Gene, 106 (1991) 165–172

GENE 06077

The isopycnic, compartmentalized integration of Rous sarcoma virus sequences

(Isochores; mammalian genome; hamster; provirus; base composition)

Alla Rynditch[a,b], Farida Kadi[a], Josef Geryk[c], Serguei Zoubak[a,b], Jan Svoboda[c] and Giorgio Bernardi[a]

[a] Laboratoire de Génétique Moléculaire, Institut Jacques Monod, 75005 Paris (France); [b] Institute of Molecular Biology and Genetics, Ukrainian Academy of Sciences, 252627 Kiev (U.S.S.R.) Tel. (7-044)266 3498, and [c] Institute of Molecular Genetics, Czechoslovak Academy of Sciences, 16637 Prague (Czechoslovakia) Tel. (42-2)331 2289

Received by D. Stehelin: 10 July 1991
Accepted: 20 July 1991
Received at publishers: 1 August 1991

SUMMARY

Rous sarcoma virus (RSV) can cause tumors in hamsters, which harbor complete or partially deleted RSV sequences, in their genomes. Here we have studied the localization of RSV sequences integrated into the genome of cell lines derived from six independent hamster tumors. We have found that integration occurred in the isochores richest in guanine + cytosine, of the host genome, as it had been previously observed for bovine leukemia and hepatitis B viral sequences. The integration of RSV proviral sequences is, therefore, 'isopycnic' (i.e., it takes place in host genome sequences which compositionally match the viral sequences) and compartmentalized (i.e., it occurs in a small compositional compartment of the host genome). The hamster genome compartment hosting RSV sequences precisely corresponds to a compartment of the human genome which is the most active in both transcription and recombination. The notion of a compartmentalized, isopycnic integration of RSV proviral sequences fits, therefore, with the viral integration into transcriptionally active and recombinogenic regions of the host genome observed by other authors, but is broader, in that it includes, in addition, the requirement for a compositional match between host genome sequences and expressed viral sequences.

INTRODUCTION

An important aspect of viral integration concerns the localization of integrated viral sequences in the host cell genome. The long held opinion that viral integration in the host genome is random (see Weinberg, 1980, for a review) was challenged twelve years ago by results showing that integration is 'isopycnic' (i.e., takes place in host genome sequences which match compositionally the viral sequences) and compartmentalized (i.e., occurs in small compositional compartments of the host genome). Indeed, bovine leukemia virus (BLV) sequences were found to be integrated into the DNA segments richest in % of guanine + cytosine (GC) of the bovine genome (Kettmann et al., 1979; 1980), which matched compositionally the viral sequences and represented less than 15% (most probably less than 8%) of the host genome. Similar results were obtained in the case of integration of sequences from human HBV and MMTV. In the first case, it was shown that eight out of the nine viral sequences integrated in the Alexander cell

Correspondence to: Dr. G. Bernardi, Laboratoire de Génétique Moléculaire, Institut Jacques Monod, 2 Place Jussieu, 75005 Paris (France) Tel. (33-1)43 29 58 24; Fax (33-1)44 27 79 77

Abbreviations: AAV, adeno-associated virus; BAMD, 3,6-bis(acetate mercuri-methyl)dioxane; BLV, bovine leukemia virus; bp, base pair(s); *env*, gene encoding retroviral envelope protein; *gag*, gene encoding retroviral capsid protein; GC, % of guanine + cytosine; kb, kilobase(s) or 1000 bp; HBV, human hepatitis B virus; LTR, long terminal repeat; MMTV, mouse mammary tumor virus; nt, nucleotide(s); *pol*, reverse transcriptase-encoding gene; ρ, buoyant density; RAV, Rous associated virus; RSV, Rous sarcoma virus; *src*, gene encoding the retroviral Src oncoprotein.

line were located in the GC-richest segments, which represented about 4% of the genome and also matched compositionally the viral sequences (Zerial et al., 1986a). In the second case, endogenous sequences were localized in the GC-poorest segments of the mouse genome, which again matched compositionally the viral sequences, whereas exogenous sequences showed a slightly broader distribution (Salinas et al., 1987).

These localization studies could be performed because vertebrate genomes can be fractionated according to base composition, or GC level, by preparative centrifugation in Cs_2SO_4 density gradients in the presence of sequence-specific DNA ligands. Indeed, these genomes consist of isochores, long (> 300 kb) segments of DNA characterized by a remarkable homogeneity of base composition; isochores from mammalian and avian genomes belong to a number of families that cover a very broad range of GC levels (for a review, see Bernardi, 1989).

The features of viral integrations just discussed are paralleled by those of mobile sequences, which integrate by retrotransposition. For instance, the very large numbers of *Alu* sequences (or *SINE*s) and long interspersed repeats (or *LINE*s) were shown to be localized in human and mouse genome segments matching their GC-rich (*SINE*s) or GC-poor (*LINE*s) compositions (see Bernardi, 1989). An isopycnic, compartmentalized integration appears, therefore, to be a very general property of viral and mobile sequences.

In the present work, we investigated retroviral integration in the genome of well-characterized clones of hamster tumor cells transformed by RSV type II, as present in rat XC cells; the latter retained a high degree of malignancy, as revealed by inoculation into hamsters (Svoboda et al., 1983). Hamster cells are nonpermissive for this virus and this excludes reinfection and the presence of unintegrated proviral sequences. Moreover, they lack endogenous RSV-like proviral sequences.

