Volume 12 Number 22 1984 Nucleic Acids Research

A retroviral provirus closely associated with the *Ren*-2 gene of DBA/2 mice

David W.Burt, Alistair D.Reith and William J.Brammar

ICI/University Joint Laboratory, Department of Biochemistry, University of Leicester, Leicester LE1 7RH, UK

Received 6 September 1984; Revised and Accepted 19 October 1984

ABSTRACT

We have determined the entire nucleotide sequence of an intra-cisternal A particle (IAP) genome, associated with the *Ren*-2 gene of DBA/2 mice. This genome (MIARN) displays features common to other IAP retroviral-like genomes. Long terminal repeats (LTRs) are approximately 430 base pairs (bp) in length and show typical retroviral U3-R-U5 organisation, though the R-region, at 120 bp, is much larger than the average IAP. This difference probably arose by the amplification of a pyrimidine-rich sequence, by a slippage-mispairing mechanism. Flanking the 5' LTR is a sequence complementary to a phenylalanine tRNA, strongly conserved in all rodent IAP genomes and probably required to prime the initiation of (-) strand synthesis. Flanking the 3' LTR, is a purine-rich sequence probably required for (+) strand synthesis. The tRNA binding site (TBS) is flanked by six tandem copies of a sequence homologous to the TBS. The relationship of the MIARN element to other IAP genomes and the significance of its association with the highly expressed *Ren*-2 is discussed.

INTRODUCTION

The genomes of most strains of mice contain copies of retroviral sequences (1,2) that code for potentially infectious virus particles, capable of replicating in murine cells. In addition, the mouse genome contains a family of approximately 1000 copies of sequences, homologous to the RNA of intra-cisternal A particles (IAPs, 3). The latter are non-infectious retrovirus-like structures (4,5) found generally in early mouse-embryos (6,7,8) and mouse tumours (9,10,11), and, very rarely, in normal tissues (12). The mode of retrovirus replication, by way of a DNA intermediate (13), enables them to insert proviral DNA sequences at random sites throughout the host genome. It has been demonstrated that the integration/transposition of these genomes can alter the expression of cellular genes, and in recent years both infectious and non-infectious forms have been associated with such changes. Murine leukemia viruses (MuLVs) have been associated with coat colour (14) and developmental mutants of the mouse (15,16). A cellular oncogene, c-*myc* has been shown to be activated in bursal lymphomas by insertion of complete or partial proviral copies after infection by avian leukosis virus (17,18,19,20).

Recently, IAP genomes have also been associated with both inactivated and activated cellular genes. For example, immunoglobulin C_K genes were inactivated by IAP integration in a hybridoma line (21,22), a transposed α-globin pseudogene (α▼3) was associated with flanking IAP genomes (23) and the c-*mos* oncogene was activated in a plasmacytoma line by insertion of an IAP element (24,25,26).

Inbred mouse strains can be divided into two groups on the basis of submaxillary gland (SMG) renin activities (27). This strain difference has been mapped to a single genetic locus, *Rnr*, the renin regulator, on chromosome 1 (28) and biochemical studies (29,30,31) show that low-renin producing strains (eg C57BL/10) contain a single renin gene, *Ren*-1, whilst high renin-producers (eg DBA/2) have two, *Ren*-1 and *Ren*-2. Comparison of renin genes from high and low SMG renin-producing strains suggest that the low producers evolved from the high renin-producer after deletion of *Ren*-2 (32). This comparison, together with the analysis of renin cDNA clones (33), suggests that *Ren*-2 is expressed at higher levels than *Ren*-1 in the submaxillary gland.

Physical comparison of the *Ren*-1 and *Ren*-2 genes from DBA/2 (31) revealed at least 14kb of sequence homology, with an interruption of 3kb in the 3' flanking region of *Ren*-2. In this paper we characterise this 3kb element further and identify it as an IAP genome. This apparent association between an IAP insertion and high level expression of a cellular gene is of considerable interest because it may be an example of gene activation by a discrete genetic element.

MATERIALS AND METHODS

General Techniques. The methods for restriction and ligation of DNAs and gel electrophoresis of DNA fragments are described in ref. (34).

DNA Sequencing. Cloning into M13mp10 and mp11 was used in conjunction with the chain-termination sequencing reactions of Sanger (35). General techniques used are described in ref. (36).

Construction of pDB*Rn*3 and pDB*Rn*3Δ. To facilitate the quantitative analysis of the 3kb element in mouse genomic DNA, an internal 1700bp *Hind*III fragment from the 3kb element was cloned into pUC8 to generate pDB*Rn*3. To remove LTR sequences from pDB*Rn*3, plasmid DNA was treated with *Eco*RI and the resulting plasmid, pDN*Rn*3Δ, was used as a probe, specific for internal MIARN DNA sequences.

