Gene, 143 (1994) 155–163

GENE 07897

Regional specificity of HTLV-I proviral integration in the human genome

(Retroviruses; isochores; chromosomal bands; leukemia; myelopathy; spastic paraparesis)

S. Zoubak[a,b], J.H. Richardson[c], A. Rynditch[a,b], P. Höllsberg[d], D.A. Hafler[c,d], E. Boeri[e], A.M.L. Lever[c] and G. Bernardi[a]

[a]*Laboratoire de Génétique Moléculaire, Institut Jacques Monod, 75005 Paris, France;* [b]*Institute of Molecular Biology and Genetics, Ukrainian Academy of Sciences, 150 Zabolotnogo str., 252627 Kiev, Ukraine. Tel. (7-044) 266-3498;* [c]*University of Cambridge Clinical School, Department of Medicine, Level 5, Addenbrooke's Hospital, Hill Road, Cambridge CB2 2QQ, UK. Tel. (44-223) 336-747;* [d]*Center of Neurologic Diseases, Brigham&Women's Hospital and Harvard Medical School, 221 Longwood Avenue, Boston, MA 02115, USA; and* [e]*Laboratorio di Virologia, Centro San Luigi, Ospedale San Raffaele, Via Stamira d'Ancona 22, 20127 Milano, Italy. Tel. (39-2) 2643-7990*

Received by K.I. Berns: 5 November 1993; Accepted: 23 November 1993; Received at publishers: 21 February 1994

SUMMARY

The location of HTLV-I (human T-cell leukemia virus type 1) proviral sequences in the genome of infected human cells was explored by hybridization of a viral probe with compositional fractions of host-cell DNAs. In the twelve cases examined, HTLV-I sequences were absent from the GC-poorest 40% of the host genome (namely, from isochores that are below 39% GC). Transcriptionally inactive proviral sequences were localized in GC-poor isochores (comprised between 39% and 42–44% GC) of the human genome, which are characterized by a constant and low gene concentration. In contrast, transcriptionally active proviral sequences were found in the GC-rich and very GC-rich isochores, which are gene rich, transcriptionally and recombinationally active, and endowed with an open chromatin structure. Since GC-rich isochores are present in R′-bands and very GC-rich isochores form T-bands, these results also provide information on the location of HTLV-I proviral sequences in human chromosomes. The results obtained with HTLV-I are in agreement with the non-random, compartmentalized integration of animal retroviral sequences that had been previously observed in other viral-host systems. They provide, however, much more detailed information on the regional location of proviral sequences in the host genome and on the correlation between their transcription and their location.

INTRODUCTION

In retroviral infection, genomic viral RNA is reverse transcribed into a double-stranded DNA copy which is then integrated, as a provirus, into the host-cell genome. The widespread opinion that retroviral integration into the host genome occurs randomly (see Weinberg, 1980) was challenged 15 years ago (Kettmann et al., 1979; 1980) by an approach based upon the compositional fractionation of mammalian genomes. These genomes are mosaics of very long (>300 kb) DNA segments that are compositionally homogeneous. They are termed isochores (see Bernardi, 1989; 1993, for reviews) and comprise a small number of families, covering a very extended GC range (30–60% GC for the human genome). The large DNA fragments, 50–100 kb in size, that form standard DNA preparations and result from random physical and enzymatic degradation of chromosomal DNA during its extraction, reflect the base composition of the

Correspondence to: Dr. G. Bernardi, Laboratoire de Génétique Moléculaire, Institut Jacques Monod, 2 Place Jussieu, 75005 Paris, France. Tel. (33-1) 4329-5824; Fax (33-1) 4427-7977; e-mail: bernardi@citi2.fr

Abbreviations: ATL, adult T-cell leukemia; BAMD, bis(acetatomercurimethyl)dioxane; BLV, bovine leukemia virus; bp, base pair(s); G-, R-, T-bands, Giemsa, reverse and telomeric bands, respectively; GC, molar % of deoxyguanosine + deoxycytidine; HAM, HTLV-I-associated myelopathy; HBV, hepatitis B virus; HTLV-I, human T-cell leukemia virus type 1; kb, kilobase(s) or 1000 bp; MMTV, mouse mammalian tumor virus; nt, nucleotide(s); RSV, Rous sarcoma virus; TSP, tropical spastic paraparesis.

SSDI 0378-1119(94)00108-5

isochores from which they derive. They can be fractionated by preparative equilibrium centrifugation in Cs_2SO_4 density gradients in the presence of sequence-specific DNA ligands. In the human genome, two GC-poor isochore families, L1 and L2, form 62% of the genome, and three GC-rich isochore families, H1, H2 and H3, make up 20%, 10% and 3–5% of the genome, respectively, the rest being represented by satellite and ribosomal DNA. Hybridization of the integrated viral sequences in compositional fractions of host-cell DNA with appropriate probes showed that, in all cases investigated so far, integration of transcribed viral sequences is 'compartmentalized' and 'isopycnic', i.e., it takes place in the isochore families that show a compositional match with the viral sequences. This compartmentalized integration was found for BLV (Kettmann et al., 1979; 1980), HBV (Zerial et al., 1986a) and RSV (Rynditch et al., 1991) sequences, which all integrate into GC-rich regions of the host genome, and for MMTV sequences (Salinas et al., 1987), which integrate into GC-poor regions.

These results can be put in a more general context by two findings. Firstly the distribution of genes in mammalian genomes is strikingly non-uniform (Bernardi et al., 1985; Mouchiroud et al., 1991). The majority of genes, and housekeeping genes in particular (Aïssani and Bernardi, 1991), are located in the GC-rich isochore families that only represent one third of the genome. Secondly, retroviral genomes fall into two compositional classes (Zoubak et al., 1992): a large GC-rich class, comprising all oncoviruses (like BLV, HTLV-I and RSV), except B-type and some D-type oncoviruses; and a minor GC-poor class formed by lentiviruses, spumaviruses, B-type oncoviruses (like MMTV) and some D-type oncoviruses.

HTLV-I is the causative agent of adult T-cell leukemia (ATL) and HTLV-I-associated myelopathy/tropical spastic paraparesis (HAM/TSP). The integration of this GC-rich retrovirus is of interest because it would be expected to occur in the compartments of the host genome that are characterized by high concentrations of genes. These compartments correspond to R-bands of metaphase chromosomes, namely to bands that show early replication and high levels of transcription and recombination (see Bernardi, 1989).