RESULTS AND DISCUSSION

(a) The hamster tumor cell lines

Six independent cell lines originated from different tumors containing different sets of integrated RSV-specific sequences (Svoboda et al., 1983; Svoboda and Lhoták, 1984; Pichrtová et al., 1989) were chosen: H-18 with one complete provirus; H-12 with one provirus deleted in 3′ *gag/pol* region, and H-14 harboring two such deleted proviruses; H-19, with one defective provirus, only comprising the *src* gene and LTRs; H-9, with one complete and two defective proviruses; H-42, containing amplified complete and deleted proviral copies (Table I). Fig. 1 presents a map of the RSV sequences in the genomes of the hamster tumor

TABLE I

Distribution of RSV-proviral sequences in hamster tumor cell DNA

Cell line[a]	DNA fractions with hybridizing fragments[b]				Hybridizing fragments[c] (kb)						
	Fraction (No.)	DNA (%)	ρ (g/ml)	GC (%)							
H-9K2	6	5.2	1.7091	50.1		3.1	2.9	2.4	1.5		
	7	5.2	1.7107	51.7		3.1	2.9	2.4			
H-12K1	7	7.6*	1.7082	49.3		3.1	2.9				
H-14K1	7	8.7*	1.7102	51.2		3.1	2.9				
H-18K1	7	3.6	1.7122	53.3		3.1		2.4	1.5		
	8	1.6*	1.7124	53.5							
H-19K1	5	10.8	1.7061	47.0			2.65				
	6	10.4	1.7083	49.3							
H-42K1	3	15.8	1.7021	43.0		3.1		(2.4)			(0.9)
	4	12.1	1.7029	43.8		3.1		(2.4)			(0.9)
	5	11.1	1.7045	45.4	(8.5)	3.1		2.4		1.4	(0.9)
	6	8.6	1.7059	46.8	(8.5)	3.1		2.4	1.5	1.4	
	7	8.9	1.7098	50.8		3.1		2.4	1.5	1.4	
	8	5.1	1.7138	54.5		3.1		2.4	(1.5)	(1.4)	

[a] K1 and K2 indicate the monoclonal cell lines used; these indications are not given in the text. For H9–H42 see Figs. 1 and 2.

[b] For fraction numbers see Fig. 2. % of DNA values refer to total DNA. Asterisked values include satellite and ribosomal DNA.

[c] See Figs. 2 and 3. Values in parentheses correspond to weak bands.