DNA Dot Blots. Dot Blots were made as follows. Spleen and pDB*Rn*3Δ DNA were digested with *Alu*I and then purified by phenol extraction and ethanol

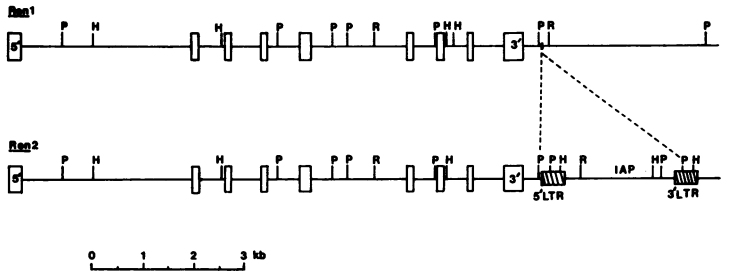

Figure 1. Physical Map of the *Ren*-1 and *Ren*-2 regions, showing the position of the 3kb element. Restriction enzyme cutting sites are represented as follows: P= *Pst*I, H= *Hind*III, R= *Eco*RI. The exons and introns of the renin genes, determined by comparitive sequencing of genomic and cDNA clones (32), are represented as open boxes and lines, respectively. The MIARN LTRs are shaded.

precipitation. DNAs were dissolved in 5µl 1M NaCl, 0.1M NaOH, 1mM EDTA and varying amounts of sheared salmon sperm DNA were added. The amount of total DNA was kept constant (5µg) for each sample applied. Samples were boiled for 5 min and applied to nitrocellulose filters, pre-hybridised as described for genomic blots in ref. (31). Membranes were air-dried for 30 min, rinsed in 3xSSC for 2 min at room temperature, blotted dry and baked at 80°C for 4 hours. Membranes were then treated in the same way as genomic blots in ref. (31).

Computer Analysis of Nucleotide Sequence Homologies. Dot-plots were obtained using a diagonal-traverse homology search algorithm based on that described in ref. (37). The program is written in compiled Microsoft BASIC and runs on a SIRIUS 1 16 bit microcomputer. Screen displays can be dumped to any Epson FX printer.

RESULTS

Association of a 3kb DNA element with *Ren*-2

The composite physical maps of *Ren*-1 and *Ren*-2, deduced from restriction enzyme mapping and DNA sequencing of genomic clones from a high renin-producing strain (DBA/2), are shown in figure 1. The extensive region of homology between *Ren*-1 and *Ren*-2 is interrupted by an extra 3kb of DNA in the 3' region flanking *Ren*-2. This was shown clearly in heteroduplexes between *Ren*-1/*Ren*-2 genomic clones as a 3kb deletion/substitution loop (31).

Estimation of the copy number of the 3kb element

A reconstruction DNA dot blot technique was used to estimate the number of sequences homologous to the 3kb element in the genome of the mouse. This

Table 1. Copy Number Measurements of the *Ren*-2 3kb element.

ng Mouse DNA	cpm probe bound	ng pDB*Rn*3Δ DNA	cpm probe bound
4000	3007	100	105432
2000	1475	50	56907
1000	1107	25	29663
500	1033	12.5	27207
250	728	6.25	17665

Dot-blots were made (see MATERIALS AND METHODS), with mouse or pDB*Rn*3Δ DNA in amounts shown above. Each blot was hybridised to an internal 3kb element probe (pDB*Rn*3Δ). The individual dots were counted in scintillant; the cpm are shown. The two sets of data were analysed by linear regression. The results of regression are as follows: 0.580 cpm probe bound/ng of mouse DNA ; 929 cpm probe bound/ng of pDB*Rn*3Δ DNA. Thus, within 1ng of mouse DNA, 0.06% (0.580/929x100%) is homologous to the pDB*Rn*3Δ probe. The mouse haploid genome is $3x10^9$ bases, so $(3x10^9)x(0.0006)=1.8x10^6$ bases are homologous to the pDB*Rn*3Δ DNA. The plasmid pDB*Rn*3Δ contains a 1300bp *Eco*RI/*Hind*III DNA fragment from within the 3kb element, thus, within the mouse genome there are approximately 1400 ($1.8x10^6/1300$) sequences related to the pDB*Rn*3Δ probe per mouse haploid genome.

involved quantitating the amounts of mouse DNA and 3kb element DNA required to hybridise to the same amount of internal probe (pDB*Rn*3Δ). The data are presented in table 1, and the computational details are in the legend. The data show that the mouse genome contains approximately 1400 copies of sequences related to the 3kb element per haploid mouse genome.

The 3kb element is an IAP genome

To examine the 3kb element further, we sub-cloned this element from λ genomic clones into the plasmid pUC8. DNA from these sub-clones was then cloned into M13 vectors for sequence analysis. To determine the boundaries of the DNA insertion, the homologous region from *Ren*-1 was also sequenced. The entire DNA sequence of the 3kb element and the target site in *Ren*-1 are shown (Figures 2 and 3, respectively).

Comparison of the 3kb DNA sequence with itself in the form of a dot-plot, (Figure 4) clearly shows directly repeating sequences of approximately 430 base pairs (bp) at each end of the unit. These LTR sequences show extensive homology (83-88%) with the LTRs of other IAP genomes from *Mus musculus* and internal sequences flanking the LTRs also show homology with regions sequenced in other IAP genomes (22, 24 and 38).

Structural features of LTR and flanking sequences

The renin IAP genome (MIARN), in common with other IAP genomes, shares many structural features with other groups of retroviral LTRs (40). These IAP LTRs include many highly conserved regions, each containing possible regulatory sequences. The sequence TTAAAA, which matches a consensus

