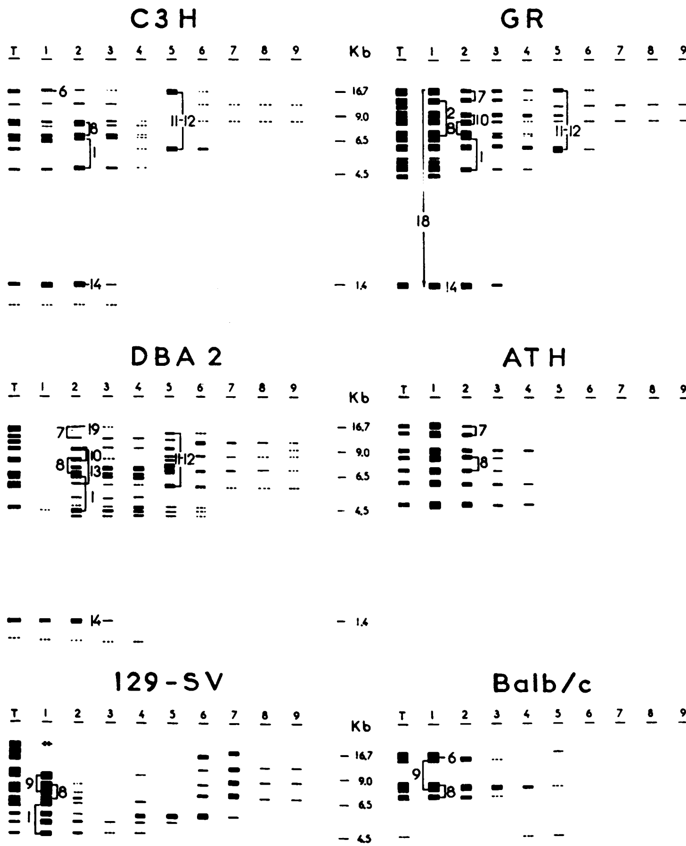

Fig. 3. Schematic representations of the distributions of MMTV endogeneous proviral sequences in fractionated mouse DNA. Some of the hybridization bands were identified with the known Mtv integration sites. Fragments originated from single loci are joined by vertical lines and labelled by loci numbers.

fragments identify a number of mouse loci called Mtv-1, 2, etc, which have been recognized in several mouse strains, and, in some cases, have also been localized chromosomally (26). Since there is no complete agreement on the Eco RI fragment sizes corresponding to different loci, we have followed here Michalides et al. (27) for Mtv-6, Mtv-8, and Mtv-9, and Gray et al. (28) for the other loci.

The localization of MMTV sequences in the Cs_2SO_4/BAMD fractions is shown in Fig. 2 and is summarized in Fig. 3.

Restriction fragments hybridizing the MMTV probe belong to two major classes and one minor class.

The first major class of fragments corresponds to known endogenous MMTV sequences. All loci previously described for strains C3H, DBA/2, Balb/c, and GR, were found (see Figs. 2 and 3; bands corresponding to these loci are numbered in Fig. 3). In the case of the latter strain, bands corresponding in size to two loci, Mtv-1 and Mtv-11-12, were detected. Since these loci were not previously reported, the corresponding bands may either correspond to the endogenous sequences just mentioned, or belong to the second class of MMTV positive fragments (see below). Moreover, the two strains which were investigated for the first time, A.TH and 129-SV, revealed the presence of loci Mtv-7, and 8 (A.TH) and Mtv-1, 8 and 9 (129-SV), respectively.

The distribution of endogenous MMTV sequences is characterized by a predominant localization in the lightest DNA fractions. Indeed, only one locus, Mtv-11-12 was found in fraction 5 (corresponding to a modal buoyant density of 1.703 g/cm^3) of strains C3H, DBA/2 and GR. All other endogenous loci were found in fractions 1 (1.698 g/cm^3) and 2 (1.700 g/cm^3). Only in a few cases, fragments showing the same mobilities as fragments from fractions 1 and 2, but much weaker intensities, were found in fraction 3 (1.701 g/cm^3; see for example Mtv-8 in strain A.TH).

