Title
A nucleoprotein complex that mediates the integraion reaction of the Moloney murine leukemia virus

Author
Bowerman, Bruce,

Publication Date
1989

Peer reviewed|Thesis/dissertation

A Nucleoprotein Complex that Mediates the Integration Reaction of the Moloney Murine Leukemia Virus

by

Bruce Alan Bowerman

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Biochemistry

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA

San Francisco

[redacted]

. . . . . .

Date **University Librarian**

SEP 03 1989

Degree Conferred: . . . . . .

I am still a learner, not a teacher, feeding somewhat omnivorously, browsing both stalk and leaves; but I shall perhaps be enabled to speak with the more precision and authority by and by,--if philosophy and sentiment are not buried beneath a multitude of details.

--Henry David Thoreau

I dedicate this thesis to my parents.

ACKNOWLEDGEMENTS

The friendly and creative atmosphere that distinguishes UCSF has made being a student here a privilege and a pleasure. I value highly the chances I have had to interact with the many professors, post-docs, technicians, and graduate students who collectively define UCSF as I know it. Many people have assisted me in many ways during my six years of graduate school, and I am grateful to all of them for the time they gave to me. I especially thank Harold Varmus, my advisor, for his always enthusiastic attention to my work and for his exemplary ability to point out ways in which I could improve myself both in science and in communication. I am also particularly grateful to Pat Brown, without whom much of this work would not have been possible, and who showed me that science can be both rigorous and fun. Finally, I thank the friends I have made here who helped me through the rough times and watched out for me when celebrating the good times.

ABSTRACT: A Nucleoprotein Complex that Mediates the Integration Reaction of the Moloney Murine Leukemia Virus

by Bruce Alan Bowerman

We have shown that in cells acutely infected by the Moloney murine leukemia virus (MLV), unintegrated viral DNA resides within a large (160S) nucleoprotein complex. In cell-free extracts, these complexes mediate authentic retroviral integration, which we were able to detect with a sensitive genetic assay. The integration-competent nucleoprotein complexes can be coprecipitated with anti-sera that recognize the viral capsid protein (CA). The presence of CA and the behavior of the complexes after treatment with nucleases suggest that the complexes are a derivative of the extracellular virion core. The DNA-containing complexes from infected-cell extracts copurify with other viral complexes that, in addition to CA, contain two other viral proteins, IN and reverse transcriptase (RT), but lack detectable quantities of matrix (MA). Our analysis of the in vitro integration reaction suggested and eventually proved that, for MLV, the immediate precursor to the provirus is the linear form of unintegrated viral DNA. Analysis of the ends of unintegrated linear viral DNA showed that most of the linear molecules have 3' ends that are recessed two bases. The 3' end of the recessed linear DNA corresponds to the position at which viral DNA becomes joined to host DNA, and the viral integration protein (IN) is required for production of the recessed ends. Unresolved issues in retroviral integration are addressed by a comparative analysis with two prokaryotic examples of recombination, lambda integration and Mu transposition.

Harold Varmus

Table of Contents

List of Figures

List of Tables

CHAPTER 1

INTRODUCTION

Animal viruses and bacteriophages have often provided relatively simple and experimentally tractable examples of important biological processes. Retroviruses have been no exception to this rule. They are perhaps best known for their ability to cause cancer in animals, a phenomenon responsible for their initial discovery and for their now outdated name, oncornaviruses. Retrovirus infections can activate a variety of cellular genes that subsequently contribute to abnormalities in growth control associated with the development of cancer. Over fifty such proto-oncogenes have now been identified, providing important insights into the molecular biology of cancer and contributing generously to our understanding of normal growth and development (Bishop 1983; Bishop and Varmus 1982, 1985, Varmus 1987).

In addition to their association with cancer, retrovirus infections are known to induce a variety of other diseases as well, including immune deficiencies, anemia, arthritis, pneumonia, and nervous system disorders (Teich et al 1982; Varmus 1988; Varmus and Brown 1989). The relatively recent discovery, in the last decade, that retrovirus infections can lead to cancer in humans, and the identification of the human immunodeficiency virus (HIV) as the causative agent of AIDS, have made the study of retrovirus replication increasingly important in efforts to develop anti-viral therapies (Weiss 1985; Varmus 1988a, 1988b; Cullen and Greene 1989).

Retroviruses are Rich in Genetic Recombination

The importance of genome stability to the proper expression of an organism's genetic information and the significance of mutation and genome rearrangement for genetic variability and evolution constitute two opposing demands upon genome structure. A desire to understand how life has evolved in response to these fundamental demands has motivated a vast array of experimental efforts directed at uncovering molecular events that affect genome structure and stability. Examples include the study of proofreading and mismatch repair mechanisms that insure sufficient fidelity in the replication of DNA, and the molecular analyses of a large variety of prokaryotic recombination reactions, such as lambda integration and excision, and Mu transposition, which cause DNA rearrangements in the bacterial genome.

The study of retroviruses, in addition to improving our understanding of carcinogenesis, has shed light on a variety of other interesting biological phenomena (Varmus 1983, 1988a; Varmus and Brown 1989). A recurring theme in retroviral replication that is of general interest is the role of genetic recombination. Four unusual features of the retrovirus life cycle make these viruses particularly attractive as a system in which to study recombination:

(i) As their name suggests, retroviruses (and their relatives, hepadnaviruses and caulimoviruses) are capable of reversing one step in what was once thought to be the unidirectional flow of genetic

information. The "central dogma" of molecular biology dictates that information flows from DNA to RNA to protein. By virtue of an activity encoded in their *pol* gene, retroviruses have the ability to reverse transcribe their RNA genomes into double stranded DNA upon infection of susceptible cells (Baltimore 1970; Temin and Mizutani 1970). These viruses reverse the flow of genetic information at the level of transcription, providing one way to generate additional copies of DNA sequences.

Reverse transcription is now known to be widespread in biology. In addition to being related to hepadnaviruses and caulimoviruses, retroviruses belong to a group of elements called retrotransposons, which includes *Ty* in yeast and *copia* in fruit flies (Varmus 1983; Boeke 1985, 1989; Bingham and Zachar 1989). These elements all reside in host genomes as DNA proviruses and replicate through RNA intermediates. Retroviruses differ from other retrotransposons in that they produce an extracellular virus that can enter new cells, whereas elements like Ty undergo only an intracellular pathway of retrotransposition. Retrotransposons, in turn, are part of an even larger family of elements that appear to generate additional copies of their DNA from RNA intermediates in much less well understood ways. Examples include LINEs and SINEs in mammalian cells, F and I elements of Drosophila, the DIRS-1 element of *Dictyostelium*, and the recently described msDNAs in myxobacteria and *E. coli* (Capello et al 1985; Weiner et al 1987; Varmus 1989; Deininger 1989; Finnegan 1989a, 1989b; Hutchinson et al 1989; Varmus and Brown 1989).

Indeed, reverse transcription appears to have had a significant impact on genome structure over the course of evolution in that retroviruses and their relatives sometimes account for a substantial fraction an organism's genetic material. For example, proviral DNA sequences account for as much as 1% of the mouse genome (Varmus and Brown 1989). Whether or not reverse transcription has played a functional role in evolution remains a matter of speculation, but clearly the relevance of this activity to gene duplication and DNA rearrangement is of interest to molecular archaeologists.

The process of reverse transcription is of additional interest in understanding how genomes evolve because of the current speculation that life may have originated through the evolution of replicating molecules of RNA (Alberts et al 1989). Models of molecular evolution hypothesize that at some point, RNA genomes were converted into the more stable form of DNA, which now predominates as the nucleic acid of choice for the storage and transmission of genetic information (only among viruses and viroids do we still see examples of RNA genomes). Reverse transcription is one example of how such a conversion to DNA can occur.

(ii) Retroviruses are unique among known animal viruses in that the orderly integration of a DNA copy of their genome into the chromosome of an infected cell is an obligate step in their life cycle. Parvoviruses may be an exception; they sometimes integrate into the host genome as part of an alternative replication pathway (Cheung 1980), but it is not yet known if this putative example of integration

occurs in a precise fashion. Although the DNA genomes of other animal viruses occasionally become joined to host chromosomes, such events appear to occur by imprecise mechanisms and are not believed to be required for viral replication (Botchan et al 1976; Wettstein 1982; Robinson and O'Callaghan 1983; Matsuo 1984). The retrovirus life cycle, then, provides an unusual example of transposition in a vertebrate system. Moreover, the study of retroviral integration is informative with respect to other retrotransposons, which appear to require a similar reaction in their transposition cycles (Eichinger and Boeke 1988).

(iii) Another interesting feature of retroviruses is their diploidy (Bender and Davidson 1976, Bender et al 1978). The extracellular virion particle contains two (or some multiple of two) copies of single-stranded, genomic RNA, which are of the same polarity as viral mRNA (and are therefore said to be plus-stranded). Infected cells which express only one proviral template bud progeny virus which contain two identical copies of viral RNA. But in a mixed infection, in which a single cell acquires proviruses from two different incoming retroviruses, heterozygous particles can form, setting the stage for genetic recombination between different viral genomes during the ensuing round of reverse transcription. This type of viral recombination, which can occur with efficiencies as high as 10 to 30%, provides another striking example of genome rearrangement in the life cycle of retroviruses (Hunter 1978; Coffin 1979; Linial and Blair 1982, 1985).

(iv) All three of the above features of retrovirus replication act to insure the establishment of proviral DNA in an infected cell. These steps in replication occasionally accomplish another remarkable genetic rearrangement by transducing cellular genes. The mechanism of transduction is complex and appears to require more than one cycle of viral replication (Varmus and Swanstrom 1982, 1985; Varmus and Brown 1989). Models for retroviral transduction require as a starting point the insertion of a provirus near the 5′ end of a cellular gene. Viral sequences are then proposed to become fused to host sequences either through the deletion of intervening DNA or through RNA processing events. Chimaeric virus/host RNA molecules are then postulated to be packaged into mixed particles that contain a wild-type viral genome in addition to the chimaeric genome. Recombination during reverse transcription can then produce either a replication-competent virus containing host sequences (such as Rous sarcoma virus) or, more typically, a replication-defective virus in which much of the viral coding potential has been replaced by the transduced cellular sequences. Such defective viruses include all of the *cis*-acting sequences in the viral genome required for replication but can be propagated only with the assistance of wild-type "helper" functions required in *trans*. This unusual capacity to transduce cellular genes highlights the features of retroviral replication that are relevant to studies of genetic recombination.

Objectives

I became interested in retroviruses because of their ability to integrate DNA copies of their genome into host chromosomes. The processes of reverse transcription, viral recombination, and integration into the genome presumably all occur within some derivative of the extracellular virus particle. The general goal of my thesis has been to better define the machinery retroviruses use inside infected cells to accomplish these important steps in replication, and to compare these intracellular viral structures to the extracellular virion particle.

I will first provide a description of the context in which integration occurs by discussing the essential features of retroviral replication (see Figure 1). I will do so by describing the viral genome and the gene products it encodes, and how these products assemble into an extracellular particle. Finally, a summary of the known structural features of the mature virion provides a point of departure for considering the events in replication that culminate in the integration of viral DNA into the host genome.

The Virion and the Provirus; Early and Late Events in Replication

Two convenient landmarks in the retrovirus life cycle are (i) the extracellular virus particle and (ii) the DNA provirus covalently linked to a host chromosome. The virus particle initiates an infection by binding to and entering susceptible cells, and the provirus provides the

template for the expression of viral gene products that eventually assemble into new virus particles.

The events that transpire during entry into a susceptible cell and conversion of the viral RNA genome into a DNA provirus constitute what are often referred to as the early events of retroviral replication. These events depend largely on the activity of viral proteins. The subsequent expression of the proviral genome, by the host transcriptional and translational machinery, and the assembly of viral gene products into progeny virions comprise the late events of replication. In this context, integration is the ultimate step of the early events in replication. It is essential in that viral gene products are expressed efficiently only from an integrated provirus.

Dividing retroviral replication into early and late stages conveniently parallels the molecular genetic methods that have been widely used to dissect the function of retroviral gene products. The existence of unintegrated viral DNA and DNA proviruses in retrovirus-infected cells has made it possible to obtain infectious clones of retroviral DNA that are fully capable of generating progeny virus when introduced back into cells by DNA transfection procedures. Otherwise normal viral genomes containing site-directed mutations can be reintroduced into cells to artificially establish a provirus, bypassing the early steps in replication. These mutant viruses can then be examined for any phenotypes that shed light on the normal function of the mutated sequences.

The early events in replication are essentially conservative in nature--one incoming virus particle probably produces a single provirus.

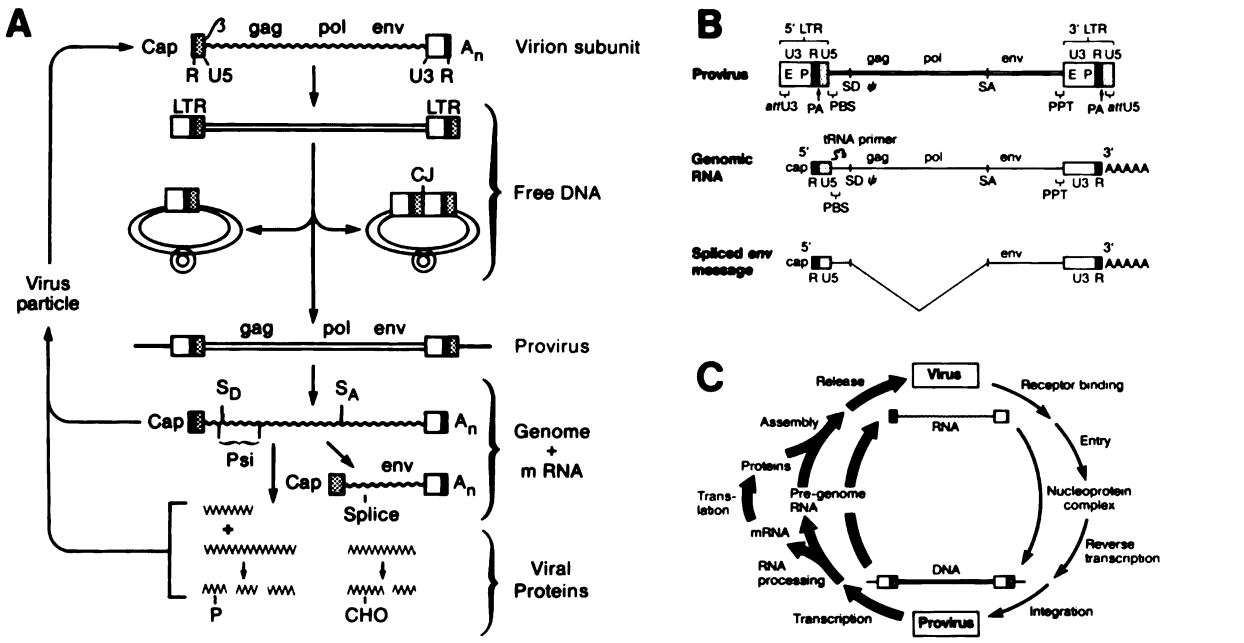


Figure 1. The Retrovirus Life Cycle.

A. The molecular transformations of the different forms of the viral genome during the life cycle are outlined. The unintegrated linear viral DNA is shown as the precursor to the provirus, and the circular forms are drawn as by-products that do not participate in the integration reaction. B. The proviral genome and the two forms of viral RNA, with cis-acting sequences and coding regions indicated. C. Two landmarks in the retrovirus life cycle; the extracellular virus and the chromosomal provirus. The early events in replication that convert the viral RNA genome into a DNA provirus are shown in thin lines to the right. The late events of viral gene expression and virion assembly and release are shown in thick lines (to emphasize that many particles are made from a single provirus) at the left. See the text for more detailed discussion. Cap, capped nucleotide at 5′ end of viral RNA; U3 and U5, unique sequences duplicated during DNA synthesis; R, a short repeated sequence at each end of the viral RNA that is required for template jumps during reverse transcription; LTR, long terminal repeat; CJ, circle junction, site of joining of ends of linear DNA; S_D and S_A, splice donor and acceptor sites, respectively; Psi and ψ, signal for packaging of viral RNA; P and CHO; modifications of viral proteins by phosphorylation and glycosylation, respectively; E and P, enhancer and promoter regions, respectively; PBS, primer binding site for the tRNA primer of first strand DNA synthesis; PPT; polypurine tract that serves as the primer of second strand DNA synthesis; *att*U3 and *att*U5, cis-acting sites at the ends of linear viral DNA that are required for integration. Adapted from Varmus and Brown 1989.

In contrast, the late events of replication function to amplify the expression of viral gene products. The amplification can be substantial, because, with important exceptions, retroviruses generally do not kill the cells they infect. In tissue culture, infected cells can produce high titers of infectious virus even after years of passaging. Each virus particle that is produced can be considered a preassembled replication machine that contains all the viral functions needed to establish a subsequent proviral template in another cell. A description of the composition and structure of the virion particle is therefore useful for any discussion of the viral structures that mediate early events inside infected cells.

Types of Retroviruses

Before discussing retroviral replication, it is important to realize that there are many different kinds of retroviruses, all of which fall into one of three general classes in terms of genome structure (Varmus and Brown 1989). The simplest class are replication-competent viruses, which package all of their gene products, expressed from *gag*, *pol*, and *env* genes, into the extracellular particle. The second class includes those retroviruses which have transduced cellular sequences. The transduced sequences usually replace viral coding sequences, and these viruses are dependent on helper virus for replication. The best known member of the third class of retroviruses is the human immunodeficiency virus, HIV. These viruses have more complex genomes that--in addition

to the products of *gag*, *pol*, and *env*--encode additiional non-structural proteins, most of which regulate viral gene expression and are not found in the extracellular virion (Varmus 1988b, Cullen and Greene 1989).

The Moloney murine leukemia virus (MLV), which I have studied for my thesis, is representative of the first class of simple, replication-competent retroviruses and has been the subject of many informative experiments concerning basic issues of retroviral replication. The MLV provirus (see Figure 1) consists of three genes, *gag*, *pol*, and *env*, which are flanked at each end by long terminal repeats (LTRs). The LTRs contain regulatory signals for the expression and processing of viral RNA. The *gag* gene encodes the major structural proteins of the virion core. The core is a nucleoprotein complex in which reside the viral RNA genome and the enzymatic activities encoded by *pol*. The virus core is surrounded by a lipid bilayer membrane. The lipids of the viral membrane are acquired from the host plasma membrane during budding, and its outer surface is studded with multiple copies of the *env*-encoded glycoprotein (see Figure 2). The MLV gene products and their arrangement in the virion particle are described in detail below.

I will also refer to other types of murine viruses, such as Friend and Rauscher MLVs; the avian sarcoma and leukosis viruses (ASLVs), which include Rous sarcoma virus (RSV) and the avian myeloblastosis virus (AMV); spleen necrosis virus (SNV), which, although it is an avian virus, is closely related to MLV; and the mouse mammary tumor virus (MMTV). Unless otherwise stated, all discussions refer to Moloney murine leukemia virus (MLV).

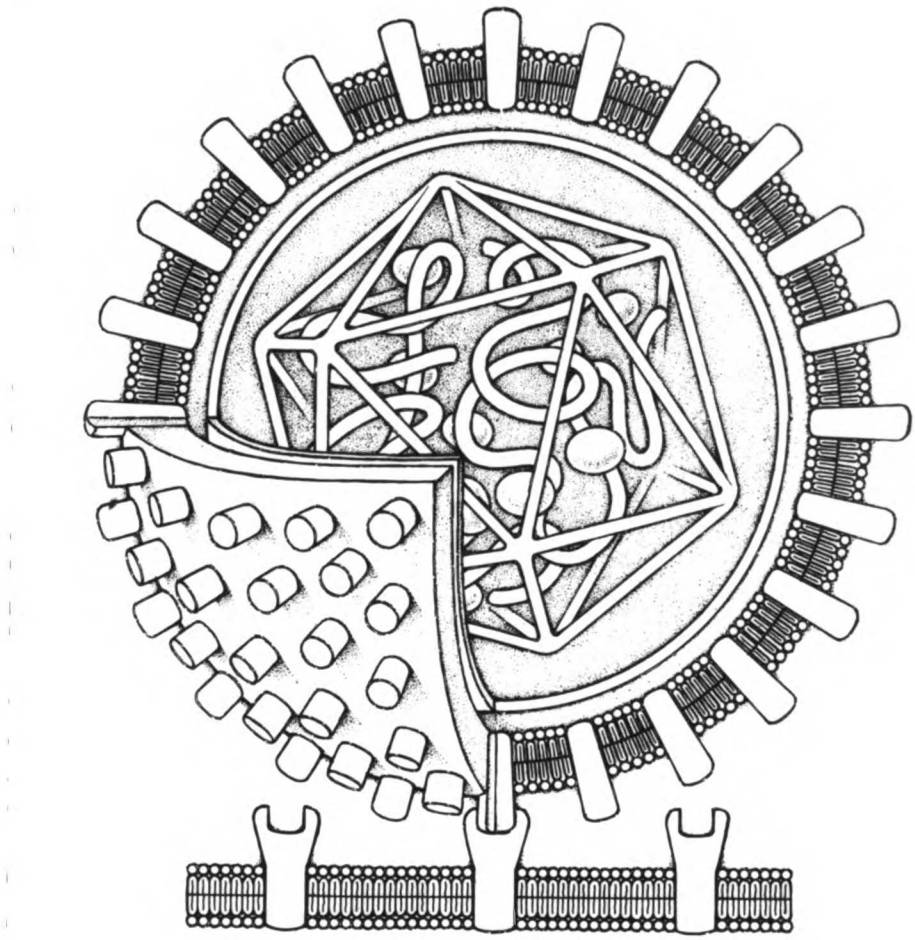


Figure 2. Schematic View of the Extracellular Retrovirus Particle.

Two identical strands of viral RNA, coated by the nucleocapsid protein, and viral enzymes (reverse transcriptase integrase, and protease) are drawn within the capsid-derived polyhedron shell; these components constitute the virion core. The core is surrounded by an envelope derived from host membranes and enriched with viral glycoproteins. Intermediate between the envelope and the core is a spherical shell representing the poorly defined matrix-derived structure that associates both with the lipid envelope and the polyhedron shell. The location of the remaining viral protein, p12, within the virus particle is not known. Reprinted with permission from the cover of *Advances in Oncology*, May, 1987 (artwork by Neil Hardy).

Viral Gene Expression

The products of *gag*, *pol*, and *env* are expressed from two mRNAs (see Figure 1). A spliced, subgenomic mRNA, in which most of the *gag* and *pol* sequences are removed as an intron, is used for expression of the viral *env* gene. In the mature virus particle, the envelope glycoprotein of MLV consists of a large, roughly 70 kD extracellular domain (SU) joined by disulfide bonds to a smaller, 15 kD transmembrane domain (TM) that is anchored in the lipid bilayer and includes a cytoplasmic tail that projects into the interior of the virus particle (Dickson et al 1982, 1985). SU and TM are translated together as a single polypeptide on membrane-bound polysomes. They are later cleaved into their mature forms after reaching the cell surface, probably by a host protease (Crawford and Goff 1985).

A single, genomic-sized transcript suffices for the expression of both *gag* and *pol*. In MLV, *gag* and *pol* are in the same reading frame but are separated by a stop codon. Roughly 95% of the time, the message is translated to produce the 65 kD *gag* polyprotein, which terminates as expected at the UAG codon separating *gag* and *pol*. However, about 5% of the time, the stop codon is informationally suppressed by a glutamine tRNA and a much larger 180 kD *gag-pol* fusion protein is made (Yoshinaka et al 1985). This approximately 20:1 ratio of *gag* to *pol* gene products persists in the composition of the extracellular virion core.

In many retroviruses, *gag* and *pol* are out of frame with respect to each other (the *pol* open reading frame being -1 with respect to the *gag*

open reading frame), and share a short region of overlap (which includes the *gag* stop codon). In these cases, expression of the *gag-pol* fusion protein requires a ribosomal frameshift within the overlap of the two reading frames, both to bypass the *gag* stop codon and to correct for the change in reading frame (Jacks and Varmus 1985; Jacks et al 1987; Jacks et al 1988). Termination suppression and frameshifting both produce *gag* and *gag-pol* proteins at ratios that reflect the efficiency of these unusual translational events.

If the *gag* and *gag-pol* polyproteins are expressed at altered ratios, virus assembly is very inefficient (Felsenstein and Goff 1988). Consequently, the ratio of *gag* to *gag-pol* is thought to be important for proper assembly and budding of infectious virus particles. Finally, although the mRNA for *gag-pol* is probably identical in sequence to the viral genomic RNA, virus-producing cells generate two distinct pools of *gag-pol* RNA, with the half-life of pregenomic RNA being shorter than that of the mRNA (Levin et al 1974; Levin and Rosenak 1976). Some investigators have claimed to find incoming genomic viral RNA associated with polysomes, suggesting that perhaps the two pools can exchange at some level (Shurtz et al 1979). The relevance of these findings to viral replication has, however, not been well documented.

Virus Assembly and Budding

Once viral proteins have been synthesized in proper and sufficient quantities, they must congregate in such a way as to form an infectious

particle. The first requirement for assembly is simply to bring all the components together. The *gag* and *pol* gene products achieve this largely through being synthesized as polyprotein precursors which share the same N-terminus (Dickson et al 1982, 1985; Schultz et al 1988). For MLV, these polyproteins (*Pr65*^{gag} and *Pr180*$^{gag\text{-}pol}$) are cotranslationally modified by the removal of the N-terminal methionine and the addition of myristate, a fatty acid, to the second *gag* amino acid, a glycine (Schultz and Oroszlan 1983). The presence of the hydrophobic myristyl moiety is required for these proteins to localize to the cytoplasmic face of the plasma membrane, and the presence of all the *gag* and *pol* products in these polyprotein precursors largely accounts for their colocalization in preparation for assembly (Rein et al 1986). Presumably these polypeptides aggregate due to self-assembly properties, accounting for their vast enrichment in virus particles. Noninfectious virus particles lacking either genomes (Linial et al 1978) or both reverse transcriptase and the *env* glycoproteins (Shields et al 1978) will bud from cells which express only the *gag* polyprotein, suggesting that the ability to assemble cores is intrinsic to this single polypeptide even in the absence of the other viral components.

Some viruses, such as MMTV and the Mason-Pfizer monkey virus, synthesize myristylated *gag* proteins but assemble their core structures in the cytoplasm instead of at the plasma membrane (Dickson et al 1982, 1985; Schultz et al 1988). Site-directed mutagenesis studies of the Mason-Pfizer monkey virus showed that intracellular cores form in the absence of myristylation, but the cores do not associate with plasma

membrane and bud; the acylation was not required for assembly of cores but was required for localization of the assembled cores to the plasma membrane (Rhee and Hunter 1987). The N-terminus of the RSV *gag* polyprotein has been shown by in vitro translation in a rabbit reticulocyte lysate to be acetylated, and the N-terminal *gag* protein purified from virions is blocked for Edman degradation (Palmiter et al 1978). However, the identity of the blocking group on the protein isolated from virions has not been determined.

Two other major viral components must join in the assembly of infectious virus particles. First, the *env*-encoded glycoproteins and the *gag* and *gag-pol* polyproteins must all become enriched in a common region of plasma membrane. While the forces that colocalize the envelope proteins with the viral structural proteins during assembly are not known, some type of recognition between the envelope proteins and the core proteins is presumed to occur. Second, the viral genomic RNA must be incorporated into the *gag* and *gag-pol* derived core. As mentioned earlier, retrovirus particles are diploid. Two copies of genomic RNA are incorporated into virions as a 70S RNA complex. If RNA purified from virions is denatured, the two viral strands can be separated from each other to generate 35S RNA. The possible mechanisms for incorporation of the viral RNA genome will be discussed below. The aggregated components undergo a poorly understood process of budding, in which the viral structural proteins organize into a core structure, which buds through the plasma membrane and thereby becomes enclosed by the glycoprotein-studded lipid bilayer.

cis-acting Components of the Viral Genome

The viral RNA genome, in addition to transmitting the coding information for viral proteins, also includes a variety of *cis*-acting sequences that are required for replication (Varmus and Swanstrom 1982, 1985; Coffin 1982, 1985; Varmus and Brown 1989; see Figure 1). Near the 5′ end of the genomic RNA is the primer binding site (PBS), which base pairs with a specific host tRNA that serves as the primer for first strand DNA synthesis in reverse transcription. Another site near the 3′ end of the viral RNA, the polypurine tract (PPT), serves as the primer for second strand DNA synthesis. This second primer molecule is derived from the viral RNA by RNAse H activity after the synthesis of first strand viral DNA. By base-pairing with its complementary sequence in the first strand DNA, a PPT-derived RNA fragment acts as a primer for second strand DNA synthesis. A short sequence, R, directly repeated at each end of the genomic RNA, is required for the template jumps that occur during reverse transcription. U5 sequences (between R and PBS at the 5′ end of the genome) and U3 sequences (between PPT and R at the 3′ end of the genome) are present only once in viral RNA, but reverse transcription results in their being present twice in the viral DNA, in the LTRs. The ends of unintegrated linear viral DNA consist of U3 sequences at the left end and U5 sequences at the right end. As described in greater detail below, these sequences at the ends of the viral DNA are required in *cis* for integration into the host genome. U3

also contains sequences characterstic of eukaryotic enhancers and promoters, which, in the provirus, direct the host transcriptional machinery in the synthesis of viral RNA. The border of R and U5 forms the viral polyadenylation signal. In addition, retroviral genomes include splice acceptor and donor signals. The synthesis of multiple viral transcripts from a single proviral template is an interesting example of differential gene expression regulated at the level of splicing. Finally, genomic viral RNA includes near the 5′ end poorly defined signals (ψ sequences) that are required for the highly selective packaging of viral RNA into virus particles.

The existence of transduced cellular genes in retroviruses, and the identification of important *cis*-acting sequences in replication, have led to the use of retroviruses as genetic vectors (Soriano et al 1989). Genes of interest can be cloned into proviral vectors from which all but the essential *cis*-acting sequences have been deleted. When introduced into "packaging" cell lines, recombinant viral genomes are incorporated into virus particles that are fully infectious for a single round of replication. In addition to aiding the study of viral replication, retroviral vectors are also useful for stably introducing foreign genes into tissue culture cells and even into developing embryos, and they may become clinically useful in gene therapy.

pol Components of the Virus Particle

Protease. The position of the protease (PR) coding sequences varies among retroviruses, although it is always present at the junction of *gag*

and *pol*. The first four amino acids of the MLV PR are encoded in *gag*; the fifth amino acid, glutamine, remains from suppression of the *gag* stop codon; and the rest of the protein is encoded in sequences at the 5' end of *pol* (Yoshinaka et al 1985). While the MLV PR straddles the reading frames of *gag* and *pol*, in MMTV, PR is encoded in a reading frame distinct from those for both *gag* and *pol*, and two ribosomal frameshifts are required to produce a *gag-pro-pol* fusion protein. In HIV, PR is encoded entirely in the *pol* frame, and, in RSV, it is mostly if not entirely encoded in the *gag* frame (Varmus and Brown 1989).

During or shortly after the process of budding, PR cleaves the *gag* and *gag-pol* polyproteins into their mature viral proteins. While the sites of cleavage vary considerably in sequence, proline is frequently at the amino terminus of the cleaved peptide, and phenylalanine or tyrosine is often at the carboxy terminus. All retroviral proteases share related amino acids, some of which are characteristic of aspartyl proteases, and they are inhibited by Pepstatin A, which is specific for aspartyl proteases (Dickson 1985; Varmus and Brown 1989). The structures of the HIV and RSV PR proteins have recently been determined by x-ray crystallography, and efforts are being made to find specific inhibitors that could be used to block viral replication (Miller et al 1989; Navia et al 1989). Electron microscopic studies of budding MLV virions showed that cleavage of the *gag* polyproteins is associated with a condensation of the core structure into a more electron dense configuration (Yoshinaka et al 1977). Viruses in which the protease has been inactivated by site-directed mutations do not undergo these

maturation events, and the virions they produce are not infectious (Crawford and Goff 1985; Katoh et al 1985).

Reverse transcriptase ($p80^{pol}$). The largest mature polypeptide encoded by MLV *pol* gene is the 80 kD reverse transcriptase (RT), which includes an N-terminal reverse transcriptase domain and a C-terminal RNAse H domain (Tanese and Goff 1988). SNV encodes a *pol* gene product of similar size, but the ASLVs are quite different in that RT exists as a dimer of two chains, called α and β. α is proteolytically processed at its C-terminus to produce the shorter β chain (Varmus and Swanstrom 1982, 1985). The two activities of reverse transcriptase and RNAse H are sufficient to reverse transcribe RNA, either endogenously in permeabilized virions, or exogenously with added synthetic templates and primers, into double stranded DNA. Analysis of replication intermediates isolated from infected cells and genetic studies in tissue culture have contributed to our understanding of reverse transcription, which is described in more detail below. This protein may also be responsible for the inclusion in virions of the host-derived tRNA primer that initiates reverse transcription, since mutants of RSV which lack RT do not incorporate the proper host tRNA into budded virions even though they package the viral genome (Sawyer and Hanafusa 1979). Similar experiments with MLV, however, do not show such a strong dependence on RT for packaging the host tRNA primer (Shields et al 1978).

Integration protein ($p46^{pol}$). The remaining 3′ sequences of the MLV *pol* gene encode the 46 kD integration protein (IN). SNV encodes an IN protein of similar size, while the ASLVs produce a smaller 32 kD

protein, which is derived from the C-terminus of the α chain (Varmus and Swanstrom 1982, 1985). The IN protein received its awkward name because even though genetic evidence and sequence similarities to transposases suggest that IN participates in recombination, such an activity has not been demonstrated biochemically. Furthermore, use of the term integrase implies a catalytic function; the protein need only act stoichiometrically during an infection, and demonstration of any catalytic potential awaits further biochemical analysis. Endonuclease activities have been demonstrated in vitro for ASLV IN proteins (both for p32 and for the α,β dimer), but not yet for MLV. The process of integration and the roles for IN are discussed in detail below.

gag Components of the Viral Particle

The viral protease cleaves the *gag* polyprotein into four mature proteins that form the structural components of the virion core. Roughly 1000 to 2000 copies of each of the *gag* proteins are present per virus particle, in contrast to the roughly 50 to 100 copies per particle of the *pol* proteins (Dickson et al 1982; 1985). The four mature *gag* proteins are found within the *gag* polyprotein precursor in roughly the same order in which they exist in the virus particle, with their order from N-terminus to C-terminus in the *gag* polyprotein (Matrix, p12, Capsid, and Nucleocapsid) corresponding to their position in the virus particle from outside to inside (see Figure 2).

Matrix (p15^{gag}). The most N-terminal MLV *gag* cleavage product is the 15 kD matrix protein (MA), which seems to reside between the virus

particle lipid bilayer and the capsid (CA)-derived polyhedron shell of the virion core. MA is the most hydrophobic *gag* product and as discussed earlier, is acylated at its N-terminus. (Dickson et al 1982, 1985; Schultz et al 1988). When purified from virions, MA stays in solution in the presence of 0.5% Triton X-100 but forms a precipitate when the detergent is removed by dialysis (Barbacid and Aaronson 1978). Gel-exclusion column chromatography in the presence of phosphatidyl choline liposomes showed that the Rauscher MLV MA binds to lipid bilayers (Barbacid and Aaronson 1978). The dependence of these properties on acylation has not been tested.

A close association of the Rauscher MLV MA with the virus envelope is also evident from its accessibility to labeling by lactoperoxidase, suggesting that MA may even penetrate the membrane in some way (Barbacid and Aaronson 1978). The identity of the MA protein was confirmed in these experiments by the use of a specific antiserum (which was important since the *env*-encoded TM is similar in size to MA), and the same results were obtained using chloramine-T labeling.

A close association with the membrane has also been shown for the avian MA protein. The ASLV p19 MA was cross-linked to the viral envelope lipids by treatment with dimethyl suberimidate, and the region of the protein that binds lipids was mapped to the most N-terminal of six cyanogen bromide cleavage products (Pepinsky and Vogt 1984).

Purification of AMV cores by density equilibrium centrifugation in sucrose gradients after disruption of virions with nonionic detergents resulted in the recovery of about 10% of MA, suggesting that it may also

associate with core proteins in addition to membrane lipids (Stromberg et al 1974). The purification of cores resulted in nearly complete recovery of RT, IN, and, NC, and roughly 25% recovery of CA, in addition to the smaller quantity of MA. Similar studies on Friend MLV showed that variable amounts of MA, from 10 to 20%, copurified with cores in equilibrium density gradients (Bolognesi et al 1973). Isolation of Rauscher MLV envelopes using a similar procedure (except the disrupted lipids were isolated instead of the cores) demonstrated that 40% of MA copurified with the viral membranes, along with all of SU and TM (Van de Ven 1978).

The RSV MA and TM proteins can be chemically cross-linked in intact virions, suggesting that MA may be involved in the enrichment of TM and SU in the region of the plasma membrane where budding occurs (Gebhardt et al 1984). RSV mutants lacking the cytoplasmic tail of TM assembled into fully infectious particles, arguing that any interactions between MA and TM probably occur with either transmembrane or extracellular domains of TM (Perez et al 1987). The ability to iodinate MA in virions suggests that such an association could occur. Affinity chromatography with purified proteins would perhaps be useful in identifying specific affinities of MA for individual viral proteins.

p12^{gag}. The next mature *gag* protein of MLV, p12, resides between MA and CA in the *gag* polyprotein, but its position and function within the extracellular virus particle are at present poorly defined. p12 does not purify with the cores from detergent-disrupted virions, but it has been shown to bind viral genomic RNA in a virus-specific manner (Sen et

al 1976). These binding studies were done primarily with Rauscher MLV and the simian sarcoma virus (SSV). p12 was purified and labeled with radioactive iodine and used for in vitro binding studies with ^{32}P-labeled genomic viral RNA. After incubation of p12 with viral RNA followed by formaldehyde cross-linking, 20% of the p12 was capable of binding either 35S or 70S viral RNA, as determined by sedimentation analysis in sucrose gradients. Furthermore, p12 would bind only its cognate viral RNA, and the binding was competed away only by p12 purified from the same virus (i.e. Rauscher MLV p12 would bind Rauscher MLV and not SSV viral RNA, and the binding was competed away by Rauscher MLV but not SSV p12). By saturating the binding of p12 to viral RNA and isolating chemically cross-linked molecules in equilibrium density gradients, the specific activity of the labeled protein and RNA suggested a binding stoichiometry of 4-5 p12 molecules per 35S RNA molecule. When intact virions were irradiated with UV light, a 12 kD viral protein became cross-linked to viral RNA, providing further evidence for an association.

A subpopulation of the p12 in virions is known to be phosphorylated on serine (Naso et al 1979; Dickson et al 1982, 1985). Purification of the phosphorylated and unphosphorylated forms of p12 from virions showed that only the unphosphorylated form would bind viral RNA in vitro (Sen et al 1977). When phosphorylated p12 was treated with phosphatase in vitro, it became active for binding, and when unphosphorylated p12 was kinased in vitro, it was no longer active for binding. Finally, UV irradiation of intact virions cross-linked a 12 kD protein that eluted

from a DEAE column at the same ionic strength as unphosphorylated p12. However, whether or not the fraction of p12 in virions that is not phosphorylated could be quantitatively cross-linked to viral RNA was not determined.

In the ASLVs, MA is a 19 kD protein, and the protein that is p12 in MLV is a 10 kD protein (Dickson et al 1982, 1985; Pepinsky and Vogt 1983). In virions, a subset of 19 kD avian MA is phosphorylated. The avian MA, like the murine p12, has been shown to bind its viral genomic RNA in a virus-specific manner. This specific binding activity was again supported by the demonstration that UV irradiation of intact virions cross-linked a roughly 19 kD protein to viral RNA (Sen and Todaro 1977). Cyanogen bromide mapping has shown that an N-terminal fragment of purified ASLV MA has an affinity for viral RNA in vitro. In contrast to the specificity demonstrated for the intact p19 MA (Sen and Todaro 1977), the binding activity of the N-terminal fragment was not specific in that MA protein from other viruses competed as well for the binding activity as did MA from the same virus (Johnson et al 1983). These observations have led to speculation that perhaps some of the functions of the MLV p12 are incorporated into the larger avian MA protein.

A trivial observation that might explain these results would be contamination of the purified p12 (or the avian p19 MA) with NC, which, as described below, can bind both single-stranded RNA and DNA. Several points argue against this possibility (Sen et al 1976, 1978; Davis et al 1976; Schulein et al 1978). p12 was well resolved from the 10 kD NC

during purification by gel filtration in guanidinium hydrochloride, and analysis by immunoprecipitation with an antiserum specific for p12 established that the purified protein was at least 90% pure. The saturation of p12 binding at 4-5 molecules per 35S RNA would not be expected if NC were present (see below), and neither would the specificity for the cognate viral RNA be expected for NC binding. Furthermore, purified NC flows through a DEAE column at 30mM KCl, while the purified p12 and the specific binding activity are retained by a DEAE column in 50mM KCl. p12 and NC also differ in the sensitivity of their nucleic acid binding activities to salt. p12 shows maximal binding from 50 to 150mM KCL, dropping to about 20% maximal levels at 250mM. NC is more sensitive to salt, showing optimal binding at 50-75mM KCl, 40% optimal binding at 100mM, and no binding at 150mM. Unfortunately, the presence or absence of NC was never directly determined in the studies of p12 or the avian p19 MA.

One curiosity in the studies of p12 is that UV irradiation of intact virions apparently did not cross-link NC to the viral RNA as might be expected, and the issue was not discussed. The identity of the cross-linked protein was determined only by its apparent size during gel electrophoresis. Similar experiments which took advantage of specific antisera, however, have shown that for RSV at least 90% of the protein that is cross-linked to viral RNA is not p19 MA but is the RSV NC (Meric et al 1984). Nevertheless, a substantial amount of signal was detected with the antiserum specific for p19, and no attempt was made to determine if NC contaminates purified preparations of p19 MA. The

confusion over the identity of the viral proteins involved in binding viral nucleic acids in these experiments emphasizes the need for studies of viral proteins purified after expression from independent recombinant vectors.