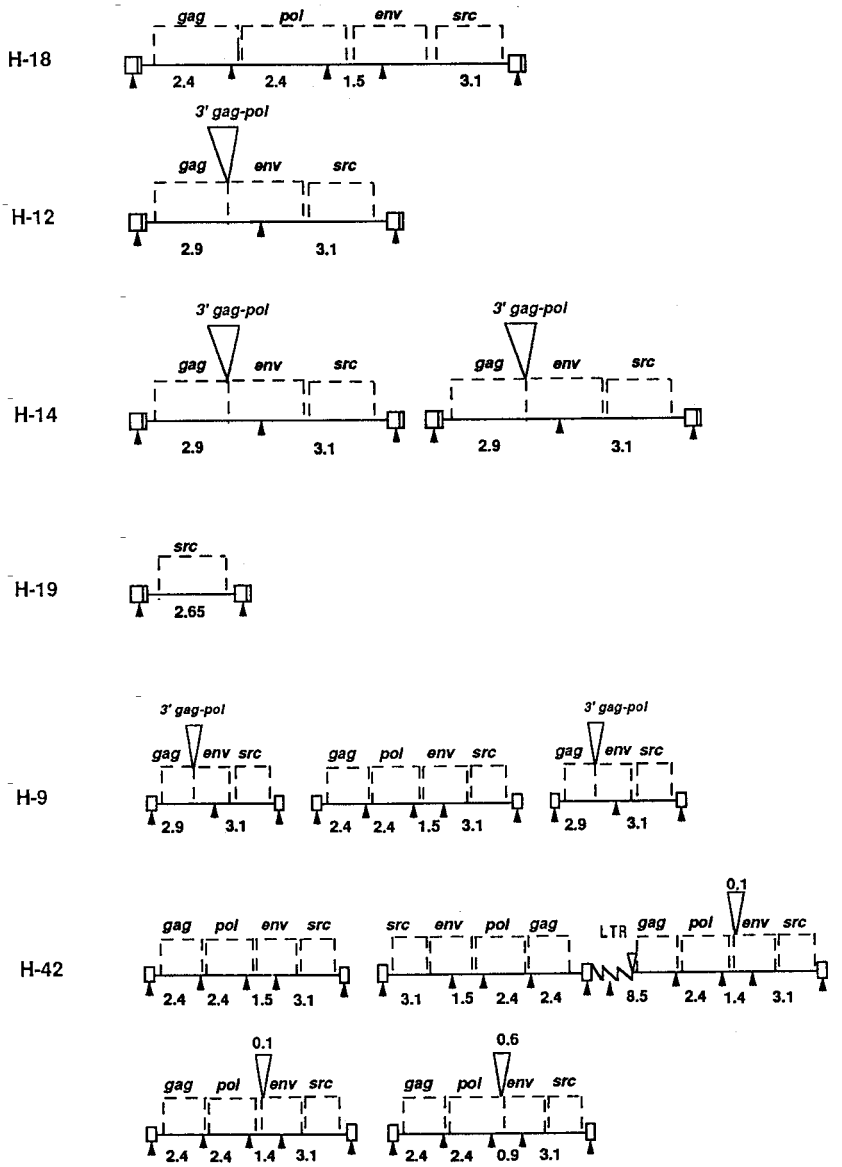

Fig. 1. Scheme of sequence arrangement of proviral copies integrated in the genomes of the hamster tumor cell lines investigated. *Eco*RI sites are indicated by upward arrowheads. Viral genes are indicated by their standard abbreviations. LTR by boxes, cellular DNA by a wavy line, proviral deletions by open arrowheads. *Eco*RI fragment sizes are given in kb.

cell lines investigated here. The data used to construct the maps are derived from the references quoted above.

The following points should be noted. DNA from the H-18 cell line exhibited the same *Eco*RI restriction fragments as the Prague RSV-C provirus type II present in rat XC cells: a 3.1-kb *env-src* fragment, two 2.4-kb fragments containing the *gag* and *pol* genes, and a 1.5-kb *pol-env* fragment. For cell lines H-9 and H-42, which harbor three and at least five proviral copies, respectively, the sequence arrangement of the copies, as shown in Fig. 1, is just probable. Transforming viruses could be rescued from all tumor lines by cell fusion with chicken fibroblasts either in the absence or presence of avian helper virus RAV-1 (Svoboda et al., 1984; Geryk et al., 1986; J.G., unpublished observations). In all five cell lines tested (H-9, H-12, H-14, H-18, and H-19), the v-*src* gene was found to be preferentially expressed, as revealed by prominent v-*src* mRNA radioactive signals in Northern blots and/or by the measurement of $pp60^{v\text{-}src}$ product. In some lines, a low degree of viral replicative gene expression and low amounts of some viral replicative gene product were detected (Svoboda et al., 1983; Geryk et al., 1984; Mazurenko et al., 1984; Grofova et al., 1985).

(b) Localization of RSV sequences in hamster tumor cells

The CsCl profiles of the DNA fractions obtained by preparative density gradient centrifugation in Cs_2SO_4/BAMD from the cell lines mentioned in section a are shown in Fig. 2.

The localization results obtained by hybridizing a viral probe on DNA fractions will be presented in an order going from the simple to the complex, namely from cell lines harboring one single complete or defective proviral copy, to cell lines containing several proviral copies (Fig. 3).

In the case of cell line H-18, the single, complete proviral copy integrated in the host genome was detected in two fractions, 7 and 8, having practically the same GC level, 53.3% and 53.5%, and representing 3.6% and 1.6% of the host genome, respectively. Fraction 8 comprised a satellite DNA which does not carry the provirus, since the provirus is mainly found in fraction 7, which does not contain the satellite. If this is taken into account, we can safely conclude that the complete provirus is present in a compositional fraction representing less than 5% of the host genome and containing the GC-richest DNA fragments. It should be noted that the fractionation of the H-18 DNA was the most favorable one in that a very large amount of DNA (63%) was present in the pellet, a situation favoring the fractionation of GC-rich fractions (and also accounting for the partial pelleting of satellite DNA; see legend to Fig. 2).

In three other cases, one or two defective proviruses were integrated. In cell line H-12, the provirus was found in a single fraction having a GC level of 49.3% and representing 7.6% of the host genome. In the case of H-14, the two defective proviral copies were found in a fraction having a GC level of 51.2% and representing 8.7% of the DNA. Again, these proviral copies were contained in less than 7–9% of host DNA (after correction for the presence of satellite DNA), this amount obviously representing an upper limit. These copies were present in fractions very rich in GC, but apparently not in the GC-richest ones, since in each case a GC-richer fraction not containing the provirus