As far as the integration of the HTLV-I proviral genome into chromosomal DNA of leukemic cells is concerned, it should be recalled (*i*) that the sites of integration have been shown to differ from case to case (Seiki et al., 1984) and to be located in different chromosomal regions; (*ii*) that polyclonal proviral integration was observed both in asymptomatic HTLV-I carriers and in patients with HAM/TSP (Greenberg et al., 1989); however, clonal integration has been reported in patients with ATL, indicating clonal expansion of infected cells (Ohshima et al., 1991; Furukawa et al., 1992); (*iii*) that integration of the proviral genome has been thought to be a prerequisite for the development of ATL (Yoshida et al., 1984); and (*iv*) that both complete and defective proviruses have been detected in cell lines of leukemic origin (Salahuddin et al., 1983; 1984; Yoshida et al., 1985; Aldovini et al., 1986; Kobayashi et al., 1986) and in cells directly isolated from patients (Hiramatsu et al., 1986; Ohshima et al., 1991). If these cells contain only defective proviral forms, the pX region that encodes regulatory proteins (see, for references, Franchini et al., 1993) was usually preserved (Ohshima et al., 1991).

RESULTS AND DISCUSSION

(a) Hybridization results

HTLV-I sequence localization was studied in five HTLV-I immortalised cell lines and seven T-cell clones from patients with HAM/TSP or HTLV-I-associated polymyositis. The transcriptional activity of the proviruses in these cells and details of virus production are presented in Table I.

TABLE I

Properties of cell lines and clones

Cell type or clone examined	Virus production	Proviral transcription
HTLV-I-transformed cell lines[a]		
C10/MJ	+	+
C8166	−	+
Hut-102	+	+
C91/PL	+	+
MT-2	+	+
T-cell clones[b]		
Ph1C		
Du4	+	+
Mu16	+	+
Du43	−	−
Mu40	−	−
Cp9		
Cp10		

High-M_r DNA samples from the cells used (> 50 kb as estimated by gel electrophoresis using whole λ DNA as a size marker) were fractionated by preparative Cs_2SO_4 density-gradient centrifugation in the presence of bis(acetatomercurimethyl)dioxane, BAMD, a sequence-specific DNA ligand (see Bernardi, 1989). Usually, seven fractions covering the 38–54% GC-range and a pellet (fraction 1) containing the most GC-poor material were prepared and analyzed in CsCl density gradients. The results of a typical experiment are presented in Fig. 1. The relative amounts of DNA found in individual fractions, as obtained from the different cells investigated, are displayed in Fig. 2.

The presence of proviral sequences in the DNA fractions was detected by Southern blot analysis on restriction digests obtained with *Eco*RI, an enzyme that does not cut the proviral sequence, so that the number of different-size bands reflects, as a rule, the number of proviral copies.

The autoradiographs of four Southern blot analyses showing hybridization patterns characterized by different features, such as different distributions in compositional fractions, different intensities and spreads of same-size bands, faint bands and smears are presented in Fig. 3. The results of all analyses are schematically presented in Fig. 4. They will be commented below.

(*i*) Since the starting population of molecules carrying a given proviral sequence is the result of random breakage occurring during DNA extraction, proviral sequences may be positioned at different locations on DNA molecules. For this reason, hybridization explores a chromosomal region almost twice as large as the average size of the DNA molecules investigated (in fact, 1.7-times as large, assuming a Poisson distribution of the starting DNA molecules). Since molecular sizes of DNA preparations were always higher than 50 kb, regions larger than 100 kb around the integrated sequence were explored in all experiments. Thus the fragments carrying the provirus are more than 6-times as large as a complete proviral sequence, 9 kb, and up to 12 or more times larger in the case of defective proviruses (HTLV-I positive fragments may be as small as 4.5–4.6 kb; see Fig. 4). This means that the contribution of composition of the provirus (53.9% GC) to the GC level of the provirus-containing fragments can be neglected. Indeed, in the most unfavorable case of a full-size provirus carried by host DNA fragments of only 50 kb and derived from the lowest GC (39% GC) isochores, the DNA fragments would be increased in GC from 39% to 41% by the provirus.

(*ii*) In most cases, a given fragment containing a proviral sequence was present in more than one fraction because Brownian diffusion of DNA molecules in the 50–100 kb size range leads to a relatively broad distribu-

