Applications of a retroviral integrase towards substrate DNA *in vivo*.

by

JOHN-PAUL HEALE

B.Sc., McMaster University, 1992

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE STUDIES

(Department of Biochemistry and Molecular Biology)

We accept this thesis as conforming
to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

April 1997

In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of Biochemistry and Molecular Biology

The University of British Columbia
Vancouver, Canada

Date April 30/98

ABSTRACT:

This study explored the efficacy of using retroviral integrase expressed transiently *in vivo* to mediate the recombination of exogenous DNA into host cell chromatin. Stable recombination of a substrate DNA into the genome of Baby Hamster Kidney (BHK) cells was achieved by co-transfecting them with a circular plasmid encoding Rous Sarcoma Virus (RSV) integrase (pIN) and a linearized plasmid (pNR) serving as a substrate for the integrase. *NdeI* linearized pNR (substrate DNA) mimicked the wild type RSV viral DNA in being a linear, double-stranded DNA molecule possessing terminal sequences recognizable by the integrase. Either a dihydrofolate reductase or neomycin cassette engineered into pNR served as a marker for recombination by conferring resistance to methotrexate or G418 selection, respectively. After 10 days growth in media supplemented with either 500μM methotrexate or 575μM G418, a small number of discrete colonies had formed on control plates containing cells which had been transfected only with substrate DNA. However, plates containing cells which had been co-transfected with both substrate DNA and pIN showed a ten-fold or higher increase in colony numbers over control plates. Sequence analysis of co-transfected BHK genomes identified many clonal recombinants; however none was a conserved, full length substrate DNA molecule expected from RSV IN mediated integration. Southern blot analysis of genomic DNA from co-transfected cells indicated that multiple copies of pNR had recombined with the BHK. Hence, *in vivo*, integrase increases the frequency of recombination of a recognizable substrate DNA molecule. However RSV IN mediated integration of substrate DNA was not observed.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS AND GLOSSARY

AIDS	acquired human immunodeficiency syndrome associated with infection by the HIV retrovirus
ALV	avian leukosis virus, a retrovirus which infects fowl
293	transformed human embryonic kidney cells
BHK	transformed baby hamster kidney cells
bp	base pair, a unit which denotes a single base pairing between nucleotides of a double stranded DNA molecule
C	Cys, cysteine, an amino acid
CA_{OH}-3'	terminal dinucleotide exposed on retroviral DNA essential for integration to occur
CV-1	transformed African green monkey kidney cells
D	Asp, aspartate, an amino acid
Da	dalton, a unit of mass referring to the atomic mass of a hydrogen atom
DHFR	dihydrofolate reductase, an enzyme responsible for reducing folate to tetrahydrofolate which is necessary in the synthesis of purine nucleotides
DMEM-F12	a media used in the propagation of mammalian cells in tissue culture
DNA	deoxyribonucleic acid
DODAC	di-oleoyldimethylammonium chloride, a positively charged lipid used in a 1:1 molar ratio with DOPE to form liposomes which will interact with DNA and facilitate its transfection to mammalian cells in tissue culture
DOPE	di-oleoylphosphatidylethanolamine, a neutral lipid used in a 1:1 molar ratio with DODAC to form liposomes which will interact with DNA and facilitate its transfection to mammalian cells in tissue culture
DTT	dithiolthreitol, a chemical which reduces disulfide bonds
ds	double stranded
E	Glu, glutamate, an amino acid

g	gravity, i.e. 3000×g is defined as 3000 times the force of gravity
G418	an analog of neomycin capable of uptake by mammalian cells which inhibits protein synthesis, the commercial name is Geneticin, marketed by Gibco-BRL
H	His, histidine, an amino acid
HIV	human immunodeficiency virus, a retrovirus
IN	integrase, a retroviral protein which is responsible for mediating integration of retroviral DNA into host cell chromatin to form the provirus
IPTG	isopropyl-β-D-thiogalactopyranoside, and inducer of the lactose operon
kb	1000 base pairs (bp)
kDa	1000 daltons
LB	Luria-Bertani, a broth used in the propagation of bacteria
LTR	long terminal repeat, *cis* acting DNA sequences found at termini of retroviral DNA
ml	milliliter, a unit of volume referring to a thousands of a liter
mRNA	messenger ribonucleic acid, transcribed from DNA, which is in turn translated by ribosomes to form proteins
Mg^{2+}	magnesium, a divalent cation
Mn^{2+}	manganese, a divalent cation
MTX	methotrexate, a drug used in cancer treatment which inhibits the enzyme DHFR, thus killing cells which are actively dividing
pIN	a plasmid based on pΔE-sp1B which is used in the construction of recombinant adenoviruses, it contains a RSV IN gene whose transcription was driven by a cytomegalovirus promotor and transcription terminated by a SV40 polyadenylation signal
PMSF	phenylmethyl sulfonyl fluoride, an inhibitor of aspartic proteases
pNRdhfr	a plasmid based on pBluescript, which contains a NDE-RSV insert such that digestion with the endonuclease *NdeI* forms linear substrate DNA, also

	contains a dihydrofolate reductase cassette which confers resistance to methotrexate
pNRneo	a plasmid based on pBluescript, which contains a NDE-RSV insert such that digestion with the endonuclease *NdeI* forms linear substrate DNA, also contains a neomycin cassette which confers resistance to G418
RCR	replication competent retrovirus, arises from recombination of a replicative deficient recombinant retrovirus which recombines at least two time with mRNA encoding viral proteins
RSV	Rous Sarcoma Virus, a retrovirus which infects chickens
RT	reverse transcriptase, a retroviral protein which converst ssRNA genome into ds viral DNA
ss	single stranded
substrate DNA	a linear molecule of DNA whose 30 bp terminal ends were identical to RSV viral DNA thus allowing recognition and binding by RSV IN
µl	microliter, a unit of volume referring to a hundred thousands of a liter

ACKNOWLEDGEMENTS

I would like to thank Dr. R.T.A. MacGillivray for the opportunities he provided for me and the patience he showed to me during my graduate studies. To the MacGillivray lab and my committee members for insight and assistance. In particular I am indebted to Dr. Mark Brown, Beatrice Tam, and Gord Rintoul.

Life outside the lab was made possible by friends and family, without whom I would have never been able to complete this work.

INTRODUCTION:

GENE THERAPY

What is gene therapy?

Gene therapy can in some forms be considered a novel method of drug delivery that enlists the machinery of a patient's cells to produce a therapeutic agent. Using the body to treat its own disease overcomes the need to manufacture highly purified protein. It also eliminates the need for repeated administration of proteins or drugs and reduces the difficulties of complying with drug regimens. Applications of gene therapy are not limited to rare inherited diseases but extend to common acquired disorders, including cancer, heart disease, and the acquired immunodeficiency syndrome (Blau *et al.* (1995)). Thus gene therapy is likely to have broad implications for the future practice of medicine

History of gene therapy.

The feasibility of gene transfer therapy was demonstrated in the early 1970s when naturally occurring DNA and RNA tumor viruses were used to introduce new genetic information into the genomes of mammalian cells. Several investigators noted that these viruses could be used as vehicles for gene transfer if their undesirable elements were replaced with genes that would yield therapeutic products (Friedmann *et al.* (1972)). In theory, vectors could be used to transfer new DNA into the cells of patients by, for example, direct injection or inhalation of viral particles. Alternatively, vectors could be used *ex vivo* to modify autologous cells destined for engraftment into patients.

In 1980, the first gene therapy trial involving humans had begun in an attempt to treat β-thalassemia (Mercola *et al.* (1980)). Widely viewed as premature on scientific grounds and lacking appropriate ethical review, the testing was eventually stopped. This was followed by a ten year moratorium on human gene therapy trials.

While ethical question were, and continue to be debated, it was decided to focus further gene therapy trials on diseases whose genetic treatment were less dependent on stringent gene control, as was the case for beta-thalassemia (Miller (1992)). The first approved gene transfer trial involving humans occurred in 1989 involving investigators from the United States National Institutes of Health (Rosenberg *et al.* (1990)). The protocol involved use of a recombinant retroviral vector to transduce a bacterial antibiotic resistance gene into autologous tumor infiltrating lymphocytes used in the immunotherapy of patients with advanced melanoma. The transferred gene was not intended for therapeutic purposes, instead it was to serve as a marker to study the activity of the tumor-infiltrating lymphocytes. In terms of gene therapy, it provided the first evaluation of the safety and efficacy of an *in vivo* gene therapy. Results showed persistence of the altered lymphocytes for up to 2 months at tumor sites and in the blood. There were no ill effects from the gene transfer procedure, laying the foundation for further clinical gene therapy trials.

Current applications of gene therapy trials.

At present there are over 100 gene therapy protocols approved worldwide. No adverse outcomes have been reported on any protocol. Ongoing protocols include gene therapy for

several inherited disorders such as adenosine deaminase deficiency, cystic fibrosis, Gaucher's disease, hemophilia B and familial hypercholesterolaemia (Table 1) (Dube *et al.* (1995)). Approaches include *in vivo* administration of adenoviral vectors to airway passages of cystic fibrosis patients, and liver resection coupled with portal vein infusion of *ex vivo*, retroviral modified, hepatocyte cultures encoding a high density lipoprotein receptor for patients suffering from hereditary hypercholesterolaemia. In the latter case, lower cholesterol levels were observed for a period of up to 18 months (Grossman *et al.* (1994)). Diseases such as cancer and AIDS are the focus of more than 80% of the clinical trials in progress. Current approaches encompass up-regulation of the immune system, targeting and killing of diseased cells, inhibition of oncogenes, and protection of normal tissue against the effect of chemotherapy. Many of the gene therapies involving cancer utilize marker genes to determine the efficacy of the gene transfer procedures. Such marker studies have proven that using marrow pre-harvested from children that undergo bone marrow transplantation will lead to relapse due to the inclusion of cancerous cells (Brenner *et al.* (1993)). Marked cells from marrow transplant procedures have been detected in the circulating blood and marrow for at least 18 months (Brenner *et al.* (1994), Dube *et al.* (1995)).

DISEASE	GENE or PRODUCT DELIVERED	TARGET	VECTOR
Inherited Diseases			
α-antitrypsin deficiency	α-antitrypsin	Respiratory tract	Liposomes
Chronic granulomatous disease	$p47^{phox}$ (oxidase)	Myeloid cells	Retrovirus
Cystic fibrosis	Cystic fibrosis transmembrane cotransductance regulator	Respiratory tract	Adenovirus, AAV, liposomes
Familial hypercholesterolaemia	Low density lipoprotein receptor	Hepatocyte	Retrovirus
Fanconi's disease	Complementation group C gene	Hematopoeitic progenitors	Retrovirus
Gaucher's disease	Glucocerebrosidase	Peripheral blood and stem cells	Retrovirus
Hunter's syndrome	Iduronate-2-sulfatase	Lymphocytes	Retrovirus
Adenosine deaminase deficiency	Adenosine deaminase	Lymphocytes	Retrovirus
Acquired diseases			
HIV	HIV clearing ribozyme, antisense RNA	Lymphocytes	Retrovirus
Peripheral-artery disease	Tumor angiogenesis factor	Endothelial cells	Plasmid DNA
Rheumatoid arthritis	Interleukin-1-receptor antagonist	Proliferating synovial cells	Retrovirus
Cancer	Tumor suppressor genes Herpes thymidine kinase Antisense RNA Multidrug resistance Tumor necrosis factor B7 co-factor for T-cell stimulation HLA-B7 for immune response Cytokines Interferon γ	Carcinoma cells Brain tumors Breast, brain tumor Hematopoeitic cells Lymphocytes Melanoma Melanoma Organ tumors Melanoma	Retrovirus, Ad Retrovirus, Ad Retrovirus Retrovirus Retrovirus Retrovirus Liposomes Retrovirus, liposomes Retrovirus

Table 1. Some gene therapy protocols approved for clinical trials by Recombinant DNA Advisory Committee

Current drawbacks of gene therapy protocols.

As the overview contained in Table 1 indicates, three major methods of gene transduction are currently being used. Adenoviral vectors are composed of a linear dsDNA viral genome which is approximately 36 kb in length. They differ from wild type adenovirus in that the E1 transcriptional regulatory proteins have been deleted from the viral genome. Recombinant viral adenovirus is packaged when the viral DNA is transfected into 293 cells which have been transformed by the *e1* genes, and therefore complement for the $e1^-$ recombinant vector. Recombinant adenoviral vectors are beneficial in that they are highly efficient at transducing and expressing high levels of a desired gene in both quiescent and actively dividing mammalian cells (Tripathy *et al* (1994)). Adenoviral vectors suffer from the drawback that their viral DNA remains extra-chromosomal providing only transient gene expression. In addition, repeated adenoviral infections cause inflammation within the tissue targeted for transduction (Zabner *et al* (1994)).

Another vector for gene transduction is the encapsulation of DNA into cationic liposomes. Liposomes are both easy to prepare and use, as well as having the most favorable safety profile *in vivo* of any of the gene transduction systems (Nabel *et al* (1993)). Similar to the adenoviral vectors, the DNA transduced remains extra-chromosomal, greatly limiting the duration of gene expression.

The most popular viral vector used in gene therapy is the recombinant retrovirus. Shortcomings of the retroviral vectors in gene therapy will be discussed, but more central to this thesis are the advantages that they confer to gene therapy protocols. Retroviral vectors are able to

stably integrate viral DNA into the host cell chromosome with the potential of stable gene expression (Nabel *et al* (1995)). The ability of the retrovirus to integrate its provirus into host cell genomes results from the retroviral IN protein and its ability to recognize the terminal ends of the retroviral LTRs. While integration of the viral DNA into the host is mediated by the entire nucleocapsid complex of the retrovirus, *in vitro* studies have shown that the IN protein and recognizable substrate DNA are all that is required to elicit auto-integration into other substrate DNA molecules.

It was decided to explore the ability of IN to be expressed separately from other retroviral proteins *in vivo*, and following substrate DNA transfection, observe its ability to integrate the substrate DNA in hopes of achieving stable gene expression.

THE RETROVIRUS

The retrovirus as a viral vector.

Recombinant retroviruses have been used extensively as vectors for the stable integration of various genes into the eukaryotic genome (Miller (1992)). Studies have explored their potential usefulness in gene therapy but, to date, clinical trials employing recombinant retroviral vectors have met with only limited success (Miller (1992), Anderson *et al.* (1992)).

The retrovirus structure.

The retrovirus is an enveloped virion with a diameter of approximately 100 nm (Figure 1). The membrane surface is decorated with a single protein structure, the product of the *env* gene. Within the viral membrane envelope, the nucleocapsid whose structure is ill defined, is comprised of the products from the *gag* gene. The nucleocapsid contains the duplicate, single stranded (ss) RNA genome and the two products of the *pol* gene: reverse transcriptase (RT), which converts the duplicate ssRNA genome into double stranded (ds) DNA, and integrase (IN) which covalently joins the viral DNA to cellular DNA, forming the provirus (Fields *et al.* (1990)).

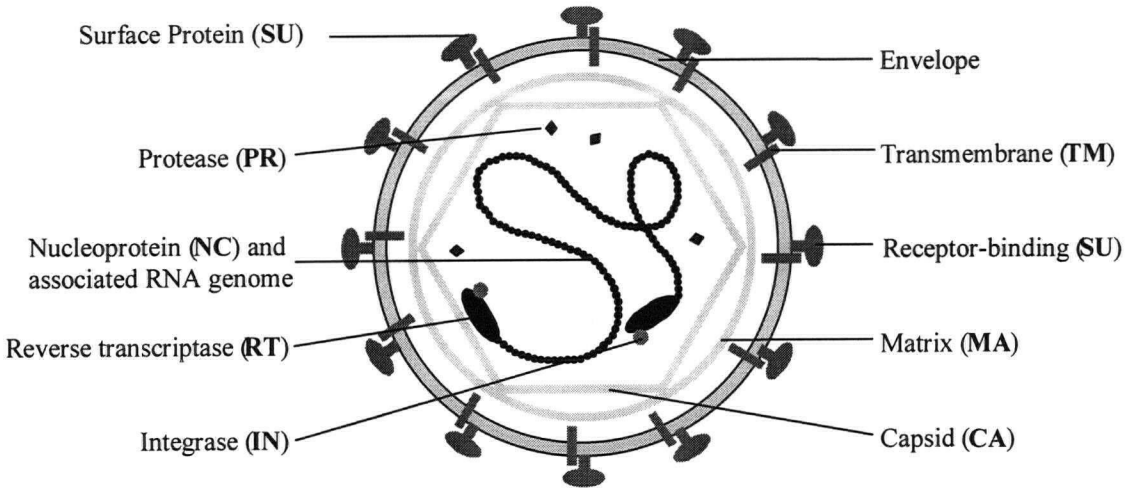

FIG. 1. **The retrovirus virion.** This highly schematic figure shows the relative locations of the various structures and proteins.

The retroviral RNA genome and conversion to double stranded DNA.

The retroviral RNA genome consists of positive strand mRNA usually between 7 and 10 kilo bases (kb) long, both 5' capped and 3' poly-adenylated as in cellular mRNA. The genes are always ordered gag-pol-env, and non-coding regions serve as essential recognition signals for DNA or RNA synthesis as well as genome packaging within developing virions (Figure 2). The viral genome does not function as a template for protein translation; instead it has the distinction of acting as a template for RNA directed DNA synthesis (Fields *et al.* (1990)).

RNA directed DNA synthesis as mediated by RT begins in the first phase of infection. RT, bound to a tRNA primer, begins negative strand DNA synthesis in a 3' direction from the primer binding region (Figure 2). As the DNA is synthesized, the ribonuclease H activity of RT degrades the RNA template. The DNA synthesis is momentarily halted, termed 'strong stop', as RT encounters the 5' end of the RNA genome. In the context of the nucleocapsid, the newly synthesized DNA strand migrates to the 3' end of the viral RNA and hybridizes to it by virtue of the repeat region found at both ends of the RNA genome. The hybridized DNA strand is then able to serve as the primer for renewed DNA synthesis which completes the negative DNA strand (Figure 2). Positive strand DNA synthesis begins from an RNA primer in a poly-purine tract that remains despite the RNase activity of RT. DNA synthesis continues to the 3' end of the DNA template where 'positive strand strong stop' occurs. Just prior to 'positive strand

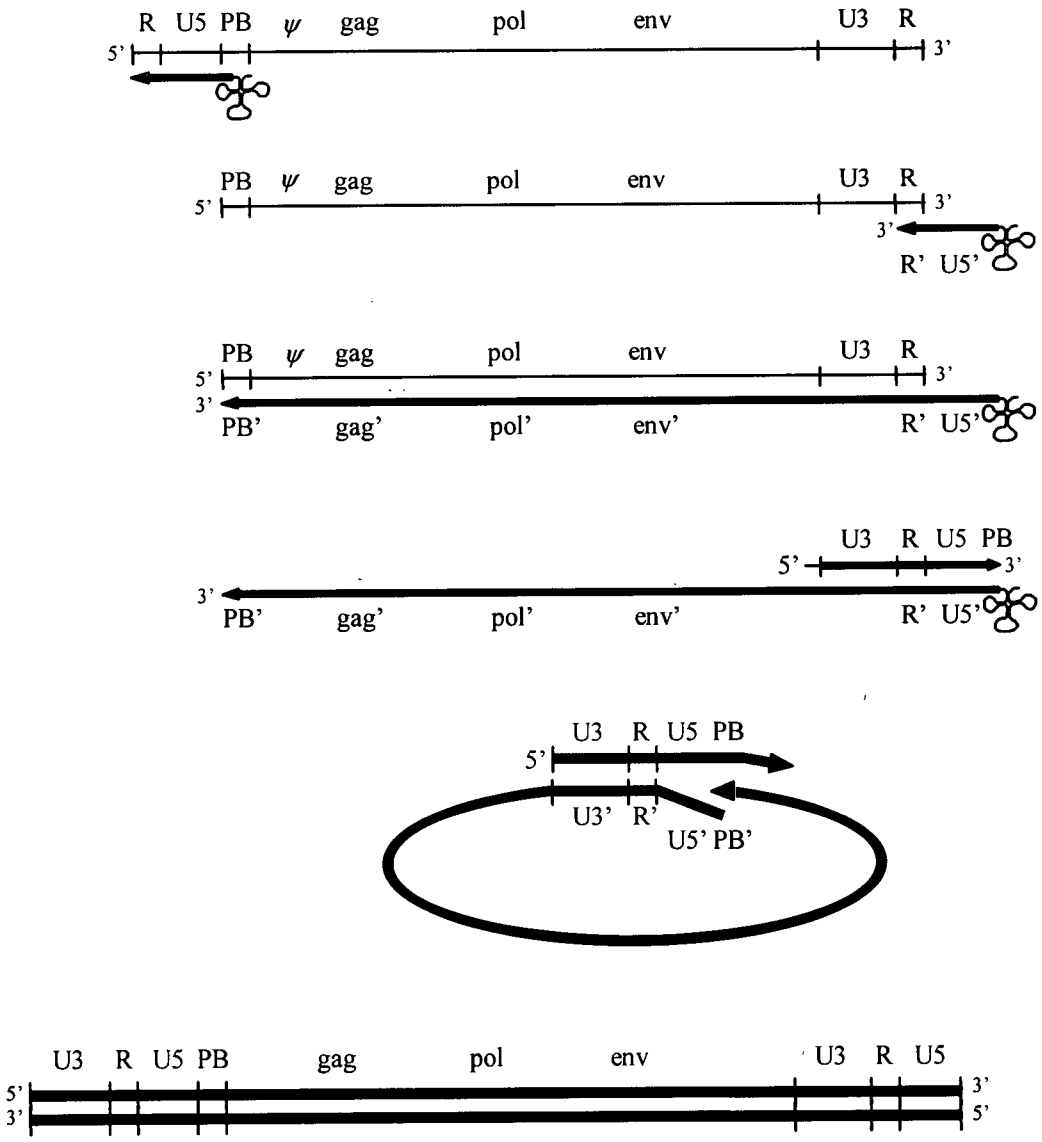

FIG. 2. **Mechanism of viral DNA synthesis.** Thin lines depict RNA, thick lines denote DNA. Negative-sense sequences are indicated by a 'prime'.Definitions; **R**-repeat, U5-U5 region of LTR, **PB**-primer binding region, ψ-packaging signal, **U3**-U3 region of LTR. **gag**, **pol**, and **env** are viral genes encoding structural and enzymatic viral proteins.

strong stop', the tRNA primer is copied to form the complement of the primer binding region. This complementary region can hybridize to the far 5' end of the negative DNA strand, and allow for completion of the positive DNA strand. The completed viral dsDNA is flanked by identical long terminal repeats (LTR) as a result of the complicated but elegant conversion of ssRNA to dsDNA (Fields *et al.* (1990)).

Functions of viral LTRs and integrase proteins.

IN binds the terminal regions of the 5' and 3' LTRs, and its endonuclease activity cleaves the TT-3' dinucleotide to expose the conserved CA_{OH}-3' dinucleotide. Changes to the CA dinucleotide abolish IN activity and thus terminate the infectious cycle. The nucleocapsid, including both IN and bound viral DNA, translocates to the nucleus where the viral DNA is covalently joined to the cellular DNA to form the provirus. The provirus is stably integrated and will persist for the life of the cell, as well as being passed on to any of the cell's progeny. Integration of the provirus marks the end of the first phase of infection, having begun with attachment of the virion to a specific cell surface receptor (Fields *et al.*(1990)).

The second phase of infection includes synthesis and processing of the viral genome RNA from the integrated provirus. The retroviral LTRs have the promoter activity necessary for mRNA transcription by cellular RNA polymerase, and the recognition sequence for integration of viral IN. mRNA production serves two purposes: first as a template for viral protein translation by host cell systems, and second, as viral genomes to be packaged within developing virions. The LTR acts as a constitutive promoter in many different cell types (Vile *et al.* (1994)) and allows transcription and subsequent protein translation of the viral genome by host cell mechanisms. A

packaging signal present on the viral genomes directs their inclusion into the virions, and a completed virus results from budding off of the host cell membrane.

THE RECOMBINANT RETROVIRAL VECTOR

Basic features of recombinant retroviral production.

In a cell containing a retroviral provirus, any RNA molecule can be packaged provided it has a packaging signal (ψ) (Figure 2). Inclusion of appropriate LTR elements in a packageable molecule of RNA form the cornerstone of recombinant retrovirus production. In addition to the packaging signal, LTR elements included within the RNA molecule ensure that following recombinant viral transduction, conversion to dsDNA by reverse transcriptase will occur. In a recombinant retrovirus, packaged RNA contains *cis* acting viral elements required for efficient virus replication, however they do not contain the viral genes, notably *gag-pol*, and *env*, necessary for structural and enzymatic proteins. These proteins are provided in *trans* by a packaging cell line whose genome includes *gag-pol*, and *env* retroviral genes.

The first generation recombinant retroviral packaging cell lines were achieved through stable introduction of a mutant Moloney Murine Leukemia virus (MoMLV) proviral genome containing a deletion of the packaging signal (ψ) (Danos *et al* (1988)). While theoretically unable to be packaged, the retroviral genome provided all of the structural and enzymatic proteins.

Subsequent transfection of those cells with DNA containing a gene of interest and the viral elements found within the LTRs would result in recombinant retroviral production which contained RNA transcripts encoding the gene of interest. Following recombinant retroviral transduction and conversion of the RNA to DNA by reverse transcriptase, the DNA would be integrated into the host cell genome ensuring persistence throughout its life span and passage to progeny cells.

Although thought to be unlikely, it was repeatedly observed that the mutant proviral genomes lacking the packaging signal were transferred to cells via the recombinant virus. Further recombination events did lead to production of replication competent retrovirus (RCR). RCR was, and still is a major concern as a study observed three of eight rhesus monkeys developing thymic lymphoma after being infected with a recombinant retroviral vector contaminated with RCR (Kolberg *et al* (1992)).

The second generation packaging cell lines for recombinant retroviruses involved further alterations to the ψ^-genome which significantly reduced the risk of co-packaging the retroviral genome into the recombinant virus. Generation of RCR was still occasionally observed (Danos *et al* (1988)).

Classical third generation packaging cell lines were constructed by successfully transfecting two plasmids containing the genes encoding *gag-pol* and *env,* respectively. Plasmid integration was selected for by drug resistance encoded within the respective plasmids and cells could again be selected based upon their protein expression levels (Danos *et al* (1988)). None of the retroviral genomic constructs contained the retroviral packaging signal, although there were regions of homology between the two plasmids. The packaging cells generated were termed ψCRE and ψCRIP (Danos *et al* (1988)). Although theoretically possible, generation of replication competent retrovirus (RCR) was extremely unlikely and observed only upon helper virus contamination or mass recombinant retroviral culture necessitated by large scale clinical requirements (Rigg *et al* (1996)).

Murine based third generation retroviral vectors were well suited for *ex vivo* applications, although they had limited capabilities *in vivo*. These limitations were the inability of the vector packaging cell systems to concentrate large amounts of the virus needed for direct *in vivo* gene transfer (Ory *et al* (1996), Cossett *et al* (1995)). Typically virus titers were in the order of 10^5 to 10^7 pfu/ml. Third generation recombinant retrovirus with amphotropic host range were also sensitive to inactivation by human serum (Takeuchi *et al* (1994), (1996)). It appears that inactivation by human serum was controlled by two elements; the retroviral envelope protein and the packaging cell line which may bear α1-3 galactosyl sugar on the cell surface membrane, thus activating complement (Cossett *et al* (1995)).

There are other detractions for recombinant retrovirus vectors. Murine based vectors were unable to integrate into quiescent cells (Lewis *et al* (1994)), and anomalies of the packaged virus have also been known to arise. Replication competent virus along with packaging of the *trans* supplied helper genomes lacking a packaging signal have occurred at low frequency (Cossett *et al* (1995). Lastly recombinant retrovirus vectors are limited by the size of the insert they can accept due to packaging limitations. Typically inserts are less than 8 kb, and if long term gene expression *in vivo* requires large flanking regions surrounding the gene of interest, retroviral vectors may prove unsuitable (Dube *et al* (1995)).

Higher recombinant viral titers and resistance to human serum have been observed when Vesicular Stomatitis virus (VSV) G protein/retroviral pseudotypes have been used. These pseudotypes vectors have been constructed when VSV and recombinant retrovirus vectors were co-transduced into cells resulting in, for the most part, a retroviral like particle exhibiting VSV G protein on its surface (Ory *et al*(1996)). These pseudotypes possessed the wide host range of VSV and were easily concentrated to titers of 10^9 pfu/ml without loss of activity. The pseudotypes were also more resistant to human serum, although the packaging cell line was derived from a 293 transformed cell line which has been shown to produce human serum resistant recombinant retrovirus (Loiler *et al* (1997), Rigg *et al* (1995)).

Other improvements made to retroviral vectors were through the replacement of MoMLV *env* gene, which was shown to promote human complement binding. The Gibbon Ape Leukemia *env* gene (Bunnell *et al* (1995)), Murine Leukemia virus *env* (Miller *et al* (1996)), and Mink Cell Focus virus *env* (Loiler *et al* (1997)) all provide more resistance to human complement than

traditional MoMLV based retroviral vectors. Typically different retroviruses utilize distinct cell surface receptors to gain entry into a cell (Loiler *et al* (1997)). Amphotropic retroviral constructs which employ MLV and GALV *env* genes are particularly useful since they can utilize either mouse or human *Glvr-1,* or rat or human *Ram1* for entry (Miller *et al* (1996)). This feature is attractive to researchers since *Glvr-1* is highly expressed in the bone marrow hematopoeitic cells and targeting bone marrow stem cells for gene therapy has long been a goal of gene transfer.

Research continues on production of retroviral vectors incapable of RCR formation. Recent efforts have focused on removing all retroviral untranslated sequences from packaging cell lines leaving only *gag-pol* and *env* genes which are transcribed by non-retroviral promoters (Cossett *et al* (1995), Rigg *et al* (1996)). Cell lines were also selected based on their lack of homology to retroviral sequences. Cell lines such as 293 were particularly useful because of their lack of homology towards retroviral sequence and production of complement resistant recombinant retroviral particles (Soreola *et al* (1995)).

THE ROUS SARCOMA VIRUS INTEGRASE PROTEIN

Retroviruses are unique among eukaryotic RNA viruses in that a DNA copy of a RNA genome is synthesized and integrated into the chromosomes of the infected cell as a normal step in the viral life cycle. In this regard, retroviruses can be viewed as a member of a family of DNA insertion elements along with Ty1, phage Mu, and Tn10 (Bushman *et al* (1991)). All of these insertion

elements contain *cis* acting terminal repeat sequences and encode for a protein which is a *trans* acting site-specific nuclease, facilitating transposition by joining the exposed 3' ends of the mobile insertion element to a 5' ends of a staggered cut made by the same protein in the target DNA (Collicelli *et al* (1985)). In the case of all retroviruses, the *trans* acting protein is the integrase (IN) protein. The general model for integration is that the IN protein acting as a dimer (Grandgennett *et al* (1993)) recognizes both ends of the linear viral DNA and cleaves the 3' ends of the termini by precisely 2 bp in preparation for integration. The IN dimer and associated viral DNA cleaves the host or target DNA and performs single stranded joining of the viral to host DNA. Upon integration, a duplication of host DNA sequences is created at the site of integration suggesting that the host DNA undergoes a staggered cut that produces 5' extensions, which following repair by cellular enzymes, results in direct repeats (Craigie *et al* (1990)). The integration reaction is a partially site-specific recombination reaction in that specific sequences on the viral DNA are joined to nearly random sites on the host DNA (Roth *et al* (1989)).