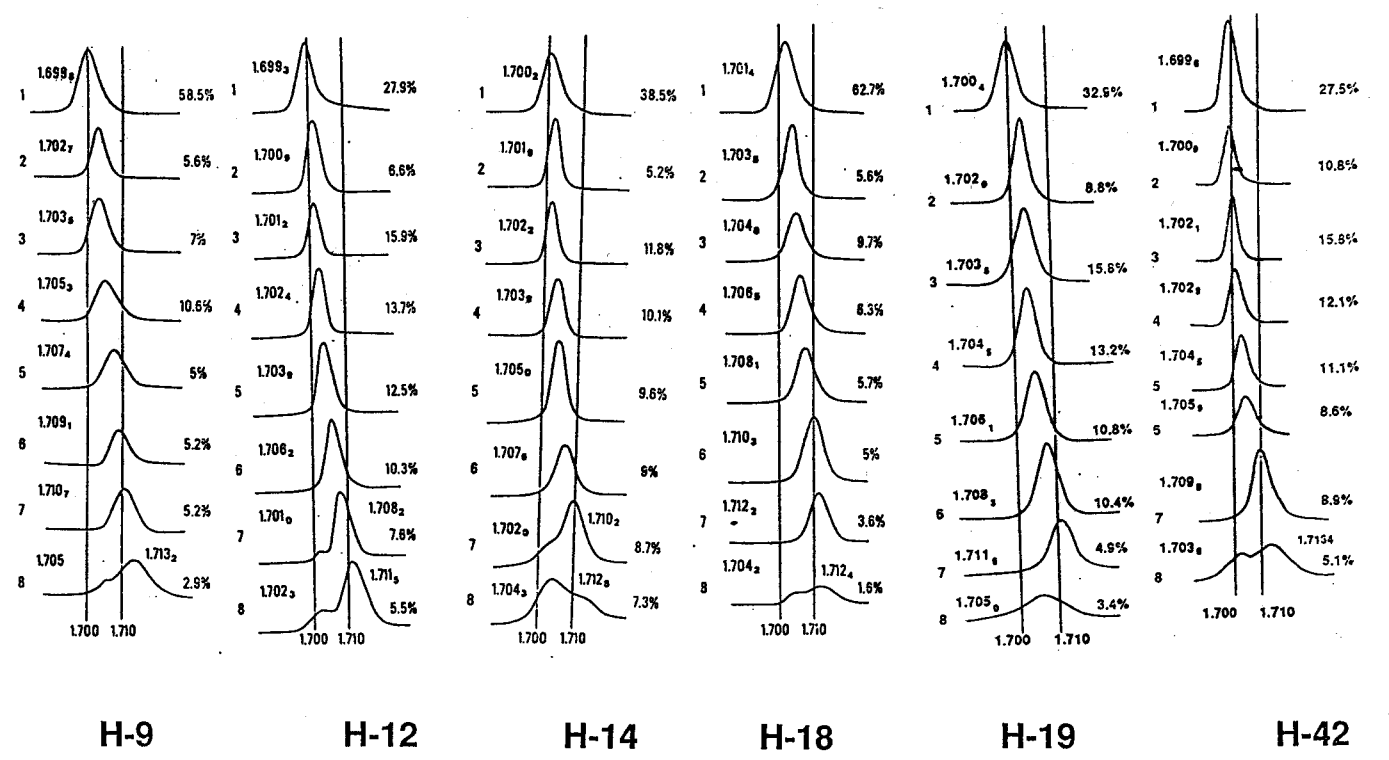

Fig. 2. Fractionation of DNAs from six hamster tumor cell lines (H9–H42) by Cs_2SO_4/BAMD density gradient centrifugation. The cell lines were derived from Syrian hamsters inoculated with chicken sarcoma tissue induced by cloned XC provirus type II rescued from XC cells by transfection. All cell lines originated from primary tumors which developed after a latent period of 5–24 months except for H-19, which derived from a renal metastasis (Svoboda et al., 1983; Pichrtovà et al., 1989). For DNA isolation, in vitro cultured monocellular clones K1 were used, except for H-9 (clone K2). DNA was isolated from hamster tumor cells either by the method of Kay et al. (1952) or that of Svoboda et al. (1983). The average sizes of DNA fragments were in the 50–100-kb range or higher, as estimated by electrophoretic mobility. DNA samples in 0.1 M Na_2SO_4/1 mM borate pH 9.4 were fractionated (Cortadas et al., 1977; Salinas et al., 1986) by preparative density gradient centrifugation at 20°C in Cs_2SO_4/BAMD. A nominal $r_f = 0.14$ (r_f is the molar ligand/nt ratio) was chosen in order to favor the resolution of GC-rich DNA fractions. Some differences in the actual r_f value led, however, to the formation of different amounts of pelleted DNA and to different qualities of separation of GC-rich fragments. A pellet (indicated as fraction 1), comprising the GC-poorest material with the largest amount of bound BAMD and seven fractions (numbered 2 to 8) were routinely obtained, and their buoyant densities were determined by analytical centrifugation in CsCl run at 25°C and 44000 rpm for 24 h. This approach also revealed the presence of a satellite DNA peak (centered at about 1.704 g/cm^3 and representing about 3% of the hamster genome) in the last fractions, which also contained ribosomal DNA.