The only locus present in all strains examined is Mtv-8. The second most common locus is Mtv-1, which is only absent from strains Balb/c and A.TH. If the loci found in different strains are examined in relationship with the known origins of the inbred strains studied here, one can observe that strains GR and C3H, which both originally derive from a DBA x Albino Bagg cross, show loci derived from one or the other parent. Two loci of GR, Mtv-2 and 18, are, however, not represented in the parental strains, as if they were recent acquisitions; the same can be said of loci Mtv-9 and 13, which are present in strains DBA/2 (directly derived from DBA) and Balb/c (directly derived from Albino Bagg), respectively, but are not present in either GR or C3H.

TABLE I

Mouse DNA fragments hybridizing MMTV sequences (a)

ATH	Balb/c	C3H	DBA/2	GR	129-SV
9.3 (1)	16.7 (5)	11.7 (2)	7.2 (2)	6.0 (1)	20 (1)
5.7 (1)	4.2 (5)	8.4 (2)	5.0 (2)	5.0 (1)	5.6 (1)
4.5 (1)		1.3 (2)	4.1 (2)	4.7 (1)	5.0 (1)
		11.4 (6-9)	1.3 (2)	4.1 (1)	5.8 (6)
		8.5 (6-9)	4.7 (4)	10.5 (5-9)	15.0 (6)
			10.0 (5)	8.5 (5-9)	18.0 (7)
			8.5 (5)		11.0 (7)
			8.3 (5)		8.5 (7)
			7.2 (5)		7.0 (7)
			7.1 (5)		
			7.0 (5)		
			5.6 (6)		
			11.0 (6)		
			10.0 (9)		

(a) Underlined values correspond to strong hybridization bands of the second class (hybridizatioin bands of the first class are only shown in Fig. 3); others values to very weak hybridization bands (see Text). Values in parentheses indicate the Cs_2SO_4/BAMD fractions in which bands were located.

The second major class of fragments corresponds to hybridization bands which have an intensity comparable to that of endogenous sequence bands and should, therefore, be present, like the latter, in one copy per haploid genome, but have sizes different from those described for Mtv loci. The distribution of these fragments in the gradient sometimes includes fractions showing a higher density than that of endogenous sequences described above. This class of fragments is, in all likelihood, represented by as yet undescribed endogenous sequences. These fragments are absent from two strains, C3H and Balb/c, are rare in strain A.TH, and DBA/2 whereas they are abundant in strains 129-SV and GR. In the latter case, some of the fragments may correspond to exogenous sequences.

A third minor class of fragments is characterized by extremely weak hybridization bands, which in most cases have been missed when unfractionated DNA was studied. The weak intensity of these bands witnesses their low homology with the probe used. Another feature of these bands is the widespread distribution in the gradient of at least several of them, as well as the presence of same-size fragments in different strains. Fig. 3 and Table I summarize the results obtained on the fragments belonging to this minor class as well as to the second major class. Fragments of the first class are only shown in Fig. 3.

DISCUSSION

Three main results have been obtained in the present work. The first one is that the combination of the density gradient centrifugation and restriction fragment electrophoresis allows an easier identification of Mtv loci than electrophoresis alone because of the resolution of DNA fragments which are close in molecular weight, but different in GC levels. Identification of Mtv loci by hybridization on unfractionated mouse DNA has required, so far, cloning of host genomic regions carrying MMTV sequences, segregation of MMTV hybridizing sequences in recombinant inbred mouse strains or isolation of single mouse chromosomes in cell fusion experiments (see ref. 14 for a review).

The second one concerns the genomic distribution of MMTV endogenous sequences corresponding to previously studied loci. Eleven of these twelve loci were found to be localized in fractions 1 and 2, characterized by modal buoyant densities of 1.698 and 1.700 g/cm^3, respectively, namely by GC contents of 37-39%. These fractions correspond to 17% and 23%, respectively, of total mouse DNA, namely to two thirds of the light isochore classes; (these represent about 60% of the mouse genome. Some slight differences in distribution were found among these fragments. For example, in Balb/c, locus Mtv-6 is exclusively present in fraction 1, whereas loci Mtv-8 and Mtv-9 are more widely distributed in fractions 1, 2 and 3. These differences in the spreading of hybridization bands across the gradient, were

already previously found in the localization of some genes (21, 22). They are, in all likelihood, due to the fact that in the first case the long DNA segment harboring the MMTV sequence is very homogeneous in BAMD binding sites (and composition), whereas in the second is less homogeneous. As a consequence, after random breakage of these long DNA segments, DNA molecules will be more or less narrowly distributed in the gradient.