Early on in the study of retroviruses, two classes of mutations were discovered which directly affected integration of viral DNA into host cell sequences. The first class of mutation mapped near the 3' end of the retroviral *pol* gene and resulted in the failure of formation of the provirus even though reverse transcription of the original RNA viral genome into dsDNA proceeded normally (Donehower *et al* (1984), Schwartzberg *et al* (1984)). The other class of mutations involves changes in the DNA sequence present at each end of the unintegrated viral DNA. Such mutations can severely reduce or abolish integration and hence viral replication (Roth *et al* (1989), Craigie *et al* (1990), Collicelli *et al* (1985)). Investigators followed the logical conclusion that the 3' *pol* region encoded a protein that interacted with viral DNA termini and

ultimately determined provirus integration (Collicelli *et al* (1985)). It was not long before viral integrases had been shown to exhibit binding to viral termini (Roth *et al* (1989)).

The wild type integrase protein.

The Rous Sarcoma Virus IN protein results from cleavage of the gag-pol precursor protein by the viral protease $Pr180^{gag\text{-}pol}$ (Eisenman *et al.* (1980)) to release *gag* encoded proteins as well as amino terminal RT and carboxy terminal IN from the *pol* encoded precursor protein. Having resulted from proteolytic processing, IN does not have an amino terminal methionine encoded by an ATG start codon in recombinant RSV integrase (Figure 3). The molecular weight of the IN protein is 32 kDa, and IN is not normally glycosylated by host cells (Grandgenett *et al* (1993)). In wild type RSV infections, the serine residue at position 85 is phosphorylated, an indirect result of the v-*src* oncogene present in the RSV genome (Mumm *et al.* (1992)). Phosphorylation of Ser-83 does not appear to affect integrase function *in vitro*.

Retroviral integrase structure.

All retroviral integrases, and many transposons have three distinct domains (Figure 3): an amino terminal Zn^{2+} binding domain which features a conserved H-$X_{(3\text{-}7)}$-H-$X_{(23\text{-}32)}$-C-$X_{(2)}$-C, a carboxy terminal region having a low degree of homology between integrase proteins, and a central catalytic domain having a conserved sequence motif D-$X_{(39\text{-}58)}$-D-$X_{(35)}$-E (Bushman *et al.* (1994)). The amino terminal HHCC motif has homology with zinc finger domains, documented to be involved with DNA binding functions (Vincent *et al.* (1993)). Typically the zinc finger domains are arranged in clusters within a single protein, a characteristic not shared with retroviral integrases. This implies that IN behaves as a multimer, a theory supported by kinetic and

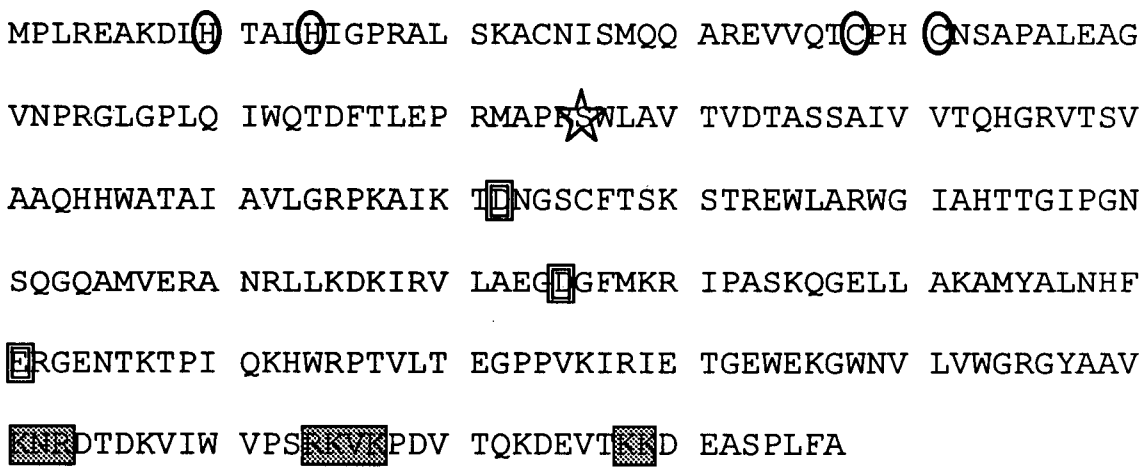

FIG. 3. **Amino acid sequence of recombinant RSV integrase protein.** Amino acid sequence of the translated *in* gene present within the pIN expression plasmid. Wild type RSV IN does not contain a Met amino acid at position 1 since the protein is processed from proteolytic cleavage of the gag-pol precursor protein. Ellipses highlight those residues necessary for formation of a zinc finger. The star highlights Ser 85, which is phosphorylated during a wild type RSV infection. Double boxed residues represent the conserved sequence motif D-$X_{(39-58)}$-D-$X_{(35)}$-E of the central catalytic domain. A non-contiguous nuclear localization signal exists near the carboxy terminus. Shaded boxes indicate a basic dipeptide separated by 10 residues from a three basic amino acids within the next five residues. This arrangement is consistent with nucleoplasmin nuclear localization signals (Kurz *et al.* (1997)).

ultracentrifugation evidence which suggests that the minimal functioning unit of IN is as a dimer (Jones *et al.* (1992)). In avian retroviruses, mutagenesis or replacement of the amino terminal HHCC motif does not significantly impair integrase function (Bushman *et al.* (1994)), while in other retroviruses, such as HIV-1, mutagenesis of the amino terminal of the integrase protein, can result in a 50% decrease in enzymatic activity (Vincent *et al.* (1993)). The carboxy terminal of the integrase protein, while having no large degree of homology, is required for DNA binding by retroviral integrases (Mumm *et al.* (1991)). Some researchers postulate that an integrase protein has two distinct DNA binding domains, one for target DNA and the other for substrate DNA (Vincent *et al.* (1993)). Others claim that there is only one DNA binding domain and that integrase functions as a multimer; in this case, the single DNA binding site consists of the substrate and target DNA for the desired integration reaction (Mizuuchi *et al.* (1992)). The central catalytic domain with the conserved D, D-35-E motif is believed to be the sole catalytic site for both 3' terminal processing and strand transfer of viral DNA to target DNA (Bushman *et al.*(1994)). This claim is strengthened by the fact that both processing and strand transfer involve the same one-step reaction (Engelmann *et al.* (1991)). Upon retroviral infection of a cell, the IN protein processes the 3' terminal ends of the viral DNA by trimming off a 3' dinucleotide, to leave a conserved and recessed CA_{OH}-3'. IN accomplishes this by a one step trans-esterification using a water molecule as a nucleophile in a S_n2 chemical reaction. Similarly, strand transfer uses the CA_{OH}-3' of the viral DNA as a nucleophile for attack of the target DNA in a one step trans-esterification. Both processing and strand transfer proceed independant of any high energy co-factor.

Integration sites and metal ion influence.

Sequence analysis from integration sites of retroviral provirus reveal that in the host, integration sites appear to be statistically non-specific (Vijaya *et al* (1986)). There are however documented 'hot spots', or repeated integration sites, observed *in vivo* involving retroviral integration.

DNase I hypersensitive sites are regions of chromatin which appear to be preferentially available for the entry of proteins that effect replication, transcription, and the re-arrangement of DNA. It also appears that these sites are preferentially available as integration sites for retroviruses (Vijaya *et al* (1986)). Avian Leukosis virus (ALV) has been shown to repeatedly integrate into DNase I hypersensitive sites 5' to the proto-oncogene *c-myc.* Such integrations cause *c-myc* to be constitutively expressed due to the promoter function of the ALV LTRs. This in turn causes uncontrolled cellular division leading towards tumor formation (Vijaya *et al* (1986)). Moloney Murine Leukemia virus (MoMLV) also preferentially integrates into a DNase I Hypersensitive site in the α-collagen gene. Other examples include retroviral integration into DNase I hypersensitive sites in proto-oncogenes such as *c-erbB* and *dsi-1*. Of note are the similarity that retrotransposons exhibit for DNase I hypersensitive sites.

An *in vitro* study of retroviral integration into mini-chromosomes show regional 'hot spots' in which integrations occur at a specific period of 10 bp. This allows for the supposition that potential integration sites are limited by the orientation of DNA wound about nucleosome cores. Recognition features of the DNA helix, such as the major groove, would only be available according to the period of the helix wound about nucleosome cores, roughly 10 bp (Pryciak *et al*

(1992)). It could be more simply stated that bend induced perturbations of DNA caused by protein binding causes preferential site use by retroviral integrase.

Related to the above work are the observations that Rous Sarcoma Virus (RSV) mediated integration events occur in a subset of chromosomal regions, and that all of the insertions within those regions were at exactly the same base (Shih *et al* (1988)).

In vitro studies also indicate that no simple consensus sequence exists for retroviral integration, although retroviral integrase proteins show a preference for A/T richness at alternating positions flanking the insertion site (Hong *et al* (1993)), Grandgenett *et al* (1993)).

The central catalytic domain, with three conserved amino acids, Asp-Asp-35 amino acids-Glu, binds two divalent cations, usually Mg^{2+}. Of note are the differing effects of different divalent metal ions upon catalysis of the integration reaction. *In vitro* studies almost exclusively use Mn^{2+} rather than Mg^{2+} which comparatively increases the level of both cleavage and strand transfer mediated by IN. The requirement for Mn^{2+} for *in vitro* reactions is not consistent with *in vivo* observations where Mg^{2+} was the divalent cation available (Vink *et al* (1991)). It would appear that IN requires Mg^{2+} for optimal specificity (Craigie *et al* (1990). Differences between Mg^{2+} and Mn^{2+} include specific DNA binding in the presence of Mg^{2+} compared to non-specific binding in the presence of Mn^{2+}. This phenomenon persists as longer LTR sequence derived deoxy-oligonucleotides were preferentially bound by IN in the presence of Mg^{2+} whereas shorter deoxy-oligonucleotides were bound in the presence of Mn^{2+} (Lee *et al* (1995)). As previously mentioned, during the course of retroviral infection, IN cleaves the terminal 3' dinucleotide from

the newly synthesized viral DNA prior to integration. Using Mg^{2+} as a divalent cation in an integration buffer, this cleavage occurs with marked specificity. IN cleavage activity is enhanced with Mn^{2+}, however an additional cleavage site at the 3 position from the 3' terminus becomes apparent (Katz *et al* (1990)). Nicking and DNA binding also occurs non-specifically throughout the substrate DNA (Lee *et al* (1995)). Mg^{2+} catalyzed IN activities are optimal at physiological pH (pH 7.5) whereas Mn^{2+} induced activity was optimal at pH 8.0 (Lee *et al* (1995)). Mn^{2+} is used in *in vitro* studies so as to promote high levels of strand transfer, whereas the greater specificity induced by Mg^{2+} translates into lower levels of IN activity. The less stringent effect of Mn^{2+} on DNA binding proteins was also observed for DNA polymerase, *EcoRV* and *E.coli* resolvase (Lee *et al* (1995)).

Requirement for DNA to act as a substrate for IN

Retroviral IN binds to the termini of the viral DNA LTR (Roth *et al* (1989)). For termini binding and subsequent integration the 3' CA dinucleotide is absolutely required although not sufficient. Alterations of nucleotides 3 through 13 bp internal to the termini negatively affected cleavage and strand transfer reactions, although with the exception of the terminal CA dinucleotide, no single base pair substitution had a strong inhibitory effect (Leavitt *et al* (1992)). Using deoxy-oligonucleotides molecules to mimic LTR termini it was found that 15 bp of the retroviral LTR was sufficient for cleavage and strand transfer, although a 28 bp sequence served as a better substrate (Vink *et al* (1991)). These data indicate that only a limited sequence is required for IN binding resulting in site-specific cleavage and strand transfer. Aside from the

terminal 3' conserved CA dinucleotide, there is very little sequence homology between retroviral LTRs (Figure 4). Logically this would indicate that a distinct LTR would only be processed by its specific IN protein. This assumption for the most part holds true although MoMLV IN was able to integrate precleaved HIV viral DNA at a low efficiency, however site specific cleavage was not observed (Vink *et al* (1991)). Precleaved substrate DNA was always integrated preferentially over a blunt ended substrate (Bushman *et al* (1991)). Precleaved substrate was derived by ligating the viral termini to form an *NdeI* site, at the same time conserving the 3' CA dinucleotide. Upon digestion with *NdeI*, a linear molecule results which displays the terminal sequences necessary for recognition by IN. In the study reported here, 28 bp of the terminal RSV LTR sequence was used as it provided the optimal length for IN recognition and activity (Vink *et al* (1991)).

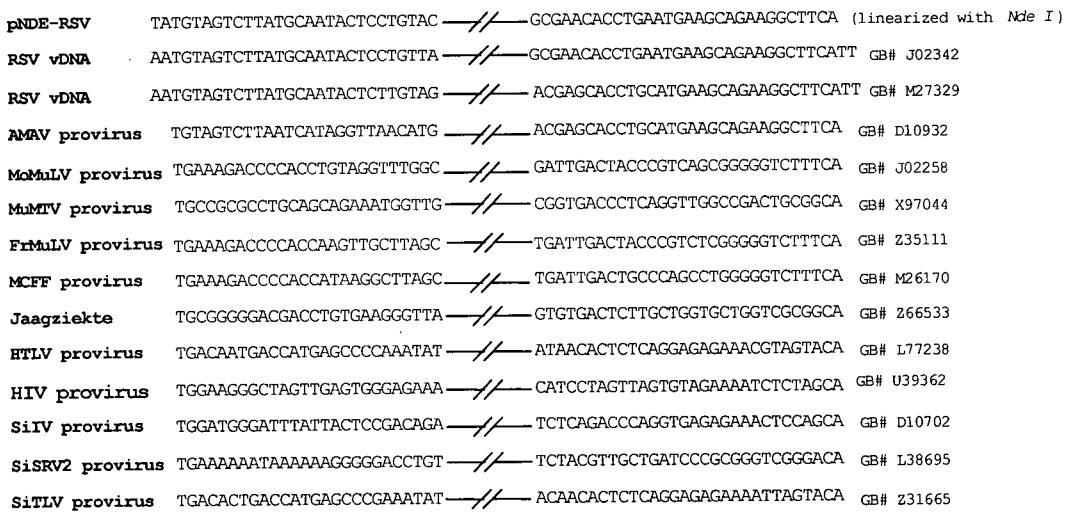

FIG. 4. **LTR comparisons.** Comparison of terminal LTR segments from a variety of retroviridae. The only conserved feature between all retroviridae is the 3' dinucleotide CA. All of the sequences were obtained from the GenBank database and the accession numbers are listed to the right of each sequence.

Objectives of the thesis.

The promise of gene therapy as a method of curing genetic disease at a molecular level is incentive for continued research. Currently the stumbling block is the cross-over between *ex vivo* studies involving gene transfer to cells in culture and *in vivo* studies in living organisms. It seems that as perfectly adapted viral vectors are towards gene transfer, the body as an organism has developed equally daunting defenses towards our crude attempts at synthetic virology. Perhaps an answer lies in a different direction.

This thesis explored alternatives to viral vectors to attempt gene transfer in mammalian cells in tissue culture. The research involved dissecting the retrovirus and using its greatest advantage over other means of gene transfer, that being viral DNA integration and persistence of the provirus in the host cell. Previous research has proven conclusively that the retroviral IN protein is responsible for mediating integration of the provirus based upon the terminal sequence of the linear viral DNA. Subsequently, using a cloned RSV *in* gene and linearized plasmid DNA mimicking retroviral DNA by virtue of homologous termini, simplified integration of a desired gene construct was attempted. While many examples of IN activity have been characterized *in vitro*, examples *in vivo* have not been characterized without the presence of a complete nucleocapsid containing many proteins in addition to IN and the viral dsDNA genome (Lee *et al.* (1995), Grandgenett *et al.* (1993)). *In vivo*, should transiently expressed RSV integrase prove capable of recognizing and integrating a substrate DNA molecule, a linear dsDNA molecule with termini identical to a RSV provirus, a major step will have been taken towards a general gene transfer protocol with therapeutic implications

The first step was to confirm the activity of the RSV integrase clone obtained from Dr. D.P. Grandgenett using a well characterized *in vitro* assay for IN. The assay would also assess the suitability of the designed substrate DNA to be used *in vivo*. Having ascertained integrase activity, co-transfections of an integrase expressing plasmid, pIN, and linearized pNR, the substrate DNA, would be attempted in Baby Hamster Kidney cells. The substrate DNA also encoded a selectable marker for drug resistance, either resistance to methotrexate or Geneticin. Following selection based upon drug resistance, and colonies that grow up would be indicative of cells which had successfully incorporated substrate DNA into their genomes. Expanded colonies provided a source of DNA which would allow characterization based on Southern blotting and sequence analysis, of the recombination sites of the substrate DNA. Owing to the complex nature of a mammalian genome, roughly 3 billion base pairs, methods for enriching clonal recombinants was required. A novel method for fast and simple purification of clonal recombinants was devised involving purification of the bacterial Lac I repressor protein which bound the Lac operon present within the substrate DNA, but not the mammalian genome. Protein/DNA complexes were bound to nitrocellulose membranes and following stringent elution of unbound genomic DNA, the Lac I operator containing sequences were eluted in the presence of IPTG which induced the Lac I repressor protein to release bound DNA. This enrichment process allowed sequence analysis of the substrate DNA terminal ends leading into the genomic DNA at the site of recombination.

Further characterization of the RSV IN, substrate DNA interaction was carried out *in vivo* when a RSV IN expression cassette was stably integrated into the BHK genome. BHK-IN clones

were isolated based on their level of IN expression, and characterized following transfection of substrate DNA encoding a neomycin cassette allowing selection for resistance to Geneticin.

The goal of these studies was to initiate a novel form of gene transfer, which after a brief period of transient viral protein expression, would lack the persistence of viral elements while stably expressing a desired gene. True IN mediated integration of substrate DNA was not observed based on sequence analysis which did not show complete conservation of the termini of the substrate DNA, but rather recombination with the BHK genome near the termini of the substrate DNA. In this study I observed a 10 fold increase in transfection efficiency based upon the number of colonies arising after selection for a recombined substrate DNA molecule conferring drug resistance. Thus *in vivo* expression of a retroviral integrase does increase the transfection efficiency of a recognizable substrate DNA.

MATERIALS AND METHODS:

A. MATERIALS

Chemicals for buffers and reagents were purchased from Fisher Scientific. Agarose for non-denaturing electrophoresis of DNA fragments was purchased from Gibco BRL, while acrylamide and bis-acrylamide for denaturing and non-denaturing electrophoresis was from Bio-Rad. Coomassie Brilliant Blue G250 was purchased from Bio-Rad. *Thermophilus aquaticus* (Taq) DNA polymerase was expressed and purified in our laboratory (Grimm *et al.* (1995)). Deoxy-ribooligonucleotides were synthesized on an Applied Biosystems PCR Mate 391 DNA synthesizer. Deoxy-ribooligonucleotides for PCR and DNA sequencing reactions, as well as dideoxy-ribooligonucleotides nucleotides were purchased from Pharmacia Biotech. A mutant form of T7 phage DNA polymerase, *Sequenase*, was purchased from United States Biochemical. Southern blot imaging and quantification were performed with a Molecular Dynamics Phosphor Screen and Phosphor Imager SI. Protein concentration was performed in a 10 ml Amicon protein concentrator with a 10 kDa cutoff filter. Autoradiograhic film was purchased from Island Scientific, Bainbridge Island, Washington. Restriction endonucleases, polymerases and ligases were purchased from Gibco BRL or New England Biolabs. The cloning vector pBluescript II and pRc/CMV was obtained from Stratagene, pTrc99A was from Pharmacia Biotech, and pNUT was a gift from Dr. R. D. Palmiter (University of Washington). The pMAL expression kit was purchased from New England Biolabs. Geneclean kits and reagents were purchased from Bio 101, Vista, California. ^{32}P labeled dCTP and Hybond N^+ nylon membrane was purchased from

Amersham. A T7 Quickprime kit for radioactively labeling DNA with ^{32}P was purchased from Pharmacia Biotech. BHK cells were obtained from American Tissue Type Collection. Falcon 6-well tissue culture plates were from Becton-Dickinson. DMEM-F12 media, newborn calf serum, glycerol, Geneticin (G418), and Trypan Blue were from Gibco BRL. DOPE:DODAC liposome mixture for cell transfections was a gift from Dr. Pieter Cullis (University of British Columbia). Methotrexate was purchased from David Bull Laboratories, Victoria, Australia.

B. STRAINS, VECTORS, AND MEDIA

1. BACTERIAL STRAINS

E. coli DH5α strain was used for most bacterial transformation using plasmids with selectable markers. The genotype is as follows; *supE*44, Δ*lac*U169 (∅80 *la*ZΔM15), *hsdR*17, *recA*1, *endA*1, *gyrA*96, *thi*-1, *relA*1. The ∅80 *la*ZΔM15 allows α-complementation with the amino terminus of β-galactosidase encoded in many vectors.

E. coli C strain was used when genomic DNA was to be transformed directly into bacteria. An example would be the direct rescue of a plasmid which had been integrated into a mammalian genome, subsequently removed by endonuclease digestion, religated to form a closed circle, and transformed. *E. coli* C is a wild type strain and lacks host restriction and modification

activity. It is also a nonsuppressing host strain used in complementation tests with amber mutants of bacteriophage λ.

2. EUKARYOTIC STRAINS

Transformed human embryonic kidney cells (293 cells) were obtained from the American Tissue Type Collection (ATCC #CRL-1573), Rockville, Maryland, U.S.A.. This cell line was transformed by a segment of adenovirus type 5 DNA containing the E1 region (Graham, F. (1991)).

BHK TK^- (defective for thymidine kinase) transformed hamster embryonic kidney cells were obtained from Dr. R. D. Palmiter (University of Washington).

Transformed green monkey embyonic kidney cells (CV-1 cells) were obtained from the American Tissue Type Collection, Rockville, Maryland, U.S.A..

3. VECTORS

pBluescript II was used for most DNA subcloning protocols and was purchased from Stratagene. The Rous Sarcoma Virus integrase gene was provided within the pGEM plasmid,

the kind gift of Dr. D. P. Grandgenett (Mumm *et al.* (1991)). pΔEsp1B was provided by Dr. Frank Graham, contained a 5 kb portion of the 5' end of the adenovirus type 5 genome from which the entire E1 genetic region has been deleted (Graham *et al.* (1991)). Eukaryotic expression vector pNUT (Palmiter *et al.* (1987)) contains the mouse metallo-thionein I promotor, which in the presence of zinc produced expression of an inserted gene. pNUT also contained a di-hydrofolate reductase (DHFR) cassette used as a selectable marker as it provides resistance for cells to methotrexate upon pNUT transfection and integration. The plasmid pMAL was purchased from New England Biolabs and upon insertion of a desired gene in frame with the plasmid's *Mal E* gene, easy purification of the expressed fusion protein was acheived by affinity chromatography through a maltose column. The plasmid pTrc99A was purchased from Pharmacia Biotech and contained a pTac promotor for regulated protein expression in bacteria upon induction with IPTG. pRc/CMV was purchased from Stratagene and encoded a neomycin cassette which was subcloned to construct the plasmid pNRneo.

4. MEDIA

Luria-Bertani (LB) broth was used in the maintenance and propagation of *E. coli* and consisted of 25 g of powdered LB base which was suspended in 1 liter of demineralized water (50% (w/w) pancreatic digested casein, 25% (w/w) autolysed yeast extract, and 25% (w/w) sodium chloride), pH adjusted to 7.5, and autoclaved for 20 minutes at 121°C. Once cooled, the

sterilized broth was ready for the addition of any desired antibiotics and the subsequent propagation of bacteria. Upon addition of 12.0 g agar powder prior to autoclaving, the LB agar solution could be poured into Petri dishes, and upon cooling, the solidified media was used as a matrix for growth of bacteria cultures.

Terrific broth was occasionally used when high bacteria cell mass was desired from overnight cultures and consisted of 47 g of Terrific broth base (25% (w/w) pancreatic digested casein, 50% (w/w) autolysed yeast extract, 20% (w/w) dipotassium hydrogen phosphate, and 5% (w/w) potassium dihydrogen phosphate) which was added to 1 liter of demineralized water, pH adjusted to 7.5 and autoclaved for 20 minutes at 121°C.

For eukaryotic cell culture, Dulbecco's modified Eagle medium, nutrient mixture F-12 (DMEM-F12) was used when supplemented with 5% (v/v) newborn bovine calf serum. 48 g of DMEM-F12 was added to 4 liters of demineralized water along with 9.2 g sodium bicarbonate, and the pH adjusted to 7.6. The medium was filter sterilized using a Gelman 0.22 μm filter in a laminar flow hood into 8, 500ml media bottles and stored at 4°C. Newborn bovine calf serum was aliquoted into 25 ml portions within a laminar flow hood, and one 25 ml aliquot was added to a DMEM-F12 media bottle to give a 5% newborn calf serum ratio (v/v).

C. GEL ELECTROPHORESIS

1. AGAROSE GELS

DNA fragments were separated according to size by employing agarose gel electrophoresis. Gels were prepared to a concentration ranging from 0.7% to 2.0% (w/v) in 1X TAE (50X TAE: 2M Tris base, 1M acetic acid, 0.1M EDTA pH 8.0), and a final ethidium bromide concentration of 1 ug/ml for DNA detection. Lower percentage agarose gels were optimal for separation of large DNA fragments, while higher percentage gels provide the small pore matrix necessary for separation of small fragments. Boiled agarose solution was cooled to approximately 48 °C and poured in an 8 ×12 centimeter tray containing a loading comb and allowed to cool to room temperature. DNA samples were mixed with 10X loading buffer (30% Ficoll, 0.2% xylene cyanol, 0.2% bromophenol blue) to a final 1X concentration and after loading into agarose wells, electrophoresis was carried out in a 1X TAE buffer at 1-3 volts/cm. The DNA fragments were visualized by irradiation over ultraviolet light at 260 nm and photographed with Polaroid 550 film using a 1 second exposure at F-stop 4.5.

2. POLYACRYLAMIDE GELS

Polyacrylamide gel electrophoresis was used to either separate small DNA fragments generated by sequencing protocols or for protein separation. Polyacrylamide gels were denaturing or non-denaturing as determined by the presence of urea.

Denaturing Polyacrylamide Gel Electrophoresis

Denaturing polyacrylamide gels used primarily to separate sequencing reactions consisted of 37.5 g of urea, 12.5 ml of 40% acrylamide (38:2 acrylamide:bisacrylamide), 7.5 ml 1X TBE (0.89 M Tris base, 0.89 M boric acid, 25 mM EDTA pH 8.0) and 28.6 ml distilled water. Polymerization was initiated by the addition of 500 µl of 0.1 g/ml fresh ammonium persulfate solution and 21.3 µl TEMED. Unpolymerized solution was poured between 0.5 mm spaced, silanized glass plates. After electrophoresis of the sequencing reactions, gels were dried under vacuum with a Bio-Rad gel drier at 80°C for an hour and the labeled nucleic acid fragments were visualized by autoradiography using film from Island Scientific.

Non-Denaturing Polyacrylamide Gel Electrophoresis

Non-denaturing polyacrylamide gel electrophoresis was carried out in order to resolve protein/DNA interactions in a Mobility Shift Assay. The gels were prepared from 13.3 mls of 30% acrylamide (40:1 acrylamide:bisacrylamide),5 mls 10X TBE (0.89 M Tris pH 8.0, 0.89 M Boric acid, 25 mM EDTA) and 81.0 mls of distilled water. Polymerization was induced by addition of 700 µl of 0.1 g/ml freshly prepared ammonium persulfate and 35 µl TEMED. Upon addition of the ammonium persulfate and TEMED, the solution was poured between 0.5 mm spaced, silanized glass plates and allowed to polymerize. After electrophoresis using 0.5 X TBE for a period of an hour and a half at 25 mA, gels were transferred to a paper backing and dried under vacuum with a Bio-Rad gel drier at 80°C. Labeled nucleic acids were detected by exposure to autoradiography film.

SDS - Polyacrylamide Gel Electrophoresis

Preparative SDS polyacrylamide gels used for protein separation consisted of 4.0 ml of 30% (w/v) acrylamide (29:1 acrylamide:biscrylamide), 3.35 ml distilled water, 2.5 ml 1.5 M Tris base pH 8.8, and 100 µl 10% SDS. Polymerization was initiated by the addition of 50 µl 0.1g/ml ammonium persulfate and 5 µl of TEMED. The unpolymerized solution was poured between glass plates separated by 1 mm spacers. The poured gels were overlaid with butanol and after polymerization, the butanol was removed and the remaining gel space was filled with a stacking gel preparation consisting of 1.3 ml of 30% (w/v) acrylamide (29:1 acrylamide:biscrylamide), 6.1 ml distilled water, 2.5 ml 0.5 M Tris base pH 6.8 and 100 µl 10% (w/v) SDS. Polymerization was initiated with 50 µl 0.1g/ml ammonium persulfate and 5 µl of TEMED. Protein samples were mixed 3:1 with 4X sample buffer (0.063 M Tris base pH 6.8, 10% glycerol, 2% SDS, 5% β-mercaptoethanol, and 0.00125% (w/v) bromophenol blue) and placed in a boiling water bath for 4 minutes before loading onto the polyacrylamide gel. Electrophoresis was carried out in 1X SDS-PAGE (25 mM Tris, 250 mM glycine pH 8.3, 0.1% SDS) running buffer. Following electrophoresis, gels were stained with Coomassie brilliant blue G250 for approximately 4 hours and destained (50% methanol, 10% glacial acetic acid) over 2 hours with several changes of destaining solution as required.

D. ISOLATION OF DNA

1. ISOLATION OF PLASMID DNA

Small scale preparations of plasmid DNA were performed using an alkaline lysis method described by Sambrook, J. *et. al.* (1989). An *E. coli* colony was used to inoculate 4 ml of Luria-Bertani (LB) broth containing 1 μg/ml of ampicillin and incubated at 37°C overnight. A 1.7 ml aliquot was transferred to a Eppendorf tube and centrifuged for 10 seconds in an Eppendorf microfuge at 10,000×g. The supernatant was discarded and another aliquot of bacterial culture added and centrifuged. The bacterial pellet was resuspended in 100 μl of glucose buffer (50 mM D-glucose, 10 mM EDTA pH 8.0, 25 mM Tris base pH 8.0, and 100 ug/ml RNAse) to which 200 μl of lysis buffer (0.2 N NaOH, 1% SDS) was added, mixed by inversion, and placed on ice for 5 minutes. Bacterial cell membranes, attached genomic DNA, and bacterial proteins were precipitated upon addition of 150 μl of Solution III (3 M potassium acetate, and glacial acetic acid to pH 4.8). Precipitate was pelleted by centrifugation in an Eppendorf microfuge for 7 minutes at 14,000 rpm. Supernatant was transferred to a fresh tube and plasmid DNA precipitated by addition of 1.3 ml of 95% ethanol, incubation at -20°C for one half hour and subsequent centrifugation for 10 minutes at 14,000 rpm. The plasmid DNA pellet was resuspended in 35 μl of distilled water and stored at -20°C. Plasmid DNA isolated in this manner was suitable for DNA sequencing protocols and restriction endonuclease digestion.