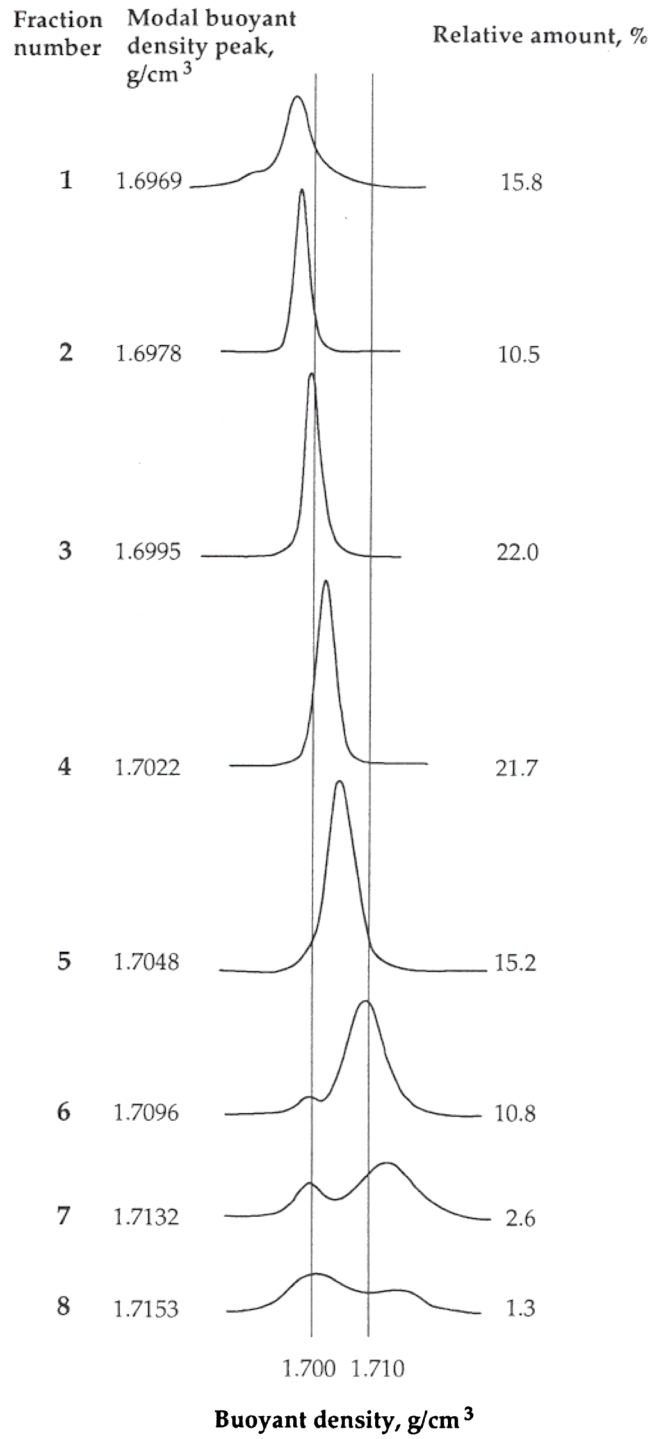

Fig. 1. Buoyant density profiles as determined by analytical centrifugation in CsCl on fractions 1–8 of MT-2 DNA (see Table I) obtained by Cs_2SO_4 /BAMD density gradient centrifugation. A molar ligand/nt ratio (R_f) of 0.12 was used. In some experiments the actual R_f was, however, different from the nominal value of 0.12, possibly because of the presence of contaminating traces of proteins and/or RNA that bind BAMD. This accounts for the fact that the pelleted DNA (which is characterized by the lowest GC level and by the highest BAMD binding) varied in amount in different experiments (see Fig. 2). As a consequence, in addition to the arbitrary cutting of the gradient, a given fraction does not generally correspond to a given GC level. A satellite DNA peak ($\rho = 1.700$ g/cm^2) is visible in fractions 6–8. The modal buoyant densities of main-band DNA and the relative amounts of DNA in the fractions are indicated.

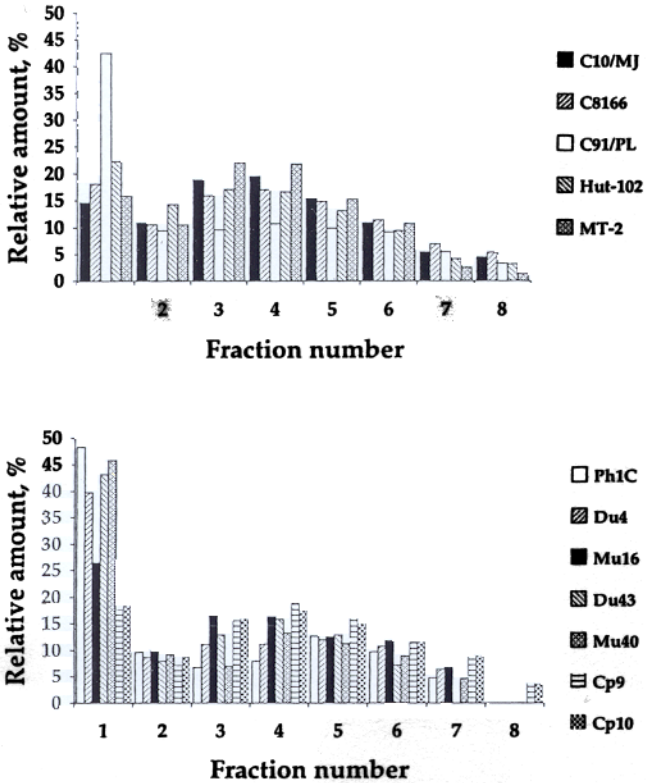

Fig. 2. Relative amounts of DNA in compositional fractions from transformed cell lines (top) and T-cell clones (bottom) described in Table I.

tion in the gradient. In some special cases, the compositional heterogeneity of the chromosomal region under study may also contribute to band spreading, as when the integrated sequence is near an isochore border characterized by a strong compositional discontinuity. A peak in the distribution of any given hybridizing fragment could always be identified by comparing band intensities of consecutive fractions and by taking into account the signal enrichments (see legend to Fig. 3).

(*iii*) The intensities of hybridization bands in DNA from established cell lines were found to correlate with the proportion of HTLV-I harboring cells in the sample (Manzari et al., 1983). Some faint bands not described so far were seen in unfractionated DNA (see, for example the 6.7-kb band of Hut-102, Fig. 4), but most of them were only detected in the DNA fractions because of the enrichment of the hybridization signals mentioned above. Such non-stoichiometric faint bands could correspond to viral sequences originating from small populations of cells that diverged from the main population because of deletions and/or rearrangements of cellular and/or viral DNA. It should be stressed that faint bands present in the last fractions correspond to extremely small populations, because of the strong enrichment of HTLV-I hybridization signals (up to 200 times; see legend to Fig. 3) occurring in these fractions.

(*iv*) In established cell lines the number of integrated proviral copies varied from two to eight, as estimated by the number of bands corresponding to different-size fragments, which may be larger or smaller than a complete provirus, 9 kb (see Fig. 4). The higher provirus copy number in immortalized cells (MT-2, C91/PL, C8166, C10/MJ, Hut-102) is a consistent finding (Markham et al., 1983; Franchini et al., 1984; Watanabe et al., 1984; Kobayashi et al., 1986; Aldovini et al., 1986; Berneman et al., 1992) and may reflect reinsertion of the provirus during cultivation (Franchini et al., 1984). In this connection, it should be mentioned that the number of proviral copies reported for Hut-102 varied from three (Franchini et al., 1984) to nine (Watanabe et al., 1984).

(*v*) The T-cell clones showed a single hybridization band for Du4 and Mu40 [if the low-M_r faint bands present in the last fractions are neglected for the reasons given in (*iii*) above]. Ph1C showed three proviral copies (neglecting again a faint band present in the last two fractions) with a size that might correspond to the complete provirus. DNA of Mu16 cells shows a smear in fraction 4 (and, to a lesser extent, in fraction 3) and two bands in fractions 6–7. This smear can be explained (Prunell et al., 1977) by the random degradation of the provirus-carrying fragment, which has a very high molecular size (>23 kb) relative to that of the host DNA preparation. The alternative explanation, namely the presence of integrated proviruses having different integration sites which belong to a polyclonal population, is most unlikely in view of the presence of the smear only in a narrow compositional range of the host genome. Cp9 and Cp10 showed three and two integrated sequences, respectively.

