Title
Cellular influences on retroviral integration

Author
Pryciak, Peter M.

Publication Date
1992

Peer reviewed|Thesis/dissertation

CELLULAR INFLUENCES ON RETROVIRAL INTEGRATION

by

Peter M. Pryciak

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Biochemistry

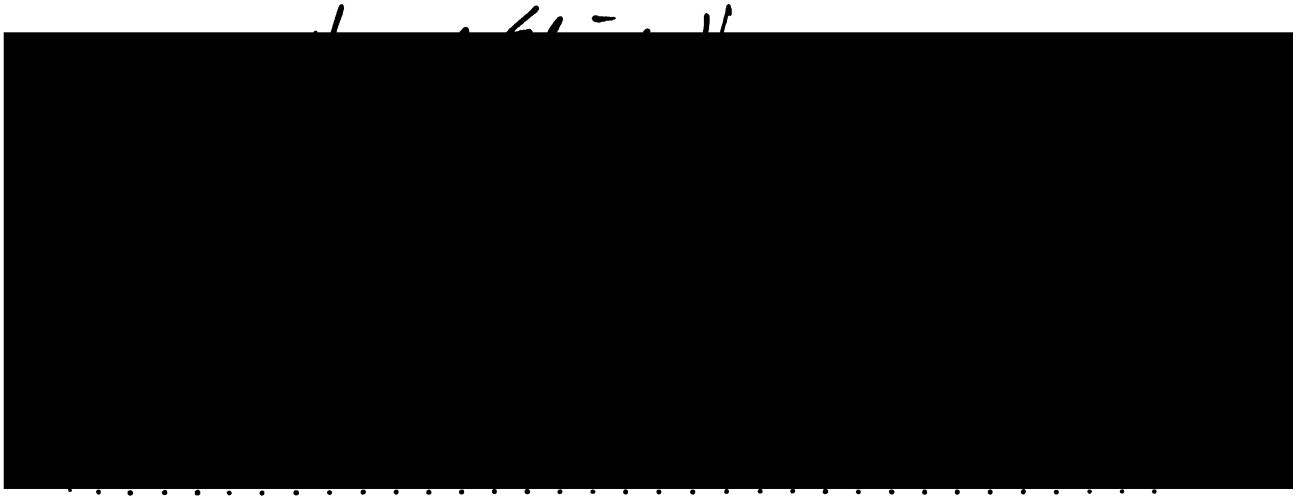
in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA

San Francisco



Date **University Librarian**

Degree Conferred: 3/23/92

PREFACE

I fluctuate between thinking that I've been at UCSF forever and that I just got here yesterday. The part of me that feels like I got here yesterday knows that I still have very much to learn and that many things went undone. The part that feels like I've been here forever remembers how enthusiastic and invincible I felt upon arrival, and the long stretches of nothing working that continuously forced me to be humble.

In general I am thankful for the learning environment provided by the Biochemistry department; there is so much interesting work being done here by so many fascinating and dynamic people, and the seminars and other forums so high-quality, that a student here receives a great education simply by immersion.

I would especially like to thank the following people for their contributions to my life here in San Francisco: Titia de Lange for scientific inspiration and remaining a challenge to figure out; Tyler Jacks for teaching me to drink champagne off of my forehead and being seemingly good at everything ; Bruce Bowerman for teaching me how to do integration assays and Southern blots, and always being entertaining; David Kaplan for toys and late night philosophy; John Young for his boundless enthusiasm and love of good beer, wine, and margaritas; Lily Shiue for her omnipotent sense of humor and never taking anything seriously; Paul Bates for his willingness to give his opinion on everything; Ken Kaplan for being generally fun; Andy Leavitt for his desire to laugh; Hans-Peter Muller for taking over where I left off; Anita Sil for being such a fabulous student and so easy to make fun of; Michael Glotzer for occasionally forcing me to take time off for recreation, especially great skiing trips; Warren Kruger for more good skiing; Delia Lakich for introducing me to Tommy's; Tommy's for making the perfect margarita; Sue Adams and Rachel Mozesson for being so on top of everything; and those

I've neglected to mention for understanding that everything I've done in my final weeks here was done in a hurry.

I thank my mother, Maria Pryciak, for always insisting that I should pursue what I found interesting, and not holding any particular expectations of me. Her incredible endurance has been an important example to me, and has often helped me to realize that no matter what happens, I've still got it pretty good.

I am particularly indebted to Caroline Goutte for making my life a better one to live by sharing it with me. Her everpresent enthusiasm, optimism, and friendliness make her a joy for everyone to have around, and her support of me during harder times was priceless--my thanks are immeasurable.

I also thank Bruce Alberts for serving on both my orals and thesis committees, and for reading my thesis so carefully on such short notice.

I am especially grateful to Pat Brown for numerous contributions to my thinking and work on integration. His enthusiasm was infectious, his wealth of ideas amazing, and his eagerness to listen and contribute invaluable. He is a model for how to place primary importance in the fun of science.

Finally, I would like to thank Harold for having me in his lab and for his contribution to my scientific development. He coached me through my first few years by constantly raising the standards and giving me something to strive for. I greatly appreciate his dedication to people in his lab, especially the students. His insistance on the highest standards of rigor and logic have molded my approach to scientific thought. I also appreciate his promotion of brevity and clarity in presentation, both in writing and speaking. In addition, he has managed to keep a delicate balance between allowing independence and providing guidance. Finally, I thank him for absorbing and ignoring my sometimes ornerous disposition.

ABSTRACT

Cellular Influences on Retroviral Integration **Peter M. Pryciak**

An essential step in the retroviral life cycle is the insertion of a DNA copy of the viral genome into the host cell DNA. This process, termed integration, is mediated by virus-encoded machinery and represents the most thoroughly understood recombination event in eukaryotic cells. Addressed here are two issues regarding the integration of retroviral DNA.

First, the mechanism of inhibition of murine leukemia virus (MLV) integration by the mouse *Fv-1* gene was investigated. During MLV infection of Fv-1 restrictive host cells, viral DNA was synthesized but was blocked from integrating into the host cell DNA, even though extracts of these cells contained viral DNA-containing nucleoprotein complexes that were fully competent to integrate their DNA in vitro. In addition, a separate event mediated by the viral integrase (IN) protein, the removal of 2 bases from the 3' ends of linear viral DNA, occurred normally during infection of restrictive cells in vivo.

Second, factors influencing the choice of integration target sites were investigated. In naked DNA targets, integration was non-uniform, perhaps dictated by sequence-dependent variation in DNA structure, rather than by recognition of DNA sequence. Most emphasis was on the effect of chromatin assembly on integration efficiency and integration site choice, which was studied using minichromosomes as targets for integration in vitro. MLV nucleoprotein integration complexes did not show a preference for nucleosome-free regions over nucleosomal regions of chromatin.

Integration into nucleosomal DNA showed a roughly 10 bp periodic distribution of preferred sites, resulting from preferential integration at positions where the major groove of the nucleosomal DNA helix was exposed. Importantly, these sites were used more frequently than the same sites in naked DNA or than analogous sites in nucleosome-free regions, demonstrating that nucleosomal assembly can increase the reactivity of sites to retroviral integration. Target site selection was further influenced by binding of non-histone proteins and by the source of integration activity. Finally, a procedure for examining integration into SV40 DNA in vivo was developed, and the distribution of these events was closely approximated by in vitro integration reactions, although there was significantly accentuated bias between strong and weak sites in vivo.

TABLE OF CONTENTS

LIST OF FIGURES AND TABLES

CHAPTER ONE

INTRODUCTION

This thesis presents two topics that concern the integration of retroviral DNA into host cell DNA. These topics share as a common feature the ability of the host cell influence the integration process, even though integration is performed using virus-specific machinery. In the first case, the ability of a mouse gene product to block the integration of murine leukemia virus (MLV) DNA is examined in molecular detail. In the second, the contribution of the components of host cell chromatin to integration efficiency and target site selection is dissected using in vitro and in vivo models.

The retroviral life cycle

The retroviral life cycle can be conveniently viewed from either of two starting points: the extracellular virus particle (virion) or the integrated provirus. In general overview, infection consists of entry of the virus into the cell, synthesis of a DNA copy of the RNA genome, and insertion of that DNA copy into the cellular DNA. These steps involved in generating the integrated provirus from virus particle are frequently referred to as the early events of the retroviral life cycle. The inserted viral DNA, or provirus, remains in the host cell genome and there serves as a template for transcription of viral genomic and messenger RNA; subsequent translation of viral proteins then encapsidate the genomic RNA and assemble new virus particles, which leave the cell by budding from the plasma membrane. The steps responsible for production of new virus particles from the integrated provirus are called the late events of the retroviral life cycle.

I will not exhaustively review here all features of the retroviral life cycle, but will instead cover only enough points to put the issues of my thesis work on retroviral integration into the proper framework. I will also concentrate on citing only those references that are crucial to the theme of this thesis--reviews can be consulted for

additional information and references (see Varmus, 1983; Varmus and Brown, 1989; Weiss et al., 1985)

The virus particle consists of an RNA genome encased in a proteinaceous core, which is then surrounded by a lipid bilayer membrane, or envelope. Infection of a host cell begins with interaction of viral proteins embedded in the envelope with receptor molecules in the cell's plasma membrane, followed by either direct fusion of the two bilayers at the cell surface or endocytosis and subsequent fusion of the viral envelope with endosomal membrane. In either case, the virion core is released into the cytoplasm of the cell, where it can then initiate synthesis of DNA from its RNA genome in a process called reverse transcription, which is mediated by the viral protein reverse transcriptase. DNA synthesis most likely takes place inside the cell in a derivative of the virion core, since the DNA product is found as a member of a large nucleoprotein complex containing viral proteins (Bowerman et al., 1989; Bowerman, 1989), and since virion cores generated by mild detergent disruption of the virion can carry out correct DNA synthesis in vitro (Rothenberg et al., 1977). The product of this synthesis is a double-stranded linear viral DNA molecule that then migrates to the nucleus while still associated with proteins of the nucleoprotein complex. In the final step of the early stage of the retroviral life cycle, this complex serves to integrate the viral DNA into the host cell DNA. This integration step will be the primary focus of subsequent sections.

The integrated viral DNA, or provirus, remains a permanent part of the host cell genome and is therefore replicated once per cell division cycle along with the rest of host cell DNA. Late events of the life cycle begin when the provirus serves as a template for transcription of viral RNA. Full-length transcripts serve both as genomic RNA, which gets packaged into virions and serves as the template for reverse transcription upon a subsequent infection cycle, as well as messenger RNA for the synthesis of the structural components of the virion core (*gag* gene products) and of the enzymatic functions involved in synthesis and integration of viral DNA (*pol* gene products). Spliced

transcripts usually serve as messenger RNA encoding the envelope glycoproteins (*env* gene products) that stud the surface of the virion membrane. In some more complex retroviral genomes, mutiply-spliced transcripts encode regulatory and other accessory proteins. Assembly of the virus particle begins by incorporation of viral genome RNA into a capsid structure formed from the *gag* and *pol* gene products, which occurs either in the cytoplasm or at the cell surface, depending on the species of virus. The virion then buds from the cell surface by association of the capsid, or core, beneath plasma membrane that is enriched in the *env* gene products, followed by outward protrusion and pinching off of membrane to release the enveloped virus particle into the extracellular medium. The product, the extracellular virus particle, can then serve to initiate a new round of infection.

The standard viral gene products are synthesized as polyproteins that are subsequently cleaved to yield the mature protein products. The env polyprotein is cleaved by a host cell protease resident in the endoplasmic reticulum, generating a transmembrane (TM) component and a surface (SU) component. The polyproteins containing the *gag* and *pol* gene products, on the other hand, are cleaved by a viral protease (PR), which is variably encoded by either the *gag* or *pol* genes or occasionally by its own gene (*pro*), depending on the species of retrovirus; this protease is also responsible for releasing itself from the polyprotein encoding it. The polyprotein cleavages takes place either during or after assembly of the virion core, and often after the virus buds from the surface.

Each polyprotein is cleaved into two or more subunits. The gag polyprotein is cleaved into the following mature products, which make up the bulk of the virion core: matrix (MA), which associates with the viral envelope and perhaps with the cytoplasmic tail of the TM subunit of the envelope glycoprotein; capsid (CA), which is the largest of the mature *gag* gene products and contains much of the virion core self-assembly determinants; nucleocapsid (NC), which is involved in binding the viral genome RNA and

perhaps also in annealing of primer tRNA molecules to genome RNA; and usually a fourth product with unidentified function, and therefore without name, that may play some role after entry of virus into the cell.

The *pol* gene products are always synthesized first as part of a gag-pol polyprotein, in which the gag and pol domains are translationally fused into a continuous reading frame by either stop codon suppression or ribosomal frameshifting. This gag-pol polyprotein, which is usually synthesized at roughly one twentieth the level of the gag polyprotein, is then cleaved to yield the same mature gag products as from the gag polyprotein, and the following mature pol products, which participate in enzymatic processes involved in synthesis and integration of viral DNA: reverse transcriptase (RT), which has both RNA-dependent and DNA-dependent DNA polymerase activity responsible for the synthesis of linear double-stranded viral DNA from single-stranded genomic RNA, as well as an RNase H activity involved in degrading genomic RNA and generation of RNA primers during reverse transcription; and integrase (IN), which is responsible for a maturational processing of the end of linear viral DNA and integration of that DNA into the DNA of the host cell.