The potential specificity of MLV p12 and the avian p19 MA in binding only to cognate viral RNAs has encouraged speculation that perhaps some fraction of these molecules in a budding virus plays a role the determining the specificity with which viral genomic RNA is packaged into virus particles, even if most of the protein is absent in cores purified from virions. Others have speculated that the RSV p19 MA may have a role in RNA processing or even in reverse transcription (Leis 1978, 1981; Darlix and Spahr 1982), but conclusive data for any of these possibilities is lacking.

If p12 and the avian p19 MA are capable of binding their cognate viral RNAs, it should be possible to map the domains in the RNA that are recognized, especially since these proteins were shown to bind viral RNA that was first denatured (to convert the 70S RNA into 35S RNA). One might expect the ψ sequences that are required for packaging of genomic RNA into virion particles to be involved in such a recognition. Such experiments might be complicated by the finding that the binding activity activity for the RSV p19 MA was inactivated by pretreatment of the RNA with RNAse III, which recognizes double stranded regions of RNA (Leis et al 1978). Binding sites for the RSV p19 MA to viral RNA after UV cross-linking have been identified (Darlix and Spahr 1982), but the apparent predominance of NC in such cross-linked complexes makes these results uninformative.

Genetic analysis suggested that the MLV p12 may function in the very early steps of "uncoating" after a virus binds to and enters a susceptible cell, although the evidence was not strong (Crawford and Goff 1984). Cell lines that expressed mutant proviral genomes with deletions in the coding sequences for p12 generated fully assembled virus particles that were active for reverse transcription when assayed in vitro. However, when these mutant virions were used to infect new cells, they were blocked for replication at a very early step. Although the mutant viruses presumably could enter cells, they never initiated reverse transcription in vivo. Three such deletions were analyzed, only one of which was restricted entirely to p12; the other two extend into the C-terminus of MA. In all three cases, the *gag* polyprotein was improperly processed, and MA and p12 were fused (the other protease-recognition sites in the mutant *gag* polyproteins were properly cleaved). One cannot interpret these mutations as being specific for either MA or p12, and the presence of such a fusion protein might be expected to have relatively non-specific effects on infectivity. Mutants in p12 that disrupt viral replication but preserve the protease cleavage site between MA and and p12 have not been identified.

Capsid ($p30^{gag}$). Proceeding toward the C-terminus of the MLV *gag* polyprotein, the next mature product is the largest structural protein of the virion core, the 30 kD capsid (CA), which forms the polyhedron shell that encompasses the viral nucleocapsid (Dickson et al 1982, 1985). In vitro studies of CA, in which its molecular weight was determined by sedimentation of increasingly concentrated samples of the

purified protein, suggested that it was capable of self-association (Burnette et al 1976). When the molecular weight of CA was plotted against the starting concentration, the shape of the curve was most consistent with the behavior predicted for a molecule capable of forming an octamer. The observations that only a portion of CA copurified with virion cores in equilibrium density gradients (for AMV, RSV, and Friend MLV: Davis and Rueckert 1972; Bolognesi et al 1973; Stromberg et al 1974) is puzzling in that one might expect the structure of the CA-derived polyhedron shell to depend on the maintenance of interactions between the CA subunits. In recent studies of the equine infectious anemia virus, CA was quantitatively recovered with cores; sedimentation in glycerol gradients instead of density equilibrium centrifugation in sucrose gradients was used as a means of purification (Roberts and Oroszlan 1989).

The idea that self-assembly of CA is important to the assembly of budding viruses is supported by genetic studies, since mutations in CA typically disrupt the ability of the virus to assemble (Lobel and Goff 1984; Goff and Lobel 1987). Three mutant virus constructs with in-frame deletion mutations and three with missense mutations in CA all were incapable of assembling and budding virus particles (Schwartzberg et al 1984; Hsu et al 1985). It would be interesting to test such mutants in vitro for self-assembly properties.

These mutants also shared another defect in that they were incapable of delivering the *gag-pol* polyprotein into virions when they were introduced into a cell line that expressed wild-type *gag* but not *gag-*

pol. One mutant construct with two missense mutations in CA is interesting in that it was capable of assembling into virions and could provide *gag-pol* to virions produced from the cell line expressing wild type *gag*, but the virions were blocked at an early step in infection (Hsu et al 1985). Presumably, the mutant particles could enter cells but were incapable synthesizing viral DNA. However, as with the deletion mutations in MA and p12 described earlier (which had an identical phenotype), the ability of the mutant viruses to enter susceptible cells was not demonstrated.

An interesting feature of CA is its involvement in the inability of certain strains of MLV, called N-type and B-type viruses, to replicate in cell lines derived from particular inbred strains of mice, a phenomenon termed *Fv-1* restriction (Jolicoeur 1979). In restricted infections, virions can enter the cell and synthesize viral DNA but are incapable of establishing proviruses. N- and B- tropism map to CA; if the coding sequences for CA are interchanged beween an N-type and a B-type virus, their tropism is reversed (DesGroseillers and Jolicoeur 1983). CA may somehow interact with components of the host cell which, in the appropriate genetic context, can inhibit infection. Affinity column chromatography might help identify cellular proteins that recognize type-specific CA proteins.

Nucleocapsid ($p10^{gag}$). The predominant components of the nucleocapsid enclosed by the CA-derived polyhedron shell are the diploid, single-stranded RNA genome, and the most C-terminally located *gag* protein, the 10 kD (for MLV) nucleocapsid (NC) (Dickson et al 1982, 1985). NC is the

most basic retroviral protein and has both single-stranded and double-stranded nucleic acid binding activities in vitro, as shown by filter-binding studies that were done with NC purified from Rauscher MLV (Davis et al 1976). A preference for single-stranded nucleic acids is apparent from the observation that 15-30 times more double-stranded DNA than single-stranded DNA was required to a reduce binding activity by 50%. These studies suggested that NC is relatively nonspecific in its binding activities since no preference was observed for viral RNA relative to single-stranded calf thymus DNA or Qβ single-stranded RNA. Saturation binding studies with radioactively labeled NC suggested a stoichiometry of about 100 molecules of NC per 35S RNA subunit (Schulein et al 1978), but this number does not account for the 100-2000 copies of NC present in a virion. Other single-stranded nucleic acids, including Qβ RNA and calf thymus DNA, were capable of binding only 4-5 NC molecules at saturation, although single-stranded bacteriophage fd DNA was bound at comparable levels. Heterologous retroviral RNAs were not tested.

Nucleocapsids can be purified from virion cores by using higher levels of nonionic detergents during density equilibrium centrifugation, which results in the loss of CA but recovery of NC and the viral RNA (Dickson et al 1982, 1985). Nucleocapsids isolated from virion cores have a density of 1.34 g/ml in cesium gradients. After cross-linking NC bound to viral RNA in vitro, the density of the reconstructed nucleocapsids was 1.55g/ml. The higher density is consistent with fewer NC molecules being present in the reconstructed complexes than in nucleocapsids purified from virions (Schulein et al 1978).

The ASLV NC, a 12 kD protein, has also been shown to bind single-stranded nucleic acids. Studies in which changes in fluorescence were measured after binding of NC to either polyethenoadenylic acid or polyuracilic acid, established that one NC molecule is bound for every six nucleotides, which is roughly consistent with the number of NC molecules present in a virus particle if all of them are used to completely coat two 35S viral RNA strands (Karpel et al 1987). The authors suggested that a difference in salt conditions might account for the different stoichiometry obtained in the studies of Rauscher MLV (100 NC molecules per 35S RNA).

Several mutations in NC have been described. The NC proteins of retroviruses all have either one (MLV) or two (RSV) copies of a sequence with characteristically spaced cysteine and histidine residues. In one study, five different mutants of MLV, with single missense mutations introduced into the codons for these amino acids, were made by site-directed mutagenesis. They were all capable of producing extracellular virions that had wild type quantities of CA protein and roughly wild type levels of reverse transcriptase activity when assayed in vitro on exogenous substrates. However, all the mutants were noninfectious and lacked viral RNA, although they contained significant quantities of cellular RNA. The defect was shown to be due to the altered protein and not to changes in the RNA sequence (Gorelick et al 1988). Surprisingly a deletion that removed the N-terminal 32 amino acids of the MLV NC had very little effect on virus production or infectivity (Schwartzberg et al 1984), even though the deletion removed two of the conserved

cysteines that both had phenotypes when they were mutated individually (Gorelick et al).

In RSV, three mutations in the region of the protein that contains two copies of the same pattern of cysteine and histidine residues had slightly different phenotypes (Meric and Spahr 1986). A two amino acid insertion led to the production of virions that contained viral RNA. The RNA appeared to be less stably dimerized than in wild type virions in that more of it migrated as 35S RNA during nondenaturing gel electrophoresis. A mutant with a 39 amino acid deletion that removed one copy of the cysteine/histidine sequence contained no viral RNA, while a duplication of the sequence contained normal amounts of viral RNA that comigrated with wild type 70S viral RNA during non-denaturing gel electrophoresis. As with MLV, in all three cases, the mutant viruses were active for reverse transcription in vitro but were noninfectious, even though two of them contained viral RNA. The authors suggested that in addition to having effects on viral RNA packaging, alterations in NC can affect the ability of the virus to activate reverse transcription upon entry into a cell.

The ASLV NC is phosphorylated on serine, and the phosphorylation is required for in vitro binding activity. When NC purified from virions is phosphatased, the binding activity is greatly reduced, and kinase treatment of the phosphatased protein restores the activity (Leis and Jentoft 1983; Leis et al 1984). Biochemical studies have also shown that two adjacent lysine residues in the avian NC are protected from chemical modification when the protein is bound to viral RNA. When the

two adjacent lysine residues were both changed to isoleucines, virus with the double mutation were noninfectious and lacked viral RNA. Moreover, when the mutant protein was expressed in *E. coli*, phosphorylation of the protein did not activate binding to viral RNA, in contrast to the activation seen when wild type protein expressed in *E. coli* was phosphorylated (Fu et al 1988a). Finally, when the serine residue in the RSV NC that is phosphorylated in vivo was changed to alanine, fully infectious virions were produced. The altered protein was constitutively active for binding--even though other amino acids in the protein were phosphorylated in vivo, phosphatase treatment, which eliminates the in vitro binding activity of the wild type protein, did not affect the activity of the serine to alanine mutant (Fu et al 1988b). Taken together, these mutant indicate that a variety of different residues in the protein are important for its ability to bind viral RNA in a manner that leads to the production of infectious virions.

Consistent with its in vitro binding activity and the behavior of the mutant proteins, wild type NC coats the single-stranded genomic RNA in virion particles. The nucleocapsid also includes the smaller quantites of RT and IN (Dickson et al 1982, 1985). A highly schematic diagram summarizing the position of the various mature *gag* and *pol* polypeptides in the mature virus particle is shown in Figure 2.

Entry Into a Host Cell

To begin an infection, the viral particle awaits only the recognition of a susceptible cell, through binding of the *env* glycoprotein to an

appropriate cell surface receptor molecule. Binding to the host receptor initiates a series of events culminating in membrane fusion and release of the virion core into the cytoplasm of the infected cell. For some retroviruses, entry into the cytoplasm occurs via a direct fusion to the plasma membrane (Stein et al 1987; McClure et al 1988), while other retroviruses appear to enter via an endocytic, low pH pathway (Redmond et al 1984). The host receptor molecules have been identified only for HIV (McDougal et al 1986; Maddon et al 1986) and more recently for ecotropic strains of MLV (Albritton et al 1989).

Very little is known in terms of exactly what components of the extracellular virion core remain associated with the particle that enters the host cell cytoplasm, or what, if any, changes in the structure of the core must occur for the viral enzymatic functions to become active. However it happens, the viral replication machine is said to be activated by the process of entry into the cytoplasm, meaning that some derivative of the extracellular virus particle proceeds to complete the remaining early events of reverse transcription, the gaining of access to the genome, and the culminating step of integration of the viral DNA into a host chromosome.

Reverse Transcription

Reverse transcription, in brief, consists of two priming events, one for the first (minus) strand and one for the second (plus) strand of DNA. Both priming events occur near the 5′ ends of their respective

templates and consequently generate strong stop products that must perform template jumps to copy the full length of the genome. The *cis*-acting sequences that are required for these events were described above. A great deal of experimental effort went into defining the complex series of steps that convert the RNA genome into double stranded DNA with terminal redundancies (Varmus 1983, 1987; Varmus and Swanstrom 1982, 1985; Varmus and Brown 1989). The early steps were first worked out both by the use of purified components in reconstituted reactions and by analyzing the products of the endogenous polymerase activity in purified virions (which can be activated by the addition of nonionic detergents and deoxynucleoside triphosphates). Analysis by solution hybridization of intermediates isolated from acutely infected cells was invaluable for working out many of the steps of reverse transcription, both because of the difficulty in reproducing the entire process in vitro and because the analysis of intermediates derived from in vivo reactions provided assurance that the structures were authentic. Predictions based on the analysis of intermediates isolated from infected cells were subsequently verified by improvements in the in vitro reactions and by the eventual cloning and sequencing of the DNA products of reverse transcription.

Viral recombination is thought to occur during reverse transcription, allowing for the generation of novel viral genomes and for the transduction of cellular genes. Viral recombination might conceivably occur at either of two stages (Hunter 1978; Coffin 1979). RT could switch templates between the two different RNA strands in the

heterozygous particle during first strand synthesis (copy choice), or two more fully replicated genomes might undergo strand exchange after partial second strand synthesis (strand invasion). These possibilities could be addressed more thoroughly with the use of recombinant viruses that carry selectable markers in their genomes and are capable of only a single round of replication.

Structure of Unintegrated Viral DNA

The principal final product of reverse transcription is the full-length linear, double-stranded DNA molecule found initially in the cytoplasm of acutely infected cells. This linear DNA product, for MLV, is about 9 kb in length and has at each end 0.6 kb long LTRs (Shinnick et al 1981). The LTRs are made up of three sequences called U3, R, and U5 (from from right to left; see Figure 1). The MLV LTRs do not encode any proteins but rather serve to regulate transcription initiation and termination. The rest of the viral sequences, most of which do code for viral proteins, are colinear in the genomic RNA and proviral DNA.

The ends of the LTRs are predicted by models of reverse transcription to be blunt-ended and to terminate at positions defined precisely by the position of first and second strand primers. The single example of linear, unintegrated viral DNA that has been cloned and sequenced, from an infection by the gibbon ape leukemia virus, conformed to these predictions (Scott et al 1981), but analyses of entire populations of linear DNA molecules have until recently been lacking (see Chapter 2).

The linear form of viral DNA is synthesized in the cytoplasm (Varmus et al 1974), although the possibility exists that some retroviruses may synthesize DNA in the nucleus. The linear DNA later begins to appear in the nucleus as well, where one also sees for the first time, at detectable levels, two different kinds of circular viral DNA, called the one LTR and two LTR circles (Shank and Varmus 1978; Shank et al 1978; Varmus and Brown 1989). The one LTR circles appear to be the product of homologous recombination between LTRs at the ends of linear DNA, and are the predominant form of circular DNA seen in infected cells. The two LTR circles in at least some cases appear to be the product of simply ligating together the blunt ends of the linear viral DNA. Bringing together the two ends of the LTRs joins the small inverted repeats present at the ends of the LTRs to form what is called the circle junction. Many two LTR circles, however, have small deletions or insertions at the junction of the ends of linear DNA, and some of them have internal inversions as a consequence of autointegration events (Shoemaker et al 1980; Varmus 1983; Varmus and Brown 1989). As an infection progresses, increasing amounts of viral DNA become covalently associated with the cellular chromosomes as proviral DNA.

Retroviral Integration; The Provirus Hypothesis

The proposal that retroviral genomes, in spite of existing as RNA in the extracellular virus particle, pass through a DNA form somehow linked to the host chromosome, was originally inspired by two observations

(Temin 1971). First, when tissue culture cells were infected by transforming retroviruses, the extremely stable transmission of the transformed state of the host cell during subsequent cell divisions argued indirectly for a close association of the viral and host genomes (Temin 1959, 1961). More important, inhibitors, such as actinomycin D, which are known to interfere with DNA mediated processes, abolished the production of retroviruses from infected cells (Temin 1964). This inhibition was in striking contrast to the ability of other RNA viruses to replicate in the presence of such compounds and strongly argued that retroviruses went through a DNA intermediate. However, the lack of precedent for any biological activities that converted RNA into DNA was an obstacle to the acceptance of such an hypothesis.

Pioneering work in the late 1960s and early 1970s led to the discovery of reverse transcription in purified retrovirus particles and defined a mechanism for the synthesis of a DNA form of the retroviral genome (Baltimore 1970; Temin and Mizutani 1970). Hybridization studies showed that viral sequences were present in cellular DNA (Baluda and Nayak 1970; Varmus et al 1973), and the ability to transfer infectious virus into new cells by the transfection of chromosomal DNA from infected cells convincingly demonstrated that viral DNA could become stably associated with the host genome (Hill and Hillova 1972). These experiments established that retroviruses, after entry into cells, reverse transcribe their RNA genome into DNA, and establish a provirus which can passively reside in the host genome and provide a template for the expression of progeny virus. While the mechanism of provirus

establishment remained ambiguous, the requirement for some type of strand-exchange reaction was apparent.

Lessons from the Provirus

Molecular analysis of retroviral infections in tissue culture revealed that retroviral integration was both efficient and specific in that every cell in an infected population readily acquired proviruses which all shared certain characteristic features (Varmus and Swanstrom 1982, 1985; Varmus 1983; Varmus and Brown 1989). Although the sites of integration in the host genome varied considerably, sequence analysis of proviral DNA revealed informative consistencies. The ends of the proviral DNA always lack two base pairs that are present at the ends of full-length blunt-ended linear viral DNA, and the host target sequences are consistently duplicated as short direct repeats of 4 to 6 base pairs flanking the viral DNA. The size of the host repeat depends on the type of retrovirus; MLV generates a 4 base pair repeat of the target DNA sequences, RSV a 6 base pair repeat. Unlike the occasional examples of imprecise integration events seen for some animal viruses, retroviruses with high efficiency insert their viral DNA into the host chromosome in a very precise fashion.

Target site selection shows no known specificity in terms of sequence, but several lines of evidence suggest that some sites and regions in the host genome may be highly preferred. For example, integration of MLV into the hypoxanthine phosphoribosyltransferase gene

in mouse cells occured about 100-fold less frequently than would be expected if integration were entirely random (King et al 1975). In contrast, integration events were shown to occur much more frequently than expected near DNase I-hypersensitive sites and in transcribed regions of the genome (Vijaya et al 1986; Rohdewohld 1987). The effects of transcription on chromatin structure have been proposed to influence the frequency of integration into particular regions of the host genome by perhaps making DNA more accessible to the integration machinery. Retroviral DNA has been shown to integrate predominantly into DNA that has replicated during an infection--perhaps replication also makes the chromosomal DNA more accessible for integration (Varmus et al 1977). Finally, cloning of multiple integration sites for Rous sarcoma virus has revealed the presence of a large class of sites that are chosen to the nucleotide with high frequency (as high as 1 in 3000 integration events, compared to the 1 in 10^9 expected for a random event: Shih et al 1988). The two sites that have been sequenced show no similarity to each other, and the reasons for their frequent use remain obscure; general effects on chromatin structure might not be expected to result in such precise use of a given sequence as a target site.

The presence of short, direct repeats of host sequences flanking the provirus implies that the host target is cleaved in a staggered fashion, and the loss of two base pairs from the tips of the viral LTRs implies that viral DNA is also cleaved during integration. The observation that the size of the host repeat is specific for the type of virus, independent of the host cell type, provides tantalizing evidence that a

virus-specific function participates in the integration reaction. Additional evidence for the participation of virally encoded functions came from the inability to reproduce either the efficiency or specificity of integration by simply introducing purified retroviral DNA into cells by microinjection or transfection (Copeland et al 1981; Luciw et al 1983; Kriegler and Botchan 1983). Such artificially introduced viral DNA integrates much less efficiently, shows no specificity in terms of the position of attachment of viral DNA, and does not cause a duplication of the host target sequences.

Genetic Analyses of Retroviral Integration

The suggestions that virally encoded activities participate in retroviral integration prompted a genetic search for such a function. Attention was focused on the *pol* gene because of studies demonstrating that *pol* products purified from ASLV virions possess both DNA binding activity and endonuclease activity. The p32 IN protein of AMV was shown to cleave supercoiled DNA that contained a viral DNA fragment encompassing the circle junction (formed by joining the two ends of linear viral DNA to make the two LTR circle). Some of the cleavages were at sites two base pairs removed from the end of each LTR, which would generate a four base pair staggered cleavage consistent with the loss of two base pairs from the end of each LTR in the integrated provirus (Grandgenett and Vora 1985; Gandgenett et al 1986). However, the protein also cleaved at many other sites in the plasmid DNA. A more

specific activity was demonstrated for the *pol*-encoded α,β dimer of ASLV, but the cleavages occurred one base pair further from the end of each LTR than would be predicted, and the activity showed an unexplained preference for single-stranded DNA (Duyk et al 1985).

The use of site-directed mutagenesis, first with MLV and later with SNV, subsequently revealed that the 3' portion of *pol* is indeed required for proper integration (Schwartzberg et al 1984; Donehower and Varmus 1984; Panganiban and Temin 1984). Viruses with mutations in the sequences encoding IN were capable of entering cells and reverse transcribing their genomes into DNA, but the efficiency with which they integrated was reduced in some cases by more than two orders of magnitude. Furthermore, on the rare occasions that viral DNA from such mutant viruses did become covalently linked to the host genome, restriction mapping and DNA sequence analysis showed that these events occured imprecisely (Hagino-Yamagishi et al 1987).

The mutations that inactivate IN are fairly limited in number and many of them disrupt substantial portions of the coding sequences. For MLV, four frameshift mutations, two small deletions, and one single amino acid substitution mutation (at a highly conserved position) were initially described. In each case, viruses with these mutations produced virions that were active for reverse transcription in vitro, but they could not productively infect cells. A single frameshift mutation was described and analyzed for SNV. The potential for multiple IN activities--viral and host DNA binding, cleavage of viral and host DNA, and strand-exchange activity--suggests that the protein may have

multiple domains involved in establishing the provirus. Also, the palindrome structure formed by bringing together the ends of linear viral DNA suggest that IN may act as a dimer, which would presumably require additional domains. More recently, a series of twelve different two and three amino acid insertions in the MLV IN protein were described (Donehower 1988). Ten of these abolished integration activity without interfering with reverse transcriptase activity, while two of them located at the very C-terminus of IN did not interfere with replication at all. The construction of additional subtle mutations, and more extensive analysis of different activities using in vitro assays, should improve our understanding of the function of this protein.

Genetic analyses have not yet demonstrated a role in integration for any other viral proteins. However, this negative evidence is qualified by the late position of integration in the early events of retroviral replication. Mutations in *gag* generally have phenotypes reflecting an inability to assembly virus particles properly, and RT mutants never generate the viral DNA substrate for integration.

The Viral *att* Site

Another class of mutant viruses that are defective for integration have genetic lesions in sequences present at the ends of the unintegrated linear viral DNA (in U5 or U3). Detailed mutational analyses of these sequences have shown that between 6 and 12 base pairs

at the tip of each LTR are required in *cis* for proper integration. As with IN, mutations in these *att* sites did not affect the assembly of extracellular virus particles or the ability of the virus to reverse transcribe its genome upon entry into a susceptible cell, but they did severely reduce the efficiency of integration (Panganiban and Temin 1984; Colicelli and Goff 1985, 1988; Cobrinik et al 1987).

Although the *att* sites of different retroviruses show very little sequence similarity (outside of the completely conserved TG and CA found at the 5′ and 3′ termini of proviral DNA), these sequences are similar in that for most retroviruses they form inverted repeats of various lengths (Varmus and Brown 1989). For MLV, the repeats form a perfect 26 base pair palindrome when the two ends of linear viral DNA are joined, while for HIV and HTLV-II, the palindromes are only 6 and 4 base pairs long, respectively. The ends of SNV form a perfect 12 base pair palindrome, and the ends of RSV form a 30 base pair palindrome with three mismatches.

Deletion analysis of the SNV LTRs showed that the terminal 5 to 12 base pairs of U3 (the end of the left LTR) and the terminal 6 to 8 base pairs of U5 (the end of the right LTR) were required for *att* site function (Panganiban and Temin 1984). In ASLV, deletion experiments showed that from 17 to 23 base pairs of U5 were required for replication (presumably for integration), but the U3 sequences were not analyzed (Cobrinik et al 1987).

The most extensive *att* site studies have been done with MLV (Colicelli and Goff 1985, 1988a, 1988b). The description of an MLV provirus in

which the right LTR was inverted provides a striking demonstration of the sufficiency of the inverted repeat for *att* site function, since the virus presumably integrated using only U3 sequences present at each end of the viral DNA (Colicelli and Goff 1988a). The most terminal 11 to 12 base pairs of the inverted repeat are sufficient for integration--a deletion that removed the most internal base pair of the 13 base pair repeat of U5 was tolerated, but removal of three base pairs abolished integration (J. Murphy and S. P. Goff, personal communication). A one base pair deletion at the terminus of U5, and a single transversion at the second base pair from the end of U5 did not affect integration, while deletion of 2 or 8 base pairs from the end of U5 eliminated integration. The insertion of two base pairs at the end of U5 did not affect replication significantly, but insertion of 10 base pairs led to a substantial delay in the appearance of progeny virus. Sequence analysis of proviruses (or two LTR circle autointegrants) generated by viruses carrying *att* site mutations that were tolerated revealed that the integration machinery made the proper cleavages in spite of the sequence changes. For example, insertion of two base pairs at the end of U5 led to a integrant in which six instead of four base pairs had been removed as a consequence of integration, and deletion of one base pair at the end of U5 led to a integrant which had lost only three base pairs (in other words, in both cases, the recombinants ended at the conserved CA, in spite of the altered numbers of base pairs present at the end of U5 beyond the CA dinucleotide in the unintegrated viral DNA). In both cases, 4 base pairs of the target sequence were duplicated as in

wild-type infections. The size of the host repeat in other integrants could be interpreted in more than one way, due to coincidences of sequence at the target sites. The finding that small deletions and insertions in the terminal two base pairs of U5 were tolerated while deletions of more internal sequences blocked replication has led to the proposal that the integration machinery recognizes the more internal sequences of the inverted repeats and "reaches out" to make cleavages near the termini of the viral DNA.

Sequence analysis of pseudorevertants derived by continued passage of cells infected by MLV *att* site mutants provided additional cases of altered *att* sites that were capable of efficient integration (Colicelli and Goff 1985, 1988b). The virus with a two base pair deletion at the end of U5 reverted to a virus with only one base pair missing from the end of U5, as did the virus that originally had a ten base pair insertion (this revertant, in addition to the one base pair deletion, also had a transversion of T to G at the terminal base pair). The virus with an 8 base pair deletion at the end of U5 reverted to the wild type sequence. Most interesting was a pseudorevertant of the ten base pair insertion mutant that underwent a single base change in the inserted sequence that allowed recombinants to form efficiently at a new site. An integrant from this pseudorevertant had lost three base pairs from the edges of the LTRs, but retained 9 of the 10 additional base pairs at the end of U5, with the breakpoint just 3' of a CA dinucleotide in the inserted sequence. This revertant was indistinguishable from wild-type in its ability to replicate. The sequence of the insertion matched the

repeat at the end of wild type U5 at 6 of 12 positions (when the two sequences were slightly staggered to optimize the comparison), suggesting that perhaps the integration machinery can function on a largely divergent sequence. However, the single base change did not improve the match. It did disrupt a perfect symmetry present in the inserted sequence, though, and the authors suggested that perhaps the change prevented the formation of an inhibitory hairpin structure. Additional mutations are needed to determine if such mismatches can be tolerated in the MLV *att* site.

The Circle Junction Experiment

The observations that a palindrome is formed by joining the ends of linear viral DNA to form a circle junction in the two LTR circle, and that the two LTR circle is seen only in the nucleus (where integration presumably occurs), encouraged efforts to implicate the two LTR circle as the viral DNA substrate that directly participates in the retroviral integration reaction. Convincing experimental evidence for the two LTR circlular form of viral DNA being the precursor to the integrated provirus came from studies of SNV (Panganiban and Temin 1985). A recombinant virus was engineered to contain an introduced circle junction sequence (about 50 base pairs in size), placed internally in the viral genome. Reverse transcription of this recombinant virus generates viral DNA that, in the form of a two LTR circle, contains two circle junctions--the one formed by ligating together the ends of the

LTRs, and the one introduced at an internal site. With the assistance of wild type helper virus, double-*att* site recombinants were observed to integrate with equal efficiency at the both the normal and the introduced circle junction sequences, suggesting that the two LTR circle directly participates in the integration reaction. Efficient use of the internal circle junction occurred with two different viruses in which the internal sites were in different positions and opposite orientations. One inconsistency in the behavior of the recombinant viruses was that no proviruses were found in which a one LTR circle had served as a precursor by virtue of containing the internal circle junction (which should be present in all forms of the unintegrated viral DNA). This deficiency was not discussed, but one could imagine that one LTR circles exist in a different pool and are not active for integration--for instance, the availability of the LTRs for homologous recombination could be a consequence of changes that also disrupt the integration machinery.

Recent attempts to repeat these results by placing circle junction sequences ectopically in MLV have failed in spite of elaborate efforts to detect such events (Lobel et al 1989; Ellis and Bernstein 1989). In the most thorough attempt, two different viral constructs were made, one with a 220 base pair circle junction insert and the other with a 60 base pair insert present at a different position (Lobel et al 1989). In the latter construct, the normal att site was mutated to favor use of the internal wild-type sequence. In both cases, the recombinant viruses also contained genes encoding selectable drug-resistance markers to

facilitate the recovery of integrated proviruses. In spite of these efforts, none of the proviruses that were analyzed had used the internal circle junction, even when the normal *att* site had been mutated. One could argue that the internal circle junctions had been placed in unfavorable contexts, but the efficient use of two different internal sites in SNV makes this seem unlikely. Given the wide variation in the sequence and structures of different viral *att* sites, it is perhaps not surprising that more than one mechanism for integration might be operative. The issue of which form of unintegrated viral DNA serves as the precursor to integration for MLV will be discussed in more detail in Chapter 2.

In brief, we know that upon entry into susceptible cells, retroviruses convert their RNA genome into viral DNA using RT, and they join the very ends of the viral DNA to the host chromosome in a strand-exchange reaction that requires the IN protein. The sequences at the ends of the LTRs are required in *cis* for integration, and the viral DNA precursor to the integrated provirus, for SNV, appears to be the 2 LTR circle.

Summary

Our knowledge of retroviral replication has traditionally come from genetic analyses of viral gene products and from the identification of the nucleic acids intermediates involved in reverse transcription and integration. These two types of analysis, while informative, suffer

from the common fault that they essentially treat the infected cell as a black box--we know what nucleic acid intermediates are generated during viral replication, and what viral gene products are required to make these intermediates, but we know very little about how these different components fit together in their native state inside infected cells. While some work has been done to define the structural features of the extracellular virus particle, very little is known about the structures inside infected cells that contain the viral nucleic acid intermediates involved in replication. As described in the remaining chapters, we have been able to isolate structures from acutely infected cells that contain unintegrated viral DNA and are active for integration in an in vitro assay. By looking inside infected cells, we have gained fundamental insights into the process of integration and have begun to understand how the different viral gene products are assembled at different stages in the retrovirus life cycle.

The following chapter will describe the collaboration I have had with Pat Brown in which we developed an assay that accurately reproduces the retroviral integration reaction in vitro. Our results using the assay suggested and eventually proved that, for MLV, the immediate precursor to retroviral integration is not the 2 LTR circular form of viral DNA, but rather is the linear form of viral DNA. This second chapter begins with a narrative that summarizes our collaboration and concludes with two published papers describing the details of that work. The third chapter, also a published paper, describes my results demonstrating that in acutely infected cells, unintegrated retroviral DNA resides within a

large nucleoprotein complex. This "integration machine" includes all the activities required for integration and appears to be a derivative of the extracellular virion core. The fourth chapter, a paper in preparation, describes my efforts to directly identify the viral protein components that make up the viral DNA-containing nucleoprotein complexes found in the cytoplasm of acutely infected cells. The final chapter summarizes these findings and their relevance to our current understanding of retrovirus replication and genetic recombination.

REFERENCES

Alberts, B., D. Bray, J. Lewis, M. Raff, K. Roberts, J. D. Watson. 1989. The evolution of the cell, pp. 3-40. In **The Molecular Biology of the Cell**, second edition. Garland Publishing, Inc., New York.

Albritton, L. M., L. Tseng. D. Scadden, and J. M. Cunningham. 1989. A putative murine ecotropic retrovirus receptor gene encodes a multiple membrane-spanning protein and confers susceptibility to virus infection. *Cell* 57:659-666.

Baltimore, D. 1970. RNA-dependent DNA polymerase in virions of RNA tumor viruses. *Nature* 226:1209-1211.

Baluda, M. and D. P. Nayak. 1970. DNA complementary to the RNA genome of avian leukosis viruses. *Proc. Natl. Acad. Sci.* 69:576-580.

Barbacid M. and S. A. Aaronson. 1978. Membrane properties of the *gag* gene-coded p15 protein of mouse type-C RNA tumor viruses. *J. Biol. Chem.* 253:1408-1414.

Bender, W., Y.-H. Chien, S. Chattopaddhyay, P. K. Vogt, M. B. Gardner, and Davidson, N. 1978. High molecular weight RNAs of AKR, NZB, and wild mouse viruses and avian reticuloendotheliosis virus all have similar dimer structures. *J. Virol.* 25:888-896.

Bender, W. and N. Davidson. 1976. Mapping of poly(A) sequences in the electron microscope reveals unusual structure of type C oncornavirus RNA molecules. *Cell* 7:595-607.

Bingham, P. M. and Z. Zachar. 1989. Retrotransposons and the FB transposon from *Drosophila melanogaster*, p. 485-502. In D. E. Berg, and M. M. Howe (eds.), **Mobile DNA**. Am. Soc. Microbiology, Washington, D. C.

Bishop, J. M. 1983. Cellular oncogenes and retroviruses. *Annu. Rev. Biochem.* 52:301-354.

Bishop, J. M., and H. E. Varmus. 1982. Functions and origins of retroviral transforming genes, pp. 999-1108. In R. Weiss, N. Teich, H. Varmus, and J. Coffin (eds.), **RNA Tumor Viruses**, vol. 1. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

Bishop, J. M., and H. E. Varmus. 1985. Functions and origins of retroviral transforming genes, pp. 246-356. In R. Weiss, N. Teich, H. Varmus, and J. Coffin (eds.), **RNA Tumor Viruses**, vol. 2. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

Boeke, J. D. 1989. Transposable elements in *Saccharomyces cerevisiae*, pp. 335-374. In D. E. Berg, and M. M. Howe (eds.), **Mobile DNA**. Am. Soc. Microbiology, Washington, D. C.

Boeke, J. D., D. J. Garfinkel, C. A. Styles, and G. R. Fink. 1985. Ty elements transpose through an RNA intermediate. *Cell* 40:491-500.

Bolognesi, D. P., R. Luftig, and J. H. Shaper. 1973. Localization of RNA tumor virus polypeptides. 1. Isolation of further virus substructures. *Virology* 56:549-564.

Botchan, M., W. C. Topp, and J. Sambrook. 1976. The arrangement of simian virus 40 sequences in the DNA of transformed cells. *Cell* 9:269-287.

Burnette, W. N., L. A. Holladay, and W. M. Mitchell. 1976. Physical and chemical properties of Moloney murine leukemia virus p30 protein: a major core structural component exhibiting high helicity and self-association. *J. Mol. Biol.* 107:131-143.

Cappello, J., K. Handelsman, and H. F. Lodish. 1985. Sequence of *Dictyostelium* DIRS-1: an apparent retrotransposon with inverted terminal repeats and an internal circle junction sequence. *Cell* 43:105-115.

Cheung, A. M. K., M. D. Hoggan, W. W. Hauswirth, and K. I. Berns. 1980. Integration of the adeno-associated virus genome into cellular DNA in latently infected Detroit 6 cells. *J. Virol.* 33:739-748.

Cobrinik, D., R. Katz, R. Terry, A. M. Skalka, and J. Leis. 1987. Avian sarcoma and leukosis virus *pol*-endonuclease recognition of the tandem long terminal repeat junction: minimum site required for cleavage is also required for viral growth. *J. Virol.* 61:1999-2008.

Coffin, J. M. 1979. Structure, replication, and recombination of retrovirus genomes: some unifying hypotheses. *J. Gen. Virol.* 42:1-26.

Coffin, J. M. 1982. Structure of the retroviral genome, pp. 261-368. In R. Weiss, N. Teich, H. Varmus, and J. Coffin (eds.), **RNA Tumor Viruses**, vol. 1. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

Coffin, J. M. 1985. Genome structure, pp. 17-74. In R. Weiss, N. Teich, H. Varmus, and J. Coffin (eds.), **RNA Tumor Viruses**, vol. 2. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

Colicelli, J. and S. P. Goff. 1985. Mutants and pseudorevertants of Moloney murine leukemia virus with alterations at the integration site. *Cell* 42:573-580.

Colicelli, J. and S. P. Goff. 1988a. Isolation of an integration provirus of Moloney murine leukemia virus with long terminal repeats in inverted orientation: integration utilizing two U3 sequences. *J. Virol.* 62:633-636.

Colicelli, J. and S. P. Goff. 1988b. Sequence and spacing requirements of a retrovirus integration site. *J. Mol. Biol.* **199**:47-59.

Copeland, N. G., N. A. Jenkins, and G. M. Cooper. 1981. Integration of Rous sarcoma virus DNA during transfection. *Cell* **23**:51-60.

Crawford, S. and S. P. Goff. 1984. Mutations in *gag* proteins P12 and P15 of Moloney murine leukemia virusw block early stages of infection. *J. Virol.* **49**:909-917.

Crawford, S. and S. P. Goff. 1985. A deletion mutation in the 5′ part of the *pol* gene of Moloney murine leukemia virus blocks proteolytic processing of the *gag* and *pol* polyproteins. *J. Virol.* **53**:899-907.

Cullen, B. R. and W. C. Greene. 1989. Regulatory pathways governing HIV-1 replication. *Cell* **58**:423-426.

Davis, N. L. and R. R. Rueckert. 1972. Properties of a ribonucleoprotein particle isolated from Nonidet P-40-treated Rous sarcoma virus. *J. Virol.* **10**:1010-1020.

Darlix, J.-L. and P.-F. Sphar. 1982. Binding sites of viral protein P19 onto Rous sarcoma virus RNA and possible controls of viral functions. *J. Mol. Biol.* **160**:147-161.

Davis, J., M. Scherer, W. P. Tsai, and C. Long. 1976. Low-molecular-weight Rauscher leukemia virus protein with preferential binding for single-stranded RNA and DNA. *J. Virol.* **18**:709-718.

Deininger, P. L. 1989. SINEs: short interspersed repeated DNA elements in higher eucaryotes, pp. 619-636. In D. E. Berg and M. M. Howe (eds.), **Mobile DNA**. Am. Soc. Microbiology, Washington, D. C.

DesGroseillers, L. and P. Jolicoeur. 1983. Physical mapping of the *Fv-1* tropism host range determinant of *Balb/c* murine leukemia viruses. *J. Virol.* **48**:685-696.

Dickson, C., R. Eisenman, H. Fan. E. Hunter, and N. Teich. 1982. Protein biosynthesis and assembly, pp. 513-648. In R. Weiss, N. Teich, H. Varmus, and J. Coffin (eds.), **RNA Tumor Viruses**, vol. 1. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

Dickson, C., R. Eisenman, and H. Fan. 1982. Protein biosynthesis and assembly, pp. 135-146. In R. Weiss, N. Teich, H. Varmus, and J. Coffin (eds.), **RNA Tumor Viruses**, vol. 2. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

Donehower, L. A. 1988. Analysis of mutant Moloney murine leukemia viruses containing linker insertion mutations in the 3′ region of pol. *J. Virol.* **62**:3958-3964.

Donehower, L. A. and H. E. Varmus. 1984. A mutant murine leukemia virus with a single missense codon in *pol* is defective in a function affecting integration. *Proc. Natl. Acad. Sci. USA* **81**:6461-6465.

Duyk, G., M. Longiaru, D. Cobrinik, R. Kowal, P. deHaseth, A. M. Skalka, and J. Leis. 1985. Circles with two tandem long terminal repeats are specifically cleaved by *pol* gene-associated endonuclease from avian sarcoma and leukosis viruses: Nucleotide sequences required for site-specific cleavage. *J. Virol.* **56**:589-599.

Eichinger, D. J. and J. D. Boeke. 1988. The DNA intermediate in yeast Tyl transposition copurifies with virus-like particles: cell-free Tyl transposition. *Cell* **54**:955-966.

Ellis, J. and A. Bernstein. 1989. Retrovirus vectors containing an internal attachment site: evidence that circle are not intermediates to murine retrovirus integtration. *J. Virol.* **63**:2844-2846.

Felsenstein, K. and S. P. Goff. 1988. Expression of the gag-gpol fusion protein of moloney murine leukemia virus without gag protein does not induce virion formation or proteolytic processing. *J. Virol.* **62**:2179-2182.

Finnegan, D. J. 1989a. The I factor and I-R Hybrid Dysgenesis in *Drosophila melanogaster*, pp. 503-518. In D. E. Berg and M. M. Howe (eds.), **Mobile DNA**. Am. Soc. Microbiology, Washington, D. C.

Finnegan, D. J. 1989b. F and related elements in *Drosophila melanogaster*, pp. 519-522. In D. E. Berg, and M. M. Howe (eds.), **Mobile DNA**. Am. Soc. Microbiology, Washington, D. C.

Fu., X.-D., P. J. Tuazon, J. A. Traugh, and J. P. Leis. 1988. Site-directed mutagenesis of the avian retrovirus nucleocapsid protein, pp12, at serine-40, the primary site of phosphorylation in vivo. *J. Biol. Chem.* **263**:2134-2139.