(b) Localization of HTLV-I sequences in the host-cell genome and proviral transcription

HTLV-I sequences are distributed in a 39–54% GC range of the host-cell genome (see Fig. 5). This means that no HTLV-I sequences are found in the GC-poor 40% of the human genome (<39% GC), all integrated sequences being present in the 60% GC-rich part of the genome. Although the three proviruses of the cell clone Ph1C are present in the lowest GC fraction (fraction 1), the spill-over hybridization in fractions 2 and 3 suggests that the proviruses are close to the upper extreme of GC content within this fraction. The lack of hybridization signals in fraction 8 (see Figs. 1 and 4) indicates that HTLV-I does not integrate in either ribosomal or satellite DNA. A first conclusion is, therefore, that 40% of the human genome, the GC-poorest part of it (as well as satellite and ribosomal DNAs), never host HTLV-I sequences in the cells investigated.

At this point, questions should be asked about the distribution of proviral sequences in the human genome and about the possible correlations between the host genome

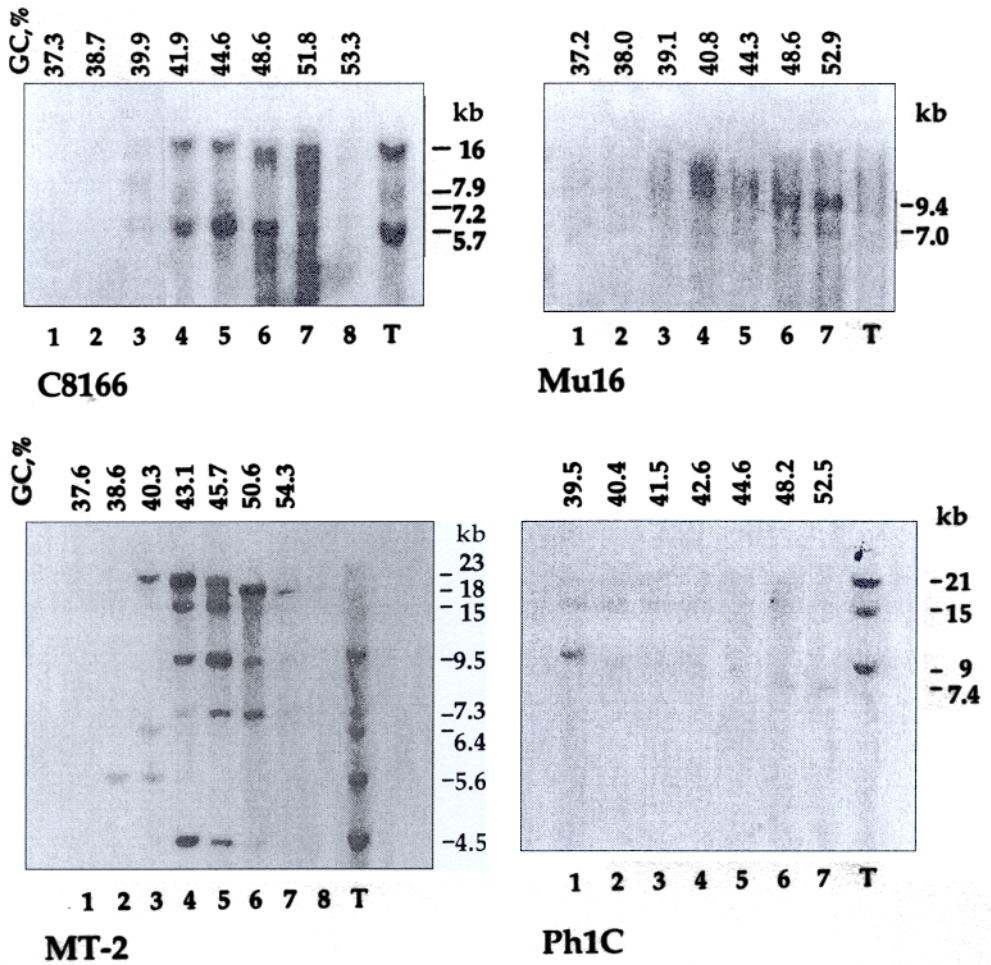

Fig. 3. Localization of HTLV-I proviral copies in DNA fractions from cell lines MT-2, C8166, Ph1C and Mu 16. Total DNA (8 µg) and equal amounts (3 µg) of DNA from each fraction were digested with *Eco*RI, electrophoresed on 0.8% agarose gels and hybridized under stringent conditions (see Rynditch et al., 1991) with the insert of plasmid pMT-2, which contains a complete HTLV-I provirus with a single LTR (Clarke et al., 1983). Loads of DNA fraction were, therefore, not proportional to their relative amounts in human DNA, but were larger for almost all fractions. As a consequence, hybridization signals were generally enriched relative to signals from unfractionated DNA. This enrichment was particulary strong for the last fractions. As it can be judged from the data of Fig. 2, while signal enrichment was practically absent for the most abundant pellets, corresponding to 50% of DNA (3 µg/8 µg/0.5=0.75), it reached high levels for the last fractions, corresponding to less than 5% of DNA (at least 3 µg/8 µg/0.05=7.5). Estimates of total signal enrichment for individual fractions should, however, not simply be calculated from Figs. 1 and 2, but should also take into account the relative intensity of the bands. If this is done, signal enrichment can reach a 50-fold level (relative to 'normal' bands) for the faint bands located in the last fractions. In fact, enrichment for those fractions is even higher, up to 100–200-fold, because they are largely made up of satellite and ribosomal DNA (see Fig. 1) which do not host HTLV-I sequences (see section b). Needless to say, the faint bands under consideration here are not those corresponding to the tails of distribution of 'normal' bands, but bands corresponding to independent integrates.

regions in which integration was found and the transcription of HTLV-I sequences. While some transformed cell lines and T-cell clones investigated contain proviral sequences integrated in both high- and low-GC regions (both of these regions being in the GC-rich 60% of the genome), several of them contain integrated HTLV-I sequences in only one of the two regions, and they may be useful starting points for answering this question.