Integration of retroviral DNA

The generation of a permanent chromosomal template for production of progeny virus is dependent upon proper synthesis and integration of viral DNA after entry into the host cell. The consequence of integration as far as the cell is concerned is usually minor; in comparison to many DNA viruses which amplify to enormous numbers inside the cell and eventually kill it in a lytic infection, the retrovirus infection is relatively benign, resorting to a continual production of virus particles from a single copy of integrated DNA rather than an enormous burst. The usual consequence, then, is that a small new bit of genetic information is added to the host cell and the cell spends a fraction of its metabolic resources producing more virus. There are, however, some

examples of cell killing by a few species of retrovirus (Varmus and Brown, 1989). In addition, the integration event can occasionally cause a genetic phenotype because of the disruption of a critical cellular gene or the inappropriate activation of expression of a cellular gene; these topics have been intensely studied, especially because of the ability of retroviruses to participate in insertional activation of cancer-causing genes (Varmus and Brown, 1989).

In order to understand retroviral integration, I will first describe the nature of the two classes of unintegrated viral DNA molecules found in newly-infected cells: linear and circular. The linear form represents the initial DNA synthesis product. It has a structure in which sequences present only once at the termini of the RNA genome are duplicated, via some complicated gymnastics of template switching during reverse transcription, to become present as direct repeats (called long terminal repeats, or LTRs) at the termini of the linear DNA; thus, the linear DNA has the structure LTR - coding region - LTR. The terminal bases of the LTRs are imperfect inverted repeats, the extent of which depends upon the species of virus. The circular DNA forms are observed only in nuclear fractions of newly-infected cells, and they represent products of reactions acting upon linear DNA. The most familiar of these circular DNA forms are the 1-LTR and 2-LTR circles, which are molecules with structures consistent with having been formed by homologous recombination between LTRs and self-ligation of the LTR termini, respectively. A significant fraction of the circular DNAs, however, have structures consistent with having been formed by intramolecular integration of viral DNA into itself, forming so-called "autointegrants" (Shoemaker et al., 1981).

For many years a subject of debate was which of these DNA forms served as the precursor to the integrated provirus. The 2-LTR circle gradually gained favor as the precursor for circumstantial reasons, including the existence of a novel, palindromic "circle junction" sequence formed by ligation of the two ends of linear DNA together and numerous situations in which inhibition of circle formation correlated with inhibition

of integration, as well as some seemingly direct evidence (Panganiban and Temin, 1984a) for the 2-LTR circle as precursor. Eventually, however, the development of in vitro integration assays and further in vivo experimentation has shown that the linear species of DNA is the major, and probably only, immediate precursor to the integrated provirus; these developments are discussed in greater detail in a subsequent section.

The product of integration in vivo, the provirus, has a structure where the termini of linear unintegrated viral DNA has been joined to host DNA. At the junction between viral and host DNAs, a set of characteristic features are always observed that are considered to be the hallmarks of legitimate retroviral integration events: in the process of integration two base-pairs (bp) at the very end of linear viral DNA are lost and a 4-6 bp length of host DNA sequences present only once in the preintegration target becomes duplicated at either end of the viral DNA. The observation that the size of the target DNA duplication is dependent upon the species of virus and not of host was an early clue that the integration event is mediated by virus-encoded machinery and not by the host cell (see Table IV in Varmus, 1983). Another clue was the orderly and regular manner in which retroviral DNA was integrated into host DNA, which resembled that of many bacterial transposable elements, several of which were known to encode their own transposition enzymes (for reviews, see Shapiro, 1983; Berg and Howe, 1989).

Mutational studies of the viral genome uncovered sequences involved in the integration process. Two classes of regions were eventually defined as being required for normal retroviral integration: (i) the imperfect inverted repeat sequences at the termini of the viral LTR sequences (Panganiban and Temin, 1983; Colicelli and Goff, 1985; Colicelli and Goff, 1988), subsequently named "*att*" sites (for "attachment") after analogous sequences important in prokaryotic transposons; and (ii) a C-terminal domain of the *pol* gene open reading frame (Donehower and Varmus, 1984; Donehower, 1988; Panganiban and Temin, 1984b; Schwartzberg et al., 1984; Quinn and Grandgenett, 1988), subsequently named "endonuclease", later changed to "integration

protein" (or IN), and finally to "integrase" (still IN) in the current terminology. Mutations in either region had essentially the same phenotype; linear viral DNA was synthesized, and circular DNA was also formed, but the formation of integrated proviruses was blocked. Thus, a viral protein had been implicated in acting upon the ends of viral DNA in order to accomplish their integration into host DNA in vivo. The validity of this idea, and further details about the integration reaction were rapidly established upon the eventual development of in vitro assays for integration, which are thoroughly reviewed in a subsequent section.

The consequence of the defect in integration caused by these mutations was a block to the expression of viral gene products and the production of virus from the newly-infected cells. This established the essential nature of the integration step in the retroviral life cycle, which is not too surprising since virtually every step in the life cycle is essential for its continuation. The universality of this requirement for integration has, however, recently been challenged in studies of some subclasses of retroviruses that had previously received little attention. In several of the most commonly studied subclass of retroviruses, the oncornaviruses, the requirement for integration to produce infectious virus has been well established (Donehower and Varmus, 1984; Donehower, 1988; Panganiban and Temin, 1984b; Schwartzberg et al., 1984; Quinn and Grandgenett, 1988). Of the two other subclasses, the lentiviruses and the spumaretroviruses (sometimes called foamy viruses), the previously poorly-studied lentiviruses have received enormous amounts of recent study due to the epidemic resulting from infection with human immunodeficiency virus (HIV), a lentivirus. There have been a few studies that report either that productive lentivirus infection can proceed in the absence of detectable integration (Harris et al., 1984), or that cells infected with integrase mutants produce infectious virus (Stevenson et al., 1990; Prakash et al., 1992). The presence of intact integrase coding domains in such viruses, and the as yet unverified and somewhat conflicting observations with such mutants

(Stevenson et al., 1990), makes it difficult to conclude whether these retrovirus groups can truly bypass the integration step; further study should resolve this issue.

Integration in vitro

Shortly after my entry into the Harold Varmus's lab, the first in vitro integration assay was developed (Brown et al., 1987). This assay resulted from a collaborative effort between Pat Brown, then a postdoctoral fellow in Mike Bishop's lab, and Bruce Bowerman, who at the time was a graduate student in the Varmus lab. Pat was trying to develop an in vitro integration assay, and had devised a powerful genetic selection for the recovery of even very infrequent recombinants that might result, and Bruce had been characterizing the native state in which newly-synthesized viral DNA was found in the recently-infected cell. Bruce had shown that murine leukemia virus (MLV) DNA was not naked in the cell but was instead found as part of a large nucleoprotein complex, with sedimentation characteristics (160 S) that suggested its size to be larger than a ribosome (Bowerman et al., 1989; Bowerman, 1989). This observation was interpreted to suggest that the structure of the native in vivo integration machinery might be very complex and therefore difficult to properly assemble in vitro. Therefore, Pat and Bruce used extracts of MLV-infected cells and asked for the integration machinery assembled in vivo to integrate its DNA into an exogenously-provided DNA target in vitro. Those efforts were successful; the DNA-containing nucleoprotein complexes present in the extracts were competent to integrate their DNA into a target provided in vitro, and the recombinant products bore the hallmarks of legitimate retroviral (specifically MLV) integration--the loss of 2 bp from the viral DNA ends and the duplication of 4 bp of target DNA sequence (for a thorough historical discussion of the development of this assay, see Bowerman, 1989).

This original assay was optimized and refined in order to establish a number of important points (Brown et al., 1987). First, the integration activity was entirely

resident in the large DNA-containing nucleoprotein complex, which could be purified away from the majority of free proteins in the extract and retain all activity (see also Bowerman et al., 1989). Second, no energy source was required for the integration reaction. Since the reaction involves the creation of new phosphodiester bonds between the viral DNA and the target DNA, this observation implied either that the energy inherent in the cleaved target DNA was conserved and utilized in the formation of the new bonds or that the integration complex contained stored energy that was spent in the course of the reaction. Third, it was found that cytoplasmic extracts that contained only linear viral DNA and no detectable circular DNA were as effective in the reaction as nuclear extracts that contained abundant circular DNA in addition to linear DNA. This showed that the linear DNA could serve as precursor to the integrated provirus, either directly or indirectly, and raised the possibility that circular DNAs did not serve as precursors, since their presence in the nuclear extracts did not lead to an increase in recombinant products.

In order to detect integration, this first assay applied a genetic selection for recombinants in bacteria. The integration target was DNA of a lambda bacteriophage that carried amber codons in three genes essential for lytic growth, and the viral DNA carried an inserted suppressor tRNA; integration recombinants in which viral DNA had been inserted into lambda DNA then gave rise to phage that could suppress their own amber codons and therefore form plaques on a suppressor-negative strain of bacteria (Brown et al., 1987). Thus, the detection of integration products required passage through bacteria, which complicated analysis of intermediates in the integration reaction. Over time, this assay served as the starting point for continual reductionism in the development of simpler assays. First, methods were developed to physically analyze the structure of DNAs present before and after the integration reaction (Fujiwara and Mizuuchi, 1988; Brown et al., 1989), which established two important points. (i) The linear DNA present in newly-infected cells undergoes a processing

reaction shortly after completion of synthesis, with two nucleotides removed from the each of the 3' ends of the originally blunt-ended molecule; this processing reaction is dependent upon wild-type IN function in vivo (Brown et al., 1989; Roth et al., 1989). (ii) The immediate product of the integration reaction is a gapped intermediate with a structure predicted from utilization of a linear DNA as the immediate precursor and inconsistent with utilization of a circular DNA. This signalled the beginning of the end for the circular DNA precursor model, since numerous subsequent experiments overwhelmingly demonstrate that the linear molecule serves as precursor in vitro (Fujiwara and Mizuuchi, 1988; Brown et al., 1989; Craigie et al., 1990; Katz et al., 1990; Bushman and Craigie, 1991), and several determined efforts (Lobel et al., 1989; Ellis and Bernstein, 1989) have failed to support the use of 2-LTR circle precursors in vivo that was observed earlier (Panganiban and Temin, 1984a). The structural features of the observed reaction intermediate (Fujiwara and Mizuuchi, 1988; Brown et al., 1989) were sufficient to explain the in vivo hallmarks that were previously noted. Specifically, the direct joining of the 2-base recessed 3' ends of MLV DNA to the 5' ends of a 4-base staggered cleavage made in target DNA accounts for the loss of 2 bases from the end of viral DNA and the duplication of 4 bp of target DNA--which is how the product would appear after repair of the gapped intermediate by DNA replication machinery.

Further refinements led to the observation that exogenously-provided viral DNA molecules or derivatives thereof would be integrated into target DNAs by the viral DNA-containing nucleoprotein complexes, and even by RNA-containing cores from extracellular virions (Fujiwara and Craigie, 1989). This led to the eventual demonstration that purified IN protein was sufficient to accomplish the 3' processing and integration of viral DNA in the absence of any other viral or cellular proteins (Katzman et al., 1989; Craigie et al., 1990; Katz et al., 1990; Bushman et al., 1990). Currently, the most common and popular integration assay is one where the purified IN

is provided with oligonucleotides resembling the viral *att* sites, and the reaction consists of IN-mediated insertion of one oligonucleotide into another. This convenient reaction was discovered in the course of attempts at observing an IN-mediated removal of 2 bases from the 3' ends of *att* site oligonucleotides; it was found that IN would indeed cleave these bases from the ends, but would then continue by integrating these cleaved substrates into other oligonucleotides present in the reaction mixture (Craigie et al., 1990; Katz et al., 1990). Many advances have resulted from the use of this assay. For instance, the viral *att* site sequences required for 3' processing and integration have been further studied; in general they confirm the earlier in vivo studies and reinforce the idea that the terminal 2 bases immediately proximal to where the 3' cleavage reaction takes place and to the junction with target DNA, which are absolutely conserved in every retrovirus and retrotransposon (sequence 5' CA 3'), are the most critical (Bushman and Craigie, 1991; LaFemina et al., 1991; Vink et al., 1991; Leavitt et al., 1992). In addition, it has been shown that the 3' cleavage reaction is not a formal requirement for the integration reaction in vitro, since an oligonucleotide that resembles the 3' cleavage product will be used as efficiently as one that is uncleaved as an integration substrate (Craigie et al., 1990; Katz et al., 1990; Bushman and Craigie, 1991; LaFemina et al., 1991; Leavitt et al., 1992).

Most recently, the exact chemistry and stereochemistry of the 3' cleavage and strand transfer reactions have been determined (Engelman et al., 1991; Vink et al., 1992). Importantly, these experiments show that the 3' end of the integrating viral DNA makes a direct nucleophilic attack upon a phosphodiester bond in the target DNA (Engelman et al., 1991); thus, the IN protein catalyzes a concerted breakage and joining reaction in a manner that directly substitutes a phophodiester bond between the viral DNA for one present between adjacent bases in target DNA. The stereochemical inversion of chirality at the critical phosphodiester bond that is transfered strongly argues against a model where a covalent IN protein-DNA intermediate is formed, such as that formed

during topoisomerase reactions (for comparison of models, see Varmus and Brown, 1989). In addition, it has been shown that the IN protein will catalyze the reverse of the integration reaction, the formation of reactants from an integration product, and that this reverse reaction has relaxed requirements for specific sequences in the *att* site (Chow et al., 1992).

