The Mechanisms of Retroviral Gene Transduction

by

Ealon Mathew Joelson

B.A. (University of California at Berkeley) 1988

A thesis in partial satisfaction of the

requirements for the degree of

Master of Science

in

Health and Medical Sciences

in the

GRADUATE DIVISION

of the

UNIVERSITY of CALIFORNIA at BERKELEY

Committee in charge:

Professor Peter H. Duesberg, Chair
Professor Gertrude C. Buehring
Professor John E. Swartzberg

1993

The thesis of Ealon Mathew Joelson is approved:

Peter Berkeley 2/July/93.
Chair Date

6/15/93
Date

Gertrude C. Buckley 7/2/93
Date

University of California at Berkeley

1993

Table of Contents

List of Tables and Figures

Introduction

Since their initial discovery over eighty years ago, retroviruses have been a focus of attention (1). Of major interest has been the presence of coding regions within some viral genomes which have the ability to transform cells in vitro and cause cancer in animal hosts after infection. These sequences are called oncogenes and over twenty of them have been discovered (2). These viral oncogenes (*v-onc*) were presumably picked up, or transduced, from proto-oncogenes (*p-onc*) present in the host cell genome, which do not normally cause cancer in the host cell milieu. For obvious reasons, the mechanisms by which these viral oncogenes act to cause cancer and their relationship to proto-oncogenes has been the subject of intense research (1,2).

Similarly, the mechanisms by which oncogenes become transduced into retroviral genomes have also been the subject of much investigation. Because techniques of molecular biology have become refined only within the last ten years, serious attempts to elucidate the steps involved in transduction of *p-oncs* have occurred only recently. Two divergent theories regarding the mechanisms of gene transduction have emerged from such research. To comprehend both these models, first a basic understanding of retroviral anatomy and life cycles is necessary.

The Normal Retroviral Life Cycle

It should be noted from the outset that, even though the DNA phases of the life cycle consist of the traditional anti-parallel double helix molecule, DNA genomes are referred to as being unipolar

in correspondence with the truly unipolar viral RNA phase. Therefore for simplification purposes viral 5' (also called upstream or leftward) DNA sequences correspond to the viral 5' RNA sequences etc.

Let us start by looking at the viral genome after it has been copied from ssRNA to dsDNA in the infected cell. The proviral genome at this point is complete, containing the appropriate viral coding regions and other sequences important in viral replication (described below) sandwiched between two LTRs (long terminal repeat) (fig. 1). Each LTR represents an identical nucleotide sequence, about 700 base pairs long, present on each end of the proviral genome (1,2). The LTRs are of the same polarity in relation to each other and are functionally identical. Moreover, as will be outlined, LTRs are multifunctional.

Fig. 1. General anatomy of the dsDNA retrovirus. Each LTR is identical and is further subdivided into functional/structural units. The U3 and U5 stand for unique 3' and 5' respectively while R means repeat. Psi (Ψ) is the packaging site and lies upstream of the protein coding sequences. Not drawn to scale.

The proviral DNA finds its way to the nucleus where it integrates at the LTRs with the aid of inverted repeat sequences found at the extreme edges of the LTRs (fig. 2) (1,2). The integration is not site specific, however, and appears to occur through a non-homologous process through the aid of a viral integrase (3). The viral DNA is now fully integrated with cellular DNA flanking the two viral LTRs.

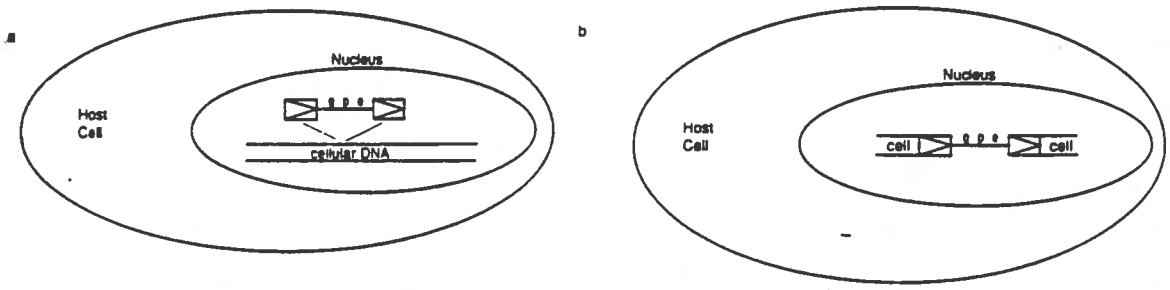

Fig. 2. Normal integration of a retrovirus after being synthesized into dsDNA. a) The retrovirus seeks to integrate into the host cell's DNA. b) The retrovirus integrated in the host cell DNA. Letter g, p and e stand for *gag pol* and *env*, respectively. Packaging sequences are not shown here. Arrows within LTRs designate polarity.

The 5' portion of the upstream LTR (the U3 region) functions as a promoter and enhancer to initiate transcription of the viral genome by normal enzymatic machinery (1,2). RNA transcription proceeds from the 5' R region of the upstream LTR through the rest of the viral genome until the 3' R region is reached, signaling the cessation of transcription (fig. 3) (1,2).

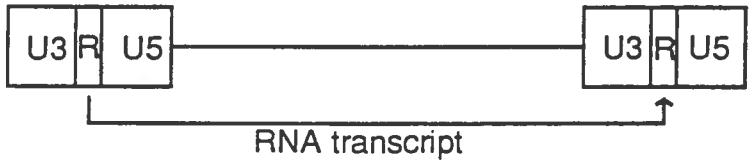

Fig. 3 Retroviral RNA synthesis

The RNA transcript may now be processed as normal cellular messenger RNA (including a cap and a poly A tail). The RNA now has one of two fates. If the RNA is appropriately spliced, multifunctional viral proteins (called *gag*, *pol* and *env*) will be synthesized which aid in packaging, infection, replication and integration of the viral genome (1,2). Alternatively, the RNA may be left unspliced and subsequently utilized as the viral genome itself. Two copies of the viral RNA (+ strand) genome are packaged together with the help of a cellular tRNA molecule which is co-

packaged (1,2). Packaging sequences found downstream of the 5' LTR, but before the *gag* region, are essential for efficient packaging (fig. 4). This region is called "Psi" and is indicated by the Greek letter Ψ (fig. 1).

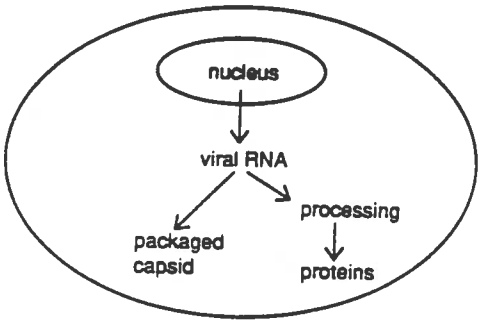

Fig. 4. The fate of RNA after transcription from the provirus.

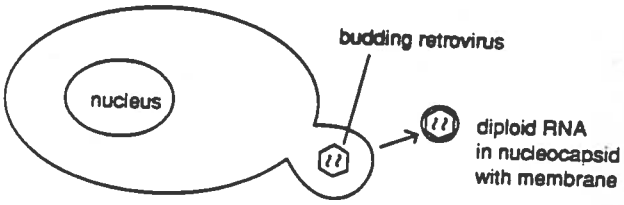

Fig. 5. Budding of the packaged retroviral RNA from the cell membrane.

The *gag* and *env* proteins (forming the core and outer envelope proteins respectively) function as the main structural elements of the virion (1,2). The virion is released from the host cell in a non-lytic fashion, acquiring host cell membrane as part of the viral envelope (fig. 5) (1,2). In a receptor specific fashion the *env* protein mediates infection of the next host cell. Upon infection of the second host cell the *pol* protein (called reverse transcriptase) begins to transcribe the viral RNA into proviral DNA (fig. 6) (1,2). The sequence of events involved are complex (2). Synthesis is initiated from a Primer Binding Site (PBS) in the packaging region which utilizes the copackaged tRNA as the primer to synthesize (-) strand DNA from the PBS toward the 5' R region. Because the R regions are identical at both ends of the RNA genome, the tRNA/DNA hybrid with R sequences can now "jump" to the 3' R region and act as a primer for DNA synthesis towards the 5' end of the viral RNA

molecule, thus completing (-) strand DNA synthesis. As outlined in figure 6, reverse transcription proceeds with similar use of primers and "jumps" as well as enzymatic degradation of the RNA genome to create the double stranded DNA described at the start which is ready for integration.

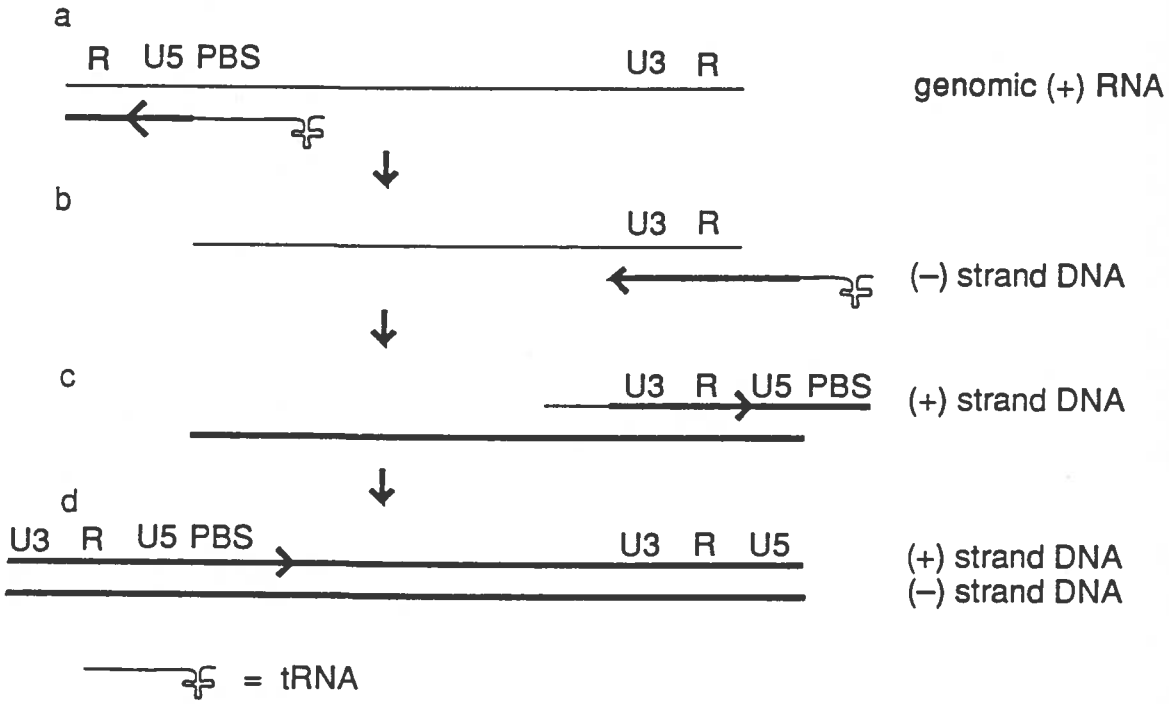

Fig. 6 Abbreviated mechanism of reverse transcription. a) The co-packaged tRNA binds to the PBS, serving as a primer for (-) strand DNA synthesis. b) Minus strand DNA synthesis continues after the tRNA/DNA strand "jumps" to the 3' (+) strand RNA R region. c) After being partially digested by reverse transcriptase (+) strand RNA serves as a primer to initiate (+) strand DNA synthesis. This (+) strand DNA then jumps to the opposite pole of the (-) strand DNA, hybridizing at the PBS. d) Complete dsDNA of retrovirus.

As will be outlined later, recombination between homologous retroviral elements may occur during reverse transcription such that information between packaged heterodimers can be exchanged. It has been suggested that such recombinations increase retroviral viability and should therefore be considered part of the normal life cycle (4).

Models of Retroviral Gene Transduction

What has been described thus far is the normal retroviral life cycle. However, there are aberrations of this life cycle which are relevant to the question of gene transduction. This involves two independent recombinatory events. The first event requires aberrant integration of proviral DNA into the host genome wherein the 3' LTR is lost, thus placing the remaining leftward viral elements, including the 5' LTR, upstream of *p-onc* sequences (fig. 7) (5,6,7).

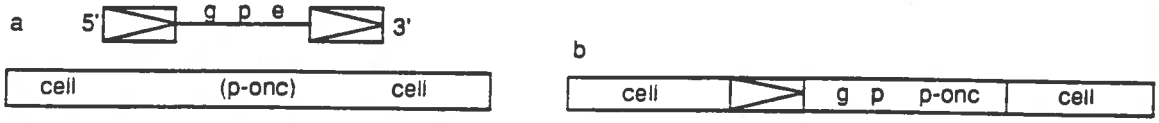

Fig. 7. Aberrant integration of retrovirus. a) Retrovirus is depicted before integrating into target cell DNA near a *p-onc* sequence. b) Retrovirus integrates in aberrant fashion losing its 3' end, situating itself just upstream of the *p-onc* sequences. As an example, *env* is lost in this event. The symbols g, p and e stand for *gag*, *pol* and *env* coding regions respectively. Packaging sequences not shown here.