Fu., X.-D., P. J. Tuazon, J. A. Traugh, and J. P. Leis. 1988. Site-directed mutagenesis of the avian retrovirus nucleocapsid, pp12. *J. Biol. Chem.* **263**: 2140-2150.

Gebhardt, A., J. V. Bosch, A. Ziemiecki, and R. R. Friis. 1984. Rous sarcoma virus p19 and gp35 can be chemically crosslinked to high molecular weight complexes. *J. Mol. Biol.* **174**:297-317.

Goff, S. P. and L. I. Lobel. 1987. Mutants of murine leukemia viruses and retroviral replication. *Biochim. et. Biophys. Acta* **907**:93-123.

Gorelick, R. J., L. E. Henderson, J. P. Hanser, and A. Rein. 1988. Point mutants of Moloney murine leukemia virus that fail to package viral RNA:

evidence for specific RNA recognition by a "zinc-finger-like" protein sequence. *Proc. Natl. Acad. Sci. USA* 85:8420-8424.

Grandgenett, D. P. and A. C. Vora. 1985. Site-specific nicking at the avian retrovirus LTR circle junction by the viral pp32 DNA endonuclease. *Nuc. Acids Res.* 13:6205-6221.

Grandgenett, D. P., A. C. Vora, R. Swanstrom, J. C. Olsen. 1986. Nuclease mechanism of the avian retrovirus pp32 endonuclease. *J. Virol.* 58:970-974.

Hagino-Yamagishi, K., L. A. Donehower, and H. E. Varmus. 1987. Retroviral DNA integrated during infection by an integration-deficient mutant of murine leukemia virus is oligomeric. *J. Virol.* 61:1964-1971.

Henderson, L. E., H. C. Krutzsch, and S. Oroszlan. 1983. Myristyl amino-terminal acylation of murine retrovirus proteins--an unusual posttranslational protein modification. *Proc. Natl. Acad. Sci. USA* 80:339-343.

Hill, M. and J. Hillova. 1972. Virus recovery in chicken cells tested with Rous sarcoma cell DNA. *Nat. New Biol.* 237:35-39.

Hsu, H.-W., P. Schwartzberg, and S. P. Goff. 1985. Point mutations in the P30 domain of the *gag* gene of Moloney murine leukemia virus. *Virology* 142:211-214.

Hunter, E. 1978. The mechanism of genetic recombination in avian viruses. *Current Topics in Microbiology and Immunology* 79:295-309.

Hutchinson, C. A., S. C. Hardies, D. D. Loeb, W. R. Shehee, and M. H. Edgell. 1989. LINEs and related retroposons: long interspersed repeated sequences in the eucaryotic genome, pp. 593-618. In D. E. Berg and M. M. Howe (eds.), Mobile DNA. Am. Soc. Microbiology, Washington, D. C.

Jacks T., M. D. Power, F. R. Masiarz, P. A. Luciw, P. J. Barr, and H. E. Varmus. 1988. Characterization of ribosomal frameshifting in HIV-1 *gag-pol* expression. *Nature* 231:280-283.

Jacks T., K. Townsley, H. E. Varmus, and J. Majors. 1987. Two efficient ribosomal frameshifting events are required for synthesis of mouse mammary tumor virus *gag*-related polyproteins. *Proc. Natl. Acad. Sci. USA* 84:4298-4302.

Jacks, T., and H. E. Varmus. 1985. Expression of Rous sarcoma virus *pol* gene by ribosomal frameshifting. *Science* 230:1237-1242.

Johnson, S. P., M. Veigl, T. Vanaman, and J. Leis. 1983. Cyanogen bromide digestion of the avian myeloblastosis virus pp19 protein: isolation of an amino-terminal peptide that binds to viral RNA. *J. Virol.* 45:876-881.

Jolicoeur, P. 1979. The *Fv-1* gene of the mouse and its control of murine leukemia virus replication. *Curr. Topics Microbiol. Immunol.* **86**:67-122.

Katoh, I., Y. Yoshinaka, A. Rein, M. Shibuya, T. Odaka, and S. Oroszlan. 1985. Murine leukemia virus maturation: *protease* region required for conversion from "immature" to "mature" core form and for virus infectivity. *Virology* **145**:280-292.

King, W., M. D. Patel, L. I. Lobel, S. P. Goff, and M. C. Nguyen-Huu. 1985. Insertion mutagenesis of embryonal carcinoma cells by retrovirisues. *Science* **228**:554-558.

Kriegler, M. and M. Botchan. 1983. Enhanced transformation by a simian virus 40 recombinant virus containing a Harvey murine sarcoma virus long terminal repeat. *Mol. Cell. Biol.* **3**:325-339.

Leis, J. P. and J. Jentoft. 1983. Characteristics and regulation of interaction of avian retrovirus pp12 with viral RNA. *J. Virol.* **48**:361-369.

Leis, J. P., S. Johnson, L. S. Collins, and J. A. Traugh. 1984. Effects of phosphorylation of avian retrovirus nucleocapsid protein pp12 on binding of viral RNA. *J. Biol. Chem.* **259**:7726-7733.

Leis, J. P., J. McGinnis, and R. W. Green. 1978. Rous sarcoma virus p19 binds to specific double-stranded regions of viral RNA: Effect on cleavage of viral RNA by RNase III. *Virology* **84**:87-98.

Leis, J. P., P. Scheible, and R. E. Smith. 1981. Correlation of RNA binding affinity of avian oncornavirus p19 proteins with the extent of processing of virus genome RNA in cells. *J. Virol.* **35**:722-731.

Levin, J. G. and M. J. Rosenak. 1976. Synthesis of murine leukemia virus proteins associated with virions assembled in actinomycin D-treated cells: evidence for the persistence of viral messenger RNA. *Proc. Natl. Acad. Sci. USA* **73**:1154-1158.

Levin, J. G., P. M. Grimley, J. M. Ramseur, and I. K. Berezesky. 1974. Deficiency of 60 to 70S RNA in murine leukemia virus particles assembled in cells treated with actinomycin D. *J. Virol.* **14**:152-161.

Linial, M., and D. Blair. 1982. Genetics of retroviruses, pp. 649-783. In R. Weiss, N. Teich, H. Varmus, and J. Coffin (eds.), **RNA Tumor Viruses**, vol. 1. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

Linial, M., and D. Blair. 1985. Genetics of retroviruses, pp. 147-186. In R. Weiss, N. Teich, H. Varmus, and J. Coffin (eds.), **RNA Tumor Viruses**, vol. 2. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

Linial, M., E. Medeiros, and W. S. Hayward. 1978. An avian oncovirus mutant (SE 21Q1b) deficient in genomic RNA: biological and biochemical characterization. *Cell* 15:1371-1381.

Lobel, L. I. and S. P. Goff. 1984. Construction of mutants of Moloney murine leukemia virus by suppressor-linker insertional mutagenesis: Positions of viable insertion mutations. *Proc. Natl. Acad. Sci. USA* 81:4149-4153.

Lobel, L. I., J. E. Murphy, and S. P. Goff. 1989. The palindromic LTR-LTR junction of Moloney murine leukemia virus is not an efficient substrate for proviral integration. *J. Virol.* 63:2629-2637.

Luciw, P. A., H. Oppermann, J. M. Bishop, and H. E. varmus. 1984. Integration and expression of several forms of Rous sarcoma virus DNA used for transfection of mouse cells. *Mol. Cell. Biol.* 4:1260-1269.

Maddon, P. J., A. G. Dalgleish, J. S. McDougal, P. R. Clapham, R. A. Weiss, and R. Axel. 1986. The T4 gene encodes the Aids virus receptor and is expressed in the immune system and the brain. *Cell* 47:333-348.

Matsuo, T., M. Heller, L. Petti, E. O'Shiro, and E. Kieff. 1984. Persistence of the entire Epstein-Barr virus genome integrated into human lymphocyte DNA. *Science* 226:1322-1324.

McClure, M. O., M. Marsh, and R. A. Weiss. 1988. Human immunodeficiency virus infection of CD4-bearing cells occurs by a pH-independent mechanism. *EMBO J.* 7:513-518.

McDougal, J. S., M. S. Kennedy, J. M. Sligh, S. P. Cort, A. Mawle, and J. K. A. Nicholson. 1986. Binding of HTLV-III/LAV to $T4^+$ T cells by a complex of the 100K viral protein and the T4 molecule. *Science* 231:382-385.

Meric, C., J.-L. Darlix, and P.-F. Spahr. 1984. It is Rous sarcoma virus p12 and not p19 that binds tightly to Rous sarcoma virus RNA. *J. Mol. Biol.* 173:531-538.

Meric, C. and P.-F. Spahr. 1986. Rous sarcoma virus nucleic acid-binding protein p12 is necessary for viral 70S RNA dimer formation and packaging. *J. Virol.* 60:450-459.

Miller, M., M. Jaskolski, J. K. Rao, J. Leis, and A. Wlodawer. 1989. *Nature* 337:576-579.

Naso, R. B., W. L. Karshin, Y. H. Wu, and R. B. Arlinghaus. 1979. Characterization of 40,00- and 25,000-dalton intermediate precursors to Rauscher murine leukemia virus *gag* gene products. *J. Virol.* 32:187-198.

Navia, M. A., P. M. D. Fitzgerald, B. M. McKeever, C.-T. Leu, J. C. Heimbach, W. K. Herber. I. S. Sigal, P. L. Darke, and J. P. Springer.

1989. Three-dimensional structure of aspartyl protease from human immunodeficiency virus HIV-1. *Nature* 337:615-620.

Palmiter, R. D., J. Gagnon, V. M. Vogt, S. Ripley, and R. N. Eisenman. 1978. The NH_2-terminal sequence of the avian oncovirus *gag* precursor polyprotein ($Pr76^{gag}$). *Virology* 91:423-433.

Panganiban, A. T. and H. M. Temin. 1983. The terminal nucleotides of retrovirus DNA are required for integration but not virus production. *Nature* 306:155-160.

Panganiban, A. T. and H. M. Temin. 1984. The retrovirus *pol* gene encodes a product required for DNA integration: identification of a retrovirus *int* locus. *Proc. Natl. Acad. Sci. USA* 81:7885-7889.

Panganiban, A. T. and H. M. Temin. 1985. Circles with two tandem LTRs are precursor to integrated retrovirus DNA. *Cell* 36:673-679.

Pepinsky, R. B. and V. M. Vogt. 1983. Purification and properties of a fifth major viral *gag* protein from avian sarcoma and leukemia viruses. *J. Virol.* 45:648-658.

Pepinsky, R. B. and V. M. Vogt. 1984. Fine-structure analyses of lipid-protein and protein-protein interactions of *gag* protein p19 of the avian sarcoma and leukemia viruses by cyanogen bromide mapping. *J. Virol.* 52:145-153.

Perez, L. G., G. L. Davis, and E. Hunter. 1987. Mutants of the Rous sarcoma virus envelope glycoprotein that lack the transmembrane anchor and cytoplasmic domains: analysis of intracellular transport and assembly into virions. *J. Virol.* 61:2981-2988.

Redmond, S., G. Peters, and C. Dickson. 1984. Mouse mammary tumor virus can mediate cell fusion at reduced pH. *Virology* 133:393-402.

Rein, A., M. R. McClure, N. R. Rice, R. B. Luftig, and A. M. Schultz. 1986. Myristylation site in $Pr65^{gag}$ is essential for virus formation by Moloney murine leukemia virus. *Proc. Natl. Acad. Sci. USA* 83:7246-7250.

Roberts, M. M. and S. Oroszlan. 1989. The preparation and biochemical characterization of intact capsids of equine infectious anemia virus. *Biochem. and Biophys. Res. Comm.* 160:486-494.

Robinson, R. A. and D. J. O'Callaghan. 1983. A specific viral DNA sequence is stably integrated in herpes virus oncogenically transformed cells. *Cell* 32:569-578.

Rohdewohld, H., H. Weiher, W. Reik, R. Jaenisch, and M. Briendl. 1987. Retrovirus integration and chromatin structure: Moloney murine leukemia virus gene expression. *Proc. Natl. Acad. Sci. USA* 84:4919-4923.

Sawyer, R. C. and H. Hanafusa. 1979. Comparison of the small RNAs of polymerase-deficient and polymerase-positive Rous sarcoma virus and another species of avian retrovirus. *J. Virol.* 29:863-871.

Schulein, M., W. N. Burnette, and J. T. August. 1978. Stoichiometry and specificity of binding of Raushcer oncovirus 10,00-dalton (p10) structural protein to nucleic acids. *J. Virol.* 26:54-60.

Schultz, A. M., L. E. Henderson, and S. Oroszlan. 1988. Fatty acylation of proteins. *Ann. Rev. Cell Biology* 4:611-647.

Schultz, A. M. and S. Oroszlan. 1983. In vivo modification of retroviral *gag* gene-encoded polyproteins by myristic acid. *J. Virol.* 46:355-361.

Schwartzberg, P., J. Colicelli, and S. P. Goff. 1984. Construction and analysis of deletion mutations in the *pol* gene of Moloney murine leukemia virus: a new viral function required for productive infection. *Cell* 37:1043-1052.

Schwartzberg, P., J. Colicelli, M. L. Gordon, and S. P. Goff. 1984. Mutations in the *gag* gene of Moloney murine leukemia virus: effects on production of virions and reverse transcriptase. *J. Virol.* 49:918-924.

Scott, M. L., K. McKereghan, H. S. Kaplan, and K. E. Fry. 1981. Molecular cloning and partial characterization of unintegrated linear DNA from gibbon ape leukemia virus. *Proc. Natl. Acad. Sci. USA* 78:4213-4217.

Sen, A., C. J. Sherr, and G. J. Todaro. 1976. Specific Binding of the type C viral core protein p12 with purified viral RNA. *Cell* 7:21-32.

Sen, A., C. J. Sherr, and G. J. Todaro. 1977. Phosphorylation of murine type C viral p12 proteins regulates their extent of binding to the homologous RNA. *Cell* 10:489-496.

Sen, A. and G. J. Todaro. 1977. The genome-associated, specific RNA binding proteins of avian and mammalian type C viruses. *Cell* 10:91-99.

Shank, P. R., S. H. Hughes, H.-J. Kung, J. E. Majors, N. Quintrell, R. V. Guntaka, J. M. Bishop, and H. E. Varmus. 1978. Mapping unintegrated avian sarcoma virus DNA: termini of linear DNA bear 300 nucleotides present once or twice in two species of circular DNA. *Cell* 15:1383-1395.

Shank, P. R. and H. E. Varmus. 1978. Virus-specific DNA in the cytoplasm of avian sarcoma virus-infected cells is a precursor to covalently closed circular viral DNA in the nucleus. *J. Virol.* 25:104-114.

Shields, A., O. N. Witte, E. Rothenberg, and D. Baltimore. 1978. High frequency of aberrant expression of Moloney murine leukemia virus in clonal infections. *Cell* 14:601-609.

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

Shinnick, T. M., R. A. Lerner, J. G. Sutcliffe. 1981. Nucleotide sequence of Moloney murine leukemia virus. *Nature* 293:543-548.

Shoemaker, C., S. P. Goff, E. Gilboa, M. Paskind, S. W. Mitra, and D. Baltimore. 1980. Structure of a cloned circular Moloney murine leukemia virus DNA molecule containing an inverted segment: implications for retrovirus integration. *Proc. Natl. Acad. Sci. USA* 77:3932-3936.

Shurtz, R., S. Dolev, M. Aboud, and S. Salzberg. 1979. Viral genome RNA serves as messenger early in the infectious cycle of murine leukemia virus. *J. Virol.* 31:668-676.

Soriano, P., T. Gridley, and R. Jaenisch. 1989. Retroviral tagging in mammalian development and genetics, pp. 927-938. In D. E. Berg and M. M. Howe (eds.), **Mobile DNA**. Am. Soc. Microbiology, Washington, D. C.

Stein, B. S., S. D. Gowda, J. D. Lifson, R. C. Penhallow, K. G. Bensch, and E. G. Engleman. 1987. pH-independent HIV entry into CD4-positive T cells via virus envelope fusion to the plasma membrane. *Cell* 49:659-668.

Stromberg, K., N. E. Hurley, N. L. Davis, R. R. Rueckert, and E. Fleissner. 1974. Structural studies of avain myeloblastosis virus: comparison of polypeptides in virion and core component by dodecyl sulfate-polyacrylamide gel electrophoresis. *J. Virol.* 13: 513-528.

Tanese, N., and S. P. Goff. 1988. Domain structure of the Moloney murine leukemia virus reverse transcriptase: mutational analysis and separate expression of the DNA polymerase and RNase H activities. *Proc. Natl. Acad. Sci. USA* 85:1777-1781.

Teich, N., A. Bernstein, T. Mak, J. Wyke, and W. Hardy. 1982. Pathogenesis of retroviral diseases, pp. 785-998. In R. Weiss, N. Teich, H. Varmus, and J. Coffin (eds.), **RNA Tumor Viruses**, vol. 1. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

Teich, N., J. Wyke, and Paul Kaplan. 1985. Pathogenesis of retrovirus-induced disease, pp. 187-248. In R. Weiss, N. Teich, H. Varmus, and J. Coffin (eds.), **RNA Tumor Viruses**, vol. 2. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

Temin, H. M. 1961. Mixed infection with two types of Rous sarcoma virus. *Virology* 13:158-163.

Temin, H. M. 1964. The effects of actinomycin D on growth of Rous sarcoma virus in vitro. *Virology* 20:577-582.

Temin, H. M. 1971. Mechanism of cell transformation by RNA rumor viruses. *Annu. Rev. Microbiol.* 25:609-649.

Temin, H. M. and S. Mizutani. 1970. RNA-directed DNA polymerase in virions of Rous sarcoma virus. *Nature* 226:1211-1213.

Temin, H. M. and H. Rubin. 1959. A kinetic study of infection of chick embryo cells in vitro by Rous sarcoma virus. *Virology* 8:209-222.

Van de Ven, W. J. M, A. J. M. Vermorken, C. Onnekink. H. P. J. Bloemers, and H. Bloemendal. 1978. Structural studies on Rauscher murine leukemia virus: isolation and characterization of viral envelopes. *J. Virol.* 27:595-603.

Varmus, H. E. 1983. Retroviruses, pp. 411-503. In J. Shapiro (ed.), **Mobile Genetic Elements**. Academic Press, Inc., New York.

Varmus, H. E. 1987. Reverse transcription. *Sci. Am.* 257:56-66.

Varmus, H. E. 1988a. Retroviruses. *Science* 240:1427-2435.

Varmus, H. 1988b. Regulation of HIV and HTLV gene expression. *Genes and Development* 2:1055-1062.

Varmus, H. E. 1989. Reverse transcription in bacteria. *Cell* 56:721-724.

Varmus, H. E., J. M. Bishop, and P. K. Vogt. 1973. Appearance of virus-specific DNA in mammalian cells following transformation by Rous sarcoma virus. *J. Mol. Biol.* 74:613-626.

Varmus, H., and P. Brown. 1989. Retroviruses, pp. 53-108. In D. E. Berg, and M. M. Howe (eds.), **Mobile DNA**. Am. Soc. Microbiology, Washington, D. C.

Varmus, H. E., R. V. Guntaka, W. J. W. Fan, S Heasley, and J. M. Bishop. 1974. Synthesis of viral DNA in the cytoplasm of duck embryo fibroblasts and in enucleated cells after infection by avian sarcoma virus. *Proc. Natl. Acad. Sci. USA* 71:3874-3878.

Varmus, H. E., T. Padgett, S. Heasley, G. Simon and J. M. Bishop. 1977. Cellular functions are required for the synthesis and integration of avian sarcoma virus-specific DNA. *Cell* 11:307-319.

Varmus, H. E., and R. Swanstrom. 1982. Replication of retroviruses, pp. 369-512. In R. Weiss, N. Teich, H. Varmus, and J. Coffin (eds.), **RNA Tumor Viruses**, vol. 1. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

Varmus, H. E., and R. Swanstrom. 1982. Replication of retroviruses, pp. 75-134. In R. Weiss, N. Teich, H. Varmus, and J. Coffin (eds.), **RNA**

Tumor Viruses, vol. 2. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

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

Weiner, A. M., P. L. Deininger, and A. Efstratiadis. 1987. Nonviral transposons: genes pseudogenes and transposable elements generated by the reverse flow of genetic information. *Annu. Rev. Biochem.* 55:631-662.

Weiss, R. 1985. Human T-cell retroviruses, pp. 405-485. In R. Weiss, N. Teich, H. Varmus, and J. Coffin (eds.), **RNA Tumor Viruses**, vol. 2. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

Wettstein, F. O., and J. G. Stevens. 1982. Variable-sized free episomes of Shope papilloma virus DNA are present in all non-virus producing neoplasms and integrated episomes are detected in some. *Proc. Natl. Acad. Sci. USA* 79:790-794.

Yoshinaka, Y., I. Katoh, T. D. Copeland, and S. Oroszlan. 1985. Murine leukemia virus protease is encoded by the *gag-pol* gene and is synthesized through suppression of an amber termination codon. *Proc. Natl. Acad. Sci. USA* **82**:1618-1622.

Yoshinaka, Y., and R. B. Luftig. 1977. Murine leukemia virus morphogenesis: cleavage of P70 in vitro can be accompanied by a shift from a concentrically coiled internal strand ("immature") to a collapsed ("mature") form of the virus core. *Proc. Natl. Acad. Sci. USA* 74:3446-3450.

CHAPTER 2

MOLECULAR ANALYSIS OF THE RETROVIRUS INTEGRATION REACTION

An important step in understanding the mechanism of any example of genetic recombination is the precise definition of the precursors and products involved. For this reason, advances in the study of recombination have often come from the development of assays that accurately reproduce in vitro the reaction of interest. In vitro assays make it possible to directly analyze the precursors and products of a reaction--a much more difficult proposition in vivo, given the ambiguities introduced by other cellular processes. The work described in this thesis began with the development of an in vitro assay for retroviral integration which resulted from a collaboration with Pat Brown.

Pat had designed an in vitro assay for retroviral integration that incorporated a powerful genetic selection in *E. coli* for the detection of recombinants, and I had identified structures inside infected cells that contain unintegrated retroviral DNA. By combining the sensitivity of genetic selection with the use of precursors that had been assembled in vivo by acute infection of tissue culture cells, we succeeded in detecting recombinants from an in vitro reaction and optimized the reaction sufficiently to permit biochemical analysis of retroviral integration. This narrative will describe the origins of this collaborative effort and summarize the findings that followed. Two publications that are reproduced at the end of this chapter provide the experimental details of our findings.

A Genetic Assay for Retroviral Integration

The definition of important *cis*- and *trans*-acting viral functions required for proper integration suggested to Pat a strategy for reproducing the MLV integration reaction in a cell-free system. He designed a DNA substrate for integration by constructing a donor plasmid that included two key components: a cloned fragment of viral DNA encompassing the MLV circle junction to serve as an *att* site (based on the SNV circle junction experiment, which predated by several years the conflicting results from studies of MLV), and a bacterial *supF* tRNA gene to provide a genetic selection in *E. coli* for the detection of recombinants (Seed 1983). Pat chose to use purified bacteriophage λ gtWES DNA as a target, or recipient, DNA molecule for two reasons. First, the nonessential regions of the λ gtWES genome (which account for about 25% of its DNA sequences) could accomodate insertion of the donor plasmid and still be within the size limits for in vitro packaging of λ DNA into phage particles. Second, the λ gtWES genome contains amber mutations in three genes (W,E, and S) essential for lytic growth in a nonsuppressor bacterial host. The suppression of these amber mutations by the *supF* tRNA gene provided a genetic selection in *E. coli* for the growth of λ gtWES bacteriophage which had acquired a copy of the donor plasmid. Finally, Pat provided viral trans-acting activities either by adding disrupted MLV virions (with or without uninfected cell extracts in case any unidentified host functions were needed) or by adding extracts made from acutely infected cells.

These components were mixed together and incubated to allow time for integration of the plasmid donor molecule into the the λ gtWES target DNA. The λ gtWES target DNA (and any recombinant products) were subsequently purified, and packaged in vitro into phage particles using commercially available reagents. Recombinants could be detected by their ability to form plaques on a nonsuppressor bacterial host. This experimental strategy provided a powerful genetic selection, since as many as 10^8 bacteriophage genomes could readily be examined from a single reaction mixture for the presence of any recombinant molecules.

Extracts from Acutely Infected Cells as a Source of Integration Activity

At the time Pat was developing a genetic assay for retroviral integration, I was studying retroviral replication by looking inside infected cells to determine what viral proteins remain associated with the viral genome, and what kinds of higher order structures, if any, are involved in retroviral replication after entry into a cell. Such an analysis at the time was limited by the absence of assays to verify that any viral structures found inside infected cells were active for subsequent steps and hence relevant to replication. Because of this, I chose to examine structures that contained full-length, unintegrated retroviral DNA, which would at least indicate that such structures had initially been competent for the completion of reverse transcription.

One simple experiment proved informative. I harvested tissue culture cells at a time after infection by MLV at which there were several

copies per cell of full-length viral DNA (12-24 hours; see Figure 3), and fractionated these infected cells into cytoplasmic and nuclear extracts using mild detergent and isotonic salt conditions. When I analyzed such extracts by velocity sedimentation through sucrose gradients, the viral DNA was found to sediment as part of a large (roughly 160S) nucleoprotein complex. If the same extracts were first treated with proteinase, the viral DNA stayed at the top of the gradient, consistent with previous work demonstrating that purified viral DNA sediments at about 20S. These experiments are described in more detail in Chapter 3.

A Successful Collaboration

Even though we could not at the time confirm that the nucleoprotein complexes I had identified were active for integration, the finding that unintegrated viral DNA inside infected cells resides within a large complex led us to suspect that such structures might be important for efficient integration and might be difficult to reconstitute in an in vitro system. In addition, other experiments had shown that when plasmid DNA containing the MLV circle junction was transfected into tissue culture cells, the circle junction was not used for integration even when viral proteins were provided by superinfection, suggesting that perhaps integration activity was confined to authentic reverse-transcripts generated during infection (Botchan and Kriegler 1983). Our suspicions were further supported by the difficulties Pat encountered in attempting to reconstitute integration using the approaches described

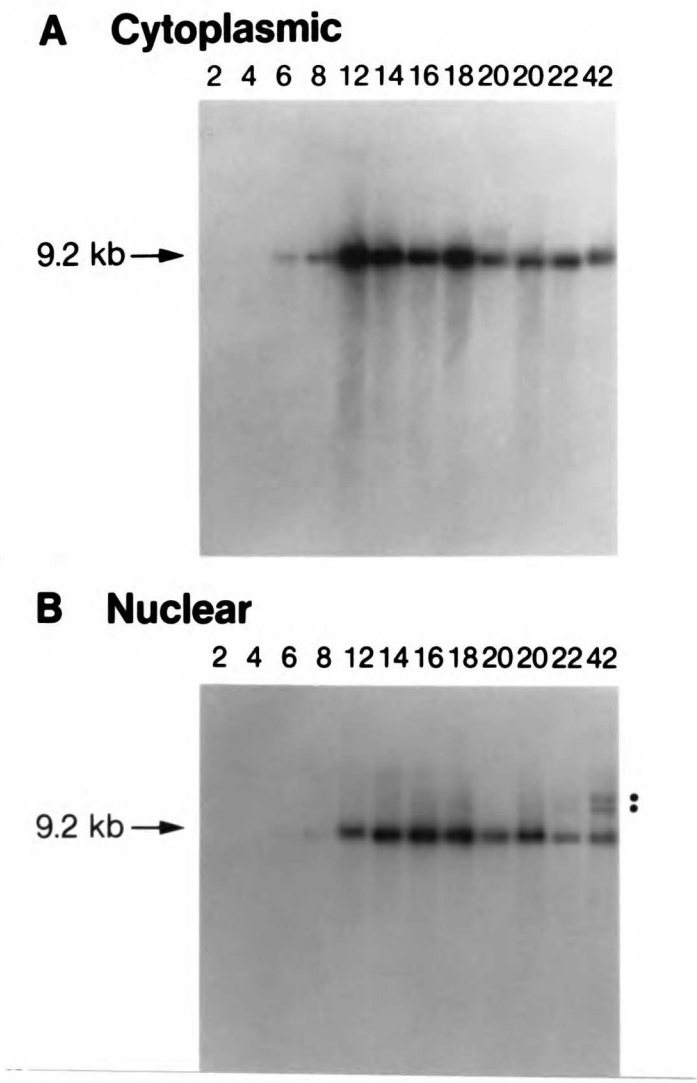


Figure 3. Unintegrated Viral DNA in Cytoplasmic and Nuclear Extracts From a Time Course of Viral DNA Synthesis During Infection by Cocultivation.

Viral DNA was prepared from cytoplasmic and nuclear extracts of NIH 3T3 cells that were acutely infected by cocultivation with an MLVsupF producer line. DNA samples were resolved by agarose gel electrophoresis, then transferred to a nylon filter. MLV DNA was identified by hybridization, using as a probe the plasmid p8.2 (a complete, permuted clone of the MLV 1-LTR circle), labeled by nick translation. The time at which the cells were harvested, in hours after beginning the cocultivation, are indicated above each lane. Each sample represents 100 microliters of extract from about 4 x 10^6 infected cells. The samples for the 22 and 42 hour time points were derived from an independent cocultivation; 20 hour time points from both sets of cocultivations are shown. A. Viral DNA from cytoplasmic extracts. B. Viral DNA from nuclear extracts. Arrows indicate the 9.2 kb full-length linear viral DNA. The relaxed one and two LTR circular forms of viral DNA are indicated at the right by two dots.

earlier. These results encouraged us to consider using extracts from acutely infected cells as a source of precursors that would be active for integration in an in vitro assay.

Conveniently, we were able to obtain a fully replication-competent derivative of MLV that stably carries a bacterial *supF* tRNA gene genetically engineered into a non-essential region of U3 (Lobel et al 1985). This virus (called MLV*supF*), which had been constructed to facilitate the cloning of proviruses from chromosomal DNA, made it possible to bridge the two approaches Pat and I had been taking toward the study of integration. By infecting cells with MLV*supF*, we could make extracts containing viral precursors that would potentially be active for integration in the genetic assay Pat had designed which used λ gtWES DNA as a target molecule.

Verifying the Authenticity of the In Vitro Reaction

In our first attempt to detect recombinants, we observed a single plaque from a reaction to which we had added nuclear extract harvested 24 hours after infection of NIH 3T3 cells with MLV*supF*. Pat amplified the phage from the potential recombinant plaque and purified the bacteriophage DNA. He then confirmed that the λ gtWES genome had acquired an authentic provirus by restriction mapping (which showed that the provirus was grossly intact) and by DNA sequence analysis of the MLV*supF*-λ gtWES junction sequences (which showed that two base pairs had been lost from each end of the viral DNA and that four base pairs at the

target site had been duplicated as short direct repeats flanking the provirus). These results proved that we had reproduced in vitro an authentic retrovirus integration reaction.

After optimizing the conditions for the reaction, we were able to obtain hundreds to thousands of recombinant plaques from a single in vitro reaction. When these numbers were corrected for the efficiency of in vitro packaging into phage particles and for other parameters, we found that between 0.1% and 0.5% of the unintegrated viral DNA molecules present in a reaction were capable of productively integrating in vitro as scored by the assay. Restriction analysis of over thirty different recombinants revealed that the proviruses had integrated throughout the nonessential regions of the λ gtWES genome with no apparent specificity either in the sites of integration, or in the polarity of the viral genome with respect to the bacteriophage genome. In addition, sequence analysis of six additional recombinants showed that in every case, two base pairs of viral DNA were lost from the ends of each LTR and four base pairs of the host target sequence were duplicated as short direct repeats flanking the provirus.

Biochemical Characterization of the In Vitro Reaction

By varying the reaction conditions, we easily defined the basic parameters of the reaction. EDTA abolishes integration activity, demonstrating the requirement for a divalent cation, which was satisfied by Mg^{++} in our reactions. The optimal K^{+} concentration is 150 mM, with

concentrations higher than 300 mM significantly reducing activity. The pH range for activity is broad, with comparable levels of activity from pH 6.6 to pH 8.6. The optimal temperature range for incubation is 15 to 43°C, with no detectable integration at 0°C. Pretreatment with proteinase and SDS, or N-ethylmaleimide, abolishes activity, while RNase A pretreatment does not.

The activity is unfortunately very labile, as shown in Figure 4. The loss of roughly 50% of the integration activity after preincubation at 37° for only five minutes limits the ease with which one can experimentally manipulate the extracts.

Linear Viral DNA Participates in the In Vitro Reaction

When developing the assay, we assumed that nuclear extracts, in which we could detect 2 LTR circles, would be the best source of integration activity. However, we were surprised to find that upon comparison of cytoplasmic and nuclear extracts made 24 hours after infection of cells in tissue culture, the cytoplasmic extracts typically were more active in our in vitro assay than were the nuclear extracts. Mixing experiments with uninfected extracts showed that this difference in activity was due neither to stimulatory factors in the cytoplasm or inhibitory factors in the nucleus. By carefully quantifying the levels of the different forms of viral DNA (linear, one LTR circle, and two LTR circle) present in the two extracts, we discovered that the level of activity correlated very well with the number of linear viral DNA

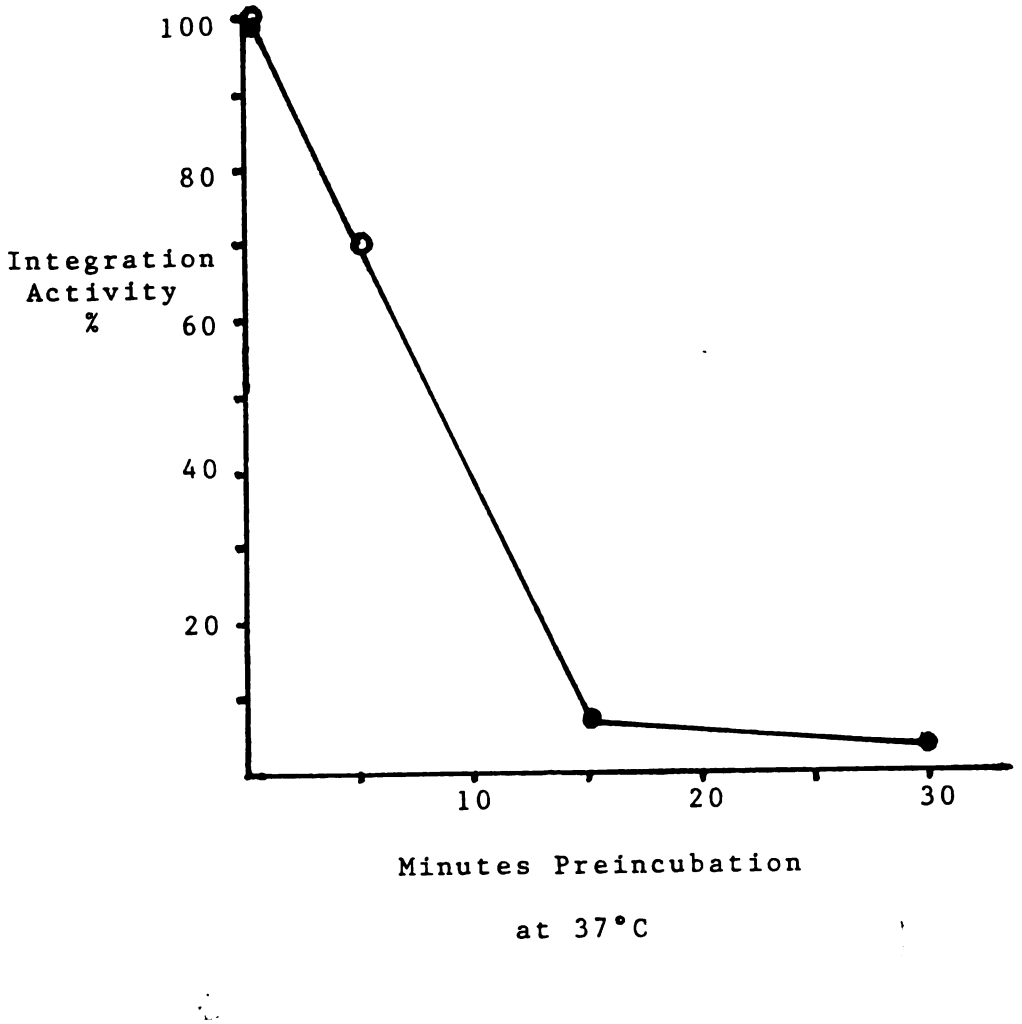


Figure 4. Loss of Integration Activity by Preincubation of Extracts before Addition of Target DNA.

Reaction mixtures were prepared using standard conditions after preincubation of the extract at 37°C for the times indicated. These data were compiled from experiments using two different preparations of extracts, indicated by the filled and empty circles. Both sets of data were normalized relative to percentage of integration activity present with no preincubation (t=0).

molecules present in an extract, and not at all well with the number of two LTR circles present. This was true both for comparisons between cytoplasmic and nuclear extracts from the same population of infected cells, and for comparisons between extracts from different populations of infected cells.

Furthermore, the number of recombinants we could score from a single reaction using cytoplasmic extracts argued either that the less-than-detectable numbers of 2 LTR circles that could have been present in cytoplasmic extracts integrated with extremely high efficiency (at least 60%), or that the linear molecules present were serving as precursors. The former possibility seemed unlikely, since the calculated integration efficiency for two LTR circles in reactions using nuclear extracts was much lower (about 4%, assuming they were the only active precursors). Given the good correlation between the levels of linear viral DNA and integration activity in both cytoplasmic and nuclear extracts, and the absence of any stimulation or inhibition of activity during mixing experiments, our results strongly suggested that linear viral DNA molecules participate in our in vitro reaction, either directly or through a circular intermediate.

Integration Activity Does Not Require an Exogenous Energy Source

To address whether or not the linear molecules pass through a circular intermediate, we asked if integration activity depends on an exogenously added energy source. Since the formation of a two LTR

circle from linear viral DNA requires the formation of new phosphodiester bonds, such a pathway would require an energy input. By passing cytoplasmic extracts containing active precursors through a gel filtration column, we could reduce the ATP levels in a reaction to about 1 nM, with no loss in the efficiency of integration. Although alternative explanations could account for this result (for example, the complex containing the unintegrated viral DNA could somehow be precharged at an earlier step) the simplest explanation was that the linear molecule could serve as a direct precursor to the integrated provirus. Such a reaction does not involve the net formation of new phosphodiester bonds and would not require an exogenous energy source if the strand breakage-and-joining reaction were coupled. We therefore proposed two possible pathways for the integration of retroviral DNA, one in which the linear viral DNA molecule directly participates in the reaction, and one in which it first passes through a 2 LTR circular intermediate.

Identification of the Immediate Viral DNA Precursor in Retroviral Integration

The two proposed pathways for retroviral integration make distinguishable predictions about the structure of the initial recombinant molecule (see Figure 7 on page 99). The presence of short direct repeats of the host target sequence flanking the provirus, and the loss of two base pairs from the tips of each LTR, together imply

that the reaction pathway requires both the staggered cleavage of the target DNA (in order to generate the host repeat), and cleavages two bases from the tips of the LTRs in the viral DNA (which would account for their loss). If the 2 LTR circle were cleaved in this way (i.e. to create a four base pair staggered cleavage), the two bases at the end of an LTR would remain covalently linked to the viral DNA at the end of the other LTR. On the other hand, if the linear form of viral DNA were cleaved in this way, the two extra bases would simply be lost. In either case, repair processes (either in *E. coli* or in an infected cell) presumably eliminate all four extra base pairs of viral DNA, and fill in the gap left from staggered cleavage of the host DNA to generate the short direct repeats. However, the structure of the initial recombinant product, before repair, should reflect the origin of the viral DNA that participates in the integration reaction. For the 2 LTR circle pathway, retention of the two bases after cleavage should result in the presence of a four base extension of viral DNA at the 5′ end of each strand of viral DNA in the recombinant molecule. The loss of the two bases after cleavage from a linear precursor should result in only a 2 base extension at each free 5′ end (assuming in both cases that only the 3′ end of viral DNA is initially joined to the target DNA). Thus by directly purifying recombinant molecules from an integration reaction and examining the sequences present at the very ends of the viral DNA, one should be able to distinguish between these two pathways and obtain direct evidence for the identity of the actual precursor to the retroviral integration reaction. Furthermore, we wanted to directly

analyze the ends of purified, unintegrated linear viral DNA to define the precise structure of the *att* sites in the entire population of potential precursors. The second of the two papers included at the end of this chapter addresses precisely these issues, and the results are summarized in the remaining paragraphs of this narrative.

A Physical Assay for Retroviral Integration

The genetic assay described earlier, while informative, suffers from the shortcoming that the recombinant products of the integration reaction can be examined only after replication in *E. coli*, by which time the information contained in the structure of the initial strand-exchange product is lost due to DNA repair. Analysis of the initial recombinant molecule required the use of a physical assay that would permit the direct purification of recombinant molecules. Pat therefore modified the genetic assay to make this possible by substituting ϕX RF1 DNA for λ gtWES DNA as a target molecule. In addition, the activity in cytoplasmic extracts was first partially purified by gel-exclusion column chromatography (using a matrix with an exclusion limit of 5×10^6 daltons) to remove any contaminating nucleases and nucleotides that might modify the structure of the initial recombinant product. Otherwise, the reaction conditions and the procedure for purification of DNA from the reaction were the same as for the genetic assay.

Instead of detecting recombinants by packaging the purified DNA into phage particles, the DNA from reaction was restricted with an

endonuclease that cleaves MLV DNA but not the ϕX target DNA. When the restricted DNA is electrophoresed in an agarose gel, three types of products are expected. (i) Nonrecombinant target molecules should remain as closed circular DNA which, due to its small size, migrates relatively rapidly in the gel. (ii) Unintegrated viral DNA, most of which is linear, should be cleaved into two or more subgenomic sized fragments, depending on which restriction enzyme is chosen. (iii) Recombinant molecules should be linearized,with the size of the recombinant fragment equal to the sum of the ϕX DNA and the terminal fragments of the MLV DNA (which are joined to the ϕX DNA at the ends of the LTRs). Finally, any circular viral DNA present in the mix should be detected by the presence of additional bands of novel size, but these will be smaller than the recombinant product. By choosing an appropriate restriction enzyme, the recombinant molecule will be larger than the other DNA fragments containing viral sequences. Since the number of recombinant molecules is relatively small, we detected them by hybridization to a ^{32}P-labled viral DNA probe after transfer from the gel to a nylon membrane. The probe was chosen to be specific for the LTR sequences, to avoid detecting internal fragments of viral DNA. Such an analysis showed that the conditions for generating recombinant molecules were the same as for the genetic assay, and that the integration events occurred at many positions in the ϕX target DNA.