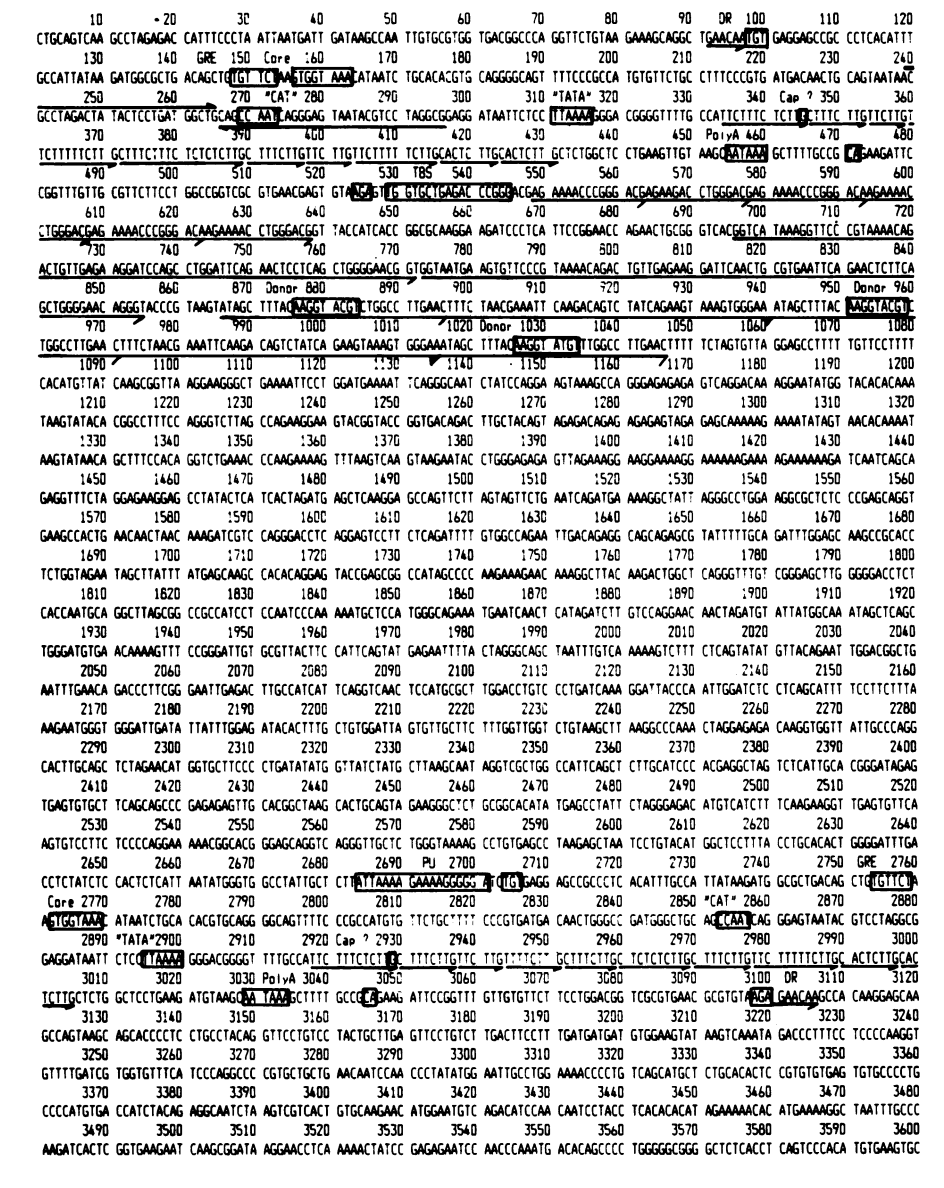

Figure 2. DNA Sequence of the *Ren*-2 IAP Genome, MIARN. Possible regulatory sequences are boxed: IR at 98 and 2704; GRE at 148 and 2754; Core Enhancer at 156 and 2762; CAT-box at 269 and 2853; TATA-box at 311 and 2895; Poly(A) recognition signal at 455 and 3029; Poly(A)-addition site at 471 and 3045; TBS at 529; 5' intron splice-sites at 876, 951 and 1026. Repeats are underlined. Direct-repeats between positions 696 and 855 are inperfect, showing 92% homology with each other. Between positions 866 and 1046, two types of direct-repeats are distinguished by open and closed arrow heads.

(a) 70 80 90 3110 3120
AGGTTCTGTAAGAAAGCAGGCTGAACAA: IAP :GAACAAGCCACAAGGAGCAAGCCAGTA *Ren*-2

AGGTTCTGTAAGAAAGCAGGCTGAACAAGCCACAAGGAGCAAGCCAGTA *Ren*-1

(b) 3130 3140 3150
AGCCAGTAAGCAGCACCCCTCCTGCCTACAGGTTCCTGTC *Ren*-2

AGCCAGTAAGCAGCACCCCTCC:ACGGCCTCTACATCAG:CTCCTGCCTACAGGTTTCTGTC *Ren*-1

Figure 3. (a) *Ren*-1 DNA sequence homologous to the *Ren*-2 region containing the site of integration of the *Ren*-2 associated IAP genome, MIARN. (b) Site of integration of the *Ren*-1 mobile-like element.

"TATA-box" ($TATA^{A}_{T}A^{A}_{T}$, 41) is present in both LTRs (Figure 2, positions 313 and 2895). The sequence CCAAT (the so called "CAAT" box, 41), is present in both LTRs 42bp upstream from each "TATA" box: together these define an RNA polymerase II promoter in both LTRs (41). The sequence AATAAA, a polyadenylation signal present at positions 455 and 3029 (Figure 2), probably directs the post-transcriptional cleavage event prior to polyadenylation (42) and precedes the dinucleotide CA, a preferred polyadenylation site (Figure 2, positions 471 and 3045). In addition, there is a conserved sequence (Figure 2, positions 156 and 2762), that matches closely a core-enhancer sequence ($GTGGT^{A}_{T}A^{A}_{T}A^{A}_{T}$, 43). Such enhancers commonly occur in the U3 region of retroviral LTRs, often (not in the case of MIARN) as part of a larger direct repeat (13). Adjacent to this conserved sequence is the sequence TGTTCT (Figure 2, positions 148 and 2754), which matches a consensus