The transcriptionally active HTLV-I sequences of C10/MJ and C91/PL (Aldovini et al., 1986; Markham et al., 1983) were only found to be integrated in high (>44%) GC-rich regions. It should be noted that the hybridization bands present in the abundant pellet (fraction 1) representing about 50% of C91/PL DNA are the result of a contamination with high-GC fractions during preparative gradient fractionation, as demonstrated by the identity of the bands in this fraction with those centered in fraction 6, as well as by the presence (data not shown) of a GC-rich component (approx. 47.2% GC) in the pellet fraction that has a modal buoyant density corresponding to 37.9% GC (this contamination may occur because the tip of the needle used to collect the fractions from the centrifuge tube is positioned just above the pellet and traces of the last GC-richest fractions may be left behind).

On the other hand, the transcriptionally inactive HTLV-I sequences in Ph1C, Du43 and Mu40 (J.H.R., P.H., D.A.H. and A.M.L.L., data not shown) were only

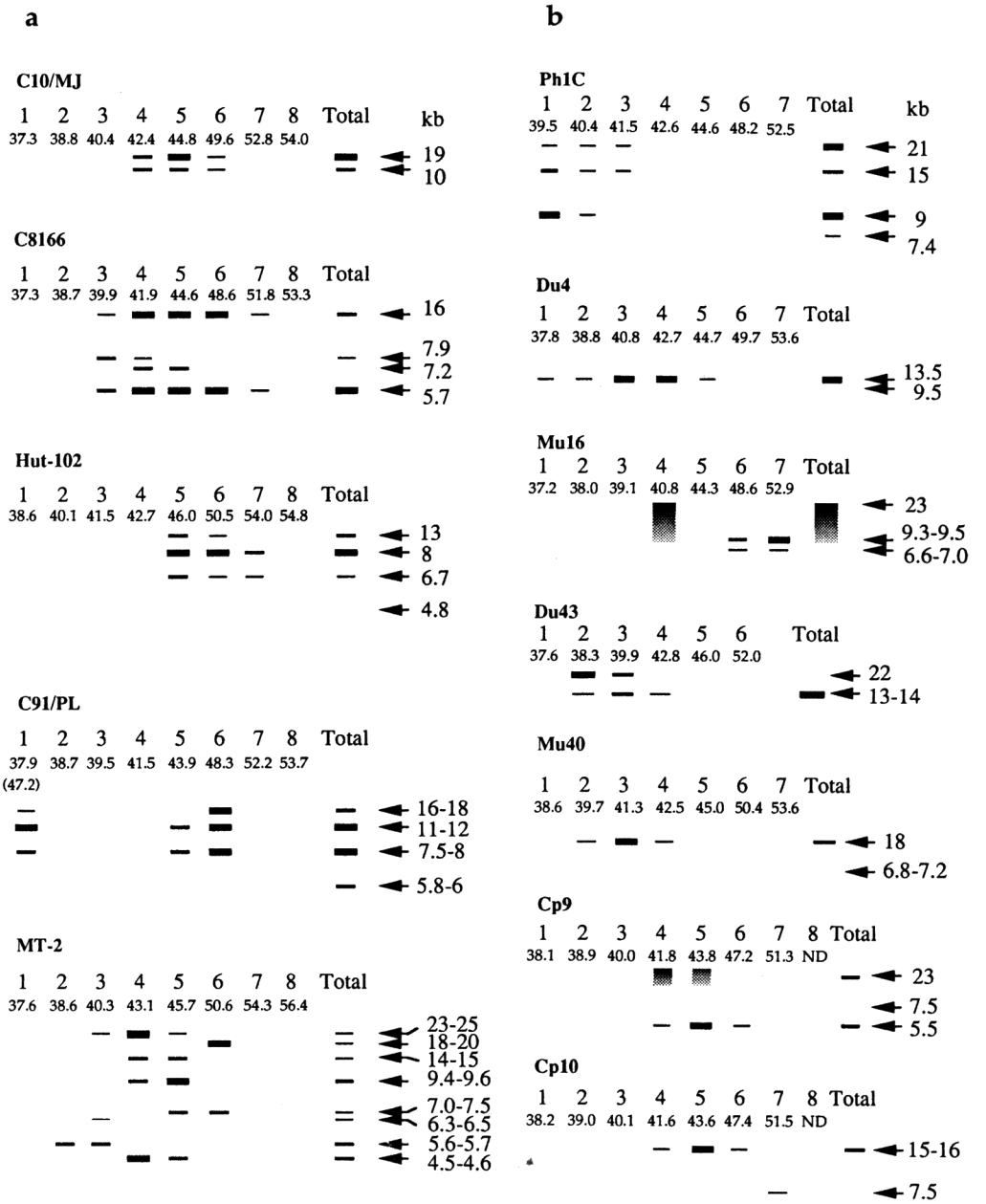

Fig. 4. Scheme of the localization of HTLV-I proviral sequences in the compositional fractions and total DNA from the genome of human cell lines (**a**) and cell clones (**b**) as obtained after centrifugation in a preparative Cs_2SO_4/BAMD density gradient. In each pattern, the first line presents the fraction number, the second line the %GC. Relative intensities of hybridization bands are indicated by lines of three different thicknesses; smears are indicated by unevenly shaded boxes. ND, buoyant density not determined.

found in the low (<42%) GC regions. (The faint bands in the high-GC region of Ph1C and Mu40 are neglected because they correspond to extremely small populations and have, in addition, a low M_r). These results indicate that, in these clear-cut cases, integrations in high (>44%) and low (<42%) GC regions correspond to transcriptional activity and inactivity, respectively.

The compositional border between these regions is likely to be around 43% GC, as suggested by the fact that the 42–44% host genome regions with GC levels of 42–44% contain integrates which are (Du4) or are not (Cp9, Cp10) transcribed. Indeed, the transcribed sequences of Du4 (Wucherpfennig et al., 1992) appear to derive from the fragments centered in fractions 3 and 4 and corresponding to a GC value slightly below 43%. The broad distribution from fraction 1 to fraction 5 might be due to compositional heterogeneity in the host chromosomal region of integration (a possibility already mentioned in section **a**). An alternative, more likely, explanation is that the signal found in fraction 1 is just due to a contamination of the pellet (representing 40% of DNA), as in the case of C91/PL. In the case of Cp9, the