Structural Analysis of the Initial DNA Product of the Integration Reaction

Recombinant molecules were purified from in vitro reactions by electrophoresing the restricted DNA in an agarose gel, followed by cutting out the region of the gel that contained the recombinants. The restriction enzyme used to cleaved the recombinant molecules before electrophoresis was chosen to cut very near the ends of both LTRs. Since we expected the initial strand-exchange product to be joined by only one strand at each end, the free end of the viral DNA was released by simply denaturing the purified recombinant molecule, and the denatured strands were electrophoresed in a 6% polyacrylamide gel containing 7 M urea. Since the viral DNA was restricted very near the ends of the LTRs, the size of the free viral strands (one from each end) were sufficiently small to allow precise identification of their size with single base resolution by comparison to an appropriate sequencing ladder run in adjacent lanes. The resolved DNA fragments were transferred from the gel to a nylon filter and detected by hybridization to strand-specific radioactive probes. An identical analysis was performed on unintegrated linear DNA, and the DNA fragments from the recombinant molecules and the unintegrated molecules were electrophoresed in adjacent lanes in the same gel to allow a direct comparison of the size of the fragments present at the ends of the viral DNA in both molecules.

Linear Viral DNA is the Immediate Precursor to the Integrated Provirus

By using probes specific for each strand of viral DNA at the ends of the LTRs, it was possible to determine both the polarity of the recombinant product (in other words, whether the 5′ or 3′ ends of the viral DNA were in fact free) and the precise size of that free strand (which allowed us to test the predictions of the two different pathways in which either linear viral DNA or the two LTR circle serve as immediate precursors). The results were very clear. The 5′ ends of each strand of the viral DNA are free in the recombinant molecule, and the size of the released fragments is consistent with models in which linear viral DNA is the direct precursor to the integrated provirus. The structure of the initial recombinant product as determined by this analysis is summarized as follows (again, see Figure 7 on p.99):

(i) The 5′ end of the plus strand of viral DNA is free at the left end of the recombinant molecule (as conventionally drawn) with its 3′ end joined to the target DNA at the end of the right LTR. The free 5′ end of the plus strand is detected only with a probe specific for the plus strand of U3 (the outermost sequences of the left LTR).

(ii) The 3′ end of the minus strand of viral DNA is joined to the target DNA at the end of the left LTR, and the 5′ end of the right LTR minus strand is free. The fragment corresponding to the 5′ end of the right LTR is detected only by a probe specific for the minus strand of U5 (the outermost sequences of the right LTR).

(iii) Analysis of both ends thus confirms that the polarity of strand-exchange is such that the 3′ end of each strand of the viral DNA is joined to target sequences, with the 5′ ends of the viral DNA remaining free as part of a gapped structure (which is predicted to result from staggered cleavage of the target DNA). Our analysis did not detect the 3′ ends of the viral DNA since their association with ϕX sequences cause a substantial reduction in electrophoretic mobility. We assume the 3′ ends are joined to the target sequences since the 5′ ends are not.

(iv) Finally, for each end of the viral DNA in the recombinant molecule, the size of the 5′ fragment was precisely that predicted if the linear viral DNA molecule is the immediate precursor to integration--with an extension of viral DNA sequences of only two bases instead of the four that would be predicted if the two LTR circle were the direct precursor.

Molecular analysis of the products of the in vitro reaction explained the surprising results we initially obtained using the genetic assay, which showed that integration activity correlated with the number of linear viral DNA molecules present in a reaction. It should be noted that we cannot eliminate the possibility that the initial recombinant contains a four base extension of viral sequences which is subsequently shortened by precisely two bases, but such an event seems extremely unlikely.

Structural Analysis of the Ends of the Unintegrated Linear Viral DNA Precursor

The results from an analysis of the ends of unintegrated linear viral DNA (using the same procedure just described for analyzing recombinants) revealed an unexpected result. The 5′ ends of the viral DNA were of the size predicted by reverse transcription priming events (with the two bases that are absent in proviral DNA being present). However, the 3′ ends were heterogeneous, with the most prominent species (roughly 90%) terminating two bases short of the 5′ ends (See Figure 2 on page 99). The remaining molecules either ended flush with or terminated one base short of the 5′ ends. The fact that most of the linear viral DNA, even in cytoplasmic extracts, is recessed two bases is of interest because the assumption that linear viral DNA is blunt-ended (see Chapter 1). The loss of two base pairs at each end of the proviral DNA is thought to result from cleavage of the blunt-ended unintegrated viral DNA at a position two bases removed the 3′ end of each strand. If such a cleavage occurs, our results suggest that it is not coupled to integration, as most molecules are already recessed before they have integrated. This issue is discussed in more detail below.

A Function for the MLV IN Protein

As a final experiment, we analyzed the ends of unintegrated linear viral DNA that was purified from cells which had been infected by an

integration-defective mutant of MLV (called SF2) that has a frameshift mutation in the IN coding sequences (Hagino-Yamagashi et al 1987). This mutant virus is capable of entering cells and reverse transcribing its genome, but is incapable of establishing a provirus. Unlike the wild-type virus, the 3′ ends of the unintegrated linear DNA made by SF2 are flush with the 5′ ends of the viral DNA, and no recessed molecules can be detected. Thus the viral IN protein is required for generating the recessed ends at the 3′ ends of unintegrated linear viral DNA, presumably by cleavage of an initially blunt-ended linear DNA molecule.

An alternative explanation for the behavior of the SF2 virus is that IN normally causes reverse transcription to terminate two bases short of full-length synthesis, which would eliminate the requirement for cleavage. However, one additional experiment supports the idea that linear viral DNA is initially synthesized as a full-length blunt-ended molecule. When the ends of unintegrated viral DNA from an unfractionated population of cytoplasmic molecules were analyzed, a larger fraction of the 3′ ends were unrecessed when compared to the 3′ ends of the gel-purified, full-length linear molecules. We presume that the other viral DNA molecules present in a cytoplasmic extract are earlier intermediates in replication. The enrichment for recessed ends in the full-length product implies a precursor product relationship where the 3′ end is initially flush with the 5′ end and is later processed to its recessed state. The ability of the ASLV IN proteins to cleave viral DNA in a manner consistent with such a model also supports this point of view (see Chapter 4).

The requirement for IN to generate a recessed linear molecule during an infection is sufficient to explain why it is required for integration, but the protein may also have additional roles in the subsequent steps of integration.

A New Model for the Integration of Retroviral DNA

Based upon the experiments summarized in this narrative, we propose that retroviral integration occurs in the following sequence of events: (i) Reverse transcription in the cytoplasm generates a blunt-ended linear DNA molecule. (ii) The viral IN protein cleaves the ends of the viral DNA to generate a two base recess at each 3′ end. (iii) This recessed linear molecule, upon gaining access to the host genome, is joined to chromosomal DNA in a strand-exchange reaction that is coupled to the staggered cleavage of the target site. (iv) Host repair process remove the protruding 2 bases of viral DNA at each 5′ end and fill in the gap generated by staggered cleavage of the target using the 3′ end of the target DNA as a primer for DNA synthesis. The results we have obtained which support this scenario are provided in detail in the two publications that constitute the remainder of this chapter.

REFERENCES

Alberts, B. 1984. The DNA enzymology of protein machines. *Cold Spring Harbor Symp. Quant. Biol.* 49:1-12.

Kriegler, M. and M. Botchan. 1983. Enhanced transformation by a simian virus 40 recombinant virus containing a Harvey murine sarcoma virus long terminal repeat. *Mol. Cell. Biol.* 3:325-339.

Hagino-Yamagishi, K., L. A. Donehower, and H. E. Varmus. 1987. Retroviral DNA integrated during infection by integration-deficient mutant of murine leukemia virus is oligomeric. *J. Virol.* 61:1964-1971.

Lobel, L. I., M. Patel, W. King, M. Nguyen-Huu, and S. P. Goff. 1985. Construction and recovery of viable retroviral genomes carrying a bacterial suppressor transfer RNA gene. *Science* 228:329-332.

Seed, B. 1983. Purification of genomic sequences from bacteriophage libraries by recombination and selection in vivo. *Nucl. Acids Res.* 11:2427-2445.

Cell, Vol. 49, 347–356, May 8, 1987, Copyright © 1987 by Cell Press

Correct Integration of Retroviral DNA In Vitro

Patrick O. Brown,*‡ Bruce Bowerman,† Harold E. Varmus,*† and J. Michael Bishop*†‡
*Departments of Microbiology and Immunology
University of California, San Francisco
San Francisco, California 94143
†Departments of Biochemistry and Biophysics
University of California, San Francisco
San Francisco, California 94143
‡G. W. Hooper Foundation
University of California, San Francisco
San Francisco, California 94143

Summary

We have developed a cell-free system for studying the integration of retroviral DNA. In our assay, amber mutations in a bacteriophage λ genome that serves as the target for integration are suppressed by integration of an MLV derivative that carries the E. coli *supF* gene. The structure of the reaction products is that expected from an authentic MLV integration reaction. Linear viral DNA from the cytoplasm of infected cells serves as a precursor, though not necessarily the immediate precursor, to the provirus integrated in vitro. The viral DNA in the infected cell appears to be tightly associated with the enzymatic machinery required for its integration. Supercoiling, chromatin structure, transcription, and replication are not required of the target DNA. Since no high-energy cofactor is necessary, the DNA breakage and joining steps in the integration reaction are probably coupled.

Introduction

Retroviruses replicate via a complex life cycle, during which the viral genome undergoes a remarkable series of information-conserving structural transformations (Varmus and Swanstrom, 1985). Two essential steps of this cycle are the synthesis of a DNA copy of the RNA viral genome and integration of this viral DNA into a chromosome of the infected cell. The resulting provirus can thereafter be transmitted and expressed as a stable genetic element of the host genome and serves as the template for the next generation of viral RNA. It is in part the ability to integrate viral DNA into the host chromosome that accounts for the capacity of retroviruses to cause tumors, in some cases by carrying a transduced cellular gene into the infected cell, in others by altering expression of a cellular proto-oncogene (Bishop, 1983).

Early in a retroviral infection, a DNA copy of the viral genome is synthesized in the cytoplasm by the viral reverse transcriptase. A series of template switches by the polymerase is required to produce a double-stranded linear DNA molecule with a terminal redundancy called the long terminal repeat (LTR). At least some of these linear DNA molecules enter the nucleus, where some are circularized by covalently joining their ends to produce a 2-LTR circle. The novel sequence created at the site of this ligation is commonly called the circle junction. DNA circles containing a single copy of the LTR (1-LTR circles) are also formed in the nucleus, perhaps by homologous recombination between the LTRs of the linear precursor (reviewed by Varmus and Swanstrom, 1985).

Little is known about the process by which retroviral integration occurs, in part because there has heretofore been no in vitro system for studying it. Restriction mapping and direct sequence analysis have shown that the provirus is colinear with the linear DNA precursor, except that it always lacks the 2 bp present at each end of the linear precursor (Hughes et al., 1978, 1981; Dhar et al., 1980; Shimotohno et al., 1980; Majors and Varmus, 1981). Work on the avian spleen necrosis virus (SNV) has provided evidence that a small sequence, found at the circle junction of the 2-LTR circle, is a site on the viral DNA at which the integration mechanism can act (Panganiban and Temin, 1983, 1984a). Similarly, for murine leukemia virus (MLV) there is genetic evidence that specific sequences at the termini of the LTRs (which, when ligated together, form the circle junction) are required in *cis* for correct integration (Colicelli and Goff, 1985). As in the transposition of both eukaryotic and prokaryotic transposable elements, integration of retroviral DNA is accompanied by a duplication of the target sequence, producing short direct repeats that flank the provirus (Shimotohno et al., 1980; Dhar et al., 1980; Hughes et al., 1981; Majors and Varmus, 1981). For most retroviruses, integration appears to occur with little or no target sequence specificity. Other important features of the target DNA remain undefined. For instance, it is conceivable that ongoing replication or transcription or a specific target DNA structure might be essential.

A variety of data suggest that viral DNA cannot serve as a substrate for normal retroviral integration unless its presence in the cell is established by infection. For example, introduction of naked viral DNA into a permissive host is not sufficient for legitimate proviral integration (Luciw et al., 1984). Even when all the viral proteins necessary for integration are provided by superinfection, plasmid DNA containing the MLV circle junction sequence, when introduced by means other than retroviral infection, does not integrate by a retroviral mechanism (Kriegler and Botchan, 1983; P. Brown, unpublished data). This apparent confinement of integration activity to authentic viral reverse transcripts is consistent with a model in which subviral structures maintain a high degree of order during replication. In particular, the product of the 3′ end of the viral *pol* gene, which has been shown to be required for integration (Donehower and Varmus, 1984; Schwartzberg et al., 1984; Panganiban and Temin, 1984b), might be active only on DNA in a specific subviral nucleoprotein complex. Indeed, most of the unintegrated viral DNA produced by reverse transcription in MLV-infected cells can be isolated in stable fast-sedimenting nucleoprotein particles that also contain virally encoded proteins (B. Bowerman,

unpublished data). No host cell functions have yet been identified as essential for integration.

There is accumulating evidence for a close similarity between retroviral integration and the intracellular transposition of retrotransposons, a class of eukaryotic transposable elements that includes the yeast Ty elements and the copia family of elements in Drosophila (Varmus, 1983; Boeke et al., 1985; Shiba and Saigo, 1983; Flavell, 1984). Compared to retroviral integration after a synchronized infection, however, transposition of these endogenous elements is a relatively infrequent event. Retroviral integration is therefore an attractive experimental model for this closely related process.

We report here our development of the first in vitro system for studying retroviral integration and its initial application to defining some key features of MLV integration.

Results

Development of an In Vitro Assay for Retroviral Integration

We anticipated that an in vitro reaction would initially give a low yield of integration products. We therefore designed an assay that exploits a powerful genetic selection in E. coli (Seed, 1983) to identify these rare events (Figure 1). The system uses the presence in λ gtWES of amber mutations in three genes that are required for lytic growth (Leder et al., 1977). As a consequence, the phage is unable to make plaques on a lawn of wild-type (*sup*^{0}) E. coli but makes plaques with normal efficiency if the *supF* amber suppressor allele is present. MLV*supF* is a fully replication-competent Moloney MLV derivative that carries the E. coli *supF* gene in its U3 region (Lobel et al., 1985). We reasoned that integration of an MLV*supF* provirus into a nonessential region of λ gtWES would produce a recombinant λ genome able to suppress its own amber mutations and thus to make plaques on a nonsuppressor strain of E. coli.

In designing the assay, we were influenced by evidence suggesting that an intact subviral particle, containing the viral DNA, might be crucial for integration. We therefore chose to provide the MLV*supF* DNA in its native state, as a nucleoprotein complex isolated from acutely infected cells, rather than as a purified DNA species. Mouse NIH 3T3 cells were infected at high multiplicity with MLV*supF* by cocultivation with a virus-producing cell line. At a time after infection when proviral integration is known to be occurring, the cells were harvested and cell-free nuclear and cytoplasmic extracts prepared. The extracts contained unintegrated MLV*supF* DNA and the enzymatic machinery required for its integration.

The cellular extracts were mixed with concatemerized λ gtWES DNA in a defined solution and incubated to allow integration to occur. After extensive digestion with proteinase K, the DNA was recovered and packaged into phage particles in vitro.

Overall recovery of the target DNA—recombinant as well as unreacted—was determined by plating on the *supF* E. coli strain KM392. To score for integration of the MLV*supF* provirus, the phage particles were plated on the *sup*^{0} lacZ^{am} strain CC114. The *supF* gene in the MLV provirus allows the recombinants specifically to make plaques on this strain. Because the *supF* gene also suppresses the amber mutation in the *lacZ* gene, recombinant plaques appear blue on our indicator plates (Figure 2). As detailed below, all such blue plaques thus far analyzed arose from λ gtWES phage carrying correctly integrated MLV*supF* proviruses. The total number of integration events in each reaction was estimated by correcting the number of blue plaques for efficiency of DNA recovery and packaging and for the inability to recover viable phage from integration events that inactivate genes essential for lytic growth (see Experimental Procedures). In a typical experiment, an extract from 4×10^6 cells, containing 7×10^7 unintegrated viral DNA molecules, when incubated with 1.4 μg (3.1×10^{10} genome equivalents) of target DNA, yielded 4×10^3 blue plaques on CC114, and 2×10^8 total plaques on KM392. This corresponds to an estimated total of 2×10^5 integration events. That is, about 0.3% of the total pool of unintegrated viral DNA present in the extract integrated productively in this in vitro reaction. Conversely, about 1 in 10^5 of the λ gtWES genomes in the mixture acquired an MLV provirus.

Integration In Vitro Produces a Provirus of Normal Structure

By design, each plaque counted in our assay represents a single recombinant clone. These can be readily amplified and their DNA analyzed. We have so far analyzed 33 clones by restriction mapping and Southern blotting. Because the complete sequences of both MLV (Shinnick et al., 1981) and λ (Daniels et al., 1983) are known, the structure of each recombinant clone can be deduced to a resolution of about 100 bp by this approach. The principal conclusions from this analysis are as follows. All clones that we analyzed contain MLV*supF* sequences. The MLV inserts terminate at or near the ends of the LTRs. No λ sequences were detectably gained, lost, or rearranged during integration. Most of the inserts appear to be complete intact proviruses; a few have internal deletions, but are terminated at each end by normal LTR sequences. Proviral integrations occurred throughout the region of the λ gtWES genome known to be nonessential for lytic growth, without any distinct target site or orientation specificity (Figure 3).

We extended this analysis by sequencing the junctions between λ and proviral DNA from seven of these clones (Figure 4). In each case, the sequences confirmed that the in vitro integrated proviruses have the structural characteristics of a normal provirus. That is, each provirus lacks the terminal 2 bp thought to be present at each end of the unintegrated linear precursor but is otherwise colinear with the precursor, and it is flanked by a precise 4 bp duplication of the sequence at the target site. The target DNA is otherwise unmodified and unrearranged.

Subcellular Localization and Structure of the Active Precursors for Integration In Vitro

In vivo, proviral integration takes place in the nucleus. Furthermore, results from experiments with SNV support the

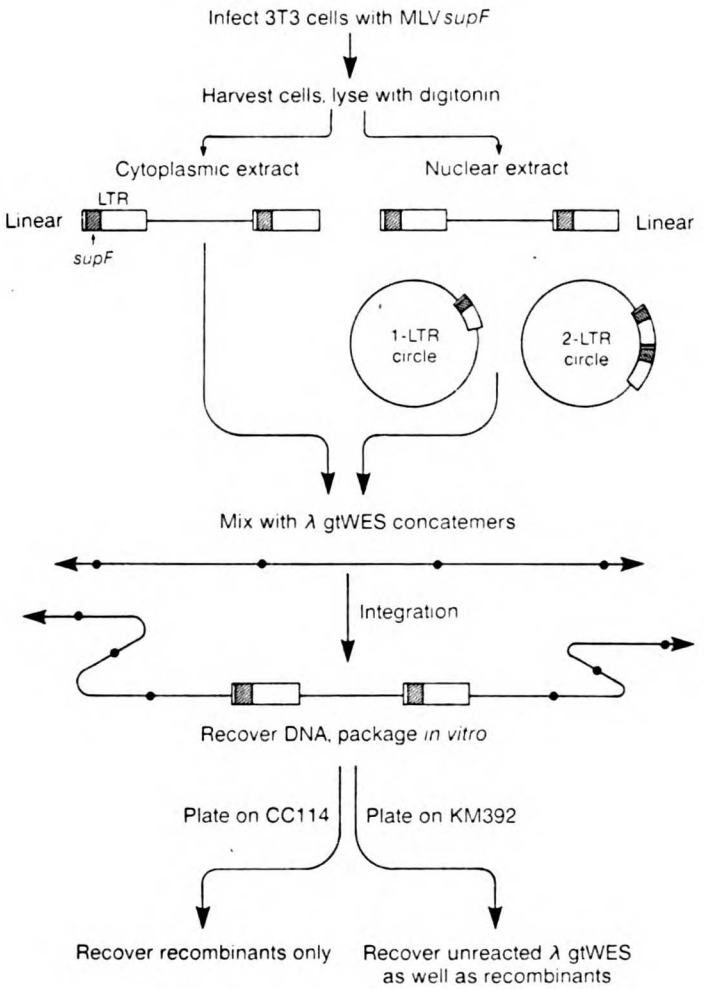


Figure 1. Flowchart of the In Vitro Integration Protocol

The *supF* gene is indicated by the shaded area in the LTR of MLV*supF*. As discussed in Results, the cytoplasmic extract contains only the linear form of viral DNA, whereas the nuclear extract contains the 1-LTR and 2-LTR circular forms as well as the linear form.

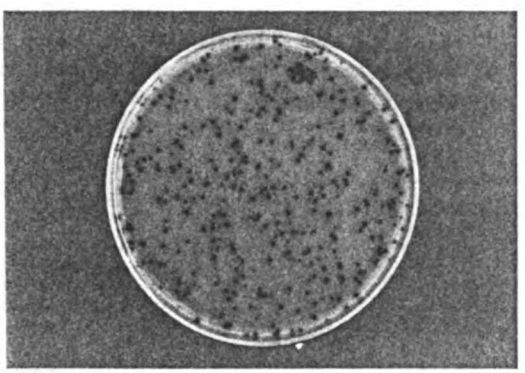


Figure 2. Assay for Integration of MLV DNA In Vitro

Integration of an MLV*supF* provirus into the λ gtWES DNA provides a suppressor for the three amber mutations in the target phage genome. This allows the recombinant phage to make plaques on the normally restrictive *sup*^{0} host. The plaques are blue on this indicator plate because an amber mutation in the host *lacZ* gene is also suppressed by the *supF* gene in the provirus.

idea that the 2-LTR circular form of unintegrated viral DNA is the immediate precursor to the integrated provirus (Panganiban and Temin, 1984a). Circular viral DNA has been detected only in the nucleus of infected cells (Guntaka et al., 1976; Varmus and Swanstrom, 1985). We therefore expected that the nuclear extract would be more active than the cytoplasmic extract in our assay. We were surprised to find that, in fact, the cytoplasmic extract was more active (Table 1). In mixing experiments (Table 2), supplementation of the nuclear extract from MLV*supF*-infected cells with cytoplasmic extract from uninfected cells failed to enhance the activity of the nuclear extract. Nor did nuclear extract from uninfected cells inhibit the activity of cytoplasmic extract from MLV*supF*-infected cells. These results argue that the greater activity of the cytoplasmic extract was not simply due to the presence of stimulatory factors in the cytoplasm or inhibitory factors in the nuclear extract.

The simplest explanation for this result is that it follows from the greater abundance of unintegrated viral DNA molecules in the cytoplasm than in the nucleus. As shown in Table 1, integration activity is proportional to the quan-

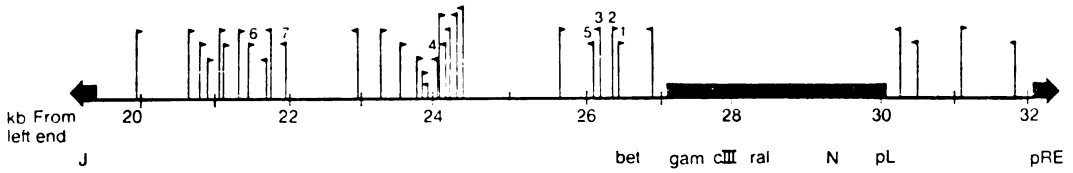


Figure 3. Sites and Orientation of MLV Proviruses Integrated In Vitro

Flags indicate orientation of proviruses, and point from 5' to 3' relative to the MLV (+) strand. Numbered flags correspond to proviruses whose junctions with λ DNA were sequenced (Figure 4). The heavy lines indicate regions essential for λ growth on a *recA*⁻ host. The positions of several λ genes are indicated as landmarks.

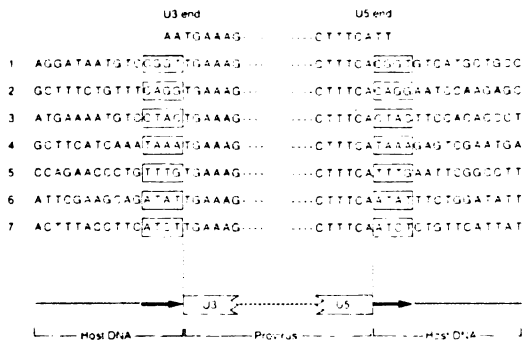


Figure 4. Sequences at the λ-MLV Junctions

The sequences shown correspond to the plus strand of the MLV DNA. The sequences of the ends of the unintegrated linear form of MLV DNA are shown at the top, for reference. The boxes enclose the 4 bp sequence, present in a single copy in the target DNA, that was duplicated upon MLV integration. Note that in the integrated provirus, the terminal 2 bp of the linear precursor have been lost.

tity of the linear form of viral DNA in the extracts, rather than the quantity of either the 1-LTR or 2-LTR circles. In fact, no circular forms of viral DNA were detected in the cytoplasmic extract by Southern blotting (Table 1 and Figure 5). Calibration of the sensitivity of our Southern blots using cloned MLV DNA showed that we could have detected as few as 4 × 10^5 circular DNA molecules (Figure 5). Even if circular molecules were present at this level in the cytoplasmic extract, they would have had to integrate with an efficiency of greater than 60% to account for the observed number of recombinants (Table 1). This seems highly unlikely, given the much lower integration efficiency—no more than 4%—calculated for the circular forms in the nuclear extract, assuming that they are able to integrate (Table 1). We therefore conclude that the linear form of viral DNA can serve as a precursor for integration in vitro.

Requirements and Optimal Conditions for Integration

A time course of the integration reaction (Figure 6) demonstrates that the yield of recombinant products increases progressively with continued incubation for about 1 hr. The apparent decline in the rate of the reaction with time is probably at least partly due to thermolability of the activity, since activity diminishes appreciably when the cellular extract is preincubated at 37°C (data not shown).

Table 3 presents the results of our preliminary characterization of the conditions that affect MLV integration in vitro. EDTA abolished activity, implying that a divalent cation is essential; we have not yet tested other divalent cations in place of Mg^{++}. The optimal K^+ concentration was about 150 mM; activity was decreased very little at 75 mM KCl, and was about 10-fold lower at a KCl concentration of 300 mM. Polyethylene glycol at 5% stimulated activity 4-fold, but was not required. The pH optimum was broad, extending at least from pH 6.6 to pH 8.6. Roughly equivalent activity was observed at an incubation temperature of 30°C or 37°C, but no detectable integration occurred in a 1 hr incubation at 0°C. The yield of recombinants was not affected by heating the DNA recovered from a reaction to 68°C for 10 min.

Table 1. Comparison of Nuclear and Cytoplasmic Extracts

	Cytoplasmic Extract	Nuclear Extract
Linear DNA molecules	8 × 10^7	2 × 10^7
2-LTR circles	None detected	1.7 × 10^6
1-LTR circles	None detected	2.6 × 10^6
Pfu on KM392	4.4 × 10^8	1.4 × 10^8
Pfu on CC114	5.7 × 10^3	4.7 × 10^2
Recombinants	2.5 × 10^5	6.7 × 10^4
Recombinants ÷ linear molecule	3.1 × 10^{-3}	3.3 × 10^{-3}
Recombinants ÷ 2-LTR circle	>6.2 × 10^{-1}	3.9 × 10^{-2}
Recombinants ÷ 1-LTR circle	>6.2 × 10^{-1}	2.6 × 10^{-2}

Nuclear and cytoplasmic extracts (100 μl) were prepared from 4 × 10^6 cells, assayed for unintegrated forms of viral DNA (as shown in Figure 5), and used for integration reactions as described in Experimental Procedures. Calculation of the number of recombinants per reaction was as described in Experimental Procedures. In the calculations for this table, the correction factor for packaging and plating efficiency (designated "E" in the formula shown in Experimental Procedures) was 0.13.

Table 2. Test for Interaction between Nuclear and Cytoplasmic Components

Source of Extracts	Pfu on KM392	Pfu on CC114	Recombinants	%
50 μl Infected Cytoplasm + 50 μl Uninfected Cytoplasm	4.2×10^8	2601	1.1×10^5	100
50 μl Infected Cytoplasm + 50 μl Uninfected Nuclei	3.4×10^8	4789	2.4×10^5	227
50 μl Infected Nuclei + 50 μl Uninfected Nuclei	1.3×10^8	305	4.0×10^4	36
50 μl Infected Nuclei + 50 μl Uninfected Cytoplasm	2.1×10^8	418	3.4×10^4	31

Preparation of extracts, integation reactions, and calculation of the number of recombinants per reaction were as described in Experimental Procedures, except that uninfected cellular extracts were prepared from NIH 3T3 cells that were cultivated in the absence of MLV*supF*-producer cells. For each reaction, extracts were mixed as indicated before addition to the remainder of the reaction mixture. In the right-hand column, results are expressed as a percent of the number of recombinants obtained with cytoplasmic extracts alone.

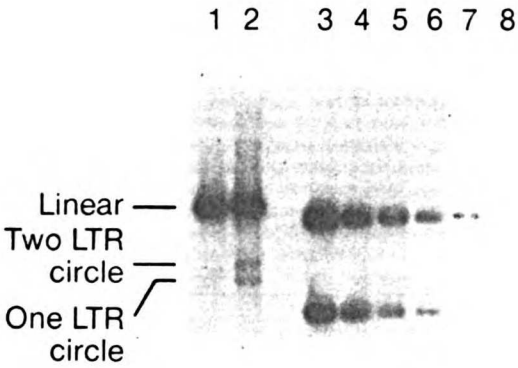


Figure 5. Quantity and Structure of Viral DNA in Nuclear and Cytoplasmic Extracts

Viral DNA was prepared from nuclear and cytoplasmic extracts of NIH 3T3 cells that were acutely infected by cocultivation with an MLV*supF* producer line. DNA samples were resolved by agarose gel electrophoresis, then transferred to a nitrocellulose filter. MLV DNA was identified by hybridization, using as a probe the plasmid p8.2 (a complete, permuted clone of the MLV 1-LTR circle), labeled by nick translation. Lanes: (1) viral DNA from 10 μl of cytoplasmic extract; (2) viral DNA from 40 μl of nuclear extract; (3–8) 160, 80, 40, 20, 10, and 5 pg, respectively, of p8.2 plasmid DNA digested with HindIII to generate 8.2 and 4.3 kb fragments as standards for quantitation. Note that 1 pg of DNA is equivalent to 1×10^5 copies of full-length viral DNA. The intensities of the viral DNA bands were compared to the standards in lanes 3–8. The results, adjusted to represent an extract volume of 100 μl, are presented in Table 1. In the gel shown here, viral DNA was overloaded with respect to linears to facilitate quantitation of the circular forms.

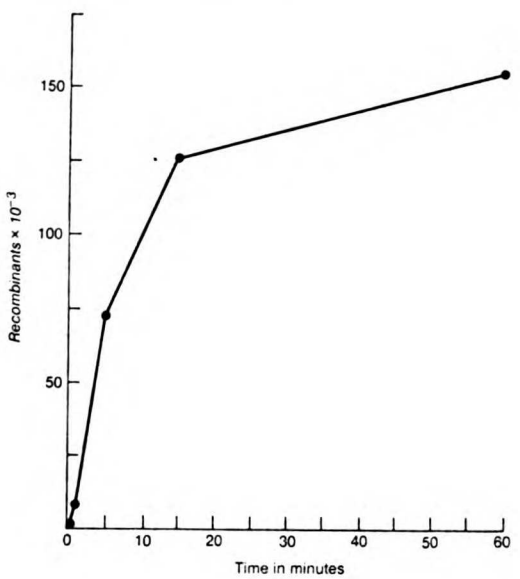


Figure 6. Time Course of Integration In Vitro

Reaction mixtures were prepared using standard conditions as described in Experimental Procedures, incubated at 37°C for the times indicated, then terminated by addition of EDTA and proteinase K.

Pretreatment of the cell extract for 5 min at 37°C with SDS and proteinase K or with 5 mM N-ethylmaleimide abolished activity, whereas pretreatment with RNAase A had little effect. Thus protein, but probably not RNA, is required for activity.

Assessment of the Nucleotide Requirements for Activity and Partial Purification of the Active Nucleoprotein Complex

To assess the role of nucleoside triphosphates in integration, we used Bio-gel A5m (Bio-Rad) gel-exclusion chromatography to remove the endogenous triphosphates from our cell extracts. Using a luciferase assay (Cheer et al., 1974), we determined that the residual ATP concentration in the column-purified extract was about 1.5×10^{-9} M. To further assess the completeness of nucleotide removal, [^{32}P]dCTP was mixed with the extract before loading on the column. Less than one part in 5×10^5 was detectable in the active, excluded volume pool. This degree of purification also translates into an ATP concentration of about 1.5×10^{-9} M, yet no additional ATP was required for integration when this pool was assayed (Table 3). Furthermore, no other nucleoside triphosphate (NTP) or deoxynucleoside triphosphate (dNTP) was required for integration when this fractionated extract was used.

Gel-exclusion chromatography through a Bio-gel A5m column also provided a substantial purification of the in-

Table 3. Conditions Affecting MLV Integration

Condition	% of Standard Activity
Standard	100
−Mg, +EDTA	<0.2
75 mM KCl	89
300 mM KCl	1
−PEG	24
pH 6.6, PIPES	119
pH 8.6, Tris	83
30°C Incubation	107
0°C Incubation	<0.2
+Proteinase K, SDS	<0.1
+5 mM N-ethylmaleimide	<2
+RNAase A	78
Heated to 68°C after reaction	93
−ATP, rNTP, creatine phosphate	216
−ATP, rNTP, dNTP, creatine phosphate	200
+pSPCJ, 10^{10} copies	78
Naked MLV*supF* DNA	
+wt MLV-infected cell extract	0.04[a]
MLV*supF* producer cell extract alone	<0.02

Except for the indicated modifications, reaction conditions were as described in Experimental Procedures. For most of the conditions tested, reactions were performed in duplicate or triplicate, in parallel with duplicate or triplicate reactions using standard conditions. The number of recombinant plaques scored from each set of standard reactions ranged from 338 to 6272, and except where the yield was less than 2% of standard, between 66 and 7340 recombinant plaques were scored from each of the reactions using modified conditions.

[a] 1 recombinant plaque was observed, not yet structurally characterized.

tegration activity away from most of the proteins in the extract, giving a roughly 100-fold increase in specific activity. All the recovered integration activity (at least 10% of the activity loaded) chromatographed with the viral DNA in the excluded volume, corresponding to a molecular weight greater than 5×10^6. This finding supports the notion that the machinery required for retroviral integration is normally stably associated with the DNA to be integrated, perhaps as a subviral particle (B. Bowerman et al., unpublished data).

Such a model is also supported by the apparent inability of the putative integration machinery to interact with analogues of the unintegrated viral DNA (Table 3). Naked MLV*supF* DNA, isolated from either the cytoplasm or nucleus of infected cells, was not efficiently integrated even when mixed with cytoplasmic extract from wild-type MLV-infected cells. Furthermore, there was little or no inhibition of integration by up to 10^{10} copies of a plasmid, pSPCJ, containing the MLV circle junction sequence.

Discussion

We have developed an extremely sensitive in vitro system for studying retroviral integration. Starting with 3×10^{10} target genomes of bacteriophage λ DNA, we can generally detect a signal (i.e., blue plaques) from any reaction that yields 100 or more MLV λ recombinants. The efficiency of the in vitro reaction under standard conditions is such that between 0.2% and 1% of the retroviral DNA molecules in a cellular extract become integrated into λ DNA during a 1 hr incubation. Thus a typical experiment starting with an extract from 4×10^6 cells, containing about 5×10^7 viral DNA molecules, yields more than 10^5 recombinants, of which over 1000 can be recovered as blue plaques. Our assay is highly specific. Under our standard reaction conditions, the only recombination event we have seen giving rise to a blue plaque in this assay produces a provirus whose structure accurately duplicates that of an in vivo provirus.

The molecular structure of the junction between an MLV provirus and host DNA is distinctive (Dhar et al., 1980; Shoemaker et al., 1981). All of the in vitro recombinants that we analyzed showed this same characteristic structure. It seems quite unlikely that an artifactual mechanism could account for this consistent pattern. We therefore conclude that retrovirus integration in our in vitro system accurately mimics the in vivo integration mechanism.

The yield of recombinant products increases progressively during incubation of λ gtWES DNA with the active cell extract. When the cell extract is pretreated with SDS and proteinase K or N-ethylmaleimide, or if the incubation is performed on ice or in the presence of EDTA, integration activity is abolished. Thus the recombination event occurs during the incubation of the λ gtWES DNA target with the cell extract and is dependent upon protein and a divalent cation. The recombinant product is stable to extensive proteolysis, followed by phenol and chloroform extraction and heating to 68°C in water for 10 min. The junction formed between provirus and λ DNA during the incubation with cell extract is therefore protein-independent and probably covalent, presumably a normal 3′–5′ phosphodiester bond.

Characteristics of the DNA Target for Proviral Integration

Since naked, linear (relaxed) DNA could act as a target for integration, using an extract that contained no detectable histones (data not shown), the DNA target for integration need be neither supercoiled nor assembled as chromatin. It remains possible, however, that a supercoiled or otherwise modified DNA molecule might be a preferred target for integration. The lack of a requirement for dNTPs or rNTPs implies that neither transcription nor extensive DNA replication per se is required for covalent linkage of the provirus to the target DNA. Clearly, generation of the flanking 4 bp repeat requires DNA synthesis, but formation of a stable linkage between provirus and target DNA does not. Thus we propose that the initial covalent linkage at each junction is on one strand only and that its formation is not intrinsically coupled with DNA synthesis (Figure 7). The DNA synthesis required to produce the mature double-stranded structure of the flanking repeat can presumably take place in E. coli after infection with the initial recombinant phage.

It is interesting that experiments in vivo show that integration occurs at least preferentially in cells that are actively replicating their DNA (Varmus et al., 1977). Perhaps, in the cell, replication physically exposes the DNA to the integration machinery—unnecessary with our naked DNA target. Alternatively, factors present only during S phase may be required. Since our extracts from unsynchronized

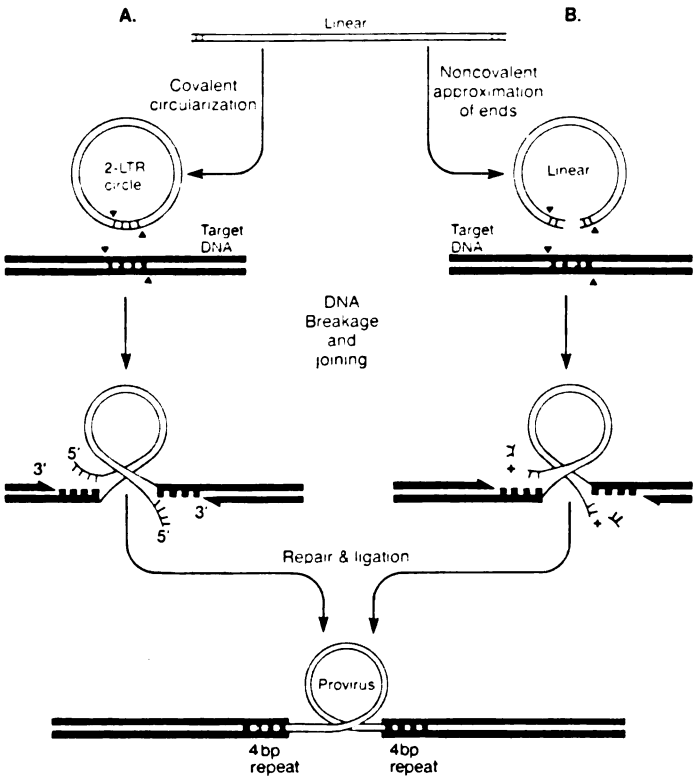


Figure 7. Two Possible Pathways for Integration of Linear Retroviral DNA

In pathway A, integration proceeds via a 2-LTR circular intermediate, formed by ligation of the ends of the linear form (Shoemaker et al., 1980). In pathway B, the linear viral DNA is itself the substrate for integration. In the DNA breakage and joining step, viral and target DNA strands are broken at the sites marked by the arrowheads. The resulting 3' ends of the viral DNA are then joined to the corresponding 5' ends of the target DNA. If the enzymatic machinery that catalyzes integration recognizes only sequences proximal to its cleavage site on the viral DNA, then it might not readily distinguish the linear form from the 2-LTR circle. Note that the final step, DNA synthesis to repair the gap or heteroduplex flanking the provirus and thus generate the 4 bp repeat, could take place in E. coli after infection with the gapped precursor. The 4 bp of viral DNA that are lost upon integration (two from each end of the linear molecule) and the 4 bp target sequence that is duplicated upon integration are indicated by the line segments perpendicular to the DNA strands.

cells are competent for integration, the apparent requirement for DNA synthesis in vivo cannot be a reflection of inhibitory factors specific to other phases of the cell cycle.

The in vitro reaction showed no apparent target sequence specificity. No consistent features could be identified in the sequences surrounding the integration sites. Furthermore, the integration sites appear to be randomly distributed throughout the dispensable portion of the λ gtWES genome. This observation is similar to the results others have reported from analysis of in vivo proviruses (Varmus, 1983). However, it has been suggested that, in vivo, retroviral DNA preferentially integrates in the upstream portion of actively transcribed genes or near DNAase-hypersensitive sites (Rohdewohld et al., 1987). Moreover, there is evidence that not all regions of the genome are equally favored for integration in vivo (King et al., 1985). The bias in target-site selection in vivo may reflect differences in accessibility of DNA in vivo, not duplicated with the naked target DNA in our in vitro system. Alternatively, there may be a subtle sequence selectivity not detected in our analysis. The possibility that nicks or gaps or other modifications of the target DNA are required for integration remains to be investigated.