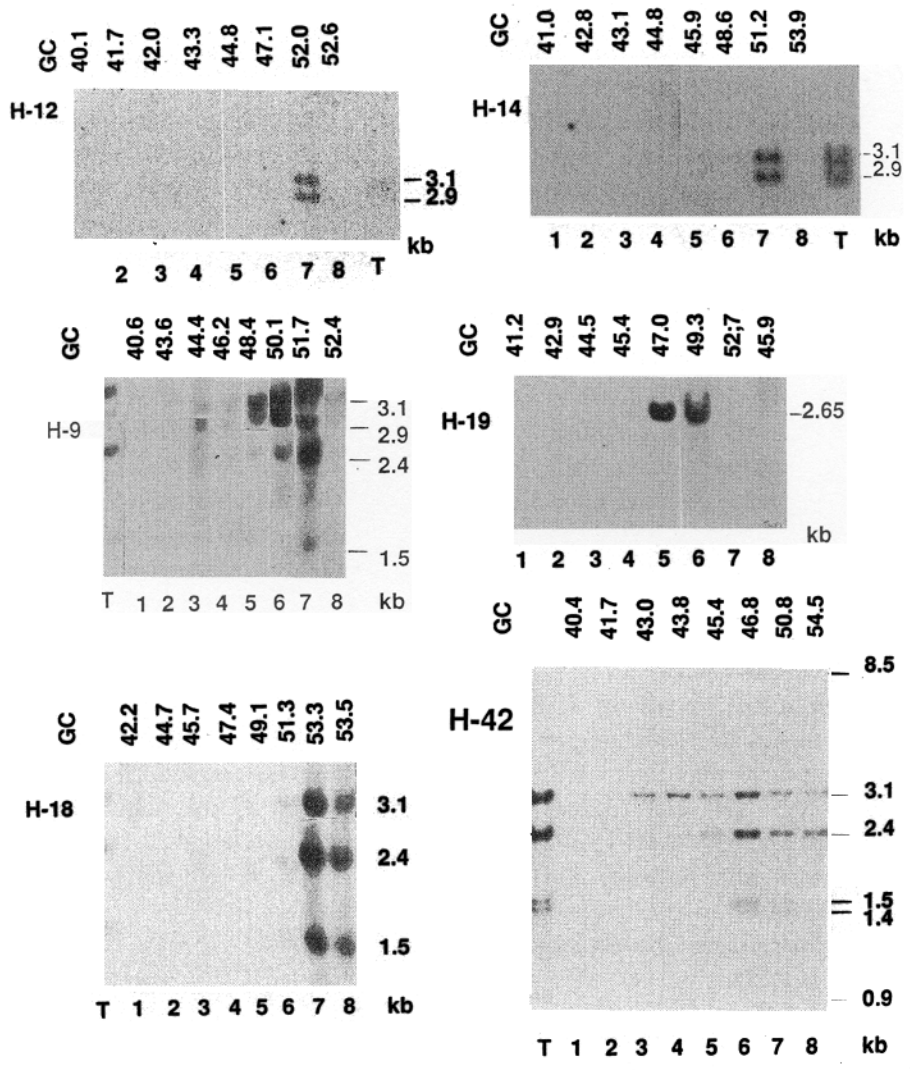

Fig. 3. Localization of RSV-specific proviral sequences in hamster tumor cell DNA fractions obtained by centrifugation in a preparative Cs_2SO_4/BAMD density gradient. After dialysis to eliminate Cs_2SO_4 and BAMD, fractions were digested with *Eco*RI and analyzed by hybridization under stringent conditions with pATV8 probe (which comprises the whole RSV genome; Guntaka and Mitsialis, 1980). Such hybridization experiments localized in DNA fractions provirus-carrying DNA segments that had up to twice the size of the DNA fragments making up the DNA samples used. This is due to the fact that RSV sequences can be located anywhere on DNA fragments, since the latter derive from the isochores harboring them by the unavoidable random enzymatic and mechanical degradation that accompanies DNA preparation. In other words, hybridization experiments localized genome segments in DNA fractions which hosted the proviral sequences and were at least 100–200 kb in size. This is to say that the GC level of the proviral sequences did not actually influence the buoyant density of the host DNA segments, that were at least 10–20 times larger than proviral genomes. It should be mentioned here that buoyant density is an additive property; viral and host sequences contribute to the buoyant densities of the DNA fragments containing them proportionally to their sizes and buoyant densities.

was obtained. In the case of H-19, the GC levels of the two fractions in which the provirus was present were lower in GC, being possibly centered at 49% (the following fraction, 52.7% GC, could not be studied because of lack of DNA).

The other cases studied were more complex because of the simultaneous presence of complete and defective copies, and because of the different distributions of those copies. In the case of H-9, a complete copy and defective copies were present in two fractions representing 10.4% of the genome and corresponding to 50.9% GC. This case was again very favorable in terms of fractionation, because of the large amount of pelleted DNA (59.5%).

Finally, in the case of H-42, complete copies were found in fractions 6–8, which ranged from 46.8% to 54.5% GC, whereas defective copies were found in these fractions, as well as in GC-poorer ones. Indeed, fractions 6, 7 and 8

exhibited all the fragments expected from complete proviruses, plus a 1.4-kb band, arising from a defective provirus carrying a 0.1-kb deletion in the 1.5-kb fragment. Fractions 3, 4 and 5 showed only some of the fragments from the complete provirus, those 3.1 and 2.4-kb in size, as well as fragments arising from other deletions: a 8.5-kb fragment (also present in fraction 6), which corresponds to internally deleted *gag*-cellular sequences, and a 0.9-kb fragment (Fig. 1). As only 1.4-kb and 3.1-fragments were found in addition to the 8.5-kb band in fraction 5 (Table I), it is likely that the proviral genome with the 8.5-kb fragment also had a 0.1-kb deletion in the *env* gene. The relatively strong 3.1-kb band in fractions 3 and 4 cannot be identified as a part of the defective provirus mentioned above, and does not correspond to any of the proviral genomes known so far. The complete proviral genomes were localized in GC-rich fractions corresponding to less than 20% of total DNA.

(c) The isopycnic integration of RSV sequences

To sum up the results presented in section **a**, RSV proviral sequences from the four cell lines containing one or two complete or defective copy(ies) were found in DNA fractions ranging from 50–53% GC and representing 5–8% of the genome, except in the case of the H-19 cell line, where RSV-containing fractions were probably close to 49% in GC and represented a larger % of the host genome. In the case of multiple proviral copies, complete copies were always found in the GC-richest fractions (47–54% GC range) and defective copies in GC-poorer ones.