Thus, much progress has been made during the time of my tenure in the Varmus lab in delineating the mechanistic aspects of retroviral integration, primarily by stripping down the reaction to its bare essentials using increasingly simplified assays. Most of my work on integration has instead tried to bridge the gap between in vitro and in vivo observations, with the general theme that more things are happening in vivo than are being reproduced in the in vitro assays. My projects, therefore, have been for the most part attempts to make some of the more complex in vivo observations amenable to in vitro dissection.

The role of the host cell in early events

After establishment of the integrated provirus, the host cell plays numerous essential roles in the late events of the retroviral life cycle, since the transcription of viral RNA and translation of viral proteins are carried out by host machinery. But the early events of the retroviral life cycle, particularly the virus-specific processes of reverse transcription and integration, are for the most part performed by machinery brought in by the viral particle. Indeed, reverse transcription and integration reactions can be carried out using purified viral proteins in vitro. This raises the question of whether there is any participation of cellular components in the early stage of the viral life cycle, or whether the cell is a relatively passive host to the infection. This is one of the more nebulous topics in the history of retroviral replication, but it is one that I have managed to become somewhat attached to. I will consider four aspects of the early retroviral life cycle that can potentially be influenced by the host cell; a post-entry but

pre-DNA synthesis stage, DNA synthesis, subcellular localization and transport, and integration. Most of my work involved two topics that can both be considered under the heading of host cell influences on integration; the first, *Fv-1* restriction, is an example of a host gene whose product can cause the inhibition of integration in vivo, and the second, major topic regards the influence of the structural and functional state of the host DNA on choice of integration site. These two topics will be discussed briefly at the end of this section and in detail in the two subsequent sections.

Entry of virus into the cell clearly requires the participation of specific cellular components. The interaction of viral envelope proteins with cellular receptor molecules on the cell surface is necessary for entry, and cells without the proper receptors are uninfectable (Varmus and Brown, 1989). Beyond entry and through integration, however, the role of the host is only vaguely understood--although it has received a reasonable amount of attention.

The first potential step after entry that could be influenced by the host cell is the vague, so-called process of "uncoating". The problem is that it is unclear, in my opinion, what "uncoating" is, if anything; yet the term is commonly used to describe some event that follows entry but precedes DNA synthesis. It perhaps makes sense to think of the release of the virion core from its membranous envelope, which would usually occur at the cell surface, as an uncoating event; although I usually regard this as part of the entry process. Alternatively, some may consider it likely that a disassembly event takes place after entry, such that the virion core breaks down somewhat to allow viral DNA synthesis to take place, in a manner analogous to such uncoating events that occur after entry of nonenveloped viruses into cells. There is virtually no direct evidence for a disassembly event after entry, but there have been some speculations that viral DNA synthesis must be somehow "activated" upon entry into the host cell. It has been shown for MLV that, after synthesis, the viral DNA is associated with a large structure that is most likely a derivative of the virion core because it contains many of

the same proteins as in the virion core (Bowerman et al., 1989; Bowerman, 1989); thus, disassembly is not extensive in this case. Perhaps the best evidence for a mild disassembly is that the DNA-containing complexes copurify with all of the mature gag proteins except the MA protein (Bowerman, 1989); since MA is involved in association with the membrane during virus assembly (and is often fatty-acylated), it may stay behind with the viral membrane after fusion, while the remainder of the virion core enters the cell. Thus, there may be a mild disassembly event that occurs upon entry, but more direct evidence is needed.

It has been frequently speculated, however, that DNA synthesis does not begin in the virion before entry into the host cell because the virion core needs to be activated by some component or condition inside the host cell. This is mostly a philosophical argument at the current time, since there is little evidence for such an activation. It is clear that little or no DNA synthesis occurs in the virion; this is in contrast to hepadnaviruses, for example, which also replicate by reverse transcription but begin their DNA synthesis before the assembled viral particle leaves the cell (Ganem and Varmus, 1987). But mild detergent permeablization of the retroviral envelope is sufficient to allow the virion to conduct DNA synthesis in vitro, as long as it is provided with deoxyribonucleotides and divalent cations (e.g. Rothenberg et al., 1977); thus it is possible that the particle needs simply to be exposed to reagents present inside the cell in order to initiate the reverse transcription reaction. On the other hand, some experiments I performed several years ago showed that the addition of cellular extracts to detergent disrupted virions greatly enhanced the yield of viral DNA in an in vitro reverse transcription reaction (unpublished observations); this boosting activity was not pursued, although it was found to be heat-stabile, and therefore possibly an effect of salt or similar less interesting buffer conditions--although there was at least one early report of enhancement of reverse transcriptase activity by histones (Manly, 1974). Among the more compelling observations suggesting an activation step in vivo is the

behavior of mutants in the MLV gag proteins MA and p12 that assemble and release virus particles that are competent to carry out viral DNA synthesis in an in vitro detergent-permeablized assay but fail to do so upon infection in vivo (Crawford and Goff, 1984). These mutants, unfortunately, have not been pursued any further.

There have been several studies on the role of the host cell in viral DNA synthesis. The most extensive experiments have characterized infections of cells that are in a quiescent state. The usual observation is that quiescent cells do not support efficient viral DNA synthesis upon infection (Fritsch and Temin, 1977; Varmus et al., 1977; Harel et al., 1981; Zack et al., 1990). Instead of the robust levels of full-length viral DNA synthesis seen upon infection of growing cells, infection of quiescent cells show mostly incomplete synthesis, with both strands of viral DNA being less than full length. The fate of such incomplete molecules upon restimulation of the cells is a matter of some dispute, with some reports claiming that synthesis of the initiated molecules can resume (Varmus et al., 1977; Harel et al., 1981; Zack et al., 1990; Zack et al., 1992) while others claiming that they are dead intermediates (Miller et al., 1990). In any case, the primary finding has often been interpreted as an in vivo requirement for some cellular activity to participate in efficient viral DNA synthesis. The nature of the in vivo requirement, however, is unknown. It is conceivable that the quiescent cells have low levels of deoxyribonucleotides or other small molecules, and therefore provide below-threshold levels of reagents for viral DNA synthesis, or alternatively that some cellular protein that is inactive or not expressed in quiescent cells directly participates in viral DNA synthesis in vivo. The issue is unresolved, but some simple experiments examining the effect of extracts from quiescent versus growing cells on reverse transcription in detergent-permeablized virions in vitro (analogous to the unpublished experiments mentioned above) might be very informative.

The clearest example that a cellular gene product can have a direct effect on viral DNA synthesis is the case of the mouse *Fv-1* gene, which has an inhibititory role in MLV

replication, operating at either the DNA synthesis or integration step depending on the virus-host combination. This phenomenon occupied a significant portion of my thesis work, and it is discussed in detail in a subsequent section and in the second chapter.

The subcellular localization of the functional particle after entry into the cell and during DNA synthesis and integration has received little attention. It is not known whether the particle diffuses randomly about in the cytoplasm and then into the nucleus or if it associates with any cytoplasmic structures, such as cytoskeletal elements, and then actively migrates in a targeted manner. The most applicable study that relates to such issues are recent experiments analyzing nuclear localization of viral DNA as a function of the stage of the cell cycle. It has been observed that entry of MLV DNA into the nucleus, as assayed by in situ hybridization, requires the passage of the cell through M phase (T. Reynolds, T. Roe, and P. Brown, personal communication). This observation is intriguing because the mammalian cell nucleus undergoes a cycle of nuclear envelope breakdown and reformation during mitosis, suggesting that the DNA-containing nucleoprotein complex, perhaps because of its large size, is incapable of entering the nucleus until the nuclear envelope barrier is broken down. Whether this model is universally true for all retroviruses remains to be determined, but there is certainly potential for variation. For example, the only other retrovirus for which there is published information about the nature of the DNA-containing complex is HIV, and in this case the complex appears to be considerably smaller than that of MLV (Farnet and Haseltine, 1990; Farnet and Haseltine, 1991); this opens the possibility that HIV is less dependent upon the cell cycle for nuclear entry, which may be relevant to non-cycling nature of many of the primary target cells for HIV (Springett et al., 1989; Stevenson et al., 1990; Cann et al., 1990; Weinberg et al., 1991). Unpublished information suggests that Rous sarcoma virus (RSV) DNA is found associated with a complex more similar to that of MLV than HIV (D. Scheinken and J. Coffin, personal communication). In addition, there is a family of elements in yeast (*S. cerevisiae*) that

are closely related to retroviruses. The Ty retrotransposons (Boeke, 1989) assemble particles and synthesize DNA in the cytoplasm but integrate into the host chromosomes in the nucleus, even though the yeast nuclear envelope does not break down during mitosis. Thus, the Ty particles probably employ a different nuclear transport mechanism, which remains to be investigated.

Other studies have addressed the role of the host cell in integration. Some early experiments showed that integration displayed a strong tendency to occur into recently-replicated DNA, which was interpreted to suggest a requirement for active replication of the target DNA or for passage of the cell through S phase (Varmus et al., 1977; Harel et al., 1981). An X-linked cellular locus that shows temperature-sensitive effects on cellular DNA synthesis also showed decreased integration at the restrictive temperature, again suggesting a requirement for DNA synthesis, although these studies were complicated by an unexplained apparent defect in the viral DNA upon extraction and introduction into new cells by transfection (Richter et al., 1984).

Experimentation with the effects of various drugs supported some of these models. Specifically, infection in the presence of aphidicolin, an inhibitor of cellular DNA synthesis, showed an inhibition of circle formation and integration (Hagino-Yamagishi et al., 1981; Hsu and Taylor, 1982; Chinsky and Soeiro, 1982). This was primarily interpreted to indicate that circles were integration precursors and that cellular DNA polymerases were involved in circle formation, or that actively replicating target DNA was a requirement for integration. In addition, infection in the presence of the protein synthesis inhibitor cycloheximide also showed decreased circle formation and integration (Yang et al., 1980b), which was interpreted as a requirement for either maintained synthesis of a cellular product or rapid synthesis of a viral product upon introduction of the viral RNA genome into cells.

All of these observations can be accounted for by the above-mentioned recent studies showing that entry of the viral DNA-containing nucleoprotein complexes into the

nucleus, and therefore integration, requires passage of the cell through M phase of the cell cycle (T. Reynolds., T. Roe., and P. Brown., personal communication). Thus, the tendency for integration to occur in recently-replicated DNA can be explained by the fact that S phase precedes M phase, and the effects of aphidicolin and cycloheximide can be explained by the fact that they will cause cells to arrest in the cell cycle and not allow a round of nuclear envelope breakdown and reformation. In addition, some unpublished experiments that I performed, which are also discussed in Chapter 2, show that DNA-containing nucleoprotein complexes present in cells infected in the presence of either aphidicolin or cycloheximide are competent to integrate their DNA in vitro. Therefore, there is little evidence for a required and direct participation of a cellular function in integration, only for indirect requirements that allow the viral functions to continue with their business. But perhaps related to the above observations are studies showing that yeast Ty1 transposition is inhibited (at a post-transcriptional stage) by arrest of the cell cycle at G1 using mating pheromones, but not by arrest at G2 using microtubule-depolymerizing agents (Xu and Boeke, 1991). In the cells treated with mating pheromones, particles assemble that contain normal amounts of RNA but are defective for reverse transcription both in vivo and in vitro, and some differences in the protein content of the RNA-containing particles was observed. Unfortunately, the lack of an extracellular phase in the yeast Ty retrotransposition cycle makes it difficult to uncouple particle assembly from DNA synthesis and integration events.

There is at least one intriguing example of a seemingly normal and actively-growing (human) cell line that allows infection and synthesis of (murine) retroviral DNA, but does not allow integration of that DNA (Collins, 1988). Unfortunately, this observation does not seem to have been pursued to elucidate the nature of the integration block. In particular, it would be interesting to know whether the lack of integration reflected the absence of a required condition or factor, or the presence of an inhibitor. This information would be particularly applicable to what is probably the most-studied

and clearest example of a particular cellular gene influencing retroviral integration, a phenomenon known as *Fv-1* restriction, in which the mouse *Fv-1* gene product is responsible for the inhibition of MLV integration.

Fv-1 restriction

As mentioned above, the first in vitro integration assay had been developed shortly after I joined the Varmus lab. To those who had followed the phenomenon known as *Fv-1* restriction, the development of an in vitro integration assay presented a critical opportunity to explore the underlying mechanism. After familiarizing myself with the problem, I agreed that this interesting in vivo phenomenon was ripe for in vitro dissection, for reasons that I will outline below. The logic was sound, but unfortunately the process does not appear to work quite the way we had expected it to.