Large scale purification of plasmid DNA was performed using a Qiagen-tip 500. These preparations began by inoculating 100 ml LB broth containing 1 μg/ml ampicillin was inoculated with an *E. coli* colony and grown at 37°C overnight. Following the Qiagen protocol for plasmid maxi preps, approximately 500 μg of supercoiled plasmid DNA was recovered from the 100 ml culture.

E. OLIGONUCLEOTIDE SYNTHESIS

1) NDE-RSV 1

5'-CGG *GGT ACC* ATT GCG AAC ACC TGA ATG AAG CAG AAG GCT T*CA TAT G*TA GTC-3'

A 51 mer whose 21 base pair 3' end is complementary to oligonucleotide 2), NDE-RSV 2, 3' end and codes for a 5' *KpnI* site and 3' *NdeI* site.

2) NDE-RSV 2

5'-CGG *GGT ACC* GAG TAC AGG AGT ATT GCA TAA GAC TA*C ATA TG*A AGC CTT CTG-3'

A 51mer whose 21 base pair 3' end is comlementary to oligonucleotide 1)'s, NDE-RSV 1, 3' end and codes for a 5' *KpnI* site and 3' *NdeI* site.

3) 5' IN

5'-AAA *GAT ATC* ccc ttg aga gag gct aaa gat-3'

A 30mer whose lowercase nucleotides correspond to base pairs 4-21 of the RSV IN gene (Mumm, S. R. (1991)) and whose uppercase nucleotides encode an *EcoRV* site with a 3 base pair overhang.

4) 3' IN

5'-AAA *AGG CCT* tca tgc aaa aag agg gct cgc-3'

A 30mer whose lowercase nucleotides correspond to base pairs 837-857 of the RSV IN gene (Mumm, S. R. (1991)) and whose uppercase nucleotides encode a *Hind III* site with a 3 base pair overhang.

5) 5' SV40

5'-TTC TGA GGC GGA AAG AAC CA-3'

A 20mer whose nucleotides correspond to the complement of base pairs 292-273 of the SV40 genome (Genbank accession # J02400).

6) 3' SV40

5'-TTT GCA AAA GCC TAG GCC TC-3'

A 20mer whose nucleotides correspond to base pairs 5177-5196 of the SV40 genome (Genbank accession # J02400).

7) 5' DHFR

5'-ATG GTT CGA CCA TTG AAC TG-3'

A 20mer whose nucleotides corresponds to base pairs 56-75 of *Mus musculus* dihydrofolate reductase mRNA (Genbank accession # L26316).

8) 3' DHFR

5'-AGA TGC TCT TCT TTC TGA TT-5'

A 20mer whose nucleotides corresponds to base pairs 600- 619 of *Mus musculus* dihydrofolate reductase mRNA (Genbank accession # L26316).

9) pBluescript Forward Primer

5'-GTA AAA CGA CGG CCA GT-3'

A 17mer whose nucleotides correspond to base pairs 614 to 598of the pBluescript II KS- polycloning site.

10) pBluescript Reverse Primer

5'-AAC AGC TAT GAC CAT G-3'

A 16mer whose nucleotides correspond to base pairs 818-802 of the pBluescript II KS- polycloning site.

11) 5' Lac I Repressor Primer

5'-GTA TCT ctc gag AAA AGA ATG AAA CCA GTA ACG TTA TA-3'

A 38mer whose 3' nucleotides correspond to base pairs 1-20 of the Lac I Repressor from the *E. coli* DH5α genome (Genbank accession # G146576). The 5' terminus contains (in lowercase) a *Xho I* restriction site and a *KEX2* coding signal sequence for export in the yeast *Pichia pastoris*.

12) 3' Lac I Repressor Primer

5'-ATA AAG AAT gcg gcc gcT CAC TGC CCG CTT TCC AGT C-3'

A 37 mer whose 3' nucleotides correspond to base pairs 1180-1200 of the Lac I Repressor from the *E. coli* DH5α genome (Genbank accession #g146576). The 5' terminus contains (in lowercase) a *NotI* restriction site.

13) pBluescript T3 Primer

5'-ATT AAC CCT CAC AAT AG-3'

A 17mer whose nucleotides correspond to base pairs 785-768 of the vector pBluescript II KS- polycloning site.

14) Anti R Primer

5'-CTG TGT GAA ATT GTT ATC CG-3'

A 20mer whose nucleotides correspond to base pairs 821-840 of the vector pBluescript II KS- polycloning site.

The synthesized oligonucleotides were removed from the synthesis column with 2 ml of 8 M ammonium hydroxide flushed through the column using two 1 cc syringes fitted to the column ends. After incubation in a 55°C water bath overnight, the ammonium hydroxide was evaporated under a vacuum in a Sorvall Speed Vac Centrifuge for 3-4 hours. The oligonucleotide pellet was resuspended in 200 µl of distilled water and the absorbance at 260 nm measured. One absorbance unit at 260 nm was equivalent to 33 µg of single stranded DNA per ml.

F. AMPLIFICATION, LABELING, AND EXTENSION OF DNA

1) PCR AMPLIFICATION OF RSV IN

The Rous Sarcoma Virus integrase gene was amplified using the polymerase chain reaction (PCR) from a plasmid provided by Dr. D. P. Grandgenett (Mumm *et al.* (1991)). Amplification was carried out in order to introduce restriction endonuclease sites to facilitate cloning into the pMAL protein expression system. The reaction consisted of 5 ng of template DNA, 100 ng of oligonucleotide 3 (5' IN), 100 ng of oligonucleotide 4 (3' IN), 0.05 mM dNTPs, 1X Buffer E (67 mM Tris Base pH 9.01, 1.5 mM magnesium sulfate, 166 mM ammonium sulfate, 10 mM β-

mercaptoethanol), and 5 units of *T. aquaticus* polymerase in a total volume of 50 µl. After overlaying with 2 drops of mineral oil, the temperature of the mixture was cycled as follows; 30 seconds at 94 °C, 30 seconds at 50 °C, 1 minute at 72 °C, for 25 cycles. The reaction mixture was electrophoresed in a 1% agarose gel as described in Section C., 'Gel Electrophoresis', to ensure amplification of the desired DNA fragment size.

2) PCR AMPLIFICATION OF THE Lac I REPRESSOR GENE

The Lac I repressor gene was amplified using PCR and the *E. coli* DH5α genome as a template. A scraping of an *E. coli* colony was smeared within a 0.7 µl microfuge tube to serve as the amplification template. To the microfuge tube was added 100 ng of both 5' and 3' Lac I Repressor oligonucleotides, 2 µl of 1.25 mM dNTPs, 5 µl of 10X Buffer E (335 mM Tris pH 9.01, 7.5 mM magnesium sulfate, 830 mM ammonium sulfate, 50 mM β-mercaptoethanol), and 5 units of *T. aquaticus* DNA polymerase in a total volume of 50 µl. After overlaying with 2 drops of mineral oil, the temperature of the mixture was cycled as follows; 30 seconds at 94°C, 30 seconds at 50°C, 1 minute at 72°C, for 25 cycles. The reaction mixture was electrophoresed in a 1% agarose gel as described in Section C., 'Gel Electrophoresis', to ensure amplification of the desired DNA fragment size.

3) FILLING IN OF NDE-RSV INSERT

Oligonucleotides 1 and 2 (5' NDE-RSV and 3' NDE-RSV) have a 12 bp 3' complementary region, and thus are able to anneal to form a region of double stranded DNA. The single stranded regions were filled in using a reaction mixture comprising of 0.5 μg of each oligonucleotide, 0.05 mM dNTPs, 1X Buffer E, and 5 units of *T. aquaticus* DNA polymerase in a total volume of 50 μl. After being overlaid with 2 drops of mineral oil, the mixture was subjected to 2 cycles at 94°C for 30 seconds, 50°C for 30 seconds, 72°C for 2 minutes. The reaction mixture was electrophoresed in a 2% agarose gel to confirm extension of the annealed oligonucleotides.

4. POLYMERASE CHAIN REACTION ^{32}P PROBE AMPLIFICATION

^{32}P probes used in Southern Blot analysis were created by PCR. Using oligonucleotides 5 and 6 (5' SV40 and 3' SV40) or oligonucleotides 7 and 8 (5' DHFR and 3' DHFR), probes were created by amplifying either the SV40 promotor or the di-hydrofolate reductase (DHFR) gene of the substrate DNA (pNRdhfr). The reaction conditions were as follows: 1 ng of template DNA (pNDE-RSV), 50 ng each of the respective 5' and 3' oligonucleotides, 0.625 mM of dATP, dGTP, and dTTP, 10 μl of ^{32}P-dCTP, 3000 mCi/mmol, and 5 units *T. aquaticus* DNA polymerase in a total volume of 20 μl. The reaction mixture was subjected to 15 cycles of: 94°C for 30 seconds, 50°C for 30 seconds, and 72°C for 1 minute, after which the reaction volume was adjusted to 100 μl. An aliquot of the reaction (1 μl) was precipitated with 1 ml of 10%

trichloroacetic acid (TCA) for 10 minutes at 4°C, and filtered through a glass filter disk, which was subsequently washed with 5 ml of 10% TCA and 5 ml of ice cold 95% ethanol. The filter disk was placed in a scintillation vial along with 4 ml of scintillation fluid and the incorporation of ^{32}P-dCTP was detected by scintillation counting. Typically, the PCR amplified probes were labeled to a specific activity of approximately 5×10^9 dpm/ug of DNA

5) LABELING THE 3' ENDS OF DNA USING THE KLENOW FRAGMENT OF *E. coli* DNA POLYMERASE I

^{32}P labeled 3' ends of DNA were generated to study DNA/protein interactions by means of a mobility shift assay. DNA was digested with the endonuclease *Taq I* generating 5' overhang ends, 5'-GC-3'. In a 20 µl reaction mixture, 4 µg of digested DNA was mixed with 2 µl of 10X Klenow buffer (10 mM Tris, pH 7.6, 5 mM $MgCl_2$, 40 mM NaCl), 1 µl of 5 mM dGTP, and 80 µCi of ^{32}P-dCTP at 3000 Ci/mmol. The Klenow fragment of DNA polymerase I (1 unit) was mixed with the reaction and incubated for 15 minutes at 30°C, followed by 10 minutes at 75°C to inactivate the enzyme. Decays per minute were counted within a scintillation counter and the specific activity was typically 10^8 dpm/µg of DNA.

G. DNA SUBCLONING

1. RESTRICTION ENDONUCLEASE DIGESTION OF DNA

Unless otherwise indicated, restriction digests were carried out in a 20 µl volume which was comprised of 2 µl of an appropriate 10X React restriction endonuclease buffer (Gibco BRL). Plasmid DNA (1 µg) was digested with 5 units of the desired restriction endonuclease, for 1 hour at 37°C. DNA fragments were recovered by electrophoresis in a 1% agarose gel, and visualized by virtue of the ethidium bromide within the agarose gel preparation and exposure of the gel to UV light at 260 nm. Selected bands were excised from the agarose gel with a standard medical scalpel, and the gel slices were dissolved in a 2M sodium iodide solution by incubation at 55 °C for 5 minutes. Geneclean glass bead solution (10 µl) was added to the dissolved agarose slice and the mixture was incubated at 4°C for 5 minutes. The glass beads and bound DNA fragments were recovered by brief centrifugation for 10 seconds in an Eppendorf centrifuge at 14,000 rpm and washed three times with New wash solution (20 mM Tris pH 7.2, 0.2 M NaCl, 2 mM EDTA, 50% ethanol (v/v)). After the final wash, the glass beads were resuspended in 10 µl of distilled water and incubated for five minutes at 55°C and re-precipitated by brief centrifugation. The distilled water containing the purified DNA fragment was drawn off and could be used directly in ligation protocols.

2. LIGATION OF DNA FRAGMENTS TO FORM PLASMID CONSTRUCTS

The insert DNA fragment (0.5 µg) and the vector DNA backbone (0.5 µg) were ligated together in a 10 µl reaction volume which included 2µl of 5X T4 DNA ligase buffer and 1 unit of

T4 DNA ligase. Successful ligations were observed provided the insert DNA fragment and the vector backbone DNA fragment had compatible cohesive ends resulting from restriction endonuclease digestion.

3. TRANSFORMATION OF *E. coli* DH5α BY CLOSED, CIRCULAR PLASMID CONSTRUCTS

Competent host bacteria were prepared by calcium chloride treatment (Sambrook *et al.* (1989)). *E. coli* DH5α cells (50 ml) were grown at 37 °C to mid-log phase (OD 600 nm ~ 0.5-0.7), and centrifuged for 10 minutes at 3000 × g at 4°C. The pellet was resuspended in 25 ml ice cold 50 mM calcium chloride and incubated on ice for approximately one half hour. The treated cells were centrifuged and resuspended gently in 5 ml ice cold 50 mM calcium chloride. At this point, the cells were used immediately, or incubated 4°C for 4 hours to increase transformation efficiency. Approximately 0.3 μg of ligated plasmid DNA in a volume of 5 μl was added to 50 μl of prepared competent cells and left on ice for one half hour. The suspension was heat shocked at 42°C for 90 seconds and subsequently replaced in an ice bath for a further 2 minutes. LB broth (200 ml) was added to the suspension and the mixture was incubate for 45 minutes at 37°C in a rotary shacking incubator. Cells (~100μl) were spread on LB-ampicillin (50 μg/ml) plates saturated with X-gal (50 μg/ml) and IPTG (1 μg/ml) and incubated at 37°C overnight. In the selection procedure used, non-recombinant colonies were blue. Recombinant clones appeared clear by virtue of an interruption in the β-galactosidase promotor of the plasmid DNA by the insertion of the recombinant DNA into the poly-cloning site.

4. CONSTRUCTION OF INTEGRASE EXPRESSION VECTOR

The integrase gene of the Rous Sarcoma Virus (RSV) was obtained from Dr. D. P. Grandgenett of St. Louis University (Mumm *et al.* (1991)). The integrase protein of the retrovirus is the product of proteolytic cleavage, and thus does not possess an amino terminal methionine or the ATG start codon normally associated with transcribed and translated proteins. Dr. Grandgenett had inserted an ATG start codon immediately 5' to the coding sequence. The integrase gene was removed from the pGEM backbone by restriction digest with *EcoRI* . The digestion mixture was electrophoresed in a 1% agarose gel and the UV-illuminated DNA fragment corresponded to the recombinant IN gene (~850 bp) was excised and purified by the Geneclean protocol. The recombinant integrase DNA was ligated to similarly digested and purified pΔEsp1B (Graham *et al.* (1991)). Once the RSV integrase gene was checked for proper orientation, the resultant plasmid was termed pΔE-IN. An early Cytomegalovirus (CMV) promotor was removed from pRc/CMV by restriction digestion with *NruI* and *HindIII* , and ligated into *EcoRV* and *HindIII* digested pBluescript, and subsequently termed pBS-CMVprom. The CMV promotor was recleaved from pBS-CMVprom by digestion with *Cla I* and *Hind III* and ligated into similarily digested pΔE-IN. Lastly, a bovine growth hormone (BGH) poly adenylation signal was removed from pRc/CMV by digestion with *Xho I* and *Xba I* and ligated into similarily digested pΔE-IN (containing CMV promotor), forming pIN.

5. CONSTRUCTION OF INTEGRASE SUBSTRATE PLASMID, OR pNRdhfr

The plasmid to serve as a substrate for RSV integrase was constructed using pBluescript as a backbone. A dihydrofolate reductase (DHFR) cassette containing an SV40 promotor, a murine DHFR cDNA, and a Bovine Growth Hormone poly adenylation signal was obtained from the plasmid pNUT (Palmiter *et al.* (1989)) as a *KpnI* - *SalI* fragment and ligated into similarily digested sites of pBluescript to produce the plasmid pBS-DHFR. Oligonucleotides 1 and 2 (NDE-RSV 1 and NDE-RSV 2) were annealed and extended as described in Section F. After amplification, labeling, and extension of DNA, and the resultant dsDNA strand was digested with *KpnI* and ligated into *KpnI* digested pBS-DHFR to create the plasmid pNRdhfr. pNRneo was constructed in an analogous fashion with the removal of a neomycin cassette from pRc/CMV by digestion with *KpnI* and *BamHI* which was then ligated into similarily digested pBluescript instead of the DHFR cassette to form the plasmid pBS-NEO. Both pNRdhfr and pNRneo plasmids were linearized by digestion with *NdeI*, thereby exposing truncated RSV LTRs at both ends (Figure 5). The linearized plasmid was used as a substrate for RSV integrase.

6. CONSTRUCTION OF INTEGRASE BACTERIAL EXPRESSION PLASMID

The RSV integrase gene was amplified by means of polymerase chain reaction (PCR) as previously described in Section F., Amplification, labeling, and extension of DNA, in order to insert restriction endonuclease sites which would facilitate cloning of the gene into the bacterial

expression vector pMAL. The amplified RSV integrase gene was digested with *EcoRV* and *HindIII* creating a 5' blunt end for ligation into pMAL which was digested with the endonuclease *XmnI* , and *HindIII*. Such a ligation produced a gene construct whose 5' end encoded a maltose binding protein (MalE) and 3' end which encoded RSV integrase. The resultant plasmid was termed pMAL-IN

7) CONSTRUCTION OF Lac I REPRESSOR PROTEIN EXPRESSION PLASMID

The Lac I repressor gene was amplified by means of PCR which inserted a 5' *XhoI* and 3' *NotI* restriction endonuclease site. The Lac I Repressor gene was subcloned into pBS using the *XhoI* and *NotI* endonuclease sites. Four pBS-Lac I clones were selected and sequenced to determine the possible presence of PCR induced errors (only clone 3 contained an error). Originally the 5' *Xho I* site was to serve for directional cloning into the yeast, *Pichia pastoris*, pPIC 9 expression vector. *P. pastoris* Lac I expression proved unsuitable as no protein was excreted. The Lac I repressor gene was then cloned into the bacterial expression vector pTrc99A. In order to ensure that the 5', ATG, start codon was within 15 bp of the ribosome landing site, Mung Bean nuclease was used to remove the 5' overhanging ends generated by *XhoI* and *NotI* endonuclease digestion. This fragment was ligated into *NcoI* digested pTrc99A which was treated with Mung Bean nuclease. Colonies formed after ligation and transformation were screened by endonuclease digestion using *HincII* which resolved those clones containing both the Lac I repressor gene and its orientation. The resultant plasmid was termed pTrc 99A-Lac I.

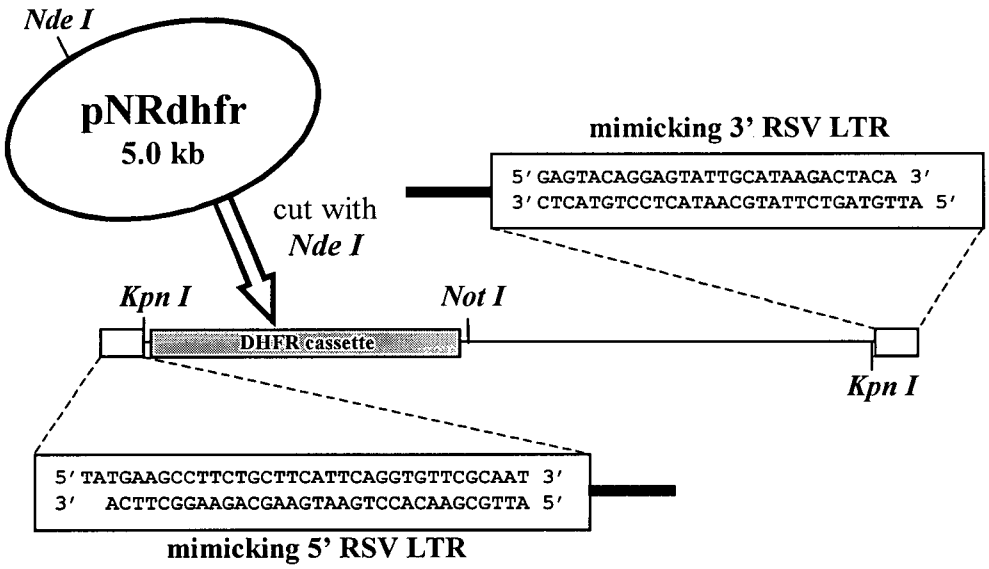

FIG. 5. **Schematic representation of the pNRdhfr substrate plasmid.** pNRdhfr contains truncated integrase recognition elements derived from the Rous Sarcoma Virus long terminal repeat (LTR) sequences which are separated by a unique *Nde*I restriction site. The plasmid also contains a dihydrofolate (DHFR) cassette to allow selection of transfected cells based on resistance to methotrexate. Digestion of pNDE-RSV with *NdeI* linearizes the plasmid and exposes a 34 bp U3 portion of the RSV LTR at one terminus of the molecule and a 29 bp U5 portion at the other terminus. Linearization with a different restriction enzyme, for example *Not*I (located 3' to the DHFR cassette), does not expose the LTR-derived elements recognizable by integrase.

H. EXPRESSION AND PURIFICATION OF PROTEINS

1) ROUS SARCOMA VIRUS INTEGRASE

Selected bacterial colonies containing pMAL-IN were grown to an OD of 0.5 at 600 nm in 1 liter of LB broth at 37°C. The expression of the maltose binding-integrase fusion protein was induced by addition of 2.5 mM IPTG, and the culture was allowed to incubate for a further two hours at 37°C. The bacteria were collected by gentle centrifugation (3000 × g in a swinging bucket rotor), washed, and resuspended in a sonication buffer (20 mM Tris pH 7.4, 200 mM NaCl, 1 mM EDTA). Bacteria cell lysis was carried out by sonication (Heat Systems Sonifier at output level 2) of the resuspended bacteria for eight bursts of 15 seconds at 4°C. Disrupted cell membranes and denatured proteins were precipitated by centrifugation in a Sorvall Super Speed centrifuge using a SS34 rotor at 5000 × g for 15 minutes. The supernatant was poured into a fresh centrifuge tube and centrifuged to remove any residual precipitate. The supernatant was then ready to be passed through a maltose column, which would bind the fusion protein. The column was washed with 8 column volumes of pMAL column buffer, after which the fusion protein was eluted from the column by the inclusion of 10 mM maltose. Fractions containing the fusion protein, as determined by absorbance readings at 280 nm, were pooled and the fusion protein cleaved by digestion with the bovine protease Factor Xa (0.1 % w/v of total fusion protein). The cleaved fusion protein was once again passed down the washed maltose column, allowing the RSV integrase to pass through while the maltose binding protein remained bound to the column. Fractions containing the RSV integrase protein were pooled and concentrated using a mini-

Amicon concentrator and a 10,000 kDa cutoff filter. Concentrated RSV integrase protein was electrophoresed on a polyacrylamide gel as described in Section C., 'Gel electrophoresis'. Once stained with Coomassie brilliant blue G250, and destained (50% methanol, 10% glacial acetic acid), the purified integrase protein was clearly visible at a position corresponding to 32 kDa.

2) Lac I REPRESSOR

Two liters of *E. coli* DH5α containing a clone of pTrc 99A-Lac I was grown to an absorbance at 600 nm of 0.5, at which time the culture was induced with 2.5 mM IPTG. After 2 hours, the cells were harvested at 3000 × g in a Beckman GSA rotor at 4°C. The culture yielded approximately 16 g of wet cells which was resuspended in 800 mls of pMAL column buffer (20 mM Tris pH 7.4, 1 mM EDTA, 200 mM NaCl) and frozen at -20°C overnight. Aliquots (35 mls) were sonicated for a total of 2 minutes in 15 second burst. The resulting lysate was precipitated by centrifugation for 30 minutes, 4°C, at 3000 × g in a Beckman GSA rotor. To the resulting supernatant was added 231 mg/ml of ammonium sulfate (18.48 g) to precipitate the Lac I repressor (Platt, T. (1973)). The precipitation was carried out for 1 hour at 4°C with gentle stirring after which time the mixture was precipitated by centrifugation for 30 minutes at 3000 × g in a Beckman GSA rotor at 4°C. The pellet was resuspended in 50 mls of Buffer B (0.2 M KCl, 0.2 M Tris pH 7.4, 0.01 MgAc, 0.3 mM M DTT and 5% glycerol). The resuspension mixture was dialyzed against 3 changes of a 0.03 M Buffer KP (0.025 M KH_2PO_4, 0.005 M K_2HPO_4, 3×10^{-4} M DTT and 5% glycerol at pH 7.4). Insoluble particulate matter was removed by centrifugation. The dialyzed extract was loaded onto an Econo-Pac High S Cartridge which was first

equilibrated with 0.03 M Buffer KP. The High S column was washed with 0.08 M Buffer KP (0.067 M KH_2PO_4, 0.013 M K_2HPO_4, 3×10^{-4} M DTT and 5% glycerol at pH 7.4), and the bound Lac I repressor protein was eluted with 25 mls of 0.20 M Buffer KP (0.167 M KH_2PO_4, 0.033 M K_2HPO_4, 3×10^{-4} M DTT and 5% glycerol at pH 7.4). The Lac I repressor protein was precipitated from the eluent by the addition 231 mg/ml (5.78g) of ammonium sulfate and precipitated for 1 hour at 4°C. The precipitate was collected by centrifugation over 30 minutes at 3000 × g in a Beckman SS34 rotor. A white pellet resulted which resuspended completely in 5 mls of a Tris buffered saline solution (20 mM Tris pH 7.6, 40 mM NaCl, 30% glycerol). A BioRad protein assay confirmed a concentration of 7.2 mg/ml, and the 36 mg total of Lac I repressor protein gave a yield of 2.25 mg of Lac I repressor protein per gram of wet cells.

3) NUCLEAR EXTRACTS OF BABY HAMSTER KIDNEY CELLS

Baby Hamster Kidney cells were grown to confluency in 10 cm tissue culture plates at which time the culture medium was removed, and the cells washed with 5 mls of PBS. The cells were scraped into a 1 ml pool of fresh PBS, and transferred to a 1.7 ml Eppendorf centrifuge tube. Cells were collected by centrifugation for 10 minutes at 500 × g in an Eppendorf microfuge. The supernatant was discarded, and the collected cell volume ascertained. The cells were quickly resuspended in 5 cell volumes of hypotonic buffer (10 mM HEPES pH 7.9 at 4°C, 1.5 mM $MgCl_2$, 10 mM KCl, 0.2 mM PMSF, and 0.5 mM DTT). The resuspended cells were collected by centrifugation for 5 minutes at 500 × g in an Eppendorf microfuge. The cell pellet was resuspended in 3 cell volumes of hypotonic buffer, and allowed to swell for 10 minutes on ice.

The cells were transferred to a glass Dounce homogenizer, and the cell membrane was disrupted with ten up-and-down strokes using a type B pestle. The mixture was transferred to a 1.7 ml Eppendorf tube, and the nuclei collected by centrifugation for 15 minutes at 3300 × g.

The collected nuclei were resuspended in a volume of low-salt buffer (20 mM HEPES pH 7.9 at 4°C, 25% glycerol, 1.5 mM $MgCl_2$, 20 mM KCl, 0.2 mM EDTA, 0.2 mM PMSF, 0.2 mM DTT) equal to one half the packed nuclear volume. In a drop wise fashion while stirring gently, a one-half packed nuclear volume of high-salt buffer (20 mM HEPES pH 7.9 at 4°C, 25% glycerol, 1.5 mM $MgCl_2$, 1.2 M KCl, 0.2 mM EDTA, 0.2 mM PMSF, 0.2 mM DTT) was added. The nuclei were incubated on ice for 30 minutes with gentle mixing. The nuclei were then collected by centrifugation for 30 minutes at 25,000 × g, and the resulting supernatant was dialyzed against a 100 mM KCl dialysis buffer (20 mM HEPES pH 7.9 at 4°C, 20% glycerol, 100 mM KCl, 0.2 mM EDTA, 0.2 mM PMSF, 0.2 mM DTT). The final nuclear extract was stored in aliquots at -80°C.

I. MAMMALIAN CELL CULTURE AND TRANSFECTION

1. MAMMALIAN CELL CULTURE

Transformed baby hamster kidney cells (BHK-21) were originally purchased from American Tissue Type Collection (ATTC) and were cultured in 1X Dulbecco's Modified Eagle Medium, nutrient mixture F-12, supplemented with 5% newborn bovine calf sera, in a 37°C incubator with a 5% carbon dioxide environment. BHK cells were cultured within Falcon 6 well plates or 90 mm dishes containing 2 ml of supplemented media per well or 10 ml per dish. Upon confluency, BHK cells were split by first removing the media, adding a minimal volume of 0.25% trypsin, removing trypsin after 30 seconds, and finally suspending the cells in 2 ml of DMEM-F12 media by passing them repeatedly through a 3 ml disposable pipette. BHK cells could then be transferred to multiple flasks and grown as before. If the resuspension mixture was supplemented with 10% dimethyl sulfoxide (DMSO), suspended cells could be transferred to a 2 ml cryovial and stored indefinitely in liquid nitrogen. Frozen cells could be thawed and cultured by transferring the cryovial contents to a flask containing media supplemented with 5% newborn calf serum.

2. MAMMALIAN CELL TRANSFECTION USING CALCIUM PHOSPHATE

Baby hamster kidney (BHK) cells were grown in 10 cm plates (Falcon) using DMEM-F12 media supplemented with 5% newborn calf serum. Transfection was initiated once the cells had attained approximately 70% confluency. Medium was changed four hours prior to the start of transfection (9 ml/plate). Typically 10 μg of the desired DNA was precipitated using 0.1 volumes of 3 M sodium acetate pH 5.6 and 2 volumes of 95% ethanol followed by centrifugation in an Eppendorf microfuge at 14,000 rpm for 15 minutes. From this point on, the DNA samples were considered sterile and the supernatant was removed in a laminar flow hood. The DNA pellet was allowed to air dry for 20 minutes, after which it was resuspended in 450 μl of sterile water, 50 μl of 2.5 M calcium chloride, and 500 μl of 2X HEPES buffered saline (2×HBS: 8 g sodium chloride, 0.37 g potassium chloride, 0.1 g di-sodium phosphate, 1 g dextrose, and 5 g HEPES in 100 ml of distilled water, pH 6.95). After a 20 minute incubation at room temperature, the solution was added directly to the plate and swirled to ensure complete mixing. The plate was placed in a incubator at 37 °C, 5% carbon dioxide, from between 4 to 16 hours, during which time the DNA/calcium phosphate co-precipitate formed and settled as a fine suspension onto the BHK cells. To terminate the transfection, the media was changed and cells left to recover for 24 hours.

3. MAMMALIAN CELL TRANSFECTION USING LIPOSOMES

A liposome mixture consisting of a 1:1 ratio of DOPE:DODAC was a gift provided by Dr. P. Cullis of the Department of Biochemistry at the University of British Columbia. As before, the

BHK cells were grown to approximately 70% confluence in Eagle's DMEM-F12 media supplemented with 5% newborn calf serum in Falcon six well plates. Four hours prior to transfection, the media was replaced with 2 ml/well of DMEM-F12 without any newborn calf serum or antibiotics. Typically 4 μg of DNA would be precipitated using 0.1 volumes of 3 M sodium acetate pH 5.6 and 2 volumes of 95% ethanol and centrifugation in an Eppendorf microfuge at 14,000 rpm for 15 minutes. From this point the DNA samples were considered sterile and the supernatant was removed in a laminar flow hood. The DNA pellet was allowed to air dry for 20 minutes, after which it was resuspended in 200 μl of serum free media. DOPE:DODAC (9 μl) was added to 200 μl of serum free media and mixed with the DNA/media resuspension. After a 20 minute incubation, 66 μl of the DNA/liposome mixture was added to each well, and the transfection was allowed to proceed overnight at which time it was terminated by replacing the media with DMEM-F12 supplemented with newborn serum.