The Precursor to the Integrated Provirus

Only the linear form of viral DNA was detectable in the cytoplasm, yet the cytoplasm was a better source of active proviral precursors than the nucleus. Furthermore, integration activity in various cell extracts paralleled the quantity of linear viral DNA in the extracts, rather than the quantity of either of the circular forms. Thus the linear MLV DNA appears to be active as a precursor for integration in vitro, just as it is in vivo. We have not, however, determined whether the linear DNA structure is the immediate precursor to the integrated provirus, or whether it must first be circularized.

The results of Panganiban and Temin (1984a) with SNV would imply that integration can occur in vivo via a 2-LTR circular intermediate. Since circular forms are found only in the nucleus (Figure 5 and Table 1), a factor or condition restricted to the nucleus may be required for circularization to occur. However, in our in vitro system, integration was efficiently carried out by the cytoplasmic extract and was not greatly stimulated by addition of nuclear extract. A simple explanation for this result would be that circularization of the viral DNA is not required for integration in our system. Alternatively, if circularization is required, our findings could indicate contamination of the cytoplasmic extract with a small but sufficient amount of nuclear contents. It is also possible that the apparent requirement for entry into the nucleus before circularization can occur might actually reflect a requirement for target DNA as a cofactor in this process. Such a requirement would be satisfied in our in vitro system, without the need to supply nuclear extract. Experiments are in progress to determine whether circularization of the linear DNA occurs in our in vitro system.

What Is the Source of Energy for Formation of the New Covalent Bonds in Integration?

Integration necessarily depends on the formation of new phosphodiester bonds linking the provirus to the host DNA. Many precedents can be found among simple recombination mechanisms for a concerted breakage and rejoining reaction, requiring no external energy source (e.g., Reed and Grindley, 1981; Vetter et al., 1983). However, formation of the bacteriophage Mu transposition intermediate—of the biochemically characterized recombination reactions, perhaps the most similar to retroviral integration—appears to require ATP (Mizuuchi, 1983).

Our results strongly suggest that MLV integration does not require added ATP or indeed any other energy-providing cofactor. Phosphodiester bond exchange via a protein–DNA intermediate in a concerted DNA breakage and joining event could obviate any requirement for an external energy source in integration per se (Reed and Grindley, 1981; Craig and Nash, 1983; Gronostajski and Sadowski, 1985). We therefore favor this concerted mechanism for integration over one involving uncoupled DNA cleavage and ligation steps.

The source of energy for formation of the new bonds when the linear viral DNA is converted into a 2-LTR circle is less apparent, since in this conversion there is a net increase in phosphodiester bonds (Figure 7A). Thus if the linear viral DNA can be integrated only via a 2-LTR circular intermediate, we have a paradox. We are examining four hypotheses that might resolve the paradox. First, the linear DNA rather than the 2-LTR circle could be the direct precursor to the integrated MLV provirus (Figure 7B). This may seem unlikely in view of the previously cited results of Panganiban and Temin (1984a), but it is worth noting that their results do not exclude the possibility that the linear form might be a direct precursor. Indeed, as these two forms differ only in the two bonds directly at the circle junction, they might not be distinguished by the integration machinery. Second, the linear DNA precursor might not be flush-ended as previously postulated, but might bear single-strand extensions that are lost by phosphodiester bond exchange in forming the 2-LTR circle. Third, the ends of the linear DNA, as isolated in our cell extracts, could already be "charged" by a high-energy bond, for example, by adenylation or by a protein–DNA bond. Fourth, the energy for ligation could be stored in a preformed ligase–adenylate intermediate, as has been demonstrated for other DNA ligase reactions (Weiss and Richardson, 1967). Our gel-exclusion chromatography results imply that such a protein, if it exists, would have to be intrinsic to the nucleoprotein complex that contains the viral DNA.

Applications

The availability of the in vitro system described here allows us now to begin to isolate and purify the components required for retroviral integration. We are currently investigating the structure of a subviral nucleoprotein complex that appears to be an active intermediate in MLV integration (B. Bowerman et al., unpublished data). We have already been able to obtain a 100-fold purification (in activity per mg of protein) by gel-exclusion chromatography. As an alternative approach to identifying or confirming the identity of essential activities, antibodies or chemical inhibitors directed against defined viral or cellular proteins can be tested for their effects on integration in vitro.

Conversely, our assay can be used to screen for agents that specifically block the integration (or circularization) of retroviral DNA. Inhibitory factors of interest include the cellular components that mediate *Fv-1* restriction of MLV host range (Jolicoeur, 1979) or interferon-induced blockade of provirus establishment (Morris and Burke, 1979), as well as synthetic compounds. Indeed, application of this assay to pathogenic retroviruses—for instance, the human immunodeficiency virus (HIV)—may be a useful approach to screening for a new class of antiviral agents.

We are currently working to develop a direct, physical assay for retroviral integration. Such an assay would be more rapid and economical than the one we have described here and, more importantly, would allow the direct detection of intermediates in the integration process. A physical assay appears quite feasible, at least for MLV, given the high efficiency of integration in vitro.

Experimental Procedures

Strains

E. coli KM392 (*supF, hsdR*$^-$, *hsdM*$^+$), a *lac*$^-$ derivative of LE392, was obtained from T. St. John. E. coli CC114 (*sup*^{0}, *lacZ*^{am}, *recA*$^-$, *hsdR*$^-$, *hsdM*$^+$) and MLV*supF* (called MLVin31sulII in Lobel et al., 1985) were provided by Steve Goff.

Target DNA

λ gtWES phage was purified by banding in a CsCl density gradient, and DNA was isolated as described in Maniatis et al. (1982). DNA was concatemerized by ligation at 37°C overnight at a DNA concentration of 150–200 μg/ml, using T4 DNA ligase (IBI) in the reaction mix provided by the manufacturer. The reactions were stopped with EDTA and heated to 68°C for 15 min to inactivate the ligase. For experiments in which ATP-free DNA was required, E. coli DNA ligase (NEB) was substituted, and the ligation mix was as recommended by the manufacturer. Alternatively, ATP was removed (to <1 nM), after ligation with T4 DNA ligase, by dialysis, followed by phenol extraction and ethanol precipitation.

Cellular Extracts

Cells (10^6) of an MLV*supF*-producing NIH 3T3 line were plated together with 4 × 10^6 uninfected NIH 3T3 cells in a 100 mm dish, in DMEM with 10% fetal calf serum and 8 μg/ml polybrene. Twenty-four hours after plating, the cells were harvested by trypsinization, washed once with buffer A (10 mM Tris-HCl, pH 7.4; 225 mM KCl; 5 mM $MgCl_2$; 1 mM DTT; 20 μg/ml aprotinin), and then lysed in 250 μl per dish of buffer A + 0.025% digitonin (from a 5% stock in DMSO). The lysate was centrifuged at 1000 × g for 3 min. The nuclear pellet from this spin was resuspended in 250 μl per dish of buffer A, gently broken in a ball-bearing homogenizer, and then centrifuged at 8000 rpm in a JA-20 rotor for 20 min to pellet chromatin. The supernatant, designated "nuclear extract," contained unintegrated linear and circular viral DNA, and was largely free of chromosomal DNA.

The supernatant from the initial low-speed spin of the cell lysate was centrifuged at 8000 rpm in a JA-20 rotor for 20 min. The supernatant from this spin, designated "cytoplasmic extract," contained full-length linear viral DNA as well as subgenomic intermediates in DNA synthesis. Extracts retained activity for at least several weeks when adjusted to 8% sucrose, frozen in liquid nitrogen, and stored at −70°C.

For some experiments, gel-exclusion column chromatography was used to remove ATP and other small molecules from the extract. Two milliliters of extract was loaded onto a 40 ml (20 cm) column of Biogel A5m (Bio-Rad), equilibrated with buffer A, and chromatographed at a

flow rate of 0.5 ml/min. The excluded volume absorbance peak was pooled. ATP in the extract was assayed using firefly lantern extract (Sigma), as described in Cheer et al. (1974).

Integration Reactions

Integration reactions were carried out in a total volume of 150 μl, in 1.5 ml microfuge tubes, and contained, except where noted (including the components of buffer A, provided by the cell extract), 100 μl cell extract (cytoplasmic or nuclear); 1.4 μg concatemerized λ gtWES DNA; 20 mM HEPES, pH 7.6; 6 mM Tris–HCl, pH 7.4; 150 mM KCl; 5 mM $MgCl_2$; 1.25 mM ATP; 50 μM each of UTP, CTP, and GTP; 100 μM each of dATP, dGTP, dCTP, and dTTP; 10 mM creatine phosphate; 1 mM DTT; 13 μg/ml aprotinin; and 5% polyethylene glycol 8000. The order of addition was the following: 15 μl 10× cocktail, 10 μl target DNA solution, 100 μl cell extract, mixed well; and 25 μl 30% polyethylene glycol. Reactions were incubated at 37°C for 1 hr, then stopped by addition of 12 μl of a solution of 250 mM EDTA, 0.5% proteinase K.

Scoring Integration

After incubation, reaction mixtures were subjected to proteolysis at 37°C for 24 hr, then diluted with 160 μl of 0.4% SDS, incubated at 37°C for an additional 24 hr, and finally diluted with 160 μl of a solution containing 10 mM Tris–HCl (pH 7.4), 1 mM EDTA, and 600 mM NaCl. Extractions were performed gently with $CHCl_3$, then with phenol, then twice with $CHCl_3$:phenol::1:1, then with $CHCl_3$. The DNA was then recovered by ethanol precipitation, washed twice with 70% ethanol, resuspended overnight in 4 μl water, and packaged in vitro into phage particles, using Gigapack Gold packaging extract (Stratagene).

To select for provirus-containing recombinants, the phage particles were plated on a lawn of CC114. Recombinant phage containing the MLV*supF* provirus produce blue plaques on IPTG/X-Gal plates. Since MLV DNA has an unmodified *EcoK* restriction site, the plating efficiency for the recombinants on the *hsdR*$^-$ strain CC114 is about three times higher than on the *hsdR*$^+$ strain LG75. DNA recovery and packaging efficiency were assessed by plating on a lawn of the *supF* strain KM392. The absolute number of recombinants produced in each reaction was calculated as (A/B) × (C/D) × E, where A = the number of blue plaques on CC114, B = the number of plaques on KM392, C = the number of λ gtWES genomes in 1.4 μg of DNA (3.1×10^{10}), D = the fraction of the λ gtWES genomes that can tolerate an insertion (estimated by examining the genetic map of λ gtWES to be 0.27). To control for differences in packaging and plating efficiencies between the provirus-containing recombinant phage genomes (50 kb) and unrecombined λ gtWES (40 kb), defined amounts of DNA from λ gtWES and an MLV*supF*-λ gtWES recombinant were ligated together, then packaged and plated in parallel with experimental samples. E = the ratio between the efficiencies with which unreacted λ gtWES and the recombinants were recovered as plaques.

DNA Analysis

Plaque amplification, λ DNA isolation, Southern blotting, and subcloning were performed using standard methods (Maniatis et al., 1982). Filter hybridizations followed the method of Church and Gilbert (1984). With repeated amplification of the provirus-containing recombinant phage, deletion derivatives are recovered in which the unique portion of the provirus is lost by recombination between the LTRs. Sequence analysis was by the dideoxy method (Sanger et al., 1977).

Acknowledgments

We thank Larry Donehower and Kimiko Hagino-Yamagishi for their generous advice, Sue Klapholz for her helpful comments on the manuscript, and Steve Goff for providing plasmids and strains. This work was supported by funds from the National Institutes of Health, the G. W. Hooper Foundation, and the Lucille P. Markey Charitable Trust. P. O. B. is a Lucille P. Markey Scholar. H. E. V. is an American Cancer Society Research Professor.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received February 9, 1987; revised March 2, 1987.

References

Bishop, J. M. (1983). Cellular oncogenes and retroviruses. Ann. Rev. Biochem. *52*, 301–354.

Boeke, J. D., Garfinkel, D. J., Styles, C. A., and Fink, G. R. (1985). Ty elements transpose through an RNA intermediate. Cell *40*, 491–500.

Cheer, S., Gentile, J., and Hegre, C. (1974). Improved methods for ATP Analysis. Anal. Biochem. *60*, 102–114.

Church, G. M., and Gilbert, W. (1984). Genomic sequencing. Proc. Natl. Acad. Sci. USA *81*, 1991–1995.

Colicelli, J., and Goff, S. P. (1985). Mutants and pseudorevertants of Moloney murine leukemia virus with alterations at the integration site. Cell *42*, 573–580.

Craig, N. L., and Nash, H. A. (1983). The mechanism of phage λ site-specific recombination: site-specific breakage of DNA by Int topoisomerase. Cell *35*, 795–803.

Daniels, D., Schroeder, J., Szybalski, W., Sanger, F., Coulson, A., Hong, G., Hill, D., Petersen, G., and Blattner, F. (1983). Complete annotated lambda sequence. In Lambda II, R. Hendrix et al., eds. (Cold Spring Harbor, New York: Cold Spring Harbor Laboratory).

Dhar, R., McClements, W., Enquist, L., and Vande Woude, G. (1980). Nucleotide sequences of integrated Moloney sarcoma provirus long terminal repeats and their host and viral junctions. Proc. Natl. Acad. Sci. USA *77*, 3937–3941.

Donehower, L., and Varmus, H. E. (1984). A mutant murine leukemia virus with a single missense codon in *pol* is defective in a function affecting integration. Proc. Natl. Acad. Sci. USA *80*, 6461–6465.

Flavell, A. J. (1984). Role of reverse transcription in the generation of extrachromosomal *copia* mobile genetic elements. Nature *310*, 514–516.

Gronostajski, R. M., and Sadowski, P. (1985). The FLP recombinase of the *Saccharomyces cerevisiae* 2 micron plasmid attaches covalently to DNA via a phosphotyrosyl linkage. Mol. Cell. Biol. *5*, 3274–3279.

Guntaka, R., Richards, O., Shank, P., Kung, H.-J., Davidson, N., Fritsch, E., Bishop, J. M., and Varmus, H. E. (1976). Covalently closed circular DNA of avian sarcoma virus: purification from nuclei of infected quail tumor cells and measurement by electron microscopy and gel electrophoresis. J. Mol. Biol. *106*, 337–357.

Hughes, S. H., Shank, P. R., Spector, D. H., Kung, H.-J., Bishop, J. M., Varmus, H. E., Vogt, P. K., and Breitman, M. L. (1978). Proviruses of avian sarcoma virus are terminally redundant, co-extensive with unintegrated linear DNA and integrated at many sites. Cell *15*, 1397–1410.

Hughes, S. H., Mutschler, A., Bishop, J. M., and Varmus, H. E. (1981). A Rous sarcoma virus provirus is flanked by short direct repeats of a cellular DNA sequence present in only one copy prior to integration. Proc. Natl. Acad. Sci. USA *78*, 4299–4303.

Jolicoeur, P. (1979). The *Fv-1* gene of the mouse and its control of murine leukemia virus replication. Curr. Topics Microbiol. Immunol. *86*, 67–122.

King, W., Patel, M. D., Lobel, L. I., Goff, S. P. and Nguyen-Huu, M. C. (1985). Insertion mutagenesis of embryonal carcinoma cells by retroviruses. Science *228*, 554–558.

Kriegler, M., and Botchan, M. (1983). Enhanced transformation by a simian virus 40 recombinant containing a Harvey murine sarcoma virus long terminal repeat. Mol. Cell. Biol. *3*, 325–339.

Leder, P., Tiemeier, D., and Enquist, L. (1977). EK2 derivatives of bacteriophage lambda useful in the cloning of DNA from higher organisms: the lambda gtWES system. Science *196*, 175.

Lobel, L., Patel, M., King, W., Nguyen-Huu, M., and Goff, S. (1985). Construction and recovery of viable retroviral genomes carrying a bacterial suppressor transfer RNA gene. Science *228*, 329–332.

Luciw, P., Oppermann, H., Bishop, J. M., and Varmus, H. E. (1984). Integration and expression of several molecular forms of Rous sarcoma virus DNA used for transfection of mouse cells. Mol. Cell. Biol. *4*, 1260–1269.

Majors, J. E., and Varmus, H. E. (1981). Learning about the replication of retroviruses from a single cloned provirus of mouse mammary tumor virus. ICN–UCLA Symp. Mol. Cell. Biol. *18*, 241–253.

Maniatis, T., Fritsch, E. F., and Sambrook, J. (1982). Molecular Cloning: A Laboratory Manual (Cold Spring Harbor, New York: Cold Spring Harbor Laboratory).

Mizuuchi, K. (1983). In vitro transposition of bacteriophage Mu: a biochemical approach to a novel replication reaction. Cell *35*, 785–794.

Morris, A. G., and Burke, D. C. (1979). An interferon-sensitive early step in the establishment of infection of murine cells by murine sarcoma/leukemia virus. J. Gen. Virol. *43*, 173–181.

Panganiban, A. T., and Temin, H. M. (1983). The terminal nucleotides of retrovirus DNA are required for integration but not for virus production. Nature *306*, 155–160.

Panganiban, A. T., and Temin, H. M. (1984a). Circles with two tandem LTRs are precursors to integrated retrovirus DNAs. Cell *36*, 673–679.

Panganiban, A. T., and Temin, H. M. (1984b). The retrovirus *pol* gene encodes a product required for DNA integration: identification of a retrovirus *int* locus. Proc. Natl. Acad. Sci. USA *81*, 7885–7889.

Reed, R. R., and Grindley, N. D. F. (1981). Transposon-mediated site-specific recombination in vitro: DNA cleavage and protein–DNA linkage at the recombination site. Cell *25*, 721–728.

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

Sanger, F., Nicklen, S., and Coulson, A. (1977). DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA *74*, 5463–5467.

Schwartzberg, P., Colicelli, J., and Goff, S. P. (1984). Construction and analysis of deletion mutations in the *pol* gene of Moloney murine leukemia virus: a new viral function required for productive infection. Cell *37*, 1043–1052.

Seed, B. (1983). Purification of genomic sequences from bacteriophage libraries by recombination and selection in vivo. Nucl. Acids Res. *11*, 2427–2445.

Shiba, T., and Saigo, K. (1983). Retrovirus-like particles containing RNA homologous to the transposable element *copia* in *Drosophila melanogaster*. Nature *302*, 119–123.

Shimotohno, K., Mizutani, S., and Temin, H. M. (1980). Sequence of retrovirus provirus resembles that of bacterial transposable elements. Nature *285*, 550–554.

Shinnick, T. M., Lerner, R. A., and Sutcliffe, J. G. (1981). Nucleotide sequence of Moloney murine leukemia virus. Nature *293*, 543–548.

Shoemaker, C., Goff, S., Gilboa, E., Paskind, M., Mitra, S. W., and Baltimore, D. (1980). Structure of a cloned circular Moloney murine leukemia virus DNA molecule containing an inverted segment: implications for retrovirus integration. Proc. Natl. Acad. Sci. USA *77*, 3932–3936.

Shoemaker, C., Hoffman, J., Goff, S., and Baltimore, D. (1981). Intramolecular integration within Moloney murine leukemia virus DNA. J. Virol. *40*, 164–172.

Varmus, H. E. (1983). Retroviruses. In Mobile Genetic Elements, J. Shapiro, ed. (New York: Academic Press), pp. 411–503.

Varmus, H. E., and Swanstrom, R. (1985). Replication of retroviruses. In RNA Tumor Viruses, Vols. 1 and 2, R. Weiss et al., eds. (Cold Spring Harbor, New York: Cold Spring Harbor Laboratory).

Varmus, H. E., Padgett, T., Heasley, S., Simon, G., and Bishop, J. M. (1977). Cellular functions are required for the synthesis and integration of avian sarcoma virus–specific DNA. Cell *11*, 307–319.

Vetter, D., Andrews, B., Roberts-Beatty, L., and Sadowski, P. (1983). Site-specific recombination of yeast 2 micron DNA *in vitro*. Proc. Natl. Acad. Sci. USA *80*, 7284–7288.

Weiss, B., and Richardson, C. C. (1967). Enzymatic breakage and joining of deoxyribonucleic acid. III. An enzyme-adenylate intermediate in the polynucleotide ligase reaction. J. Biol. Chem. *242*, 427.

Proc. Natl. Acad. Sci. USA
Vol. 86, pp. 2525–2529, April 1989
Biochemistry

Retroviral integration: Structure of the initial covalent product and its precursor, and a role for the viral IN protein

(transposition/recombination/retrotransposon/AIDS)

PATRICK O. BROWN*†, BRUCE BOWERMAN‡, HAROLD E. VARMUS‡§, AND J. MICHAEL BISHOP*‡§

*The G. W. Hooper Research Foundation, ‡Department of Biochemistry and Biophysics, and §Department of Microbiology and Immunology, University of California, San Francisco, CA 94143

Contributed by J. Michael Bishop, December 6, 1988

ABSTRACT **An essential step in the life cycle of a retrovirus is the integration of a DNA copy of the viral genome into a host cell chromosome. We have analyzed the structure of the initial covalent product of an *in vitro* retroviral integration reaction and determined the structure of the ends of the unintegrated linear viral DNA molecules present *in vivo* in cells infected with murine leukemia virus (MLV). Our results lead to the following conclusions: (*i*) Circularization of viral DNA plays no role in integration. The direct precursor to the integrated MLV provirus is a linear molecule. (*ii*) The initial step in the integration reaction is probably a cleavage that removes the terminal 2 bases from each 3′ end of the viral DNA. This cleavage depends on a virally encoded protein, IN, that has previously been shown genetically to be required for integration. (*iii*) The resulting viral 3′ ends are joined to target DNA to form the initial recombination intermediate.**

Retroviruses are ubiquitous pathogens. Consequences of retroviral infection in humans include AIDS, leukemia, lymphoma, and degenerative diseases of the central nervous system. To multiply, a retrovirus must synthesize a DNA copy of its RNA genome and integrate the viral DNA into a chromosome of the infected host cell. The integrated viral genome, or provirus, is thereafter transmitted as a stable element of the host genome. The provirus provides the sequences required in *cis* for its own expression and the template for the next generation of viral RNA (1).

Soon after a retrovirus enters the cytoplasm of its host cell, the viral reverse transcriptase synthesizes a terminally redundant double-stranded linear DNA copy of the viral genome. The terminal redundancies are called long terminal repeats (LTRs). These linear DNA molecules enter the nucleus, where some are circularized by covalently joining their ends to produce molecules called 2-LTR circles. The sequence created at the site of this ligation is called the circle junction. Both the linear molecule and the 2-LTR circle are reasonable candidates for the ultimate precursor to the integrated provirus. Which of these topological forms is actually used for integration?

In one experiment, the circle junction sequence of spleen necrosis virus (SNV), when inserted at an internal site in the SNV genome, appeared able to serve as a viral attachment site for integration (2). This result pointed to the 2-LTR circle as the ultimate precursor in SNV integration. However, results of an analogous experiment with murine leukemia virus (MLV) do not support the 2-LTR circle as the proximal precursor in MLV integration (L. Lobel, J. Murphy, and S. Goff, personal communication).

We have investigated the mechanism of retroviral integration by using an *in vitro* system that faithfully reproduces the *in vivo* integration reaction (3). *In vitro*, the linear form of unintegrated viral DNA can serve as a precursor to the integrated MLV provirus (3). It does not follow that the immediate precursor is linear, since the linear molecule could, in principle, be covalently circularized *in vitro* prior to integration.

Our results *in vitro* led us to reinvestigate the possibility that the direct precursors to integrated proviruses might be linear molecules. In this report, we use a detailed analysis of the structure of the viral DNA ends in the unintegrated precursor and the initial covalent integration product to show that the immediate precursor to the integrated provirus is a linear molecule. We can also deduce from our data the polarity of the initial bonds between viral and target DNA, the probable source of the energy for formation of these initial bonds, and a specific role for the viral IN protein in preparing the viral DNA ends for integration. A similar analysis of a MLV integration intermediate has recently been reported (4).

MATERIALS AND METHODS

General Methods. General methods were as described (5). Strand-specific probes were prepared as described (6). Electroblotting onto nylon membranes, crosslinking, and hybridizations were as described (7).

Defining the Ends of Unintegrated Linear MLV DNA. Sixteen hours after infection, MLVsupF (8) DNA was isolated from the cytoplasm of infected cells (3), and full-length linear [≈9.2 kilobases (kb)] molecules were purified by agarose gel electrophoresis. This DNA was digested with *Pvu* II or *Sac* I (New England Biolabs), denatured, and electrophoresed through a 6% polyacrylamide/7 M urea gel. Sequencing ladders were prepared using as templates single-stranded M13mp18 phage DNA containing a cloned copy of either the plus or minus strand of the MLVsupF LTR (8), or the *Xba* I/*Xba* I circle junction fragment from p8.8 (9), cloned in pSP64. The primers matched the known (Fig. 1 *C* and *D*) or predicted (Fig. 1 *A* and *B*) 5′ ends of the DNA fragments being measured. The primers were as follows: (*A*) pAATGAAAGACCCCCGCTGAC; (*B*) pAATGAAAGACCCCACCTGTA; (*C*) pCAATAAAAGAGCCCAC; (*D*) pCTGTTCCATCTGTTCCTGA. After electrophoresis, the resolved DNA fragments were electroblotted to nylon filters (NEN Genescreen) and detected by hybridization using strand-specific probes for the MLVsupF LTR.

Preparation of Unintegrated DNA from the SF2 Mutant Virus. A molecular clone of Moloney MLV DNA, containing the SF2 mutation (10), but otherwise identical to wild-type [clone 1 (11)] MLV was transfected into NIH 3T3 mouse fibroblasts. Supernatant virus was used to infect fresh NIH

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: MLV, murine leukemia virus; LTR, long terminal repeat; SNV, spleen necrosis virus; ASLV, avian retrovirus; RFI, replicative form I.

†Present address: Departments of Pediatrics and Biochemistry, Stanford University School of Medicine, Stanford, CA 94305.

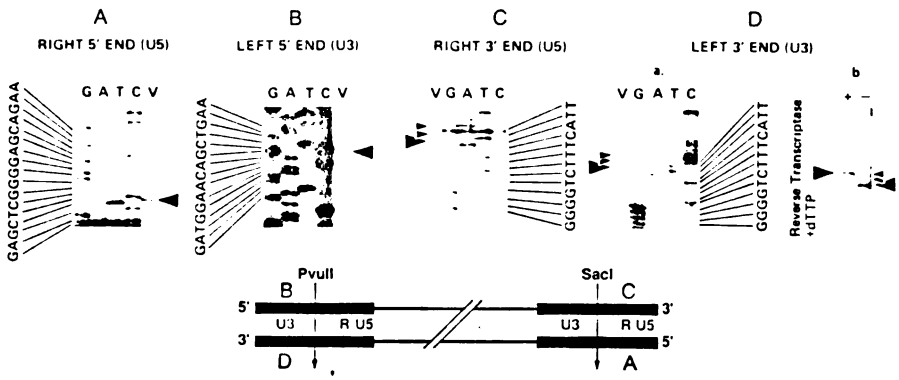


FIG. 1. Defining the ends of unintegrated linear MLV DNA. Full-length MLVsupF DNA molecules were purified from the cytoplasm of infected cells, digested with *Pvu* II (*B* and *D*) or *Sac* I (*A* and *C*), denatured, and electrophoresed through a 6% polyacrylamide/7 M urea gel. Sequencing ladders prepared from cloned copies of the viral DNA ends served as size markers. Resolved DNA fragments were transferred to nylon filters and detected by hybridization using strand-specific probes for the LTR. The origins of the DNA fragments analyzed in each panel are indicated in the diagram at bottom (only the restriction sites nearest the ends are shown). In *A*, *B*, *C*, and *Da* the lane marked V contains viral DNA, and the lanes marked G, A, T, and C contain the corresponding sequencing reactions. Arrowheads point to the viral DNA fragments. The interpretation of the sequencing ladder is shown to one side. (*Db*) The effect of treatment of the viral DNA with avian myeloblastosis virus reverse transcriptase (Boehringer Mannheim) + TTP (dATP, dCTP, and dGTP were omitted) (lane +). Untreated viral DNA is in lane −.

3T3 cells. Cytoplasmic viral DNA molecules were recovered 16 hr after infection.

Physical Assay for Retroviral Integration. Reactions were as described (3), except that ϕX replicative form I (RFI) DNA was used as target instead of λgtWES DNA. DNA products were recovered (3) and digested with *Pvu* II and then electrophoresed through a 0.7% agarose gel. Resolved DNA fragments were blotted to a nylon filter (HybondN), crosslinked to the filter with UV light, and then detected by hybridization with a ^{32}P-labeled probe.

Analysis of the Free Viral DNA End in an Integration Intermediate. Integration-competent nucleoprotein complexes were purified from a cytoplasmic extract by Bio-Gel A5m chromatography (3). Integration reactions used ϕX RFI DNA as a target. Products were digested with *Pvu* II and electrophoresed through a 0.7% agarose (Seaplaque) gel. The 6.1-kb recombinant molecules containing the junctions between viral and target DNA were recovered from the gel, denatured, and electrophoresed through a 6% polyacrylamide/7 M urea gel. Unintegrated linear viral DNA was digested with *Pvu* II, denatured, and electrophoresed in neighboring lanes. Resolved DNA fragments were blotted to a nylon filter and detected by hybridization.

RESULTS

Structure of the Ends of Unintegrated Linear Viral DNA Molecules. Whether they are joined directly to target DNA or first joined to one another to form a 2-LTR circle, the ends of the unintegrated linear viral DNA molecules participate directly in an essential step in retroviral integration. Furthermore, their structure is fundamental to our analysis of the role of circularization in the integration process. Therefore, we isolated linear viral DNA molecules from the cytoplasm of MLVsupF-infected cells (3) and determined the structure of their ends of an indirect sequencing method. Bands corresponding to each strand of each of the terminal restriction fragments were resolved by electrophoresis and identified by their alignment with an adjacent sequencing ladder (Fig. 1). Our conclusions are illustrated in Fig. 2. The left and right ends of the viral DNA molecule are essentially identical. At each end, the 5'-terminal base corresponds to the first nucleotide (deoxyadenosine) that is joined to the RNA primer at the start of plus or minus strand DNA synthesis (1). The 3' termini are heterogeneous, with the most abundant population (generally >90% of the total) terminating 2 bases short of the 5' end, as illustrated in Fig. 2, thus matching the boundaries of the integrated provirus. The recessed 3' ends could be filled in completely by avian myeloblastosis virus reverse transcriptase in the presence of dTTP (Fig. 1*Db*), establishing that the ends have unblocked 3'-OH groups and confirming that the missing bases are both thymidines. The remainder of the 3' ends terminate flush with, or 1 base short of, the corresponding 5' ends. The 2-base-recessed structure, henceforth called structure 1, is particularly intriguing since it matches a predicted intermediate in the integration reaction, as discussed below.

The IN Protein Is Required for Formation of a Recessed 3' End. Viruses with mutations in the 3' portion of the *pol* gene can carry out the early steps of the viral life cycle, from transcription of the provirus through DNA synthesis and nuclear entry, but their integration is markedly impaired, and replication is blocked (9, 10, 12). The protein encoded by this domain is called IN. To define the function of IN more precisely, we investigated the effect of the SF2 mutation, a frameshift mutation in the IN coding region of MLV (10), on the structure of the ends of linear viral DNA molecules (Fig. 3). As also shown in Fig. 1, the 3' ends of most wild-type MLV DNA molecules were recessed by 2 bases from the 5' end (Fig. 3, lane 2). In contrast, the 3' ends of almost all unintegrated SF2 DNA molecules were flush with the 5' ends, and the band corresponding to a 2-base recess was undetectable (lane 3). Thus, the IN protein is required for formation of the recessed 3' end.

What is the role of the IN protein in generating the recessed 3' end? One possibility is that it directs termination of DNA synthesis at a point 2 bases short of the end of the template.

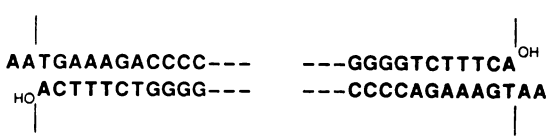


FIG. 2. The ends of unintegrated MLV DNA molecules. In cells infected with wild-type MLV, most of the unintegrated linear DNA molecules have the structure shown. The 5' ends extend 2 bases beyond the boundaries of the integrated provirus (indicated by vertical lines). The 3' ends are recessed by 2 bases, terminating precisely at the 3'-OH group that is ultimately joined to target DNA.

FIG. 3. Ends of unintegrated viral DNA molecules from wild-type and IN MLV. Cytoplasmic viral DNA molecules were harvested 16 hr after infection with wild-type (11) (lanes 1 and 2) or SF2 mutant (lane 3) Moloney MLV. Total cytoplasmic DNA (lane 1) or purified full-length (8.8 kb) linear viral DNA molecules (lanes 2 and 3) were analyzed. DNA was digested with *Sac* I, denatured, then electrophoresed and electroblotted. Filters were hybridized with a non-strand-specific probe specific for the MLV LTR. The upper band in each lane corresponds to the right 5' end of the viral DNA, and the lower cluster of bands corresponds to the right 3' end. The band indicated by the open arrowhead corresponds to the 2-base-recessed 3' end typical of wild-type viral DNA (structure 1); the band indicated with the solid arrowhead corresponds to a 3' end that is flush with the 5' end. Bands representing the 3' and 5' ends are well separated because *Sac* I cuts DNA with a 4-base stagger, with the 5' end recessed.

Alternatively, DNA synthesis could continue unimpeded to the end of the template, producing a flush-ended molecule from which the 3'-terminal dinucleotide is subsequently removed in a reaction that depends on IN. We favor the latter hypothesis for two reasons. First, 3' ends recessed by 2 bases are more frequent in purified full-length linear molecules than in an unfractionated population of viral DNA molecules from an infected cell cytoplasm (Fig. 3, lanes 1 and 2). Although the structures of the other DNA molecules in the unfractionated population are not known in detail, they are presumed to be intermediates in viral DNA synthesis. The fact that, compared to its precursors, the ultimate product of viral DNA synthesis is enriched for the 2-base-recessed 3' end implies that the recessed end is the result of a late processing event. Second, enzymological analyses of the IN protein from avian retroviruses (ASLV) point to the possibility of a role for this protein in preparing the viral 3' end for integration (1, 13, 14). The ASLV IN protein has an endonuclease activity that yields products terminated by a 5'-P and 3'-OH. The ends of linear ASLV DNA have not been investigated as substrates for this endonuclease activity, but under some conditions model substrates containing the ASLV circle junction are cleaved preferentially to expose the precise 3'-OH group expected to be joined to target DNA. Taken together, these data and our results are most consistent with a model in which the 2-base recess results from cleavage of the 3' end of the viral DNA molecule by the IN protein.

Deducing the Topology of the Viral DNA Precursor and Polarity of the Initial Linkage. Based on existing knowledge of the structure of the integrated provirus, the DNA target, and the putative viral DNA precursors, one can outline a plausible reaction pathway invoking either the linear viral DNA molecule or the 2-LTR circular form as the proximal precursor to the integrated provirus (Fig. 4) (3, 9). In pathway A, the proximal precursor is the 2-LTR circle (1A), formed by ligation of the ends of the linear molecule. In pathway B, the linear viral DNA molecule (1B) is itself the substrate for integration. In the DNA breakage and joining steps, viral and target DNA strands are broken at the sites marked by arrowheads. The 3'-OH ends of the viral DNA are then joined to the corresponding 5'-P ends of the target DNA. The cuts in the viral and target DNA molecules do not need to occur in concert, but the joining reaction is probably coupled to cleavage of the target DNA, as discussed below. In the

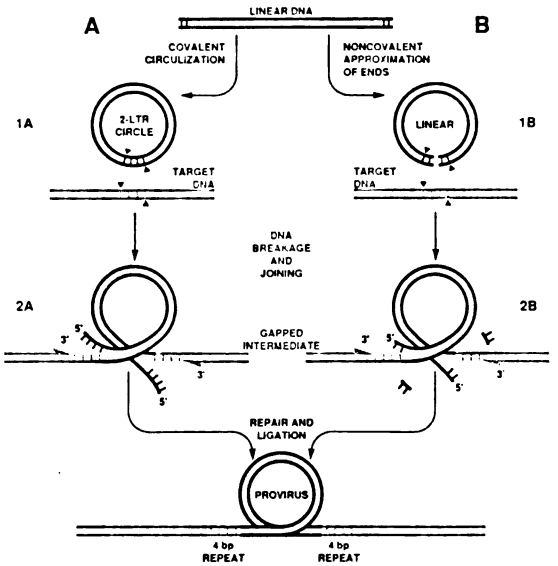


FIG. 4. Two possible pathways for integration of linear retroviral DNA. The steps are described in the text. The 4 base pairs (bp) of viral DNA that are ultimately lost upon integration (2 from each end of the linear molecule) and the 4-bp target sequence that is duplicated upon integration of MLV DNA are indicated by line segments perpendicular to the DNA strands.

resulting intermediate (2A or 2B), the provirus is flanked by short gaps that are the precursors to the flanking repeats in the final product. DNA synthesis primed by the 3'-OH on the target side of the gap can initiate repair to yield the mature integrated provirus. DNA breakage and joining with the opposite polarity, joining an overhanging 3' end at the target site to a recessed 5' end produced by cleavage of the viral DNA, is also mechanistically plausible.

In the gapped initial product of the integration reaction, the expected structure of the free viral DNA end differs depending on whether the ultimate precursor is linear or a 2-LTR circle (Fig. 4, 2A and 2B). In pathway A, this free end is longer by 2 bases than its counterpart in the linear precursor. These 2 extra bases are covalently joined to the end in the circularization step and left attached when the viral DNA is cleaved 2 bases to one side of the circle junction. In pathway B, the free viral DNA end in the gapped intermediate is identical to the corresponding end of the linear precursor, having never been covalently modified. The strict dependence of the structure of this intermediate on the topology of its precursor allows us to deduce the topology of the precursor by comparing the free viral DNA ends in the gapped intermediate with their counterparts in the unintegrated linear molecule.

Our previous work used a genetic assay to detect the products of retroviral integration *in vitro* (3). This assay was highly sensitive and allowed us to determine the sequences at the junctions between viral and target DNA in the cloned products of the integration reaction. However, it did not allow us to examine intermediates in the reaction. We have therefore developed physical methods that enable us to detect and characterize these intermediates.

In one such assay, we used gel electrophoresis to separate the products of the integration reaction from unintegrated viral DNA molecules (Fig. 5). In the experiment shown in Fig. 5, the target for integration was ϕX RFI DNA. By carrying out the integration reaction in the absence of nucleotides, using nuclease-free preparations of integration-

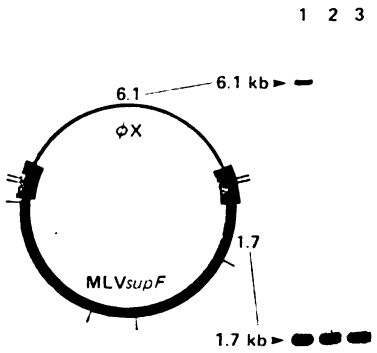


FIG. 5. Physical assay for retroviral integration. Expected structure of a recombinant molecule produced by integration of MLVsupF into φX RFI DNA. *Pvu* II cuts at several sites in MLVsupF DNA (tick marks perpendicular to the circular map), but does not cut φX RFI. Reactions using φX RFI DNA as the target for integration (lane 1) or no target DNA (lane 2) were carried out as described in *Materials and Methods*. After stopping the DNA reaction (lane 2), φX RFI DNA was added prior to recovery of the DNA products. DNA from each reaction (lanes 1 and 2) or an equivalent amount of DNA isolated directly from the cytoplasmic extract without any incubation (lane 3) were digested with *Pvu* II and then electrophoresed through a 0.7% agarose gel. Resolved DNA fragments were transferred to a nylon filter and detected by hybridization with a ^{32}P-labeled probe specific for the LTR. *Hin*dIII fragments of λ DNA served as size markers. The electrophoretic mobility of the upper band, seen only in lane 1, indicates a size of 6.1 kb, the sum of the size of φX RFI plus the two terminal *Pvu* II fragments of MLVsupF. As illustrated in the accompanying diagram, this corresponds to a novel restriction fragment produced by integration of MLVsupF into φX RFI. The lower band, present in all three lanes, corresponds to the 1.7-kb *Pvu* II fragment from either integrated or unintegrated viral DNA molecules. Smaller fragments and fragments lacking LTR sequences are not visualized in this blot.