However, this integrated, mutated provirus is defective in two regards. First, because it has lost its downstream genome, it has most likely truncated at least one of its viral genes, thus disabling its ability to replicate. (An exception to this is the Rous Sarcoma Virus which has the *v-src* oncogene downstream of a full complement of viral genes). Second, because it has lost its 3' LTR, the virus would be unable to participate in reverse transcription (see above) and would thus be unable to integrate into the genome of the next cell to be infected even if all the appropriate viral proteins were present. For both these reasons the integrated, mutated provirus therefore needs the aid of another, similar retrovirus (called a helper virus) to "rescue" it from the host genome. By gene

complementation, the helper virus provides the requisite gene products necessary for packaging, transport and processing of the transducing virus transcript. Additionally, through a second recombinatory event, the helper virus provides the transducing virus with a 3' LTR necessary not only for reverse transcription, but integration as well. Because the elements undergoing recombination are dissimilar (i.e. random cellular sequences recombining with viral sequences), the processes concerned utilize non-homologous recombination (6,7,8). It is at that point (the second recombination) that the two theories diverge.

The so called "DNA model" holds that the helper virus rescues the integrated, truncated provirus when the former integrates downstream of the latter (fig. 8 a,b) (8,9). Now, because both LTRs are identical, the 5' LTR of the helper virus functions as the 3' LTR of the truncated provirus, thus giving it all the functional components necessary for reverse transcription in the next cell. In the process a cellular gene (in this case the *p-onc*) is placed between two retroviral LTRs and the sequence has been transduced. Additionally, gene products requisite for the viral life cycle are provided for by the helper virus. Similarly, although less likely, were the helper virus to also integrate in an aberrant fashion downstream of the truncated provirus, the helper virus would contribute its 3' LTR as well as some upstream helper sequences to rescue the truncated provirus along with *p-onc* (fig. 8c) (9). It is also conceivable that the first retrovirus might integrate normally upstream of a *p-onc* sequence thus providing the upstream LTR (fig. 8d). A subsequent integration by a helper virus downstream as

described above would again place the *p-onc* sequence between two LTRs, thus giving rise to a transduction.

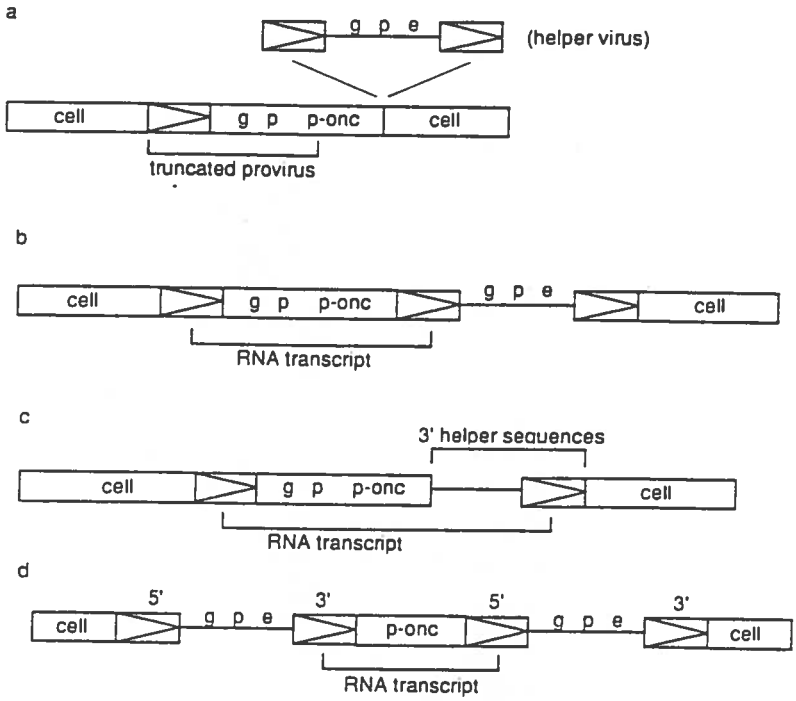

Fig. 8. DNA model of LTR recovery. a) and b) Helper virus integrates downstream of the integrated, truncated provirus thus providing the downstream LTR. c) and d) are possible alternative methods of recombination described in text. Packaging sequences not shown here. Again, the *env* gene is lost as an example.

In contrast, the so called "RNA model" holds that rescue of the truncated provirus occurs post-transcriptionally after infection of the secondary cell (6,7,10,11). In this model RNA which has been transcribed from the 5' LTR of the truncated provirus through the *p-onc* sequences is successfully packaged into a virion with the help of gene products provided by the helper virus. Because retroviruses are diploid, two different viral RNAs can be packaged at once, thus allowing a heteroduplex of helper and mutant RNA to be packaged at the same time (1). It has been proposed that during the process of reverse transcription in the secondary cell, viral sequences are recombined to restore the 3' LTR of the truncated retrovirus thus

placing the *p-onc* again between two LTRs (6,7). The main mechanism proposed for this is "copy choice" of RNA transcripts during reverse transcription (fig. 9) (4,10). (Other mechanisms, such as strand displacement {see below for details} {11}, have also been suggested, however no evidence to support such notions have been forthcoming).

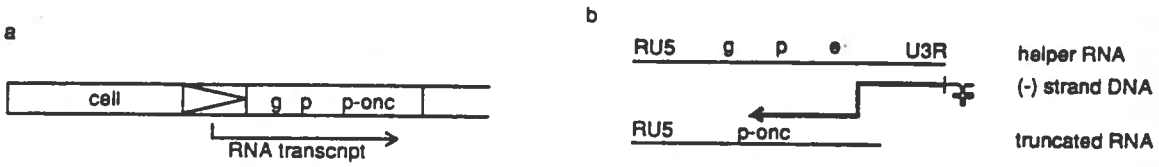

Fig. 9. Copy choice model of LTR recovery. a) RNA is transcribed from the upstream LTR and through the *p-onc*. b) During (-) strand DNA synthesis in the second host cell reverse transcriptase jumps from the helper RNA to the RNA template created by the truncated retrovirus. This is done after the helper LTR has already been transcribed but downstream of the picked up *p-onc* sequences. The tRNA acts as a primer at the downstream R region after the first jump. DNA is in bold lines while RNA is in thin lines. The symbols g, p and e stand for *gag*, *pol* and *env* coding regions respectively. Packaging sequences not shown here.

Relevance of Homologous Recombination

The question being discussed in gene transduction is how non-homologous sequences (one viral and one cellular) might recombine to rescue a transducing virus. This is important because once significant homology is introduced, the rate of recombination increases -- recombination rates reaching as high as 40% *in vitro* (11). Even as little as 50% homology between sequences shows significant recombination (9). Furthermore, it is recognized that these homologous recombinations occur at the level of reverse transcription via either copy choice (see above) or a strand displacement mechanism (4,11,12). In the latter case, the two dsDNA strands which are made after reverse transcription invade

each other, thereby cross hybridizing, giving rise to "H" structures of DNA (so named because of appearance of the cross over structures by electron microscopy)(4). Upon resolution of the cross hybrid, homologous sequences have allowed for the incorporation of non-homologous sequences from the other viral genome to be assimilated through mismatch repair (4,11) (fig. 10). This second mechanism, however, implies that there are two intact retroviruses in order for two separate strands of dsDNA to be created (because two LTRs are needed for synthesis) and that regions of homology bound the transduced cellular elements. Therefore, this mechanism is not germane to gene transduction. It should not be overlooked, however, that the benefits of homology could also be conferred upon constructs which might tend to recombine at the DNA level as is recognized with regular chromosomal recombinations.

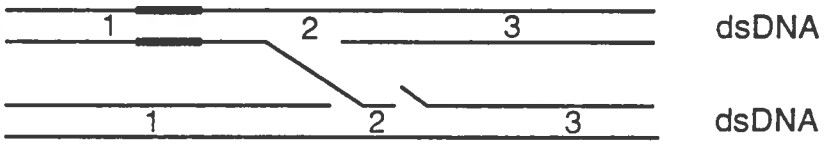

Fig. 10 Mechanism of homologous recombination among retroviruses using the strand displacement model. After synthesis of dsDNA by reverse transcriptase, one strand of DNA invades and cross hybridizes to a homologous region (depicted as "1, 2 and 3") on the other retrovirus. Incorporation of the non-homologous (bold-faced) region between homologous regions 1 and 2 is accomplished through mismatch repair.

While there has been much investigation on the mechanisms of homologous recombinations between retroviruses at the RNA level, the system involved in gene transduction does not naturally involve homologous elements. In contrast to the high rates noted above for homologous recombination, the rates of recombination among non-

homologous elements may be as low as .005% per replication cycle (13). Therefore drawing conclusions based on analogies between these two discordant systems is not helpful.

Experimental Design of Previous Research

The general experimental design for investigations which have examined retroviral recombination are essentially the same (5,8,9,13,14,15,16). Because it was recognized that the first step of recombination according to either model was the placement of an LTR upstream of a *p-onc* sequence, this basic structure was utilized as the starting point for experiments. A plasmid was constructed which contained an upstream LTR, a packaging sequence, and a selectable marker (such as an oncogene) from 5' to 3' respectively (fig. 11a). No downstream LTR is present. The plasmid DNA was then introduced into the cellular genome by transfecting the DNA into cells grown in culture. The gene product was synthesized efficiently because the marker is downstream of the LTR, which functions as a promoter and enhancer. After some time, cells which have taken up the plasmid DNA and appropriately express the marker gene will form a focus of transformed cells which are distinguishable from their non-transformed neighbors. These foci (called the I° cells) were picked and grown up (fig. 11b). These cell clones were as yet unable to produce any virus.

These I° cells were then infected with a helper virus which was appropriate in host range for the cells being utilized (fig. 11c). The helper virus had no selectable marker, but did have a complete retroviral genome. Through complementation, helper virus and

transducing virus genomes were elaborated into the culture medium. After a few days the supernatant (called the I° supernatant) was harvested and an aliquot was added to fresh cells. After some time new foci arose in these cells (called the II° foci) and they were picked as above and analyzed for evidence of recombination, usually by Southern Blot analysis or PCR (fig. 11d).

The only way that the original transfected construct could have successfully integrated into the next host cell genome to cause the selectable marker to be expressed was if the 3' LTR had been recovered from the helper virus by a recombinatory event. Without the 3' LTR, integration would be impossible because a DNA copy could not be made from the (+) strand RNA.

Clearly there are variations on this approach, but in general they all follow this basic model.

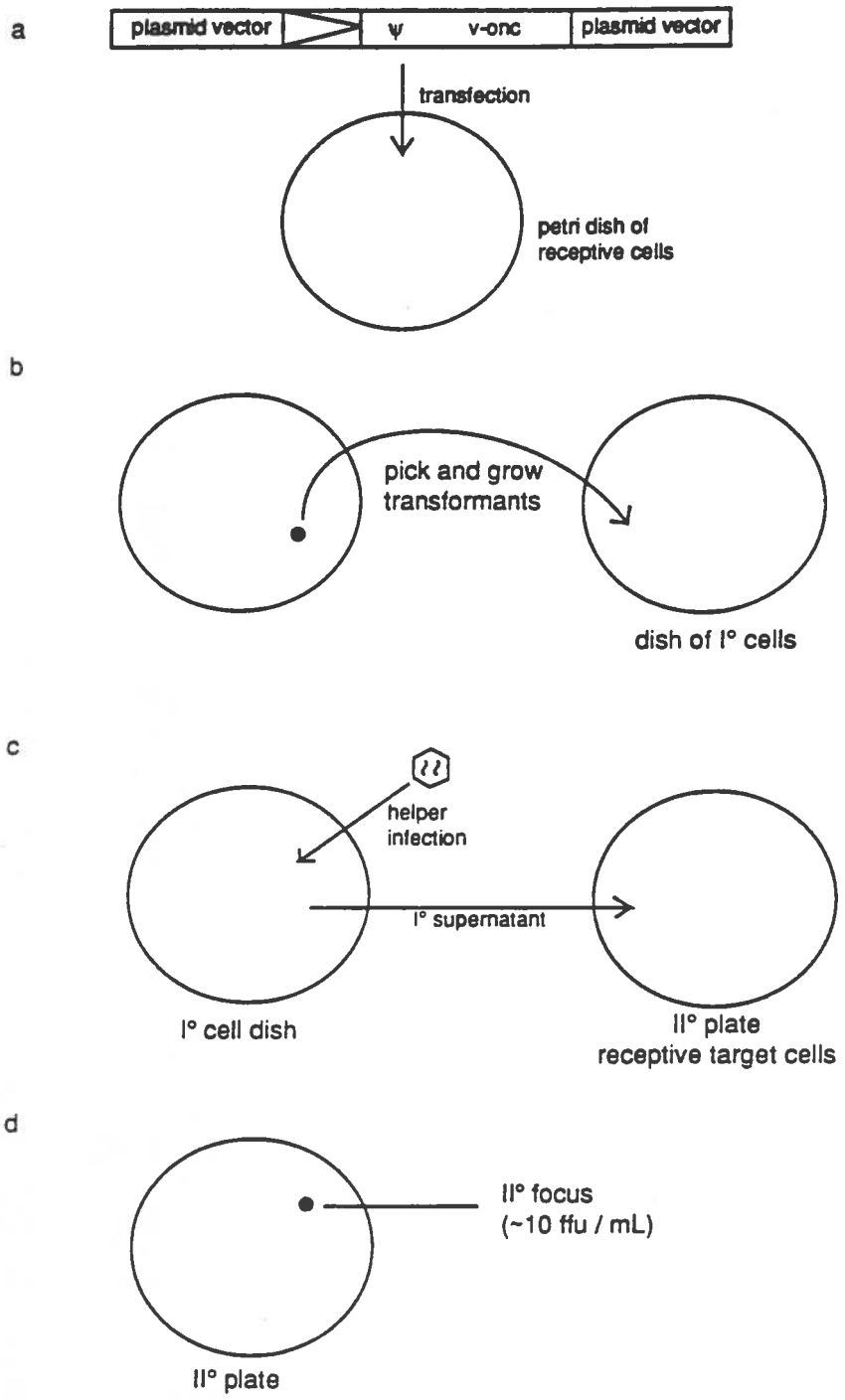

Fig. 11. General experimental design. a) A truncated retrovirus is introduced into a petri dish of receptive cells via transfection. b) Foci expressing the oncogene (and therefore having successfully taken up the desired construct) grow and are picked and expanded. c) A helper virus infects these primary cells. The primary supernatant is harvested and used to infect fresh target cells. d) Secondary foci arise (approx. 10 ffu/ml) signifying a recombinatory event and 3' LTR recovery.