4. SELECTION OF INTEGRATED PLASMID

Transfected plasmids which integrated into the BHK cell genome could be selected by virtue of their dihydrofolate reductase (DHFR) cassette. Selection was carried out by the inclusion of 500 μM methotrexate with the DMEM-F12, 5% newborn calf serum media. After approximately 2 days, the BHK cells became visibly abnormal (rounding up, and lifting from the well surface), and died in increasing numbers until day 10, after which surviving colonies were clearly visible.

As the cells died, it was important that the media be changed every day, as dead cell matter was observed to inhibit the growth of the surviving cells. Surviving colonies were visible to the eye but were counted under a microscope at 100X magnification while the well was placed on a transparent 4 mm grid.

5. EXPANSION OF SELECTED COLONIES

Colonies were 'picked' by addition of a drop of 0.25% trypsin to the colony after removal of the media. Immediately following the trypsin exposure, a disposable 3 ml pipette containing 1 ml of DMEM-F12 media was used to remove the colony from the plate; gently squeezing the bulb of the pipette, a half drop was formed at the end of the pipette and was placed on top of the colony and the bulb released so that the colony was removed from the surface into the pipette. The contents of the pipette were then expelled into a Falcon 25 cm^2 T-flask containing 5 ml of DMEM-F12 media supplemented with 5% newborn calf serum. The flasks were incubated at 37°C incubator in a 5% carbon dioxide environment.

J. GENOMIC DNA ISOLATION AND BLOT HYBRIDIZATION

1. PREPARATION OF GENOMIC DNA FROM TISSUE CULTURE

BHK cells grown as a monolayer on 10 cm plates. Genomic DNA was isolated by first removing the medium from the monolayer of cells and overlaying with 4 mls of extraction buffer (20 mM Tris pH 7.6, 1 mM EDTA, 100 μg/ml pronase, 10 μg/ml RNAse, 0.1% SDS). After a digestion period of no less than 4 hours, the viscous mixture of lysed cells was removed from the plate and transferred to a 14 ml disposable Falcon tube. The viscous lysate was extracted with an equal volume of phenol:chloroform, and after gentle mixing over 5 minutes, the phases were resolved with a 10 minute spin in a bench top, swinging bucket centrifuge at 3000 × g at room temperature. The upper aqueous phase was removed and placed in a 14 ml disposable Falcon culture tube, and the genomic DNA precipitated with 5 ml of ice cold 95% ethanol. The DNA formed a white mass which sunk to the bottom of a tube and was removed with a sterile glass Pasteur pipette. The DNA pellet was washed 5 times in 5 ml of 95% ethanol, until the DNA formed a small condensed pellet which was transferred to a 5 ml disposable Falcon tube containing 3 ml of distilled water. The DNA pellet was re-suspended by slow rotation of the tube over 2 days at 4°C

2. SOUTHERN BLOT ANALYSIS

Southern blots were prepared as described in Current Protocols in Molecular Biology, Chapter 2 (Ausubel *et al.* (1996)). Typically 10 μg of genomic DNA was digested with restriction enzymes and electrophoresed in a 0.7% agarose gel at low voltage over a period of 16 hours.

The agarose gel was submerged in 0.1 N hydrochloric acid, causing limited depurination, for 15 minutes. The gel was then transferred to a tray above a 0.4 N sodium hydroxide reservoir from which the sodium hydroxide was wicked to the tray by 3 overlaid rectangles of 3MM Whatman paper. This method ensured continuous flow of the buffer through the gel, on which was stacked a positively charged nylon membrane (Hybond N+, Amersham), 4 more rectangles of 3MM Whatman paper and 5 inches of paper towels cut to size. The 0.4 N sodium hydroxide served to denature the DNA within the agarose gel, but also to cause strand breakage where the HCl had caused depurination. By creating smaller denatured fragments, the transfer onto the charged nylon membrane was accomplished in approximately 5 hours. After washing the nylon membrane with 5X SSC (20X SSC: 3.0 M sodium chloride, 3.0 M trisodium citrate, pH 7.5) and allowing excess moisture to evaporate, the nylon membrane was placed in a microwave and heated on the highest setting for two and a half minutes to irreversibly bind the DNA to the membrane.

The nylon membrane was placed in a hybridization tube (Hybaid) with 10 ml of pre-hybridization solution (7% SDS, 0.25 M di-sodium phosphate, and 1 mM EDTA pH 8.0) and incubated in a rotating chamber at 68 °C for 15 minutes. The probe, prepared as described in Section F., 'Amplification, labelling and extension of DNA', was diluted in 1 ml of pre-hybridization solution, denatured by placing it in a boiling water bath for 10 minutes, and added directly to the pre-hybridization solution already within the tube containing the nylon membrane. The tube was replaced within the rotating chamber and left to hybridize overnight, or for a period no less than 8 hours, at 68 °C.

The hybridization solution was poured off and replaced with 100 ml of 2X SSC and 2% SDS, preheated to 68 °C. After rotation at 68 °C for 15 minutes the solution was replaced with fresh 2X SSC and 2% SDS, and again incubated for 15 minutes. Subsequent washes included two with 1X SSC and 1% SDS, two with 0.2X SSC and 1% SDS, and two with 0.1X SSC and 1% SDS, all in the rotating incubator at 68 °C.

The washed blot was sandwiched between two sheets of Saran Wrap and imaged by placing under a sheet of autoradiography film and developed after a 1 to 7 day exposure, or on a phosphorimaging screen and imaged after 24 hours using a Molecular Dynamics Phosphorimager.

K. DNA SEQUENCE ANALYSIS

1. DNA SEQUENCE ANALYSIS USING MODIFIED T7 DNA POLYMERASE

Sequenase is a modified version of T7 DNA polymerase. The properties of Sequenase include high processivity which enable it to process through self-complementing loop structures of a DNA strand which would normally impede a DNA polymerase *in vitro*. After the addition of 100 ng of a sequencing primer, DNA (10 μg) was denatured with 0.2 N sodium hydroxide in a volume of 20 μl by placing in a boiling water bath for 2 minutes and snap cooled on ice. The denatured DNA was precipitated along with the annealed sequencing primer by the addition of

0.1 volumes of 3 M sodium acetate pH 5.6, 2.5 volumes of 95% ethanol, and centrifugation in an Eppendorf micro-centrifuge for 15 minutes at 14,000 rpm. DNA was resuspended in 10 µl of 1X Annealing Buffer (5X: 200 mM Tris base pH 7.5, 100 mM magnesium chloride, 250 mM sodium chloride). To the DNA template and annealed primer was added 1 µl of 0.1 M di-thiothreitol, 2µl 1X labelling mix, 5 µCi α^{35}S-dATP, 2 µl of a 1:8 dilution of *Sequenase* enzyme, and this 'labelling reaction' was left to incubate at room temperature for 2 minutes. In a 60 well micro-plate (Nunc), 2.5 µl of 80 µM dideoxyguanosine, dideoxyadenosine, dideoxythymidine, and dideoxycytidine in 50 mM sodium chloride were added to separate wells and prewarmed to 37 °C on a hot plate. 3.5 µl of the labelling reaction was added to each of the wells containing the respective dideoxy-nucleotides and left to incubate at 37 °C for 5 minutes, after which 4µl of stop solution (95% formamide, 20 mM EDTA, 0.05% bromophenol blue, and 0.05% xylene cyanol FF) was added to terminate the reaction. The Nunc plate, wrapped in Saran Wrap, was placed in a boiling water bath for 2 minutes and immediately transferred to ice. The reactions were then electrophoresed on a 10% denaturing polyacrylamide gel with a constant power of 55 watts applied for at least 90 minutes. For sequences distant to the annealed primer, the polyacrylamide gels were electrophoresed for up to 5 hours at 55 watts.

L. *IN VITRO* ASSAY FOR INTEGRASE ACTIVITY

Either purified integrase or Baby Hamster Kidney cell nuclear extracts were used to detect *in vitro* integrase activity. Purified integrase was added to an integration buffer (50 mM

HEPES pH 8.0, 2 mM β-mercaptoethanol, 0.01% BSA, 3 mM $MnCl_2$, 4% glycerol, 50 mM NaCl) to give a final concentration between 1 and 10 μM. Alternatively, 1 μl of either 100 fold diluted, 10 fold diluted or straight BHK nuclear extract was added to the integration buffer. *NdeI* linearized pNRdhfr, displaying the NDE-RSV terminal ends, was added to a concentration 25 μM, and the final volume adjusted to 10 μl. The mixture was incubated at 30°C for 30 minutes, and electrophoresed through a 1% agarose gel in preparation for transfer and Southern blotting .

M. MOBILITY SHIFT ASSAY

The basic protocol for the mobility shift assay was obtained from Current Protocols in Molecular Biology (Ausubel *et al.* (1996)). This protocol was divided into 4 parts: preparation of a radioactively labeled probe (see section F. 'Amplification, Labelling, and Extension of DNA'); preparation of a non-denaturing gel (see section C. 'Gel Electrophoresis'); a binding reaction; and electrophoresis of the protein/DNA complex through the gel, which was then dried and autoradiographed (see section C. 'Gel Electrophoresis').

1) BINDING REACTION FOR MOBILITY SHIFT ASSAY

For each reaction tube, 3 μl of 5X Binding Buffer (30% glycerol, 100 mM Tris pH 7.6, 50 mM KCl, 5 mM EDTA) was mixed with 20,000 cpm of radioactively labeled DNA (~0.5 ng) containing the lactose operon, 2 μg nonspecific carrier DNA, 300 ug/ml bovine serum albumin, and brought to a total reaction volume of 14 μl. 1 μl of Lac I Repressor protein (~5 to 25 ng) was then added to the tube which was incubated for 30 minutes at room temperature. The entire reaction was then pipetted into a 0.75 cm spaced comb atop a non-denaturing polyacrylamide gel

and electrophoresed at 25 mA for an hour and a half, the gel dried and subsequently autoradiographed.

N. NITROCELLULOSE CAPTURE ASSAY

This protocol relies on the ability of nitrocellulose to bind proteins but not double-stranded DNA. Purified Lac I Repressor protein was mixed with DNA containing the lactose operon, and after incubation, the mixture was filtered through nitrocellulose under suction, allowing unbound DNA to pass through the filter while the Lac I Repressor protein and any bound DNA was retained. The bound DNA was subsequently eluted from the filter with 10 mM IPTG.

1) BIND DNA TO PROTEIN

DNA (1 to 50 μg) which had previously been digested by a restriction endonuclease, was mixed with 1X Binding Buffer (12% glycerol, 20 mM Tris pH 7.6, 50 mM KCl, 1 mM EDTA) in a total volume of 49 μl. Lac I Repressor protein (1-10 μg in a 1 μl volume) was added, and the mixture allowed to incubate at room temperature for 30 minutes.

2) FILTER DNA/PROTEIN COMPLEX

Using a 1 cm diameter, two piece Millipore filter apparatus, and a side arm flask connected to a vacuum source, the DNA/protein mixture was added to a nitrocellulose membrane sandwiched within the Millipore filter apparatus. The filter was washed three times with 0.5 ml

of 1X Binding Buffer, and the bound DNA was eluted from the membrane bound protein by washing with 0.5 ml of 1X Binding Buffer with 100 mM IPTG. The IPTG eluent containing the previously bound DNA was precipitated upon addition of 0.1 volumes of 3M sodium acetate and 2.5 volumes of 95% ethanol. The precipitate was resuspended in a total of 20 µl of distilled water. Using 1 ug aliquots of pNRdhfr, which contained a lactose operon, approximately 75 % of the original DNA was recovered from the precipitated IPTG eluent using 7 µg of Lac I Repressor protein.

O. ANTIBODY PRODUCTION AND *IN SITU* ANTIBODY APPLICATIONS

Norwegian white rabbits were used to produce a polyclonal antibody against Rous Sarcoma Virus integrase. The protocol was approved by the local Animal Care Committee prior to use.

1) INTRAMUSCULAR INJECTION OF PROTEIN/ADJUVANT MIXTURE.

Complete Freund's adjuvant is a water-oil emulsion prepared with nonmetabolizable oils containing killed *M. tuberculosis*. Freund's adjuvant produces a strong and prolonged immune response. RSV integrase (100 µg) in a 200 µl volume was mixed with 200 µls of complete Freund's adjuvant. Using glass syringes, equal volumes of the protein/adjuvant mixture were taken up into different syringes. Air was removed from the syringes and they were connected

through a luer fitting. Alternatively pushing the plungers, the solution was mixed to form an emulsion. The total emulsion was transferred to one syringe, the air removed and an 18 gauge needle attached.

With one person immobilizing the rabbit on a rough plastic surface, a rear leg was grasped from the front and the injection was made into the rear of the thigh muscle near the hip. Once the needle was inserted, the plunger was withdrawn slightly. If resistance was encountered then the injection proceeded, if blood appeared, the syringe was withdrawn and the injection moved to a nearby site. The innoculum was added slowly in a steady motion and the needle withdrawn. Two boosts were given as described above at 4 week intervals.

2) SERUM COLLECTION

With one person restraining the rabbit on a rough plastic surface, the marginal vein was located on the inner edge of the ear (a lamp often facilitates its location). A patch was shaved around the vein about two-thirds of the distance from the head to the tip of the ear. At a 45° angle to the vein, a cut was made with a clean sterile scalpel. It is essential to keep the rabbit calm and warm. The flow of blood could be increased by gentle heating with a lamp. The blood was allowed to drip into a clean test tube, with a maximum volume of 50 mls collected. The flow of blood was stopped by applying gentle pressure to the cut with sterile gauze. The blood was allowed to clot for 1 hour at 37° C, at which time the clot was removed from the sides of the tube with a Pasteur pipette and the clot placed at 4° C overnight. The serum was then removed and centrifuged at 10,000g for 10 minutes at 4° C to remove any particulate matter. The serum was distributed into 1 ml aliquots and was stored at -70° C indefinitely.

3) *IN SITU* CELL STAINING WITH ANTIBODIES

Baby Hamster Kidney cells were plated on a tissue culture dish and allowed to adhere for at least 24 hours. To avoid overcrowding, the cells were plated at a low density. When the cells had reached the desired confluency, the media was removed and the cells were rinsed with two changes of ice cold Tris buffered saline (TBS: 25 mM Tris pH 8.0, 200 mM NaCl). Cells were fixed for four minutes using a 4% formaldehyde/TBS solution, after which they were rinsed with two washes of ice cold TBS, and the cells permeabilized by incubation by immersion in 100% methanol for 2 minutes at room temperature.

Using TBS containing 3% bovine serum albumin, the polyclonal antibody for RSV integrase was diluted 1000 fold and added in minimal amounts to the fixed cells within the tissue culture plate. The plate was left to incubate overnight at room temperature with gentle agitation provided by a rotary shaker. The primary antibody solution was washed with three changes of TBS. The secondary anti rabbit-alkaline phosphatase conjugated antibody was diluted 1000 fold in a TBS-3% BSA solution and once again applied to the fixed cells which were left to incubate for at least 8 hours at room temperature with gentle agitation. The secondary antibody solution was washed with three changes of TBS.

In order to visualize the binding of the primary and secondary antibodies, bromochloroindoyl phosphate/nitro blue tetrazolium (BCIP/NBT), a substrate for alkaline phosphatase, was used to generate an intense black-purple precipitate at the site of alkaline

phosphatase binding. The reaction proceeded at a steady rate which allowed for accurate control of the development of the reaction.

Just prior to developing, a fresh substrate solution was prepared by adding 66 µl of NBT stock (0.05 g/ml, 70% dimethylformamide) to 10 ml of alkaline phosphatase buffer (100 mM NaCl, 5 mM Mg Cl_2, 100 mM Tris pH 9.5). BCIP stock (33 µl) (0.05 g/ml, 100% dimethylformamide) was added last, mixed well, and used within 1 hour. The washed cells were overlaid with the substrate solution and allowed to develop at room temperature with gentle agitation. Typically a suitable stain develop within a half hour, at which time the reaction was stopped by addition of EDTA to 100 mM, the substrate solution drained, and the cells washed with three changes of TBS.

RESULTS:

A. *IN VITRO* ASSAY FOR ROUS SARCOMA VIRUS INTEGRASE ACTIVITY.

The activity of the Rous Sarcoma Virus integrase clone provided by Dr. D. P. Grandgenett (Grandgenett *et al.* (1993)) was assayed for integration activity by *in vitro* analysis. The RSV IN protein, substrate DNA (lacking a selection cassette), and an integration buffer were combined and incubated for 30 minutes at 30°C. The products of the reaction were electrophoresed through an agarose gel and blotted onto a charged nylon membrane for Southern analysis. While the substrate DNA electrophoresed to a position corresponding to 3000 bp (Figure 6), two other bands were evident as the concentration of both IN and substrate DNA increased. The bands were situated at approximately 6000 and 9000 bp and were diffuse in appearance (Figure 6). These higher molecular weight DNA bands were the product of integration of substrate DNA molecules. The diffuse band corresponding to 3000 bp was the result of a substrate DNA molecule integrating into a second substrate DNA molecule. As observed by others (Bushman *et al.* (1994)), the diffuse nature of the band was probably due to small deletions and branched molecules which arose from *in vitro* integration. Similarly, the band corresponding to 9000 bp was probably the result of the integration of 3 substrate molecules. As concentrations of both IN and substrate DNA increased, aggregates of DNA/protein complexes formed that were unable to enter the gel under these electrophoresis conditions. These complexes were evident as a band at the position

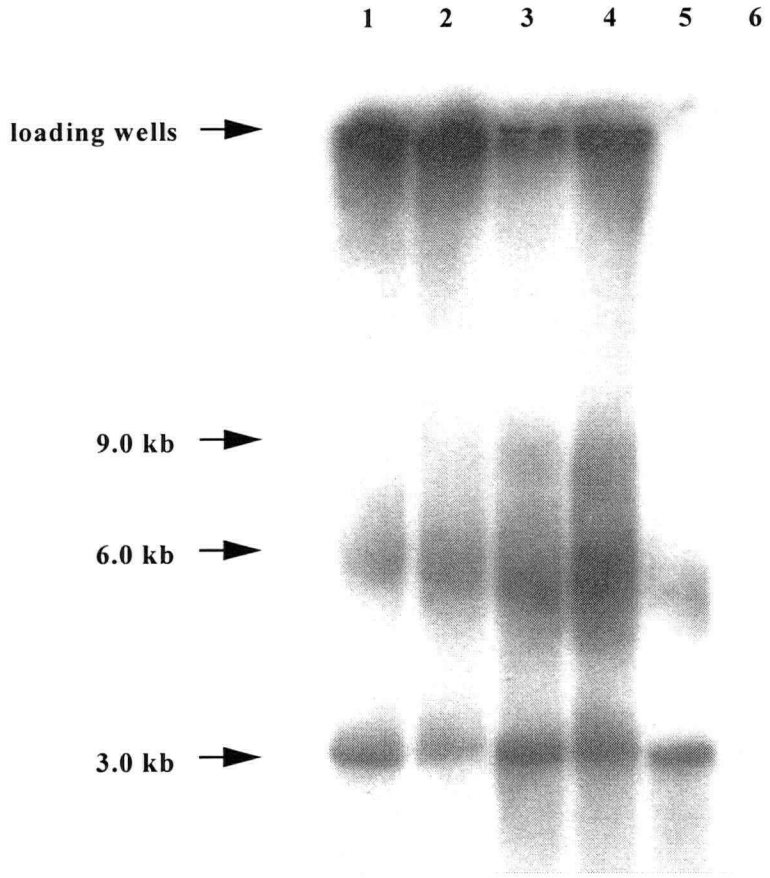

FIG. 6. ***In vitro* assay for RSV IN activity.** RSV IN and substrate DNA was incubated in a ligation buffer, and the products electrophoresed and blotted on a nylon membrane for Southern analysis. The band at 3 kb corresponds to the substrate DNA while higher molecular weight bands are the products of integration events. *Lane 1* contained 2 μM RSV IN, 12.5 μM substrate DNA. *Lane 2* contained 4 μM RSV IN, 12.5 μM substrate DNA. *Lane 3* contained 2 μM RSV IN, 25 μM substrate DNA. *Lane 4* contained 4 μM RSV IN, 25 μM substrate DNA. *Lane 5* contained 12.5 μM substrate DNA, the fuzzy band near 6 kb was relaxed circular DNA which has been observed in other *in vitro* assays (Bushman *et al* (1991)), and *lane 6* contained 2 μM RSV IN.

corresponding to the loading lanes. If either the IN protein or substrate DNA was not included in the reaction mixture, no higher molecular weight bands corresponding to integration events were evident.

B. EFFECT OF CO-TRANSFECTION OF BHK CELLS WITH pIN AND pNR.

BHK cells were co-transfected with liposomes containing an integrase expression vector (pIN) and *NdeI*-digested substrate plasmid (pNR). Linearization of pNR exposes integrase (IN) recognition elements at both termini of the molecule (Figure 5) and it was reasoned that, in the presence of intracellular integrase, these elements would promote incorporation of the substrate DNA into the host cell genome. Either a DHFR cassette or a neomycin cassette was engineered into pNR which served as a selectable marker for integration by conferring resistance to either methotrexate or G418.

As shown in Figure 7, *plate 6*, cells co-transfected with pIN and substrate DNA (*NdeI*-linearized pNRdhfr) prepared together in the same liposome mixture produced 10-fold more colonies than cells transfected with only substrate DNA, *plate 2* (6.56×10^{-7} versus 6.44×10^{-6} colonies per cell); the latter transfections provided an index of the number of colonies resulting from random recombination events. When cells were co-transfected with pIN and *NotI*-linearized pNRdhfr (which does not possess terminal IN recognition elements; see Figure 5), the integration efficiency was comparable to that

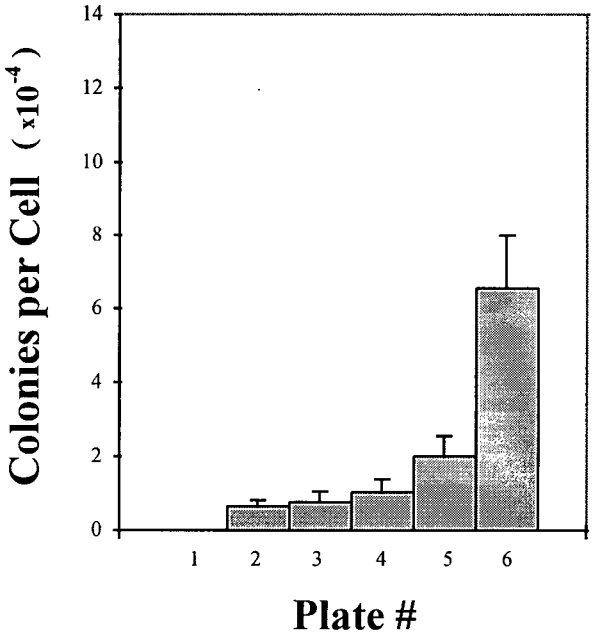

FIG. 7. **Transfection of BHK cells with liposome-loaded plasmids.** Liposome mediated transfections and colony counts were carried out as described in Experimental Procedures. Results are plotted as the mean for the 6 wells of each culture plate and error bars indicate one standard deviation from the mean. Results for each well were standardized by dividing the number of colonies formed by the number of cells originally seeded into each well. Cells in each well of plates 1 were transfected with 0.33 μg pNR$^-$ (i.e. lacking a DHFR cassette) and cells in each well of plate 2 were transfected with 0.33 μg pNRdhfr. Cells in each well of plates 3-6 were co-transfected with: *plate 3*, 0.33 μg pNRdhfr and 0.33 μg pCMV-A1; *plate 4*, 0.33 μg pIN and 0.33 μg *NotI*-linearized pNRdhfr; *plate 5*, 0.33 μg pIN and 0.33 μg pNRdhfr (precipitated individually and resuspended in separate liposome mixtures); *plate 6*, 0.33 μg pIN and 0.33 μgpNRdhfr. Error bars represent standard deviation of n = 6 (Table 2).

A)

	Colony counts (6 well)						mean	std. dev.
Plate 1, pNR (no DHFR)	0	0	1	0	0	0	0.2	0.4
Plate 2, pNR	11	9	13	14	10	9	11.0	2.1
Plate 3, pNR, pCMV-A1 (no IN)	15	19	12	10	15	13	14.0	3.1
Plate 4, pNR (*NotI* linearized), pIN	24	17	21	11	15	21	18.2	4.8
Plate 5, pNR, pIN (separate liposomes)	41	25	34	30	48	40	36.3	8.3
Plate 6, pNR, pIN (co-transfections)	115	88	119	150	118	116	117.7	19.7

B)

	(1.8 E06 cells per well) colonies per cell ($\times 10^{-6}$)	std. dev. ($\times 10^{-6}$)
Plate 1	0.093	0.23
Plate 2	6.1	1.2
Plate 3	7.8	1.7
Plate 4	10	2.6
Plate 5	20	4.6
Plate 6	65	1.1

TABLE 2. **Transfection of BHK cells with liposome-loaded plasmids.** Chart A) indicates the actual colony counts per well of a six well Falcon Plate as well as mean and standard deviation. Chart B) contains the average colonies per well which were standardized for the amount of cells present at the time of transfection.

observed when transfecting with *NdeI*-linearized substrate plasmid alone (*plate 4*, 1.01×10^{-5} colonies per cell). Similarly, low transfection efficiencies were seen if pIN was substituted with a different plasmid of comparable size (pCMV-A1, *plate 3*, 7.78×10^{-6} colonies per cell). Delivery of pIN and substrate plasmid in separate liposome preparations produced 2.5-fold fewer colonies than delivering both plasmids in the same liposome preparation. It was also observed that colonies which formed from cells co-transfected with pIN and *NdeI*-linearized pNRdhfr consistently exhibited more robust growth than the controls (Fig. 8).

In order to determine conditions for optimizing the co-transfection protocol, two experiments were performed. In the first, the ratio of pIN:pNRdhfr was varied without altering the total amount of DNA delivered to the cells. This showed that the concentration of substrate DNA was a limiting factor; the number of colonies formed decreased as the amount of substrate DNA was reduced, even though increasing amounts of pIN were being delivered to the cells. Figure 9a depicts the reduction in colonies formed as less substrate DNA and more pIN were transfected. With a total of 0.66 μg of DNA transfected into each well, the ratio of substrate DNA to pIN changed: *plate 1* 7:1, 1.14×10^{-4} colonies per cell, *plate 2* 3:1, 8.92×10^{-5} colonies per cell, *plate 3* 1.7:1, 7.3×10^{-5} colonies per cell, *plate 4* 1:1, 6.47×10^{-5} colonies per cell, *plate5* 1:1.6, 5.73×10^{-5} colonies per cell, *plate 6* 1:3, 3.99×10^{-5} colonies per cell, and *plate 7* 1:7, 2.76×10^{-5} colonies per cell. Figure 9b served as the control reactions since *plates 1-7* were transfected with the same amount of substrate DNA as Figure 9a, however no pIN was co-transfected.

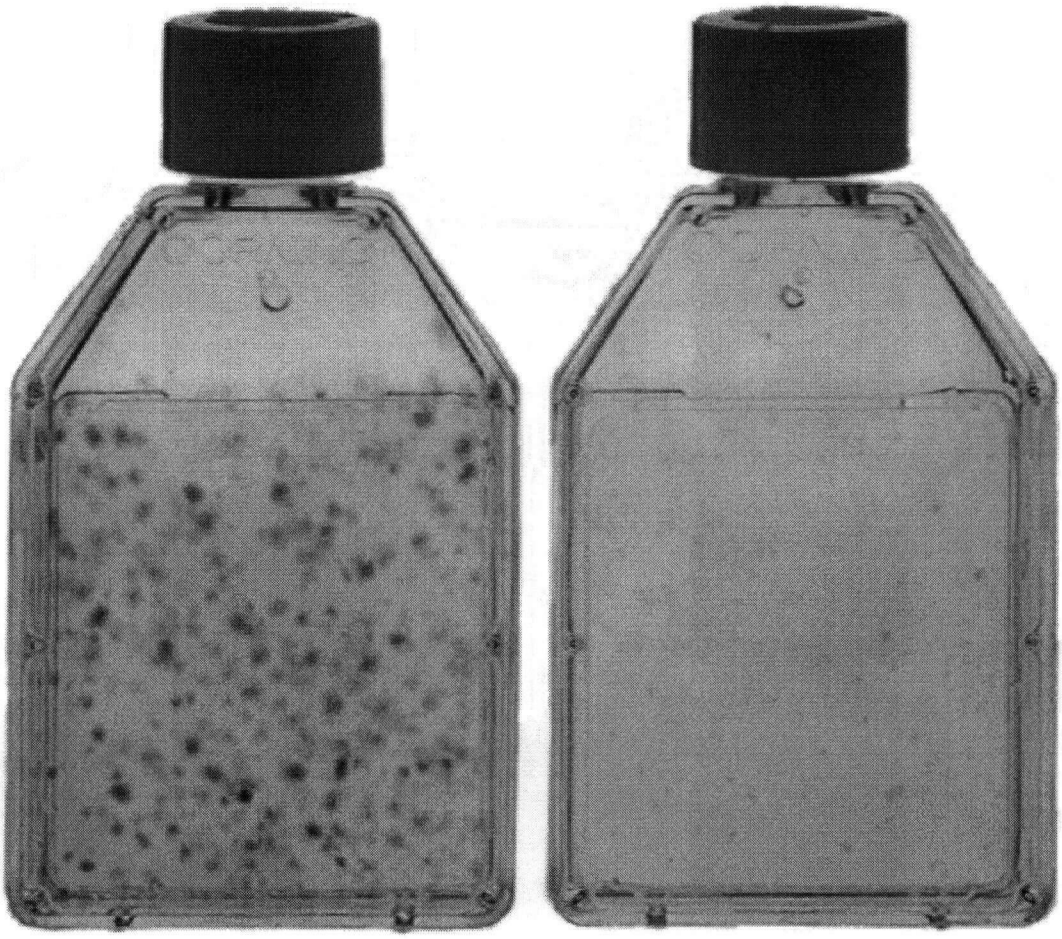

FIG. 8. **Comparison of efficiency of co-transfection with pIN and pNRdhfr to transfection only with pNRdhfr.** BHK cells at 70% confluency were transfected, grown for 10 days in media supplemented with methotrexate, and stained with Trypan Blue as described in Experimental Procedures. Cells in the left flask, co-transfected with 2.44 μg *NdeI*-linearized pNRdhfr and 0.81 μg pIN produced 335 colonies. Cells in the right flask, transfected only with 2.44 μg *NdeI*-linearized pNRdhfr produced 31 colonies

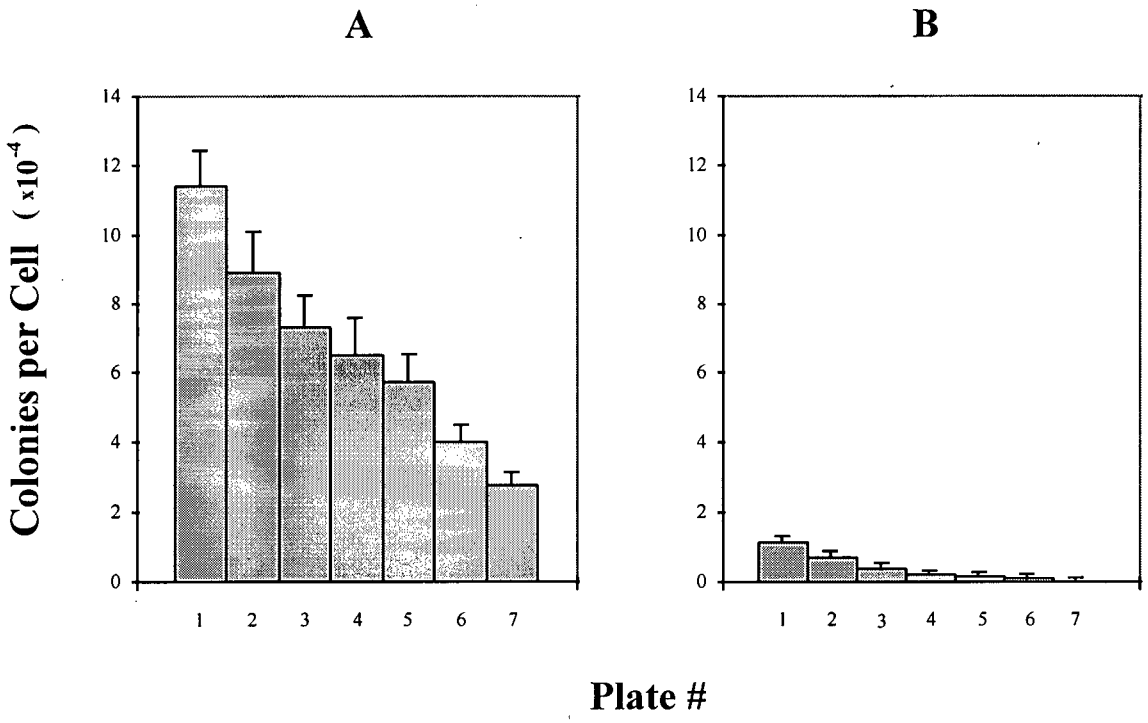

FIG 9. **Effect of varying the pIN:pNRdhfr ratio on transfection efficiency.** The amounts of pIN and *Nde*I-linearized pNRdhfr used in co-transfections were varied (A) and compared to transfections using only *Nde*I-linearized pNRdhfr (B). In A, cells in each well of a 6-well culture plate were co-transfected with: *plate 1*, 0.08 μg pIN and 0.58 μg pNRdhfr; *plate 2*, 0.17 μg pIN and 0.50 μg pNRdhfr; *plate 3*, 0.25 μg pIN and 0.42 μg pNRdhfr; *plate 4*, 0.33 μg pIN and 0.33 μg pNRdhfr; *plate 5*, 0.42 μg pIN and 0.25 μg pNRdhfr; *plate 6*, 0.50 μg pIN and 0.17 μg pNRdhfr; *plate 7*, 0.58 μg pIN and 0.08 μg pNRdhfr. In B, cells on plates 1-7 were transfected with 0.58, 0.50, 0.42, 0.33, 0.25, 0.17, 0.08 μg/well of pNRdhfr, respectively (i.e. amounts equivalent to those used in A). After growth for 10 days in media supplemented with methotrexate, the number of colonies in each well was counted. The results were plotted as described in the legend to Fig. 7. Error bars represent a standard deviation of n = 12 (Table 3).