If one recalls that the GC level of the RSV genome is 54.2% (with minor differences for the partially deleted copies), these data indicate an isopycnic, compartmentalized viral integration, which is very similar to those previously reported for BLV and HBV (Kettmann et al., 1979; Zerial et al., 1986).

The GC level of DNA fragments harboring proviral sequences (considering only the complete ones when multiple sequences were present), was found to be slightly lower (47–54% GC) than the GC level of the integrated proviral genomes. The compositional features of viral, transcribed sequences appear, therefore, to be similar to those of cellular transcribed sequences in general. Indeed, transcribed sequences, like mammalian exons, are 5–10% higher in GC than the flanking intergenic sequences (Aïssani and Bernardi, 1991; D'Onofrio et al., 1991), this difference tending to increase when the GC levels of genes increase (paper in preparation). A very similar situation had been previously found for BLV and HBV.

The GC-richest compositional compartment of the hamster genome in which integrated RSV sequences were found does not reach the GC level of the GC-richest compartment of the human or bovine genomes, because the hamster genome, like those of mouse and rat, belongs to a group of mammalian genomes (those of murids, cricetids and spalacids) which exhibit a narrower compositional pattern than that generally found in mammals and do not reach very high GC levels (Salinas et al., 1986; Zerial et al., 1986b; Mouchiroud et al., 1987; 1988). The GC-richest compositional compartment of the mouse and rat genomes has been shown, however, to correspond precisely to the GC-richest compartment of the human genome by (*i*) demonstrating a conservation of the order of GC levels of homologous genes and of their different codon positions (Mouchiroud et al., 1988); and (*ii*) showing that in the species under consideration (mouse/rat and man) these compartments are the richest in genes (Bernardi et al., 1988; Mouchiroud et al., 1991) and in CpG islands (Aïssani and Bernardi, 1991a,b). Since CpG islands are regularly associated with housekeeping genes (Gardiner-Garden and Frommer, 1987), this compositional compartment is also most likely to host the highest concentration of housekeeping genes, which are constitutively expressed in all cells. These facts indicate that the RSV sequences are present in a compartment of the mammalian genome which is the most transcriptionally active. This compartment is also the most recombinogenic one (Bernardi, 1989), and coincides with T-bands at the chromosomal level (paper in preparation), as previously suggested (Gardiner et al., 1990).

(d) Conclusions

The conclusion that the integration of viral sequences can occur at multiple sites of host DNA and that there are no preferred target sequences for viral insertion was based on the primary structure of DNA stretches in host-virus junctions and on analyses of restriction fragments of cellular regions flanking integrated proviral sequences (Hayward et al., 1981; Payne et al., 1982; Brown et al., 1987; Fujiwara and Mizuuchi, 1988). These lines of evidence ruled out the existence of a small number of target sites of integration, but they did not, in fact, demonstrate that integration was random at sites distributed all over the genome, even if they were generally so interpreted. In fact, random integration is not only contradicted by the previous and present results on the isochore localization of integrated viral sequences, but also by other more recent data.

Indeed, viral genomes appear to integrate preferentially into DNase I-hypersensitive chromosomal regions (Vijaya et al., 1986; Rohdewohld et al., 1987), and into transcriptionally active genome regions, as shown by studies on adenovirus integration sites (Schulz et al., 1987), the analyses of the transcriptional activity of cellular target sequences for retroviruses in mice (Scherdin et al., 1990), the preferential insertion of Moloney murine leukemia sequences in the vicinity of transcribed DNA (Mooslehner

et al., 1990), and of transcribed Rous sarcoma proviruses near CpG-rich islands in infected rat cells (Fincham and Wyke, 1991).

At this point, we can draw the major conclusion of this paper, namely that the isopycnic, compartmentalized proviral localization found here for RSV, and in previous investigations for BLV and HBV fits, in fact, with the recent findings reported by other investigators. Indeed, as mentioned in the preceding section, the compositional genome compartment in which integrated viral sequences were found is characterized by the highest transcriptional activity. It should be stressed, however, that the concept of isopycnic integration of viral sequences is broader than that of integration into transcriptionally active genome regions, because it includes an additional key requirement, that of a compositional match between the host genome sequences and the expressed viral sequences (see below). Incidentally, even if the wording 'isopycnic, compartmentalized' integration has been generally used in this paper, it is obvious that what really matters is the compositional match. In fact, because of the great compositional heterogeneity of the genome of warm-blooded vertebrates, any isopycnic integration will be a compartmentalized one.