Results of Prior Research

For some time it has been assumed that the second recombinatory event in retroviral gene transduction involves either the DNA mechanism or the RNA mechanism, both of which have been outlined. When these studies began over 10 years ago a "surprisingly high frequency" (on the order of 1,000 ffu/ml) of apparent recombinants were noted, values inconsistent with what was assumed to be a relatively rare event (5,6). It was later pointed out that the large number of recombinants were due to transfection artifacts and that the conclusions drawn from prior experiments were likely false (8,9). During transfection of the truncated viral genome (described above) artifactual recombination among the transfected elements is known to take place while DNA is integrating into the host genome (9). Therefore, as proviral DNA is introduced into the cell a truncated provirus may artifactually recombine with another truncated provirus thus recovering the downstream LTR far before a helper virus has been introduced into the system.

Another problem is that many experiments fail to examine the intermediate steps in the recombination process. Nonetheless, investigators attempt to draw conclusions about the nature of these intermediate steps by looking at the final products of recombination (5,8,9,13,14,15,16). In this regard, inferences about presumed steps in the mechanism of recombination are based on the analysis of the final product (i.e. by analysis of the II° foci--well after recombination is presumed to have taken place in both models).

Lastly, studies are often designed to support one model or the

other. In reality, most data offered in reports are not specific and could be explained by either model (5,8,9,13,14,15,16). In other instances, data is presented which directly contradicts the theory that is being favored (16). And still in other cases, reports which do have strong evidence for the theory hypothesized also present data that supports the opposite view (13).

While no definitive conclusions can be drawn, recent evidence lends credence to both the RNA and DNA models. Direct evidence for the RNA model is twofold. Firstly, while recovery of helper virus elements upstream of the 3' LTR could take place with aberrant helper virus integration into DNA, this is not generally considered to be a common event. However, in one study nearly 25% of the recombinants recovered showed evidence of such recovery upstream of the 3' LTR (13). Through copy choice this type of recovery would be more common because strand switching could theoretically take place at any point in minus strand synthesis. More importantly, these recombinants were of differing lengths. Therefore, they arose from separate recombination events--a fact which tends to support the RNA rather than the DNA model. However, while these points are indicative of the RNA model they do not conclusively prove the RNA model. As explained, recovery of sequences upstream of the 3' helper LTR is possible in the DNA model as well (9). If the system in question demonstrated some increased proclivity for aberrant helper integration and/or if helper infections were performed in separate experiments, then it would be possible to record multiple, separate recombination events in the fashion described.

Secondly, and more convincingly for the RNA model, in transfected constructs which were missing the 3' LTR, a few recombinants were recovered which had significant stretches of adenine (A) residues sandwiched between the original construct and the recovered LTR (13,15). In one report the A's correctly followed a poly-adenylation signal which had been placed at the downstream end of the originally transfected construct (13). This indicates the recombination seems to have taken place post-transcriptionally, after poly-adenylation of the RNA. Recovery of the LTR could have taken place only after DNA recombination is theorized to occur. Therefore, recombination during reverse transcription appears to be the only explanation for these recombinants. However, that report did not exclude the possibility that pseudogenes may arise from poly-adenylated, reverse transcribed and re-integrated retroviral sequences. Subsequent integration of a helper virus downstream of the poly-A sequence at the DNA level could then initiate rescue. In fact, those constructs with polyadenine residues also had only an LTR with no upstream viral sequences incorporated (13). This pattern of recombination is more indicative of DNA recombination, as will be described below.

One of the major criticisms of the DNA model is that helper virus integration downstream of a truncated provirus is not sufficient for rescue. It has been argued that a polypurine track (ppt) just upstream of the 3' LTR in all retroviruses is required in the replication cycle (1). Without the ppt it is said that plus strand synthesis during reverse transcription can not occur effectively because the ppt serves as an essential primer sequence. The

truncated virus as it would exist in the primary cell would be unlikely to have this track and therefore the virus would still not be rescueable. However, recent evidence has shown that rescued proviruses without a ppt do exist (in up to 50% of recombinants) and it has been suggested that further recombination events or mutations which might occur at a later time would select for viruses that regain their ppt (13). In addition, there is evidence that viral constructs created through molecular cloning techniques which do not have a ppt and were transfected into NIH3T3 cells are capable of being transmitted to fresh 3T3 cells (Duesberg and Schwartz, unpublished data). Therefore, this argument against the DNA model is not valid.

More direct evidence that the DNA model is operating is a study which examined the effect of packaging site deletions on the frequency of recovery of a 3' LTR from a helper virus. They showed only a 5 to 50 fold reduction in recombination events when compared to a 1,000-10,000 fold reduction in packaging (16). As will be outlined later, these data are consistent with the DNA rather than the RNA model.

More interestingly, and as alluded to above, one paper reported that nearly 50% of the recombinants had recombination junctions which included only the helper virus LTR with no other helper virus elements as determined by PCR (13). No consensus sequences or homologies between the helper virus LTR and the truncated provirus were described to explain why copy choice might induce recombination exactly at the 5' end of the helper LTR (13). The most reasonable interpretation of this phenomenon is that normal helper

integration downstream of the integrated proviral sequences provided the LTR, as proposed by the DNA model. No helper sequences upstream of the helper 3' LTR could be recovered in the recombinant and therefore these elements were absent in nearly 50% of the recombinants analyzed. In reality, because PCR was used to amplify the recombinants, recombinations which took place such that the primers were far apart would make their detection technically less likely (17). Therefore, the incidence of such recombinants that contain only the helper LTR and no other helper sequences is probably even higher than reported. However, this evidence is also circumstantial and does not lend definitive proof for the DNA model.

In summary, there are no solid conclusions which can be drawn about the mechanisms of retroviral gene transduction because of the lack of unambiguous evidence. In reality, much of this paradoxical data may be resolved once it is considered that both models are not mutually exclusive and that both may in fact be operating.

The issue of DNA verses RNA recombination may be a bit of a moot point altogether when considering the recognized role of homologous recombination during reverse transcription. It has been pointed out that regions which share up to 50% homologies have a greatly increased rate of recombination during reverse transcription. However, it has also been reported that regions of homology between the two viruses as short as 5 to 8 base pairs might be sufficient in copy choice (13,16). Therefore, homologous recombinations between nucleotide stretches that are no longer than

common restriction sites might serve as the true mediators of efficient recombination.

Proposed Experiments

The hypothesis of this report is that retroviral gene transduction occurs at the DNA level. Experiments were primarily designed to provide experimental evidence for the DNA model of recombination. Unlike prior research, the experimental designs presented here were based on a novel set of approaches which examined the intermediate steps in recombination. The experimental designs were essentially identical to what is depicted in figure 11 with pertinent modifications noted below.

The first set of experiments was modeled after a construct created by Stuhlmann et. al. (16) wherein an intact retrovirus with both LTRs, but no packaging site, was introduced into cells via electroporation, a transfection method which introduces DNA into cells in low copy number. Because there is no packaging site, packaging efficiency is greatly reduced, from 1,000 to 10,000 fold (16). Recombination with the helper virus with recovery of the packaging site would reconstitute packaging efficiency. Our system utilized the Harvey Sarcoma Virus (HaSV) containing the *ras* oncogene. The packaging site between the 5' LTR and the oncogene was removed (see Material and Methods) (fig. 12) and cloned into a plasmid vector (the construct was name p$\Delta\Psi$(-)A). Rescue was attempted with Moloney Murine Leukemia Virus (MoMLV) which gave rise to recombinant secondary foci. Biological evidence of packaging site recovery in secondary foci was then determined by high titer

transformations of C3H cells during tertiary infection. The MoMLV and HaSV genomes share no significant homology in the region between the LTRs; however, the LTRs are nearly identical. C3H mouse fibroblast cell lines were utilized as the target cells in all experiments.

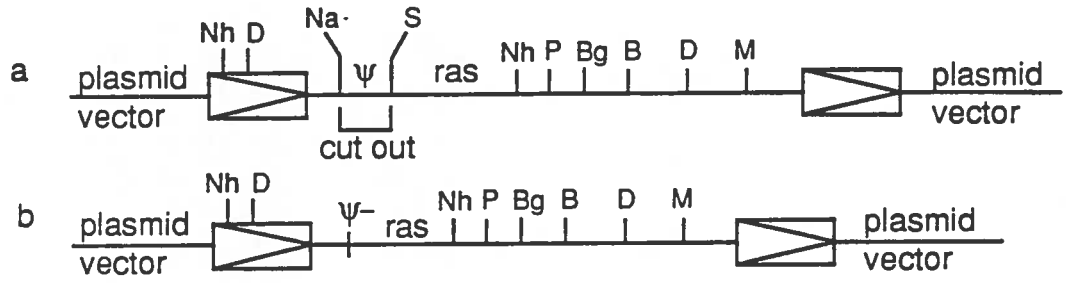

Fig. 12. a) HaSV and b) pΔΨ(-)A genomes with appropriate restriction sites. Na, S, P, Bg, D, Nh and M stand for NaeI, SstII, PstI, BglII, DraI, NheI and MstII restriction sites, respectively. See material and methods for pΔΨ(-) A construction. Not drawn to scale.

The RNA model dictates that the RNA transcript must first be packaged before it is capable of recombining. Therefore, if the RNA model were operating, a reduction in packaging would be expected to exhibit a parallel reduction in recombination. In other words a 10,000 fold reduction in packaging would give a 10,000 fold reduction in II° foci formation.

In contrast, the DNA model dictates that recombination occurs prior to packaging. Therefore, were a single I° cell containing the packaging site minus {Ψ(-)} construct to recombine with the helper virus and recover the packaging site, it would now selectively be packaged efficiently. However, this single cell would still be able to contribute only a relatively small number of virions compared to the entire plate of cells ($5x10^6$ cells) which are elaborating their viruses 10,000 fold less efficiently. On average, therefore, the

packaging efficiency still appears to be about 10,000 fold lower for that entire plate. The DNA model then predicts that a 10,000 fold reduction in overall packaging does not necessitate a 10,000 fold reduction II° foci formation.

Another avenue for investigating which mechanism is operating relies on looking at the products of recombination. Because DNA recombinations occur at the level of the I° cell the rescued construct is fixed at the DNA level and each recombinant progeny would be of the same length. In contrast, because the RNA model theorizes recombinations occur independently of each other in separate II° cells, each recombinant should be of a different length.

These experiments explore the phenomenon of recombination in general but do not specifically answer the question at hand, that is, how does a provirus with a truncated 3' LTR recover that downstream LTR and include intervening oncogenic sequences? To answer that question by looking at an intermediate step in the recombination process, a novel construct was created where an intact helper virus was placed just upstream of an intact packaging site and oncogene (i.e. a truncated provirus) (fig. 13). Therefore both the helper and truncated virus' RNAs and gene products may be produced from their respective upstream LTRs because the 3' LTR of the helper also functions as the 5' LTR of the truncated provirus. In this system an intact AKR virus genome was placed upstream of the 5' portion of the HaSV genome (fig. 13). The AKR virus provided the helper sequences while the HaSV provirus which is missing its 3' LTR functions as the truncated provirus. The 3' LTR of the AKR

sequences is therefore the same as the HaSV 5' LTR (see Materials and Methods).

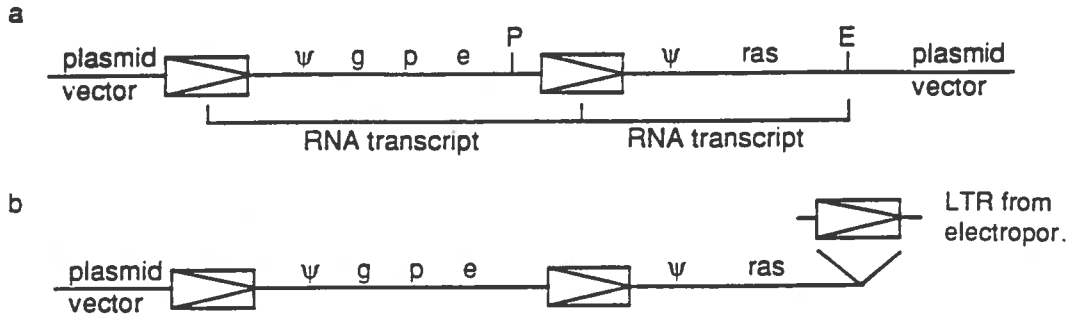

Fig. 13. a) pEMJ construct with important restriction sites and sites of RNA transcription. P and E stand for PstI and EcoRI restriction sites. b) pEMJ with an LTR being transfected downstream of *ras*. Letters g, p an e stand for *gag*, *pol* and *env* respectively. Not drawn to scale.