A)	Colony counts (6 well)						mean	std. dev.
Plate 1 (0.5 ug pIN, 3.5 ug pNR)	137	108	121	134	119	118	122.8	10.8
Plate 2 (1.0 ug pIN, 3.0 ug pNR)	107	80	77	110	80	89	90.5	14.5
Plate 3 (1.5 ug pIN, 2.5 ug pNR)	102	79	100	76	73	81	85.2	12.6
Plate 4 (2.0 ug pIN, 2.0 ug pNR)	94	72	69	89	76	91	81.8	10.8
Plate 5 (2.5 ug pIN, 1.5 ug pNR)	79	57	72	81	73	75	72.8	8.5
Plate 6 (3.0 ug pIN, 1.0 ug pNR)	60	58	61	53	49	63	57.3	5.3
Plate 7 (3.5 ug pIN, 0.5 ug pNR)	46	39	41	40	31	35	38.7	5.2

B)	(1.2 E06 cells per well) colonies per cell	std. dev.
Plate 1 ($\times 10^{-6}$)	102	9
Plate 2 ($\times 10^{-6}$)	75	12
Plate 3 ($\times 10^{-6}$)	71	10
Plate 4 ($\times 10^{-6}$)	68	9
Plate 5 ($\times 10^{-6}$)	61	7
Plate 6 ($\times 10^{-6}$)	48	4
Plate 7 ($\times 10^{-6}$)	32	4

C)	Colony counts (6 well)						mean	std. dev.
Plate 1 (0.5 ug pIN, 3.5 ug pNR)	119	92	119	127	90	108	109.2	15.3
Plate 2 (1.0 ug pIN, 3.0 ug pNR)	110	78	81				89.7	17.7
Plate 3 (1.5 ug pIN, 2.5 ug pNR)	68	64	52	82	58	68	65.3	10.3
Plate 4 (2.0 ug pIN, 2.0 ug pNR)	79	57	43	63	38	40	53.3	16.0
Plate 5 (2.5 ug pIN, 1.5 ug pNR)	66	38	44	46	42	45	46.8	9.8
Plate 6 (3.0 ug pIN, 1.0 ug pNR)	37	30	24	22	26		27.8	5.9
Plate 7 (3.5 ug pIN, 0.5 ug pNR)	22	20	25	14	18	21	20.0	3.7

D)	(8.7 E05 cells per well) colonies per cell	std. dev.
Plate 1 ($\times 10^{-6}$)	125	18
Plate 2 ($\times 10^{-6}$)	103	20
Plate 3 ($\times 10^{-6}$)	75	12
Plate 4 ($\times 10^{-6}$)	61	18
Plate 5 ($\times 10^{-6}$)	54	11
Plate 6 ($\times 10^{-6}$)	32	6.8
Plate 7 ($\times 10^{-6}$)	23	4.3

E)	Average of two experiments colonies per cell	std. dev.
Plate 1 ($\times 10^{-6}$)	114	13
Plate 2 ($\times 10^{-6}$)	89	16
Plate 3 ($\times 10^{-6}$)	73	11
Plate 4 ($\times 10^{-6}$)	65	14
Plate 5 ($\times 10^{-6}$)	57	9
Plate 6 ($\times 10^{-6}$)	40	6
Plate 7 ($\times 10^{-6}$)	28	4

TABLE 3. **Effect of varying the pIN:pNRdhfr ratio on transfection.** Chart A) indicates the actual colony counts per well of a six well Falcon Plate of trial 1, as well as

Table 3 con't.

mean and standard deviation. Chart B) contains the mean colonies per well of trial 1, which were standardized for the amount of cells present at the time of transfection. Chart B) indicates the actual colony counts per well of a six well Falcon Plate of trial 2, as well as mean and standard deviation. Chart D) contains the mean colonies per well of trial 2, which were standardized for the amount of cells present at the time of transfection. Chart E) contains the corrected mean and standard deviation from trials 1 and 2.

In the second experiment, increasing amounts of both pIN and pNRdhfr were delivered to the cells. This resulted in a linear increase in the number of colonies formed (Fig. 10). A three to one ratio of substrate DNA to pIN was used with the total DNA transfected increasing with plate number. Beginning with *plate 1*, 0.66 μg DNA transfected, 8.38×10^{-5} colonies per cell developed after methotrexate selection. *Plate 9* had double the amount of DNA transfected, 1.32 μg, and 1.72×10^{-5} colonies per cell developed, almost exactly double that of *plate 1*. The linear increase was observed until such a point that the levels of lipid present for transfection became toxic to the cells (results not shown).

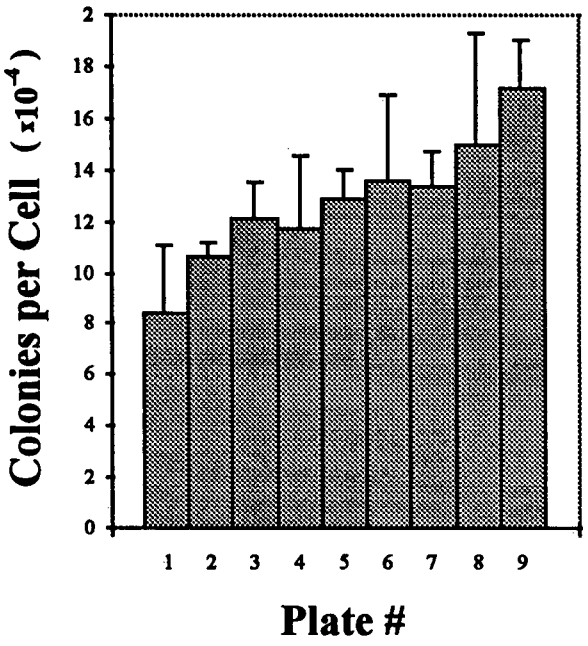

FIG. 10. **Effect of concomitantly increasing the amounts of pIN and pNRdhfr on transfection efficiency.** Cells in each well of a 6-well culture plate were co-transfected with increasing amounts of DNA as follows: *plate 1*; 0.17 μg pIN and 0.50 μg pNRdhfr; *plate 2*, 0.19 μg pIN and 0.56 μg pNRdhfr; *plate 3*, 0.21 μg pIN and 0.63 μg pNRdhfr; *plate 4*, 0.23 μg pIN and 0.69 μg pNRdhfr; *plate 5*, 0.25 μg pIN and 0.75 μg pNRdhfr; *plate 6*, 0.27 μg pIN and 0..81 μg pNRdhfr; *plate 7*, 0.29 μg pIN and 0.88 μg pNRdhfr; *plate 8*, 0.31 μg pIN and 0.93 μg pNRdhfr; *plate 9*, 0.33 μg pIN and 1.0 μg pNRdhfr. After growth for 10 days in media supplemented with methotrexate, the number of colonies in each well was counted. The results were plotted as described in the legend to Fig. 7. Error bars represent a standard deviation of n = 3 (Table 4).

	25 cm T-flasks			mean	std.dev.	2.1 E06 cells per flask colonies per cell ($\times 10^{-6}$)	std. dev. ($\times 10^{-6}$)
Flask 1 (0.5 ug pIN, 1.5 ug pNR)	195	113	220	176.0	56.0	84	27
Flask 2 (0.56 ug pIN, 1.68 ug pNR)	228	231	211	223.3	10.8	110	5.1
Flask 3 (0.63 ug pIN, 1.88 ug pNR)	279	224	259	254.0	27.8	120	13
Flask 4 (0.69 ug pIN, 2.06 ug pNR)	309	199	230	246.0	56.7	120	27
Flask 5 (0.75 ug pIN, 2.25 ug pNR)	286	256		271.0	21.2	130	10
Flask 6 (0.81 ug pIN, 2.44 ug pNR)	335	311	211	285.7	65.8	140	31
Flask 7 (0.88 ug pIN, 2.63ug pNR)	257	275	309	280.3	26.4	130	13
Flask 8 (0.94 ug pIN, 2.8 ug pNR)	226	402	316	314.7	88.0	150	42
Flask 9 (1.0 ug pIN, 3.0 ug pNR)	400	326	355	360.3	37.3	170	18

TABLE 4. **Effect of concomitantly increasing the amounts of pIN and pNRdhfr on transfection efficiency.** The r^2 value as determined by a Pearson product was 0.91, indicating a strong linear trend.

C. CO-TRANSFECTION OF pIN AND pNR IN THREE MAMMALIAN CELL LINES.

To determine if increased transfection efficiencies carried over to other cell lines, three mammalian cell types were co-transfected with pIN and an *NdeI* digested substrate plasmid (pNRneo). The first cell line was Baby Hamster Kidney cells, already well characterized with respect to co-transfections of pIN and pNR; it acted as a control. The second cell type was 293, a transformed human embryonic kidney cell line developed by Dr. F. Graham (Graham *et al.* (1991)). Transformed by the E1 region of adenovirus type 5, 293 cells were primarily used to rescue recombinant adenoviruses and propagate them in cell culture conditions. The third cell line was CV-1, transformed African green monkey kidney cells.

pNRneo was used as a selectable marker since the 293 and CV-1 cell lines were resistant to methotrexate selection. G418 was used to select for pNRneo recombinants which contained the neomycin cassette. G418 resistant clones had an altered response to selection when compared to methotrexate. When pNRneo and pIN were co-transfected at levels comparable for methotrexate selection, selection under G418 resulted in a confluent lawn. The amount of transfected pNRneo which produced distinct colonies was at or below 0.17 μg/well. G418 selection also tended to produce 'fuzzy' colonies whose borders were not distinct, but tended to blend with other colonies. This effect was minimized by transfecting cells at a confluency level below 60%. At low confluency, cells tended to sucuumb to G418 selection quickly since division was occurring rapidly, and colonies which formed were far enough removed from neighbors such that distinct colonies formed.

All three cell types responded to G418 selection quite differently. BHK cells with the fastest doubling time of 12 hours were the quickest to die off after selection had commenced at 400 μg/ml G418. All cells which had not incorporated a neomycin cassette were dead by day six of selection. HEK 293 cells and CV-1 cells did not succumb to G418 concentrations of 400 μg/ml until at least day ten. This was presumably due to their slower doubling time of approximately 24 hours. When G418 levels were increased to 800 μg/ml, 293 and CV-1 recombinants were selected much more quickly; however, 293 cells exibited ten fold fewer colonies and CV-1 cells exhibited no survivors.

Figure 11 indicated that increased transfection efficiencies are recorded for all three cell lines when co-transfected with pIN and substrate DNA (*NdeI* linearized pNRneo) versus substrate DNA alone. When BHK cells were co-transfected with 1.62 μg pIN and 0.08 μg pNRneo, 494.7 ± 40.8 colonies resulted per well after G418 selection. BHK cells transfected with 0.08 μg pNRneo, resulted in 212.7 ± 19.9 colonies per well. pIN had caused a 2.3 fold increase in BHK cell colony formation. When 293 cells were co-transfected with 1.62 μg pIN and 0.08 μg pNRneo, 287.0 ± 50.6 colonies resulted per well. 293 cells transfected with 0.08 μg pNRneo, resulted in 170.0 ± 31.5 colonies per well. pIN had caused a 1.7 fold increase in 293 cell line colony formation. When CV-1 cells were co-transfected with 1.62 μg pIN and 0.08 μg pNRneo, 225.3 ± 21.1 colonies

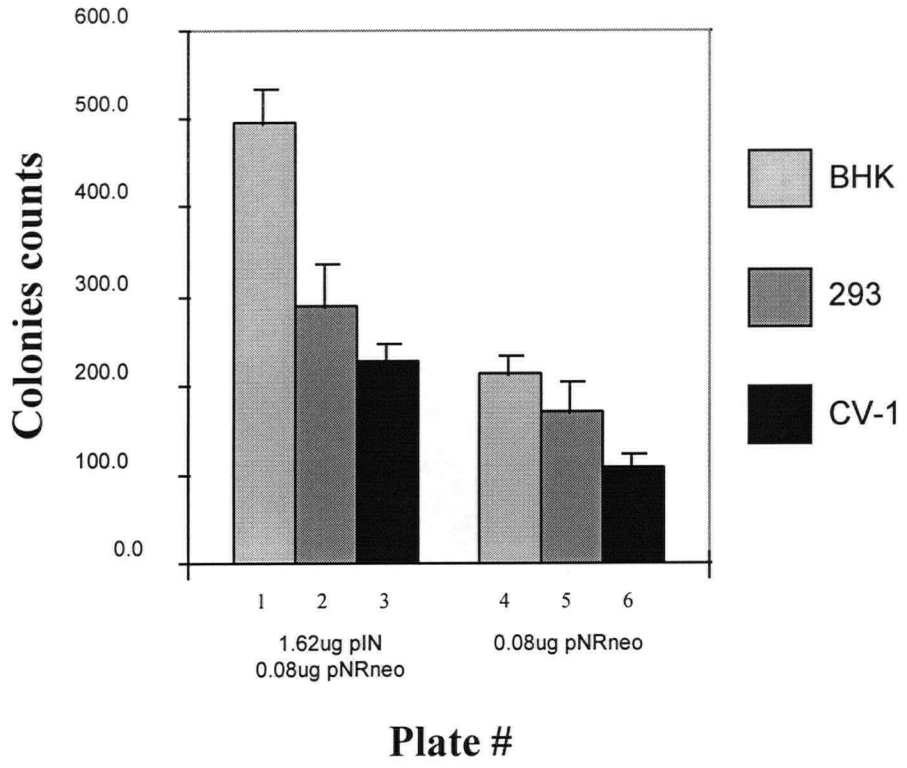

FIG. 11. **Effect of co-transfection with pIN and pNRneo on three different cell lines.** Co-transfections were carried out in order to assess wether the increased transfection efficiency caused by pIN and substrate DNA occurs in different cell lines. Transfections were carried out when the cells had reached 60% confluency. *Plates 1* through *3* were co-transfected with 1.62 μg/well of pIN and 0.08 μg/well *NdeI* linearized pNRneo. *Plate 1* was comprised of BHK cells, *plate 2* was 293 cells, and *plate 3* was CV-1 cells. *Plates 4* through *6* were transfected with 0.08 μg/well *NdeI* linearized pNRneo. *Plate 4* was comprised of BHK cells, *plate 5* was 293 cells, and *plate 6* was CV-1 cells. Selection was carried out with G418 at a concentration of 575 μM.

resulted per well. 293 cells transfected with 0.08 μg pNRneo, resulted in 109.7 ± 14.2 colonies per well. pIN had caused a 2.1 fold increase in 293 cell colony formation.

D. SOUTHERN BLOT ANALYSIS OF BHK GENOMIC DNA FROM COLONIES PREVIOUSLY CO-TRANSFECTED WITH pIN AND pNR.

Genomic DNA isolated from cells co-transfected with pIN and *NdeI*-linearized pNRdhfr was digested with *KpnI*, Southern-blotted, and probed with a ^{32}P-labeled SV40 probe (Figure 12). If the 5 kb substrate DNA had been preserved intact during recombination, it would be expected that *KpnI* digestion would release a fragment of ~5 kb, representing almost all of the pNR plasmid (Figure 5). It was apparent from Figure 12, *lanes 4, 6, 7, 8,* and *9* that much of the signal on the blot was concentrated at ~5 kb. These bands migrated slightly more quickly after electrophoresis than *lane 1*, the control substrate DNA. This was due to the digestion of *lanes 4, 6, 7,* and *9* with the endonuclease *KpnI* which cuts approximately 30 bp within the terminal ends of *NdeI* digested pNRdhfr. The main 5 kb signal of the autoradiogram indicated that much of the substrate DNA was incorporated into the host cell genome intact. Genomic DNA from untransfected cells did not bind the probe. Several lighter bands above and below the 5 kb main signal were detected in *lanes 4, 6, 7, 8, and 9* containing DNA from transfected cells. Genomic DNA arising from individual colonies which had been transfected with pIN and substrate DNA displayed distinct banding for each individual clone. As expected, pooled colonies did not

display significant, distinct banding since the pooling of upwards of 200 colonies created a smear of all the possible fragment sizes from separate clonal recombinants.

A second Southern Blot (Figure 13) was performed to analyze the banding pattern of BHK genomic DNA transfected with only substrate DNA. Figure 13 shows identical banding patterns for genomic DNA collected from pooled colonies arising from transfection of substrate DNA (*lanes 5* and *6*) alone or co-transfection of both pIN and substrate DNA (*lanes 8* and *9*). This indicated that the intact recombination of pNR occurred in the absence of RSV IN expressed from pIN, even though inclusion of pIN increased the formation of colonies approximately ten fold after selection with 500 μM methotrexate.

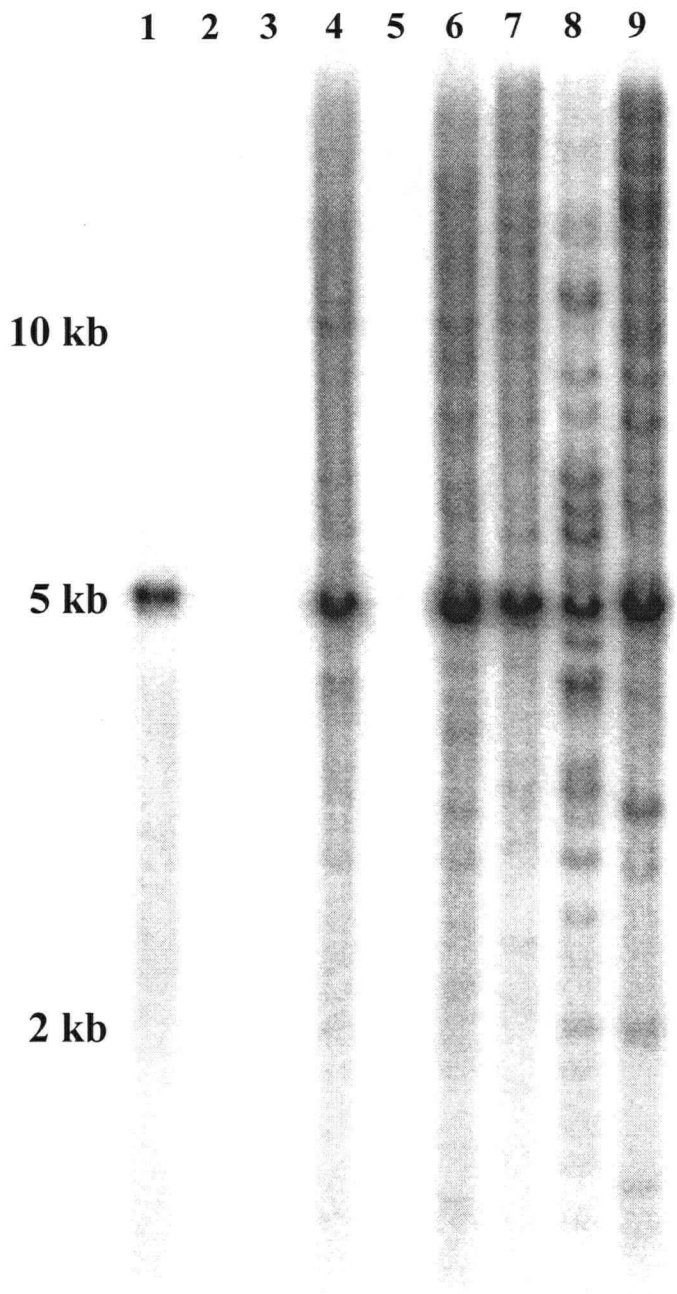

FIG.12. **Southern blot analysis of genomic DNA.** Genomic DNA from BHK cells was digested with *KpnI*, Southern-blotted, and probed with a ^{32}P-labeled SV40 promotor

(FIG. 12. continued)

probe. *KpnI* sites in pNRdhfr are located ~30 bp inside each LTR, and thus *KpnI* digestion of genomic DNA containing an intact substrate DNA molecule would be expected to release a 5 kb fragment representing almost all of the integrated pNRdhfr molecule (Fig. 6). *Lane 1*, 1 ng *NdeI*-digested pNRdhfr; *lane 3*, *KpnI*-digested DNA from untransfected BHK cells; *lane 4*, *KpnI*-digested DNA from pooled confluent colonies arising from cells co-transfected with pIN and *NdeI*-linearized pNRdhfr; *lanes 6-9*, *KpnI*-digested DNA from 4 separate, expanded colonies arising from cells co-transfected with pIN and *NdeI*-linearized pNRdhfr. *Lanes 2* and *5* contained no DNA.

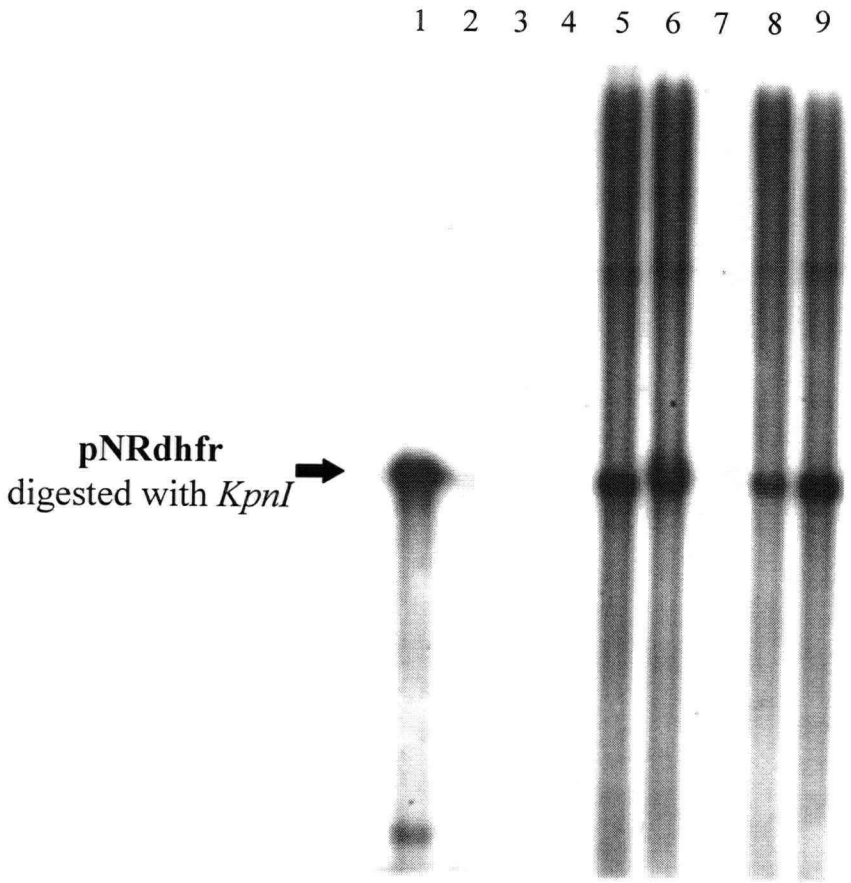

FIG. 13. **Southern blot analysis of genomic DNA.** *Lane 1*, 1 ng *KpnI*-digested pNRdhfr. *Lane 3*, *KpnI*-digested DNA from untransfected BHK cells. *Lane 5*, *KpnI*-digested DNA from pooled colonies arising from cells transfected with 0.17 μg/well *NdeI*-linearized pNRdhfr. *Lane 6*, *KpnI*-digested DNA from pooled colonies arising from cells transfected with 0.34 μg/well *NdeI*-linearized pNRdhfr. *Lane 8*, *KpnI*-digested DNA from pooled colonies arising from cells co-transfected with 1.62 μg/well pIN and 0.17 μg/well *NdeI*-linearized pNRdhfr. *Lane 9*, *KpnI*-digested DNA from pooled colonies arising from cells co-transfected with 1.62 μg/well pIN and 0.17 μg/well *NdeI*-linearized pNRdhfr. *Lanes 2, 4* and *7* contained no DNA.

E. PURIFICATION OF Lac I REPRESSOR PROTEIN.

The Lac I repressor protein was needed to purify fragments of DNA containing a portion of the pNR backbone and adjoining genomic DNA by virtue of a lactose operon present within the pNR backbone. This clonal recombination site in turn gives information on the type of recombination which had occured as well as the position within the BHK genome.

The Lac I repressor protein expressed by the bacterium *E. coli* is comprised of 360 amino acids that associate to form a homotetramer of 154 520 dalton molecular mass. The lactose operator region which binds the Lac I repressor is a 27 bp DNA segment which is bound by the Lac I repressor protein and is situated 11 bp downstream from the start site of the *lacZ* gene of *E. coli*. This operator has an axis of approximate dyad symmetry, a common feature of cis-acting sequences which bind transcription factors, especially in prokaryotes. The natural inducer of the repressor molecule is allolactose; however a gratuitous inducer, isopropyl-β-D-1-thiogalactoside (IPTG), can substitute for allolactose and cause the release of the LacI repressor from the lac operator DNA (Lewis *et al.* (1996)).

The *lacI* gene was amplified from the *E. coli* DH5α genome by PCR. The 5' oligonucleotide primer contained a portion of the α-factor export signal for the yeast *Pichia pastoris* as well as a Kex 2 cleavage signal to remove the export signal peptide from the expressed Lac I repressor protein. To check for the absence of PCR errors, four clones of the *lac I* gene (Genbank accession # J01636) were ligated into pBluescript and their nucleotide sequence determined. Of the four clones analysed, only one clone contained a nucleotide

substitution, a cytosine nucleotide was replaced by a thymidine nucleotide at position 204. This nucleotide switch in turn would cause a threonine → glycine substitution upon translation. Clone 4, which contained no PCR errors, was subcloned into the yeast expression vector pPIC9. Further studies on Lac I repressor protein expression in the yeast *P.pastoris* failed as no protein was expressed.

As a result it was decided that the *lac I* gene would be expressed in *E. coli* bacteria using the bacterial expression plasmid pTrc99A. The *lac I* gene was cloned into pTrc99A and *lac I* transcription and subsequent translation was induced by the addition of IPTG (Figure 14). The first purification step was an ammonium sulfate precipitation of the crude extract. The Lac I repressor protein at a concentration of 37.5% (Figure 15), which precipitated all of the Lac I repressor protein, together with a number of other proteins. Further purification was carried out by column chromatography through a Bio-Rad High S column, a strong cation exchanger which binds the basic N-terminal region of the Lac I repressor protein. The column chromatography was sufficent to purify the Lac I repressor protein completely as determined by gel electrophoresis through a SDS polyacrylamide gel followed by Coomassie brilliant blue staining (Figure 16). The eluant was concentrated by a second ammonium sulfate precipitation, which resuspended completely to yield 36 mg of pure protein in 5 mls of TBS.

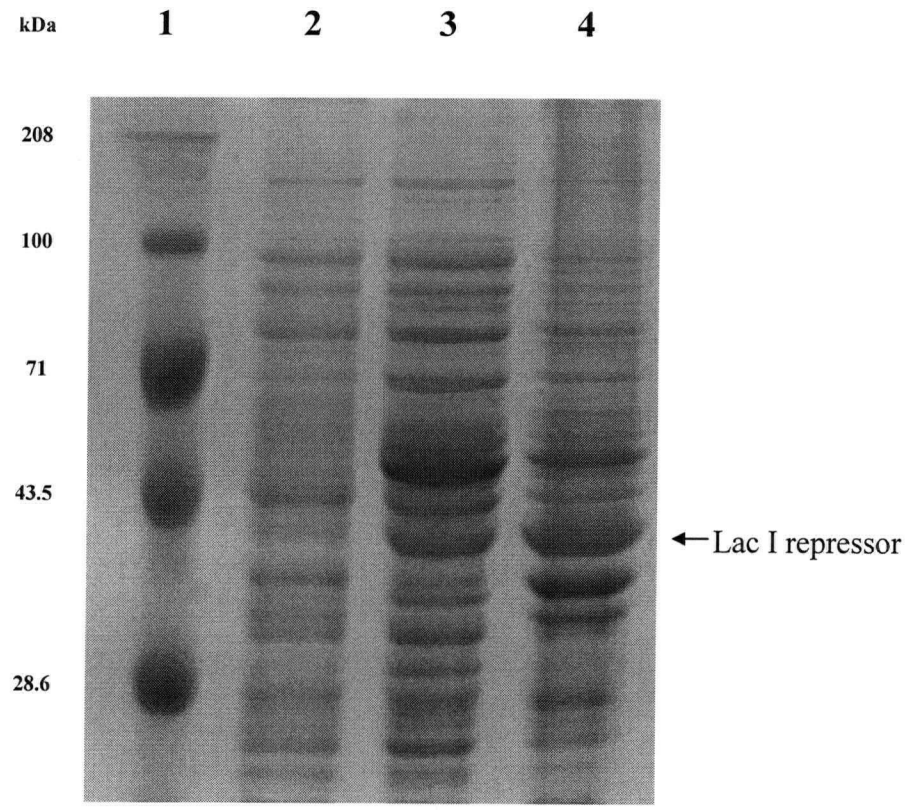

FIG. 14. **Induction of Lac I repressor expression with IPTG.** *E. coli* DH5α containing a clone of pTrc99a-Lac I was induced with 2.5 mM IPTG. After 2 hours, the cells were harvested and subsequently lysed by sonication. *Lane 1* contained a high molecular weight protein marker. *Lane 2* contained 5 µl of uninduced *E. coli* grown to an A_{260nm} of 0.5, immediately prior to IPTG induction. *Lane 3* contained 5 µl of supernatant from an IPTG induced culture following centrifugation of the sonicated bacterial suspension. *Lane 4* contained 5 µl of resuspended insoluble particulate matter following centrifugation of the sonicated suspension. Samples were electrophoresed at 40 mA through a 12% SDS polyacrylamide gel.