The conclusion just presented deserves the following remarks. (*i*) The viral sequences under consideration here, like those investigated in previous studies along the same line, are expressed sequences. Nonexpressed sequences may exhibit a broader compositional range of integration sites. For instance, the only nonexpressed sequence from HBV integrated in the Alexander cell line was located in a GC-poor host genome compartment (Zerial et al., 1986a). It is not impossible that some amplified, deleted copies of RSV which were found here in lower GC fractions are also nonexpressed. In this connection, the report of Fincham and Wyke (1991) that expressed RSV sequences are close to CpG islands in the rat genomes, whereas silent copies are far from them is of great interest, because it strongly suggests that the former are localized in very GC-rich regions and the latter in GC-poorer regions. Indeed, in the mouse and rat genomes CpG islands are heavily concentrated in the GC-richest isochores (Aïssani and Bernardi, 1991a,b). (*ii*) The different localization of MMTV provirus in the GC-poorest isochores of the mouse genome matches the distribution of most tissue-specific genes and is apparently related to the hormone-dependent, tissue-specific expression of MMTV sequences. (*iii*) A final point to be considered is whether there are any preferential sites for integration in the compositional compartment in which viral sequences are found. In all likelihood, recombination-prone sites may be preferentially used in proviral integration, even under nonselective conditions (Shih et al., 1988). Along this line, preferential integration of DNA virus sequences at specific cytogenetic sites associated with constitutive chromosomal fragility was reported (Popescu et al., 1987). More recently, it has been shown that the AAV genome and the Ad5/SV40 hybrid virus sequences preferentially integrate at a specific site on human chromosome 19 and at a highly recombinogenic site on human chromosome 1, respectively (Kotin et al., 1990; Romani et al., 1990). Needless to say, these observations fit with the fact that the GC-richest compartment of the human genome is also the most recombinogenic one (Bernardi, 1989).

ACKNOWLEDGEMENTS

We wish to express our appreciation to our colleagues Jan Filipski, Gabriel Macaya, Georgette Sabeur for their help in some of the experiments, to the French Ministry for Research and Technology (MRT) for a Fellowship to A.R., and to the Association pour la Recherche sur le Cancer (ARC) for financial help.

REFERENCES

Aïssani, B. and Bernardi, G.: CpG islands: features and distribution in the genome of vertebrates. Gene 106 (1991a) 173–183.

Aïssani, B. and Bernardi, G.: CpG islands, genes and isochores in the genome of vertebrates. Gene 106 (1991b) 185–195.

Bernardi, G.: The isochore organization of the human genome. Annu. Rev. Genet. 23 (1989) 637–661.

Bernardi, G., Mouchiroud, D., Gautier, C. and Bernardi, G.: Compositional patterns in vertebrate genomes: conservation and change in evolution. J. Mol. Evol. 28 (1988) 7–18.

Brown, P.O., Bowerman, B., Varmus, H.E. and Bishop, J.M.: Correct integration of retroviral DNA in vitro. Cell 49 (1987) 347–356.

Cortadas, J., Macaya, G. and Bernardi, G.: An analysis of the bovine genome by density gradient centrifugation: fractionation in Cs_2SO_4/3,6-bis (acetato-mercuri-methyl) dioxane density gradient. Eur. J. Biochem. 76 (1977) 13–19.

D'Onofrio, G., Mouchiroud, D., Aïssani, B., Gautier, C. and Bernardi, G.: Correlations between the compositional properties of human genes, codon usage and aminoacid composition of proteins. J. Mol. Evol. 32 (1991) 504–510.

Fincham, V.J. and Wyke, J.A.: Differences between cellular integration sites of transcribed and non transcribed Rous sarcoma proviruses. J. Virol. 65 (1991) 461–463.

Fujiwara, T. and Mizuuchi, K.: Retroviral DNA integration: structure of an integration intermediate. Cell 54 (1988) 497–504.

Gardiner, K., Aïssani, B. and Bernardi, G.: A compositional map of human chromosome 21. EMBO J. 9 (1990) 1853–1858.

Gardiner-Garden, M. and Frommer, M.: CpG islands in vertebrate genomes. J. Biol. Chem. 196 (1987) 261–282.

Geryk, J., Sainerovà, H., Sovovà, V. and Svoboda, J.: Characterization of cryptovirogenic, virus-productive and helper-dependent virogenic hamster tumor cell lines. Folia Biol. 30 (1984) 152–164.

Geryk, J., Pichrtovà, J., Guntaka, R.V., Gowda, S. and Svoboda, J.: Characterization of transforming viruses rescued from a hamster tumor cell lines harbouring the v-*src* gene flanked by long terminal repeats. J. Gen. Virol. 67 (1986) 2395–2404.

Grofova, M., Bizik, J., Blahovà, S., Geryk, J. and Svoboda, J.: Identifi-

cation of transformation-specific proteins synthesized in cryptovirogenic mammalian cells. Folia Biol. 31 (1985) 152–160.

Guntaka, R.V., Rao, P.Y., Katz, R.A. and Mitsialis, S.A.: Binding of *Escherichia coli* RNA polymerase to a specific site located near the 3′ end of the avian sarcoma virus genome. Biochim. Biophys. Acta 607 (1980) 457–469.

Hayward, W.S., Neel, B.G. and Austrin, S.M.: AVL-induced lymphoid leukosis: activation of a cellular *onc* gene by promoter insertion. Nature 290 (1981) 475–480.

Kay, E.R.M., Simmons, N.S. and Dounce, A.C.: An improved preparation of sodium desoxyribonucleate. J. Am. Chem. Soc. 74 (1952) 1724–1726.

Kettmann, R., Meunier-Rotival, M., Cortadas, J., Cuny, G., Ghysdael, J., Mammerickx, M., Burny, A. and Bernardi, G.: Integration of bovine leukemia virus DNA in the bovine genome. Proc. Natl. Acad. Sci. USA 76 (1979) 4822–4826.