The mouse gene *Fv-1* was discovered during a screen for genetic contributors to retrovirus-induced disease in the mouse (for review, see Jolicoeur, 1979). Infection with Friend virus (FV), which turned out to be a mixture of MLVs, was found to cause leukemia and related diseases in some strains of mouse but not others. Several genetic loci were found to contribute to this differential susceptibility, and one, *Fv-1*, was studied with particular intensity. It was found that some MLVs, termed N-tropic, replicated well on NIH strains of mice and poorly on BALB strains, whereas other MLVs, termed B-tropic, replicated well on BALB strains but poorly on NIH strains. This reciprocal pattern was determined to result from the dominant ablility of products of a single genetic locus, called *Fv-1*, to inhibit the replication of MLVs in an allele-specific manner; thus, NIH mice carry the $Fv\text{-}1^{n}$ allele, which inhibits B-tropic MLVs, and BALB mice carry the $Fv\text{-}1^{b}$ allele, which inhibits N-tropic MLVs, and $Fv\text{-}1^{n/b}$ heterozygotes inhibit both N- and B-tropic MLVs. The stage in the MLV life cycle inhibited was found to be after entry of virus into the cell and before or including integration. Molecular analysis eventually revealed that in most cases viral DNA was

synthesized but integrated proviruses did not appear upon infection of restrictive hosts (Jolicoeur and Baltimore, 1976; Sveda and Soeiro, 1976; Yang et al., 1980a; Jolicoeur and Rassart, 1980; Jolicoeur and Rassart, 1981; Chinsky and Soeiro, 1981; Chinsky and Soeiro, 1982). Thus, this presented an intriguing example of what appeared to be a direct intervention of cellular factors into the retroviral integration process.

There were a few complications to this story. First, it was found that infection of restrictive cells would occasionally show decreased levels of viral DNA synthesis, which was therefore sufficient to explain the decrease in integrated proviruses (Yang et al., 1980a). The explanation for this pleiotropy is not entirely clear, but it seems to depend upon the particular combination of host cell and virus isolate. Second, as mentioned above, the usual observation was that viral DNA synthesis was not inhibited in restrictive infections, but there were decreases in levels of circular forms of unintegrated viral DNA, whereas linear forms were unaffected (Yang et al., 1980a; Jolicoeur and Rassart, 1980; Jolicoeur and Rassart, 1981; Chinsky and Soeiro, 1981; Chinsky and Soeiro, 1982). As described in a previous section, circumstantial evidence had been building to support the 2-LTR circle as the precursor to the integrated provirus. Since circle formation was inhibited by *Fv-1*, then decreased integration seemed a natural outcome of this inhibition. Upon development of the in vitro integration assay, evidence began to mount against the circle and for the linear as the integration precursor; eventually it was proven that linear molecules are the immediate precursor to the integrated provirus. Thus, it appears that *Fv-1* inhibits both the circularization and the integration of linear DNA. Nevertheless, these considerations did not overwhelmingly affect our approach to *Fv-1* restriction. We were presented with a situation in which the conversion from unintegrated DNA to integrated DNA was inhibited in vivo, and the newly-developed in vitro integration assay offered the opportunity to attempt to duplicate the inhibition in vitro.

The original plan was to conduct parallel infections of restrictive and permissive cells, and then from each newly-infected cell prepare extracts of the type that had been shown to carry out integration in vitro. If extracts from restrictive infections showed decreased integration in vitro compared to those from permissive infections, then this would constitute an in vitro restriction assay. Subsequently, extracts from permissively infected cells could be mixed with extracts of restrictive uninfected cells, to see if the restricting activity could be provided *in trans*; if so, this would have provided an assay for the identification and purification of the restricting cellular component. Unfortunately, when the experiments were done, there was no evidence for any restriction in vitro. Viral DNA-containing nucleoprotein complexes from restrictive infections were just as competent to integrate their DNA as those from permissive infections (see Chapter 2 for details). This showed that the in vivo integration machinery was not drastically or irreversibly damaged, but it left open the possibility of a labile or easily disassociated restriction activity to explain the failure to observe restriction in vitro. An alternative, and perhaps more optimistic, view was that restriction did not operate to directly inhibit integration activity in vivo. Importantly in this regard, I found that another function of the viral IN protein, the 3' cleavage of linear viral DNA ends, occurred normally in restrictive infections; this demonstrated that not all integration functions are inhibited in vivo. These experiments and their interpretations are presented in Chapter 2.

But if *Fv-1* does not directly inhibit integration activity in vivo, what could it inhibit? It was this question that led me to the general point of view that there may be additional events in the in vivo integration pathway that were not being duplicated in the test tube. The inability to observe *Fv-1* restriction in vitro might be a hint that there are hurdles that need to be jumped over in vivo for successful establishment of a provirus, with many of them bypassed by the use of increasingly simple integration assays. This idea was supported by some additional in vitro experiments showing that in

cases where integration was inhibited in vivo--namely, infection in the presence of aphidicolin or cycloheximide--integration was not inhibited in vitro using extracts of these cells (discussed in Chapter 2). Thus, I decided to try to address the effects on integration of features present in the in vivo situation that were not being duplicated in any of the in vitro integration assays, but that would be amenable to in vitro analysis. Eventually I shifted all of my efforts toward trying to understand how chromatin assembly and other physiological changes in target DNA could affect the use of DNA as an integration target and the precise choice of target sites.

Integration target site choice

Retroviruses do not integrate their DNA at a specific acceptor site in host cell DNA, but instead can integrate at many positions (for reviews see Sandmeyer et al., 1990; Varmus and Brown, 1989). Early restriction mapping of integration sites showed that a large number of sites are available in the host genome, and the cloning and sequencing of a small number of proviruses showed no consensus base sequence at the insertion sites (Shimotohno and Temin, 1980; Shoemaker et al., 1981). These properties can be compared to those of other, particularly prokaryotic, transposable elements for which there is information about target site specificity (see Berg and Howe, 1989). Prokaryotic transposons display a wide spectrum of target site preferences, from specificity for a single site in the bacterial genome (lambda: Thompson and Landy, 1989; Tn7: Craig, 1989) to use of large numbers of sites (Mu: Castilho and Casadaban, 1991; Tn10: Bender and Kleckner, 1992; Kleckner, 1989; Tn5: Berg, 1989; IS4: Klaer et al., 1980). Those elements that transpose to many different sites tend to show variable degrees of usage of "hot spots" and consensus sequence at the integration sites. In at least one case (Tn10) the transposase can sustain mutations that alter the target site specificity (Bender and Kleckner, 1992).

Retroviruses, as mentioned above, in general show little consensus sequence at integration sites, but some preferences most likely exist. The most extreme examples are retrotransposons, such as Ty3 from yeast (Chalker and Sandmeyer, 1990; Chalker and Sandmeyer, 1992) and the DRE element from *Dictyostelium discoideum* (Marschalek et al., 1992), that integrate exclusively in a small subpopulation of regions, which in both of these two cases are the upstream regions of tRNA genes. Such preferences are clearly not simply imposed by the host, since the yeast *S. cerevisiae* harbors two other retrotransposons, Ty1 and Ty2, which do not show the tRNA preferences of Ty3 (Boeke, 1989). In true retroviruses, however, the evidence of bias for some integration target sites over others comes from a handful of observations that all point toward the same general but rather vague notion that integration sites are not distributed completely randomly. What exactly determines the non-randomness has not been determined, only hinted at.

Several observations have suggested that retroviral integration target site selection is non-random. A particularly striking example of target site bias came from a study in which a large number of proviruses resulting from infection of chicken cells with a Rous sarcoma virus (RSV) derivative were cloned into a lambda phage library, and then randomly-picked clones were screened for their frequency of representation in the library (Shih et al., 1988). The degree of overrepresentation of some clones suggested that a small subset of integration sites accounted for roughly 20% of the total integration events. Sequencing of the recombinant junctions in two such high-frequency sites revealed that all insertions occur into the same site to the base pair, representing a bias of a million-fold over random since these sites represented roughly 1 in every 3000 events occurring in a genome of 10^9 bp. It is not yet known if such preferential sites will be observed for other retroviral species besides RSV. In another study, the frequency of knockout mutations generated by insertion of retroviral DNA into a particular gene was found to be far lower than expected if integration occurred randomly

throughout the genome (King et al., 1985). In other studies, measurements were made of the distances between integration sites and the nearest DNase I hypersensitive sites, or of the transcriptional activity of the integration region in the parental uninfected cells (Vijaya et al., 1986; Rohdewohld et al., 1987; Scherdin et al., 1990; Mooslehner et al., 1990). In these cases, there was a greater correlation between integration sites and DNase I hypersensitive or transcriptionally active regions than expected if integration sites were chosen randomly.

These observations have most frequently been interpreted to suggest that integration prefers transcriptionally active or simply more structurally open regions of chromatin. Such suggestions receive some support from the tendency of yeast Ty1 insertions to occur at the 5' ends of transcription units (Natsoulis et al., 1989). Studies in *Drosophila* have also revealed that 5' ends of transcription units tend to be preferred target sites by a variety of different kinds of transposable elements (Voelker et al., 1990; Berg and Spradling, 1991). In addition, the above-noted strict preference of yeast Ty3 for insertion adjacent to tRNA genes has been shown to be dependent upon the maintainance of intact binding sites for RNA polymerase III transcription factors (Chalker and Sandmeyer, 1992). In none of these cases, however, is it clear what mechanistic feature controls insertion site preference. The retroviral integration examples cited in general suffer from small numbers of events that were analyzed, the difficulty of comparing more than one event into single loci (except for the RSV hot spots), and the uncertainties inherent in attempting to correlate integration site choice with any one feature of target DNA when numerous physiological processes occur simultaneously in the cell. Therefore, an in vitro approach was needed.

In the development of the original in vitro integration assay, a reasonably large number of integration events (33 total) into roughly 9 kb of naked lambda DNA were mapped and seven had their recombinant junctions sequenced (Brown et al., 1987). This analysis showed that many sites were available as targets in vitro (and, in fact,

none was used more than once), and no consensus sequence appeared to govern the target sites used. We wondered if the in vivo observations could be modelled by modifications of the in vitro system to include more complex targets. Some of the in vivo observations had hinted at a preference for structurally open regions of chromatin, including perhaps nucleosome-free regions that are often associated with transcriptional control regions. Therefore we decided to attempt to use, as targets for in vitro integration, simple chromatin substrates, and preferentially ones that contained both nucleosomal and nucleosome-free regions that could be compared on the same target. The choice of chromatin substrates to use was facilitated by one of my fellow students, Jay Thomas, who had been studying glucocorticoid receptor-mediated induction of transcription of genes using minichromosome templates in Keith Yamamoto's lab. Based upon Jay's experience with a couple of different candidates, and some perusal of the literature for general characteristics of different minichromosomes, I decided to use not one but two different minichromosomes, the yeast TRP1ARS1 minichromosome and the SV40 viral minichromosome, as potential integration targets. This decision, in retrospect, may have been critical to my having found out anything interesting about integration into chromatin, since much of the eventual findings were dependent on the extremely homogeneous nature of one of the targets (TRP1ARS1) that is not inherent in the other (SV40).

I started this chromatin integration target project as I was still conducting *Fv-1* and related projects, so progress was greatly accelerated by the arrival and hard work of Anita Sil, who came as a medical student from Michigan to work in the lab for the summer. (Amazingly, in spite of her working with one of the more cynical graduate students in the department, Anita subsequently decided to take a sabbatical from medical school and come to UCSF to earn her Ph.D.) That summer Anita managed to purify enough minichromosomes for us to do some experiments, and she was able to show that the minichromosomes would indeed serve as integration targets in vitro. This was a

significant observation, since at the time we were concerned that the minichromosome DNA might be very inefficiently used because of steric inaccessibility of the nucleosome-bound DNA. This worry turned out not to be true, and instead the minichromosomes were used as efficiently as naked DNA targets, as described in Chapter 3 (Pryciak et al., 1992). Because of the success in using minichromosomes as integration targets, I set out to map where in the targets the integration events occured. There were several potential methods of mapping the insertions; none of the clever methods worked, so I instead turned to the brute-force method of cloning and sequencing large numbers of recombinants. The position of insertion in each clone could have been located by restriction mapping, but I decided that sequencing was just as easy; this was another important decision, in retrospect, since many (but not all) of the interesting observations were dependent upon the nucleotide-level resolution mapping information generated by sequencing. The integration events were distributed such that it was clear that there was no preference for nucleosome-free regions. This was a surprise, as we had partly expected to be able to explain some of the in vivo observations by preferential integration into nucleosome-free regions of chromatin. In addition, the high-resolution mapping data allowed for the detection of a periodic pattern to the distribution of integration sites, where the period matched that of the DNA helix, roughly 10 bp. This and other features, such as a high proportion of sites used more than once and some sequence bias at the integration sites in minichromosomes, led us to propose a model where the integration machinery preferentially uses the exposed face of the nucleosomal DNA helix (see Chapter 3).

Because the lessons from these experiments were interesting and informative, I made a more determined effort to develop an assay for the distribution of integration sites that would give the same single-base resolution information but would allow the mapping of large numbers, perhaps hundreds or thousands, of integration events at once. I reasoned that any method that measured the distance from a fixed position in the viral

DNA to a fixed position in the target DNA in theory would be able to give the desired sort of information. In an institution where the "awesome power" of yeast genetics, and occasionally affinity chromatography, are so highly touted, I instead managed to discover the awesome power of PCR (the polymerase chain reaction). Using PCR to amplify integration products between primer binding sites in viral and target DNA, an assay was developed that could measure the frequency of integration at individual positions in any target DNA region of interest (see Chapter 4). This assay allowed us to confirm our previous observations and show our predictions to be correct with regard to the geometrical relationship between target sites used and the rotational orientation of the nucleosomal DNA helix.