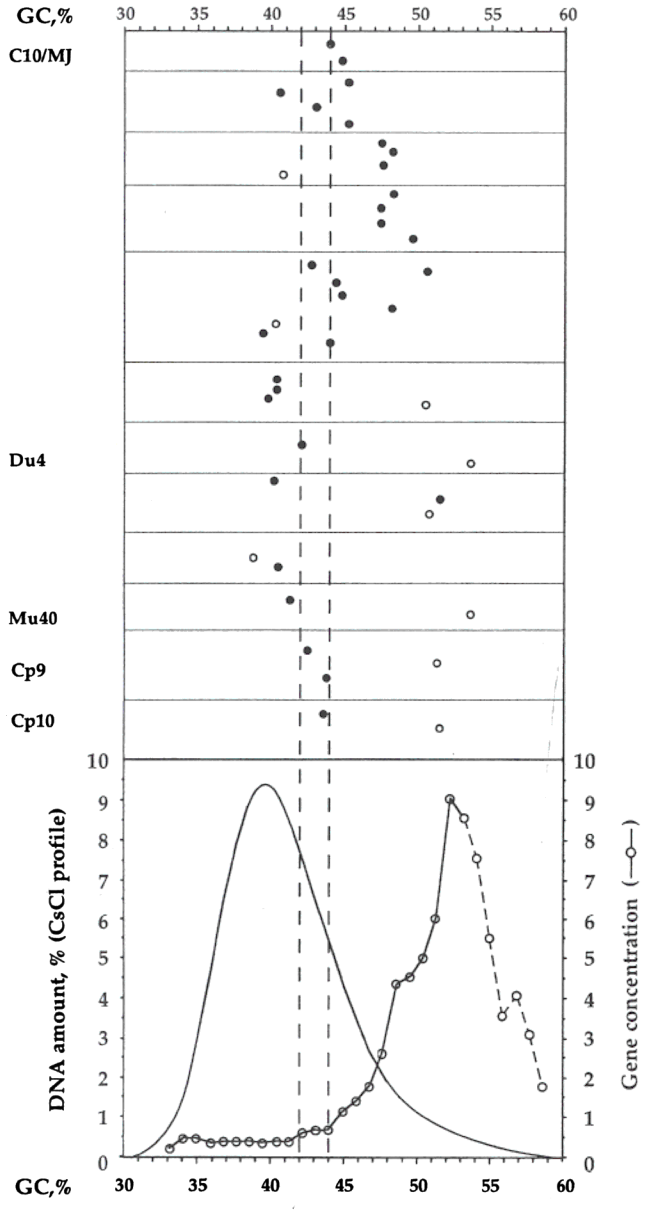

Fig. 5. The distribution of proviral sequences in the genome of the human cell lines investigated (upper frame) is compared with the compositional distribution (CsCl profile, solid line) of human DNA and with the concentration profile (-○-) of human coding sequences (bottom frame from Bernardi, 1993; the final drop, broken line, in gene concentration is due to the presence of ribosomal genes in the GC-richest part of the human genome). The centers of distribution of proviral sequences in the fractions are shown by points put at different levels on the ordinate axis to reflect band sizes. These were assessed by taking into account the hybridization intensities of bands corresponding to same-size fragments. Open circles in the top part correspond to faint bands. The two vertical broken lines indicate the GC range of the host genome in which both transcribed and non-transcribed sequences were found. Only non-transcribed sequences were found below that range, only transcribed sequences above.

two integrated sequences were found in fractions centered at 42–43% and at 43.8% GC, respectively. In Cp10, integration occurred in genomic compartments of about 43.5% GC. (In both cases the faint bands of fraction 7 are neglected).

The remaining four cell lines exhibit HTLV-I sequences that are present both in low (<42%) and high (>44%) GC regions of the host genome. The Hut-102 line shows three proviral sequences in the high-GC region; a fourth one, corresponding to a low-M_r band (4.8 kb), is present as a faint band in low-GC fragments. There is no evidence for Hut-102 transcripts originating from this sequence.

In the case of C8166, in which only *tax* and part of *rex* regulatory genes are expressed (Lee et al., 1984; Bhat et al., 1993), three out of four proviral sequences are again in the high-GC region; the fourth low-M_r one is located in fractions 3 and 4. Transcription of this fourth integrate was not reported, whereas the transcription of two different defective proviruses in addition to the transcript from full-size provirus was shown (Bhat et al., 1993).

For MT-2, where eight proviral copies were detected (seven of them being defective; Kobayashi et al., 1984), only two are in the low-GC compartment. Northern blotting experiments show that three viral mRNA are transcribed from the complete and defective proviruses. However, again we do not know (Kobayashi et al., 1984) whether the low-M_r integrates present in low-GC compartments (one of which corresponds to a faint band) are transcribed.

The clone Mu16 contains at least one transcriptionally active provirus and shows integrates both in the high-GC fractions and in fraction 4. No information is available as to which of these proviruses is transcribed.

The results just presented indicate that non-transcribed HTLV-I sequences are integrated in human genome regions below 42% GC, whereas transcribed sequences are integrated in regions above 44% GC. In the 42–44% GC range both transcribed and non-transcribed sequences were found. These results suggest that in those cases in which integration takes place in both GC-poor and GC-rich regions, the proviruses contained in the former regions are transcriptionally inactive and those contained in the latter are active. While a firm conclusion on this point can only come from the identification of the proviral sequences which are transcribed, an analysis of the relevant HTLV-I transcription data (on Hut-102, C8166, MT2 and Mu16) is compatible with this suggestion. This viewpoint is in agreement with the finding that the transcribed sequences of GC-rich proviruses BLV and RSV are localized in the GC-richest compartments of mammalian genomes (Kettmann et al., 1979; 1980; Rynditch et al., 1991).

(c) Localization of transcribed and non-transcribed HTLV-I sequences and gene distribution in the human genome

If the localization of non-transcribed and transcribed HTLV-I sequences in the human genome is compared with the distribution of human genes (Mouchiroud et al.,

1991; Bernardi, 1993; S.Z. and G.B., data not shown), it appears that the location of transcriptionally inactive and active proviral sequences corresponds to the isochores having a low and constant gene concentration and to those exhibiting a higher gene concentration, respectively. In fact, gene concentration increases in parallel with GC content to reach, in the GC-richest isochores, a level that is over 20-times higher than the low, constant level present in GC-poor isochores (Bernardi, 1993; S.Z. and G.B., data not shown). The latter isochores are also characterized by very high transcription and recombination activity (see INTRODUCTION) and by a very high *Alu* sequence concentration (Soriano et al., 1983; Zerial et al., 1986b). Interestingly, the majority of HTLV-I sequences integrated in GC-rich isochores (above the 44% GC level) are, in fact, located in regions above 47% GC. These correspond to the GC-richest 14% of the genome, which in turn make up T-bands of human chromosomes (Saccone et al., 1992; 1993).