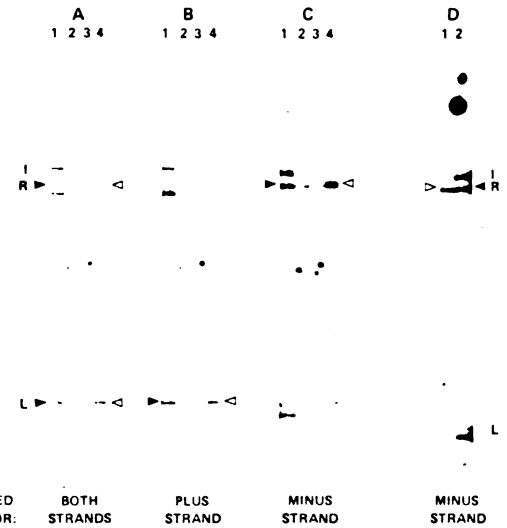


FIG. 6. The free viral DNA end in an integration intermediate. Products of an integration reaction using φX RFI DNA as a target were digested with *Pvu* II and molecules containing the junctions between viral and target DNA were purified, denatured, and electrophoresed through a 6% polyacrylamide/7 M urea gel. Unintegrated linear viral DNA was digested with *Pvu* II, denatured, and electrophoresed in neighboring lanes. Resolved DNA fragments were blotted to a nylon filter and hybridized with ^{32}P-labeled probes for the viral plus strand (*B*) or minus strand (*C* and *D*) or both (*A*). (*A–C*) Successive hybridizations to a single filter. Lanes 1 and 2 contain unintegrated DNA molecules; lanes 3 and 4 contain the isolated recombination intermediate. In *A*, *B*, and *C*, lane 3 contains recombinants produced using an extract from 4×10^6 infected cells. Lane 4 in *A*, *B*, and *C* and lane 1 in *D* contain products of a 5-fold larger reaction. In *A*, *B*, and *C*, lanes 1 and 2 contain viral DNA from 2×10^5 and 4×10^4 cells, respectively. In *D*, lane 2 contains viral DNA from 2×10^5 cells. Because of a small amount of residual signal from the plus strand probe in *C*, a separate filter hybridized exclusively with a probe for the minus strand is shown in *D*. In *D* the recombination intermediate is in lane 1 and unintegrated viral DNA is in lane 2. Clusters of bands arising from the right end, left end, and an internal fragment of the viral DNA molecule are indicated by R, L, and I, respectively. Solid arrowheads indicate bands representing the 5′ ends of viral DNA molecules. The other bands in the right and left end clusters correspond to 3′ ends. Open arrowheads indicate bands corresponding to the free viral DNA ends from the gapped intermediate.

competent viral nucleoprotein complexes prepared by gel-exclusion chromatography (3), we can recover intact the initial covalent product of the integration reaction (Fig. 4 structure 2A or 2B). After digestion of the reaction products with *Pvu* II, a recombinant fragment containing the junctions between MLV and target DNA is resolved from the internal MLV fragments, and more importantly from the unintegrated MLV ends, by agarose gel electrophoresis (Fig. 5, Lane 1). This recombinant fragment was purified by agarose gel electrophoresis, denatured, and electrophoresed through a denaturing polyacrylamide gel. Unintegrated MLVsupF DNA, similarly digested with *Pvu* II and denatured, was electrophoresed in an adjacent lane to allow direct comparison of the lengths of the terminal *Pvu* II fragments from the integrated and unintegrated viral DNA molecules. The resolved fragments were transferred to a nylon filter and detected by hybridization with radioactive probes (Fig. 6) (7). By probing successively with probes specific for the plus and minus strands of viral DNA, we could also determine the polarity of the linkage. Fig. 6*A* shows the pattern of hybridization with probes recognizing both DNA strands. Whereas bands representing both the 3′ and 5′ fragments from each end of the viral DNA are visible in the lanes containing unintegrated viral DNA (lanes 1 and 2) only a single band from each viral DNA end (open arrowheads) is visible in the lanes containing the isolated recombination intermediate (lanes 3 and 4). The bands from the recombination intermediate (open arrowheads) comigrate precisely with the bands representing the 5′-terminal fragment from each end of the unintegrated linear molecule (solid arrowheads). That these bands are indeed from the 5′ ends of the viral DNA molecule is demonstrated by hybridization with strand-specific probes. The viral plus strand contributes the left 5′ end and the right 3′ end of the viral DNA. Thus, from Fig. 6*B* it is apparent that, in the gapped integration intermediate, the 5′ end of the viral DNA plus strand is not joined to the target DNA, and that it is identical in length to the corresponding end of the unintegrated linear molecule. Similarly, the free end of the viral minus strand in the gapped intermediate is the 5′ end, and it is identical to the corresponding 5′ end of the linear precursor (Fig. 6 *C* and *D*).

We conclude that in the gapped intermediate the 3′ end of each viral DNA strand must be linked to target DNA, since the 5′ end is not. As an expected consequence of their attachment to target DNA, the 3′-terminal fragments have a reduced electrophoretic mobility such that they are not detected in this analysis. Furthermore, in the gapped intermediate, the free 5′-terminal fragment from each end of the viral DNA is identical in length to the corresponding fragment from the unintegrated precursor. Thus, the ultimate precursor to the integrated provirus must be a linear molecule and not the 2-LTR circle.

DISCUSSION

Our conclusion that the precursor to the integrated provirus is exclusively a linear molecule is in agreement with that of Fujiwara and Mizuuchi (4) and consistent with *in vivo* studies of MLV integration (L. Lobel, J. Murphy, and S. Goff, personal communication) but contrasts with the deduction by Panganiban and Temin (2) that the 2-LTR circle is a precursor in SNV integration. This apparent inconsistency might be reconciled in any of several ways. Integration of a 2-LTR circle, accompanied or followed by absolutely efficient and precise removal of the terminal 2 bases from the viral 5' end, could account for the observed structure of the MLV integration intermediate, but such a processing reaction appears gratuitous and therefore implausible. Alternatively, the two viruses might differ in the preferred topology of the integrative precursor. However, since the evidence for a circular intermediate in SNV integration is equivocal (A. Panganiban, personal communication), and we can see no compelling mechanistic purpose for the circularization step, we favor the hypothesis that both viruses ordinarily integrate via a linear precursor.

MLV integration does not require a high-energy cofactor (3). The DNA breakage and joining reactions are therefore probably coupled. One possible coupling mechanism involves a transient high-energy protein–DNA bond (15). An alternative solution to the energy coupling problem (16) calls for an enzyme-catalyzed nucleophilic attack by the viral 3'-OH terminus on a phosphodiester bond in the target DNA, resulting in an essentially isoenergetic transesterification, much like that which occurs in RNA splicing (17). The viral 3' ends that are joined to target DNA in the initial covalent joining reaction determine the boundaries of the integrated provirus. Accordingly, to account for both the observed structure of the gapped intermediate and the structure of the integrated provirus by any model, the viral 3' end that is joined to the target DNA needs to be recessed by 2 bases from the 5' end. Thus, we surmise that structure 1, a linear molecule with 3' ends that are recessed by 2 bases from the 5' ends, is the viral DNA precursor that participates directly in the joining reaction. Since most viral DNA molecules have this structure even before they enter the nucleus, cleavage of the viral 3' end must not generally be coupled to the joining reaction. Thus, it is likely that the joining reaction is energetically coupled to cleavage of the target DNA rather than viral DNA. If the exigencies of energy conservation therefore cannot explain a requirement for cleavage of the viral 3' ends, why is a recessed 3' end used for integration rather than a flush 3' end? Perhaps the 5' extension that distinguishes the two structures is important for an effective interaction with the integration machinery.

On the basis of this work, we propose that retroviral integration involves the following sequence of events: (*i*) Viral DNA synthesis produces a blunt-ended linear molecule. (*ii*) The IN protein cuts the 3' end of each viral DNA strand, removing 2 bases and exposing the 3'-OH group that is to be joined to the target DNA. The result is structure 1 (Fig. 2). This activity is sufficient to explain why IN is required for integration, although IN might also play an essential role in later steps. (*iii*) The target DNA is cut with a 4-base-pair stagger. This cleavage is energetically coupled to joining of the viral 3'-OH ends to the target 5'-P ends, resulting in a gapped intermediate (Fig. 4, structure 2B). (*iv*) Repair synthesis across the gap is then primed by the target 3'-OH group.

This model points to specific interactions between IN and the viral DNA ends that should be detectable in the native nucleoprotein complexes or by using purified IN protein and model DNA substrates. The spatial and temporal separation of the cleavage of the viral 3' ends from their joining to target DNA raises the possibility that these steps might be carried out by different proteins. It will be interesting to determine whether these activities can be separated genetically or by using specific inhibitors, and whether different active sites are involved. The sequence-sensitive interaction of the integration apparatus with the viral DNA ends and the apparently sequence-insensitive recognition of the DNA target for integration are likely to be mediated by different DNA binding domains, if not different proteins. Ultimately, the detailed biochemistry of retroviral integration can best be illuminated by using defined purified components to reconstruct the complete integration reaction and its individual steps. Stages of the integration reaction that might usefully be studied in isolation include viral DNA binding and cleavage, target DNA binding and cleavage, and the joining reaction itself. Knowledge of the structures of the viral DNA substrate and intermediates in retroviral integration provides a foundation for work aimed at reconstituting these activities.

We thank Sue Klapholz for her helpful comments on the manuscript. This work was supported by funds from the National Institutes of Health, the G. W. Hooper Foundation, the Howard Hughes Medical Institute, and the Lucille P. Markey Charitable Trust. H.E.V. is an American Cancer Society Research Professor. P.O.B. was a Lucille P. Markey Scholar for a portion of this work and an assistant investigator of the Howard Hughes Medical Institute for the remainder.

1. Varmus, H. & Brown, P. (1989) in *Mobile DNA*, eds. Howe, M. & Berg, D. (ASM, Metals Park, OH), in press.
2. Panganiban, A. & Temin, H. (1984) *Cell* **36**, 673–679.
3. Brown, P., Bowerman, B., Varmus, H. & Bishop, J. M. (1987) *Cell* **49**, 347–357.
4. Fujiwara, T. & Mizuuchi, K. (1988) *Cell* **55**, 497–504.
5. Ausubel, F., Brent, R., Kingston, R., Moore, D., Seidman, J., Smith, J. & Struhl, K., eds. (1987) *Current Protocols in Molecular Biology* (Wiley, New York).
6. Hu, W.-T. & Messing, J. (1982) *Gene* **17**, 271–277.
7. Church, G. & Gilbert, W. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 1991–1995.
8. Lobel, L., Patel, M., King, W., Nguyen-Huu, M. & Goff, S. (1985) *Science* **228**, 329–332.
9. Shoemaker, C., Goff, S., Gilboa, E., Pasking, M., Mitra, S. & Baltimore, D. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 3932–3936.
10. Donehower, L. & Varmus, H. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 6461–6465.
11. Fan, H. & Paskind, M. (1974) *J. Virol.* **14**, 421–429.
12. Panganiban, A. & Temin, H. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 7885–7889.
13. Grandgenett, D. & Vora, A. (1985) *Nucleic Acids Res.* **13**, 6205–6221.
14. Grandgenett, D., Vora, A., Swanstrom, R. & Olsen, J. (1986) *J. Virol.* **58**, 970–974.
15. Craig, N. L. (1988) *Annu. Rev. Genet.* **22**, 77–105.
16. Craigie, R. & Mizuuchi, K. (1987) *Cell* **51**, 493–501.
17. Padgett, R., Grabowski, P., Konarska, M., Seiler, S. & Sharp, P. (1986) *Annu. Rev. Biochem.* **55**, 1119–1150.

Chapter 3

A Nucleoprotein Complex that Mediates the Integration of Retroviral DNA

A nucleoprotein complex mediates the integration of retroviral DNA

Bruce Bowerman,[1] Patrick O. Brown,[3,4] J. Michael Bishop,[1-3] and Harold E. Varmus[1,2]

[1]Department of Biochemistry and Biophysics; [2]Department of Microbiology and Immunology; [3]G.W. Hooper Foundation, University of California, San Francisco, California 94143 USA

The integration of viral DNA into the host genome is an essential step in the retrovirus life cycle. To understand this process better, we have examined the native state of viral DNA in cells acutely infected by murine leukemia virus (MLV), using both a physical assay for viral DNA and a functional assay for integration activity (Brown et al. 1987). The viral DNA and integration activity copurify during velocity sedimentation, gel filtration, and density equilibrium centrifugation, indicating that viral DNA is in a large (~160S) nucleoprotein complex that includes all functions required for integration activity in vitro. Analysis by immunoprecipitation shows that the viral capsid protein is part of the active nucleoprotein complex, but recognition of the complex by only a subset of anti-capsid sera implies that the protein is constrained conformationally. The viral DNA within this structure is accessible to nucleases; the effects of nucleases on the integrity of the complex suggest that the integration-competent particle is derived from and similar to the core of extracellular virions.

[*Key Words*: Retrovirus; integration; nucleoprotein complex; transposition]

Received January 13, 1989; revised version accepted February 24, 1989.

Early events in the retrovirus life cycle convert the viral RNA genome, originally within an extracellular virion, into a DNA provirus joined to a host chromosome (Fig. 1). Upon entry into a cell, viral RNA is transcribed in the cytoplasm to form a linear double-stranded DNA molecule with long terminal repeats (LTRs), and the viral DNA eventually is integrated into a chromosome (for reviews, see Varmus 1983; Varmus and Brown 1989). In addition, some of the linear DNA that enters the nucleus is converted into circular molecules with either one or two copies of the LTR.

Little is known about the intracellular structures that mediate these early events in replication. The retroviral integration reaction, however, recently has been recapitulated in vitro using cell-free extracts made from acutely infected cells as a source of precursors active for integration (Brown et al. 1987). Analysis of murine leukemia virus (MLV) integration intermediates observed in vitro has identified the linear viral DNA molecule as the immediate precursor in retroviral integration (Fujiwara and Mizuuchi 1988; Brown et al. 1989). The circular forms of viral DNA that are present in the nuclei of infected cells appear not to participate in the in vitro reaction, and their role in the virus life cycle is unresolved (Panganiban and Temin 1984a; Brown et al. 1987, 1989; Fujiwara and Mizuuchi 1988; S. Goff, pers. comm.). The viral integration protein (IN), encoded by the 3′ end of the *pol* gene, has been implicated in the reaction through site-directed mutagenesis (Donehower and Varmus 1984; Schwartzberg et al. 1984; Panganiban and Temin 1984b), and through demonstration of an associated endonuclease activity with the integrase protein of avian retroviruses (Duyk et al. 1983; Grandganett and Vora 1985; Grandganett et al. 1986).

We wish to improve our understanding of the structures that carry out important steps in retrovirus replication by characterizing the native state of viral DNA during its synthesis in the cytoplasm, entry into the nucleus, and integration into the genome. The intracellular structures that contain unintegrated viral DNA presumably are derived from components of the extracellular virion core, in which the viral RNA genome initially resides. The MLV virion core consists of an icosahedral protein shell, made up predominantly of multiple copies of the 30-kD viral capsid protein (CA) encoded by the *gag* gene. This shell encloses the viral nucleocapsid, a complex in which two copies of the single-stranded genomic RNA are associated with the highly basic viral nucleocapsid protein, another product of *gag*. Also included in the nucleocapsid, though less tightly associated with viral RNA, are reverse transcriptase and the integration protein encoded by *pol*, probably the viral protease, and perhaps other *gag* proteins (Dickson et al. 1985). On the basis of knowledge of the structure and components of the extracellular virion core, we have begun to examine the intracellular structures that contain unintegrated retroviral DNA and are active for integration in an in vitro assay.

Previous work has suggested that retroviral DNA in

[4]Current address: Departments of Pediatrics and Biochemistry, Stanford University School of Medicine, Stanford, California 94305 USA.

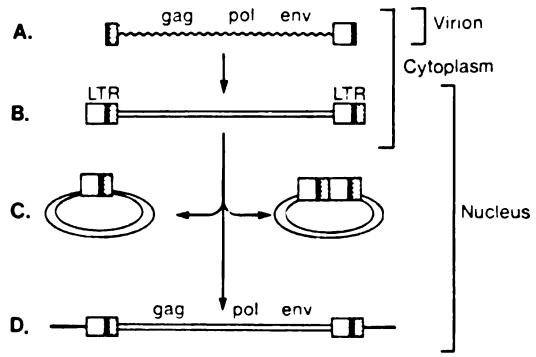


Figure 1. Nucleic acid intermediates in retroviral replication. The early events of retroviral replication convert the single-stranded, genomic RNA (*A*) into a double-stranded DNA provirus covalently linked to the host genome (*D*). The viral genomic RNA is present both in extracellular virion particles and in the cytoplasm of acutely infected cells immediately after viral entry. The viral RNA is reverse-transcribed in the cytoplasm into a full-length linear double-stranded DNA molecule (*B*), which also appears in the nucleus. In the nucleus, some of the linear DNA is converted into circular molecules with one or two LTRs (*C*). After integration of the linear DNA into a host chromosome, the provirus (*D*) can serve as the template for expression of progeny virus. The approximate positions of the three retroviral genes—*gag*, *pol*, and *env*—are indicated in structures *A* and *D*. The compartments in which the different forms of the viral genome appear are indicated to the right.

an infected cell is tightly associated with the enzymatic machinery required for its integration (Brown et al. 1987). To characterize these putative complexes, we have made extracts from MLV-infected cells at a time when several incoming viral genomes per cell have been converted into full-length double-stranded DNA molecules. We show that this linear DNA intermediate in viral replication is part of a large, 160S nucleoprotein complex that contains all the machinery required for integration activity in an in vitro assay. Immunoprecipitation and nuclease sensitivity experiments indicate that the complex is constrained conformationally and also permeable to macromolecules. The results we present are consistent with models in which the intracellular nucleoprotein complex that mediates retroviral integration is derived from and resembles the extracellular virion core.

Results

MLV DNA in acutely infected cells resides within a large nucleoprotein complex

To study retroviral integration, we use a replication-competent derivative of MLV, called MLVsupF, which carries a bacterial *supF* tRNA gene in its LTRs (Lobel et al. 1985). This marker enables us to score integration events by a genetic assay in which the *supF* gene in the MLV LTR suppresses amber mutations in λ gtWES genes required for lytic growth in *Escherichia coli* (Brown et al. 1987). To obtain active precursors for the in vitro integration reaction, NIH-3T3 cells are infected by cocultivation with an MLVsupF-producing cell line. After the infection has proceeded for a period of time sufficient to generate multiple copies per cell of full-length, unintegrated viral DNA (about 18 hr), cell-free cytoplasmic and nuclear extracts are made (see Experimental methods). These extracts contain viral precursors that can reproduce accurately the retroviral integration reaction in vitro (Brown et al. 1987).

To determine whether the viral DNA in acutely infected cells is associated with other macromolecules, we subjected extracts to velocity sedimentation through sucrose gradients (Fig. 2), and detected the viral DNA by both a physical assay (DNA hybridization after gel electrophoresis and transfer to a nylon membrane) and a functional assay (in vitro integration activity). By both tests, the viral DNA is observed to sediment as part of a discrete complex with a sedimentation velocity of roughly 160S, in comparison to ribosomes (80S) and polyribosomes in the same gradients. Deproteinized unintegrated viral DNA, mixed with an extract of uninfected NIH-3T3 cells, sediments much more slowly, at about 20S (Fig. 2D), as anticipated from earlier analyses (Smotkin et al. 1975; Varmus et al. 1978). These results indicate that in an infected cell, unintegrated viral DNA resides within a large nucleoprotein complex of discrete size.

The complexes in cytoplasmic extracts typically sediment in a more discrete peak than those from nuclear extracts (see Fig. 2A,B). The circular forms of viral DNA, less abundant than the linear DNA, also sediment as large complexes. Because the linear form of DNA has been shown to be the direct precursor to MLV integration in vitro, and because complexes with circular DNA from nuclear extracts do not appear to participate in the integration reaction as studied in vitro (Fujiwara and Mizuuchi 1988; Brown et al. 1989), we have concentrated our efforts on characterizing complexes from cytoplasmic extracts in which only the linear form of viral DNA is present at detectable levels.

Integration activity copurifies with the nucleoprotein complex

We have used three independent methods of fractionation to ask if the viral nucleoprotein complexes present in our extracts are sufficient to mediate integration in vitro. Assays of integration activity across the fractions of a sucrose gradient through which a cytoplasmic extract had been sedimented reveal that the peak for activity coincides precisely with the position of the nucleoprotein complex, as measured by detection of viral DNA (Fig. 2A,C). Furthermore, the activity in the peak fractions was not stimulated significantly by the addition of either uninfected cell extracts or extracts made from cells infected by MLV-Clone 1, a virus that does not carry the *supF* gene and therefore does not score in our assay (data not shown). These results suggest that all of the components required for integration are included in the nucleoprotein complex that contains the viral DNA.

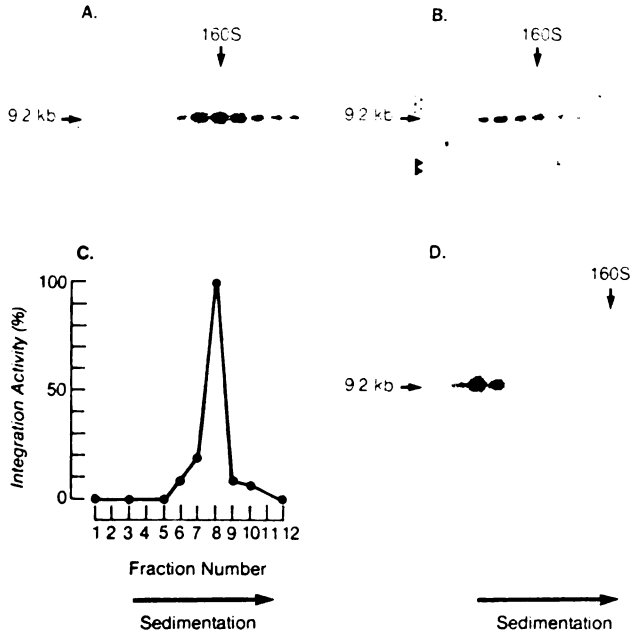


Figure 2. Retroviral DNA and integration activity cosediment in sucrose gradients. Eighteen hours after cocultivation of an MLVsupF-producing cell line with uninfected NIH-3T3 cells, cell-free extracts were prepared as described in Experimental methods, and 0.5-ml aliquots were loaded onto 15–30% sucrose gradients, which were then centrifuged at 35,000 rpm for 3 hr at 4°C in a Beckman SW-41 rotor. The gradients were divided into 12 fractions of 1 ml each. Viral DNA samples prepared from each fraction (see Experimental methods) were resolved by gel electrophoresis in 0.8% agarose and transferred to a nylon membrane. MLV DNA was identified by hybridization, using as a probe the plasmid p8.2 (a complete, permuted clone of circular MLV DNA with one LTR) labeled by nick-translation. Arrows indicate the full-length 9.2-kb linear viral DNA, the 160S peaks of polysome profiles determined by absorbance at 260 nm during the fractionation, and the direction of sedimentation. Integration activity is expressed as a percentage of the most active fraction (the numbers of recombinant plaques are presented in Table 1 in Experimental methods). (*A*) Viral DNA from gradient fractions after sedimentation of a cytoplasmic extract. (*B*) Viral DNA from gradient fractions after sedimentation of a nuclear extract. The open arrowheads indicate relaxed circles with one and two LTRs, and the closed arrowheads indicate supercoiled circles. (*C*) Integration activity across the gradient fractions shown in *A*; the fraction numbers in *B* correspond to the lanes in *A*. (*D*) The profile of viral DNA obtained after sedimentation of an uninfected cytoplasmic extract to which purified cytoplasmic viral DNA had been added.

We obtained similar results using gel-exclusion column chromatography (Fig. 3). When a cytoplasmic extract was passed through a gel filtration column with an exclusion limit of 5×10^6 daltons, the DNA-containing complexes appeared in the void volume, coincident with the integration activity. By comparison to serial twofold dilutions of the extract loaded onto the column, we estimate that we recovered between 25% and 50% of the viral DNA (Fig. 3B). A pool of the peak fractions was 36% as active for integration in vitro as the extract that was loaded onto the column (see Table 1). Thus, the recovered amounts of activity and viral DNA were roughly the same. The fractionation achieved by the column resulted in more than a thousandfold reduction in protein concentration (data not shown, but see absorbance profile, Fig. 3A), providing substantial purification. Again, addition of crude extracts did not augment the activity of the partially purified complexes.

As a further test of the association of integration activity with the nucleoprotein complex, we characterized nuclear extracts by density equilibrium centrifugation in nonionic density gradients. Analysis of gradient fractions for the presence of viral DNA showed that both the linear and circular forms produce bands at a density characteristic of a nucleoprotein complex (Fig. 4). The peak of integration activity again coincided with the position of the nucleoprotein complex (Fig. 4B), and the concentration of activity achieved (see Table 1) was comparable to that of the viral DNA (compare the intensities of bands representing the full-length linear DNA in the load and in the peak fraction of the gradient in Fig. 4A). The copurification of integration activity with the nucleoprotein complex during velocity sedimentation, gel-exclusion column chromatography, and density equilibrium centrifugation argues strongly that the nucleoprotein complex that contains the viral DNA includes all the machinery required for integration into the host genome.

After density equilibrium centrifugation, the DNA-containing complexes from cytoplasmic extracts banded at densities similar to those of complexes from nuclear extracts but lost over 90% of their integration activity. We have been unable to restore integration activity reproducibly by the addition of crude extracts (data not shown). Currently, we are unable to explain this difference in behavior between cytoplasmic and nuclear extracts.

The viral capsid protein is a component of the nucleoprotein complex

Polyclonal antisera against the viral capsid protein, the major structural component of extracellular virion cores

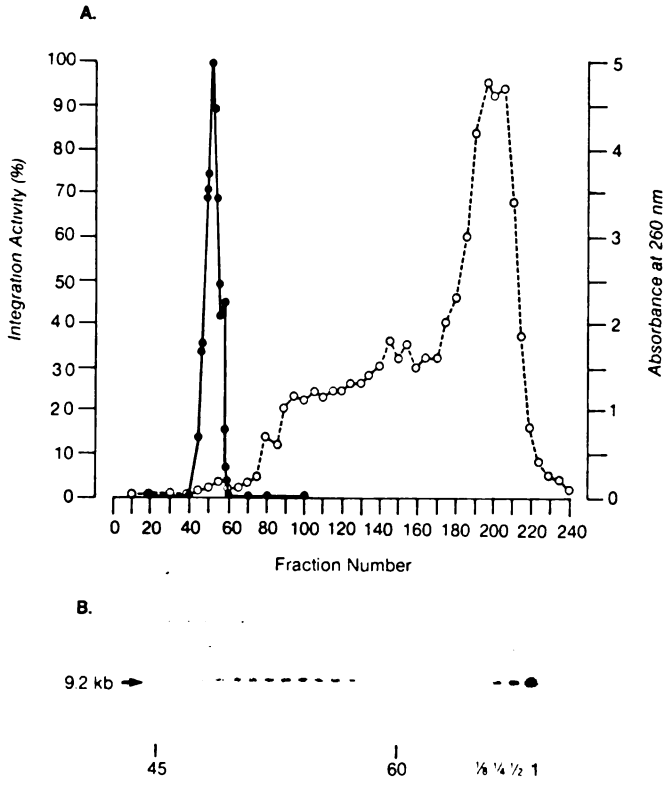


Figure 3. Copurification of the nucleoprotein complex and integration activity during gel-exclusion column chromatography. After fractionation of MLVsupF-infected NIH-3T3 cells, 5 ml of cytoplasmic extract were passed through a 200-ml BioRad A5M column. Next, 1.5-ml fractions were collected and aliquots were used to prepare viral DNA and assay integration activity. (*A*) (●) Integration activity expressed as percentage of the most active fraction. Fraction 45 is the first point with detectable integration activity, and activity is plotted for fractions 10, 40, 45–60, 70, 80, and 100. (○) Absorbance, with units shown to the right. For primary data from integration assays of the extract loaded onto the column and a pool of fractions 45–60, see Table 1. (*B*) Viral DNA samples from fractions 45–60 are to the left, after electrophoresis and hybridization as described in Fig. 2. Flanking fractions did not contain detectable levels of viral DNA (data not shown). To the right are serial twofold dilutions of viral DNA from the cytoplasmic extract that was loaded onto the column.

(see introductory section), coprecipitated both viral DNA (Fig. 5) and integration activity (Table 2) from cytoplasmic extracts of cells infected with MLVsupF, indicating that the viral capsid protein is part of the integration-competent nucleoprotein complex. However, two of the tested anti-capsid sera were unable to coprecipitate the viral DNA, even though they recognize the capsid protein from denatured virions when assayed under identical conditions (data not shown). Recognition of the nucleoprotein complex by only a subset of antisera suggests that some capsid epitopes may not be exposed. As evidenced by the presence of integration activity in the immune complex pellets, none of the antisera tested inhibited the ability of the complexes to integrate into target DNA. In addition, pretreating the extracts with quantities of antisera that are sufficient to precipitate the complexes had no effect on integration activity (data not shown).

Genetic analyses have demonstrated a role for the viral integration protein in retrovirus integration (see introductory section), but we have been unable to immunoprecipitate complexes with antisera that recognize the integration protein, reverse transcriptase, or both (data not shown). Because the nucleoprotein complex appears to include all of the activities required for integration, we suspect that at least the integration protein is included in these complexes but is inaccessible to the antibodies.

Viral DNA in nucleoprotein complex is accessible to nuclease

We have used nucleases as probes of the arrangement of DNA and protein in the complex. First, infected cell extracts were purified partially by gel filtration through a BioRad A5M column (as described in Fig. 3) and incubated with micrococcal nuclease in the presence of Ca^{2+}. Under these conditions, much of the viral DNA was digested (see Fig. 6A). Disappearance of full-length viral DNA, measured by densitometry of the 9.2-kb band remaining at each time point, roughly paralleled the loss of integration activity (Fig. 6B), suggesting that active complexes are similarly sensitive to nuclease digestion. As a control, complexes were treated with micrococcal nuclease in the absence of Ca^{2+} and shown to retain full activity throughout the time of incubation (Fig. 6B). Furthermore, addition of complexes extensively digested by micrococcal nuclease to extracts containing untreated complexes did not inhibit the integration activity of the untreated extracts (data not shown). Thus, the loss of integration activity appears to be due to cleavage of viral DNA and not to inhibition by products of the nuclease treatment.

The pattern of degradation of viral DNA by micrococcal nuclease shows a general lack of specificity. The degraded viral DNA, when visualized by hybridization to a probe for the entire viral genome, consists of a broad

Table 1. *Integration assays: recovery of target DNA and number of recombinant plaques scored*

Experiment	Packaging efficiency (pfu/μg λ target DNA)	Recombinant plaques
Velocity sedimentation (Fig. 2C)		
Fraction 6	1.0×10^8	2
Fraction 7	3.0×10^8	27
Fraction 8	1.2×10^8	64
Fraction 9	2.5×10^8	22
Fraction 10	1.1×10^8	3
Gel filtration (Fig. 3)		
Cytoplasmic extract load	4.3×10^7	467
Pool of fractions 45–60	6.2×10^7	239
Density gradient (Fig. 4)		
Nuclear extract load	3.3×10^7	357
Fraction 5	2.0×10^7	285
Fraction 6	3.4×10^7	1687
Fraction 7	3.3×10^7	139
Fraction 8	3.2×10^7	61
Fraction 9	2.4×10^7	25
Micrococcal nuclease (Fig. 6B)		
0 min	7.7×10^7	332
0.5 min	7.3×10^7	126
1.0 min	4.0×10^7	40
2.5 min	1.1×10^8	52
5.0 min	6.3×10^7	11
7.5 min	1.1×10^8	11
10.0 min	2.1×10^8	9
30.0 min	2.0×10^8	1
60.0 min	7.5×10^6	0

The numbers shown are averages for either duplicate or triplicate assays. Integration reactions were done with 100 μl of sample in a total reaction volume of 150 μl. Control reactions showed that the levels of sucrose and Nycodenz present in some samples did not affect integration activity (data not shown). Packaging efficiency (pfu/μg target DNA) reflects the recovery of phage DNA during in vitro packaging. Using this value, the number of recombinant plaques can be normalized to give the percentages shown in Figs. 2, 3, 4, and 6.

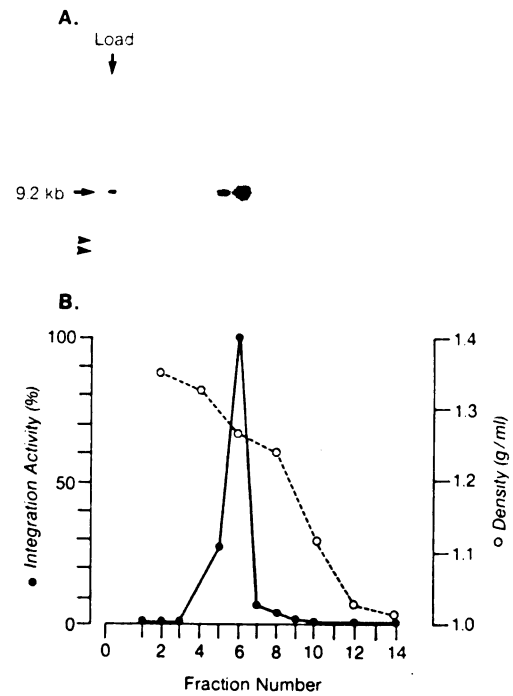


Figure 4. Retroviral DNA and integration activity band at a density appropriate for a nucleoprotein complex. After infection of NIH-3T3 cells with MLVsupF, nuclear extracts were made by breaking open nuclei and pelleting the cellular chromatin (see Experimental methods). The nuclear extract supernatant was mixed with Nycodenz, a nonionic density gradient medium, to form a 12-ml step gradient, which was centrifuged for 8 hr at 50,000 rpm at 4°C in a Beckman Ti-50 fixed-angle rotor. The gradient was divided into 14 fractions of 0.85 ml each, and aliquots were removed for measuring viral DNA, integration activity, and density. (*A*) Viral DNA in gradient fractions. Viral DNA from the extract used to make the gradient is shown at the left (lane designated 'Load') for comparison with the peak fractions. Arrow indicates the full-length 9.2-kb linear viral DNA, and the arrowheads indicate the supercoiled circles with one and two LTRs. (*B*) Plots of integration activity (•) and density (o) across the gradient. Purified DNA bands at 1.13 to 1.17 grams/ml and most proteins at densities between 2.0 and 3.0 grams/ml. The fraction numbers designated in *B* correspond to the lanes above in *A*. For primary data from integration assays, see Table 1.

smear that becomes smaller in size with increasing time of digestion (Fig. 6A and 7A). To substantiate the absence of products of discrete sizes after nuclease digestion, we mixed Rat-1 nuclei with complexes purified partially by gel filtration and treated the mixture with micrococcal nuclease. Digestion of cellular chromatin generated the expected nucleosome ladder (Fig. 7B), but the viral DNA again was degraded into a broad smear that bore no resemblance to a nucleosome ladder (Fig. 7A). The only indication of specific sensitivity to nuclease cleavage was the early appearance of a band representing a roughly 0.6-kb fragment of viral DNA discernible after hybridization to a probe for the entire viral genome (Fig. 7A). This band was more prominent—relative to the band representing full-length 8.8-kb viral DNA—when detected by hybridization to a probe specific for the viral LTR sequences (Fig. 7C). Thus, although the majority of the viral DNA in the complex appears to be uniformly sensitive to micrococcal nuclease, at least some of the viral LTR sequences are associated with a more specialized structure.

To examine further the effects of nuclease treatment on the structure of these complexes, we incubated complexes purified partially by gel filtration with a mixture of *Bgl*II and *Bam*HI restriction endonucleases (for which there are a total of seven sites in the viral DNA), and then analyzed them by velocity sedimentation in sucrose gradients. Analysis of gradient fractions for the presence and state of viral DNA showed that the complexes could still sediment as discrete particles, even

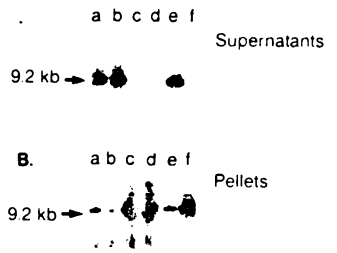


Figure 5. Coprecipitation of viral DNA with antisera that recognize the viral capsid protein. To identify viral components of the retroviral nucleoprotein complex, antisera that recognize the virion capsid protein were tested for their ability to precipitate the nucleoprotein complex and integration activity (see Table 2 for integration activity data). For each immune precipitation, 300 μl of cytoplasmic extract from cells acutely infected by MLVsupF and 5 μl of antiserum were mixed on ice in buffer A with 0.5% NP-40, and immune complexes were formed by the addition of fixed *S. aureus* cells. The complexes were pelleted by centrifugation, and the supernatants were saved for analysis of viral DNA and integration activity. The pellets were washed three times with buffer A containing 0.5% NP-40, and the washed pellets were resuspended in the same buffer and also analyzed for viral DNA and integration activity. (*A*) Viral DNA from supernatants after removal of immune complexes. The arrow designates the full-length 9.2-kb linear viral DNA. Samples: Goat nonimmune serum (lane *a*); goat anti-MLV (whole virus) serum (lane *b*); goat anti-MLV capsid serum #1 (lane *c*); rabbit anti-MLV capsid serum #1 (lane *d*); rabbit anti-MLV capsid serum #2 (lane *e*); goat anti-MLV capsid serum #2 (lane *f*). (*B*) Viral DNA from immune complex pellets. Lanes are same as in *A*. Viral DNA samples were prepared and analyzed as described in Fig. 2.

when the viral DNA had been cleaved at several sites, generating fragments as small as 1 kb in size (Fig. 8). Complexes that were mock-treated for the same period of time sedimented to the same position in control gradients (data not shown). Although we have seen sometimes a slight staggering of smaller fragments toward the

Table 2. *Coprecipitation of retroviral integration activity by anti-capsid antiserum*

Experiment	Packaging efficiency (pfu/μg target DNA)	Recombinant plaques
Immune precipitations		
Goat nonimmune serum		
Supernatant	8.3×10^7	240
Pellet	2.5×10^7	3
Goat anti-MLV capsid serum #1		
Supernatant	4.5×10^7	9
Pellet	2.5×10^6	90

Integration reactions were done with 100 μl of sample in a total volume of 150 μl. See Fig. 5, Table 1, and Experimental methods for descriptions of immunoprecipitation procedures and in vitro integration assays.

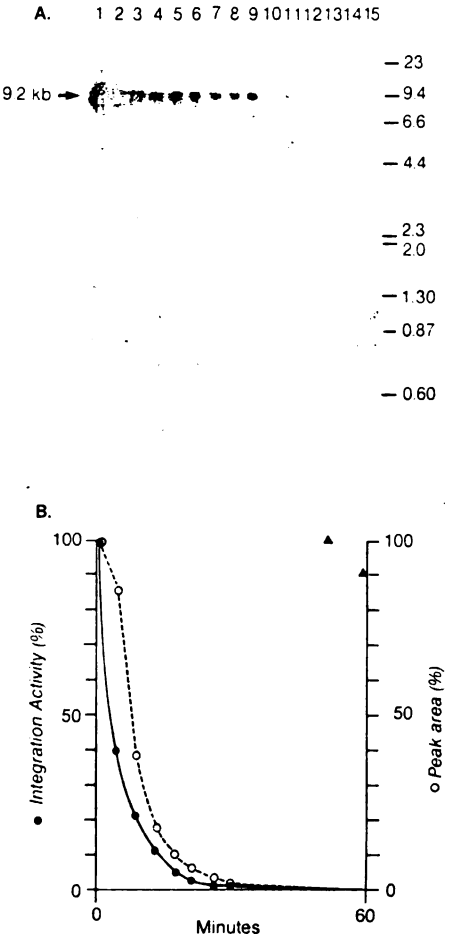


Figure 6. Viral DNA in the integration-competent nucleoprotein complex is accessible to micrococcal nuclease. A cytoplasmic extract made from cells acutely infected by MLVsupF was purified partially by gel filtration (as described in Fig. 3). A total of 3.5 ml of the partially purified extract was incubated on ice with micrococcal nuclease in the presence of 1 mM $CaCl_2$. Digestion was stopped by removal of samples at the times indicated below; these were adjusted to 2 mM EGTA and analyzed for viral DNA and integration activity. The relative quantities of full-length viral DNA remaining at each time point were determined by densitometry to obtain peak areas of the 9.2-kb band (indicated by arrow) remaining at each time point. (*A*) Viral DNA from time course of micrococcal nuclease incubation, after electrophoresis in 1.0% agarose and hybridization, as described in Fig. 2. (Lanes *1–15*) Digestions stopped at 0 (before addition of $CaCl_2$), 0.5, 1.0, 2.5, 5.0, 7.5, 10, 12.5, 15, 17.5, 20, 30, 45, and 60 min. Size markers (in kilobases) are to the right. (*B*) Plot of integration activity (●) and peak area (○) after scanning the 9.2-kb band shown in *A*. Time points plotted correspond to lanes *1–7* shown in *A*. (▲) Integration activity of control samples incubated on ice with micrococcal nuclease in the absence of $CaCl_2$ for 30 and 60 min. For the primary data from integration assays, see Table 1.

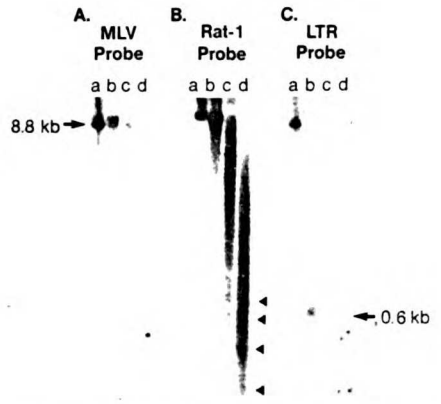


Figure 7. Viral DNA in the nucleoprotein complex differs from chromatin in its pattern of sensitivity to micrococcal nuclease. Nuclei from uninfected Rat-1 cells, prepared as described in Experimental methods, were resuspended in buffer B and mixed with an equal volume of cytoplasmic extract made from cells acutely infected by MLV-Clone 1 (a wild-type MLV strain that lacks the *supF* gene) and purified partially by gel filtration through a BioRad A5M column (see Fig. 3). The mixture was adjusted to 1 mM $CaCl_2$ and incubated on ice with micrococcal nuclease. Digestion was stopped by addition of 2 mM EGTA (final concentration) to aliquots removed at the times indicated. DNA was prepared and samples resolved by gel electrophoresis in 1.5% agarose. After transfer to a nylon membrane, sequential hybridization to nick-translated probes was used to distinguish viral and Rat-1 chromosomal sequences. (Lanes *a–d*) 0, 5, 20, and 40 min time points. (*A*) Hybridization to the probe for the entire MLV genome (made by nick-translation of the plasmid p8.2). An arrow indicates full-length 8.8-kb linear viral DNA. (*B*) Hybridization to nick-translated Rat-1 chromosomal DNA. The closed arrowheads indicate the nucleosome ladder generated by nuclease digestion of cellular chromatin. (*C*) Hybridization to the probe specific for viral LTR sequences (made by nick-translating the plasmid pGEMLTR, a permuted clone of the MLV LTR). An arrow indicates the ~600-bp fragment specific for the LTR probe.

top of the gradient in such experiments (data not shown), the sedimentation properties of the complex appear to be largely unaffected by cleavage of the viral DNA within them.

Discussion

The principal result we present here is that unintegrated retroviral DNA isolated in its native state from acutely infected cells resides within a large nucleoprotein complex. This complex sediments at 160S and carries with it all of the activities necessary for integration in vitro. Copurification of the complex and integration activity also is observed during gel-exclusion column chromatography and density equilibrium centrifugation.