The conclusion to be drawn here seems to be, therefore, that in the great majority of cases, endogenous MMTV sequences are localized in DNA fractions ranging in GC content from 37 to 39%. This result confirms the non-randomness of MMTV integration, and also the match between the GC levels of host DNA segments harboring the MMTV sequences, and the GC levels of viral sequences, 44% (as determined on 68% of the viral genome; 9-13). The slightly higher GC value of viral sequences compared to the isochores harboring them has been found to be the general rule for a number of coding sequences (1). It is of interest to note that such integration does not appear to concern the totality of the light isochores in the mouse genome, but only the lightest ones, so providing an independent indication for the existence (21) of subclasses in the light components of the mouse genome.

Only one of the previously described loci, Mtv-11-12, is localized in fractions exhibiting a higher buoyant density, 1.703 g/cm^3. An explanation for this situation must await for further information on the MMTV sequences present in Mtv-11-12. The main question here concerns the nature of the sequences neighboring the MMTV LTR's which are, in fact, the only sequences actually localized by the probe used.

The third result concerns MMTV sequences which do not belong to any of the loci harboring endogenous sequences identified so far and which show strong hybridization bands. Some of these sequences might well correspond to other endogenous sequences not yet identified. This might be true, in particular, for sequences localized, like endogenous sequences, in fractions 1 and 2, and showing fragments exhibiting the same size in different strains. Some of these sequences might, however, correspond to more recent integration events. This is suggested by the practical absence of these sequences in some strains,

Balb/c, C3H and A.TH. A third possibility, at least in the case of the mammary tumor cell line from GR mice, is that these sequences correspond to exogenous sequences. Interestingly, some of these sequences are present in heavy fractions (fraction 5).

The latter point is of special interest in three respects. First of all, it fits with the idea that viral integration initially may occur more randomly in the genome, integration stability preserving over time only some sequences matching in composition the host genome isochores in which they are located. Second, the presence of these MMTV sequences in heavy isochores suggests that sequences of this kind may be responsible for the activation of certain genes playing a role in the mammary tumorigenesis process. Indeed, both the int-1 domain of the C3H strain (15) and the int-2 domain of the BR6 strain (16) are high in GC, 67% and 55% respectively, and therefore, are most likely present in heavy isochores, (as almost all of the localized oncogenes are, 1). Third, heavy isochores, although much higher in CpG than the light isochores (1), are known to be undermethylated, a point certainly favoring the enhancer role of MMTV LTR's located in such regions. Indeed, it is possible that the known undermethylation of exogenous MMTV proviruses relative to the endogenous copies (29, 30) corresponds to the localization of such exogenous MMTV proviruses in the heavy isochores. In fact, it is known that provirus methylation matches the methylation of surrounding regions from the host genome (31). The high methylation level of the light isochores would then account for the lack of transcription in most tissues of the endogenous sequences which are located in that genome compartment.

A final remark is that the AT-rich MMTV sequences, which are expressed in a tissue-specific, hormone-dependent and developmentally regulated manner, are mostly localized in the AT-rich part of the mouse genome which almost only carries genes expressed in a tissue-specific way (1, 32). In contrast, the GC-rich HBV sequences, which are constitutively expressed in the host cells, are essentially localized in the GC-rich compartment of the genome (7) which carries most of the localized housekeeping genes and most of the oncogenes (1, 32). The

general relationships between genome compartimentalization and functional aspects will be discussed in more detail elsewhere.

ACKNOWLEDGEMENTS

We thank Ministerio de Educacion y Ciencia, Madrid, Spain and FEBS (J.S.) and Associazione Italiana per la Ricerca sul Cancro, Milan, Italy (M.Z.) for financial suppoort.