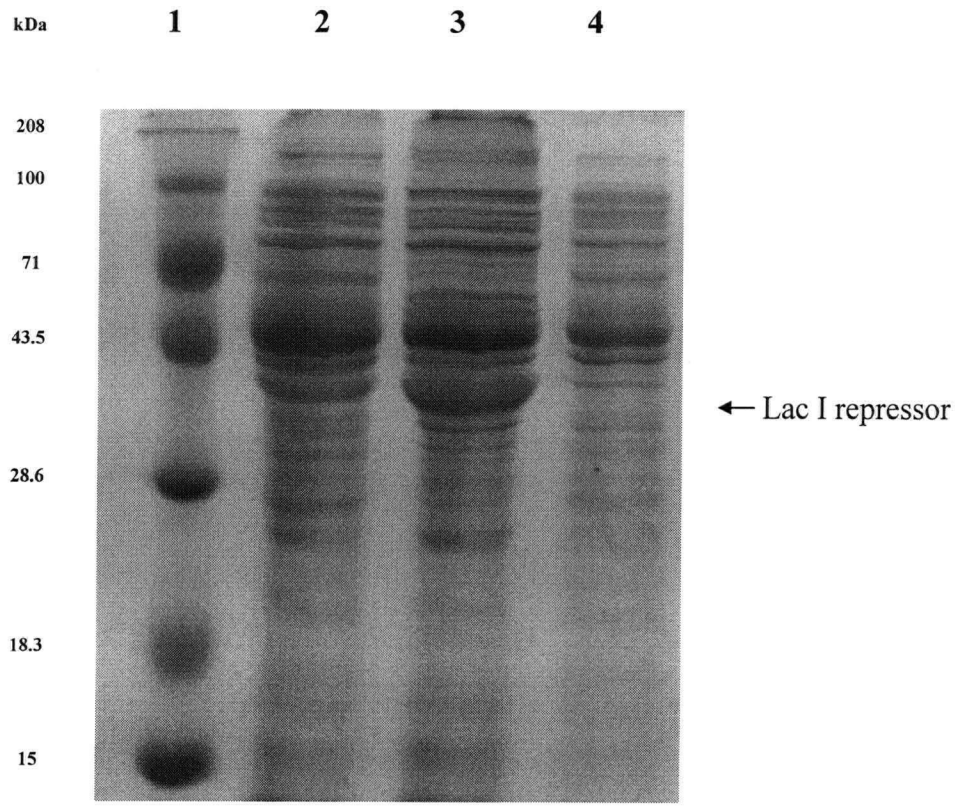

FIG. 15. **Ammonium sulfate percipitation of Lac I repressor.** To the resulting sonicated supernatant, ammonium sulfate was added to 37.5% (231 mg/ml) which in turn precipitated the Lac I repressor. *Lane 1* contained a high molecular weight protein marker. *Lane 2* contained 5 μl of supernatant following centrifugation of the sonicated bacterial suspension. *Lane 3* contained 5 μl of the resuspended ammonium sulfate precipitated pellet. *Lane 4* contained 5 μl of the supernatant from the ammonium sulfate percipitation following centrifugation. The samples were electrophoresed at 40 mA through a 10% polyacrylamide gel.

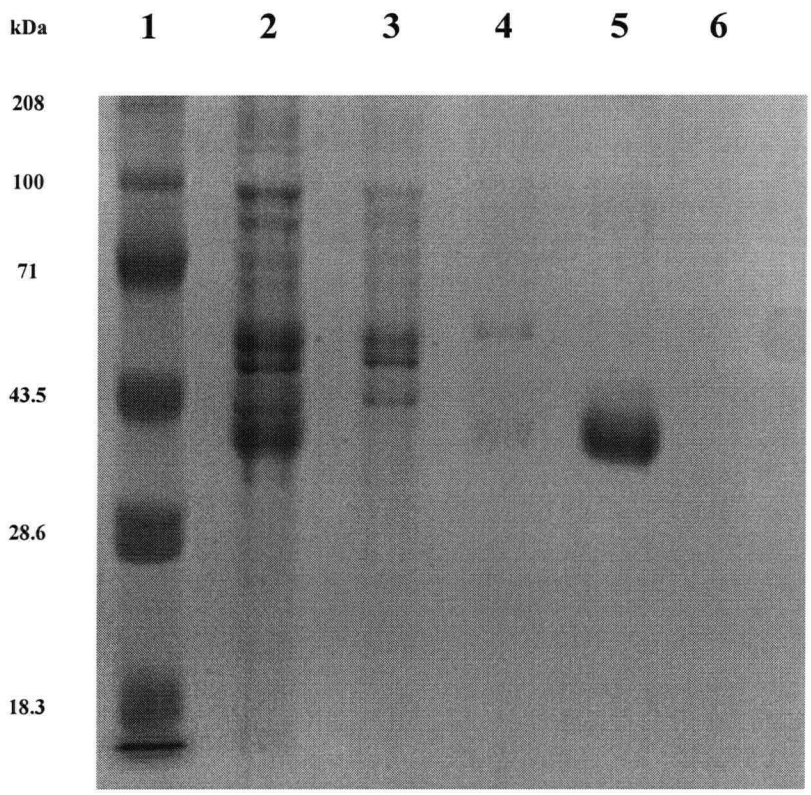

FIG. 16. **Purification of Lac I repressor by column chromatography through a Bio-Rad High S column.** The resuspended ammonium sulfate precipitate was dialysed against 0.03 M potassium phosphate buffer and subsequently loaded onto a High S column. *Lane 1* contained a high molecular weight protein ladder. *Lane 2* contained 5 µl of the resuspended and dialysed ammonium sulfate precipitate. *Lane 3* contained 5 µl of the flow through from the High S column washed with 0.03 M potassium phosphate buffer. *Lane 4* contained 5 µl of eluant from the column washed with 0.08 M potassium phosphate. *Lane 5* contained 5 µl of eluant from the column washed with 0.20 M potassium phosphate, the Lac I repressor fraction. *Lane 6* contained 5 µl of eluant from the column washed with 0.50 M potassium phosphate. The samples were electrophoresed at 40 mA through a 10% SDS polyacrylamide gel and stained with Coomasie Blue.

F. ASSAYING THE ACTIVITY OF THE PURIFIED Lac I REPRESSOR.

In order to ascertain if any activity was present within the purified Lac I repressor, a mobility shift assay was performed. The plasmid pBluescript was digested with the endonuclease *Taq I*, which among other DNA fragments, produced three which were approximately 500 bp in length. The 512 bp fragment contained the lactose operator. The fragments were purified and end labelled with ^{32}P-dCTP, and were mixed with the Lac I repressor. Following electrophoresis through a non-denaturing polyacrylamide gel, specific binding was observed to the fragment containing the lactose operator by autoradiographic imaging (Fig. 17). In both the three day and 5 hour exposure, the gradual disappearance of the 512 bp fragment was visible as greater amounts of the Lac I repressor protein was added. The 3 day exposure indicates that all of the 512 bp fragment was bound by 20 ng of Lac I repressor, and elevated to a band shift position. The 3 day exposure also showed a higher super shift position. This band was explained by the homotetrameric nature of the Lac I repressor. A dimer of the Lac I repressor is able to effectively bind a single Lactose operator due to the operon's dyad symetry, each monomer binds one half of the operon. The tetramer is able to bind two separate operons, as would be seen within an uninduced *E. coli* bacteria (Lewis *et al.* (1996)). Thus the tetrameric form of the Lac I repressor binds two 512 bp fragments and accounts for the elevated super shift. Higher concentrations of Lac I repressor caused aggregation of both the protein and DNA which remained near the fringe of the loading gel.

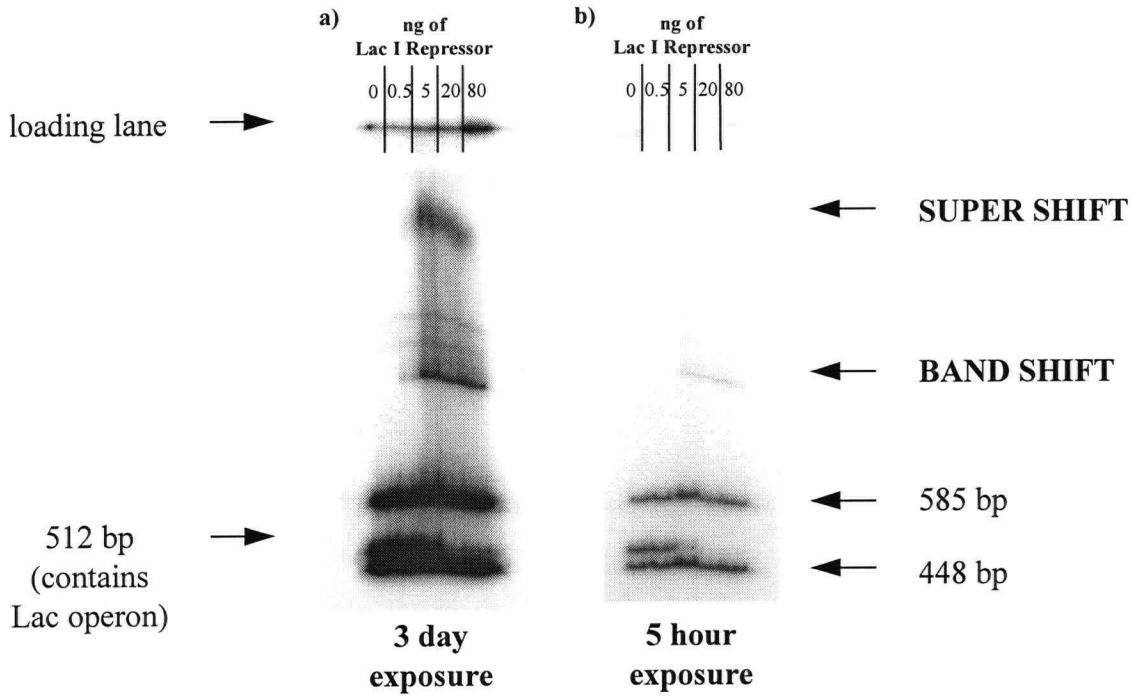

FIG. 17. **Mobility shift assay for the Lac I repressor and lactose operator.** Lac I repressor was mixed with three radio-labelled DNA fragements, one of which contained a lactose operator. Following non-denaturing gel electrophoreses, the fragments were visualized by audioradiography. Autoradiogram a) was developed following a three day exposure. Gradual disappearance of the 512 bp fragment containing the lactose operator was observed as the amount of Lac I repressor was increased. As the disappearance of the 512 bp fragment occured, the appearance of a band shift and super shift developed. The super shift was a product of the Lac I repressor tetramer binding two DNA molecules. Aggregation of protein/DNA complexes occured at high levels of Lac I repressor and were seen as a strong signal near the fringe of the loading lane. Autoradiogram b) was a 5 hour exposure and more clearly indicates the disappearance of the 512 bp fragment and appearance of the band shift.

G. NITROCELLULOSE CAPTURE ASSAY USING Lac I REPRESSOR AND pNRneo.

The Lac I repressor protein is able to bind with high affinity to the lactose operator present within the *E. coli* bacterial genome (Platt *et al.* (1973)); this lactose operator is also present in many plasmids including pNRdhfr and pNRneo. This was the basis for developing an assay to purify the pNR clonal recombinant from within the BHK genome. This would increase the likelyhood of obtaining a rare integrant from a population of mostly randomly recombined clones within the BHK genome.

Previous papers have described methods to rescue retroviral integrants from mammalian genomes using the Lac I repressor. Examples employed antibodies directed towards the Lac I repressor protein or a β-galactosidase/Lac I repressor fusion protein to construct affinity columns or magnetic bead precipitation (Gossen *et al.* (1993)). These options were considered too costly.

A nitrocellulose membrane effectively binds protein while DNA is able to pass through (Ausubel *et al.* (1996)). Therefore, if purified Lac I repressor protein was bound to DNA containing a lactose operator, that DNA molecule would remain bound to a nitrocellulose membrane by virtue of its association with the Lac I repressor. Other DNA fragments which did not contain a lactose operator would pass through the nitrocellulose membrane, resulting in the rescue of the clonal recombinant from the overwhelming mass of genomic DNA. Upon addition of IPTG to the buffer, the nitrocellulose bound Lac I repressor would release the lactose operator containing DNA fragment which would subsequently be collected and concentrated.

Figure 18 demonstrates the ability of the nitrocellulose capture assay to purify a DNA fragment containing a lactose operator away from genomic DNA digested with the endonuclease *EcoRI*. *Lanes 2* and *3* resulted from a binding reaction containing 0.02 μg of Lac I repressor, 1 μg of *NdeI* digested pNRdhfr, and 2.5 μg of BHK genomic DNA digested with *EcoRI*. *Lane 2* was the first 500 μl wash which was collected and ethanol precipitated, while *lane 3* was the fourth 500 μl wash which contained an additional 100 mM of IPTG. Clearly 0.02 μg binds little of the lactose operator within pNRdhfr which eluted with the first wash. Hardly any pNRdhfr remained with the Lac I repressor bound to the nitrocellulose membrane which would have been released upon addition of IPTG. Levels of Lac I repressor increased up to 20 μg for *lanes 8* and *9*. Once again *lane 8* contained the first wash while *lane 9* contained the fourth wash which contained an additional 100 mM IPTG. 20 μg of Lac I repressor was sufficient to bind 1 μg of pNRdhfr as none was visible in the first wash. Approximately 75% of the initial 1 μg of pNRdhfr was eluted in the fourth wash which contained 100 mM IPTG as judged by the 5 kb band present within *lane 9*. As more Lac I repressor was added to the binding reaction, the digested genomic DNA gradually disappeared from the first wash. This was most likely due to the basic N-terminal domain of Lac I repressor non-specifically binding the genomic DNA. Further washes stripped away the non-specific interaction such that no genomic DNA was visible in the fourth wash containing IPTG, meant to elute the pNRdhfr molecule bound to the lactose operator within the pNRdhfr molecule.

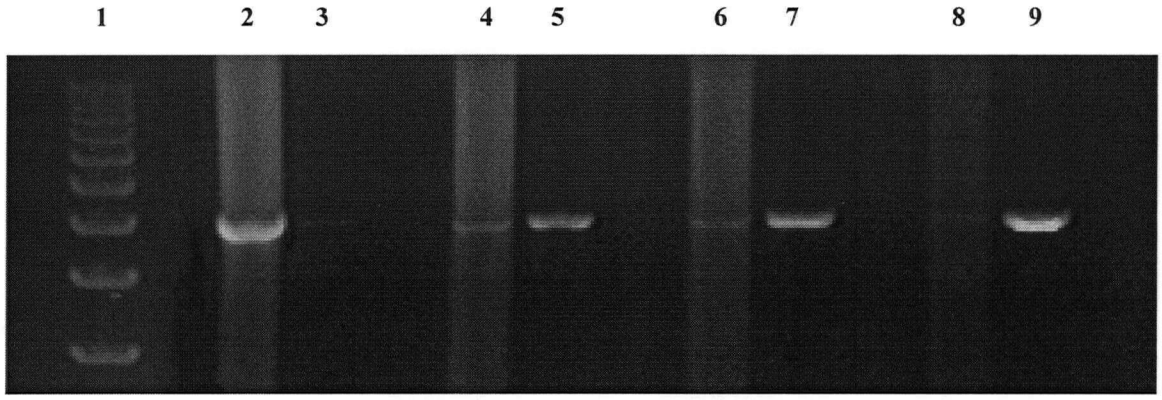

FIG. 18. **Nitrocellulose capture assay using Lac I repressor and pNRdhfr.** The assay was performed in order to determine if the Lac I repressor would purify pNRdhfr away from digested genomic DNA. *Lane 1* contains a *HindIII* digested lambda DNA ladder. *Lanes 2* and *3* arose from a binding reaction containing 0.02 μg of Lac I repressor, 1 μg of *NdeI* pNRdhfr, and 2.5 μg of *EcoRI* digested BHK genomic DNA. *Lane 2* contained the ethanol precipitated 500 μl first wash while *lane 3* contained the ethanol precipitated 500 μl IPTG eluant. *Lanes 4* and *5* arose from a binding reaction containing 0. 2 μg of Lac I repressor, 1 μg of *NdeI* pNRdhfr, and 2.5 μg of *EcoRI* digested BHK genomic DNA. *Lane 4* contained the ethanol precipitated 500 μl first wash while *lane 5* contained the ethanol precipitated 500 μl IPTG eluant. *Lanes 6* and *7* arose from a binding reaction containing 2 μg of Lac I repressor, 1 μg of *NdeI* pNRdhfr, and 2.5 μg of *EcoRI* digested BHK genomic DNA. *Lane 6* contained the ethanol precipitated 500 μl first wash while *lane 7* contained the ethanol precipitated 500 μl IPTG eluant. *Lanes 8* and *9* arose from a binding reaction containing 20 μg of Lac I repressor, 1 μg of *NdeI* pNRdhfr, and 2.5 μg of *EcoRI* digested BHK genomic DNA. *Lane 8* contained the ethanol precipitated 500 μl first wash while *lane 9* contained the ethanol precipitated 500 μl IPTG eluant.

H. ANALYSIS OF SEQUENCE INFORMATION FROM RESCUED pNR RECOMBINANTS.

Development of the nitrocellulose capture assay enabled small restriction endonuclease generated fragments of the pNR DNA molecule to be purified from the overwhelming majority of genomic DNA. This assay was ideal when attempting to analyze the site of clonal recombination of the pNR molecules with the BHK genome.

The lactose operator sequence was present within 100 bp of the 3' terminal end of the pNR molecule when digested with the endonuclease *NdeI.* Endonucleases with a four base pair recognition sequence digest a random piece of DNA on average every 256 bp. There are many such endonucleases, such as *HhaI,* which digest just upstream of the lactose operator within *NdeI* linearized pNR, but not 3' of that site. Thus *HhaI* would digest within an area of genomic DNA immediately 3' of the clonal recombination site (Fig. 19). Such an endonuclease digestion would produce a DNA fragment which contains the terminal end of pNR, whose sequence included the lactose operator, as well as genomic sequence from the site of clonal recombination.

Purified lactose operator containing DNA fragments were required for sequence analysis. In order to accomplish this, the fragments were first circularized, having been diluted to 0.5 μg/ml and their cohesive *HhaI* digested ends ligated with T4 DNA ligase. The closed circular DNA fragments were then amplified by inverse PCR (Nolta *et al.* (1996)). Inverse PCR used primer pairs complementary to the pNR backbone near the

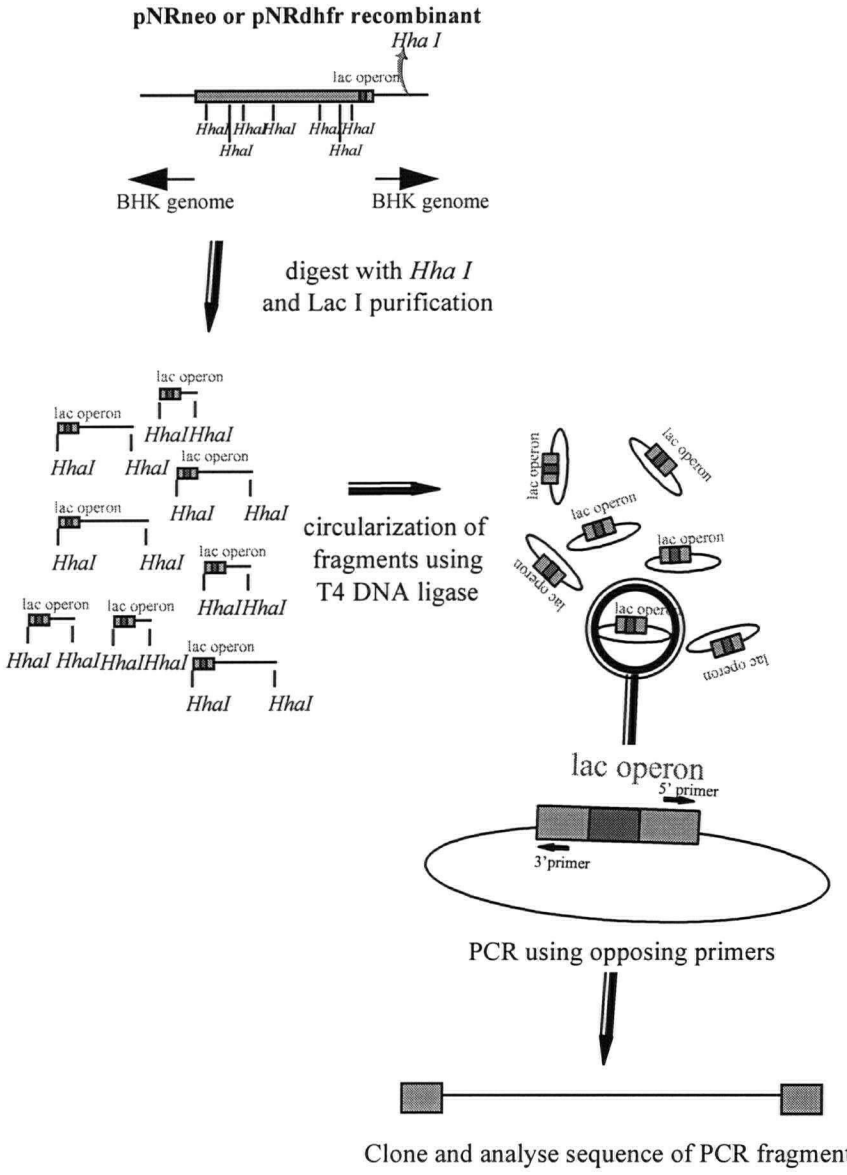

FIG. 19. **Strategy to assess pNR clonal recombination sites.** Genomic DNA was digested with the endonuclease *HhaI*, and the resulting fragments of various sizes containing a lactose operator were purified from BHK genomic DNA. The purified fragments were self ligated and inverse PCR was performed to generate fragments containing genomic DNA from the site of clonal recombination, flanked by pNR segments. The linear DNA fragments were then cloned and their sequence analyzed.

lactose operator: pBluescript reverse oligonucleotide primer and Anti R oligonucleotide primer. These primers annealed in opposite directions around the closed circle to produce a linear PCR product from a circular template. Further rounds of PCR would proceed from the exponentially growing population of linear fragments. Nested PCR using pBluescript T3 oligonucleotide primer and Anti R oligonucleotide primer, was performed to further amplify the products from clonal recombination sites.

Amplified fragments were then cloned and their sequence was analyzed. PCR fragments were first blunt ended by the addition of Klenow fragment, and after a phenol extraction and DNA precipitation, digested by the endonuclease *BamHI*. The resulting directional fragments were ligated into a pBluescript vector. Positive clones of transformed *E.*coli DH5α were selected by PCR of colonies which served as a template for amplification with an internal fragment oligonucleotide Anti R and a pBluescript 'reverse' primer. The DNA sequences of 80 clones were analyzed by ^{35}S-dideoxy nucleotide sequencing. The overwhelming majority of the clones analyzed indicated that the pNR molecule, although transfected in a linear form, had re-ligated, presumably within the nucleus of the host cell. Inverse PCR would amplify these products to the exclusion of all others. The re-ligated form of rescued clonal recombinants was removed from the inverse PCR template by endonuclease digestion with *NdeI*. Such a digestion would linearize any of the closed circles which contained an *NdeI* site; principally those clones resulted from religation of pNR post-transfection.

Having removed the re-ligated template, there were many examples of pNR molecules which had undergone intra-molecular recombination including recombination of pNR molecules

with other pNR molecules before or after recombination with the BHK genome and examples of pNR which had recombined within the BHK chromatin (Figure 20). Of those clones analyzed, some yielded presumably genomic sequence but did not show homology to any other gene within the BLAST database (National Center for Biotechnology Information, http://www.ncbi.nlm.gov/). Others clones were the result of recombination with specific genes within the BHK genome as determined, with the BLAST search. Two of these recombination sites were encountered more than once, even though the clones resulted from separate transfections (Fig. 20). These hotspots were homologous to the rat histamine receptor gene (RATH1SIR, average 85% homology) and homologous to the murine mutant tetrahydrofolate reductase gene (MMDHF4, average 99% homology).

An estimate as to the enrichment factor of the nitrocellulose capture assay towards lactose operator containing DNA fragments was 1×10^7 times (BHK genome 3×10^{12} bp / 10 integrants at 5000 bp each = 7×10^7). This enrichment factor made it possible to rescue recombinants from a *NdeI* digested template prior to inverse PCR. Previously, inverse PCR of a template treated with *NdeI*, but still principally composed of genomic DNA, would have yielded no amplification products, even after a second nested PCR was performed.

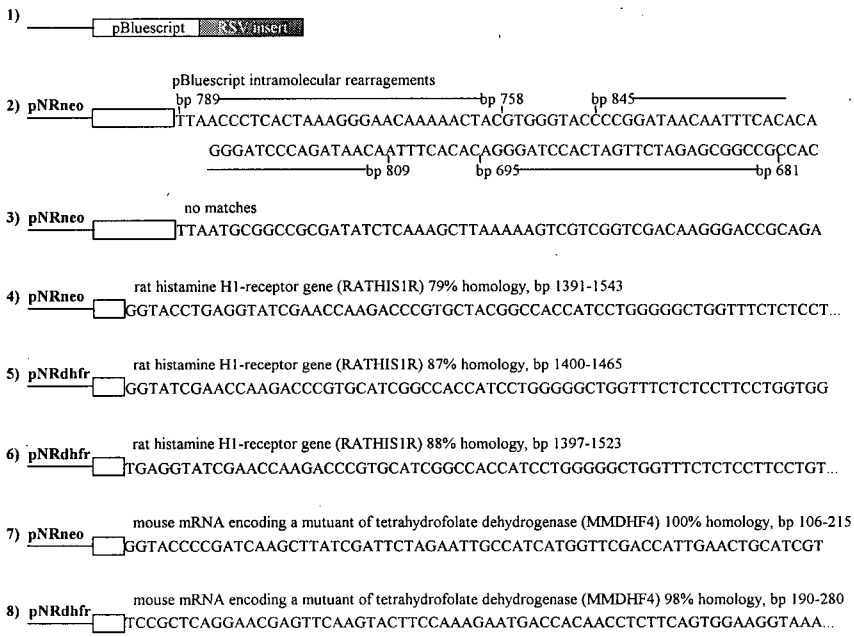

FIG. 20. **Location of recombination sites of pNR clones rescued from the BHK cell genomes.** Unknown DNA sequences from pNR clonal recombinants were entered into a BLAST search to define regions of homology within the database. *Diagram 1)* represents the expected terminal region would RSV IN have mediated integration of the substrate plasmid. Instead, only examples of random recombination or religation of the substrate plasmid were observed. *Diagram 2)* represents one of many examples of intramolecular recombination which resulted from many recombination events within the pNR plasmid. *Diagram 3)* was one example of an analyzed sequence which displayed no sequence homology. Three clones from separate transfections recombined within the same region of the histamine receptor gene of BHK chromatin (Genbank RATHIS1R); *diagram 4)* bp 1391-1456, *diagram 5)* bp 1400-1465, and *diagram 6)* bp 1397-1462 . Two clones

FIG. 20 continued

recombined within the mutant tetrahydrofolate reductase gene of BHK chromatin (Genbank MMDHF4); *diagram 7)* bp 106-171, and *diagram 8)* bp 190-255. Base pairs (bp) denote position of clonal recombinant within the specified gene.

I. CHARACTERISTICS OF A STABLE BHK CELL LINE EXPRESSING ROUS SARCOMA VIRUS INTEGRASE.

To further study the effect of RSV IN on substrate DNA *in vivo*, a stable BHK cell line which constituitively expressed RSV IN was constructed. Subsequent transfection of substrate DNA (*NdeI* linearized pNRneo) were carried out to study the effect of high levels of pre-existing intra-cellular RSV IN towards the substrate DNA and subsequently the cellular chromatin.

The stable BHK-IN expressing cell line was constructed by transfecting normal BHK cells with pNUT-IN. The vector pNUT (Palmiter *et al.* (1987)) contained a zinc activated metallothionein promoter which remained constituatively active due to the presence of zinc within the culture media, thus ensuring constant, high level expression of the inserted *in* gene. The vector pNUT also contained a dihydrofolate reductase cassette, allowing cells containing the vector to grow in media containing methotrexate. As cells containing the vector grew, only those with a stable integrated copy would survive to form colonies as extra-chromosomal DNA would be lost upon successive divisions.

Twenty-four colonies were picked and expanded, and the level of integrase expression assessed by *in situ* cell staining with a polyclonal antibody directed towards RSV IN. Although the antibody exhibited some cross-reactivity, staining was sufficient to divide the expanded clones into three separate categories (Figure 20): those clones

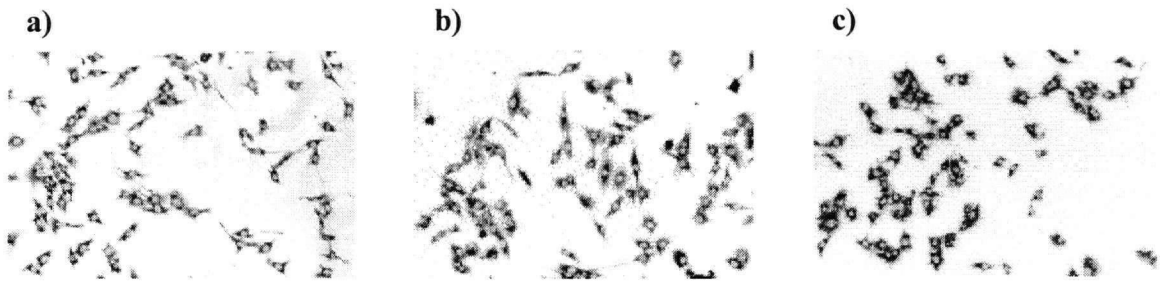

FIG. 21. ***In situ* staining of BHK-IN clones with polyclonal antibody towards RSV IN.** BHK cells were transfected with the vector pNUT-IN, which having recombined with BHK chromatin, stably expressed RSV IN. After selection of recombinants with 500 μM methotrexate, clones were divided into three categories after *in situ* cell staining with a polyclonal antibody directed against RSV IN. *a)* depicts BHK-IN+ cells which expressed little or no IN. *b)* depicts BHK-IN++ cells which expressed moderate levels of IN. *c)* depicts BHK-IN+++ cells which expressed high levels of IN

expressing little or no RSV IN (BHK-IN+, *plate a*), clones expressing moderate levels of IN (BHK-IN++, *plate b*), and clones expressing high levels of integrase (BHK-IN+++, *plate c*). While BHK-IN+ cells appeared essentially normal in terns of morphology and growth rates, BHK-IN+++ cells were more rounded in appearance than a healthy 'fibroblast' appearance of BHK cells. The rounded appearance of BHK cells occurs just prior to cell death, although the BHK-IN+++ cells continued to divide and were passaged numerous times. Their dividing time was also slower than untransfected BHK cells, around 24 hours. BHK-IN++ cells, had a similar 'fibroblast' appearance to untransfected BHK cells, although their dividing time was slightly greater.