These results suggest that retroviral DNA within infected cells remains associated with the protein machinery responsible for integration into the host genome. Such a solution to the requirement for integration is consistent with the idea that retroviruses bring into infected cells all the activities needed for reverse transcription and integration (Varmus and Brown 1989). A tight association of the viral nucleic acids with the viral replication machinery avoids any dilution of essential factors in the host cytoplasm and nucleus, allowing the virus to perform essential steps in replication with only a limited number of the necessary proteins. If any host components participate in these events, presumably they do so through association with these complexes.

Purification of the active complex eventually should allow the direct identification of its components. However, the extracts we make from acutely infected cells exhibit substantial variability in their activity, with as few as 0.1% to as many as 50% of the complexes present in an extract being active for integration in vitro (Brown et al. 1987, 1989; Fujiwara and Mizuuchi 1988; P. Pryciak and P. Brown, unpubl.). Because about 1% of the complexes present in the extracts used here were active in vitro, conclusions based on direct purification of the entire population might be misleading. Instead, we have focused initially on indirect methods of analysis that allow us to examine active complexes by the use of an in vitro assay for integration activity. In these experiments, we also have examined the entire population present in cell-free extracts by detection of the viral DNA after gel electrophoresis. In general, the active population and the

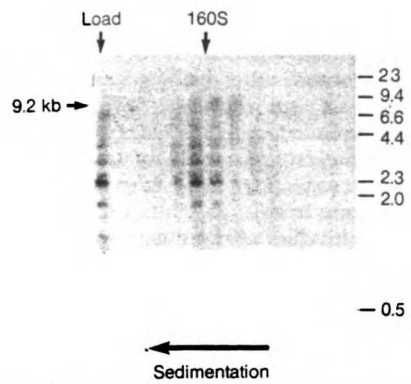


Figure 8. Cleavage of viral DNA in the nucleoprotein complex does not affect its sedimentation behavior in sucrose gradients. Cytoplasmic extract made from cells acutely infected by MLVsupF was purified partially by gel filtration through a BioRad A5M column (see Fig. 3) and digested on ice with *Bgl*II and *Bam*HI restriction enzymes for 3 hr. The digestion was stopped by the addition of 10 mM EDTA. After removing an aliquot for the lane designated 'Load', the remaining extract was subjected to velocity sedimentation in a 15–30% sucrose gradient (see Experimental methods). The gradient was divided into 13 fractions of 1 ml each. Viral DNA was prepared from each fraction, resolved by gel electrophoresis in 1.0% agarose, and analyzed as described in Fig. 2. The arrow indicates the position of full-length 9.2-kb linear viral DNA. Size markers (in kilobases) are to the right. The 160S arrow indicates the position of the nucleoprotein complex in a control gradient in which the extract was mock-treated. Direction of sedimentation is indicated by the arrow at the bottom.

total population show similar behavior. The one exception has been our inability to concentrate integration activity when we band complexes from cytoplasmic extracts in density equilibrium gradients. This is in contrast to banding nuclear complexes in density gradients, in which case both integration activity and the nucleoprotein complex are concentrated.

Immunoprecipitation of integration-competent particles has enabled us to show that the viral capsid protein is one component of the active complex. The capsid protein of MLV is the major structural component of the extracellular virion core, in which it forms the icosahedral shell that encases the viral nucleocapsid. The ability to precipitate the nucleoprotein complex and integration activity with antisera that recognize capsid suggests that, after entry into the cell and completion of reverse transcription, unintegrated viral DNA remains associated with a structure that still retains at least some of the capsid protein. In addition, the failure of some anti-capsid sera to precipitate the complexes implies that conformational constraints obscure some antigenic determinants.

The presence of at least some of the capsid in the nucleoprotein complex is of interest because of the role this protein is thought to play in the inability of certain strains of MLV, called N-type and B-type viruses, to replicate in cell lines derived from particular inbred strains of mice, a phenomenon termed *Fv-1* restriction (Jolicoeur 1979). In restricted infections, virus can enter the cell and synthesize viral DNA but does not appear capable of establishing a provirus. N- and B-tropism map to the viral capsid protein; if the coding sequences for capsid are interchanged between an N- and a B-type virus, their tropism is reversed (Jolicoeur 1979; DesGroseillers and Jolicoeur 1983). The presence of the viral capsid protein in the nucleoprotein complex that contains unintegrated viral DNA raises the interesting possibility that *Fv-1* restriction may operate through an interaction with the capsid protein components of the nucleoprotein complex that contains the viral DNA inside infected cells.

The use of nucleases as a probe of structure has revealed that much of the unintegrated viral DNA in these complexes is accessible to exogenous macromolecules. The pattern of degradation of the viral DNA within the complexes shows little pattern, unlike the nucleosome ladder generated by nuclease treatment of host chromatin. At least some of the viral LTR sequences, however, show a more specific sensitivity to treatment with micrococcal nuclease. The LTRs are the only viral sequences required in *cis* for integration, and their pattern of sensitivity to micrococcal nuclease suggests the presence of specialized structures associated with LTR sequences.

Viral DNA in the complex also is accessible to the restriction endonucleases *Bgl*II and *Bam*HI, but complexes that have been cleaved at several sites have unaltered sedimentation properties. This result argues against models for the structure of these complexes in which large stretches of viral DNA are left essentially free, with the protein components arranged only at the ends of the viral DNA. Models in which the proteins and viral DNA are in some type of 'beads on a string' arrangement also would be inconsistent with these results. The viral DNA in the complex is not assembled into nucleosomes, in contrast to host chromatin (Kornberg 1977; Felsenfeld 1978) and papovavirus minichromosomes (Griffith 1975; Muller et al. 1978). More consistent with our results are models in which the viral DNA resides within a protein shell that is permeable to nucleases. Given that the MLV capsid protein forms an icosahedral shell in the virion core and appears to be a component of the integration-competent complex, we favor this possibility. The simplest view is that the integration-competent complex is similar to and derived from the extracellular virion core. However, we cannot rule out models in which the viral DNA is on the outside of a nucleoprotein complex that remains intact after DNA cleavage because of other interactions. We hope to resolve these issues by electron microscopy of purified complexes.

The best-understood example of DNA recombination that resembles retroviral integration is the nonreplicative transposition reaction of the bacteriophage Mu. Reconstitution in vitro of the Mu transposition reaction has shown that a highly ordered and tightly bound complex of DNA and protein is required for activity (Craigie and Mizuuchi 1985; Surette et al. 1987). Highly ordered nucleoprotein complexes also are involved in a variety of other functions that involve precise manipulation of DNA. Examples include the multicomponent arrangements of proteins at replication origins in *E. coli* and phages λ and T4, and the intasome intermediate in the integration reaction of phage λ (for reviews, see Alberts 1984; Echols 1986; Bramhill and Kornberg 1988; Friedman 1988). Further characterization of the integration-competent nucleoprotein complexes described here may prove useful in designing attempts to reconstruct MLV integration activity with purified components, and contribute to a description of the molecular events that constitute retroviral integration. Finally, a better understanding of retroviral integration should prove relevant to studies of retrotransposons, such as Ty elements in yeast, which also employ a nucleoprotein complex to integrate into the host genome (Boeke et al. 1985; Eichinger and Boeke 1988), but which do not have an extracellular phase in their transposition cycle (for review, see Roeder and Fink 1983).

Experimental methods

Cell culture, cellular extracts, and integration reactions

Procedures for infection of NIH-3T3 cells by cocultivation with MLVsupF-producing NIH-3T3 cells, harvesting and fractionation of infected cell cultures into cytoplasmic and nuclear extracts, and integration reactions were nearly the same as described previously (Brown et al. 1987). However, cells were harvested at 18 hr postinfection instead of 24 hr, and integration reactions included 1.0 μg of λ target DNA instead of 1.4 μg. Integration reactions were normalized for recovery of the target DNA after in vitro packaging of the DNA recovered from reac-

tions. In Figures 2, 3, 4, and 6, integration activity was plotted as a percentage of the most active fraction in each experiment. Recovery of target DNA and the number of recombinant plaques scored are indicated in Table 1.

DNA analysis

For preparation of DNA, samples were adjusted to 10 mM EDTA, 0.5% *N*-lauroyl sarcosine, and 0.5 mg/ml of proteinase K; incubated for 2 hr at 55°C; extracted twice with phenol/chloroform (1 : 1); and precipitated by adjusting the final salt concentration to 300 mM with NaOAc, adding 4 μg of yeast RNA carrier and 2.5 volumes of 95% ethanol. Nick-translation for hybridization probes, agarose gel electrophoresis, and Southern blotting were performed using standard methods (Maniatis et al. 1982). Filter hybridization followed the method of Church and Gilbert (1984). Hybond-N nylon membranes were used for transfer of DNA following gel electrophoresis. Probes were removed from the filter between hybridizations by washing in a solution of 80% formamide, 10 mM Tris-HCl (pH 7.5), 12.5 mM EDTA, 1% SDS in a shaking water bath at 65°C. Comparisons of the amount of viral DNA in the starting extracts with the amounts of viral DNA recovered after gel filtration (Fig. 3) and density equilibrium centrifugation (Fig. 4) were done using equal volumes of the corresponding samples. Although these data do not allow a determination of the total recovery in the two experiments, they demonstrate that the specific activity (integration events/linear DNA molecule) of the starting material and the recovered material are roughly equivalent.

Velocity sedimentation gradients

Continuous sucrose gradients were poured with a two-chamber, 15-ml Hoeffer Gradient maker using 15% and 30% sucrose solutions in buffer A [10 mM Tris-HCl (pH7.4), 225 mM KCl, 5 mM $MgCl_2$, 1 mM DTT, 20 μg/ml aprotinin] and kept on ice. Gradients were overlayed with 0.5 ml of cytoplasmic or nuclear extract and centrifuged at 35,000 rpm for 3 hr at 4°C in a Beckman SW-41 rotor. Gradients were fractionated from the top into 12 fractions of 1 ml with a Buchler Densi Flow II apparatus. Polyribosome peaks were detected during the fractionation by absorbance at 260 nm with an ISCO UA-5 flowthrough spectrophotometer system. Each lane in Figure 2, A, B, and D, represents one-quarter of the corresponding fraction.

Gel-exclusion column chromatography

A 200-ml BioRad A5M column 30 cm in height was equilibrated at 4°C with 2 liters of buffer B [10 mM Tris-HCl (pH 7.4), 100 mM KCl, 5 mM $MgCl_2$, 1 mM DTT, 20 μg/ml aprotinin]. The lower salt concentration was chosen to facilitate nuclease treatments in the experiments described below. Integration reactions were adjusted to return salt levels to standard assay conditions. Then, 5.0 ml of cytoplasmic extract was loaded onto the column and 1.5-ml fractions were collected with an ISCO Retriever II fraction collector. The flow rate was 1.0 ml/min. Next, 100-μl aliquots of the fractions were diluted to 500 μl in water, and absorbance at 260 nm was determined with a Beckman DU-40 Spectrophotometer. The absorbance values shown in Figure 3A were corrected for the dilution. Each lane in Figure 3B represents 25 μl of the corresponding fraction.

Equilibrium density gradients

A three-step discontinuous gradient was poured using approximately 4 ml each of nuclear extract adjusted to 60%, 30%, and 0% Nycodenz by mixing nuclear extract with an appropriate volume of 80% Nycodenz made in buffer A (Nycodenz is a nonionic density gradient medium available from Accurate Scientific and Chemical Corporation). The gradient was poured in a heat-seal tube and kept at 4°C throughout all manipulations. Centrifugation was at 50,000 rpm in a Beckman Ti-50 fixed-angle rotor at 4°C for 8 hr. Fractions were collected by dripping from the bottom of the gradient. Density was determined by measuring absorbance at 360 nm and using the formula: Density (gm/ml) = 0.135 × OD_{360} + 1.0. Density of the peak fraction was confirmed by refractometry. Density values are affected by the fact that macromolecules are hydrated extensively in the nonionic medium. Each lane in Figure 3A represents 50 μl of the corresponding fraction.

Immune precipitations

For each sample tested, 300 μl of cytoplasmic extract was adjusted to 0.5% NP-40, mixed with 5 μl of antiserum, and incubated on ice for 1 hr. Thirty microliters of 10 mg/ml Pansorbin (fixed *Staphylococcus aureus* cells from CalBiochem) was added and again the mixture was incubated for 1 hr on ice. Immune complexes were pelleted for 20 sec in an Eppendorf centrifuge at 4°C. The supernatants were saved for analysis of viral DNA and integration activity. The immune complex pellets were washed three times with 1.0 ml of buffer A containing 0.5% NP-40. The final washed pellets were resuspended in 300 μl of buffer A with 0.5% NP-40 and analyzed for viral DNA and integration activity. Each lane in Figure 5 represents 50 μl of the corresponding sample. The nonimmune goat serum, the goat anti-MLV (whole virus) serum, and the goat anti-MLV capsid serum m1 were from the Biological Carcinogenesis Branch, DCCP, at the National Cancer Institute. The rabbit anti-MLV capsid sera m1 and m2, and the goat anti-MLV capsid serum m2 were gifts from Hung Fan (University of California, Irvine).

Nuclease treatments

All nuclease treatments were done in buffer B with cytoplasmic extract that had been purified partially by gel-exclusion column chromatography as described above. Micrococcal nuclease (Boehringer–Mannheim) was added to the partially purified cytoplasmic extract, which was then adjusted to 1 mM $CaCl_2$ and incubated on ice. Aliquots were removed at indicated times, adjusted to 2 mM EGTA, and kept on ice until all samples were collected. Then samples were analyzed for viral DNA and integration activity. Each lane in Figure 6A represents 50 μl of the corresponding sample. For the experiment shown in Figure 7, nuclei were prepared by trypsinizing a confluent monolayer of Rat-1 cells (100 mm in diameter). The cells were washed twice in buffer C [10 mM Tris-HCl (pH 7.4), 100 mM KCl, 1 mM EDTA, 1 mM DTT, 20 μg/ml aprotinin], lysed with 0.2% Triton X-100 in buffer C, the nuclei pelleted at 1000*g*, and washed twice with buffer C. The final pellet of roughly 5 × 10^6 nuclei was resuspended in 1.0 ml of buffer B and mixed with 1.0 ml of partially purified cytoplasmic extract. Each lane in Figure 7 represents 250 μl of the corresponding sample. For the experiment shown in Figure 8, the restriction digest with *Bgl*II and *Bam*HI (New England Biolabs) was done on ice for 3 hr with 100 units of each enzyme in 0.5 ml of partially purified cytoplasmic extract. The digestion was stopped by the addition of 10 mM EDTA and the sample was loaded onto a 15–30% sucrose gradient in buffer A and centrifuged for 3 hr at 35,000 rpm in an SW-41 rotor at 4°C.

Acknowledgments

We thank Titia de Lange and Peter Pryciak for helpful comments on the manuscript, and Hung Fan and Stephen Goff for providing antisera. This work was supported by funds from the National Institutes of Health, the G.W. Hopper Foundation, the Howard Hughes Medical Institute, and the Lucille P. Markey Charitable Trust. H.E.V. is an American Cancer Society Research Professor. P.O.B. was a Lucille P. Markey Scholar for a portion of this work and an assistant investigator of the Howard Hughes Medical Institute for the remainder.

References

Alberts, B. 1984. The DNA enzymology of protein machines. *Cold Spring Harbor Symp. Quant. Biol.* **49**: 1–12.

Boeke, J.D., D.J. Garfinkel, C.A. Styles, and G.R. Fink. 1985. Ty elements transpose through an RNA intermediate. *Cell* **40**: 490–500.

Bramhill, D. and A. Kornberg. 1988. A model for initiation at origins of DNA replication. *Cell* **54**: 915–918.

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

———. 1989. Retroviral integration: Structure of the initial covalent product and its precursor, and a role for the viral integrase protein. *Proc. Natl. Acad. Sci.* (in press).

Church, G.M. and W. Gilbert. 1984. Genomic sequencing. *Proc. Natl. Acad. Sci.* **81**: 1991–1995.

Craigie, R. and K. Mizuuchi. 1985. Mechanism of transposition of bacteriophage Mu: Structure of a transposition intermediate. *Cell* **41**: 867–876.

DesGroseillers, L. and P. Jolicoeur. 1983. Physical mapping of the *Fv-1* tropism host range determinant of *Balb/c* murine leukemia viruses. *J. Virol.* **48**: 685–696.

Dickson, C., R. Eisenman, H. Fan, E. Hunter, and N. Teich. 1985. Protein biosynthesis and assembly. In *RNA tumor viruses* (ed. R. Weiss et al.), vols. 1 and 2. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

Donehower, L. and H.E. Varmus. 1984. A mutant murine leukemia virus with a single missense codon in *pol* is defective in a function affecting integration. *Proc. Natl. Acad. Sci.* **80**: 6461–6465.

Duyk, G., J. Leis, M. Longiaru, and A.M. Skalka. 1983. Selective cleavage on the avian retroviral long terminal repeat sequence by the endonuclease associated with the form of avian reverse transcriptase. *Proc. Natl. Acad. Sci.* **80**: 6745–6749.

Echols, H. 1986. Multiple DNA-protein interactions governing high-precision DNA transactions. *Science* **233**: 1050–1056.

Eichinger, D.J. and J.D. Boeke. 1988. The DNA intermediate in yeast Ty1 element transposition copurifies with virus-like particles: Cell-free Ty1 transposition. *Cell* **54**: 955–966.

Felsenfeld, G. 1978. Chromatin. *Nature* **271**: 115–121.

Freidman, D.I. 1988. Integration host factor: A protein for all reasons. *Cell* **55**: 545–554.

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

Grandganett, D.P. and A.C. Vora. 1985. Site specific nicking at the avian retrovirus LTR circle junction by the viral pp32 endonuclease. *Nucleic Acids Res.* **13**: 6205–6221.

Grandganett, D.P., A.C. Vora, R. Swanstrom, and J. Olsen. 1986. Nuclease mechanism of the avian retrovirus pp32 endonuclease. *J. Virol.* **58**: 970–974.

Griffith, J.D. 1975. Chromatin structure: Deduced from a minichromosome. *Science* **187**: 1202–1203.

Jolicoeur, P. 1979. The *Fv-1* gene of the mouse and its control of murine leukemia virus replication. *Curr. Topics Microbiol. Immunol.* **86**: 67–122.

Kornberg, R.D. 1977. Structure of chromatin. *Annu. Rev. Biochem.* **46**: 931–955.

Lobel, L., M. Patel, W. King, M. Nguyen-Huu, and S. Goff. 1985. Construction and recovery of viable retroviral genomes carrying a bacterial repressor transfer RNA gene. *Science* **228**: 329–332.

Maniatis, T., E.F. Fritsch, and J. Sambrook. 1982. *Molecular cloning: A laboratory manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

Muller, U., H. Zentgraf, I. Eicken, and W. Keller. 1978. Higher order structure of simian virus 40 chromatin. *Science* **201**: 406–414.

Panganiban, A.T. and H.M. Temin. 1984a. Circles with two LTRs are precursors to integrated retrovirus DNAs. *Cell* **36**: 673–679.

———. 1984b. The retrovirus *pol* gene encodes a product required for DNA integration: Identification of a retrovirus *int* locus. *Proc. Natl. Acad. Sci.* **81**: 7885–7889.

Roeder, G.S. and G.R. Fink. 1983. Transposable elements in yeast. In *Mobile genetic elements* (ed. J. Shapiro), pp. 300–328. Academic Press, New York.

Schwartzberg, P., J. Colicelli, and S.P. Goff. 1984. Construction and analysis of deletion mutations in the *pol* gene of murine leukemia virus: A new viral function required for productive infection. *Cell* **37**: 1043–1052.

Smotkin, D., A. Gianni, S. Rozenblatt, and R. Weinberg. 1975. Infectious viral DNA of murine leukemia virus. *Proc. Natl. Acad. Aci.* **72**: 4910–4913.

Surette, M.G., S.J. Buch, and G. Chaconas. 1987. Transpososomes: Stable protein-DNA complexes involved in the in vitro transposition of bacteriophage Mu DNA. *Cell* **49**: 253–262.

Varmus, H.E. 1983. Retroviruses. In *Mobile genetic elements* (ed. J. Shapiro), pp. 411–503. Academic Press, New York.

Varmus, H.E. and P.O. Brown. 1989. Retroviruses. In *Mobile DNA* (ed. M. Howe and P. Berg), Amercian Society for Microbiology Publications, Washington, D.C. (in press).

Varmus, H.E. and R. Swanstrom. 1985. Replication of retroviruses. In *RNA tumor viruses*, vols. 1 and 2, (ed. R. Weiss et al.). Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

Varmus, H.E., S. Heasley, K.-J. Kung, H. Oppermann, V.C. Smith, J.M. Bishop, and P.R. Shank. 1978. Kinetics of synthesis, structure, and purification of avian sarcoma virus-specific DNA made in the cytoplasm of acutely infected cells. *J. Mol. Biol.* **120**: 55–82.

Chapter 4

Unintegrated Viral DNA Isolated from Cells Acutely Infected by Murine Leukemia Virus Copurifies with Derivatives of the Extracellular Virion Core

ABSTRACT.

After infection of NIH 3T3 cells with the Moloney murine leukemia virus (MLV), three viral proteins--reverse transcriptase (RT), the integration protein (IN), and capsid (CA)--persist in cytoplasmic extracts made 14 hours after initiation of infection. Two of these proteins, RT and IN, are stably associated with large complexes that copurify with unintegrated viral DNA during gel exclusion column chromatography and density equilibrium centrifugation. CA appeared to be less stable in its association with large complexes during centrifugation in density gradients. The identity of these three viral proteins was confirmed by the use of specific antisera. One viral protein, matrix (MA), was not detected in partially purified complexes. The stoichiometry of the viral proteins and the unintegrated viral DNA that copurify in density gradients suggest that most of the viral proteins detected are associated with complexes that do not contain full-length, unintegrated viral DNA.

Upon entry into susceptible cells, retroviruses reverse transcribe their plus-stranded RNA genome into a linear, double-stranded DNA molecule with LTRs at each end (Varmus and Swanstrom 1982, 1985). For the Moloney murine leukemia virus (MLV), the linear form of viral DNA integrates into the host genome to form the provirus (Brown et al 1987, 1989; Fujiwara and Mizuuchi 1988; Varmus and Brown 1989). The integration reaction requires the integration protein (IN), encoded by the viral *pol* gene, and is essential for the subsequent expression of viral gene products (Schwartzberg et al 1984; Donehower and Varmus 1984; Panganiban and Temin 1984).

We have recently shown that integration of viral DNA is mediated by a large (160S) nucleoprotein complex that includes all the activities required for integration in an in vitro assay (Bowerman et al 1989). We showed by immunoprecipitation that unintegrated viral DNA remains associated with (CA), the major structural protein of the extracellular virion core. CA, encoded by the viral *gag* gene, forms the polyhedron shell that encloses the viral nucleocapsid. The nucleocapsid contains the viral RNA genome and additional viral proteins encoded by *gag* and *pol*. These include IN, reverse transcriptase (RT), and nucleocapsid (NC). Another *gag* protein, matrix (MA) is thought to reside between the CA-derived polyhedron shell and the lipid bilayer envelope (Dickson et al 1982, 1985).

Here, we present evidence that IN and RT copurify in density gradients with the DNA-containing nucleoprotein complex. In these experiments, the stoichiometry of the copurifying viral proteins and DNA

suggests that other viral complexes (which do not contain full-length viral DNA) copurify with the DNA-containing complexes and probably account for most of the viral proteins we are able to detect. At least one viral protein, the *gag*-encoded MA, is not present in these complexes at detectable levels.

To identify viral proteins that are present in acutely infected cells before the majority of viral DNA is integrated, we labeled virions (MLVsupF; Brown et al 1987) by growing a producer cell line in Dulbecco's Modified Eagle's medium (DME) lacking methionine and cysteine, supplemented with 2% dialyzed fetal calf serum (Gibco), ^{35}S-methionine and -cysteine (Amersham Tran^{35}Slabel), and 0.1% sodium bicarbonate. 40 nearly confluent 150mm plates of producer cells were grown in the radioisotope-containing medium (20ml/plate) for eight hours at 37°C, at which time the medium was removed. This pooled stock of labeled virus was supplemented with 1mM each of unlabeled methionine and cysteine, 10% fetal calf serum, and 8 micrograms/ml polybrene, and used to infect 40 150mm plates of NIH 3T3 cells. After allowing one hour for the virus to adsorb onto the cells, the radiolabeled virus stock was removed, the plates washed with PBS, and the medium replaced with DME and 10% fetal calf serum.

Fourteen hours after infection, the cells were fractionated as previously described into 12.5 ml of a cytoplasmic extract (Bowerman, et al, 1989). The nucleoprotein complexes containing unintegrated viral DNA from 10 ml of the extract were partially purified by gel exclusion column chromatography through a BioRad A5M column, which has an

exclusion limit of 5 x 10^6 daltons. 9 ml of the partially purified complexes were concentrated by centrifugation in a nonionic density gradient, which was dripped into 12 fractions of one ml (see Bowerman et al 1989).

The gradient fractions were analyzed for the presence of viral DNA and for labeled viral proteins. The results of the DNA and protein analysis are shown in Figure 1. The DNA-containing complexes were concentrated in the gradient, as shown by the analysis of viral DNA in Figure 1A. Three viral proteins--**CA**, **IN**, and **RT**--can be identified based on their comigration with the corresponding virion proteins. An additional protein of about 55 kD is also present. This 55 kD protein is recognized by antisera raised against an **RT/IN** fusion protein (data not shown). This protein has been observed in virus stocks by other investigators, but varies in quantity from preparation to preparation (Stephen Goff, Bob Craigie; personal communications). **RT** and **IN** are both present in the fractions that contain viral DNA, while **CA** appears throughout most of the gradient.

Quantitation of the viral DNA and protein present in the peak fraction of the density gradient (lane 6 in both Figure 1A and 1B) revealed that most of the viral proteins we can detect are probably associated with structures that do not contain viral DNA. The quantitation was done by determination of the specific activity of the labeled viral proteins and densitometric scanning of autoradiograms (data not shown). The peak fraction of the density gradient was found to contain about 1 x 10^7 copies of linear viral DNA molecules per 100

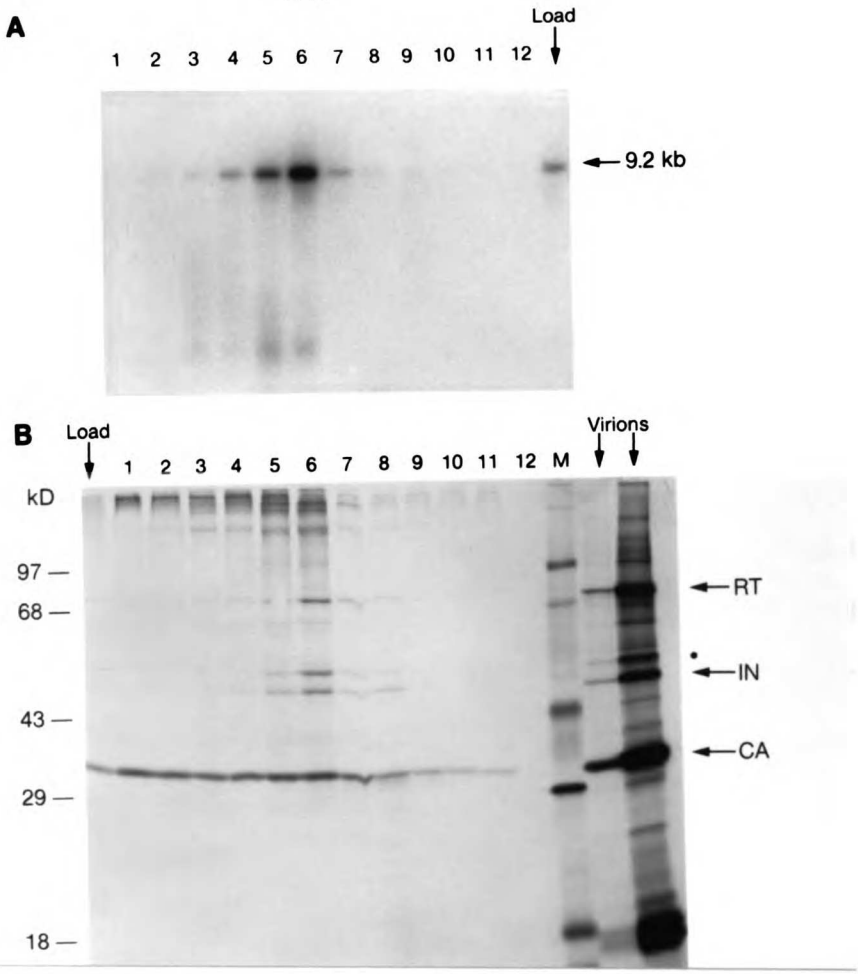


Figure 1. Unintegrated Viral DNA and a Subset of Viral Proteins Copurify after Centrifugation in Nonionic Density Gradients.

Cytoplasmic extract from cells acutely infected by metabolically-labeled MLVsupF was partially purified by gel exclusion column chromatography and density equilibrium centrifugation (Bowerman et al 1989). The gradient was dripped from the bottom into twelve fractions of one ml. (A) Viral DNA was prepared from 100 microliters of each fraction and resolved by agarose gel electrophoresis. The viral DNA was transferred to a nylon membrane and detected by hybridization to a ^{32}P-labeled probe. 100 microliters of extract partially purified by gel filtration, which was used to make the gradient, is shown in the lane marked "load." An arrow indicates the 9.2 kb full-length linear viral DNA. (B) Metabolically-labeled viral proteins were TCA-precipitated from 100 microliters of each gradient fraction and resolved by electrophoresis in a 10% denaturing polyacrylamide gel. The two lanes at the right, labeled "virions," contain virus particles pelleted from the medium used for the infection. The position of the viral proteins **RT**, **IN**, and **CA**, are indicated by arrows at the right. The 55 kD protein, which is not always present in virus preparations (see Figure 2B and the text), is indicated by a dot at the right. Fraction numbers are indicated above each lane. The lane marked "M" contains molecular weight markers, the size of which are shown in kD to the left.

microliters. Both RT and IN are present at roughly 1000 copies per DNA molecule, about ten times more than would be expected based on the quantities of these proteins that are present in extracellular virus particles (Dickson et al 1985). Dot blot analysis after density equilibrium centrifugation of acutely infected cell extracts has shown that complexes with viral RNA peak in the same fractions as those containing unintegrated viral DNA (data not shown).

CA was calculated to be present at about 3200 molecules per linear DNA molecule in the peak gradient fraction. If the calculation includes the CA present in all of the gradient fractions, then there are about 21,000 copies per linear DNA molecule. The latter figure, in comparison to the quantities of IN and RT in the peak fraction, is consistent with the ratio of *gag* to *pol* proteins seen in the extracellular virus particle (Dickson et al 1982, 1985).

To confirm the identity of the viral proteins shown in Figure 1, we repeated the experiment with the following differences to allow detection of the viral proteins using immunoblotting techniques. The virus stock was not labeled with radioisotope, and the proteins from the density gradient fractions, after electrophoresis in a 12% denaturing polyacrylamide gel, were transferred to a nitrocellulose filter. The filter was preincubated with 2% gelatin in TBST (10 mM TRIS.HCl, pH 8.0; 150 mM NaCl; 0.5% Triton-X100), and then probed with a rabbit anti-serum that recognizes both RT and IN (a gift from Stephen Goff). After washing the filter in TBST, the filter was treated with a second anti-rabbit antibody conjugated to alkaline phosphatase, and RT and IN were

detected by exposing the blot to the color reagents NBT and BCIP (Promega). To detect CA, the antibodies were removed from the filter with two fifteen minute washes at 68°C in TSB (100 mM TRIS.HCl, pH 6.8; 2% SDS; 100 mM BME). The filter was again preincubated with 2% gelatin in TBST and probed with a goat antiserum that recognizes CA (NCI 77S195). The filter was washed with TBST and the bound antibodies detected with ^{125}I-Protein G (Amersham), followed by autoradiography.

As in the previous experiment, the nucleoprotein complexes were concentrated in the density gradient, as shown by the analysis of viral DNA in Figure 2A. Panels B and C in Figure 2 show the proteins detected by the RT/IN and CA antisera, respectively. These data confirm the identity of the viral proteins suggested by the metabolic labeling experiment shown in Figure 1. The 55 kD protein was not detected in this experiment.

We have also found that one viral protein, MA, is not present at detectable levels in association with the viral complexes present in the cytoplasm of acutely infected cells. These results are shown in Figure 3. The immunoblotting analysis was performed in the same manner as described for Figure 2, except that we examined extracts that had been partially purified only by gel filtration. In Figure 3, the first two lanes in each panel contain two different amounts of purified virions to serve as markers. The proteins in 250 microliters of the partially purified cytoplasmic extract were TCA precipitated for electrophoresis in the third lane. The panel on the left was probed with the CA antiserum, and the panel on the right with an antiserum that recognizes

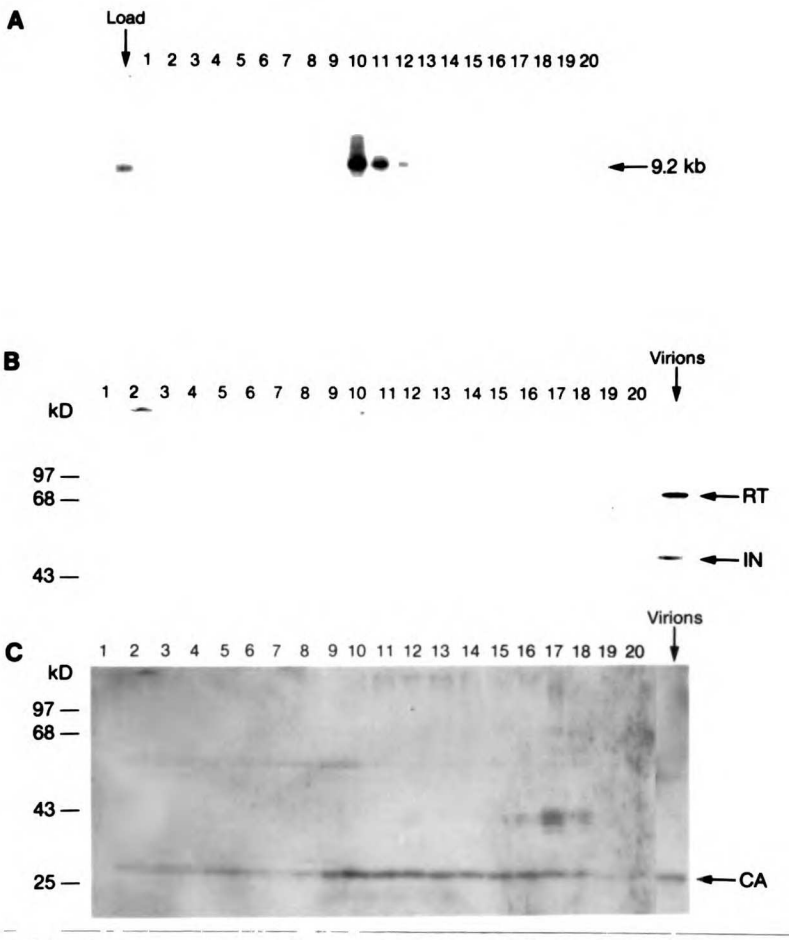


Figure 2. Specific Antisera Confirm the Identify of the Viral Proteins that Copurify with Unintegrated Viral DNA in Nonionic Density Gradients.

Cytoplasmic extracts made from acutely infected NIH 3T3 cells were partially purified by gel-exclusion column chromatography and density equilibrium centrifugation. The density gradient was dripped from the bottom into twenty fractions of 0.6 ml. (A) Viral DNA was detected as described in Figure 1. 100 microliters of extract partially purified by gel filtration, which was used to make the gradient, is shown in the lane marked "load." An arrow indicates the 9.2 kb full-length linear viral DNA. (B) Viral proteins were TCA-precipitated from 100 microliters of each gradient fraction and resolved by electrophoresis in a 12% denaturing polyacrylamide gel. The proteins were transferred to a nitrocellulose filter, and viral proteins were detected with a primary rabbit antiserum specific for **RT** and **IN**, and a second anti-rabbit antibody conjugated to alkaline phosphatase. (C) Antibodies from the filter shown in (B) were removed, and probed with a goat antiserum specific for **CA**. Bound antibody was detected with ^{125}I-Protein G. The position of molecular weight markers in kD are shown to the left. The lanes at the right, labeled "virions," contain purified virus particles. The positions of the viral proteins **RT**, **IN**, and **CA** are indicated by arrows at the right. Fraction numbers are indicated above each lane.

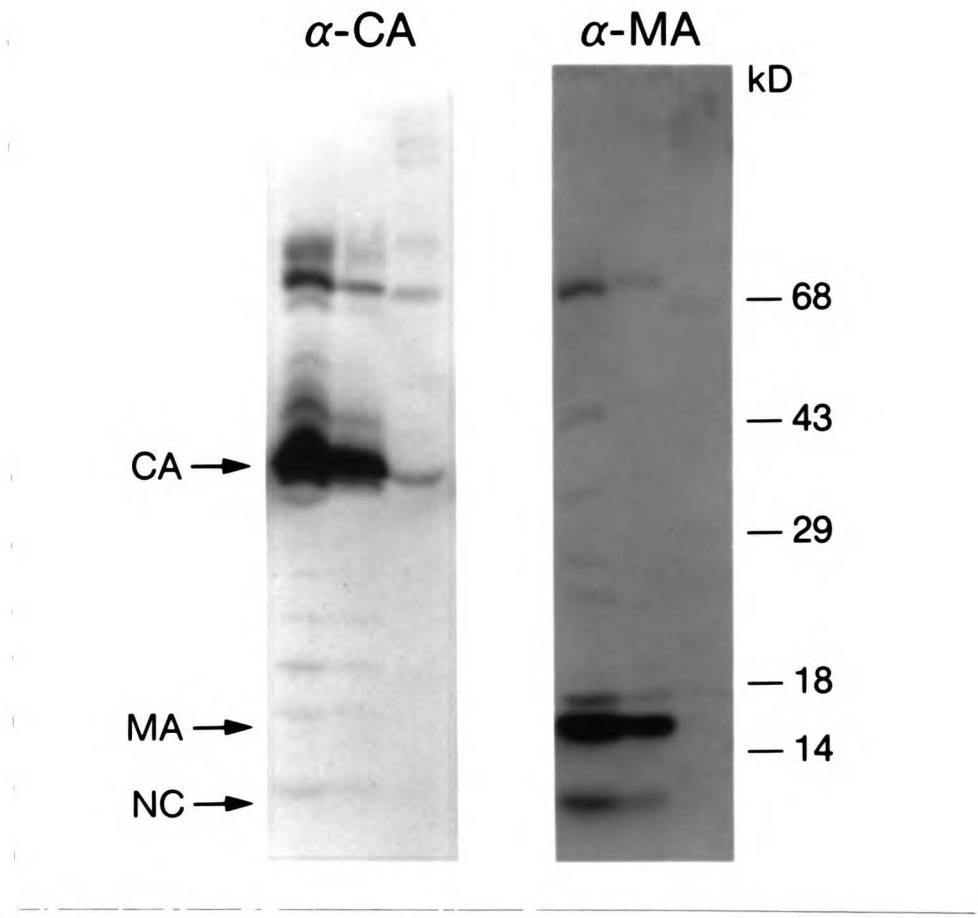


Figure 3. Matrix is Absent from Viral Complexes Partially Purified by Gel Exclusion Column Chromatography.

Cytoplasmic extracts made from acutely infected NIH 3T3 cells were partially purified by gel-exclusion column chromatography. Viral proteins were resolved by electrophoresis in 10-14% denaturing polyacrylamide gradient gel and transferred to nitrocellulose filters. (A) Viral proteins were detected with ^{125}I-Protein G after the filter was probed with a goat antiserum specific for **CA**. The first two lanes contain purified virus particles. The third lane represents 250 microliters of TCA-precipitated proteins from the partially purified extract. (B) Viral proteins were detected with ^{125}I-Protein G after probing the filter with a goat antiserum specific for both **MA** and **NC**. Lanes are the same as in (A). The viral proteins **CA**, **MA**, and **NC** are indicated by arrows at the left, and size markers in kD are at the right.

MA and NC (NCI 78S282). CA, as expected from the results in Figures 1 and 2, is present in the partially purified preparation of complexes, but MA is either absent or present in reduced amounts.

These results demonstrate that in acutely infected cells, a subset of viral proteins persist in the cytoplasm of acutely infected cells and copurify with the DNA-containing nucleoprotein complexes that we have previously described (Bowerman et al 1989). Andersen has shown that while gp70 (SU), encoded by the viral *env* gene, is rapidly degraded inside infected cells, CA and MA persist in infected cell extracts at roughly equivalent levels for at least 18 hours (Andersen 1985). No attempt was made in these earlier studies to demonstrate the association of CA or MA with any type of complex, and RT and IN were not detected. Our results extend these findings by demonstrating that at least three viral proteins-CA, RT, and IN--remain associated with large complexes that copurify with the DNA-containing nucleoprotein complexes. The absence (or reduced amount) of MA is of particular interest since it suggests that at least some of the viral complexes that enter cells undergo changes which involve the loss of some and the retention of other viral proteins. If, as shown by Andersen, MA from incoming virions is stable inside infected cells, it does not appear to be associated with complexes that copurify with unintegrated viral DNA. The data we have obtained with antisera that recognize NC and p12 (both *gag*-encoded) have not been adequate to allow us to determine their presence or absence in these complexes.

The observation that CA is distributed throughout the gradient suggests that it is a relatively unstable component of these complexes.

A similar result would be expected if CA were present as monomers, because small proteins in these gradients do not reach equilibrium after only eight hours of centrifugation. However, the complexes were first purified by gel filtration with a column that resolves the complexes from all proteins not associated with large structures (Bowerman et al 1989). The behavior of CA is puzzling in that one might expect the stability of the shell formed by CA to depend upon the maintenance of interactions between the different CA subunits.