Perhaps the biggest surprise was that many of the target sites in nucleosomal DNA do not simply remain accessible, but are instead made more reactive than the same sites in naked DNA or than analogous sites in the nucleosome-free region of the same minichromosome molecule. Thus, in blatant contrast to our original naive expectations, preference for particular sites is created to the greatest degree not because sites in nucleosome-free regions are more accessible than their nucleosomal neighbors, but instead because sites in nucleosomal regions are made exceptionally reactive. This assay also allowed the demonstration that target sites can be blocked by the presence of bound factors, that purified IN molecules as well as the large integration complex will respond in these ways to the assembly of target DNA into nucleosomes, and that IN proteins from different species of retrovirus will show differences in their integration site preferences--which was nontrivial since the preferences could have simply reflected the chemical reactivity of different target DNA positions to nucleophilic attack.

Finally, I spent some time trying to develop a system where one could easily study large numbers of integration events into a single locus in vivo, and also compare the distribution of integration events in vivo with those into the same target in vitro. This turned out to be amazingly straightforward. By infecting with MLV cells that had

recently been infected with SV40, in vivo MLV integration events into SV40 minichromosomes were easily detected by the PCR method invented for the in vitro work (see Chapter 5). These experiments also represented the first comparison of in vivo and in vitro retroviral integration events into the same target. Remarkably, the distribution of integration events in vivo could be very closely mimicked by in vitro reactions using isolated SV40 minichromosomes as the target. Importantly, however, there were significant differences, hinting once again that there is more happenning in vivo than there is in vitro.

This last set of studies provides a starting point for others who wish to further examine the role of increasingly complex physiological changes in chromosomal DNA in determining integration site preference. In addition, the lessons learned from the response of retroviral integration to the assembly of DNA into chromatin has important ramifications for many cellular functions that must learn to deal with DNA in its native state; these opinions are repeated in the Discussion sections of the manuscripts (Chapters 3, 4, and 5). It is generally expected that packaging of DNA into chromatin decreases the availability of that DNA for cellular metabolic activities. Such restriction of access probably serves a useful regulatory role governing gene expression (Gross and Garrard, 1988; Morse and Simpson, 1988; Grunstein, 1990; Kornberg and Lorch, 1991; Wolffe, 1991; Felsenfeld, 1992). However, the identification of a function that is relatively indifferent to encounters with nucleosomes, and in fact prefers many sites in nucleosomal DNA over nucleosome-free DNA, suggests that there may be classes of cellular functions which have the ability to overcome nucleosomal inhibition. Further work may elucidate whether the behavior of retroviral integration machinery is more of an exception or a rule.

REFERENCES

Bender, J. and Kleckner, N. (1992). IS10 transposase mutations that specifically alter target site recognition. EMBO J. *11*, 741-750.

Berg, C. A. and Spradling, A. C. (1991). Studies on the rate and site-specificity of P element transposition. Genetics *127*, 515-524.

Berg, D. E. (1989). Transposon Tn5. In Mobile DNA, D. E. Berg and M. M. Howe, eds. (Washington, D.C.: American Society for Microbiology), pp. 185-210.

Berg, D. E. and Howe, M. M. (1989). Mobile DNA (Washington, D.C.: American Society for Microbiology).

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

Bowerman, B. (1989) A nucleoprotein complex that mediates the integration reaction of Moloney murine leukemia virus (Ph.D. Thesis: University of California, San Francisco)

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

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

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

Bushman, F. D. and Craigie, R. (1991). Activities of human immunodeficiency virus (HIV) integration protein in vitro: specific cleavage and integration of HIV DNA. Proc. Natl. Acad. Sci. USA *88*, 1339-1343.

Bushman, F. D., Fujiwara, T. and Craigie, R. (1990). Retroviral DNA integration directed by HIV integration protein in vitro. Science *249*, 1555-1558.

Cann, A. J., Zack, J. A., Go, A. S., Arrigo, S. J., Koyanagi, Y., Green, P. L., Koyanagi, Y., Pang, S. and Chen, I. S. Y. (1990). Human immunodeficiency virus type 1 T-cell tropism is determined by events prior to provirus formation. J. Virol. *64*, 4735-4742.

Castilho, B. A. and Casadaban, M. J. (1991). Specificity of mini-Mu bacteriophage insertions in a small plasmid. J. Bacteriol. *173*, 1339-1343.

Chalker, D. L. and Sandmeyer, S. B. (1990). Transfer RNA genes are genomic targets for *de novo* transposition of the yeast retrotransposon Ty3. Genetics *126*, 837-850.

Chalker, D. L. and Sandmeyer, S. B. (1992). Ty3 integrates within the region of RNA polymerase III transcription initiation. Genes Dev. *6*, 117-128.

Chinsky, J. and Soeiro, R. (1981). *Fv-1* host restriction of Friend leukemia virus: analysis of unintegrated proviral DNA. J. Virol. *40*, 45-55.

Chinsky, J. and Soeiro, R. (1982). Studies with aphidicolin on the *Fv-1* host restriction of Friend murine leukemia virus. J. Virol. *43*, 182-190.

Chow, S. A., Vincent, K. A., Ellison, V. and Brown, P. O. (1992). Reversal of integration and DNA splicing mediated by integrase of human immunodeficiency virus. Science *255*, 723-726.

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.

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

Collins, S. J. (1988). Different mechanisms account for the relative resistance of KG-1 and HL-60 cell lines to retrovirus infection. J. Virol. *62*, 4346-4348.

Craig, N. L. (1989). Transposon Tn7. In Mobile DNA, D. E. Berg and M. M. Howe, eds. (Washington, D.C.: American Society for Microbiology), pp. 211-226.

Craigie, R., Fujiwara, T. and Bushman, F. (1990). The IN protein of moloney murine leukemia virus processes the viral DNA ends and accomplishes their integration in vitro. Cell *62*, 829-837.

Crawford, S. and Goff, S. P. (1984). Mutations in *gag* proteins p12 and p15 of Moloney murine leukemia virus block early stages of infection. J. Virol. *49*, 909-917.

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.

Donehower, L. A. 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 *81*, 6461-6465.

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

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

Farnet, C. M. and Haseltine, W. A. (1990). Integration of human immunodeficiency virus type 1 DNA in vitro. Proc. Natl. Acad. Sci. USA *87*, 4164-4168.

Farnet, C. M. and Haseltine, W. A. (1991). Determination of viral proteins present in the human immunodeficiency virus type 1 preintegration complex. J. Virol. *65*, 1910-1915.

Felsenfeld, G. (1992). Chromatin as an essential part of the transcriptional mechanism. Nature *355*, 219-224.

Fritsch, E. F. and Temin, H. (1977). Inhibition of viral DNA synthesis in stationary chicken embryo fibroblasts infected with avian retroviruses. J. Virol. *24*, 461-469.

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

Ganem, D. and Varmus, H. E. (1987). The molecular biology of the hepatitis B viruses. Annu. Rev. Biochem. *56*, 651-693.

Gross, D. S. and Garrard, W. T. (1988). Nuclease hypersensitive sites in chromatin. Annu. Rev. Biochem. *57*, 159-197.

Grunstein, M. (1990). Histone function in transcription. Annu. Rev. Cell Biol. *6*, 643-678.

Hagino-Yamagishi, K., Kano, K. and Mano, Y. (1981). Effects of aphidicolin on retrovirus DNA synthesis in vivo. Biochem. Biophys. Res. Comm. *102*, 1372-1378.

Harel, J., Rassart, E. and Jolicoeur, P. (1981). Cell cycle dependence of synthesis of unintegrated viral DNA in mouse cells newly infected with murine leukemia virus. Virology *110*, 202-207.

Harris, J. D., Blum, H., Scott, J., Traynor, B., Ventura, P. and Haase, A. (1984). Slow virus visna: reproduction in vitro of virus from extrachromosomal DNA. Proc. Natl. Acad. Sci. USA *81*, 7212-7215.

Hsu, T. W. and Taylor, J. M. (1982). Effect of aphidicolin on avian sarcoma virus replication. J. Virol. *44*, 493-498.

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

Jolicoeur, P. and Baltimore, D. (1976). Effect of *Fv-1* gene product on proviral DNA formation and integration in cells infected with murine leukemia viruses. Proc. Natl. Acad. Sci. USA *73*, 2236-2240.

Jolicoeur, P. and Rassart, E. (1980). Effect of Fv-1 gene product on synthesis of linear and supercoiled viral DNA in cells infected with murine leukemia virus. J. Virol. *33*, 183-195.

Jolicoeur, P. and Rassart, E. (1981). Fate of unintegrated viral DNA in *Fv-1* permissive and resistant mouse cells infected with murine leukemia virus. J. Virol. *37*, 609-619.

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

Katzman, M., Katz, R. A., Skalka, A. M. and Leis, J. (1989). The avian retroviral integration protein cleaves the terminal sequences of linear viral DNA at the in vivo sites of integration. J. Virol. *63*, 5319-5327.

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.

Klaer, R., Kuhn, S., Fritz, H.-J., Tillmann, E., Saint-Girons, I., Habermann, P., Pfeifer, D. and Starlinger, P. (1980). Studies on transposition mechanisms and specificity of IS4. Cold Spr. Harb. Symp. Quant. Biol. *45*, 215-224.

Kleckner, N. (1989). Transposon Tn10. In Mobile DNA, D. E. Berg and M. M. Howe, eds. (Washington, D.C.: American Society for Microbiology), pp. 227-268.

Kornberg, R. D. and Lorch, Y. (1991). Irresistible force meets immovable object: transcription and the nucleosome. Cell *67*, 833-836.

LaFemina, R. L., Callahan, P. L. and Cordingley, M. G. (1991). Substrate specificity of recombinant human immunodeficiency virus integrase protein. J. Virol. *65*, 5624-5630.

Leavitt, A. D., Rose, R. B. and Varmus, H. E. (1992). Both substrate and target oligonucleotide sequences affect in vitro integration mediated by HIV-1 integrase protein (IN) produced in Saccharomyces cerevisiae. J. Virol. in press.

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

Manly, K. F. (1974). Histones stimulate polyribonucleotide-directed polydeoxyribonucleotide synthesis by murine leukemia virus. J. Virol. *13*, 305-311.

Marschalek, R., Hofmann, J., Schumann, G., Gosseringer, R. and Dingermann, T. (1992). Structure of DRE, a retrotransposable element which integrates with position specificity upstream of Dictyostelium discoideum tRNA genes. Mol. Cell. Biol. *12*, 229-239.

Miller, D. G., Adam, M. A. and Miller, A. D. (1990). Gene transfer by retrovirus vectors occurs only in cells that are actively replicating at the time of infection. Mol. Cell. Biol. *10*, 4239-4242.

Mooslehner, K., Karls, U. and Harbers, K. (1990). Retroviral integration sites in transgenic mov mice frequently map in the vicinity of transcribed DNA regions. J. Virol. *64*, 3056-3058.

Morse, R. H. and Simpson, R. T. (1988). DNA in the nucleosome. Cell *54*, 285-287.

Natsoulis, G., Thomas, W., Roghmann, M.-C., Winston, F. and Boeke, J. D. (1989). Ty1 transposition in Saccharomyces cerevisiae is nonrandom. Genetics *123*, 269-279.

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

Panganiban, A. T. and Temin, H. M. (1984a). Circles with two tandem LTRs are precursors to integrated retrovirus DNA. 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.

Prakash, K., Ranganathan, P. N., Mettus, R., Reddy, P., Srinivasan, A. and Plotkin, S. (1992). Generation of deletion mutants of simian immunodeficiency virus incapable of proviral integration. J. Virol. *66*, 167-171.

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

Quinn, T. P. and Grandgenett, D. P. (1988). Genetic evidence that the avian retrovirus DNA endonuclease domain of *pol* is necessary for viral integration. J. Virol. *62*, 2307-2312.

Richter, A., Ozer, H. L., DesGroseillers, L. and Jolicoeur, P. (1984). An X-linked gene affecting mouse cell DNA synthesis also affects production of unintegrated linear and supercoiled DNA of murine leukemia virus. Mol. Cell. Biol. *4*, 151-159.

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

Roth, M. J., Schwartzberg, P. L. and Goff, S. P. (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.

Rothenberg, E., Smotkin, D., Baltimore, D. and Weinberg, R. A. (1977). In vitro synthesis of infectious DNA of murine leukemia virus. Nature *269*, 122-126.

Sandmeyer, S. B., Hansen, L. J. and Chalker, D. L. (1990). Integration specificity of retrotransposons and retroviruses. Annu. Rev. Genet. *24*, 491-518.

Scherdin, U., Rhodes, K. and Breindl, M. (1990). Transcriptionally active genome regions are preferred targets for retrovirus integration. J. Virol. *64*, 907-912.

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.

Shapiro, J. A. (1983). Mobile Genetic Elements (New York: Academic Press).

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

Shimotohno, K. and Temin, H. M. (1980). No apparent nucleotide sequence specificity in cellular DNA juxtaposed to retrovirus proviruses. Proc. Natl. Acad. Sci. USA *77*, 7357-7361.

Shoemaker, C., Hoffmann, J., Goff, S. P. and Baltimore, D. (1981). Intramolecular integration within moloney murine leukemia virus DNA. J. Virol. *40*, 164-172.