Recovery of a downstream LTR by the truncated provirus would constitute a rescue. The advantage of this system is that no helper virus infection is required. We need not concern ourselves with where the helper virus integrates because we know it is upstream of the truncated provirus. Therefore, DNA recombination is virtually ruled out by this system. Because RNA transcripts of both viruses are being produced, the only possible avenue for recombination is at the RNA level during secondary infection. DNA recombination via retroviral integration may be simulated, however, if a second LTR is introduced into cells containing the construct via a second electroporation procedure (fig. 13). Integration of this LTR downstream of the truncated provirus would simulate helper virus integration, thus providing rescue and evidence for a DNA mechanism.

Materials and Methods

Cell culture

Complete cell culture medium consisted of Dulbecco's Modified Eagles Medium (Sigma Chemical Co, St. Louis, MO) supplemented with 50 U/ml nystatin, 100 U/ml penicillin G,. 100 mcg/ml streptomycin, 0.5 mcg/ml amphotericin B, 20 mcg/ml gentamycin, 60 mcg/ml L-glutamine and 10% fetal bovine serum. The 2% medium was exactly the same except that 2% fetal bovine serum was the supplement.

All transformed cell clones were maintained in 2% medium while non-transformed or newly transformed cells were maintained in complete medium as specified above. Culture medium was changed every two to three days. Cells were grown on 100mm Falcon petri dishes (Fisher Scientific, Pittsburgh, PA) in a 37°C incubator supplemented with 6.5% CO_2.

Harvesting of cells from petri dishes was performed with 1x trypsin-EDTA (GIBCO/BRL, Gaithersburg, MD). Cells were washed with 2-5ml trypsin solution for one minute. Cells were then harvested from petri dishes by adding fresh trypsin for 3-6 minutes at 37°C.

Determination and Selection of Focus Forming Units

Focus forming units were assessed on 100mm petri dishes. Retroviral titers were determined visually (fig.14). Foci selected for cloning were removed with the use of glass cloning wells and trypsin solution. Cloned foci were initially propagated in complete medium for about one week, but thereafter in 2% medium.

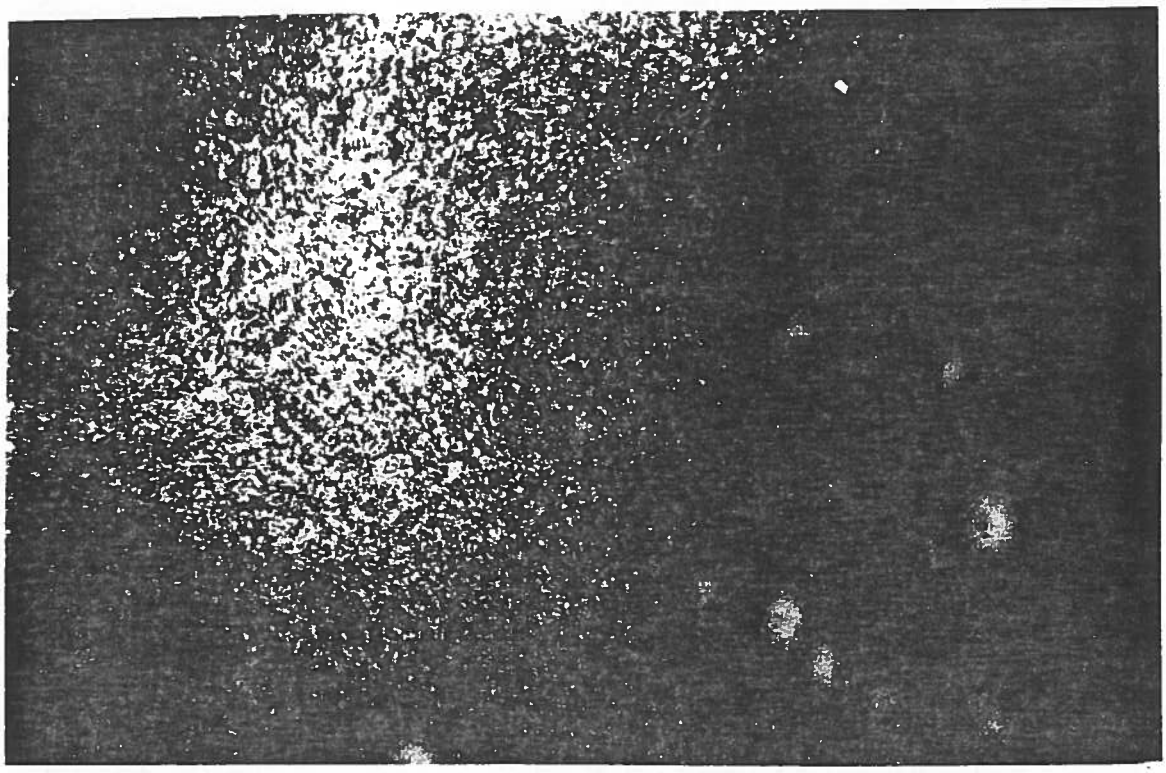

Fig. 13. The *ras* focus on the left is more refractile and has cells pilled upon one another. Contrast with the normal monolayer of C3H cells on the right. 20x magnification.

DNA Extraction

DNA extraction was performed by standard protocol (18). Cells were digested overnight in a proteinase K/SDS solution and two equal volume phenol/chloroform (1:1) extractions were performed followed by a single equal volume chloroform extraction. DNA was then precipitated in 95% ethanol and resuspended in 100-200 mcl of 1xTE buffer.

DNA Fragment Purification

DNA fragments which were purified for ligation procedures or probe purification were cut with enzymes as indicated below. DNA was fractionated on .6 to 1.2% agarose gels run at 20-40 V

overnight. The desired bands were visualized with ethidium bromide and was cut out of the gel. The DNA was then purified using the Prep-a-gene system (BioRad, Richmond, CA).

Southern Blotting and Hybridization

DNA (10-15 mcg) was digested with indicated restriction enzymes (New England Biolabs, Beverly, MA) as per manufacturer's instructions in a 3 to 5 times (enzyme units to mcg DNA) over-digestion for 12-24 hours. The digests were fractionated on .9-1.3% agarose gels run between 20 and 40V overnight. DNA was blotted onto a Nytran membrane (Schleicher and Schuell, Keene, NH) as per manufacturer's instructions.

Filters were probed with a 605 bp v-*ras* sequence from HaSV purified as described above from a HindIII/NheI digest of pHa#7. The probe was labeled by random priming with a multiprime labeling kit (Amersham, Arlington Heights, IL) as per manufacturer's instructions. Prehybridization of filters was carried out as per Schleicher and Schuell's instructions (42°C for 2 hrs). For hybridization the probe was denatured in 2.5ml of 100% formamide for ten minutes and an equal volume of 2x hybridization buffer (manufacturer's formula) was subsequently admixed. Hybridization was performed for 20-24 hours at 42°C. Washing of filters was performed as per manufacturer's instructions.

Plasmid Constructions

It should be noted that a lower case "p" refers to plasmid. For example pHa#7 is read as plasmid Ha#7.

Construction of the Ψ minus plasmid utilized pHa#7 (8), which has a partially redundant Harvey Sarcoma Virus genome cloned into a pBR322 vector (fig. 12). There is a 930 bp stretch of DNA between the U5 and the *ras* start codon corresponding to the packaging sequences of HaSV. The 9.2 kb pHa#7 was cut with SstII (GIBCO/BRL, Gaithersburg, MD) at the only site 940 bp into the HaSV genome. The DNA was then partially digested with NaeI (a blunt cutter) and the DNA was separated on an agarose gel. A DNA fragment nearly 8.5 kb (corresponding to the NaeI site 214 bp into the HaSV genome) was cut from the gel and purified as indicated above. The 3' overhang of the SstII cut was then filled with Klenow fragment (New England Biolabs, Beverly, MA) fragment and T4 ligase (GIBCO/BRL, Gaithersburg, MD) was used to perform a blunt end ligation. Ligated plasmid was used to transform HB101 competent E.Coli (GIBCO/BRL, Gaithersburg, MD) as per manufacturer's instructions. Verification of desired construct was confirmed by multiple restriction enzyme digests. The resultant plasmid contains a mutant HaSV genome with 726 base pairs of its packaging site missing. The resulting plasmid was nearly 8.5 kb and was named pΔΨ(-)A (fig. 15).

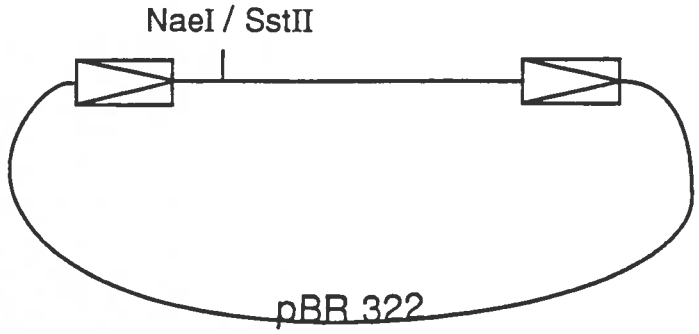

Fig. 15 Ligation sites for pΔΨ(-)A construct.

Construction of the plasmid containing HaSV sequences downstream of a Helper virus utilized pHa#7 and pAKR/NheI (fig. 13). The latter plasmid has an AKR Murine Leukemia Virus genome which was isolated from genomic DNA which contained proviral DNA. Approximately 400 and 500 bp of undefined cellular DNA flanks the 5' and 3' LTRs respectively. The viral DNA was cloned into a pBR322 vector at an EcoRI cut in the cellular sequences downstream of the 3' LTR and a HindIII cut in the cellular sequences upstream of the 5'LTR for a total linear length of 14.1 kb.

The creation of the desired plasmid involved a four piece ligation. Fragment #1 (nearly 6.4 kb) was isolated from an SstII/EcoRI double digest of pAKR/NheI which spanned from the EcoRI site in pBR322 through the 5'LTR and into the AKR *gag* region. Fragment #2 (over 6.6 kb) was isolated from a SstII/PstI double digest of pAKR/NheI extending from the *gag* region to the border of the 3' LTR. Fragment #3 (nearly 2.3 kb) was isolated from a XmaIII/PstI double digest of pHa#7 which spanned from the PstI site at 3436 bp to the XmaIII site just downstream of the 3'LTR. Fragment #4 (over 2.3 kb) was isolated from a XmaIII/EcoRI double digest of pHa#7 which went from the XmaIII site downstream of the 5' LTR to the EcoRI site downstream of the *ras* gene.

These four fragments were isolated on agarose and purified as described. Ligation was performed by T4 ligase(GIBCO/BRL, Gaithersburg, MD). The ligation product was used to transform HB101 E.Coli cells and verification of desired construct was confirmed by multiple restriction enzyme digests of selected clones. The plasmid was over 17.5 kb and was named pEMJ (fig. 16). The

construct contained an AKR genome which was complete from its 5' LTR through its coding region and up to the border of its 3'LTR. A HaSV LTR then formed the 3' LTR of AKR while at the same time functioning as the 5' LTR of HaSV. The HaSV genome was intact through the *ras* coding sequences but had no 3' LTR.

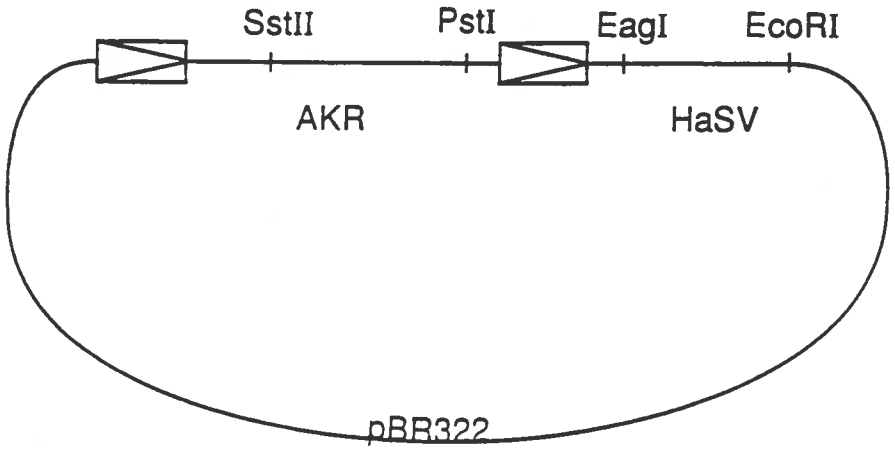

Fig. 16 Ligation sites for pEMJ construct. Not drawn to scale.

Construction of a plasmid appropriate for the second electroporation phase of the experiments designed to examine 3'LTR recovery was created through a single intramolecular ligation. pAKR/NheI was cut with AatII such that the single site in the pBR322 vector and the single site in the AKR genome at 381 bp were intramolecularly ligated. The fragment containing the 5' LTR of AKR and the ampicillin resistance gene in pBR322 was intramolecularly ligated. This plasmid was named pAKR/LTR.

Large Scale Plasmid Preparation

Bacteria containing the plasmid to be purified were inoculated into 1 L of Luria Broth and incubated overnight at 37° C, shaking at 225 rpm. Cells were pelleted at 7,000 rpm in 250 ml GSA tubes for

10 minutes at 4°C. Cells were resuspended in 40 ml of a 50 mM glucose, 25mM Tris-Cl, 10 mM EDTA solution containing lysozyme (10mg/ml). After a ten minute incubation, 80 ml of a 0.2 N NaOH solution containing 1% SDS was added and the mix was incubated on ice. After 10 minutes 60 ml of a 3M potasium acetate, 2 M glacial acetate solution was added and left on ice. After 10 minutes the mix was spun for 10 minutes (4°C) at 10,000 rpm. The supernatant was collected and precipitated in 95% ethanol. The precipitant was further purified by two CsCl spins and ethanol precipitation as is described in standard protocols (18).