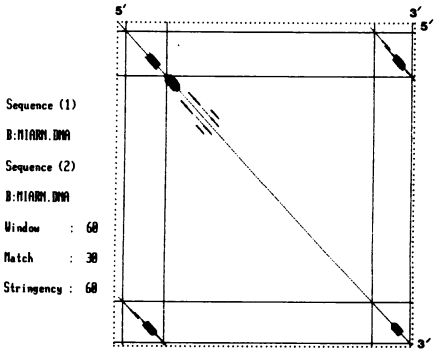

Figure 4. Dot-plot of the DNA sequence of the 3kb element to itself, showing the long terminal repeats (LTRs). Dots around the border mark off increments of 50bp. The limits of the LTRs are indicated by four sets of parallel lines on each axis.

glucocorticoid-responsive element (GRE, 44, 45). In the LTRs of MMTV, there is a similar close association between enhancer and GRE sequences (46) and such an arrangement is hormone-responsive (47). It will be of interest to determine whether the MIARN sequences also define a hormone-responsive element and whether this may influence the expression of neighbouring genes.

The retroviral LTR structure can be sub-divided into 3 regions, U3-R-U5 (40). The lengths of U3 regions are 247bp for the 5'LTR and 224bp for the 3'LTR. The difference in size is not due to a simple tandem duplication of DNA in the 5'LTR, but an imperfect, palindromic duplication of 26bp centred at position 265bp. The R region always starts with a G and is usually followed by C, approximately 30bp downstream of the "TATA" box and ends with the poly(A) addition site, CA. The length of the R region varies between IAP genomes (39), that in MIARN being the largest. This size variability is the result of amplification of an 8 bp unit (TTCTCTTG), probably involving a slippage-mispairing mechanism (48), which would be facilitated by the base asymmetry and high A/T composition of this region. The lengths of the U5 regions are short, 54bp for both 5' and 3'LTRs, a feature common to all IAP genomes.

Four kinds of tRNA: Trp, Pro, Lys (40) and Phe (39), have been identified as primer tRNAs for reverse transcription of retroviral genomes. The nucleotide sequences of the tRNA-binding site (TBS) of MIARN were complementary to the last 17 nucleotides of a mammalian phenylalanine tRNA (50 and 51) again a conserved feature for IAP genomes (22, 24 and 38). Adjacent to the TBS are six tandem repeats of a 15bp basic unit, closely related to the TBS itself. As with other tandem repeats, this probably evolved by a base slippage-mispairing mechanism (48).

Adjacent and upstream from the 3'LTR is a conserved purine-rich sequence, 17bp in length. It has been speculated that this region might be involved in the initiation of plus strand synthesis (40).

Most retroviral genomes contain a 5' intron splice-site near the TBS region. A search for probable 5' intron splice sequences (52) reveals three possibilities at positions 873, 947 and 1027 with 8/9 matches with the published consensus, AAGGTAAGT (52). Sequences extending from the 5' terminus of viral RNA up to the first 5' intron splice-site are thought to serve as an untranslated leader sequence which is spliced onto sub-genomic mRNAs (53).

The target site for MIARN integration

Because the *Ren*-2 gene and its flanking regions are highly homologous with the closely linked *Ren*-1 gene, presumably via tandem duplication, it is possible to deduce the sequence of the target site into which the MIARN

proviral sequence integrated. Comparison shows a duplication of 6bp, GAACAA, flanking both LTRs, in the *Ren*-2 region, and that is present in only one copy in the homologous *Ren*-1 region (Figure 3a). Although the sequences flanking the integrated IAP genomes are non-homologous, the duplication of target DNA is always 6bp (22, 24 and 38). Such an arrangement is typical of prokayotic and eukaryotic mobile elements (40).

In addition, this comparison also shows a 16bp sequence in *Ren*-1 that is not present in the *Ren*-2 sequence (Figure 3b). A duplication of 4bp, CTCC, flanks this sequence, however, only one copy of which was present in the homologous *Ren*-2 region (at position 3138). These features are characteristic of insertion elements and suggest an insertion event may have occurred in the 3' flanking region of the *Ren*-1 gene.

DISCUSSION

Comparison of the structure of IAP genomes

IAP genomes have been classified into two groups on the basis of restriction mapping and DNA heteroduplex analysis (54, 55 and 56). Type I, the major group (91% of total) of IAP genomes, are mostly 7.2kb in length. Genomes in the minor group, type II, are chiefly 4.8kb in length and contain a 500bp sequence not present in type I genomes. The restriction map of MIARN has been compared to the *Hind*III/*Eco*RI/*Pst*I maps of other IAP genomes (Figure 6) and internal homologies are evident indicating their retroviral origin. This conclusion was supported by copy number measurements using an internal MIARN probe, a value was obtained close to that found for other IAP geneomes. As with other rearrangents involving IAP genomes, MIARN has probably undergone a deletion of internal sequences. A deletion of 4kb, beginning approximately 1-2kb from the 5' LTR, has removed internal IAP sequences from the original MIARN IAP genome. However, a 1.5kb region flanking the 3' LTR has been conserved in all re-arranged IAP genomes (57, 58 and 59). This feature is also true of other retrovirus genomes and the retained element is known as the constant region, C. The internal rearrangements of IAPs are generally in the same position and inspection of the MIARN DNA sequence reveals three pertinent features. A pair of 77bp imperfect direct-repeats between positions 696 and 855, display 92% homology with each other and are each flanked by 3bp direct-repeats (Figure 2). Direct-repeats are also found between positions 866 and 1046 and sequences between 1050 and 1450 are A/T-rich, with A and T residues showing DNA strand asymmetry (Figure 2). It is interesting to note the unusual pattern of these direct-repeats, consisting of three repeats of 31bp interrupted by repeats of 75bp (Figures 2 and 4). The 31bp repeats

Table 2. A pairwise comparison of IAP LTR sequences.