All three divisions of IN expressing cell lines were transfected with substrate DNA (*NdeI* linearized pNRneo). BHK-IN+ cells responded in a similar fashion to normal BHK with respect to substrate transfection (Figure 22). BHK-IN+++ cells responded very dissimilarily to normal BHK cells towards transfection of substrate DNA. Those BHK-IN+++ cells transfected with 0.34 μg/well of substrate DNA had a mortality rate of 90% by day 3, before the neomycin selection had begun to occur. BHK-IN+++ cells transfected with 0.17 μg/well of substrate DNA had a 80% mortality by day 3 (Figure 22, *line d*), and those transfected with uncut pNRneo or none at all exhibited a normal reaction to G418 selection (Figure 22, *line a*) by day 3. BHK-IN++ cells responded to transfection as an average between BHK-IN+ and BHK-IN+++. BHK-IN++ cells transfected with 0.34 μg of substrate DNA had a mortality rate of 50% by day 3, those transfected with 0.17 μg of substrate DNA had a 30% mortality by day 3 (Figure 22, *line c*), and those transfected with uncut pNRneo or none at all exhibited a normal reaction to

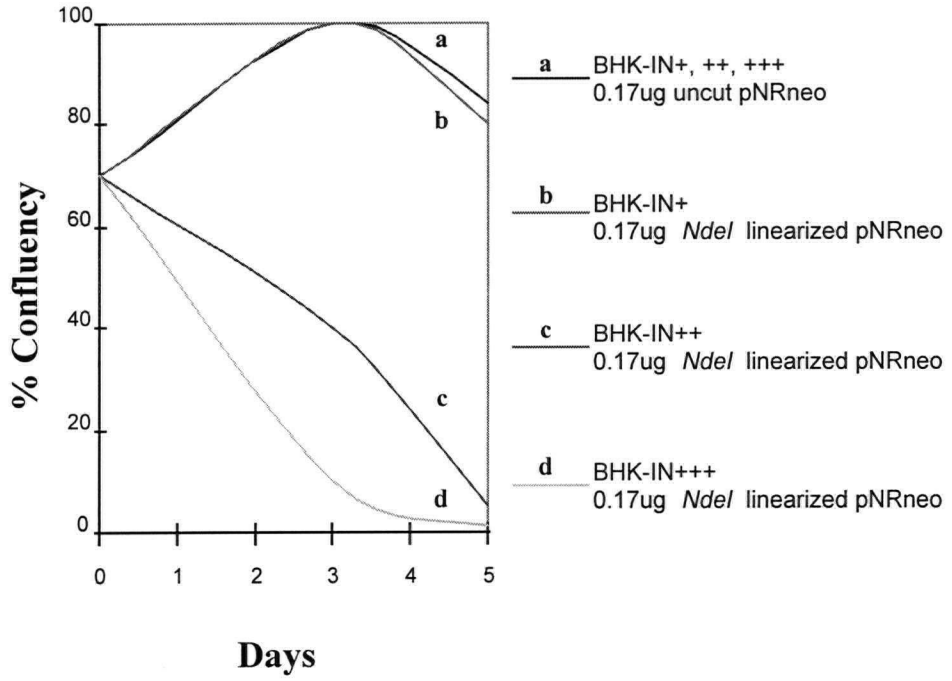

FIG. 22. **Mortality curve of BHK-IN clones transfected with substrate DNA.** Both uncut pNRneo and substrate DNA (*NdeI* linearized pNRneo) were transfected into three BHK-IN cell lines. All three cell lines; BHK-IN+, BHK-IN++, and BHK-IN+++ displayed the same mortality curve when transfected with 1 μg of uncut pNRneo under G418 selection. BHK-IN+, a BHK cell line which expresses little or no IN, responded similarily to uncut transfections when transfected by 1 μg of substrate DNA (*NdeI* linearized pNRneo). BHK-IN++, a BHK cell line expressing moderate levels of IN displayed a 30% mortality rate by day 3, before G418 selection had begun to induce cell death. BHK-IN+++, a BHK cell line expressing high levels of IN displayed a 80% mortality by day 3. G418 induced cell death typically began on day 4 for normal BHK cells after which the majority of cells would have died by day 5.

G418 selection by day 3 (Figure 22, *line a*). BHK-IN+ responded similarily to uncut transfections when transfected with substrate DNA (Figure 22, *line b*).

Previous co-transfections with pIN and substrate DNA resulted in more numerous colony formation, and thus greater transfection efficiency, than with substrate DNA alone. High level IN expressing BHK cell lines exhibited an increased mortality rate when transfected with substrate DNA. Corresponding to this finding, levels of colony formation also dropped off as stable cells lines expressed higher amounts of RSV IN (Figure 23). BHK-IN+ cells transfected with 0.17 μg/well of uncut pNRneo produced 157.0 ± 11.3 colonies after selection with G418. Transfections with 0.17 μg/well of substrate DNA, pNRneo previously digested with *NdeI*, produced almost the identical amount of colonies, 149.3 ± 8.4 colonies. Increasing the amount of DNA transfected for BHK-IN+ cells to 0.34 μg/well resulted in 216.8 ± 15.0 colonies. Higher levels of IN which were observed in BHK-IN++ cells resulted in fewer colonies as the levels of substrate DNA increased. When 0.17 μg of uncut pNRneo was transfected, 172.5 ± 7.8 colonies were produced, but digestion of pNRneo with *NdeI* prior to transfection resulted in 129.8 ± 12.4 colonies with 0.17 μg/well substrate DNA. Increasing the amount of substrate DNA transfected to 0.34 μg/well of substrate DNA resulted in 128.5 ± 8.3 colonies per well. Yet higher levels of IN expression found in BHK-IN+++ resulted in ever fewer colonies formed when transfected with substrate DNA. Transfections with 0.17 μg/well uncut pNRneo resulted in 144.5 ± 10.6 colonies, but when pNRneo was digested with *NdeI*, 0.17 μg/well of substrate DNA produced 85.8 ± 4.7 colonies, a decrease by nearly half. Increasing the

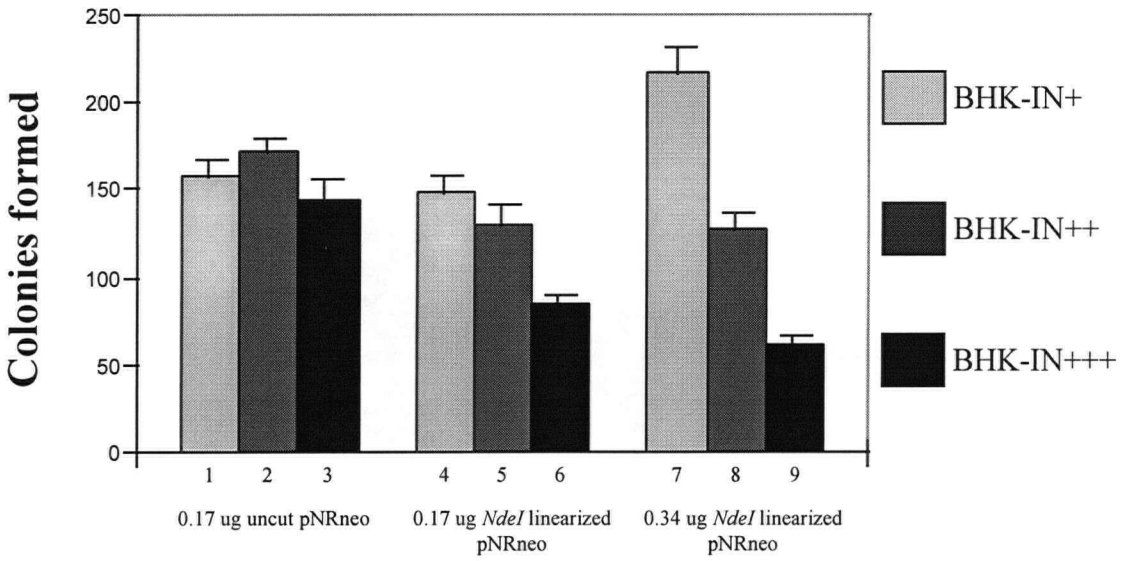

FIG. 23. **Colonies formation of BHK-IN clones transfected with substrate DNA.** For all three BHK-IN cell lines, transfections with 0.17 μg/well of uncut pNRneo resulted in similar colony formation; BHK-IN+ 157.0 ± 11.3, BHK-IN++ 172.5 ± 7.8, BHK-IN+++ 144.5 ± 10.6. BHK-IN+ cells transfected with 0.17 μg/well substrate DNA formed 149.3 ± 8.4 colonies, similar levels to uncut transfections. BHK-IN++ cells transfected with 0.17 μg/well substrate DNA formed 129.8 ± 12.4, significantly less colonies than the same amount of uncut pNRneo transfected into BHK-IN++. BHK-IN+++ cells transfected with 0.17 μg/well of substrate DNA formed 85.8 ± 4.7, fewer colonies still than transfection of uncut pNRneo into BHK-IN+++. This trend continued as greater amounts of substrate DNA were transfected. Transfections with 0.34 μg/well increased the number of colonies formed for BHK-IN+ cells, up to 216.8 ± 15.0, while decreasing the colonies formed in BHK-IN++ (128.5 ± 8.3) and BHK-IN+++ (60.3 ± 6.3).

amount of substrate DNA transfected to 0.34 μg/well further decreased colony formation to 60.3 ± 6.3.

J. SEQUENCE ANALYSIS AND SOUTHERN BLOTTING OF GENOMIC DNA FROM STABLE BHK CELL LINES EXPRESSING ROUS SARCOMA VIRUS INTEGRASE.

Having isolated genomic DNA from the three stable cell lines expressing RSV IN, sequence analysis was carried out to determine if in fact the characteristics of clonal recombination found previously upon co-transfection with pIN and substrate DNA was shared by IN expressing cell lines transfected with substrate DNA. Using the methods previously described, the nitrocellulose capture assay allowed purification of the terminal ends of clonal recombinants. After sequence analysis of the rescued clones, no full length substrate molecules were recorded. A full length substrate molecule, pNRneo digested by the endonuclease *NdeI*, displays terminal ends whose final 30 bp are identical to the terminal ends of RSV viral DNA immediately before integration to form a provirus. If substrate DNA was integrated by RSV IN within the BHK chromatin, rescued clones would have contained the terminal ends containing a CA dinucleotide, conserved in every example of retroviral mediated proviral integration (Grandgenett *et al.* (1993)). As before, examples of intra-molecular recombination occurred: recombination of pNRneo molecules with each other either before or after recombination with the BHK DNA and also random recombination near the terminal ends of the substrate DNA with the BHK

chromatin. No repeated recombination with BHK 'hotspots' was observed within the BHK-IN expressing cell lines, as were seen in BHK chromatin co-transfected with pIN and substrate DNA, although only 20 clones were analyzed.

Further analysis of the genomic DNA from BHK-IN expressing cell lines transfected with substrate DNA was carried out by a Southern blot (Figure 24). Genomic DNA from the cell lines BHK-IN+, BHK-IN++ and BHK+++ were digested by the endonuclease *KpnI*. *KpnI* would release an intact substrate molecule due to sites within 30 bp of the terminal ends of an *NdeI* linearized substrate molecule. The strongest signal recorded, just above 5 kb (control *lane 1*, 1 ng *NdeI* linearized pNRneo at 5 kb) were seen in *lanes 4, 5,* and *6*. These lanes contained 10 μg of *KpnI* digested genomic DNA from the three BHK-IN cell lines. The strongest signal results from the integrated pNUT-IN vector. With two *KpnI* sites in pNUT-IN, at the 5' end of the metallo-promoter and 800 bp within the *in* gene, a large 5.5 kb DNA fragment results after digestion, accounting for the main signal. As a previous Southern blot (Figure 24) indicated a strong signal was at ~5 kb (Fig. 24, *lanes 3, 4,* and *6*), indicating that the majority of the substrate DNA had recombined with the BHK-IN genome intact.

Comparing Figure 12, a Southern blot of genomic DNA from BHK cells co-transfected with pIN and substrate DNA (*NdeI* linearized pNRdhfr), with Figure 24, a Southern blot of genomic DNA from BHK cell lines expressing IN and transfected with substrate DNA (*NdeI* linearized pNRneo) it is obvious that pNRdhfr has many more recombinants than pNRneo. Both *lane 1* of Figure 12 and Figure 24 contained 1 ng of *NdeI* linearized pNRdhfr or pNRneo, while the experimental lanes from both figures were comprised of 10 μg of electrophoresed genomic

DNA. Those lanes resulting from pNRneo transfected clones had far fewer recombinants (Figure 24) than lanes resulting from pNRdhfr transfected clones (Figure 19). The background smear which developed from electrophoresed DNA of *KpnI* digested pNRdhfr transfections (Figure 19) was quite intense, showing many bands in addition to the main band at ~5 kb. pNRneo transfections also featured the ~5 kb band , but the background smear was almost non-existent (Figure 24).

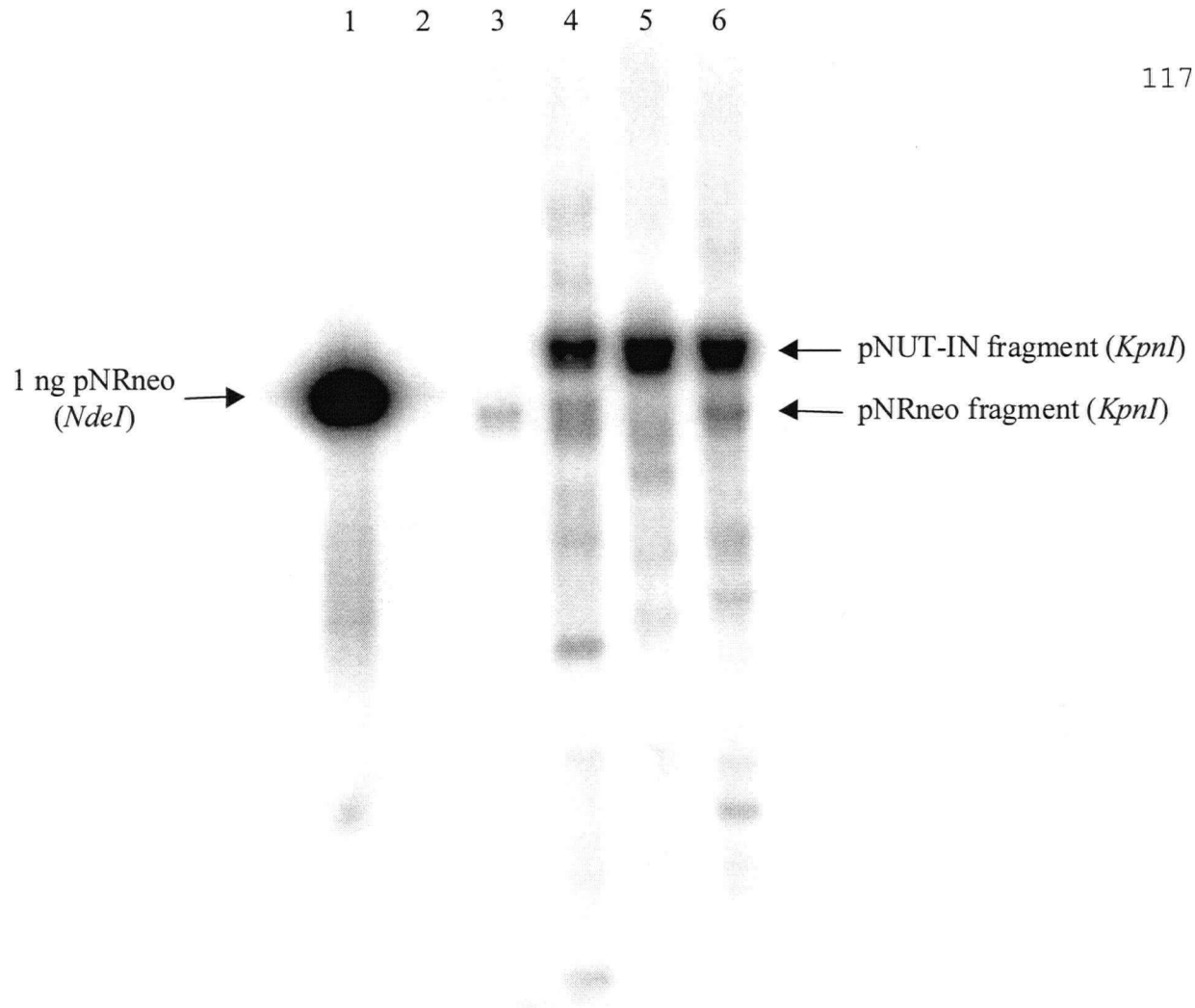

FIG. 24. **Southern blot analysis of pNRneo recombinants from BHK-IN cell lines.** Genomic DNA from BHK and BHK-IN cell lines were blotted and probed with a ^{32}P labeled SV40 promotor probe. *Lane 1* contained 1 ng of substrate DNA (*NdeI* linearized pNRneo). *Lane 2* contained 10 μg of untransfected BHK genomic DNA, while *lane 3* contained 10 μg of genomic DNA from BHK cells previously transfected with substrate DNA and selected under 500 μM G418.. *Lanes 4, 5,* and *6* also contained 10 μg of genomic DNA isolated from BHK-IN+, BHK-IN++, and BHK-IN+++ cells previously transfected with substrate DNA and selected under 500 μM G418.

K. ANALYSIS OF NUCLEAR EXTRACTS FROM POOLED CLONES OF THE CELL LINE BHK-IN.

Evidence from sequence analysis and Southern blots indicated that a re-circularized substrate DNA molecule was recombining with BHK or BHK-IN chromatin. In order to gain insight to this phenomenon, nuclear extracts from pooled clones of BHK-IN cell lines were gathered. The extracts were combined with substrate DNA and following a 30 minute incubation at 37° C, were electrophoresed and Southern blotted. Figure 25 indicated that specific and non-specific endonucleases digested the substrate molecule, at higher concentrations of nuclear extract, the substrate begins to disappear with no smaller fragments remaining (Figure 24, *lane E3*). There was no discernible difference between nuclear extracts from normal BHK cells (Figure 25, *lanes* A) and nuclear extracts from BHK-IN pooled clones. It was expected that a ligation product would appear above the main substrate band, as would be expected for non-supercoiled, closed, circular DNA molecules. If the nuclear extracts promoted ligation of the substrate, the presence of endonucleases quickly removed any trace.

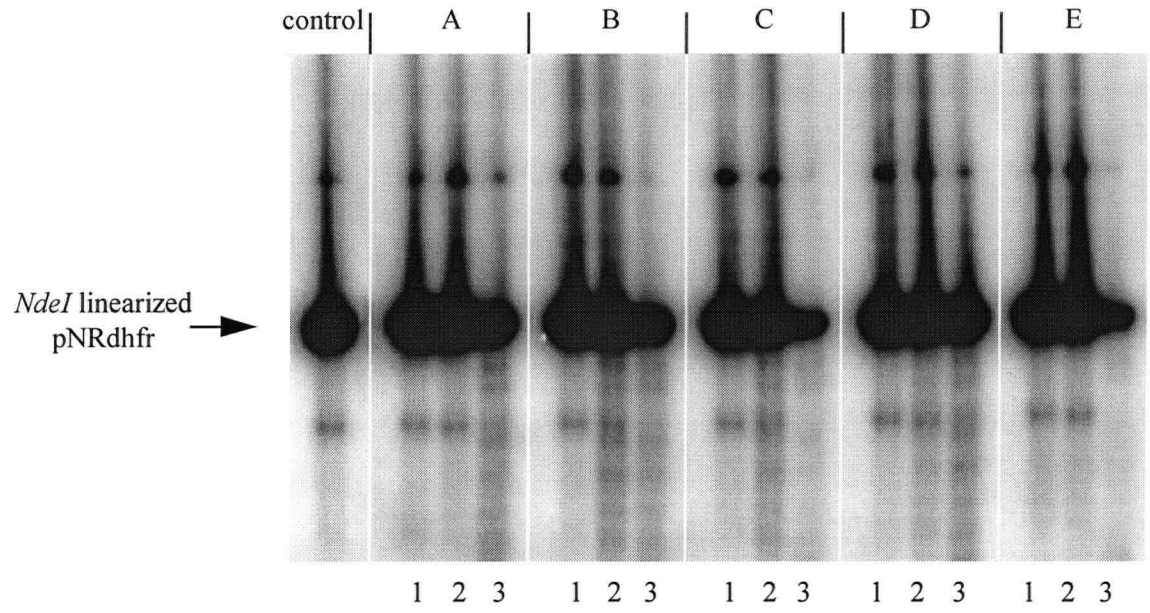

FIG. 25. **Effect of nuclear extracts of BHK and BHK-IN cell lines on substrate DNA.** Nuclear extracts were isolated to study their effect on substrate DNA *in vitro.* All lanes contained 0.1 μg of electrophoresed *NdeI* linearized pNRdhfr, and *control* contained no nuclear extract. L*ane 2*, 1 μl of 100× dilution of nuclear extract; *lane 2*, 1 μl of 10× dilution of nuclear extract; *lane 3*, 1 μl of nuclear extract. Group A used nuclear extract isolated from normal BHK cells, Group B used nuclear extract isolated from BHK-IN-1 pooled clones, Group C used nuclear extract isolated from BHK-IN-2 pooled clones, Group D used nuclear extract isolated from BHK-IN-3 pooled clones, Group E used nuclear extract isolated from BHK-IN-4 pooled clones. Reaction volume was 10 μl.

DISCUSSION:

A. EFFECT OF CO-TRANSFECTION OF BHK CELLS WITH pIN AND pNR.

This study explored the efficacy of using recombinant retroviral integrase expressed transiently *in vivo* to mediate the incorporation of an exogenous substrate DNA molecule into host cell chromatin. By co-transfecting BHK cells with an integrase (IN) expression plasmid and an engineered DNA substrate, I have shown that linear double-stranded DNA possessing truncated RSV LTRs at both termini effectively mimics the RSV provirus such that intracellular IN is able to bind the DNA molecule and increase the transfection efficiency ten fold.

The DNA substrate was constructed by ligating an NDE-RSV insert into a pBluescript backbone as well as an expression cassette capable of conferring drug resistance to transfected cells. The NDE-RSV insert contained the nucleotide sequence which was encoded by the terminal ends of the Rous Sarcoma Virus proviral DNA prior to integration. These terminal sequences were joined to form an *NdeI* restriction site, which upon *NdeI* digestion would result in a linear DNA molecule, effectively mimicking the RSV provirus such that IN sequence recognition could effect a higher transfection efficiency.

A functional recombinant IN protein was encoded by the expression vector (pIN) employed in our protocol. The *in* gene was expressed in bacteria and the IN protein was purified. As has previously been shown for integrase (Grandgenett *et al.* (1993)), the purified protein catalyzed the integration of substrate DNA molecules *in vitro.* Integration was visualized on a

Southern blot which illustrated the appearance of 2 diffuse bands above the substrate DNA which migrated to approximately 3000 bp after electrophoreses. The higher signals appeared on the autoradiogram at approximately 6000 bp and at 9000 bp, indicating that substrate DNA had integrated with other substrate molecules to give the corresponding ladder. The diffuse nature of the bands was probably due to Y-shaped DNA molecules and a range of deletion which occurred as a result of the *in vitro* integration assay (Vincent *et al.* (1993)). Thus a functional RSV IN protein was encoded.

Co-transfection of pIN and substrate DNA resulted in a ten fold increase in colony formation over transfections of substrate DNA alone (Figure 7., *plate 6* vs. *plate 2*). Co-transfection of pIN and *NotI* linearized pNRdhfr resulted in colony formation equivalent to transfections of substrate DNA alone (Figure 7., *plate 4* vs. *plate 2*). This result highlights the importance of the substrate DNA being a DNA molecule derived from the RSV provirus DNA model. *NotI* digestion of pNRdhfr results in a linear dsDNA molecule; however the RSV-LTR elements are contained within the DNA molecule, not displayed at the termini of the linear molecule as with *NdeI* digestion (Figure 5). *In vivo*, IN was only effective in increasing transfection efficiencies when terminal LTR elements were displayed as part of a linear dsDNA molecule.

Co-transfection of substrate DNA with plasmid DNA of a similar mass to pIN but not encoding an *in* gene, resulted in background levels of methotrexate resistant colonies (Figure 7., *plate 3*). It can be concluded that pIN did not serve as 'carrier' DNA whose presence increased transfection efficiencies. When co-transfections were carried out with pIN and substrate DNA

contained within separate liposome mixtures, colony formation increased 2.5 fold over background levels (Figure 7, *plate 5*), but not the ten fold increase observed when co-transfection of pIN and substrate DNA were within the same liposome preparation (Figure 7, *plate 6*). It would seem likely that co-transfection of pIN and substrate DNA in a single liposome mixture would greatly increase the likelihood of any liposome containing both DNA molecules. Having fused with a target BHK cell, both DNA molecules would be released, whereas co-transfection with separate liposome preparations would require at least two fusion events to occur in order that both pIN and substrate DNA be delivered to the same cell. The observed 25% transfection efficiency observed with liposome preparations containing a plasmid encoding β-galactosidase (results not shown) only serves to emphasize this effect.

In addition to producing ten fold more colonies over transfections of substrate DNA, co-transfection of pIN and substrate DNA yielded larger colonies after methotrexate selection representing the products of more robust growth (Figure 8). It can be concluded that the improved response to methotrexate selection was due to increased di-hydrofolate reductase expression. Southern blots of *KpnI* digested genomic DNA isolated from both co-transfected clones and clones arising from transfections of substrate DNA alone revealed a relatively uniform amount of pNRdhfr or pNRneo recombinants for both types (Figure 12 and 24). The increased DHFR expression was likely due to an increased transient expression of DHFR, brought about by IN mediated concentration of substrate DNA within the transfected cell nucleus. If the concentration of substrate DNA within the nucleus of co-transfected cells was higher, an increased level of colony formation would by the same token be expected and was observed.

I sought to determine some of the factors affecting the transfection efficiencies observed with the co-transfection protocol. Varying the ratio of pIN to substrate DNA, without altering the total amount of DNA delivered to cells indicated that the intracellular concentration of substrate DNA is a key factor affecting the transfection efficiency. Figure 9a illustrates the colonies formed per transfected cell. *Plate 1*, co-transfected with a 7:1 ratio of pIN and substrate DNA produced 6 fold more colonies than transfections with a 1:7 ratio. Furthermore, it appeared that on a mole:mole basis, pIN need only represent a small fraction of the total number of DNA molecules delivered to the cells to mediate efficient transfection. Figure 9b displays the colonies per cell formed when substrate DNA was transfected alone. *Plates 1* - 9 were transfected with the same amount of substrate DNA as *plates 1* - 9 of Figure 9a, however the transfection lacked pIN. The difference in colony formation was 10 fold greater for Figure 9a than Figure 9b. Increasing the amount of DNA delivered while keeping the ratio of pIN to substrate DNA constant resulted in a linear increase in transfection efficiency as greater amounts of DNA were delivered to the cells. Figure 10 illustrates the increasing colony formation from *plate 1* through *plate 9*. *Plate 9* was transfected with double the amount of DNA as *plate 1* which resulted in an almost exact doubling of colonies formed. This trend corresponds to the direct relationship of increasing liposome and DNA concentrations towards increasing transfection efficiencies. However, at higher concentrations of liposomes, the amounts of DOPE:DODAC reached a critical point and cell death resulted from toxicity due to high liposome concentration (results not shown).

In addition to co-transfection of pIN and substrate DNA in the BHK cell line, two additional cell lines were co-transfected to determine if the increased transfection efficiencies

were also observed. The 293 cell line was derived from transformed human kidney cells, whereas the CV-1 cell line was derived from transformed African green monkey kidney cells. Figure 11 illustrates that the increase in transfection efficiencies observed in the BHK cell line was conserved for the 293 and CV-1 cell lines. Both the 293 and CV-1 cell lines did not produce as many colonies as the BHK cell line. This was principally due to the general decrease in transfection efficiencies observed with, in this instance, calcium phosphate transfection using a β-galactosidase reporter plasmid (results not shown).

The investigation of additional cell lines was the first instance that *NdeI* linearized pNRneo was used in co-transfections as substrate DNA. The construction of pNRneo was carried out for two reasons. The first being that both the 293 and CV-1 cell lines were resistant to methotrexate due to the presence of a dihydrofolate reductase gene within their genome. The BHK cell line contained a *dhfr* gene as well, but a nucleotide substitution caused a mutation which drastically altered the affinity of the expressed protein to methotrexate, thus making the cell line sensitive to methotrexate selection. The construction of pNRneo also confirmed that the pNR backbone could be altered by substitution of cassettes, and increased transfection efficiencies in the presence of RSV IN were still observed. The response of the three cell lines to G418 selection differed greatly. Selection of 293 and CV-1 cell lines was approximately double in length of time (10 days) as compared to the BHK cell line (5 days). This in turn corresponds to the approximate doubling time of 24 hours for 293 and CV-1 cell lines in comparison to the 12 hour doubling time of the BHK cell line. If transfections were carried out at the same levels of confluency prior to transfection as with pNRdhfr (70-80%) colonies formed by G418 selection tended to be diffuse and difficult to count accurately. Transfections of pNRneo at high cell

confluency tended to produce a patchy lawn of cells post selection compared to very distinct colonies arising from methotrexate selection. Transfections using pNRneo were carried out at both low confluency (<60%) and approximately half as much substrate DNA (0.14 μg pNRneo in contrast to 0.33 μg of pNRdhfr). Following these guidelines, typically 3 times as many colonies would result from G418 selection of pNRneo as compared to methotrexate selection of pNRdhfr.

It can be concluded that the effect of RSV IN expression *in vivo* towards substrate DNA transfection is not species specific, nor limited to the BHK cell line. Although the cell line selection was by no means exhaustive, the effect of co-transfection of pIN and substrate DNA towards transfection efficiency would seem to result irrespective of the cell line chosen. The majority of experiments were conducted within the BHK cell line due to its well established behaviour and predictable response to selection.

B. SOUTHERN BLOT ANALYSIS OF THE TRANSFECTED BHK GENOME.

In probing a blotted, *KpnI* digested, co-transfected BHK genome, a distinct signal appeared at a position corresponding to approximately 5 kb after electrophoreses(Figure 12). This band corresponded to the full length substrate DNA, as *KpnI* sites were present 30 bp from either termini of the linear substrate DNA molecule. Thus, transfected substrate DNA had recombined with the BHK genome in such a fashion that the entire sequence was conserved. The observed sequence conservation of the substrate DNA molecule was not a function of the

expressed RSV IN. Transfections of substrate DNA alone resulted in a similar banding pattern of blotted and probed *KpnI* digested genomic DNA as had developed with co-transfections of pIN and substrate DNA (Figure 20). Additional bands appeared above and below the main 5 kb signal; particularly distinct in expanded clones (Figure 12, *lanes 6-9*), they were evident as a background smear in pooled and expanded colonies whose genomic DNA was probed (Figure 19, *lane 4*, Figure 20, *lanes 5,6,8,* and *9*). These bands were a result of recombination with the BHK genome which did not conserve the substrate molecule. Recombination between the substrate DNA and BHK genome which resulted in a partial deletion of either termini would effect the loss of a terminal *KpnI* site. Depending on the location of the closest *KpnI* site within the BHK genome, the mass of the *KpnI* restriction fragments containing a both a segment of the substrate DNA molecule and genomic DNA could effectively range across all observed bands on the Southern blot (Fig. 19).