Psi Minus Experiments

One million actively dividing, fresh C3H cells (a mouse fibroblast line) were electroporated (BioRad electroporation chamber, 280 or 310 Volts) with 50 ng of pΔΨ(-)A DNA which had been linearized with AatII (which cuts in the pBR322 sequences only). The surviving cells were plated onto four 100mm petri dishes in complete media. Cell were thereafter maintained in 2% media. Foci phenotypically consistent with *ras* transformation arose 7 to 10 days later and were picked 15 to 20 days after electroporation. These primary (I°) foci were expanded and labeled pΔΨ(-)A"accession #"I° (e.g. pΔΨ(-)A5I°).

For recombination experiments primary cells were plated in complete media such that they would be 50-70% confluent by the next day. Each plate was then superinfected with 1.0 ml of Moloney Murine Leukemia Virus stock harvested from C3H cells. Media was changed once, and on the fourth or fifth day after superinfection

10ml of complete media was seeded per plate. The medium (i.e. the primary supernatant) was collected twenty-four hours later and clarified in a table-top centrifuge at 1,750xg for 5 minutes.

Secondary infections were then performed which gave rise to secondary foci and hopefully a gainful recombination. Fresh C3H cells which had been growing for 24 hours in complete media, and were 50-70% confluent, were infected with 1 ml of the primary supernatant. Complete medium was changed every two to three days. On the sixth to eighth day, 10 ml of complete medium was added per plate. The medium (secondary supernatant) was collected twenty-four hours later. Plates were then maintained on 2% media.

Tertiary infections were then performed which, upon observing high titers, served as an indication of Ψ recovery. Fresh C3H cells which had been growing for 24 hours in complete media, but were 50-70% confluent, were infected with 1 ml of the primary supernatant. Complete media was changed after two days but thereafter cells were maintained on 2% media. Tertiary foci arose within 8 to 9 days and bioactivity was scored as follows: 0=no tertiary foci; 1+= less than 10 tertiary foci; 2+= 11-25 tertiary foci; 3+= 26-100 tertiary foci; 4+= greater than 100 tertiary foci.

Tertiary plates which had a bioactivity rating of 4+ (but sometimes 3+) then had their corresponding secondary plates examined. Secondary foci were picked from those corresponding plates approximately 15 to 18 days after secondary infection. These secondary (II°) foci were expanded and labeled pΔΨ(-)A"accession #"II°-"accession#" (e.g. pΔΨ(-)A5II°-13).

Tandemized Retrovirus Experiments

Fresh C3H cells were electroporated with 50 ng of pEMJ DNA which had been linearized with the single EcoRI site as described above. Cells were maintained as described above.

Foci were picked and expanded as described above and labeled pEMJ-"accession #" (e.g. pEMJ-6). Ten ml of complete media was added onto confluent plates, collected 24 hours later and centrifuge clarified (as described). As a screen, 1 ml of this supernatant was used to infect fresh C3H cells which were 50-70% confluent. After 8 to 9 days multiple foci arose in some of the cells which had been so infected. These cell clones were immediately excluded from further experimentation because, even though it was clear that some recombination had occurred the exact mechanism of such recombination could not be determined (e.g. it could have simply been an artifact of transfection). A second screen was then performed. One ml of supernatant from cell clones which were not competent to transmit a transforming virus in the initial screen was then used to infect 50-70% confluent HaSV non-producer cells (i.e. C3H cells which had the pHa#7 genome transfected into it, but had no helper virus present). After four days 10 ml of complete medium was seeded onto such cells, collected 24 hours later and centrifuge clarified as described. This HaSV producer supernatant was now added to fresh C3H cells as has been described. Foci arose 8-9 days later.

Now cell clones which had viable helper virus (as evidenced by the ability to turn HaSV non-producer clones into producer clones) and a viable 5' transducing virus sequence (as evidence by

transformation after electroporation), but had not undergone a recombination event (excluded by the initial screen) have been selected.

Such selected pEMJ cells clones were split and seeded onto petri dishes in 10 ml of complete media. Twenty-four hours later the supernatant was collected and the cells (which were actively dividing) were harvested. This pre-electroporation media was centrifuge clarified as described. One million pEMJ cells were electroporated (180 to 290 volts) with 25 to 30 micrograms of pAKR/LTR linearized with either PstI or NarI. Electroporated cells were then split evenly among four petri dishes with complete media. Four days later 10ml of complete media was seeded and collected 24 hours later. This post-electroporation supernatant was pooled and centrifuge clarified. One ml of pre- and post-electroporation supernatant was used to infect in parallel replicate dishes of C3H cells which had been actively dividing for 24 hours and were 50-70% confluent. Monitoring for foci took place over the next several weeks.

Results and Discussion

Packaging Site Recovery

Experiments which were designed to examine packaging site recovery were more consistent with the DNA model than the RNA model.

Most experiments performed yielded no recombinations as assayed by tertiary infection of C3H cells (see Materials and Methods). However, even in experiments where no recombination was evident, secondary foci still arose which represented a sort of "background" packaging of the original construct. Even though the transfected constructs were without a packaging site, the packaging efficiency was not zero. In the experiments which did exhibit recombinatory events these background foci made up the great majority of secondary foci (table 1). Overall, for those experiments which were indeed successful, the percent of recombined foci as compared to total foci was about 5-15% (table 2). A prior study showed that 66-71% of packaging minus constructs regained their packaging site through recombination (16). The higher rate may have been facilitated by homologous sequences which bounded the packaging region (16). Here, foci which were expanded and showed evidence of recombination on Southern Blot, maintained their ability to produce a high virus titer in tertiary infections whereas background foci did not (data not shown).

The secondary foci from five separate experiments which exhibited evidence of recombination were analyzed. Multiple recombinant foci arose in each initial secondary infection (from 2 to 5 per ml). More foci were harvested as needed by using the same

primary supernatant used for the initial secondary infection to infect fresh C3H cells. In those instances 0 to 2 recombinant foci per ml were recorded and titers for those experiments are detailed in table 1.

I° cell clone	II°dish#	III° plate score	# of recombinants confirmed by So. Blot	
a) pDY(-)A9	1	5	0	
	2	3	0	
	3	3	1+	
	4	4	4+	
	5	5	0	
	6	2	0	
	7	4	0	
	8	4	4+	
total		30		2
avg ffu/ml		3.8		.25
b) pDY(-)A10a	1	2	1+	
	2	3	1+	
	3	5	4+	
	4	4	0	
	5	0	0	
	6	3	0	
	7	3	0	
	8	4	3+	
total		24		1
avg ffu/ml		3.0		.13

I° cell clone	II°dish#	III° plate score	# of recombinants confirmed by So. Blot	
c) pDY(-)A5	1	14	4+	
total		14		2
avg ffu/ml		14		2
d) pDY(-)A10b	1	6	1+	
	2	2	1+	
	3	3	0	
	4	6	4+	
	5	3	4+	
	6	2	0	
	7	6	0	
	8	2	0	
total		30		3
avg ffu/ml		3.8		.38
e) pDY(-)A8 *	1	5	0	
	2	7	0	
	3	7	3+	
	4	6	3+	
total		25		4
avg ffu/ml		3.1		.5

Table 1: Secondary foci arising from five successful experiments (i.e. where recombination was noted) are detailed here. Primary supernatant harvested from superinfected primary clones was used to infect fresh C3H cells (i.e. a secondary infection). One ml was seeded per plate for lettered experiments a-c while 2ml was seeded per plate for experiment e*. See Materials and Methods for tertiary plate scoring.

Table 2 shows that packaging efficiency was reduced for each of the primary cell lines which were capable of producing secondary recombinants. As indicated this corresponds to a 100,000 fold reduction in packaging. (In comparison, MoMLV titers were only 3-10 fold lower as assessed by RNA dot blot analysis of secondary supernatants--data not shown). The reduction in packaging efficiency is at least an order of magnitude (and as high as two) greater than had been reported by Stuhlmann et. al. (16) for Ψ minus constructs. This discrepancy can be explained on two levels. First, the "wild type" construct used by Stuhlmann et. al. (16) was not a natural retrovirus, but rather a chimeric construct. It might be imagined therefore that packaging efficiency of this construct might not be that of a natural retrovirus. Indeed, their "wild type" titers were nearly 10 fold lower than is reported for HaSV here, values which are consistent with prior reporting (9). In addition, Stuhlmann et. al. (16) included a short *gag* region in place of the packaging site. While *gag* sequences can not replace Ψ function they do seem to augment packaging efficiency (19). Therefore it might be expected that the titers of their Ψ(-) construct would be somewhat higher than reported here. Indeed, their secondary foci titers are several fold higher. These factors (i.e. a relatively reduced packaging of the "wild type" and an increased packaging of the Ψ(-) construct) combined to yield a relatively smaller decrease in packaging efficiency in that report when compared to what is reported here.

I° cell clone	Avg. total II° ffu/ml	Avg. recombi/ml	% recombinants (col. 3/col.2)	Wild Type HaSV titer ffu/ml	relative packaging reduction (col. 2/col. 5)
a) pΔΨ(-)A9	3.8	.25	6.6%	10^6	$3x10^5$
b) pΔΨ(-)A10α	3.0	.13	4.3%	10^6	$3x10^5$
c) pΔΨ(-)A5	14	2.0	14.2%	10^6	$7x10^4$
d) pΔΨ(-)A10β	3.8	.38	10%	10^6	$3x10^5$
e) pΔΨ(-)A8	3.1	.50	16.1%	10^6	$3x10^5$

Table 2: Summary of secondary focus titers in relation to wild type HaSV titers.

reference	% expts. showing recombi.	range of secondary ffu/ml
Goodrich et. al. (9)	50%	0-11
Stuhlmann et. al. (16)	83%	0-69
Zhang et. al. (13)	100%	1-48
Swain et. al. (15)	n/d	10-80
This report	11%	0-5

Table 3: Rates of secondary focus formation from various sources.

As outlined in the introduction, the RNA model would predict a parallel drop in recombination rates whereas the DNA model would not. In looking at the experiments which did undergo recombination, the number of recombinants in secondary plates was generally ten fold lower than in prior reports (table 3). This may simply represent differences in experimental design, the types of recombinations which are being examined, helper virus titer differences, and the systems being utilized. Indeed, the results presented here are more consistent with other experiments which have used the HaSV virus as opposed to other virus constructs (9). Even though packaging efficiency was reduced, there was not a parallel drop in

recombination efficiency, thus supporting the DNA hypothesis (tables 2 and 3).

Recombination efficiency may also be examined from the standpoint of number of successful experiments, each one representing a gainful recombination. All told, 11% of experiments showed recombination under the conditions specified in Materials and Methods. If looked at in this context, the chance of getting at least one recombinant focus in any initial experiment was lower than prior reports (table 3). However, even in the report where 100% of the experiments showed at least one successful recombination a reduction in the success rate represents only a 10 fold drop in recombination incidence. Assuming the DNA model is operating here, the more difficult recombination which is involved in recovery of a packaging site (via aberrant helper virus integration) when compared to the simple recovery of the 3' LTR by downstream integration of a helper virus might explain this discrepancy with other studies. Indeed this small reduction (5-50 fold) in recombination efficiency in packaging site minus constructs has been noted before, and might reflect the more difficult recombination described (16). Even still, the reduction in efficiency of recombination does not parallel the reduction in packaging efficiency and therefore the RNA model is not supported.

Assuming the RNA model was correct and there was a 100,000 fold reduction in packaging, then it would take multiple experiments on average before even one recombination would be seen. Furthermore, because predicted rates of recombination based on reduced packaging are so low, finding more than one recombinant in

a single experiment is unlikely. In contrast, a single recombination in the DNA model would result in multiple recombined progeny. Therefore, the number of II° foci formed in that individual successful experiment would be consistent with recombination efficiencies of prior reports. This latter pattern of sporadic recombination with multiple recombinants in II° plates is what is seen rather than single, infrequent recombinations that are consistent with a reduction in overall packaging efficiency.

Furthermore, recent analysis of recombination based on the RNA model shows that non-homologous recombination occurs at a rate of once per 10^5 replication cycles (13). Therefore, if one experiment produces on the order of 10 secondary foci per experiment (representing 10 replication cycles of background infection), then it is reasonable to assume that one successful recombination would occur in 10,000 experiments via the RNA mechanism. This phenomenon was not seen in my set of experiments.

No matter how this issue was examined recombination rates do not drop 10^4-10^5 fold, as would be expected were the process packaging dependent. Therefore the RNA model is not supported.

Further evidence favoring the DNA mechanism is that secondary foci formed are all of monotonous (i.e. identical) length (figs. 17 and 18). As outlined in the introduction, these results are consistent with the DNA mechanism and not the RNA mechanism. Restriction mapping correctly places MoMLV packaging sequences in the HaSV genome from where the packaging site had been removed (figs. 19,20,21).