	MIARN 3'	MIA14 5'	MIA14 3'	MIAX24 5'	MIAX24 3'	L20 5'	H10 5'
MIARN 5'	.02±.01	.17±.03	.16±.02	.15±.02	.19±.03	.22±.03 (2.78)	.61±.06 (1.53)
MIARN 3'		.16±.02	.15±.02	.14±.02	.18±.03	.22±.03 (2.75)	.61±.06 (1.53)
MIA14 5'			.08±.02	.07±.02	.10±.02	.25±.03 (3.13)	.65±.07 (1.63)
MIA14 3'				.02±.01	.08±.02	.23±.03 (2.85)	.62±.07 (1.56)
MIAX24 5'					.07±.01	.22±.03 (2.74)	.61±.06 (1.53)
MIAX24 3'						.25±.03 (3.11)	.60±.06 (1.51)
L20 5'							.58±.06 (1.45)

The number of base substitutions per site (together with standard errors) as estimated by using the three-substitution-type model of Kimura (64). These were calculated using aligned LTR sequences (Figure 5). In brackets estimates of base substitution rate are given as number of base substitutions per site per year ($x10^{8}$), where divergence times are known (*Mus musculus* vs *Mus caroli*, 4 million years and *Mus* species vs *Mesocricelus auratus*, 20 million years. 49). References for LTRs follow :- MIARN this work; MIA14 (22), MIAX24 (24), *Mus caroli*, L20 (63) and *Mesocricelus auratus*, H10 (39).

contain consensus 5' splice sites (Figure 2). These sequence features may be related to the apparent instability of this region.

If the general organisation of IAP genomes is similar to that of other retroviral genomes, 5'LTR:PBS,gag,pol,env,c,PU:3'LTR, (40) then our findings would suggest that only the terminal sequences remain within the MIARN element. This suggestion is supported by the lack of open reading frames between the two LTRs (data not shown).

Evolution of the *Rnr* locus in the mouse

A model of renin gene evolution in the mouse involving gene duplication of an ancestral gene approximately 13 million years ago has been proposed (60). Recently, we have isolated and characterised a renin gene from a low renin-producing mouse strain (32). This gene was calculated to be more closely related to the *Ren*-1 (1.5% sequence divergence) gene of a high renin-producer than a *Ren*-2 (10% sequence divergence) gene. These findings supported a model of gene duplication approximately 9 million years ago, followed by a recent