(d) Conclusions

The present results are of interest in several respects:

(*1*) They show that only the GC-richer 60% of the host-cell genomes harbor integrated HTLV-I sequences. This may mean either that HTLV-I sequences cannot be integrated in GC-poor sequences, or that, if integrated, they are highly unstable and are readily excised. The first alternative appears to be favored by the recombination-proneness and by the richness in *Alu* sequences of GC-rich isochores, *Alu* sequences possibly facilitating integration events.

(*2*) Transcriptionally active HTLV-I sequences appear to be located in the GC-richer regions of the host genome, whereas inactive HTLV-I sequences appear to be present in GC-poorer regions (it should be stressed again that both these GC-poor and GC-rich regions are in the GC-richer part of the human genome which makes up 60% of it). These results indicate, in fact, that HTLV-I transcriptionally active sequences are located in those regions of the host genome which are characterized by a high gene concentration, a high transcriptional and recombinational activity, an open chromatin structure with a larger spacing of nucleosomes, by the absence of histone H1, and the acetylation of histones H3 and H4 (Tazi and Bird, 1991; Aïssani and Bernardi, 1991). Cytogenetically, these regions correspond to the GC-rich isochores present in R′-bands (namely, the R-bands that are not T-bands) and to the very GC-rich isochores forming T-bands (Saccone et al., 1992; 1993).

(*3*) The GC level of the HTLV-I provirus itself, 53.9%, is higher than that of the isochores (44–53%) in which integration takes place. This situation is similar to that found for GC-rich genes, whose coding sequences are higher in GC than the flanking sequences and is in contrast with the uniformity of GC levels of GC-poor genes and their surrounding sequences (Aïssani et al., 1991).

Lastly, these results are in full agreement with results obtained on the different viral host-cell systems previously studied (see Rynditch et al., 1991, for a discussion on this point). They go much further, however, for several reasons. The first one is that a much larger number of integrates were investigated compared to previous studies, justifying a generalization of the results. The second reason is that integration was investigated in host cells whose genome is compositionally more heterogeneous (Bernardi, 1989; 1993) than those of the rodents investigated in the case of MMTV and RSV. The third and major reason is that retroviral integration could be interpreted in the context of a detailed knowledge on the distribution of coding sequences of the host genome, i.e., the human genome.

ACKNOWLEDGEMENTS

S.Z. was supported by an EMBO Fellowship for East-European Countries. A.R. acknowledges the support of a NATO Collaborative Research Grant. G.B. is indebted to the Association pour la Recherche sur le Cancer (ARC) for financial support. We thank Dr. Genoveffa Franchini for the gift of HTLV-I-infected cell lines and for critical reading of this paper.

REFERENCES

Aïssani, B., D'Onofrio, G., Mouchiroud, D., Gardiner, K., Gautier, C. and Bernardi, G.: The compositional properties of human genes. J. Mol. Evol. 32 (1991) 493–503.

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

Aldovini, A., De Rossi, A., Feinberg, M.B., Wong-Staal, F. and Franchini, G.: Molecular analysis of a deletion mutant provirus of type 1 human T-cell lymphotropic viruses: evidence for a doubly spliced x-lor mRNA. Proc. Natl. Acad. Sci. USA 83 (1986) 38–42.

Bernardi, G., Olofsson, B., Filipski, J., Zerial, M., Salinas, J., Cuny, G., Meunier-Rotival, M. and Rodier, F.: The mosaic genome of warm-blooded vertebrates. Science 228 (1985) 953–958.

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

Bernardi, G.: The isochore organization of the human genome and its evolutionary history – a review. Gene 135 (1993) 57–66.

Berneman, Z.N., Gartenhaus, R.B., Reitz, M.S., Blattner, W.A., Manns, A., Hanchard, B., Ikehara, O., Gallo, R.C. and Klotman, M.: Expression of alternatively spliced human T-lymphotropic virus type 1 pX mRNA in infected cell lines and in primary uncultured cells from patients with adult T-cell leukemia:lymphoma and healthy carriers. Proc. Natl. Acad. Sci. USA 89 (1992) 3005–3009.

Bhat, N.K., Adachi, Y., Samuel, K.P. and Derse, D.: HTLV-I gene expression by defective proviruses in an infected T-cell line. Virology 196 (1993) 15–24.

Franchini, G., Wong-Staal, F. and Gallo, R.C.: Human T-cell leukemia virus (HTLV-I) transcripts in fresh and cultured cells of patients

with adult T-cell leukemia. Proc. Natl. Acad. Sci. USA 81 (1984) 6207–6211.

Franchini, G., Mulloy, J.C., Koralnik, I.J., Lo Monico, A., Sparkowski, J.J., Andresson, T., Goldstein, D.J. and Schlegel, R.: The human T-cell leukemia/lymphotropic virus type I $p12^I$ protein cooperates with the E5 oncoprotein of bovine papillomavirus in cell transformation and binds the 16-kilodalton subunit of the vacuolar H^+ ATPase. J. Virol 67 (1993) 7701–7704.

Furukawa, Y., Fujisawa, J., Osame, M., Teita, M., Sonoda, S., Kubota, R., Ijichi, S. and Yoshida, M.: Frequent clonal proliferation of human T- cell leukemia virus type 1 (HTLV-I) infected T-cells in HTLV-I associated myelopathy (HAM/TSP). Blood 80 (1992) 1012–1016.

Greenberg, S.J., Jacobson, S., Waldmann, T.A. and McFarlin, D.E. : Molecular analysis of HTLV-I proviral integration and T-cell receptor arrangement indicates that T-cells in tropical spastic paraparesis are polyclonal. J. Infect. Dis. 159 (1989) 741–743.

Hiramatsu, K. and Yoshikura, H.: Frequent partial deletion of human adult T-cell leukemia type I proviruses in experimental transmission: pattern and possible implication. J. Virol. 58 (1986) 508–512.

Höllsberg, P., Wucherpfennig, K.W., Ausubel, L.J., Calvo, V., Bierer, B.E. and Hafler, D.A.: Characterization of HTLV-I in vivo infected T cell clones. IL-2-independent growth of nontransformed T cells. J. Immunol. 148 (1992) 3256–3263.