Kettmann, R., Cleuter, Y., Mammerickx, M., Meunier-Rotival, M., Bernardi, G., Burny, A. and Chantrenne, H.: Genomic integration of bovine leukemia provirus: comparison of persistent lymphocytosis with lymph node tumor form of enzootic bovine leukosis. Proc. Natl. Acad. Sci. USA 77 (1980) 2577–2581.

Kotin, R.M., Siniscalco, M., Samulski, R.J., Zhu, X., Hunyer, L., Laughlin, C.A., McLaughin, S., Muzyczka, N., Rocchi, M. and Berns, K.: Site-specific integration by adeno-association virus. Proc. Natl. Acad. Sci. USA 87 (1990) 212–215.

Mazurenko, N., Assanova, A., Svoboda, J. and Kisselyov, F.L.: Expression of RSV-specific proteins in RSV-transformed mammalian cells. Folia Biol. 30 (1984) 259–266.

Mooslehner, K., Karl, U. and Harbers, K.: Retroviral integration sites in transgenic Mov mice frequently map in the vicinity of transcribed DNA region. J. Virol. 64 (1990) 3056–3058.

Mouchiroud, D., Fichant, G. and Bernardi, G.: Compositional compartmentalization and gene composition in the genome of vertebrates. J. Mol. Evol. 26 (1987) 198–204.

Mouchiroud, D., Gautier, C. and Bernardi, G.: The compositional distribution of coding sequences and DNA molecules in humans and murids. J. Mol. Evol. 27 (1988) 311–320.

Mouchiroud, D., D'Onofrio, G., Aïssani, B., Macaya, G., Gautier, C. and Bernardi, G.: The distribution of genes in the human genome. Gene 100 (1991) 181–187.

Payne, G.S., Bishop, J.M. and Varmus, H.E.: Multiple arrangements of viral DNA and an activated host oncogene in bursal lymphomas. Nature 295 (1982) 209–213.

Pichrtovà, J., Geryk, J. and Svoboda, J.: Structural analysis of proviruses in additional hamster tumour cell lines transformed by provirus II rescued from XC cells and definition of new cell line harbouring amplified proviruses. Folia Biol. 35 (1989) 239–252.

Popescu, N.C., Amsbaugh, S.C. and DiPaolo, J.A.: Human papillomavirus type 18 DNA is integrated at a single chromosome site in cervical cell line SW756. J. Virol. 51 (1987) 1682–1685.

Rohdewohld, H., Weiher, H., Reik, W., Jaenisch, R. and Breindl, M.: Retrovirus integration and chromatine structure: Moloney murine leukemia proviral integration sites map near DNase I-hypersensitive sites. J. Virol. 61 (1987) 336–343.

Romani, M., De Ambrosis, A., Alhadeff, B., Purrello, M., Gluzman, Y. and Siniscalco, M.: Preferential integration of the Ad5/SV40 hybrid virus at the highly recombinogenic human chromosomal site 1p36. Gene 95 (1990) 231–241.

Salinas, J., Zerial, M., Filipski, J. and Bernardi, G.: Gene distribution and nucleotide sequence organization in the mouse genome. Eur. J. Biochem. 160 (1986) 469–478.

Salinas, J., Zerial, M., Filipski, J., Crepin, M. and Bernardi, G.: Nonrandom distribution of MMTV proviral sequences in the mouse genome. Nucleic Acids Res. 15 (1987) 3009–3022.

Schulz, M., Freiser-Rabien, U., Jessberger, R. and Doerfler, W.: Transcriptional activities of mammalian genomes at sites of recombination with foreign DNA. J. Virol. 61 (1987) 344–356.

Sherdin, V., Rhodes, K. and Brendl, M.: Transcriptionally active genome regions and prefered targets for retrovirus integration. J. Virol. 64 (1990) 907–912.

Shih, C.-C., Stoye, J.P. and Coffin, J.M.: Highly preferred targets for retrovirus integration. Cell 53 (1988) 531–537.

Svoboda, J. and Lhotàk, V.: Characterization of incomplete proviruses integrated in two helper-dependent hamster tumor cell lines. Folia Biol. 30 (1984) 165–167.

Svoboda, J., Lhotàk, V., Geryk, J., Saule, S., Raes, M.B. and Stehelin, D.: Characterization of exogenous proviral sequences in hamster tumor cell lines transformed by Rous sarcoma virus rescued from XC cells. Virology 128 (1983) 195–209.

Svoboda, J., Dvořák, M., Guntaka, R. and Geryk, J.: Transmission of (LTR, v-*src*, LTR) without recombination with a helper virus. Virology 153 (1986) 314–317.

Vijaya, S., Steffen, D.L., Kozak, C. and Robinson, H.L.: Acceptor sites for retroviral integrations map near DNase I-hypersensitive sites in chromatin. J. Virol. 60 (1986) 683–692.

Weinberg, R.A.: Integrated genomes of animal viruses. Annu. Rev. Biochem. 49 (1980) 197–226.

Zerial, M., Salinas, J., Filipski, J. and Bernardi, G.: Genomic localization of hepatitis B virus in a human hepatoma cell line. Nucleic Acids Res. 14 (1986a) 8373–8386.