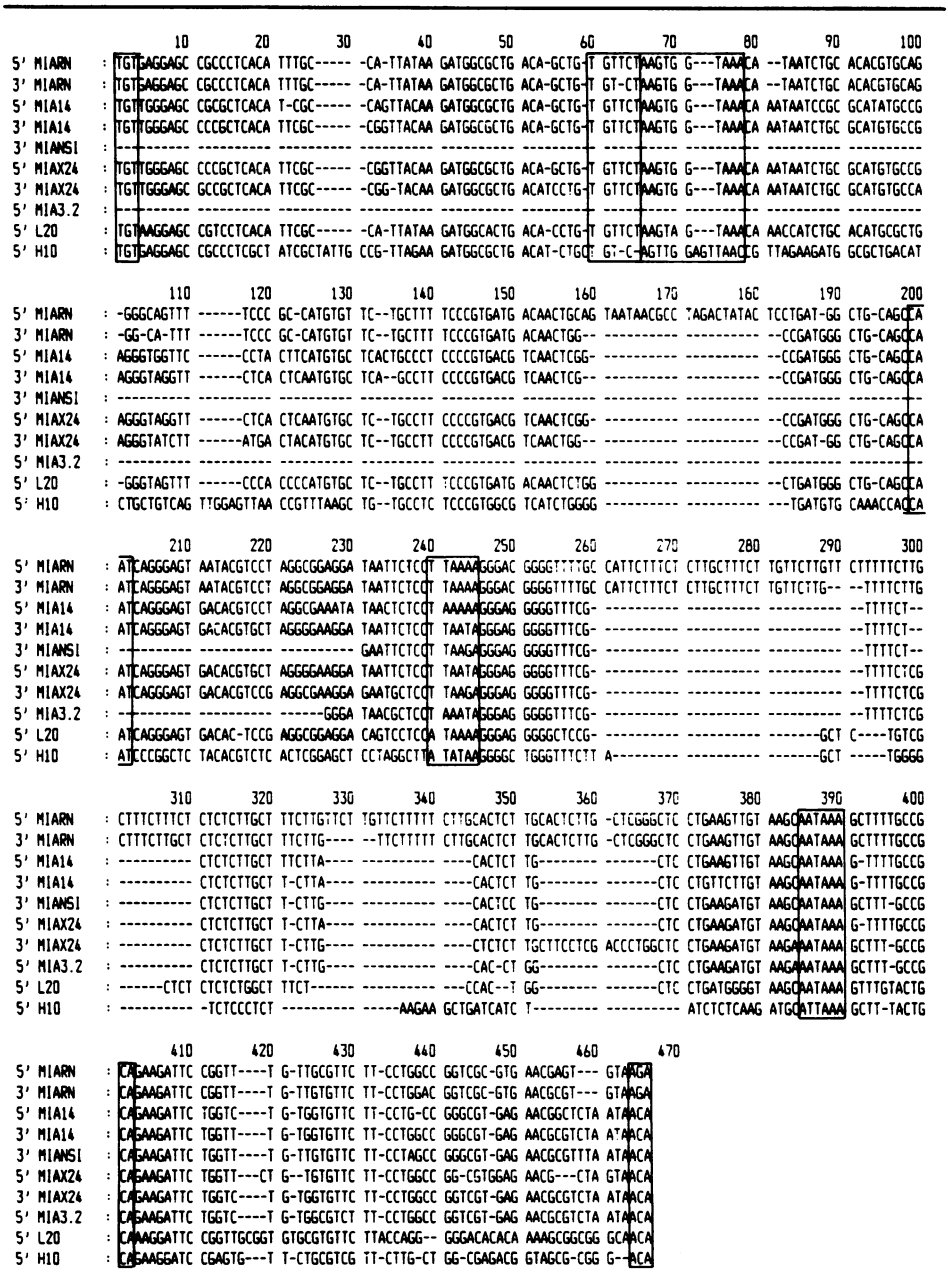

Figure 5. Comparison of LTR sequences from IAP genomes. Dashes indicate gaps inserted to maximise the alignment of homologous regions of the sequences. Conserved sequences at the terminii and possible regulatory sequences are boxed. References for LTR sequences are as follows :- MIA14 (22); MIANSI (38); MIA3.2 (38); L20 (63) and H10 (39).

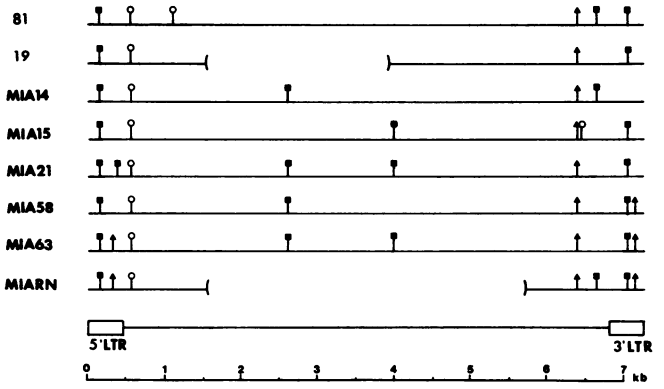

Figure 6. Physical comparison of IAP genomes. Restriction enzyme cutting sites are represented as follows: ■ = *Pst*I, ▲ = *Hind*III, O= *Eco*RI. ()= deletion. References: clones 81 and 19 (53); clones MIA14-63 (65).

(1.2 million years) deletion event removing the *Ren*-2 gene.

The MIARN genome can be treated as a member of a large multigene family, the IAP genomes and sequence divergence between the various units may be used to estimate when integration and divergence of various IAP genes took place. An important consequence of the mechanism of retrovirus replication and integration is the generation of identical LTR sequences (13). Thus, the difference in sequence between 5' and 3' LTRs must be a result of random mutation in each LTR, following integration. We have calculated 0.02±0.01 base substitutions per site from a comparison of the MIARN LTRs (Table 2). Comparison of MIARN and L20 LTRs, from species thought to have diverged from a common ancestor 4 million years ago (49), yielded an estimated mutation rate of 2.8×10^{-8} base substitutions per site per year. Assuming that the mutation rate of the MIARN LTRs is equal to this, then the integration event occurred approximately 0.7 million years ago. This date must be a minimum estimate, since homogenisation events (61), between and within IAP genomes, would reduce the degree of divergence seen in pairs of LTRs. This date would suggest that both the *Ren*-2 deletion and MIARN integration are both relatively recent events. These conclusions then support the following model of evolution for the *Rnr* locus in the mouse. Approximately 9-13 million years ago a *Ren*-1 like gene was duplicated to generate the *Ren*-1/*Ren*-2 gene arrangement, followed recently by the integration of the MIARN genome into the 3' flanking region of *Ren*-2. If this event took place before the *Ren*-2 deletion, then the latter must also remove the MIARN genome. Alternatively, the two events took place in separate groups of mice.

The number of base substitutions per site observed between pairs of IAP LTRs (Table 2), together with an estimated mutation rate of 2.8×10^{-8}/site/year, provide minimum estimates of the times of divergence. Apparently, the *Musculus musculus* IAP genomes have diverged from a common ancestor, approximately 3-7 million years ago. Also, these LTR sequences are all equally related to the L20 LTR sequence from *Musculus caroli* (Table 2), suggesting a common ancestor for these LTRs, 4 million years ago (49). These mouse LTRs all seem to share a common ancestor with the H10 LTRs of *Mesocricelus auratus*, some 20 million years ago (39).

Comparisons of LTRs within individual IAP genomes suggest that the times of integration are relatively recent (0.7 to 2.9 million years ago), however these times are probably gross underestimates due to processes of homogenisation (61) between pairs of LTRs. The comparison of LTRs within and between the IAP genomes of MIA14 and MIAX24 seems to show that the 3' LTR of MIA14 is more similar to the 5' LTR of MIAX24 than it is to that of MIA14. This could be evidence of homogenisation between IAP genomes within a given species, possibly by a mechanism of gene conversion.

The rates of base substitutions observed between pairs of IAP LTRs is at least 10-fold greater than that characteristic of functional genes (64). This would suggest that the terminal repetitions of the proviral elements examined are no longer subject to strong functional constraints.

Association of an IAP genome with the *Ren*-2 gene.