C. USE Lac I REPRESSOR PROTEIN TOWARDS IMPROVED RESCUE OF pNR CLONAL RECOMBINANTS.

Previous attempts to analyze the sequence of pNR clonal recombinants were limited as all of the clones rescued contained a re-ligated NDE-RSV insert. The re-ligated insert indicated that at some point prior to recombination with the BHK chromatin, the substrate DNA had re-circularized within the BHK nucleus. If the template for inverse PCR was digested with *NdeI* prior to clonal amplification (digesting the re-ligated NDE-RSV insert and disrupting all re-

ligated clones from the circular template population) no amplification product was observed. It was therefore necessary to purify the clonal recombinant population away from the overwhelming mass of genomic DNA.

Purification schemes for the rescue of clonal recombinants from transgenic animals have employed affinity chromatography utilizing the Lac I repressor protein (Gossen *et al.* (1993)). A crucial enzyme within the lactose operator complex, the Lac I repressor has been well characterized for nearly three decades (Platt *et al.* (1973)). A tetramer whose molecular mass is 154,520 Da, the lac I repressor binds a lactose operator within the *E.coli* genome 11 bp downstream from the start of transcription of the *lacZ* gene. The operator has an approximate dyad symmetry, one half of the operator binding an N-terminal alpha helix present on a Lac I monomer. Thus the Lac I repressor is capable of binding two complete operators, as recently illustrated by Lewis *et. al.* (1996), whose crystallographic data illustrated the potential for a striking DNA loop having formed between two operators within the *E.coli* genome upon binding of a single Lac I repressor molecule. A gratuitous inducer, isopropyl-β-D-1-thiogalactoside (IPTG), inhibits binding of the Lac I repressor to the lactose operator. Columns to which Lac I repressor protein has been covalently bound, or whose attachment was mediated by antibodies towards Lac I repressor (Gossen *et al.* (1993)), are able to rescue clonal recombinants by virtue of a lactose operator present within many of the plasmids used in construction of expression vectors. A lactose operator was present within the backbone of both pNR constructs, a convenient 150 bp from the 3' terminus of a *NdeI* linearized substrate DNA molecule. Digestion of the co-transfected BHK genome with the endonuclease *HhaI* would release small linear DNA fragments containing the immediate 3' terminal of a substrate DNA molecule, including the

lactose operator, and a section of genomic DNA preceding the next *HhaI* site. Purification of the clonal recombinants and circularization of the DNA fragments provided for a suitable template for inverse PCR (Nolta *et al.* (1996)). Inclusion of the lactose operator allowed purification of these products away from digested genomic DNA, removing competitors for amplification and enriching the desired template DNA.

While construction of a Lac I repressor column was cost prohibitive, an effective substitute was developed. A nitrocellulose membrane is capable of binding protein; however dsDNA fragments remain unbound (Ausubel *et al.* (1996)). Nitrocellulose membranes were used to characterize interactions between purified DNA-binding proteins and sequence specific DNA sites, as well as sequence analysis of those sites retained upon the membrane. Its application in the rescue of clonal recombinants became obvious, as purified Lac I repressor, bound to the lactose operator present on a substrate/genomic DNA fragments, would retain the DNA on the nitrocellulose membrane. Contaminating genomic DNA could be effectively washed away, and protein bound DNA fragments could be eluted from the membrane by addition of IPTG to the washing buffer. This process, termed the nitrocellulose capture assay, became an effective means of purifying a lactose operator containing DNA fragment away from non-specific, genomic DNA fragments. Figure 18 affinity illustrates the result of the nitrocellulose capture assay upon substrate DNA. Approximately 75% of the bound pNRdhfr was recovered from a nitrocellulose membrane following elution with IPTG having previously bound 1 μg of substrate DNA with 20 μg of Lac I repressor.

An enriched population of rescued clonal recombinant fragments (once re-ligated NDE-RSV inserts were removed by *NdeI* digestion) provided the desired amplification of fragments after inverse PCR and subsequent nested PCR. These fragments were cloned and their sequence analyzed. The desired product was to have been evidence of a RSV IN mediated, integrated substrate molecule. Such a molecule would have consisted of a conserved 3' half of the NDE-RSV insert preceding genomic DNA, marking the site of integration. The terminal 3' dinucleotides would have been CA, as has been the case in every documented integration event involving a retrovirus (Fitzgerald *et al.* (1993)), recombinant or otherwise. No such construct was observed. As was previously stated, the overwhelming majority of rescued clonal recombinants had undergone re-ligation of the NDE-RSV insert, thus re-circularizing the substrate DNA. Southern blot analysis of genomic DNA isolated from cells co-transfected with pIN and substrate DNA revealed that much of the substrate was conserved intact during recombination. When combined with the sequence data indicating that the majority of transfected *NdeI*-linearized pNR had re-ligated, this fact points towards a single re-ligated pNR molecule recombining with the BHK genome followed by multiple recombination events between further molecules of re-ligated pNR and previously recombined pNR due to the increasing degree of homology. Such events would produce a concatomeric series of pNR molecules arranged head to tail, which upon release by *KpnI* digestion, would result in release of a conserved substrate molecule, accounting for the major 5 kb signal illustrated upon Southern blots (Figure 19, 20, and 24). The postulated concatomeric series would also release *HhaI* fragments which contained the religated NDE-RSV insert as a result of the head to tail arrangement of recombinants in a series.

The activity of the RSV *in* gene has been established *in vitro*, and co-transfections of pIN and substrate DNA yield ten fold more colonies post selection than transfections of substrate DNA alone. What effect does expressed RSV IN have *in vivo* which increases transfection efficiencies of substrate DNA? It can be postulated that the RSV IN produced within the cell was acting to translocate bound substrate DNA to the nucleus. Indeed there exists a nuclear localization signal near the carboxy terminus of IN which consists of a basic dipeptide separated by 10 residues and then three basic amino acids within the next five residues (Figure 5), consistent with the nucleoplasmin protein nuclear localization signal (Kurz *et al.* (1997)). As our results have shown, the concentration of substrate DNA transfected was the limiting factor as only a small amount of RSV IN was required to effectively translocate the substrate to the nucleus. Once within the nucleus the RSV IN was probably unable to out compete endogenous nuclear enzymes, ultimately determining the fate of the substrate DNA. Translocation was only accomplished when the substrate DNA molecule was linear and displayed terminal elements which were normally observed in viral RSV DNA prior to formation of the provirus. The absence of RSV IN or the presence of a pNR molecule linearized with an endonuclease other than *NdeI* resulted in colony formation equal to those cells transfected with only substrate DNA

Other examples of recombination between the substrate DNA and the BHK genome were analyzed. Using the GENBANK sequence database as a reference, BLAST searches provided homologous matches when queried with sequence data from clonal recombinants. Second in frequency to examples of re-ligated substrate DNA were clones arising from intra-molecular recombination between substrate molecules either preceding or post-recombination with the BHK genome. All of these clones contained small segments of the pNR backbone which were on

average 35 bp in length and arranged in a random manner (Figure 18). More surprising were the examples of recombination which occurred repeatedly with the same gene within the BHK genome. Two 'hotspots' were recorded: they yielded sequence data with a high degree of homology to the rat histamine receptor and a mutant tetrahydrofolate reductase gene (present within the BHK genome making the cell line methotrexate sensitive). The assumption that these repeated recombinations within the same gene were an aberration from a single transfection would prove to be false, as the examples arose from separate transfection and amplification experiments that were separated by a period of months. Considering the size of the BHK genome at approximately 3 billion base pairs, chance would not be statistically feasible; instead recombination within these genes was non-random. Previous studies have shown that a retroviral provirus will recombine within actively transcribing regions of the host cell genome (Kirchner *et al.* (1995)), but such a result within this study was unexpected. Perhaps the histamine receptor gene and the tetrahydrofolate reductase gene were being actively transcribed at the time of transfection, and upon entry of the substrate DNA into the nucleus via RSV IN, recombination occurred with such an obvious target. The examples of histamine receptor recombination were odder still as three separate recombination events, from three separate transfections, were recorded to have occurred in a 9 base pair region of the histamine receptor gene. The final group of clonal recombinant rescued were examples where no homologous sequence was returned from the BLAST search; these may represent examples of recombination where substrate DNA had inserted into non-homologous portions of the BHK genome.

D. CHARACTERISTICS OF A STABLE BHK CELL LINE EXPRESSING ROUS SARCOMA VIRUS INTEGRASE.

Since no examples of conserved integration of the substrate DNA were recorded, a cell line was constructed which expressed high levels of RSV IN on the premise that IN mediated integration of transfected substrate DNA could be recorded. Stable expressors of IN were selected from a population of pNUT-IN transfected BHK cells. Colonies of methotrexate resistant cells appeared by virtue of a DHFR cassette contained within the vector pNUT-IN. Colonies were picked, expanded, and were stained by *in situ* hybridization of a polyclonal antibody directed towards IN. *In situ* cell staining proved sensitive enough to divide expanded colonies into three categories of IN expressing cell lines; BHK-IN+ cells expressed little or no integrase, yet were methotrexate resistant, BHK-IN++ cells were moderate expressors of IN, while BHK-IN+++ cells expressed high levels of IN. (Figure 21). These three cell lines were cultured and transfected with substrate DNA.

There was a direct correspondence for BHK-IN cell survival following transfection of substrate DNA and G418 selection to intracellular levels of RSV IN. Surprisingly, high levels of IN did not increase cell survival under G418 selection; instead BHK-IN+++ cells transfected with substrate DNA died before G418 selection had taken hold. There are several possible explanations for this phenomenon. If as proposed IN acts to translocate recognizable substrate DNA to the nucleus, high levels of IN would concentrate transfected substrate DNA within the nucleus. Such abnormally high levels could lead to increased recombination events with an increased likelihood of disrupting a critical gene. This is a reasonable hypothesis, although a

Southern blot of genomic DNA from resultant G418 resistant colonies indicated that the amount of recombined pNRneo was not greater than levels of recombination seen from selected BHK colonies transfected solely with substrate DNA (Figure 23). Indeed the level of recombination seen for pNRneo within BHK-IN cells was nowhere near the amount of recombination and/or amplification of the pNUT-IN construct within those same cells (Figure 23). If IN concentrated pNRneo within the nucleus, increased expression of neomycin could prove toxic to the cell. This phenomenon has been noted in gene transfer trials where there appeared to be selective pressure against clones expressing high levels of neomycin (personal communication, Dr. K. Humphries). It also stands to reason that abnormally high levels of extra-chromosomal DNA within the nucleus would disrupt normal nuclear processes, for example competing away DNA binding proteins from BHK chromatin or possibly signaling apoptosis. Transfections of uncut pNRneo did not illicit cell death, instead BHK-IN cells transfected with uncut pNRneo responded similarily to normal BHK cells transfected with uncut pNRneo or substrate DNA. When G418 was not included in the media of BHK-IN++ and +++ cells following transfection of substrate DNA, noticeable cell death occurred, although not as markedly as before since selective pressure had been removed from non-transfected cells.

G418 resistant colonies developed from all three BHK-IN cell lines transfected with substrate DNA. The trends evident in premature cell death of substrate transfected BHK-IN cells were also observed in colony formation for those same cells. BHK-IN+ cells displayed no ill effects towards substrate DNA transfection due to little or no IN expression. Resultant colonies post-selection were equal in numbers to transfections of BHK-IN+ cells with uncut pNRneo, or transfections of normal BHK cells with substrate DNA alone. BHK-IN++ and BHK-IN+++ cells

developed fewer colonies under selection when transfected with substrate DNA; the more intracellular IN expressed, the fewer colonies resulted from selection. Furthermore, transfections of uncut pNRneo in either BHK-IN++ or +++ cell lines produced colonies in numbers equivalent to either similarily transfected BHK-IN+ or normal BHK cells. These observations also support the theory that high levels of IN expression ultimately lead to cell death upon transfection of a suitable substrate DNA due to increased translocation of substrate DNA to the nucleus.

Clonal recombinants were also rescued from genomic DNA collected from G418 resistant colonies arising from BHK-IN cell lines. The population of rescued clonal recombinants was the same as those rescued from BHK cells co-transfected with pIN and substrate DNA. The majority of products were re-ligated substrate DNA, many examples of intramolecular recombination with other substrate DNA molecules, and occasional random recombinants. One rescued clone showed a high degree of homology with the mutant tetrahydrofolate reductase cassette, although it should be remembered that the previously recombined pNUT-IN construct contained a DHFR cassette.

To gain insight as to the fate of substrate DNA within the nucleus of cells expressing IN nuclear extracts were collected from pooled methotrexate resistant colonies arising from BHK cells transfected with pNUT-IN. Analysis of nuclear extracts incubated with substrate DNA were performed. It was hoped that evidence of the re-ligated substrate DNA would be visible on a Southern blot of substrate DNA incubated with the nuclear extracts. No such result was visible—only evidence of specific and non-specific nucleases which degraded the substrate DNA. Aside from the rescued re-ligated clonal recombinants, such nuclease activity would help explain why

no products on IN mediated integration were rescued. The terminal ends, with associated IN protein, were cleaved from the remainder of the substrate DNA which proceeded to recombine with the BHK genome. Identical nuclease activity which resulted in substrate DNA degradation was observed in normal BHK cells as in BHK-IN cell lines.

E. CONCLUSIONS AND FUTURE DIRECTIONS.

Co-transfections of pIN and substrate DNA lead to increased transfection efficiencies, and thus expression of genes contained within the substrate DNA molecule. Transfection efficiencies increased as the levels of substrate DNA transfected were increased, and only a small amount of transfected pIN was required to effect a ten-fold increase in colony formation. Indeed over expression of IN followed by transfection of substrate DNA was toxic to cells. The means by which IN elicits the increased transfection efficiencies has not been positively confirmed, although evidence collected allows for the hypothesis that IN acts to translocate a recognizable substrate DNA to the nucleus of co-transfected cells. Once within the nucleus, IN is presumably unable to out compete the endogenous nuclear enzymes which ultimately determined the fate of the delivered substrate DNA.

Several potential applications for the co-transfection protocol are foreseen. The improved transfection efficiencies and high expression resulting from co-transfection of pIN with substrate DNA would facilitate the construction of recombinant cell lines. Plasmids of up to 40 kb can be

propagated in *E. coli* (Graham *et al.* (1991)), and hence plasmids could be engineered to contain not only cDNAs, but entire genes including their regulatory regions. Large genetic regions may prove to be essential for long term expression of genes targeted by gene therapies (Anderson *et al.* (1992)) and currently all recombinant viral vectors are limited by the size of the insert which can be packaged (~7 kb). There exist gene therapy protocols that currently use liposomes to deliver DNA *in vivo* (Dubé *et al.*(1995)) which would certainly benefit from the added transfection efficiency provided by substrate DNA and pIN. In non-dividing cells, the substrate DNA would most likely remain extrachromosomal (Dubé *et al.*(1995)), and the higher levels of substrate DNA within the nucleus would provide extended gene expression. The protocol may also provide a means for more efficient genetic manipulation of cells leading to the creation of transgenic animal models.

There exists the possibility that IN mediated integration of substrate DNA could be effected within a co-transfected cell. Selective expression of retroviral *gag* genes along with retroviral IN could provide the framework necessary to protect the IN-substrate DNA interaction within a cell nucleus such that IN could effect integration of the substrate DNA. Ultimately such constructs could result in a person made gene transfer vehicle—a targeted liposome containing genetic elements for expression of viral proteins which would recognize, transfer, and protect substrate DNA such that stable integration could occur. Such constructs would resemble a 'synthetic' virus.

The long-term goal of these studies was to provide a new and subtle form of gene transfer which would allow integration of discreet pieces of DNA containing only the gene(s) of interest.

By expressing IN only transiently inside the cell, the protocol resulted in viral elements not persisting in the cell indefinitely. On this point our protocol differs greatly from those which rely on recombinant retroviral infection to deliver and integrate the gene of interest. Further study will be aimed to ensuring integrase association with the substrate DNA within the nucleus of transfected cells. As improvements in the method of DNA delivery are made, and the integrase/substrate DNA binding becomes more controllable, a future goal is to provide a general protocol for use in gene therapy.

BIBLIOGRAPHY:

Anderson, W. F. **Human gene therapy** *Science* **256,** 808-813 (1992)

Ausubel, F. A., Brent, R. Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A., Struhl, K. (eds) 1996. **Current protocols in Molecular Biology**. John Wiley & Sons, New York.

Bally, M. B., Zhang, X. P., Wong, F. M. P., Kong, S., Wasan, E., Reimer, D. L. **Lipid/DNA complexes as an intermediate in the preparation of particles for gene transfer: an alternative to cationic liposome/DNA aggregates.** *Advanced Drug Delivery Reviews* **24**,275-290 (1997)

Blau, H. M., Springer, M. L. **Gene therapy: a novel form of drug delivery** *New England J. Med.* **333,**1204-1207 (1995)

Bowerman, B., Brown, P. O., Bishop, J. M., Varmus, H. E. **A nucleoprotein complex mediates the integration of retroviral DNA** *Gene Dev.* **3**, 469-478 (1989)

Brenner, M. K., Rill D. R., Heslop, H. E., *et al* **Gene marking after bone marrow transplantation** *Eur. J. Cancer* **30,** 1171-1176 (1994)

Brenner, M. K., Rill D. R., Holladay M. S., *et al* **Gene marking to determine wether autologous marrow infusion restores long-term haemopoieses in cancer patients** *Lancet* **342,** 1134-1137 (1993)

Bunnell, B.A., Muul, L.M., Donahue, R.E., Blases, R.M., Morgan, R.A. **High efficiency retroviral mediated gene transfer into human and non-human primate peripheral blood lymphocytes.** *Proc.Natl. Acad. Sci. USA.* **92**(77), 39-43 (1995)

Bushman, F. D., Wang, B. **Rous Sarcoma Virus integrase protein: mapping functions for catalysis and substrate binding** *J. Virol.* **68,** 2215-2223 (1994)

Bushman, F. D., Craigie, R. **Activities of HIV integration *in vitro*: specific cleavage and integration of HIV DNA.** *Proc. Natl. Acad. Sci. USA* **88,** 1339-43 (1991)

Collicelli, J., Goff, S.P. **Mutant and pseudo revertants of Moloney murine leukemia virus with alterations at the integration site.** *Cell* **42,** 573-80 (1985)

Cossett, F.-L.,Takeuchi, Y., Battini, J.-L., Weiss, R.A., Collins, M.K. **High titer packaging cells producing recombinant retroviruses resistant to human serum.** *J. Virol.* **69**(12), 7430-36 (1995)

Craigie, R., Fujiwara, T., Bushman, F. **The IN protein of MoMLV processes the viral DNA ends and accomplishes their integration *in vitro*.** *Cell* **62,** 1829-37 (1990)

Danos, O., and Mulligan, R. C. **Safe and efficient generation of recombinant retroviruses with amphotropic and ectopic host ranges** *Proc. Natl. Acad. Sci. USA* **85,** 6460-6464 (1988)

Dubé, I. D., Cournoyer, D. **Gene therapy: here to stay** *Can. Med. Assoc.* **152,** 1605-1613 (1995)

Eisenman, R. N., Mason, W. S., Linial, M. **Synthesis and processing of polymerase proetins of wild-type and mutant avian retroviruses** *J. Virol.* **36,** 62-78 (1980)

Engelmann, A., Mizuuchi, K., Craigie, R. **HIV-1 DNA integration: mechanism of viral DNA cleavage and DNA strand transfer** *Cell* **67,** 1211-1221 (1991)

Fields, B. N., and Knipe, D. M. **Field's Virology**, 2nd Ed., Raven Press, New York (1990)

Fitzgerald, M. L., Grandgenett, D. P. **Retroviral integration: *in vitro* host site selection by avian integrase** *J. Virol.* **68,** 4314-4321 (1994)

Friedmann, T., Robin, R. **Gene therapy for human genetic disease?** *Science* **175,** 949-955 (1972)

Fujiwara, T., Craigie, R. **Integration of mini-retroviral DNA: a cell-free reaction for biochemical analysis of retroviral integration.** *Prceedings of the National Academy of Sciences of the U.S.A.* **86**(9), 3065-3069 (1989)

Gossen, J. A., de Leeuw, W. J. F., Molijn, A. C., Vijg, J. **Plasmid rescue from transgenic mouse DNA using Lac I repressor protein conjugated to magnetic beads.** *Biotechniques* **14**(4), 624-629 (1993)

Grandgenett, D. P., Inman, R. B., Vora, A. C., Fitzgerald, M. L. **Comparison of DNA binding and integration half-site selection by Avian Myeloblastosis Virus integrase** *J. Virol.* **67,** 2623-2636 (1993)

Graham, F. L., and Prevec, L. (1991) **Methods in Molecular Biology**, Vol. 7, Murray, E.J. (Eds.), The Humana Press Inc., Clifton, N.J.

Grimm, E., Arbuthnot, P. **Rapid purification of recombinant *Taq* DNA polymerase by freezing and high temperature thawing of bacterial expression cultures** *Nucl. Acids Res.* **23**(21), 4518-9, (1995)

Grossman, M., Raper, S. E., Kozarsky, K., *et al* **Successful *ex vivo* gene therapy directed to liver in a patient with familial hypercholesterolaemia** *Nat. Genet.* **6,** 335-341 (1994)

Hong, T., Murphy, E., Groarke, J., Drlica, K. **HIV type ! DNA integration: fine structure target analysis using synthetic oligonucleotides.** *J. Virol.* **67**(2), 1127-31 (1993)

Invitrogen. ***Pichia* expression kit instruction manual**. Version 3.0, San Diego, CA. (1995)

Jones, K. S., Coleman, J., Merkel, G. W., Laua, T. M., Skalka, A. M. **Retroviral integrase functions as a multimer and can turn over catalytically** *J. Biol. Chem.* **2647,** 16037-16040 (1992)

Jonsson, C. B., Donzella, G. A., Roth, M. J. **Characterization of the forward and reverse integration reactions of the Moloney Murine Leukemia virus integrase protein purified from *Excherichia coli*** *J. Biol. Chem.* **268,** 1462-1469 (1993)

Katz, R.A., Merkel, G., Kulkosky, J., Leis, J., Skalka, A.M. **The avian retroviral IN protein is both necessary and sufficient for integrative recombination *in vitro*.** *Cell* **63,** 87-95 (1990)

Kirchner, J., Connolly, C. M., Sandmeyer, S. B. **Requirement of RNA polymerase III transcription factors for *in vitro* position specific integration of a retrovirus like element.** *Science* **267**, 1488-1491 (1995)

Kolberg, R. *J. NIH Res.* **4,** 43-44 (1992)

Kurz, M., Doenecke, D., Albig, W. **Nuclear transport of H1 histones meets the criteria of a nuclear lacalization signal-mediated process** *J. Cell. Biochem.* **64**, 573-578 (1997)

Lebedeva, L., Fujita, K., Nihrane, A., Silver, J. **Infectious particles derived from Semliki Forest Virus vectors encoding murine leukemia virus envelopes.** *J. Virol.* **71**(9), 7061-67 (1997)

Lee, S. P., Censullo, M. L., Kim, H. G., Han, M. K. **Substrate length dependant activities of Human Immunodeficiency Virus type 1 integrase *in vitro*: differential DNA binding affinities associated with different lengths of substrates** *Biochemistry* **34,** 10215-10223 (1995)

Lee, S. P., Censullo, M. L., Kim, H. G., Han, M. K. **Substrate length dependant activities of Human Immunodeficiency Virus type 1 integrase *in vitro*: differential DNA binding affinities associated with different lengths of substrates.** *Biochemistry* **34,** 10215-10223 (1995)

Lewis, M., Chang, G., Horton, N. C., Kercher M. A., Pace, H. C., Schumacher, M. A., Brennan, R. G., Lu, P. **Crystal structure of the lactose operator repressor and its complexes with DNA and inducer.** *Science* **271**, 1247-1254 (1996)

Lewis, P.F., Emmerman, M. **Passage through mitosis is required for oncoretroviruses but not for the human immunodeficiency virus.** *J. Virol.* **68,** 510-16 (1994)

Leavitt, A.D., Rose, R.B., Varmus, H.E. **Both substrate and target oligonucleotides sequences affect *in vitro* integration mediated by human immunodeficiency virus type 1 integrase protein produced in *S. cerevisiae.*** J. Virol. **66**(4), 2359-68 (1992)

Loiler, S.A., DiFronzo, N.L., Holland, C.A. **Gene transfer to human cells using retrovirus vectors produced by a new polytropic packaging cell lines.** *J. Virol.* **71**(6), 4825-28 (1997)

Mercola, K. E., Cline, M. J. **The potentials of inserting new genetic information** *N. Engl. J. Med.* **292,** 1297-1300 (1980)

Miller, A.D., Chen, F. **Retroviral packaging cells based on 10A1 murine leukemia virus for production of vectors that use multiple receptors for cell entry.** *J. Virol.* **70**(8), 5564-71 (1996)

Miller, D. A. **Human gene therapy comes of age** *Nature* **375,** 455-460 (1992)

Mizuuchi, K. **Polynucleotidyl transfer reactions in transpositional DNA recombination** *J. Biol. Chem.* **267,** 21273-21276 (1992)

Mumm, S. R., Grandgenett, D. P. **Defining nucleic acid binding properties of avian retrovirus integrase by deletion analysis** *J. Virol.* **65,** 1160-1167 (1991)

Mumm, S. R., Horton, R., Grandgennett, D. P. **v-Src enhances phosphorylation at Ser-83 of the Rous Sarcoma Virus integrase** *J. Virol.* **66,** 1995-1999 (1992)

Nabel, E. G. **Gene therapy for cardiovascular disease.** *Circulation* **91**(2), 541-548 (1995)

Nabel, G.J., Nabel, E.G., Yang, Z., Fox, B., Plautz, G., Gao, X. **Direct gene transfer with DNA liposomes complexes in melonoma: expression, biologic activity and lack of toxicity in humans.** *Proc. Natl. Acad. Sci. USA* **91,** 11557-61 (1994)

Nolta, J. A., Dao, M. A., Wells, S., Smogorzewska, M., Kohn, D. B. **Transduction of pluripotent stem cells demonstrated by clonal analysis after engraftment in immune deficient mice** *Proc. Natl. Acad. Sci. USA* **93,** 2414-2419 (1996)

Ory, P.S., Neugeborey, B.A., Mulligan R.C. **A stable human derived packaging cell line for production of high titer retroviral/vesicular stomatitis virus G pseudotypes.** *Proc. Natl. Acad. Sci. USA* **93,** 11400-406 (1996)

Palmiter, R. D., Begringer, R. R., Quaife, C. J., Maxwell, F., Maxwell, I. H., Brinster, R. L. **Cell lineage ablation in transgenic mice by cell-specific expression of a toxin gene** *Cell.* **50**(3), 435-443 (1987)

Pear, W. S., Nolan, G. P., Scott, M. L., Baltimore, D. **Production of high-titer helper free retrovirus by transient transfection** *Proc. Natl. Acad. Sci. USA* **90,** 8392-8396 (1993)

Platt, T., Files, J., Weber, K. **Lac repressor; specific proteolytic destruction of the NH_2-terminal region and loss of the DNA binding activity.** *Journal of Biol. Chem.* **248**(1), 110-121 (1973)

Price, J., Turner, D., Cepko, C. **Lineage analysis in the vertebrate nervous system by retrovirus-mediated gene transfer** *Proc. Natl. Acad. Sci. USA* **84,** 156-160 (1987)

Pryciak, P.M., Sil, A., Varmus, H.E. **Retroviral integration into mini-chromosomes *in vitro*.** *EMBO J.* **11**(1), 291-303 (1992)

Rigg, R.J., Dando, J.S., Forestell, S.P., Plavec, L., Bohnlein, E. **A novel human amphotropic packaging cell line: high titer, complement resistance and improved safety.** *Virology* **218,** 290-295 (1996)

Rosenberg, S. A., Aebersold, P., Cornetta, K. **Gene transfer into humans: immunotherapy of patients with advanced melanoma using tumour infiltrating lymphocytes modified by retroviral gene transduction** *N. Engl. J. Med.* **323,** 570-578 (1990)

Roth, M.J., Schwartzberg, P.L., Goff, S.P. **Structure of the termini of DNA intermediates in the intgeration of retroviral DNA: dependance on IN function and terminal DNA sequence.** *Cell* **58,** 47-54 (1989)

Sambrook, J., Fritsch, E. F., Maniatis, T. (eds.) 1989. **Molecular cloning, a laboratory manual.** 2nd Edition. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Soneoka, Y., Cannon, P.M., Ramsdale, E.E., Griffiths, J.C., Romano, G. **A transient three-plasmid expression system for the production of high titer retroviral vectors.** *Nucl. Acids Res.* **23**(4), 628-633 (1995)

Takeuchi, Y., Cossett, F.-L.C., Lachman, P.J., Okada, H., Weiss, R.A., Collins, M.K.L. **Sensitization of cells and retroviruses to human serum by (alpha 1-3) galactosyltransferase.** *J. Virol.* **68,** 8001-7 (1994)

Tripathy, S.K., Goldwasser, E., Lu, M.M., Barr, E., Leidew, J.M. **Stable delivery of physiological levels of recombinant erythropoietin to the systemic circulation by intramuscular injection of replication defective adenovirus.** *Proc. Natl. Acad. Sci. USA* 91, 11557-61 (1994)

Vijaya, S., Steffen, D.L., Robinson, H.L. **Acceptor sites of retroviral integrations map near DNase I hypersensitivity sites in chromatin.** *J. Virol.* **60**(2), 683-92 (1986)

Vile, R., Russell, S. J. **Gene transfer technologies for the gene therapy of cancer**. *Gene Therapy* 1, 88-98 (1994)

Vincent, K. A., Ellison, V., Chow, S. A., Brown, P. O. **Characterization of Human Immunodeficiency Virus type 1 integrase expressed in *Escherichia coli* and analysis of variants with amino terminal mutations** *J. Virol.* **67,** 1562-1469 (1993)

Vincent, K. A., Ellison, V., Chow, S. A., Brown, P. O. **Characterization of Human Immunodeficiency Virus type 1 integrase expressed in *Escherichia coli* and analysis of variants with amino terminal mutations** *J. Virol.* **67,** 1562-1469 (1993)

Vink, C., van Gent, D.C., Elgersma, Y., Plasterk, R.H.A. **HIV integrase protein requires a subterminal position of its viral DNA recognition sequence for effcient cleavage.** *J. Virol.* **65**(9), 4636-44 (1991)

Yu, H., Rabson, A.B. Kaul, M., Ron, Y., Dougherty, J.P. **Inducible human immunodeficiency virus type-1 packaging cell lines.** *J. Virol.* **70**(7), 4530-37 (1996)

Zabner, J., Petersen, D.M., Puga, A.P., Graham, S.M., Couture, L.A., Keye, L.D. **Safety and efficacy of repetitive adenovirus mediated transfer of CFTR cDNA to airway epethilia of primates and cotton rats.** *Nat. Genet.* **6,** 75-83 (1994)