Springett, G. M., Moen, R. C., Anderson, S., Blaese, R. M. and Anderson, W. F. (1989). Infection efficiency of T lymphocytes with amphotropic retroviral vectors is cell cycle dependent. J. Virol. *63*, 3865-3869.

Stevenson, M., Stanwick, T. L., Dempsey, M. P. and Lamonica, C. A. (1990). HIV-1 replication is controlled at the level of T cell activation and proviral integration. EMBO J. *9*, 1551-1560.

Sveda, M. M. and Soeiro, R. (1976). Host restriction of Friend leukemia virus: synthesis and integration of the provirus. Proc. Natl. Acad. Sci. USA *73*, 2356-2360.

Thompson, J. F. and Landy, A. (1989). Regulation of bacteriophage lambda site-specific recombination. In Mobile DNA, D. E. Berg and M. M. Howe, eds. (Washington, D.C.: American Society for Microbiology), pp. 1-22.

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

Varmus, H. E. and Brown, P. (1989). Retroviruses. In Mobile DNA, D. E. Berg and M. M. Howe, eds. (Washington D.C.: American Society for Microbiology), pp. 53-108.

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.

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

Vink, C., vanGent, D. C., Elgersma, Y. and Plasterk, R. H. A. (1991). Human immunodeficiency virus integrase protein requires a subterminal position of its viral DNA recognition sequence for efficient cleavage. J. Virol. *65*, 4636-4644.

Vink, C., Yeheskiely, E., vanderMarel, G. A., vanBoom, J. H. and Plasterk, R. H. A. (1992). Site-specific hydrolysis and alcoholysis of human immunodeficiency virus DNA termini mediated by the viral integrase protein. Nucl. Acids Res. *19*, 6691-6698.

Voelker, R. A., Graves, J., Gibson, W. and Eisenberg, M. (1990). Mobile element insertions causing mutations in the Drosophila *suppressor of sable* locus occur in DNase I hypersensitive subregions of 5'-transcribed nontranslated sequences. Genetics *126*, 1071-1082.

Weinberg, J. B., Matthews, T. J., Cullen, B. R. and Malim, M. H. (1991). Productive human immunodeficiency virus type 1 (HIV-1) infection of nonproliferating human monocytes. J. Exp. Med. *174*, 1477-1482.

Weiss, R., Teich, N., Varmus, H. and Coffin, J. (1985). RNA Tumor Viruses (Cold Spring Harbor: Cold Spring Harbor Laboratory).

Wolffe, A. P. (1991). Activating chromatin. Current Biology *1*, 366-368.

Xu, H. and Boeke, J. D. (1991). Inhibition of Ty1 transposition by mating pheromones in Saccharomyces cerevisiae. Mol. Cell. Biol. *11*, 2736-2743.

Yang, W. K., Kiggans, J. O., Yang, D.-M., Ou, C.-y., Tennant, R. W., Brown, A. and Bassin, R. H. (1980a). Synthesis and circularization of N- and B-tropic retroviral DNA in *Fv-1* permissive and restrictive mouse cells. Proc. Natl. Acad. Sci. USA *77*, 2994-2998.

Yang, W. K., Yang, D.-M. and Kiggans, J. O. (1980b). Covalently closed circular DNAs of murine type C retrovirus: depressed formation in cells treated with cycloheximide early after infection. J. Virol. *36*, 181-186.

Zack, J. A., Arrigo, S. J., Weisman, S. R., Go, A. S., Haislip, A. and Chen, I. S. Y. (1990). HIV-1 entry into quiescent primary lymphocytes: molecular analysis reveals a labile, latent viral structure. Cell *61*, 213-222.

Zack, J. A., Haislip, A. M., Krogstad, P. and Chen, I. S. Y. (1992). Incompletely reverse-transcribed human immunodeficiency virus type 1 genomes in quiescent cells can function as intermediates in the retroviral life cycle. J. Virol. *66*, 1717-1725.

CHAPTER TWO

Fv-1 RESTRICTION AND ITS EFFECTS ON MLV INTEGRATION IN VIVO AND IN VITRO

ABSTRACT

We have investigated the molecular mechanisms underlying the restriction of murine leukemia virus (MLV) replication by the mouse *Fv-1* locus. Restriction in culture can be fully accounted for by a block to the accumulation of integrated proviruses. Inhibition of productive infection is closely paralleled by a reduction in levels of linear DNA in a cytoplasmic fraction; but nuclear DNA levels are nearly normal and total DNA levels are only mildly affected in restrictive cases. Although integration is blocked in vivo, the integrase (IN)-dependent trimming of 3' ends of viral DNA occurs normally in vivo during restrictive infections; thus, not all IN-mediated events are prevented in vivo. Viral DNA-containing nucleoprotein complexes in nuclear extracts from infections of restrictive cells are fully competent to integrate their DNA in vitro. Furthermore, the complexes are of normal size, and capsid (CA) protein is intact and present in normal amounts in both cytoplasmic and nuclear fractions. Thus, although integration is inhibited in vivo, nucleoprotein preintegration complexes are structurally and functionally normal when removed from the cell and analyzed in vitro. Restricting activity may be lost in vitro or, alternatively, *Fv-1* restriction may prevent a prerequisite step for integration in vivo that is bypassed in vitro.

INTRODUCTION

Fv-1 is a normal mouse gene that encodes the ability to inhibit the replication of certain classes of mouse retroviruses, the murine leukemia viruses (MLVs; for reviews, see Jolicoeur, 1979; Yang et al., 1983). It is one of several genes originally identified in a genetic screen for factors that contribute to the susceptibility to disease induced by the Friend virus (FV) MLV complex. There are two main alleles of *Fv-1*, $Fv\text{-}1^n$ and $Fv\text{-}1^b$, so called because of their presence in prototypical mouse strains NIH and BALB, respectively. Each allele encodes the ability to inhibit a particular class of MLV strain: $Fv\text{-}1^n$ inhibits B-tropic MLVs and $Fv\text{-}1^b$ inhibits N-tropic MLVs. Thus, N-tropic MLVs replicate well on NIH cells but poorly on BALB cells, whereas B-tropic MLVs replicate well on BALB cells but poorly on NIH cells. This inhibition, known as *Fv-1* restriction, is dominant--$Fv\text{-}1^{n/b}$ heterozygotes inhibit both N- and B-tropic MLV replication. Virtually all inbred strains of laboratory mice carry one of these two *Fv-1* alleles (Jolicoeur, 1979), although this prevalence is decreased in wild mice or other *Mus* species besides *Mus musculus* (Kozak, 1985); thus there exist mouse strains and cell lines that are phenotypically negative ($Fv\text{-}1^-$).

The determinants of viral tropism lie within the capsid (CA) protein (formerly called $p30^{gag}$) encoded by the *gag* gene, and a swap of two adjacent amino acids in CA between N and B sequences can completely reverse viral tropism (Boone et al., 1983; DesGroseillers and Jolicoeur, 1983). Nearly all isolates of ecotropic MLVs from laboratory strains of mice are either N- or B-tropic (Jolicoeur, 1979), although again in wild or other species of mice some isolates are insensitive to both alleles of *Fv-1* (Lieber et al., 1975; Voytek and Kozak, 1988). In addition, many common laboratory strains of MLV, such as Moloney MLV, have acquired insensitivity to *Fv-1* restriction, presumably through mutation during multiple passages in culture (Hopkins et al., 1977; Duttagupta and Soeiro, 1981). MLV strains that are not restricted by either allele of *Fv-1* are termed NB-tropic. Such strains would appear to be missing

determinants for restriction, since sensitivity to restriction is dominant; that is, NB-tropic MLVs can acquire sensitivity to restriction by phenotypic mixing with N- or B-tropic MLVs (Kashmiri et al., 1977), and mixed viral particles containing both N- and B-tropic *gag* gene products are dually tropic (i.e. are sensitive to both *Fv-1* alleles; Rein et al., 1976).

The ability to study *Fv-1* restriction in cell culture led to the present understanding that restriction operates after entry of virus into the cell but before or including integration of viral DNA into the host cell genome (Huang et al., 1973; Jolicoeur and Baltimore, 1976; Sveda and Soeiro, 1976; Jolicoeur and Rassart, 1980; Yang et al., 1980a; Chinsky and Soeiro, 1981; Jolicoeur and Rassart, 1981). Restriction is not absolute; usually between 10- and 1000-fold fewer cells are productively infected in a restrictive host than in a permissive host. The exact point of the viral replication cycle that it is inhibited is subject to some pleiotropy, apparently dependent on the particular host-virus combination. In some cases, restriction can be accounted for by an inhibition of viral DNA synthesis (Yang et al., 1980a). In most cases, however, restriction cannot be explained by decreased viral DNA synthesis--normal or nearly normal levels of linear viral DNA are synthesized and yet the acquisition of integrated proviruses (usually measured by virus production from the recently-infected cells) is severely decreased (Jolicoeur and Rassart, 1980; Yang et al., 1980a; Chinsky and Soeiro, 1981; Jolicoeur and Rassart, 1981). Transfection experiments also show that virus production from integrated genomes is not inhibited in restrictive hosts (Hsu et al., 1978; Chinsky et al., 1984).

Thus, *Fv-1* restriction can result in the prevention of integration, providing an intriguing example of host cell participation in the early stages of the retroviral life cycle. An additional observation from the earlier studies is that while linear viral DNA appears in normal amounts in most restrictive infections, levels of circular viral DNA forms are decreased in a manner that correlates with restriction (Jolicoeur and

Rassart, 1980; Yang et al., 1980a; Chinsky and Soeiro, 1981; Jolicoeur and Rassart, 1981). At that time, it was generally believed that circular viral DNA (particularly the "2-LTR circle") was the precursor to integrated DNA, and therefore that *Fv-1* restriction of integration proceeded by inhibition of precursor circle formation. It is now clear that linear viral DNA, and not circular DNA, is the precursor to integrated DNA (Fujiwara and Mizuuchi, 1988; Brown et al., 1989; Ellis and Bernstein, 1989; Lobel et al., 1989; Craigie et al., 1990; Katz et al., 1990); thus, decreased circle formation is more likely a reflection of the same inhibitory action that prevents integration in restrictive cells.

The ability to learn more about the contribution of the *Fv-1* gene product (the identity of which remains unknown) to the control of MLV replication is offerred by the recent development of in vitro assays for retroviral integration (Brown et al., 1987; Fujiwara and Mizuuchi, 1988; Brown et al., 1989; Fujiwara and Craigie, 1989; Bushman et al., 1990; Craigie et al., 1990; Katz et al., 1990). We report here our efforts to duplicate *Fv-1*-mediated inhibition of MLV integration in a cell-free system, using extracts of restrictively or permissively infected cells and assaying for integration of MLV DNA into an exogenously-provided target DNA in vitro. We have found that the cell-free assay fails to duplicate *Fv-1* restriction; instead, extracts from restrictively infected cells contain fully functional pre-integration nucleoprotein complexes. Thus, the restricting activity present in vivo is either inactivated in our in vitro system, or it does not operate by directly inhibiting integration activity in vivo. Importantly, we find that a separate process also mediated by the viral integrase (IN) protein, the removal of 2 bases from the 3' end of linear viral DNA, occurs normally in vivo in restrictive cells; this shows that not all integration functions are inhibited in vivo. We also present some novel observations of accumulation of linear viral DNA in restrictive infections uncovered by fractionation of newly-infected cell extracts.

RESULTS

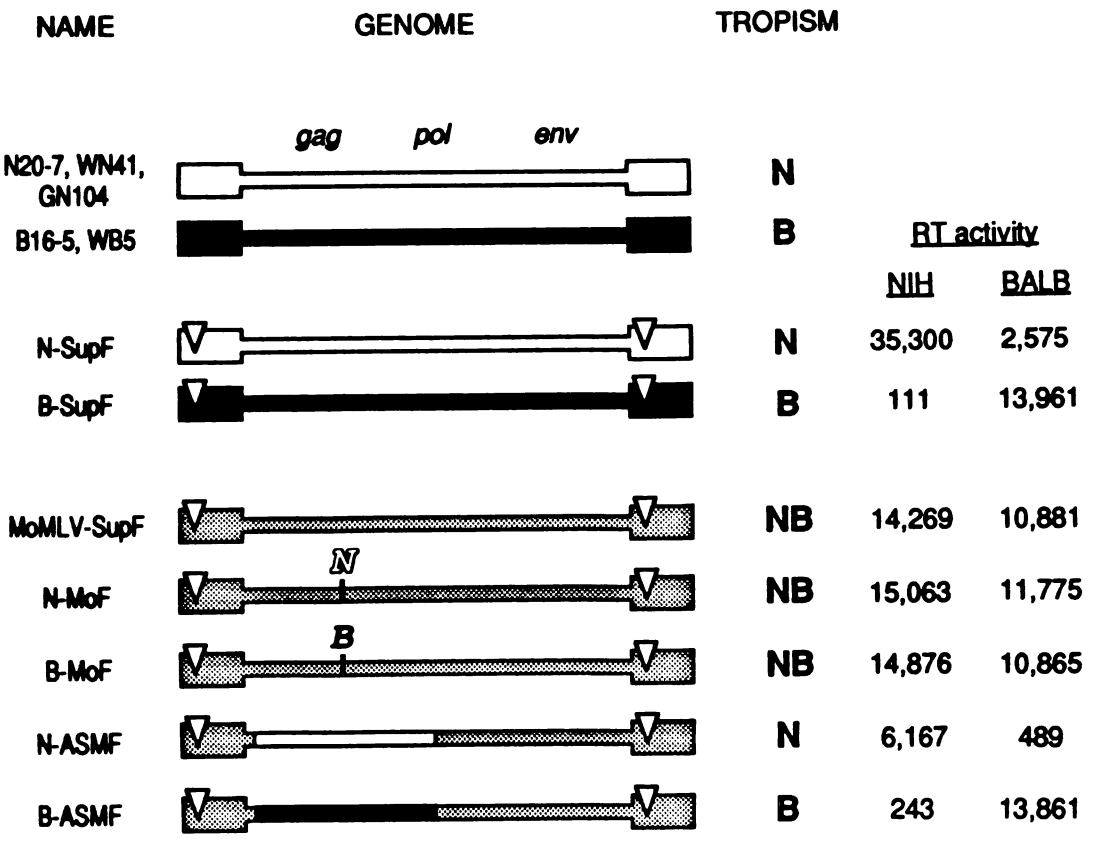
In order to study *Fv-1* restriction by in vitro integration, we required (at the time that these experiments were initiated) N- and B- tropic viruses carrying the SupF bacterial amber suppressor tRNA gene in order to apply a genetic selection of integration products (Brown et al., 1987). We generated such SupF-containing viruses in two ways: (i) insertion of the SupF gene into the N- and B-tropic MLV LTRs (to generate N-SupF and B-SupF; see Materials and Methods); and (ii) conversion of MoMLV-SupF from NB-tropic to N- and B-tropic versions by changing *gag* sequences (see below and Materials and Methods). The parental and derivative viruses used in this study are shown in Figure 1.