CA in virus particles also appears to be relatively unstable, for only a portion of CA copurified with cores when virions disrupted with nonionic detergents were centrifuged in density gradients (Davis and Reuckert 1972; Bolognesi et al 1973; Stromberg et al 1974). In recent studies of the equine infectious anemia virus, CA was quantitatively recovered with cores. In these latter experiments, sedimentation in glycerol gradients instead of density equilibrium centrifugation in sucrose gradients was used as a means of purifying cores from detergent-disrupted virions (Roberts and Oroszlan 1989).

The stoichiometry of viral protein and DNA suggests that most of the proteins we can detect are associated with structures that do not contain unintegrated viral DNA. We suspect that most of the proteins we can detect in the gradient fractions are associated with complexes that contain viral RNA and have not undergone reverse transcription to generate full-length viral DNA.

These results establish that CA, RT, and IN remain associated with large complexes for several hours after entering cells. These viral

structures copurify with the unintegrated viral DNA complexes in density gradients. At least one viral protein, MA, is not present at detectable levels in the copurifying complexes. The presence in extracts from infected cells of structures that do not contain full-length viral DNA suggests that some of the viral complexes that enter cells might not be fully active for replication. Purification procedures that can resolve these structures will be necessary for direct characterization of the different kinds of viral complexes inside acutely infected cells.

REFERENCES

Andersen, K. B. 1985. The fate of the surface protein gp70 during entry of retrovirus into mouse fibroblasts. *Virology* 142:112-120.

Bowerman, B., P. O. Brown, J. M. Bishop, and H. E. Varmus. 1989. A nucleoprotein complex mediates the integration of retroviral DNA. *Genes and Development* 3:469-478.

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

Brown, P. O., B. Bowerman, H. E. Varmus, and J. M. Bishop. 1989. Retroviral integration: structure of the initial covaqlent product and its precursor, and a role for the viral integration protein. *Proc. Natl. Acad. Sci. USA* 86:2525-2529.

Bolognesi, D. P., R. Luftig, and J. H. Shaper. 1973. Localization of RNA tumor virus polypeptides. 1. Isolation of further virus substructures. *Virology* 56:549-564.

Davis, N. L. and R. R. Rueckert. 1972. Properties of a ribonucleoprotein particle isolated from Nonidet P-40-treated Rous sarcoma virus. *J. Virol.* 10:1010-1020.

Dickson, C., R. Eisenman, H. Fan, E. Hunter, and N. Teich. 1982. Protein biosynthesis and assembly, pp. 513-648. In R. Weiss, N. Teich, H. E. Varmus, and J. Coffin (eds.), **RNA Tumor Viruses**, vol. 1. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

Dickson, C., R. Eisenman, and H. Fan. 1985. Protein biosynthesis and assembly, pp. 135-146. In R. Weiss, N. Teich, H. E. Varmus, and J. Coffin (eds.), **RNA Tumor Viruses**, vol. 2. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

Donehower, L. D. and H. E. Varmus. 1984. A mutant murine leukemia virus with a single missense codon in pol is derfective in a function affecting integration. *Proc. Natl. Acad. Sci. USA* 80:6461-6465.

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

Panganiban, A. T. and H. M. Temin. 1984. The retrovirus *pol* gene encodes a product required for DNA integration: identification of a retrovirus *int* locus. *Proc. Natl. Acad. Sci. USA* 81:7885-7889.

Roberts, M. M. and S. Oroszlan. 1989. The preparation and biochemical characterization of intact capsids of equine infectious anemia virus. *Biochem. and Biophys. Res. Comm.* 160:486-494.

Schwartzberg, P., J. Colicelli, and S. P. Goff. 1984. Construction and analysis of deletion mutations in the pol gene of Moloney murine leukemia virus: a new viral function required for productive infection. *Cell* **37**:1043-1052.

Stromberg, K., N. E. Hurley, N. L. Davis, R. R. Rueckert, and E. Fleissner. 1974. Structural studies of avian myeloblastosis virus: comparison of polypeptides in virion and core component by dodecyl sulfate-polyacrylamid gel electrophoresis. *J. Virol.* **13**:513-528.

Varmus, H. and P. Brown. 1989. **Retroviruses**, pp. 53-108. In D. E. Berg and M. M. Howe (eds.), **Mobile DNA**. Am. Soc. Microbiol., Washington, D. C.

Varmus, H. E. and R. Swanstrom. 1982. Replication of retroviruses, pp. 369-512. In R. Weiss, N. Teich, H. E. Varmus, and J. Coffin (eds.), **RNA Tumor Viruses**, vol. 1. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

Varmus, H. E. and R. Swanstrom. 1985. Replication of retroviruses, pp. 75-134. In R. Weiss, N. Teich, H. E. Varmus, and J. Coffin (eds.), **RNA Tumor Viruses**, vol. 2. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

Chapter 5

Summary and Conclusions

The principal achievement described in this thesis was the development of an in vitro assay that accurately reproduces the retrovirus integration reaction. By combining the use of precursors from infected cell extracts with a sensitive genetic assay, we have been able to show that the integration of retroviral DNA is mediated by a large nucleoprotein complex that appears to be a derivative of the extracellular virion core. The most important outcome of our work has been to prove that, at least for MLV, unintegrated linear viral DNA can serve directly as the precursor to the integrated provirus without going through a circular intermediate.

In this concluding chapter, I will describe recent advances in the study of retroviral integration that have followed from the development of an in vitro assay. For consideration of important issues in retroviral integration that have yet to be resolved, it is useful to discuss the bacterial recombination reactions of phages lambda and Mu. Finally, I will describe ways in which the study of viral structures as they exist inside infected cells can continue to contribute to our understanding of retroviral integration.

Confirmation by the Competition

Work in another laboratory has also shown that linear MLV DNA is the immediate precursor to the integrated provirus. Fujiwara and Mizuuchi, using procedures essentially identical to those we described (see Chapter 2), also showed that the structure of the initial recombinant

product of MLV integration is consistent with models in which the linear viral DNA is the immediate precursor to the integrated provirus (Fujiwara and Mizuuchi 1988). In addition, while we concluded that the 3′ end of viral DNA in the recombinant product must be joined to the target DNA since the 5′ end was not, they showed this directly by the use of denaturing agarose gels. They also showed, as we did, that the 3′ end of most of the unintegrated linear viral DNA is recessed two bases.

As described in Chapter 2, the finding that the 3′ ends of unintegrated viral DNA are recessed two bases can be explained in at least two ways. The two bases could be cleaved from an initially blunt-ended molecule, or reverse transcription could somehow be stopped two bases short of full-length synthesis. Our finding that IN is required for the production of linear DNA with recessed 3′ ends did not favor either mechanism. When we examined the ends of unintegrated viral DNA from an unfractionated population of cytoplasmic molecules, a larger fraction of the 3′ ends were unrecessed compared to the 3′ ends of gel-purified, full-length linear DNA molecules. Presumably, the less-than-full-length viral DNA molecules present in cytoplasmic extracts represent earlier intermediates in replication. If so, then the different ratios of recessed ends in the two populations implies a precursor-product relationship in which blunt-ended linear viral DNA is processed to generate the recessed ends.

Recently, eight more mutant viruses with lesions in IN that block integration were also found to synthesize unintegrated linear viral DNA

with full-length blunt ends (Roth et al 1989). Thus far, no IN mutations have been found that block integration but permit formation of the recessed 3′ ends. These investigators were able to show more convincingly that blunt-ended linear viral DNA is cleaved to give a recessed 3′ end. Cells infected with MLV were harvested at various times after infection, unintegrated viral DNA was prepared (including both full-length viral DNA and less-than-full-length intermediates), and the ends were examined as described in Chapter 2. Early in the infection (2.5 hours), most of the 3′ ends were full-length, with only a small fraction of the molecules recessed. At five hours, roughly equal levels of full-length and recessed molecules were seen, and late in infection (10 to 34 hours) at least 95% of the molecules were recessed, implying that viral DNA is initially made blunt-ended and subsequently cleaved to its recessed form. These experiments were done with a replication-defective virus (pseudotyped with wild type viral proteins) that was competent for only a single round of replication. This strategy avoided the introduction, by superinfection, of any early replication intermediates at late time points in the infection.

In addition to examining the effect of IN mutations, these investigators examined several *att* site mutants to determine if sequence changes at the ends of linear DNA affect the formation of recessed ends. As described in Chapter 1, small insertions and deletions at the very ends of viral DNA can be tolerated by the integration machinery such that proviruses are joined to host DNA at the same conserved CA dinucleotide found at the 3′ end of wild type proviruses. When the

unintegrated DNAs made by viruses with *att* site mutations were examined, in every case but one, the ends of the *att* site mutants were found to be cleaved after the same CA dinucleotide, in spite of alterations in the terminal sequences. In some cases, the cleavage site was only one base removed from the 3′ end of the LTR (instead of two bases, as in wild type). In one case, the cleavage site was four bases removed from the 3′ end. Since the mutations were all in U5, these results also showed that the cleavage events can remove different numbers of nucleotides from each end; two bases were always removed from the U3 end. In addition, some of the mutants had transversions at the U5 end that did not interfere with cleavage even when they were directly adjacent to the CA dinucleotide. A double mutant in which the last four bases at the 3′ end were changed from CATT to TATA was fully functional both for cleavage and replication. Surprisingly, the cleavage was made two bases from the end in spite of the mutation in the highly conserved CA dinucleotide.

A mutant with ten extra bases pairs inserted at the end of the LTR, which is integration-defective, synthesized only blunt-ended viral DNA. The one mutant in which cleavage was observed to occur at a new position was a psuedorevertant of this 10 base pair insertion mutant. Integrants of this pseudorevertant end just 3′of a CA dinucleotide present in the inserted sequences; the cleavage site at the 3′ end of unintegrated DNA was just 3′ of the same CA dinucleotide. As described in Chapter 1, these inserted sequences bear some resemblance to the normal att site. In all cases, for both mutant and wild type viruses, the sites of

cleavage have corresponded exactly with the sites of joining of the viral DNA to the target DNA. These results suggest that the cleavage event that generates the recessed end determines the site at which viral DNA becomes joined to the host DNA.

ASLV Endonuclease Activity Revisited

The realization that linear DNA can integrate directly into the genome has helped resolve some of the inconsistencies in the in vitro assays that detect ASLV endonuclease activity. As described in Chapter 1, both the α,β dimer and the p32 IN protein have been shown to cleave DNA molecules that include the ASLV circle junction sequence. p32 IN has been shown to cleave the circle junction at positions that generate a four base pair staggered cut that is consistent with the loss of two base pairs from the end of each LTR, but many other sites are also nicked in such experiments. The α,β dimer has also been shown to cleave a circle junction sequence, but the sites of cleavage were three instead of two bases removed from the edges of the LTRs. Also, single-stranded DNA was cleaved more efficiently than double-stranded DNA. In recent experiments, when the U5 *att* site was placed at the end of linear, double-stranded DNA molecule, the α,β dimer was shown to precisely remove two bases from the 3′ end of the linear *att* site; such an activity is more consistent with our current understanding of retroviral integration (J. Leis, personal communication). It will be interesting to see if the cleavages at both ends of a linear molecule are coupled or can occur independently of each other.

An In Vitro Activity for the MLV IN Protein

While the MLV IN protein has not yet been shown to cleave blunt-ended molecules to generate a two base recess at the 3′ end, recent work has shown that the IN protein binds the ends of linear viral DNA in a gel-mobility shift assay (S. Basu, personal communication). The binding activity is specific for MLV sequences, and *att* sites at the end of a molecule are recognized better than an internally-positioned circle junction. In all, three different in vitro assays now exist for analyzing retroviral integration: (i) Binding of IN to *att* site sequences; (ii) cleavage of *att* site sequences to generate a two base recess at the 3′ ends of linear DNA; and (iii) integration of a donor molecule into a DNA target.

Limitations of Infected-Cell Extracts; Reconstituting an Integration Machine

While the use of infected cell extracts was important for developing an in vitro assay for retroviral integration, the results presented in Chapter 4 demonstrate some of the difficulties inherent in using extracts as a source of complexes. One can imagine subjecting integration-competent complexes purified from extracts to various disruptive manipulations to define essential components of the integration machinery. But the presence of additional viral structures

in these extracts that copurify with the DNA-containing complexes makes it difficult to specifically characterize complexes that are active for integration. Both determining the protein content of these complexes and obtaining images of them in the electron microscope will require a purification protocol that can resolve the different viral structures that apparently exist inside infected cells.

Definition of the protein components required for retroviral integration is likely to come from reactions reconstituted with purified components. Indeed, integration activity has already been reconstructed in vitro using exogenously added plasmid DNA as a donor molecule (Fujiwara and Craigie 1989). Authentic integration activity was reconstituted when the viral *att* site sequences were present at the ends of a linearized plasmid that encodes resistance to ampicillin and tetracycline. The 3′ ends of the linearized plasmid were already recessed two bases, reproducing the structure seen at the ends of most unintegrated linear viral DNA during an infection. Amp^R and Tet^R provide genetic markers for the detection of recombinants. Extracts made both from acutely infected cells and from detergent-disrupted virions could provide trans-acting factors that reconstituted integration activity using the exogenously added donor molecule.

When infected-cell extract was first passed through a Sephacryl-400 gel filtration column, the trans-acting components were resolved into two poorly defined fractions. The DNA-containing complexes in the extract appeared in the void volume. Some reconstituted activity, scored by the genetic assay, was obtained using these void volume

fractions. However, about twice as much activity was obtained using slightly more included fractions from the column, which contained very little unintegrated viral DNA. When the void volume and the included fractions were mixed, the level of reconsitituted activity was ten-fold higher than with either fraction alone. No explanation for this fractionation of the activity was offered. One question concerning these results is whether integration activity was reconstituted by exogenously added DNA entering the viral complexes in the extract, or by proteins in the viral complexes diffusing and reassembling on the exogenously added DNA.

Low levels of integration activity were obtained when detergent-disrupted virions were the only source of trans-acting factors, but activity was stimulated fifty-fold by the addition of uninfected cell extracts with the disrupted virions. The stimulatory factor was inactivated by heating the uninfected cell extracts at 60°C for 15 minutes. The ability to stimulate activity with the uninfected cell extract suggests that host components may enhance integration activity, although one can imagine indirect explanations for the stimulation. For example, activities in the cell extract might disrupt the viral cores in way that liberates the trans-acting factors more effectively. When the disrupted virions, uninfected cell extract, and the donor and target DNA were preincubated together on ice for one hour, activity was stimulated three-fold compared to reactions which were not preincubated. Any effects of preincubation when only disrupted virions were added or when infected cell extracts were used were not discussed.

Integration activity was reconstituted only when the 3' ends of the linearized plasmid were already recessed two bases. A blunt-ended molecule was not active in the assay, and neither was a circular plasmid that contained an intact MLV circle junction sequence. Two possible explanations were offered for the lack of activity using the blunt-ended linear molecule. First, the two bases at the 3' ends of the linear donor DNA were GG instead of TT. While work from Goff's group has shown that changes in the sequence of these two bases can be tolerated, this particular mutation has not been tested. Furthermore, the reaction mixtures in these assays included a large excess of donor molecules. If cleaving the ends is not linked to the strand-exchange reaction, then both ends might never get processed on the same molecule. Alternatively, even if both ends were properly processed, the high concentration of uncleaved molecules might compete out any further interactions required to integrate the molecules with recessed ends.

The ability to reconstitute integration activity using exogenously added DNA is encouraging for efforts to reconstitute activity with completely defined, purified components. Recently, integration activity has been reconstituted using extracts from insect cells that express the MLV IN protein from a baculovirus vector, suggesting that IN may be the only viral protein absolutely required for integration activity (Craigie, personal communication).

Precursors Assembled In Vivo Integrate Viral DNA Efficiently

From the perspective of defining the essential components of the integration machinery, reconstitution experiments are very appealing. However, efforts to reconsititute activity thus far suffer from the shortcoming that the reactions are very inefficient; only a very small fraction of the donor DNA molecules become activated for integration. Although several hundred recombinants per reaction were detected in the assay described by Fujiwara and Craigie, the donor plasmid was present in microgram quantities, and only about one in 10^7 donor molecules actually integrated. In contrast, as many as 50% of the viral nucleoprotein complexes present in extracts made from acutely infected cells have been shown to be active for integration in vitro (P. Pryciak and P. Brown, personal communication). Even though IN might be the only protein absolutely required for integration, it seems likely that additional proteins will be required to form structures that integrate viral DNA with maximal efficiency.

Bacterial Recombination: The Role of Higher Order Structures

A focus of recent work on a variety of prokaryotic recombination events has been to define "higher order," three-dimensional arrangements of DNA and protein that are required for efficient strand-exchange (Craig 1988). Several experiments suggest that even though a single protein can be sufficient to break and join DNA strands, efficient

recombination activity requires additional proteins that assemble into relatively complex structures. Lambda integration and Mu transposition are two of the most thoroughly studied examples of bacterial recombination. For both of them, highly ordered arrangements of protein and DNA have been shown to be important for efficient strand-exchange activities.

Bacteriophage Lambda Integration: The Intasome

Lambda integrates with high specificity in the bacterial genome at *attB*, a 25 base pair region which shares a 15 base pair core sequence with *attP*, a 240 base pair region in the bacteriophage genome (see Craig 1988; Thompson and Landy 1989). Integration occurs within the core sequence to generate a prophage flanked by two new sites called *attL* and *attR* that are distinct hybrids of the sequences in *attP* and *attB* (all share the same core sequences). Lambda can reverse the direction of strand-exchange in a conservative excision, a reaction not known to occur with either Mu or retroviruses. Integration requires the λ Int gene product, and excision requires both Int and an additional phage gene product, Xis. Both reactions require the host factor IHF, and excision is enhanced by an additional host protein, FIS. IHF is similar to HU and other bacterial histone-like proteins, but is unusual in that it binds to specific sequences. FIS was first identified as a stimulator of Gin inversion in Mu and Hin inversion in *Salmonella*. The intracellular concentration of FIS drops considerably as bacterial

cultures approach stationary phase; FIS is thought to favor the lytic pathway by enhancing excision during exponential growth. There are multiple binding sites for all of these proteins in the *att* sequences. Although λ Int has an intrinsic type I topoisomerase activity (Kikuchi and Nash 1979; Craig and Nash 1983), the requirement for additional viral and host proteins to reconstitute highly efficient integration and excision reveals the importance of highly ordered nucleoprotein complexes that include multiple components with the potential for a large number of different interactions.

Electron microscopy, topological studies, and DNAase protection experiments have been used to show that the integration machinery of bacteriophage lambda assembles with *attP* to form a highly ordered complex of protein and DNA. The precise arrangement of the DNA and protein appears to be dictated both by the arrangement of protein binding sites in *attP* and by interactions between bound components. The interactions among the different components appear to be important in two ways. First, they have cooperative and competitive effects on the use of the different binding sites in *attP*. Second, they might be involved in bringing together in three dimensions sites on the DNA that are widely separated in the linear DNA sequence.

One simple way to generate higher order structures is the presence in a single protein of two DNA binding domains that recognize distinct sequences. The existence of core-type and arm-type sites for λ Int suggests that widely separated sites in *attP* can be brought together through binding to different domains of the same protein (Moitoso de Vargas et al 1988, 1989).

Electron microscopic studies showed that about 240 base pairs encompassing attP are occupied by Int and IHF in a condensed structure called the "intasome" (Better et al 1982; Echols 1986). Such structures do not readily form on *attB*, but when Xis is added, complexes can be formed that bring together *attL* and *attR* (Better et al 1983). Topological studies have taken advantage of the ability of λ integration to cause an inversion when *attP* and *attB* are placed in appropriate orientation on the same plasmid. Even when such a plasmid is relaxed by introduction of several nicks, the products of the inversion reaction include trefoils, arguing that the intasome wraps the DNA around itself in such a way that integration creates a knot in the plasmid (Pollock and Nash 1983). The structure and sign of the trefoil suggest that a left-handed coil is formed, as in a nucleosome (Griffith and Nash 1985). Other investigators have interpreted similar results to mean that the structure must be plectonemic and not, like nucleosomes, solenoidal (Spengler et al 1985).

DNAse protection studies have shown that binding of Int and IHF at certain sites is cooperative, and that Xis influences the nature of this cooperativity in a way that determines whether excision or integration is favored. Int has been shown to bind certain sites in *attP* with higher affinity by cooperatively interacting with IHF, although this cooperativity may not involve a direct protein/protein interactions (Richet et al 1986). Another detailed analysis of cooperative binding has shown that different sites in *attP* are occupied by Int and IHF with different affinities depending on whether Xis is also present (Bushman

et al 1984; Thompson et al 1987). Both sets of results could be explained by models involving either (i) changes in the structure of DNA induced by proteins binding at other sites, or (ii) simple occlusion or enhancement of binding by proteins bound to adjacent sequences. Thompson and Landy have argued that the simplest explanation for the results is that different combinations of proteins favor mutually exclusive choices in the three dimensional arrangement of the DNA and proteins. They suggest that competition between mutually exclusive higher order structures might account for the ability of lambda to choose either integration or excision depending on the physiological state of the host cell (Thompson and Landy 1989).

The need for regulation of a reversible integration reaction has been used to explain why lambda uses such a complex intasome structure to integrate and excise its genome. Other members of the λ Int family perform unidirectional reactions that need not be tightly coupled to the physiology of the host. Two such Int-like proteins, FLP in yeast and CRE in bacteria, appear not to require accessory factors, consistent with the idea that a complex intasome is needed only for the more highly regulated recombination reactions involved in lambda integration and excision (Thompson and Landy 1989). Others have argued that the intasome is structurally complex to insure that when integration is needed, it occurs with high efficiency (Echols 1986; Richet et al 1986).

The regulation and reversibility of bacteriophage λ integration make it seem very different from retroviral integration, and these two activities have other substantial differences (see Chapter 1; Craig

1988; Thompson and Landy 1989). In contrast to the requirement for core homology in *attP* and *attB*, retroviruses integrate into an extremely large number of apparently unrelated sites. The use of a covalent linkage between protein and DNA as an intermediate in strand-exchange has been demonstrated for lambda, but not for retroviral integration. The λ *att* site sequences are located internally in the linear bacteriophage genome, whereas retroviral *att* sequences are located at the ends of the linear viral DNA. Nevertheless, λ integration does provide compelling support for the ability of higher order structures to influence strand-exchange reactions.

Additional examples of recombination that appear to involve long-range interactions between proteins bound to sites separated by several turns of the DNA helix are the resolution reactions of the closely related transposons Tn3 and γδ. Tn3 brings together widely separated DNA sites by interactions between resolvase proteins bound to the two sites that recombine during resolution of cointegrates (Benjamin and Cozarelli 1988). γδ appears to use protein-induced DNA bending, in addition to protein/protein interactions, to generate complex structures in three dimensions (Salvo and Grindley 1988). AttP in λ contains intrinsically bent DNA sequences, and IHF can induce DNA bending at three different binding sites in *att* (Thompson and Landy 1989). Recently, IHF-induced bending has been shown to facilitate the ability of Int to bind both core-type and arm-type sites in *attL*. The authors proposed that protein-induced DNA bending and binding of two different sequences by a single protein operate together to form a higher order structure with *attL* (Moitoso de Vargas et al 1989).

Mu Transposition

Mu is unusual in that it is both a bacteriophage that can lytically infect cells to produce extracellular virus particles, and a transposon, in that during a lytic infection it transposes to many new sites in the host genome before being packaged into new phage particles. Three types of DNA recombination are important in the Mu life cycle (see Pato 1989). (i) During an infection, the incoming Mu genome is covalently joined to the bacterial chromosome in a conservative integration reaction. How Mu chooses between residing in the host genome as a lysogen or proceeding with a lytic infection is not known, but the choice is thought to be made after the initial integration reaction. (ii) Transposition events that occur during a lytic infection are replicative--prophages are inserted into new sites without loss of the donor molecule. The structure of the initial product of integration in both the replicative and nonreplicative pathway is called the Shapiro intermediate. This intermediate is identical in structure to the intermediate formed by retroviral integration, except that the 5′ ends of the Mu DNA, instead of being free (as in retroviral integration), remain joined to flanking sequences from the bacterial chromosome. Cleavage of the 5′ ends of the Mu DNA in the Shapiro intermediate produces a conservative transposition reaction, while replication primed by the free 3′ ends of the target DNA results in a replicative transposition event. (iii) Mu particles possess one of two types of

tail fibers depending on the orientation of an invertible 3 kilobase piece of DNA call the G segment. The Gin recombination system mediates the inversion of this piece of DNA.

The transposition reaction of bacteriophage Mu is more reminiscent of retroviral integration than is λ integration. Like retroviral integration, the initial integration of incoming Mu DNA into the bacterial chromosome is not known to be a reversible reaction. Integration can occur at many places in the bacterial genome, a phenomenon responsible for its name (Mutator). Some experiments have suggested that transcriptional activity can influence the suitably of a sequence as a target. In contrast to retroviral integration (see Chapter 1), Mu integration is thought to be inhibited rather than enhanced by transcriptional activity (Pato 1989). Both Mu and retroviral integration have the same polarity (the 3′ end of viral DNA is joined to the 5′ end of the target DNA), and both reactions generate short duplications of the host target sequence. Mu enters the cell as a linear molecule, but the ends of the DNA are of variable sequence because of a "headful" mechanism of packaging. Nevertheless, Mu DNA always becomes joined to host DNA at the same position in the Mu genome (which defines the ends of the Mu sequences). Thus, even though phage Mu does not lose any of its own DNA sequences during integration, both retroviral and Mu integration require cleavages at sites near the ends of a linear DNA molecule, and the cleavage sites are made at the sites where the incoming DNA becomes joined to the host chromosome. The sequences required in *cis* for Mu integration, like those in MLV, are

present at the ends of the Mu DNA. As with retroviruses, no covalently linked DNA/protein intermediates have been identified, and strand-exchange has been proposed to occur by a direct nucleophilic attack of the 3' OH at end of each strand of Mu DNA on the phosphodiester bond in the target DNA (Craigie and Mizuuchi 1987). A similar mechanism has been proposed for retroviral integration (Varmus and Brown 1989). Although Mu integration in vitro is substantially stimulated by ATP, an exogenous energy source is not absolutely required; ATP is probably involved in target-site selection rather than in strand-exchange (Maxwell 1987).

Higher Order Structures in Mu Transposition: The Transpososome

As with λ integration, higher order structures appear to be important in Mu transposition. Mu A, like λ Int, can bind two different sequences; sites at the ends of Mu DNA and an unrelated site that overlaps the Mu operator (Craigie et al 1984; Leung et al 1989; Mizuuchi and Mizuuchi 1989). Efficient integration activity in vitro and in vivo depends not only on the presence of the sites in the terminal sequences of Mu, but also on the site that overlaps the Mu operator. In addition, the Mu repressor can also bind the A sites at the ends of Mu DNA, and Mu repressor inhibits integration in vitro (Craigie et al 1984). The ability of Mu A to bind more than one kind of sequence, and competition for the two kinds of sites with the Mu operator have been proposed to be important for Mu integration per se and for determining

whether a lytic or lysogenic pathway is chosen.

Under some conditions in vitro, the Mu A protein and *E. coli* HU, can carry out strand-exchange activity inefficiently. The reaction is greatly stimulated by the Mu B protein and ATP (Maxwell et al 1987). Moreover, reconstitution of Mu transposition in vitro requires that the donor molecule be supercoiled (Craigie et al 1985). When Mu A and HU are mixed with a supercoiled donor molecule, an intermediate called the Type I complex assembles, and nicks are introduced at each 3′ end of the Mu DNA (Craigie et al 1987; Surette et al 1987). After nicking, the flanking plasmid sequences in the donor molecule become relaxed, but the Mu sequences internal to the two Mu ends remain supercoiled, suggesting that the Type I complex restricts movement of the free 3′ ends of the Mu DNA. Once the nicks have been made, removing the supercoiling by cleavage with a retriction enzyme does not reduce the efficiency of the subsequent strand exchange (Surette et al 1987). The supercoiling requirement appears to be important in forming the nicked precursor to integration, but not for integration itself. The degree of supercoiling that is required in vitro can be reduced by the addition of IHF; the supercoiling requirement in the presence of IHF is more consistent with the in vivo supercoiling requirement, suggesting that an additional host protein might be involved in Mu transposition (Surette and Chaconas 1989). If cuts are first introduced into the donor molecule at the appropriate positions with a restriction endonuclease, Mu A and B with ATP can integrate a linear Mu genome into target DNA in the absence of HU and supercoiling (Craigie and Mizuuchi 1987).

When the Type I complex is supplied with target DNA, Mu B, and ATP, a poorly defined type II complex forms, which mediates strand-exchange to yield the Shapiro intermediate (Craigie and Mizuuchi 1985; Craigie et al 1987; Surette et al 1987). Mu B is a nonspecific DNA binding protein, and promotes intermolecular events in vitro presumably by binding to both target DNA and the type I complex. In the absence of Mu B and ATP, strand transfer is less efficient in vitro and occurs primarily by intramolecular events into other sites in the donor plasmid (Maxwell et al 1987). The role of these "transposome" complexes in fully reconstituting Mu integration demonstrates once again the importance of higher order structures in mediating strand-exchange reactions.

Unique Features of Retroviral Integration

While it is useful to consider these other examples of integration reactions, retroviral integration differs from both λ integration-excision and Mu transposition in at least two ways. First, both Mu and λ express the viral components of the recombinational machinery from unintegrated genomes after entry into a bacterium. Retroviruses are thought to bring all the viral proteins required for integration into the cell with their incoming genome, and unintegrated DNA is expressed inefficiently if at all. Second, while Mu and λ inject a double-stranded DNA molecule directly into a bacterium, retroviruses must assemble their integration machinery on a genome that undergoes transformation from a single-stranded RNA molecule to a double-stranded

DNA molecule after entry into a cell. λ and Mu, then, both assemble integration complexes on pre-existing DNA molecules with newly synthesized proteins, while retroviruses appear to assemble complexes on newly synthesized DNA molecules with pre-existing viral proteins. Moreover, the functionality of the retroviral integration machinery may depend on events required for assembly of the *gag* and *gag-pol* polyproteins during viral assembly and budding. These features make retroviral integration distinct from most bacterial examples of transposition.

A Dual Approach to the Study of Retroviral Integration

Further progress in understanding the mechanism of retroviral integration will likely ensue both from the reconstitution of integration activity in vitro using purified components and from further characterization of the integration-competent complexes found inside infected cells. Our development of an in vitro assay is testimony to the value of such a dual approach. Moreover, most of what we know about many transposition reactions has come from the study of in vitro systems. As Bruce Alberts has argued, even though one might be able to reproduce a biological activity in vitro using only a few components, the complex environment of the cell requires understanding how that activity fits into the context of an intact cell (Alberts 1984). By comparing results from assays reconstructed in vitro with results gained by analysis of activities and structures isolated from inside cells, one can gain a more complete understanding of biological functions.

Even though direct purification of the complexes found inside infected cells may prove difficult, analysis of such structures may continue to be informative. Indirect end-labeling could be used to define sequences in the viral DNA that are protected by components of the nucleoprotein complex from nuclease digestion. Results from such an analysis could be compared to similar protection studies carried out in vitro with purified components. Additional structural information might be obtained by using monoclonal antibodies to determine what epitopes of the viral CA protein are exposed in the complexes. For example, it might be possible to determine if the amino acids in CA that are involved in *Fv-1* restriction are exposed in the nucleoprotein complex that is active for integration. Antibodies might also be useful for analyzing the structure of the nucleoprotein complex after integration--perhaps additional viral proteins or a distinct set of CA epitopes would be exposed by conformational changes that might accompany the strand-exchange reaction.

The idea that in vitro activities that can be more fully understood by examining structures inside intact cells can also be applied to studies of the target of retroviral integration. Recent experiments have shown that integration-competent complexes isolated from infected cells can integrate into both SV40 minichromosomes and minichromosomes isolated from yeast (A. Sil and P. Pryciak, personal communciation). It will be interesting to see whether or not chromatin structure can influence either the efficiency of integration or the selection of target sites.

A better understanding of retroviral integration is desirable for several reasons. It is an important event in the retrovirus life cycle that is shared by other retrotransposons, and it provides an unusual example of transposition in higher eukaryotes. In addition to the insight we can gain into the fundamental process of DNA recombination, the study of retroviral integration could have important clinical consequences. Inhibitors of the integration reaction would be good candidates for anti-viral therapy because the reaction does not resemble any known cellular processes. By taking a dual approach to the study of retroviral integration that takes advantage of both reconstituted reactions and characterization of complexes isolated from infected cells, it should be possible to significantly improve our understanding of the mechanism of retroviral integration.

REFERENCES

Alberts, B. 1984. The DNA enzymology of protein machines. *Cold Spring Harbor Symp. Quant. Biol.* **49**:1-12.

Benjamin, H. W. and N. R. Cozzarelli. 1988. Isolation and characterization of the Tn3 resolvase synaptic intermediate. *EMBO* **7**:1897-1905.

Better, M. C. Lu, R. C. Williams, and H. Echols. 1982. Site-specific DNA condensation and pairing mediated by the int protein of bacteriophage λ. *Proc. Natl. Acad. Sci. USA* **79**:5837-5841.

Better, M., S. Wickner, J. Auerbach, and H. Echols. 1983. Role of the Xis protein of bacteriophage λ in a specific reactive complex at the attR Prophage attachment site. *Cell* **32**:161-168.

Bushman, W., S. Yin., L. L. Thio, and A. Landy. 1984. Determinants of directionality in lambda site-specific recombination. *Cell* **39**:699-706.

Craig, N. 1988. The mechanism of conservative site-specific recombination. *Ann. Rev. Genetics* **22**:77-106.

Craig, N. L. and H. A. Nash. 1983. The mechanism of phage λ site-specific recombination: site-specific breakage of DNA by int topoisomerase. *Cell* **35**:795-803.

Craigie, R., D. J. Arndt-Jovin, and K. Mizuuchi. 1985. A defined system for the DNA strand-transfer reaction at the initiation of bacteriophage Mu transposition: protein and DNA substrate requirements. *Proc. Natl. Acad. Sci. USA* **82**:7570-7574.

Craigie, R., M. Mizuuchi, and K. Mizuuchi. 1984. Site-specific recognition of the bacteriophage Mu ends by the Mu A protein. *Cell* **39**:387-394.

Craigie, R. and K. Mizuuchi. 1985. Mechanism of transposition of bacteriophage Mu: structure of a transposition. *Cell* **41**:867-876.

Craigie, R. and K. Mizuuchi. 1987. Transposition of Mu DNA: joining of Mu to target DNA can be uncoupled from cleavage at the ends of Mu. *Cell* **51**:493-501.

Echols, H. 1986. Multiple DNA-protein interactions governing high-precision DNA transactions. *Science* **233**:1050-1056.

Fujiwara, T. and R. Craigie. 1989. Integration of mini-retroviral DNA: a cell-free reaction for biochemical analysis of retroviral integration. *Proc. Natl. Acad. Sci. USA* **86**:3065-3069.

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

Griffith, J. D. and H. A. Nash. 1985. Genetic rearrangement of DNA induces knots with a unique topology: implications for the mechanism of synapsis and crossing-over. *Proc. Natl. Acad. Sci. USA* **82**:3142-3128.

Kikuchi, Y. and H. A. Nash. 1979. Nicking-closing activity associated with bacteriophage λ *int* gene product. *Proc. Natl. Acad. Sci. USA* **76**:3760-3764.

Leung P. C., D. B. Teplow, and R. M. Harshey. 1989. Interaction of distinct domains in Mu transposase with Mu DNA ends and an internal transpositional enhancer. *Nature* **338**:656-658.

Maxwell, A., R. Craigie, and K. Mizuuchi. 1987. B protein of bacteriophage Mu is an ATPase that preferentially stimulates intermolecular DNA strand transfer. *Proc. Natl. Acad. Sci. USA* **84**:699-703.

Mizuuchi, M. and K. Mizuuchi. 1989. Efficient Mu transposition requires interaction of transposase with a DNA sequence at the Mu operator: implications for regulation. *Cell* **58**:399-408.

Moitoso de Vargas, L., C. A. Pargellis, N. M. Hasan, E. W. Bushman, and A. Landy. 1988. Autonomous DNA binding domains of k integrase recognize two different sequence families. *Cell* **54**:923-929.

Moitoso de Vargas, L., S. Kim, and A. Landy. 1989. DNA looping generated by DNA bending protein IHF and the two domains of lambda integrase. *Science* **244**:1457-1461.

Pato, M. L. 1989. Bacteriophage Mu, pp. 23-52. In D. E. Berg and M. M. Howe (eds.), **Mobile DNA**. Am. Soc. Microbiol., Washington, D.C.

Pollock, T. J. and H. A. Nash. 1983. Knotting of DNA caused by a genetic rearrangement: evidence for a nucleosome-like structure in site-specific recombination of bacteriophage lambda. *J. Mol. Biol.* **170**:1-18.

Richet, E., P. Abcarian, and H. A. Nash. 1986. The interaction of recombination proteins with supercoiled DNA: defining the role of supercoiling in lambda integrative recombination. *Cell* **46**:1011-1021.

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

Salvo, J. J. and N. D. F. Grindley. 1988. The γδ resolvase bends the *res* site into a recombinogenic complex. *EMBO* **7**:3609-3616.

Spengler, S. J., A. Stasiak, and N. R. Cozzarelli. 1985. The stereostructure of knots and catenanes produced by phage λ integrative recombination: implications for mechanism and DNA structure. *Cell* 42:325-334.

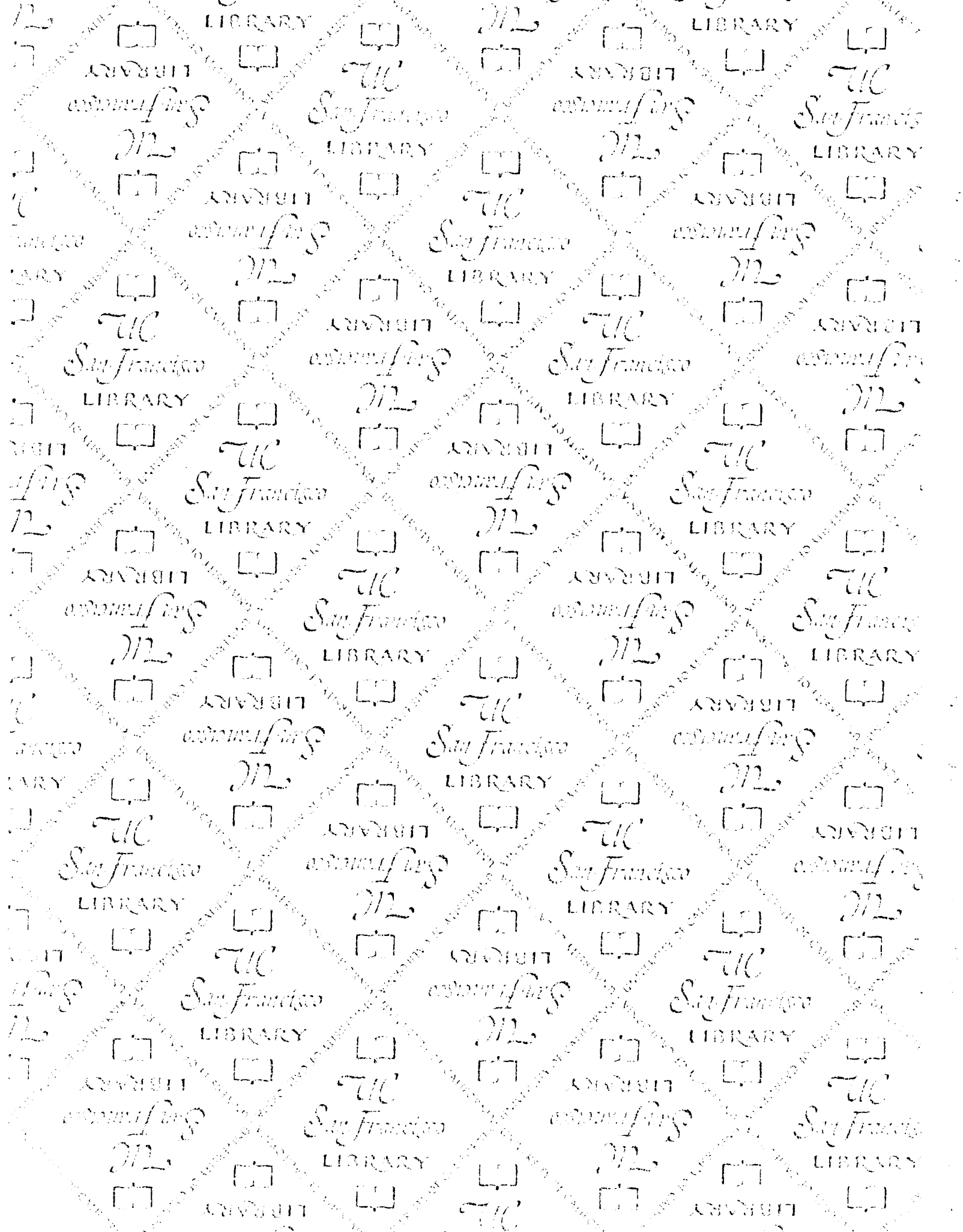
Surette, M. G., S. J. Buch, and G. Chaconas. 1987. Transposomes: stable protein-DNA complexes involved in the in vitro transposition of bacteriophage Mu. *Cell* 49:253-262.

Surette, M. G. and G. Chaconas. 1989. A protein factor which reduces the negative supercoiling requirement in the Mu DNA strand transfer reaction is *Escherichia coli* integration host factor. *J. Biol. Chem.* 164:3028-3034.

Thompson, J. F. and A. Landy. 1989. Regulation of bacteriophage lambda site-specific recombination, pp. 1-22. In D. E. Berg and M. M. Howe (eds.) **Mobile DNA**. Am. Soc. Microbiol., Washington, D.C.

Thompson J. F., L. Moitoso de Vargas, S. E. Skinner, and A. Landy. 1987. Protein-protein interactions in a higher-order structure direct lambda site-specific recombination. *J. Mol. Biol.* 195:481-493.

Varmus, H. and P. Brown. 1989. Retroviruses, pp. 53-108. In D. E. Berg and M. M. Howe (eds.), **Mobile DNA**. Am. Soc. Microbiol., Washington, D.C.


San Francisco
LIBRARY