Renin cDNA clones have been used to investigate the organization of renin gene sequences in mice of high and low renin strains (29 and 30). The physical maps given for *Ren*-1 and *Ren*-2 (Figure 1), show that the 8.8kb, 3.9kb pair are derived from the *Ren*-1 region and the 9.2kb, 4.4kb pair from the *Ren*-2 region. Sequence analysis of renin cDNAs and *Ren*-1/*Ren*-2 genomic sequences (32 and 33), indicate that *Ren*-1 codes for a kidney renin and *Ren*-2 codes for the major renin mRNA species in the SMG. The physical map of the *Ren*-2 region also indicates that the 4.4kb *Eco*RI fragments detected in high renin strains is due to the 3kb insertion and is therefore diagnostic for the association of this element with the highly expressed *Ren*-2 gene. The three low renin-producing strains Balb/c, C3H and C57BL/6 all have 3.9kb and 8.8kb *Eco*RI fragments in Southern blots (30), while the three high renin-producers DBA/2, AKR and SWR have four *Eco*RI fragments of 3.9kb, 4.4kb, 8.8kb and 9.2kb (29 and 30). These findings then demonstrate an association between high SMG renin expression and the presence of an IAP genome flanking the highly expressed gene, *Ren*-2. Given the well established correlation between proviral LTRs and altered expression of closely linked genes (17-20), it seems possible that the elevated

expression of *Ren*-2 in the SMG of high-producer strains is due to the close proximity between that structural gene and the provirus. Since *Ren*-2 expression has only been described in the SMG, it is not yet possible to argue whether any possible enhancing effect is tissue-specific.

Sequences homologous to IAP genomes have been detected in a wide range of mammals, including many rodent species, the bat, cat and monkey (62). These mobile genomes and their transcription-control signals may therefore play a significant role in altering the expression of cellular genes in mammalian genomes.

ACKNOWLEDGEMENT

We are grateful for the skilled technical assistance of J. Brooks, H. George, J. Walker and J. Bridges, to many colleagues in Leicester for fruitful discussion and to I.C.I for encouragement and financial support.

REFERENCES

1. Chen, H.W., Bryan, T., Moore, J.L., Staal, S.P., Rowe, W.P and Martin, M.A. (1980) Proc. Natl. Acad. Sci. U.S.A.77, 5779-5783.
2. Jenkins, N.A., Copeland, N.G., Taylor, B.A. and Lee, B.K. (1982) J. Virol. 43, 26-36.
3. Lueders, K.K and Kuff, E.L. (1977) Cell 12, 963-972
4. Hall, W.T., Hartley, J.W. and Sanford, K.K. (1968) J.Virol. 2, 238-247.
5. Kuff, E.L., Wivel, N.A. and Lueders, K.K. (1968) Cancer Res. 28, 2137-2148.
6. Biczysko, W., Pienkowski, M., Solter, D. and Koprowsky, H. (1973) J. Nat. Cancer Inst. 51, 1041-1050.
7. Calarco, P.G. and Szollosi, D. (1973) Nature New Biol. 243, 91-93.
8. Chase, D.G. and Piko, L. (1973) J. Natl. Cancer Inst. 51, 1971-1973.
9. de Harven, E. and Friend, C. (1958) J. Biophys. Biochem. Cytol. 4, 151-156.
10. Perk, K. and Dahlberg, J.E. (1974) J. Virol. 14, 1304-1306.
11. Dalton, A.J., Potter, M. and Merwin, R.M. (1961) J. Natl. Cancer Inst. 26, 1221-1267.
12. Wivel, N.A. and Smith, G.H. (1971) Int. J. Cancer 7, 167-175.
13. Temin, H.M. (1981) Cell 27, 1-3.
14. Copeland, N.G., Hutchinson, K.W. and Jenkins, N.A. (1983) Cell 33, 379-387.
15. Jaenisch, R., Harber, S.K., Schnieke, A., Lohler, J., Jahner, D., Grotkopp, D. and Hoffman, E. (1983) Cell 32, 209-216.
16. Schieke, A., Harbers, K. and Jaenisch, R. (1983) Nature (London) 304, 315-320.
17. Fung, Y,-K., Fadly, A.M., Crittenden, L.B. and Kung, H.,-J. (1981) Proc. Natl. Acad. Sci. U.S.A 78, 3418-3422.
18. Hayward, W.S., Neel, B.G. and Astrin, S.M. (1981) Nature (London) 290, 475-479.
19. Jenkins, N.A., Copeland, N.G., Taylor, B.A. and Lee, B.K. (1981) Nature (London) 293, 370-374.
20. Westaway, D., Payne, G. and Varmus, H.E. (1984) Proc. Natl. Acad. Sci. U.S.A 81, 843-847.