Determinants of NB-tropism

The extent of change in Moloney NB-tropic viral DNA that was required in order to generate N- and B-tropic viruses was informative with regard to the determinants of viral tropism. The determinants of N- and B-tropism had been previously mapped to two adjacent amino acid residues in CA (positions 109 and 110); swapping only this dipeptide between N and B viral DNAs completely reversed, and did not eliminate, viral tropism (Boone et al., 1983; DesGroseillers and Jolicoeur, 1983). The NB-tropic MoMLV contains a different dipeptide at this position than either N or B, which could account for its NB-tropism, but it also has other differences in CA and in the rest of *gag* (DesGroseillers and Jolicoeur, 1983). We made site-specific mutations in MoMLV CA to change positions 109/110 to each of the N and B dipeptides (generating N-MoF and B-MoF; see Materials and Methods and Figure 1 legend). This change was not sufficient to convert NB-tropic MoMLV-SupF to either N- or B-tropism (Figure 1). In addition, Rauscher MLV has the same dipeptide as N-tropic viruses (DesGroseillers and Jolicoeur, 1983) and yet is NB-tropic. Together, these observations suggest that the N

Figure 2-1. Viruses used in this study. The names, genome organization, and tropism of parental and derivative viruses are given. The genomes of prototypical virus clones N20-7 and B16-5 (DesGroseillers and Jolicoeur, 1983) and clones WN41, WB5, and GN104 (Boone et al., 1983) are schematically indicated as white or black for N- or B-tropic viruses, respectively. N-SupF and B-SupF are derivatives of N20-7 and B16-5, respectively, constructed by insertion of the SupF tRNA gene into the Pst I site of the viral LTRs (see Materials and Methods). Examples of reverse transcriptase (RT) activity (given in cpm) released 24 hr after infection of either NIH or BALB cells are given to demonstrate that the derivatives retain their tropism. A derivative, MoMLV-SupF (see Brown et al., 1987), of a prototypical NB-tropic virus (shaded genome) served as the parent for additional derivatives: N-MoF and B-MoF were created by site-directed mutagenesis of tropism-determining codons 109 and 110 (DesGroseillers and Jolicoeur, 1983) of MoMLV to those of N- and B-tropic viruses, respectively (see Materials and Methods); N-ASMF and B-ASMF were created by replacing sequences from the Aat II to Sal I sites in MoMLV-SupF with the analogous Aat II to Sal I fragment from N20-7 and B16-5, respectively (see Materials and Methods). Examples of RT activity (given in cpm) released 24 hr after infection of NIH or BALB cells with these viruses demonstrates that the point mutations (N- and B-MoF) were insufficient to confer N- or B-tropism, while the larger chimeric swaps (N- and B-ASMF) were sufficient to confer N- or B-tropism upon the derivative viruses.

Figure 2-1


NAME
GENOME
TROPISM
gag
pol
env
N20-7, WN41, GN104
N
B16-5, WB5
B
RT activity
NIH
BALB
N-SupF
N
35,300
2,575
B-SupF
B
111
13,961
MoMLV-SupF
NB
14,269
10,881
N
N-MoF
NB
15,063
11,775
B
B-MoF
NB
14,876
10,865
N-ASMF
N
6,167
489
B-ASMF
B
243
13,861

and B dipeptides are not sufficient determinants, but probably must be recognized in a particular CA context.

A larger change in NB-tropic viral DNA was made by swapping all of *gag* plus part of *pol* of MoMLV for N or B counterparts (generating N-ASMF and B-ASMF); the lack of corresponding restriction sites between the N- and B-tropic MLV sequences and the MoMLV sequence made this the smallest convenient swap. This swap was sufficient to convert the NB-tropic MoMLV-SupF to N- or B- tropism in the chimeras (Figure 1). Comparing the two experiments, it would appear that mutations in regions of *gag* other than at the 109/110 dinucleotide can result in loss of sensitivity to *Fv-1* restriction. Notably, the swaps in the chimeras did not include the IN domain of *pol*; therefore the absence of sensitivity to *Fv-1*-mediated inhibition of integration in the NB-tropic MoMLV is not due to differences in the IN coding region.

Accumulation of viral DNA in permissive and restrictive infections

Earlier work showed that, in some host-virus combinations, *Fv-1* restriction could be accounted for by inhibition of viral DNA synthesis (Yang et al., 1980a). In other host-virus combinations, however, no defect in accumulation of linear viral DNA synthesis was seen, but circular DNA formation and integration were inhibited (Jolicoeur and Rassart, 1980; Yang et al., 1980a; Chinsky and Soeiro, 1981; Jolicoeur and Rassart, 1981). We specifically wanted to address the nature of the integration block, using an in vitro integration assay, in situations where linear DNA synthesis is normal in restrictive infections. Therefore, we analyzed the accumulation of viral DNA after infection of permissive and restrictive host cell lines by the viruses shown in Figure 1. Since our intent was to test integration in vitro of DNAs synthesized in vivo, we fractionated newly-infected cells into cytoplasmic and nuclear extracts of the sort that would normally serve as the source of integration machinery in unrestricted MoMLV infections (Brown et al., 1987). The accumulation of DNA in these fractions was

then studied. Note that these fractions are operationally defined, and do not necessarily constitute solely the contents of either the cytoplasm or the nucleus.

Surprisingly, we observed a novel DNA accumulation phenotype: linear DNA in the cytoplasmic fraction, but not the nuclear fraction, was decreased upon infection of restrictive cells, as compared to permissive cells (Figure 2). This effect is shown for 3 different viruses on NIH and BALB cell lines (Figure 2A) and for 2 additional viruses on NIH, BALB, and SC-1 (*Fv-1*$^-$) cell lines (Figure 2B); it was also observed for the two AKV/MoMLV chimeric viruses (see Figure 6). Importantly, the degree of effect on cytoplasmic DNA correlated well with the degree of actual restriction (as measured by virus production in duplicate infections). For example, roughly 20-fold differences were seen in both virus production and cytoplasmic DNA comparing each virus on permissive hosts (Figure 2A). Linear DNA in the nuclear fraction, on the other hand, was only slightly or not at all decreased (Figures 2A, 2B). Circular DNA in the nuclear fraction was usually inhibited, although its appearance even in permissive infections was often difficult to detect. The pattern of decreased linear DNA in the cytoplasmic fraction was reproducibly observed in over 20 different experiments (not shown). The effect was independent of whether cell lysis was achieved by digitonin, NP-40, or dounce homogenization (not shown). In addition, the pattern of DNA accumulation was unaffected when infections were carried out in the presence of cycloheximide (not shown), demonstrating that this restriction-associated phenomenon does not require induction of new protein synthesis in response to infection.

A time course analysis of appearance of viral DNA in cytoplasmic and nuclear fractions shows that cytoplasmic DNA is decreased and nuclear DNA is uninhibited in restrictive infections even at the earliest times that viral DNA appears in those fractions (Figure 3). The absence of DNA in both cytoplasmic and nuclear fractions of restrictive infections at the earliest time (4 hr) suggests that the decrease in cytoplasmic DNA is not due to a more rapid migration of DNA to the nuclear fraction.

Figure 2-2. Accumulation of DNAs in cytoplasmic and nuclear fractions during permissive and restrictive infections. **A.** Two N-tropic viruses (WN41 and GN104) and one B-tropic virus (WB5) were used to infect NIH and BALB cells, as indicated. Nuclear and cytoplasmic extracts were prepared 14 hr after infection. Virus was harvested 36 hr after infection from separated plates infected in parallel and a dot-blot of serial 1:5 dilutions of virion RNA, hybridized with an AKV MLV probe, is shown. DNAs present in the various extracts were analyzed by agarose gel electrophoresis and Southern blotting, using an AKV MLV probe. Equal proportions of cytoplasmic and nuclear DNAs were loaded onto a single gel, and the hybridized blot was exposed to film for a short period and a ten-fold longer period. The positions of linear and circular DNAs are indicated (black arrows). (The hybridizing material between linear and circular DNAs in the nuclear fraction corresponds to small amounts of fragmented chromosomal DNA present in the extract.) **B.** An N-tropic virus (N-SupF) and a B-tropic virus (B-SupF) were used to infect NIH, BALB, and SC-1 cell lines, as indicated. Nuclear and cytoplasmic extracts were prepared 12 hr after infection, and the DNAs present were analyzed as in (A). Virus harvested from parallel plates 24 hr after infection gave reverse transcriptase activities of 8112, 155, 246, and 3746 cpm for N on NIH, N on BALB, B on NIH, and B on BALB, respectively (virus from SC-1 was not assayed). Amounts of DNA loaded on the gels in both (A) and (B) corresponds to approximately 10^6 cell equivalents per lane.