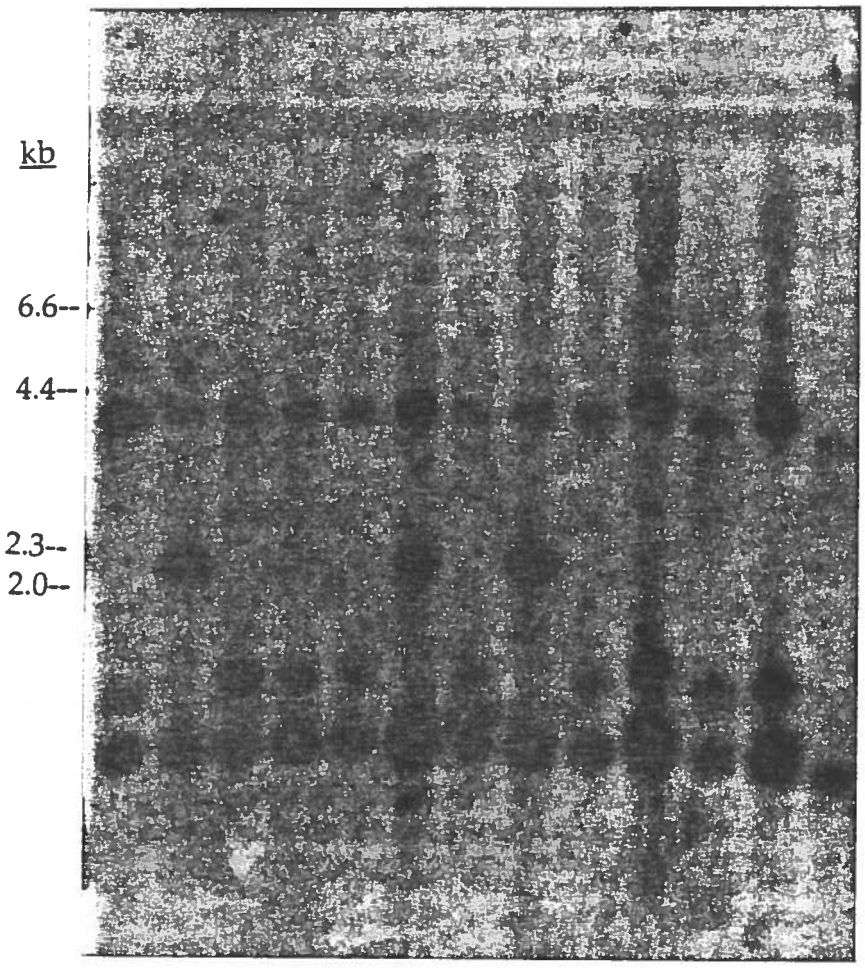

Fig. 17 Blot showing secondary foci cut with NheI to explore changes in restriction length. Bands at about 1,350 bp represent viral *ras* sequences of background secondary foci which retains the transfected pΔΨ(-)A construct (lanes 1,3,4,5,7,9,10,11,12). Bands at about 2,200 bp represent recombined secondary foci (lanes 2, 6, 8). Cell clones and lanes are as follows: pΔΨ(-)A9II° #32,33,34,35,47 and 48 are in lanes 1 thru 6 respectively. pΔΨ(-)A10αII° #11,12,13,14,15,16 are in lanes 7 thru 12 respectively. Lane 13 has C3H as negative control.

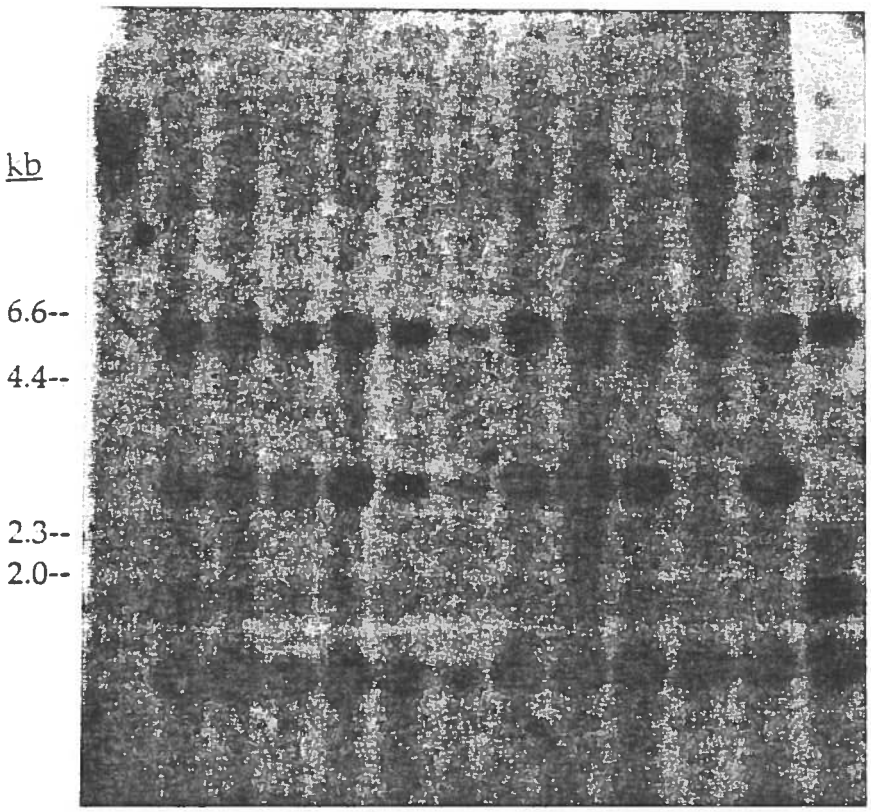

Fig. 18: Recombined secondary foci from four different successful recombination experiments. Cell clones and lanes are as follows: pΔΨ(-)A9II° # 1,9,10,33 and 48 are in lanes 1 thru 5; pΔΨ(-)A5II° # 1,13 are in lanes 6 and 7; pΔΨ(-)A10αII° #2 and 12 are in lanes 8 and 9; pΔΨ(-)A10ßII° #42,43,48 are in lanes 10,11 and 12. Lane 13 contains pΔΨ(-)A5I° DNA which has only the unrecombined construct. Size considerations are as in fig. 17. Note that lane 1 does not show very strongly.

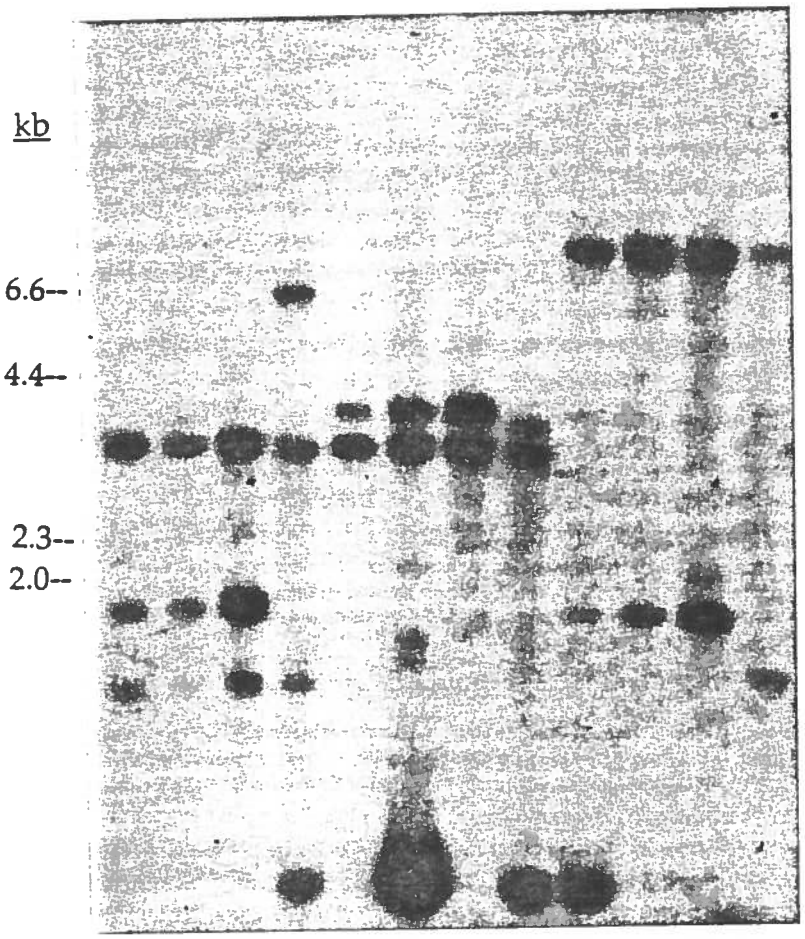

Fig. 19 Blot to determine packaging site recovery in selected secondary foci from three different experiments. In lanes 1 thru 4 cell clones pΔΨ(-)A5II° #1, pΔΨ(-)A9II° #9, pΔΨ(-)A10αII° #12 and C3H (negative control) are cut with BalI. Important band at approximately 1,850 bp. In lanes 5 thru 8 cell clones pΔΨ(-)A5II° #1, pΔΨ(-)A9II° #9, pΔΨ(-)A10αII° #12 and C3H (negative control) are cut with MstII. Important band at approximately 3,850 bp. In lanes 9 thru 12 cell clones pΔΨ(-)A5II° #1, pΔΨ(-)A9II° #9, pΔΨ(-)A10αII° #12 and C3H (negative control) are cut with EagI/BglII double digest. Important band at approximately 1,800 bp. Also see fig. 20 for PstI cuts which appropriately fit the restriction map. See fig. 21 for compilation of this data.

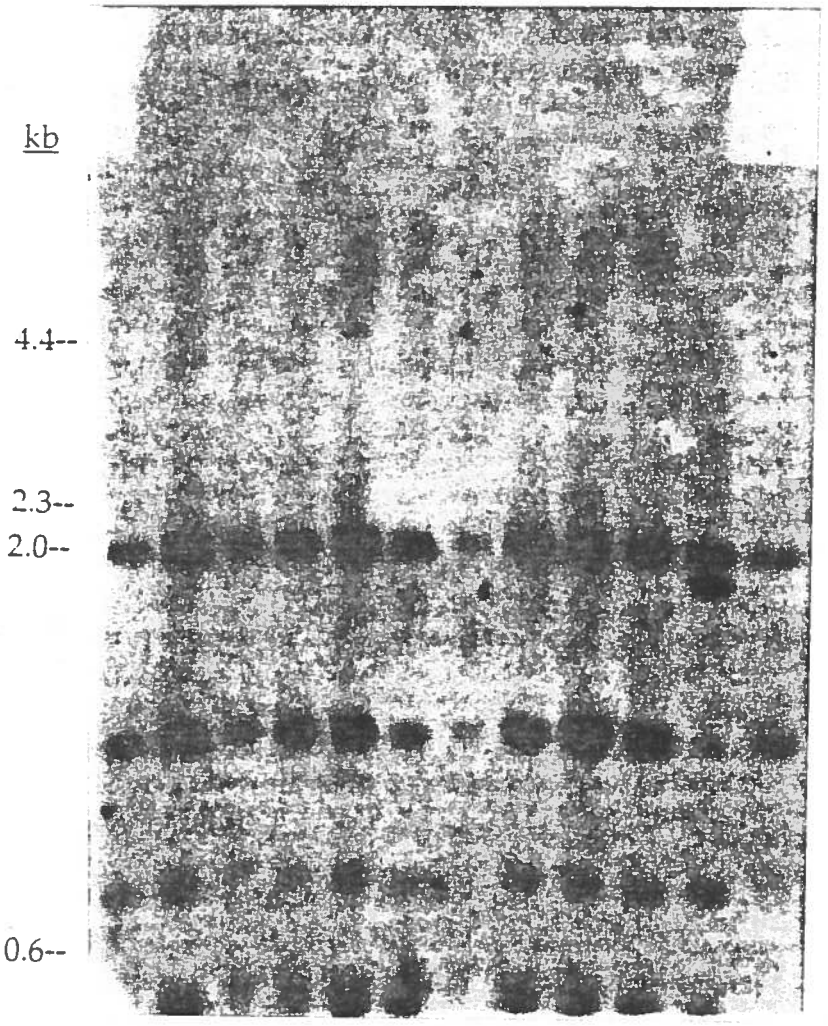

Fig. 20 Blot to determine small differences in recombination length at about 1 kb. Secondary foci from four different experiments cut with PstI show no discernible difference between lanes. Cell clones and lanes are as follows: pΔΨ(-)A9II° # 1,9,10,33 and 48 are in lanes 1 thru 5; pΔΨ(-)A5II° # 1,13 are in lanes 6 and 7; pΔΨ(-)A10αII° #2,12 are in lanes 8 and 9; pΔΨ(-)A10ßII° #42,43,48 are in lanes 10,11 and 12. Apparent differences in length are due to uneven electrophoresis of DNA in the gel. This is verified by the fact that endogenous *ras* sequences follow the same pattern of unevenness.

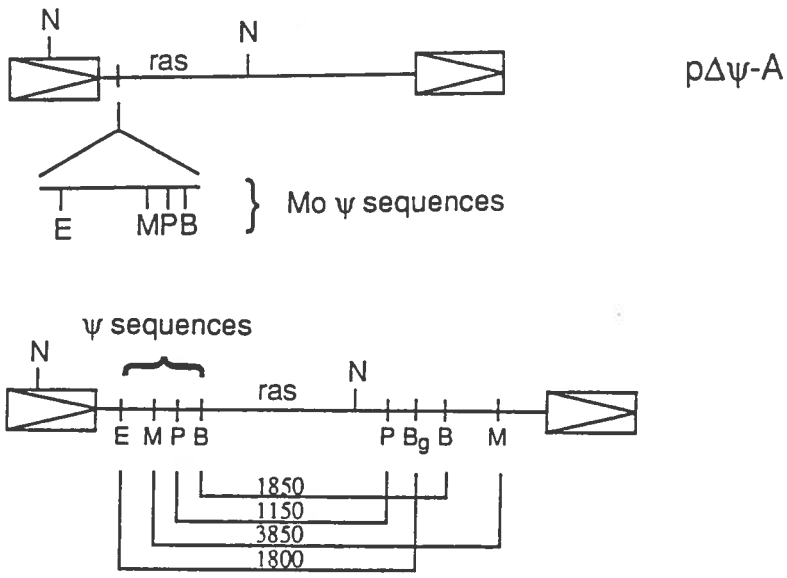

Fig. 21. Mapping of the recovered packaging sequences (figs. 19 and 20). Cuts with EagI/BglII (double digest), MstII, PstI and BalI yield restriction fragments of approximately 1800, 3850, 1150 and 1850 bp respectively. These lengths map from the appropriate MoMLV packaging sites to the corresponding respective sites in the HaSV genome and account for the 850 bp picked up by HaSV in the recombination event. The integrity of the intersite distances (e.g. the distance between PstI and BalI and BalI and MstII) is maintained by these numbers. E, M, P, B and Bg stand for EagI, MstII, PstI, BalI and BglII respectively. Not drawn to scale.