21. Hawley, R.G., Shulman, M.J., Murialdo, H., Gibson, D.M. and Hozumi, N. (1982) Proc. Natl. Acad. Sci. U.S.A 79, 7425-
22. Kuff, E.L., Feenstra,A., Lueders, K., Smith, L., Hawley, R., Hozumi, N. and Shulman, M. (1983) Proc. Natl. Acad. Sci. U.S.A 80, 1992-1996.
23. Lueders, K., Leder, A., Leder, P. and Kuff, E. (1982) Nature (London) 295, 426-428.
24. Canaani, E., Dreazen, O., Klar, A., Rechavi, G., Ran, D., Cohen, J.B., and Givol, D. (1983) Proc. Natl. Acad. Sci. U.S.A 80, 7118-7122.
25. Kuff, E.L., Feenstra, A., Lueders, K., Rechavi, G., Givol, D. and Canaani, E. (1983) Nature (London) 302, 547-548.
26. Rechavi, G., Givol, D. and Canaani, E. (1982) Nature (London) 300, 607-611.
27. Wilson, C.M., Erodos, E.G., Dunn, J.F. and Wilson, J.D. (1977) Proc. Natl. Acad. Sci. U.S.A 74, 1185-1189.
28. Wilson, C.M., Erdos, E.G., Wilson, J.D. and Taylor, B.A. (1978) Proc. Natl. Acad. Sci. U.S.A 75, 5623-5626.
29. Piccini, N., Knopf, J.L. and Gross, K.W. (1982) Cell 30, 205-213.
30. Panthier, J.J., Holm, I. and Rougeon, F. (1982) EMBO J. 1, 1417-1421.
31. Mullins, J.J., Burt, D.W., Windass, J.D., McTurk, P., George, H. and Brammar, W.J. (1982) EMBO J. 1, 1461-1466.
32. Burt, D.W., Beecroft, L.J., Mullins, J.J., Pioli, D. and Brammar, W.J. (1984) in preparation.
33. Panthier, J.J. and Rougeon, F. (1983) EMBO J. 2, 675-678.
34. Burt, D.W. and Brammar, W.J. (1982) Mol. Gen. Genet. 185, 462-467.
35. Sanger, F. (1980) J. Mol. Biol. 143, 161-178.
36. Sanger, F., Coulson, A.R., Hong, C.F., Hill, D.F. and Peterson, G.B. (1982) J. Mol. Biol. 162, 729-773.
37. White, C.T., Hardies, S.C., Hutchison III, C.A. and Edgell, M.H. (1984) Nucleic Acids Res. 12, 751-766.
38. Cohen, J.B., Unger, T., Rechavi, G., Canaani, E. and Givol, D. (1983) Nature (London) 306, 797-799.
39. Ono, M. and Ohishi, H. (1983) Nucleic Acids Res. 11, 7169-7179.
40. Chen, H.R. and Barker, W.C. (1984) Nucleic Acids Res. 12, 1767-1778.
41. Breathnach, R. and Chambon, P. (1981) Annu. Rev. Biochem. 50, 349-362.
42. Montell, C., Fisher, E.F., Caruthers, M.H. and Berk, A.J. (1983) Nature (London) 305, 600-605.
43. Laimins, L.A., Kessel, M., Rosenthal, N. and Khoury, G. (1983) in Enhancers and Eukaryotic Gene Expression, Gluzman, Y. and Shenk, T. Eds., pp28-37, Cold Spring Harbor, New York.
44. Buetti, E. and Diggelmann, H. (1983) EMBO J. 2, 1423-1429.
45. Scheidereit, C., Geisse, S., Westphal, H.M. and Beato, M. (1983) Nature (London) 304, 749-752.
46. Scheidereit, C. and Beato, M. (1984) Proc. Natl. Acad. Sci. U.S.A. 81, 3029-3033.
47. Chandler, V.L., Maler, B.A., and Yamamoto, K.R. (1983) Cell 33, 489-499.
48. Streisinger, G., Okada, Y., Emrich, J., Newton, J., Tsugita, A., Terzaghi, E. and Inouye, M. (1966) Cold Spring Harbor Symp. Quant. Biol. 31, 77-84.
49. Benveniste, R.E., Callahan, R., Sherr, C.J., Chapman, V. and Todaro, G.J. (1977) J. Virol. 21, 849-862.
50. Keith, G. and Dirheimer, G. (1978) Biochim. Biophys. Acta. 517, 133-149.
51. Roe, B.A., Anandaraj, M.P.J.S., Chia, L.S.Y., Randerath, E., Gupta, R.C. and Randerath, K. (1975) Biochem. Biophys. Res. Commun. 66, 1097-1105.
52. Mount, S.M. (1982) Nucleic Acids Res. 10, 459-472.
53. Van Beveren, C., Rands, E., Chattopadhyay, S.K., Lacy, D.R. and Verma, I.M. (1982) J. Virol. 41, 542-556.

54. Cole, M.D., Ono, M. and Huang, K.C. (1981) J. Virol. 38, 680-687.
55. Ono, M., Cole, M.D., White, A.T. and Huang, R.C.C. (1980) Cell 21, 465-473.
56. Sheng-Ong, G.L.C. and Cole, M.D. (1982) J. Virol. 42, 411-421.
57. Kominami, R. and Hatanaka, M. (1979) J. Virol. 32, 925-933.
58. Kominami, R., Tomita, Y., Connors, E.C. and Hatanaka, M. (1980) J. Virol. 34, 684-692.
59. Suzuki, A., Kitasato, H., Kawakami, M. and Ono, M. (1982) Nucleic Acids Res. 10, 5733-5746.
60. Holm, I., Ollo, R., Panthier, J.-J. and Rougeon, F. (1984) EMBO J. 3, 557-562.
61. Brown, S.D.M. and Dover, G. (1981) J. Mol. Biol. 150, 441-466.
62. Lueders, K.K. and Kuff, E.L. (1981) Nucleic Acids Res. 9, 5917-5930.
63. Ono, M., Kitasato, H., Ohishi, H. and Motobayashi-Nakajima, Y. (1984) J. Virol. 50, 352-358.
64. Kimura, M. (1981) Proc. Natl. Acad. Sci. U.S.A. 78, 454-458.
65. Lueders, K.K. and Kuff, E.L. (1980) Proc. Natl. Acad. Sci. U.S.A. 77, 3571-3575.