Hyer, S., Richardson, J.H., Brown, M., Lever, A.M., Nussey, S.S. : Progressive flaccid myelopathy associated with HTLV-I. Lancet 338 (1991) 944.

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.

Kobayashi, N., Konishi, A., Sabe, H., Shigesada, K., Noma, T., Honjo, T. and Hatanaka, M.: Genomic structure of HTLV (human T-cell leukemia virus): detection of defective genome and its amplification in MT-2 cells. EMBO J. 3 (1984) 1339–1343.

Lee, T.H., Coligan, J.E., Sodroski, J.G., Haseltine, W.A., Salahuddin, S.Z., Wong-Staal, F., Gallo, R.C. and Essex, M.: Antigenes encoded by the 3'-terminal region of human T-cell leukemia virus: evidence for a functional gene. Science 226 (1984) 57–61.

Manzari, V., Wong-Staal, F., Franchini, G., Colombini, S., Gelmann, E.P., Oroszlan, S., Staal, S. and Gallo, R.: Human T-cell leukemia-lymphoma virus (HTLV): Cloning of an integrated defective provirus and flanking cellular sequences. Proc. Natl. Acad. Sci. USA 80 (1984) 1574–1578.

Markham, P.D., Salahuddin, S.Z., Kalyanaraman, V.S., Popovic, M., Sarin, P. and Gallo, R.C.: Infection and transformation of fresh human umbilical cord blood cells by multiple sources of human T-cell leukemia-lymphoma virus (HTLV). Int. J. Cancer 31 (1983) 413–420.

Miyoshi, I., Kubonishi, I., Yoshimoto, S., Akagi, T., Ohtsuki, Y., Shiraishi, Y., Nagata, K. and Hinuma Y.: Type C-virus particles in a cord T-cell line derived by co-cultivating normal human cord leukocytes and human leukemic T cells. Nature 294 (1981) 770–771.

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.

Ohshima, K., Kikuchi, M., Masuda, Y., Kobari, S., Sumiyashi, Y., Eguchi, F., Mohtai, H., Yoshida, T., Takeshita, M. and Kimura, M.: Defective provirus form of human T-cell leukemia virus type 1 in adult T-cell leukemia/lymphoma: clinicopathological features. Cancer Res. 57 (1991) 4639–4642.

Poiesz, B.J., Ruscetti, F.W., Gazdar, A.E., Bunn, P.A., Minna, J.D. and Gallo, R.C.: Detection and isolation of type C retroviruses particles from fresh and cultured lymphocytes of a patient with cutaneous T-cell lymphoma. Proc. Natl. Acad. Sci. USA 77 (1980) 7415–7419.

Popovic, M., Sarin, P.S., Robert-Gurroff, M., Kalyanaraman, V.S., Mann, D., Minowada, J. and Gallo, R.C.: Isolation and transmission of human retrovirus (human T-cell leukemia virus). Science 219 (1983) 856–859.

Prunell, A., Goutorbe, F., Strauss, F. and Bernardi, G.: Yield of restriction fragments from yeast mitochondrial DNA. J. Mol. Biol. 110 (1977) 47–52.

Rynditch, A., Kadi, F., Geryk, J., Zoubak, S., Svoboda, J. and Bernardi, G.: The isopycnic compartmentalized integration of Rous sarcoma virus sequences. Gene 106 (1991) 165–172.

Saccone, S., De Sario, A., Della Valle, G. and Bernardi, G.: The highest gene concentrations in the human genome are in T-bands of metaphase chromosomes. Proc. Natl. Acad. Sci. USA 89 (1992) 4913–4917.

Saccone, S., De Sario, A., Wiegant, J., Raap, A.K., Della Valle, G. and Bernardi, G.: Correlations between isochores and chromosomal bands in the human genome. Proc. Natl. Acad. Sci. USA 90 (1993) 11929–11933.

Salahuddin, S.Z., Markham, P.D., Wong-Staal, F., Franchini, G., Kalyanaraman, V.S. and Gallo, R.C.: Restricted expression of human T-cell leukemia-lymphoma virus (HTLV) in transformed human umbilical cord blood lymphocytes. Virology 129 (1983) 51–64.

Salahuddin, S.Z., Markham, P.D., Linder, S.G., Gootenberg, J., Popovic, M., Hemmi, H., Sarin, P.S. and Gallo, R.C.: Lymphokine production by cultured human T-cells transformed by human T-cell leukemia-lymphoma virus. Science 223 (1984) 703–707.

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

Seiki, M., Eddy, R., Shows, T.B. and Yoshida, M.: Non-specific integration of the HTLV provirus genome into adult T-cell leukemia cells. Nature 309 (1984) 640–642.

Soriano, P., Meunier-Rotival, M. and Bernardi, G.: The distribution of interspersed repeats is non-uniform and conserved in the mouse and human genomes. Proc. Natl. Acad. Sci. USA 80 (1983) 1816–1820.

Tazi, J. and Bird, A.P.: Alternative chromatin structure at CpG islands. Cell 60 (1990) 909–920.

Watanabe, T., Seiki, M. and Yoshida, M.: HTLV Type I (U.S. isolate) and ATLV (Japanese isolate) are the same species of human retrovirus. Virology 133 (1984) 238–241.

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

Wucherpfennig, K., Höllsberg, P., Richardson, J.H., Benjamin, D. and Hafler, D.A.: T-cell activation by autologous human T-cell leukemia virus type-infected T-cell clones. Proc. Natl. Acad. Sci. USA 89 (1992) 2110–2114.

Yoshida, M., Seiki, M., Yamaguchi, K. and Takatsuki, K.: Monoclonal integration of human T-cell leukemia provirus in all primary tumors of adult T-cell leukemia suggests causative role of human T-cell leukemia virus in the disease. Proc. Natl. Acad. Sci. USA 81 (1984) 2534–2537.

Yoshida, M., Hattori, S. and Seiki, M.: Molecular biology of human T-cell leukemia virus associated with adult T-cell leukemia. Curr. Top. Microbiol. Immunol. 115 (1985) 157–210.

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.

Zerial, M., Salinas, J., Filipski, J. and Bernardi, G.: Gene distribution and nucleotide sequence organization in the human genome. Eur. J. Biochem. 160 (1986b) 479–485.

Zoubak, S., Rynditch, A. and Bernardi, G.: Compositional bimodality and evolution of retroviral genomes. Gene 119 (1992) 207–213.