The recovered length is about 850 bp long and extends into the 5' *gag* region. This is reasonable considering that important packaging sequences are said to extend from the U5 into the *gag* region (19,20). This does raise an interesting question of how *ras* sequences could be correctly translated considering the first start codon encountered would be that of *gag* and not *ras*. There are several possibilities. First, *gag* sequences might be spliced out as is normally done in processing of retroviral RNA prior to translation. Second, mistakes in recognizing start codons may occasionally initiate *ras* and not *gag* synthesis. Lastly, an in frame fusion protein might be created which does not effect *ras* function as long as too much *gag* is not appended. In fact, many *v-oncs* are expressed as *gag-onc* fusion proteins (1). In particular the *ras* oncogene from

the Rasheed Rat Sarcoma Virus has an appended *gag* portion (21,22). Therefore, the creation of a similar protein here is a distinct possibility.

As can be seen from figs. 17 and 18 lengths of recombinant provirus appear to be the same, not only within experiments, but also between experiments. DNA from secondary foci was recleaved with an enzyme (PstI, which cuts in the HaSV genome as well in recovered packaging sequences) which would allow for better resolution of small differences in recombination length. The results depicted in fig. 20 show that there is no discernible differences in the recombinatory lengths. This can be explained in several ways. First, it may be chance that all have recombined at apparently the same length (to the degree that Southern blots have the ability to resolve differences in fragment length). Second, it is possible that a specific recombination was selected for because of its biologic viability. For example, recombinations which included too much of the MoMLV *gag* coding region might not allow for proper *ras* function as indicated above. This might set the upper limit for recombination length. Conversely, a recombination which did not include the entirety of the MoMLV packaging sequences would be less likely to be packaged and might not therefore be noticed under the conditions specified in this study. Only secondary infections which went on to exhibit a high titer of tertiary foci were investigated by Southern Blot (see Materials and Methods). Perhaps, those with intermediate titers on tertiary plates would show recombinations which took up smaller amounts of the MoMLV packaging sequences. Therefore, the allowable length of recombination permitted for

these experiments to work might be constrained by an upper and lower limit of recoverable MoMLV DNA. Lastly, there might be some element of recombination bias which is taking place. Because the pattern of recombination still fits the DNA rather than the RNA model (as discussed previously), however, these recombinations would take place in the I° cell. Indeed, there is evidence which shows that recombination between chromosomal proviral elements and integrating retroviral DNA can occur in some biased fashion. The degree of the recombination efficiency was dependent upon the constructs being utilized but it was unclear what mechanism was at work, including the possible role homology played (23,24). Indeed, a second construct created for this study which had a slightly different portion of the HaSV packaging site removed (the ClaI/Nar1 fragment was removed) never showed biological evidence of recombination in a similar number of experiments, even though phenotypic transformation of primary cells and secondary cells was possible (data not shown). In contrast to what is presented here, prior work which did describe packaging site recovery demonstrated different lengths among its recombinants (16). This indicates that the recombinations noted here were indeed system specific and not due to chance recombination events.

Were the last possibility described above also acting in this system, the essentially identical lengths of recombinants between experiments tends to invalidate the conclusions which would otherwise be drawn. At this point all that can be said with certainty is that recombination was at the DNA level. However, the mechanistic relevance to retroviral recombination is unclear.

Possible Homologous Recombination

It is interesting to note that for at least one set of secondary foci analyzed, homologous recombination at the LTR may have occurred. The evidence for this phenomenon is that a single DraI site present in the LTR of HaSV, but not in that of MoMLV, is lost in several of the of the "background" secondary foci which did not otherwise demonstrate packaging site recovery. Southern Blots of secondary focus DNA cleaved by DraI gave lengths greater than that predicted for the unrecombined secondary foci (fig. 22). This could be interpreted to mean that the helper virus packaging sequences were picked up during recombination such that the packaging site was introduced leaving the original HaSV LTR intact at the 5' end. Thus the DraI site would be retained while increasing the overall length of the secondary virus consistent with the length of the packaging site recovered. However when DNA was again cleaved by enzymes found in both the HaSV and MoMLV LTRs (KpnI and NheI), 4 out of the 6 secondary foci which had appeared to have recombined via the above mechanism turned out to exhibit a length consistent with a background virus. This indicated that indeed no recombination leading to packaging site recovery had occurred (figs. 23 and 24). Even though retroviruses do mutate at a high rate due to the imprecise action of reverse transcriptase (25), mutations are too frequent here to explain the loss of the DraI site by that route. Overall, the recombination rate via this mechanism for all secondary foci examined was 36%.

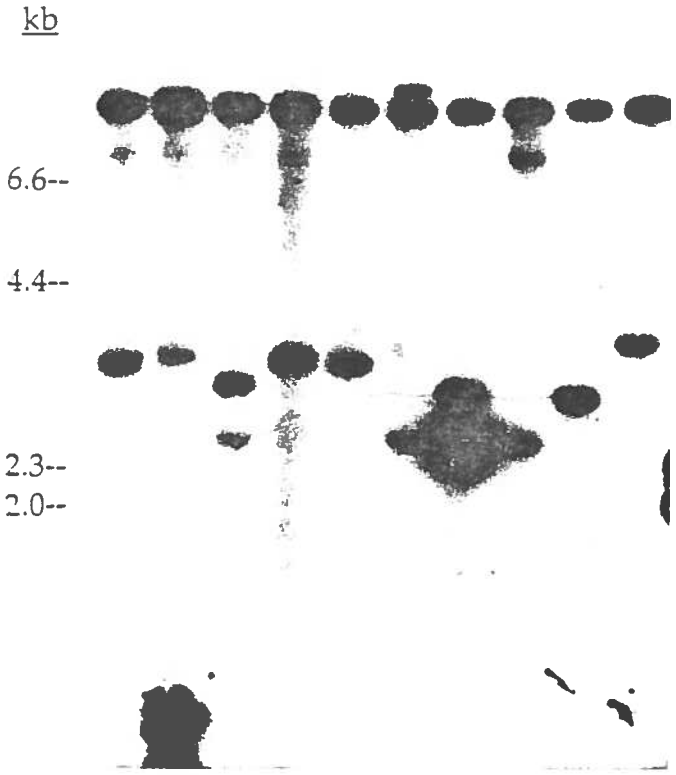

Fig. 22 Blot showing DraI cuts in selected secondary foci. Cell clones which show an increased length as compared to background foci are in lanes 1,2,3,4,5 and 10 (cell clones pΔΨ(-)A5II° # 1,10,12,13,16 and 15 respectively). Lane 6 has pΔΨ(-)A5II° # 17 (a background focus), lane 7 has pΔΨ(-)A plasmid DNA admixed with C3H DNA, and lane 8 has pΔΨ(-)A5I° cell DNA, all of which cut according to expected DraI HaSV cuts. Lane 9 has C3H DNA as a negative control.

The increased length consistent with the DraI cuts in these foci could be explained if it were assumed that the upstream cut fell in the cellular DNA upstream of the LTR. But if the background virus did not recombine to regain the packaging site, then how is it explained that the HaSV LTR lost its DraI site to allow an upstream cut? This seeming paradox might be resolved when the concept of homologous recombination is employed. Homologous recombination which may have occurred at the LTRs would exchange MoMLV and HaSV sequences, thus losing the upstream DraI site while retaining the rest of the restriction sites in common to both LTRs (fig. 25). Theoretically, this could occur via strand displacement of homologous elements between heteroduplex LTRs, however this phenomenon has not been noted in other studies designed to examine LTR synthesis during reverse transcription (4,26).

Another possible explanation is that the mechanisms of retroviral replication which have been previously described are not accurate. Presently, it is believed that retention of HaSV's U3 region (where DraI is located), and not the switching of U3s with MoMLV, follows the normal pattern of intramolecular or intermolecular DNA synthesis during reverse transcription (4,26). That is to say that retroviral U3s are not switched between heteroduplex RNA during normal reverse transcription. (This, however, is not necessarily true of U5 regions). Assuming the results presented here are simply the outcome of reverse transcription, intermolecular plus strand DNA synthesis may have

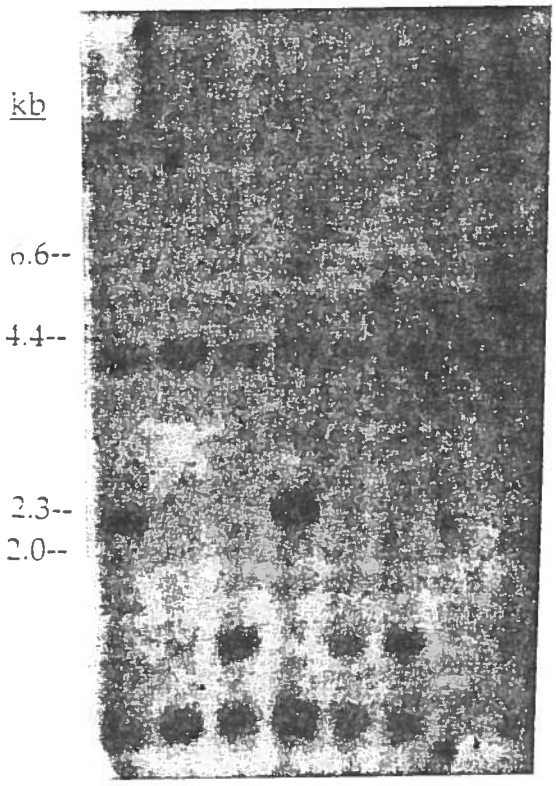

Fig. 23 Blot cutting same cell clones as in fig. 22 with Nhe1. pΔΨ(-)A5II° # 1,10,12,13, 15 and 16 are in lanes 1 thru 6. Note that only clones pΔΨ(-)A5II° # 1 and 13 retain their recombined lengths while the remaining four foci revert to a "background" status. Lane 7 is C3H and lane 8 is pΔΨ(-)A5I° DNA. Length considerations are as in fig. 17

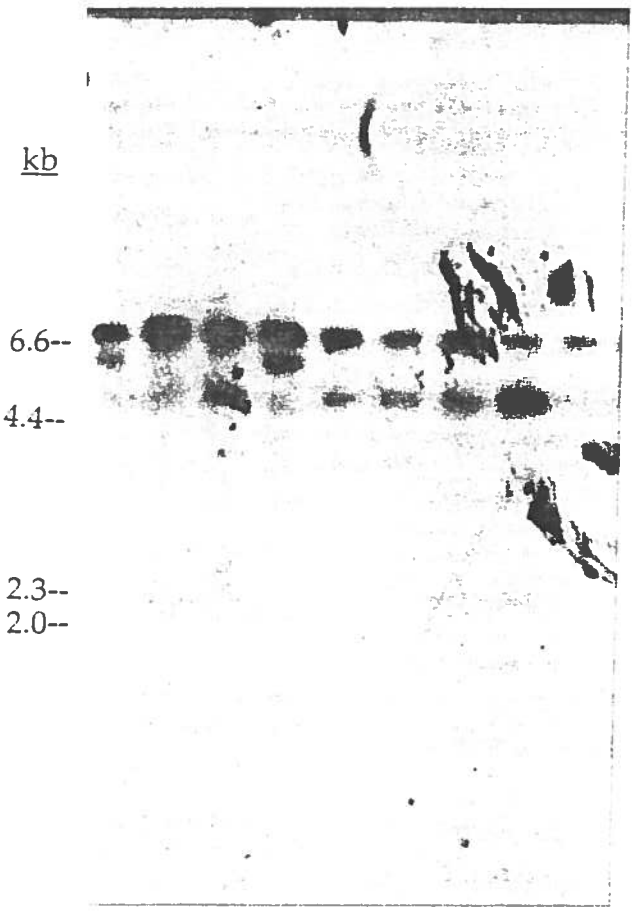

Fig. 24 Blot cutting same cell clones as in fig. 22 with Kpn1. pΔΨ(-)A5II° # 1,10,12,13, 15 and 16 are in lanes 1 thru 6. Note that only clones pΔΨ(-)A5II° # 1 and 13 retain their recombined lengths while the remaining four foci revert to "background" status. Lane 7 has pΔΨ(-)A5I° DNA, lane 8 has pΔΨ(-)A plasmid DNA admixed with C3H DNA and lane 9 has C3H DNA. Unrecombined lengths at approximately at 5 kb while recombined lengths are at approximately 5.9 kb.