*Present address: EMBL, Postfach 10.2209, D-6900 Heidelberg, FRG

+To whom correspondence should be addressed

REFERENCES

1. Bernardi, G., Olofsson, B., Filipski, J., Zerial, M., Salinas, J., Cuny G., Meunier-Rotival, M. and Rodier, F. (1985) Science 228, 953-958.
2. Singer, M. (1982) Inter. Rev. Cytol. 76, 67-112.
3. Singer, M. and Skowronski, J. (1985) Trends Biochem. Sci. 10, 119-122.
4. Rogers, J. (1983) Nature 306, 113-114.
5. Kettmann, R., Meunier-Rotival, M., Cortadas, J., Cuny, G., Ghysdael, J., Mammerickx, M., Burny, A. and Bernardi, G. (1979) Proc. Natl. Acad. Sci. USA 76, 4822-4826.
6. Sagata, N., Yasunaga, T., Tsuzuku-Kawamura, J., Ohishi, K., Ogawa, Y. and Ikawa, Y. (1985) Proc. Natl. Acad. Sci. USA 82, 677-681.
7. Zerial, M., Salinas, J., Filipski, J. and Bernardi, G. (1986) Nucl. Acids Res. 14, 8373-8386.
8. Ono, Y., Onda, H., Igarashi, K., Sugino, Y., Sishioka, K. and Sasadra, R. (1983) Nucl. Acids Res. 11, 1747-1757.
9. Donehower, L.A., Huang, A.L. and Hager, G.L. (1981) J. Virol. 37, 226-238.
10. Kennedy, N., Knedlitschek, G., Groner, B., Hynes, N.E., Herrlich, P., Michalides, R. and Van Ooyen, A.J.J. (1982) Nature 295, 622-624.
11. Majors, J.E. and Varmus, H.E. (1983) J. Virol. 47, 495-504.
12. Redmond, S.M.S. and Dickson, C. (1983) EMBO J. 2, 125-131.
13. Fasel, N., Buetti, E., Firzlaff, J., Pearson, K. and Diggelmann, H. (1983) Nucl. Acids Res. 11, 6943-6955.
14. Hynes, N.E., Groner, B. and Michalides, R. (1984) Advances in Cancer Research 41, 155-184.
15. van Ooyen, A., Kwee, V. and Nusse, R. (1985) EMBO J. 4, 2905-2909.
16. Moore, R., Casey, G., Brookes, S., Dixon, M., Peters, G. and Dickson, C. (1986) EMBO J 5, 919-924.
17. Etkin, P.R. and Sarkar, N.H. (1983) J. Virol. 45, 114-123.
18. Garcia, M., Wellinger, R., Vessaz, A. and Diggelmann, H. (1986) EMBO J. 5, 127-134.
19. Gray, D.A., McGrath, C.M., Jones, R.F. and Morris, V.L. (1986) Virology 148, 360-368.
20. Kay, E.R.M., Simmons, N.S. and Dounce, A.C. (1982) J. Amer. Chem. Soc. 74, 1724-1726.

21. Salinas, J., Zerial, M., Filipski, J. and Bernardi, G. (1986) Eur. J. Biochem. 160, 469-478.
22. Zerial, M., Salinas, J., Filipski, J. and Bernardi, G. (1986) Eur. J. Biochem. 160, 479-485.
23. Cuny, G., Soriano, P., Macaya, G. and Bernardi, G. (1981) Eur. J. Biochem. 115, 227-233.
24. Shank, P.R., Cohen, J.C., Varmus, H.E., Yamamoto, K.R. and Ringold, G.M. (1978) Proc. Natl. Acad. Sci. USA 75, 2112-2116.
25. Cohen, J.C. and Varmus, H.E. (1979) Nature 278, 418-423.
26. Traina, V.L., Taylor, B.A. and Cohen, J.C. (1981) J. Virol. 40, 735-744.
27. Michalides, R., Verstraeten, R., Shen, F.W. and Hilgers, J. (1985) Virology 142, 278-290.
28. Gray, D.A., Lee Chan, E.C.M., MacInnes, I.J. and Morris, V.L. (1986) Virology 148, 237-242.
29. Cohen, J.C. (1980) Cell 19, 653-662.
30. Fanning, T.G., Vassos, A.B. and Cardiff, R.D. (1982) J. Virol. 41, 1007-1013.
31. Günzburg, W.H. and Groner, B. (1984) EMBO J. 3, 1129-1135.
32. Goldman, M.A., Holmquist, G.P., Gray, M.C., Caston, L.A., Nag, A. (1984) Science 224, 686-692.