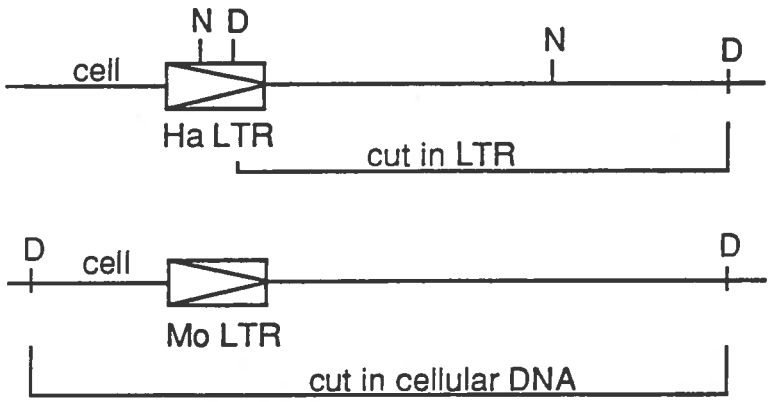

Fig. 25. HaSV and MoMLV LTR sequences and expected cuts with DraI (D). MoMLV LTR sequences found in place of HaSV LTR sequences demonstrate DraI cuts of extended lengths, falsely indicating packaging site recombination. N is NheI. Not drawn to scale.

resulted in the switching of the MoMLV U3 for that of HaSV's. This conclusion is at variance with prior reports which have said that plus strand DNA synthesis is solely an intramolecular process, thereby excluding switching of U3s during normal reverse transcription (4,26). It also should be considered that homologous recombinations could have occurred at the DNA level.

The implications of possible homologous recombination in the LTRs in this system are unclear. As discussed previously, homologous recombination is considered to be a phenomenon occurring mainly at the RNA level. Were the recombination to take place via copy choice as has been suggested, reverse transcription through the packaging minus region would have already occurred before the mutated virus would be able to benefit from homologous recombination at the LTR. Copy choice mediated homologous recombination between LTRs would therefore by necessity take place only after the packaging region had already been copied into DNA. Similarly, strand displacement after dsDNA synthesis can not explain how homologous recombination initiated at the 5' LTR might

continue into the non-homologous packaging site thereby providing packaging site recovery (4,11).

Aside from the fact that there is no theoretical mechanistic explanation for LTR homology-initiated packaging site recovery, it does not appear that homologous recombination at the RNA level plays a role in this situation for two reasons. First, as discussed previously, the profile of recombination does not fit the RNA model, but rather the DNA model. Because homologous recombination is a hallmark of RNA rather than DNA recombination, the pattern of recombination should be more like that predicted by the RNA model were homologous recombination operating. Second, of the 2 remaining foci which did indeed exhibit genuine recombination (i.e. packaging site recovery), both retained their DraI sites in the 5' LTR (fig. 22). Therefore, this indicates the 5' LTR remained that of HaSV, as opposed to those which had switched their LTR elements as described. Mechanisms which might involve U3 switching do not therefore appear to play a role in packaging site recovery in this system.

3'LTR Recovery

Experiments designed to examine the mechanisms of 3' LTR recovery could not support either the DNA or RNA mechanisms.

Three C3H cell lines were selected which had been successfully electroporated with the pEMJ construct as described in Materials and Methods. Five individual experiments where the 5' LTR of the AKR virus was secondarily transfected into established cells clones were performed (i.e. fifteen experiments total). In all

experiments the supernatant collected from each cell clone before electroporation showed no ability to transform fresh C3H cells. Likewise when fresh C3H cells were inoculated with supernatant harvested from pEMJ cells five days after the electroporation procedure no foci arose.

In regard to the DNA model, the indeterminate outcome may simply be a consequence of the low chance of experiencing a successful integration in this system. If it is assumed that a successful rescue would occur if the LTR were to land within 1 kb of the truncated provirus and that 50% of the 1,000,000 cells per experiment randomly integrated electroporated DNA (a reasonable assumption based on electroporation data), and that the incorporated DNA only had a 50% chance of being oriented in the proper direction, and that the mouse genome is 10^9 base pairs in length, then the overall chance of having one successful integration per experiment is about 1%. A reduction in the assumed efficiency of electroporation may significantly reduce the likelihood of experiencing a successful experiment.

One might reason that because both the truncated RNA is being produced (as indicated by cellular transformation) and that helper virus is being produced (as evidenced by biologic activity--see Materials and Methods), the RNA model is invalid since no recombinants arose with all the requisite elements present. This conclusion is ill founded. First, it is impossible to determine anything from a negative result in this system. Second, even though helper virus is being produced, its activity may not be that of the wild type virus. The helper virus in this situation is not the natural

AKR virus, but rather is using a HaSV LTR as its 3' LTR (see Material and Methods). Just as in the case of the packaging site, a reduction in the ability to proceed through reverse transcription because of the presence of non-conventional AKR sequences would reduce the frequency of RNA dependent recombination as well. If this were the case it would be unfair to impugn the RNA model.

A single focus from pEMJ cells which did show an ability to transform fresh C3H cells in the initial screening process (see Materials and Methods) was analyzed for LTR recovery. The restriction cuts were consistent HaSV LTR recovery (data not shown). No conclusions can be drawn from this data however because it is impossible to tell exactly how the LTR recombined downstream of the *ras* sequences because there was no baseline status for the cell clone when the initial screening took place. It could have been through copy choice or through aberrant helper integration downstream of the pEMJ construct or it could have been a transfection artifact as described in the introduction.

Conclusion

The constructs created for this report were a novel attempt to elucidate the intermediate steps involved in retroviral gene transduction. The experiments were designed to support the DNA model of recombination, but evidence for this was not forthcoming. While recombination did apparently take place at the DNA level in Ψ minus constructs, the uniform nature of the recombinants between experiments indicates that recombination occurred in some biased fashion. Because recombinations are assumed to be random and not directed, the relevance to DNA recombination as a model is unclear. There was no evidence to support the RNA model either.

The experiments designed to examine recombination at the packaging site should be viewed as a first step in the utilization of such constructs, rather than as definitive tools. Even though retroviral recombination is being examined in the general sense, the specific question of 3'LTR recovery is not being addressed. Constructs which have no 3' LTR and are with or without a packaging site should be the next step in investigation. When such constructs have been created, the investigators found that packaging reductions far exceeded reductions in recombination(16). These results were not consistent with the RNA model. Similar experiments should be performed independently to see if the same results would occur in a HaSV based system.

The system utilizing tandemized retroviral constructs still holds promise despite the fact it didn't work in these experiments. Clearly, the mechanism of viral integration employing a viral

integrase differs from the mechanisms which incorporate random sequences of DNA into the genome after electroporation. However, what is important for the DNA model is that an LTR integrates downstream of the truncated provirus--the exact mechanism by which this is done is irrelevant. In all respects this integration mimics a normal biological integration of a virus except that a virus did not perform the integration. If a 5' LTR is integrated (by whatever mechanism) downstream of a truncated provirus that provirus will be rescued. Even though there is a small risk of reinfection of cells by helper virus (due to receptor blockade), this may occur and recombinations may take place (1). However, their nature would still remain undetermined. This system could therefore be improved by adding anti-viral antibodies to the culture medium to prevent reinfection of cells by helper virus (27).

Another technical improvement would be to determine the exact nature of recombinants involved in transduction through PCR as has been done in prior reports (13,15). Subsequent sequence determination then reveals the exact substrates involved in recombination.

Although not the focus of this report, two interesting findings did emerge from these experiments. First, the odd switching of U3 regions between the helper and the mutant retrovirus indicates an as yet undescribed recombination at the LTR has occurred or that our thinking about the mechanisms of reverse transcription should be re-examined. Second, the possible creation of a *gag-ras* fusion protein in this system is intriguing. The implications for how *gag-*

onc fusion proteins might naturally arise through packaging site minus intermediates could be worth further investigation.

References:

1. Weiss, R, Teich, N, Varmus, H, Coffin J: *RNA Tumor Viruses*; second edition; Cold Spring Harbor Lasboratory, Cold Spring Harbor, NY. pp. 3,4, 227-248, 397, 423-36, 444-49, 453-467, 514-528, 537 999-1108. (1984).

   Weiss, R, Teich, N, Varmus, H, Coffin J: *RNA Tumor Viruses*; supplements appendixes; Cold Spring Harbor Lasboratory, Cold Spring Harbor, NY. pp. 46-49, 76-80. (1985).

2. Darnell, J, Lodish, H, Baltimore, D, eds: *Molecular Cell Biology*, first edition; Scientific American Books Inc, USA. pp. 1046-1063.

3. Goff SP: Integration of Retroviral DNA into the Genome of the Infected Cell. *Cancer Cells* **2(6)**: 172-178 (1990).

4. Hu WS, Temin HM: Retroviral Recombination and Reverse Transcription. *Science* **250**: 1227-1232 (1990).

5. Goldfarb MP, Weinberg RA: Generation of Novel, Biologically Active Harvey Sarcoma Viruses via Apparent Illegitimate Recombination. *Journal of Virology* **38(1)**: 136-150 (1981).

6. Goldfarb MP, Weinberg RA: Structure of the Provirus within NIH 3T3 Cells Transfected with Harvey Sarcoma Virus DNA. *Journal of Virology* **38(1)**: 125-135 (1981).

7. Varmus H: Form and Function of Retroviral Proviruses. *Science* **216**: 812-820 (1982).

8. Goodrich DW, Duesberg PH: Retroviral Transduction of Oncogenic Sequences Involves Viral DNA Instead of RNA. *Procedings of the National Academy of Sciences USA* **85**: 3733-3737 (1988).

9. Goodrich DW, Duesberg PH: Evidence that Retroviral Transduction is Mediated by DNA, not by RNA. *Procedings of the National Academy of Sciences USA* **87**: 3604-3608 (1990).

10. Coffin JM: Structure, Replication, and Recombination of Retrovirus Genomes: Some Unifying Hypotheses. *Journal of General Virology* **42**: 1-26 (1979).

11. Junghans RP, Boone LR, Skalka AM: Retroviral DNA H Structures: Displacement-Assimilation Model of Recombination. *Cell* **30**: 53-62 (1982).

12. Stuhlmann H, Berg P: Homologous Recombination of Copackaged Retrovirus RNAs During Reverse Transcription. *Journal of Virology* **66(4)**: 2378-2388 (1992).

13. Zhang J, Temin HM: Rate and Mechanism of Nonhomologous Recombination During a Single Cycle of Retroviral Replication. *Science* **259**: 234-238 (1993).

14. Goff SP, Tabin CJ, Wang JY, Weinberg R, Baltimore D: Transfection of Fibroblasts by Cloned Abelson Murine Leukemia virus DNA and Recovery of Transmissible Virus by Recombination with Helper Virus. *Journal of Virology* **41(1)**: 271-285 (1982).

15. Swain A, Coffin JM: Mechanism of Transduction by Retroviruses. *Science*. **255**: 841-845 (1992).

16. Stuhlmann H, Dieckmann M, Berg P: Transduction of Cellular *neo* mRNA by Retrovirus-mediated Recombination. *Journal of Virology* **64(12)**: 5783-5796 (1990).

17 Erlich, HA ed: *PCR Technology: Principles and Applications for DNA Amplification*; Stockton Press, USA: p. 3 (1989).

18. Ausubel, FM, Brent, R, Kingston, RE, Moore, DD, Seidman, JG, Smith, JA, Struhl, K, eds: *Current Protocols in Molecular Biology*, vols 1 and 2; John Wiley and Sons, USA: pp. 2.2.1-2.2.2, 2.4.1-2.4.4 (1989).

19. Bender MA, Palmer TD, Gelinas RE, Miller AD: Evidence that the Packaging Signal of Moloney Murine Leukemia Virus Extends into the *gag* Region. *Journal of Virology* **61(5)**: 1639-1646 (1987).

20. Murphy JE, Goff SP: Construction and Analysis of Deletion Mutations in the U5 Region of Moloney Murine Leukemia Virus: Effects on RNA Packaging and Reverse Transcription. *Journal of Virology* **63(1)**: 319-327 (1989).

21 Young, HA, Rasheed S, Sowder, R, Benton CV, Henderson, LE: Rat Sarcoma Virus: Further Analysis of Individual Viral Isolates and the Gene Product. *Journal of Virology* **38**: 653-654 (1981).

22. Rasheed, S, Norman, GL, Heideker, G: Nucleotide Sequence of the Rasheed Rat Sarcoma Virus Oncogene: New Mutations. *Science* **221**: 155-157 (1983).

23. Olson P, Temin HM, Dornburg R: Unusually High Frequency of Reconstitution of Long Terminal Repeats in U3-minus Retrovirus

Vectors by DNA Recombination or Gene Conversion. *Journal of Virology* **66(3)**: 1336-1343 (1992).

24. Dougherty, JP, Temin, HM: A Promoter-less Retroviral Vector Indicates that there are Sequences in U3 Required for 3'RNA Processing. *Proc. of the Natl. Acad. of Sci. USA* **84**: 1197-1201 (1987).

25. Pathak, VK, Temin HM: Broad Spectru of *in vivo* Forward Mutations, Hypermutations, and Mutational Hotspots in a Retroviral Shuttle Vector After a Single Replication Cycle: Substitutions, Frameshifs, and Hypermutations. *Proc. of the Natl. Acad. of Sci. USA* **87**: 6019-6023. (1990).

26. Panganiban AT, Fiore D: Ordered Interstrand and Interstrand DNA Transfer During Reverse Transcription. *Science* **241**: 1064-1069 (1988).

27. Hu, W, Temin HM: Genetic Consequences of packaging Two RNA Genomes in One Retroviral Particle: Pseudodiplody and High Rate of Genetic Recombination. *Proc. of the Natl. Acad. of Sci. USA* **87**: 1556-1560. (1990).