Figure 2-2

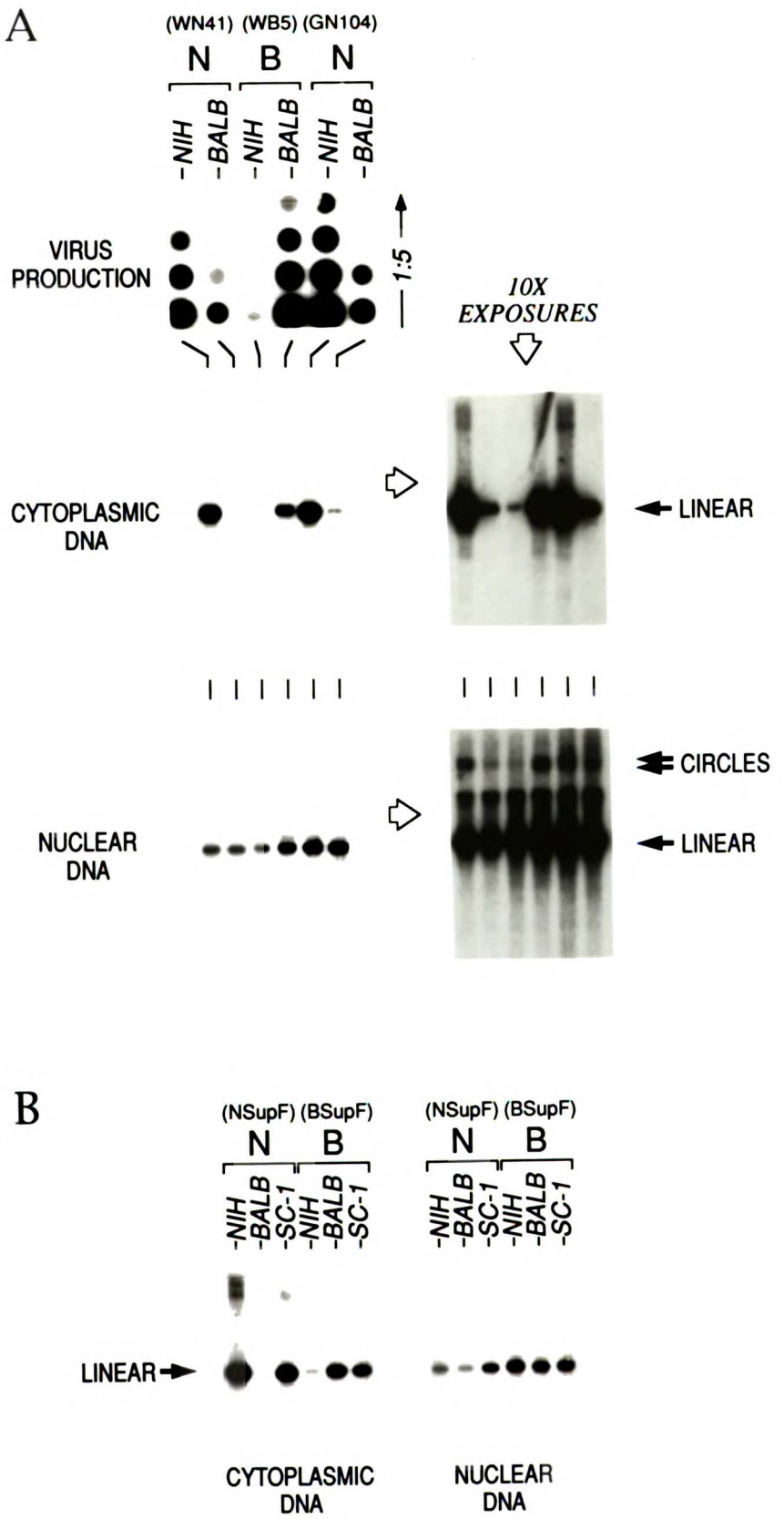
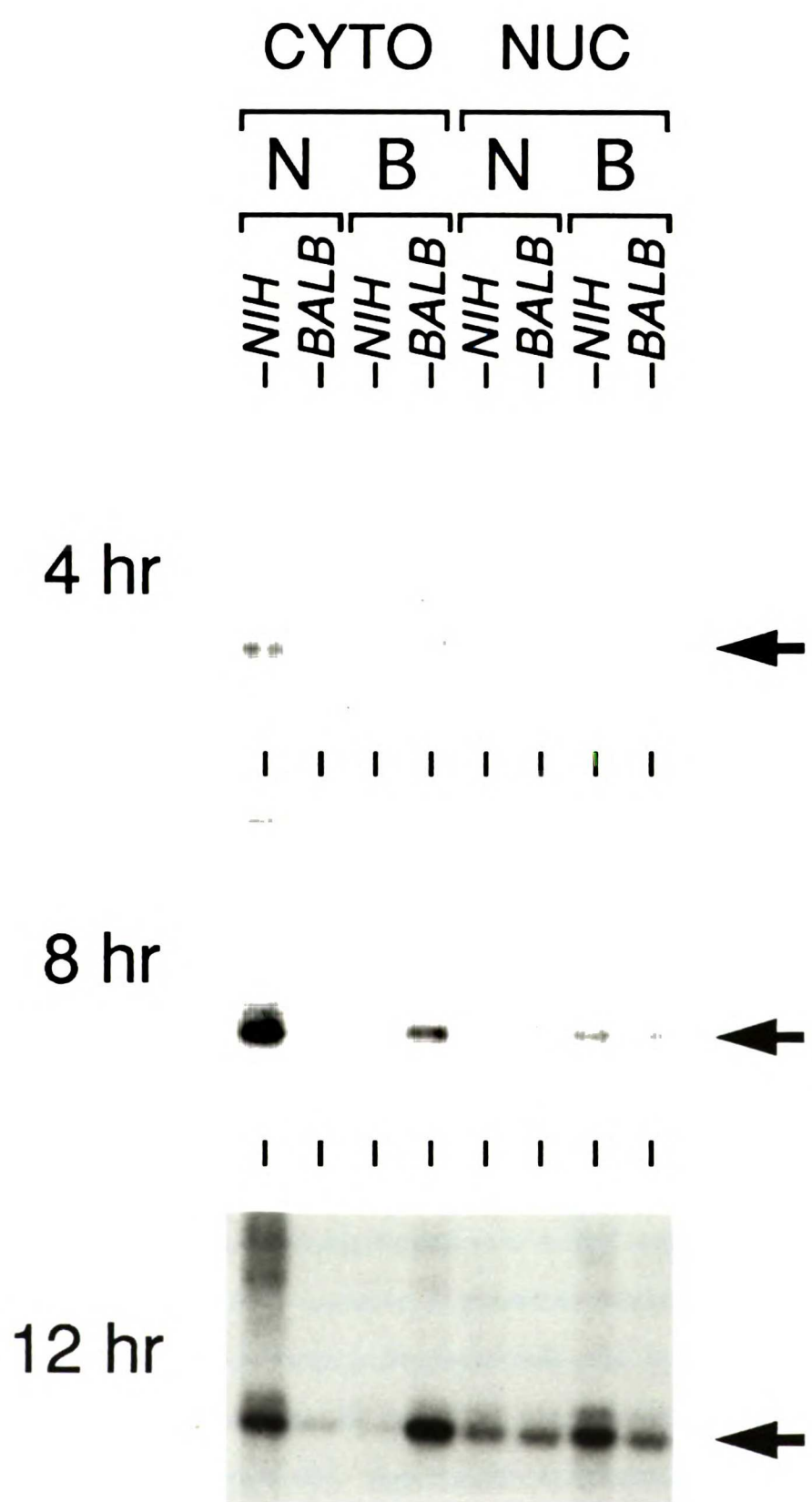

A
(WN41) (WB5) (GN104)
N B N
-NIH -BALB -NIH -BALB -NIH -BALB
VIRUS PRODUCTION
1:5
10X EXPOSURES
CYTOPLASMIC DNA
LINEAR
NUCLEAR DNA
CIRCLES
LINEAR
B
(NSupF) (BSupF)
N B
-NIH -BALB -SC-1 -NIH -BALB -SC-1
(NSupF) (BSupF)
N B
-NIH -BALB -SC-1 -NIH -BALB -SC-1
LINEAR
CYTOPLASMIC DNA
NUCLEAR DNA

Figure 2-3. Time course of appearance of DNAs in cytoplasmic and nuclear extracts during restrictive or permissive infections. Cytoplasmic and nuclear extracts were prepared at the indicated time (4, 8, or 12 hr) after infection of NIH and BALB cells with N-SupF or B-SupF. Virus harvested from parallel plates 32 hr after infection gave reverse transcriptase activities of 10527, 2674, 665, and 14275 cpm for N on NIH, N on BALB, B on NIH, and B on BALB, respectively.

Figure 2-3



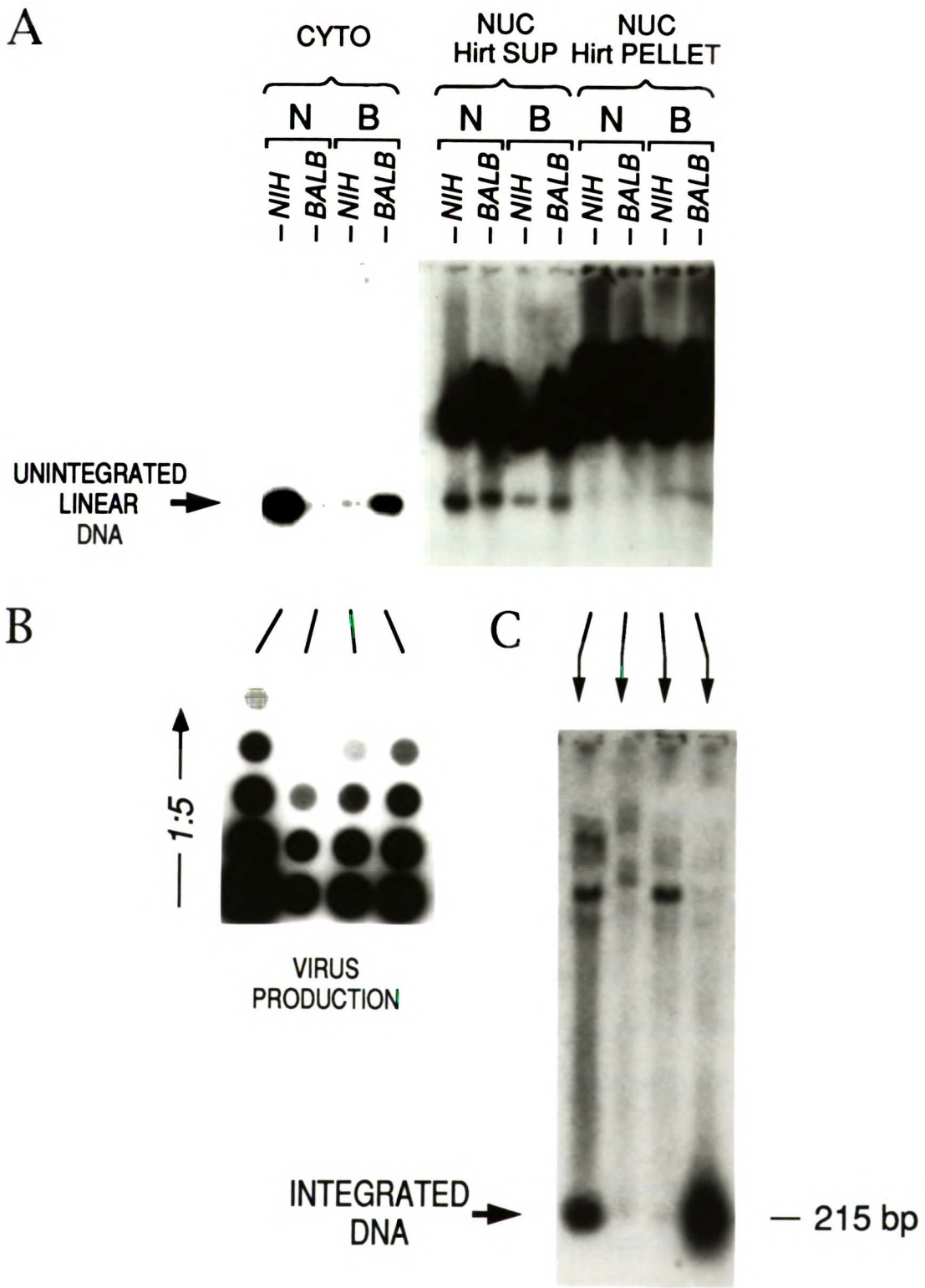
This time course experiment also shows that the effect of *Fv-1* on total DNA, measured by summation of the cytoplasmic and nuclear DNAs, depends on the time after infection. At the earlies time point, when no nuclear DNA is seen, total DNA is decreased in restrictive infections. Eventually, however, at later time points the level of total DNA in restrictive cells is comparable (2-5 fold) to that in permissive cells. A particularly clear example is seen in Figure 2A: comparing GN104 infection of NIH and BALB cells, the nuclear DNA levels in the two infections are roughly equal to the cytoplasmic DNA level in the permissive infection; thus, total DNA is decreased by roughly 2-fold in the restrictive case, even though restriction (as measured by virus production) was roughly 25-fold.

Restriction of integration in vivo

To confirm that the accumulation of integrated proviruses is inhibited in restrictive infections, we took advantage of the foreign insert (the SupF gene) present in our N-SupF and B-SupF viruses. Hybridization to high molecular weight genomic DNA with the SupF insert as a probe avoids the problem of high background hybridization to endogenous MLV sequences that compromised previous measurements of integration in *Fv-1* restriction (Jolicoeur and Baltimore, 1976; Sveda and Soeiro, 1976). The efficiency of integration in vivo was compared for permissive and restrictive infections by first preparing Hirt pellet DNA (Hirt, 1967) from the nuclei of newly-infected cells. This DNA was then cleaved with a restriction enzyme that releases the SupF insert from integrated proviruses, and was hybridized to a SupF probe after electrophoresis and blotting (Figure 4). *Fv-1* restriction of productive infection was clearly accounted for by a block to the accumulation of integrated proviruses, since the 215 bp SupF fragment was released from the Hirt pellet DNA from permissive infections but not restrictive infections (Figure 4C). Hybridization of cytoplasmic, nuclear Hirt supernatant, and nuclear Hirt pellet DNA using an MLV probe showed that the Hirt pellet

Figure 2-4. Restriction is correlated with a decrease in integrated proviruses. NIH and BALB cells infected with N- or B-SupF were harvested after 36 hr. Upon harvest, a cytoplasmic fraction was prepared, and then the nuclear pellet was resuspended and fractionated into Hirt supernatant and pellet fractions. **A.** The uncut DNAs are shown, hybridized with an AKV MLV probe, demonstrate the absence of detectable unintegrated MLV DNA. The strong signal above the position of linear DNA in Hirt supernatant and pellet sample is background hybridization to endogenous MLV-related sequences in the mouse cell chromosomal DNA. The amounts of DNA loaded on this gel were five times greater for the cytoplasmic DNA than for the nuclear Hirt DNAs ($2.5x10^6$ cells/lane for cyto, $5x10^5$ cells/lane for nuc Hirt DNAs). **B.** Virus harvested from parallel plates 63 hr after infection was analyzed by dot-blot hybridization to serial 1:5 dilutions of virion RNA. **C.** The nuclear Hirt pellet DNAs were then digested with Eco RI to release the 215 bp SupF fragment from integrated proviruses, and the resulting Southern blot was hybridized with a labelled SupF probe. Additional blots showed the amounts of integrated proviruses in these permissive Hirt pellet fractions to be roughly comparable to the amounts of unintegrated linear DNA in the Hirt supernatant fractions (not shown).

Figure 2-4


A
CYTO
NUC Hirt SUP
NUC Hirt PELLET
N
B
N
B
N
B
NIH
BALB
NIH
BALB
NIH
BALB
NIH
BALB
UNINTEGRATED LINEAR DNA
B
C
1:5
VIRUS PRODUCTION
INTEGRATED DNA
215 bp

fractions were not significantly contaminated with differential amounts of unintegrated DNA and that they contained relatively equal amounts of genomic DNA (Figure 4A).

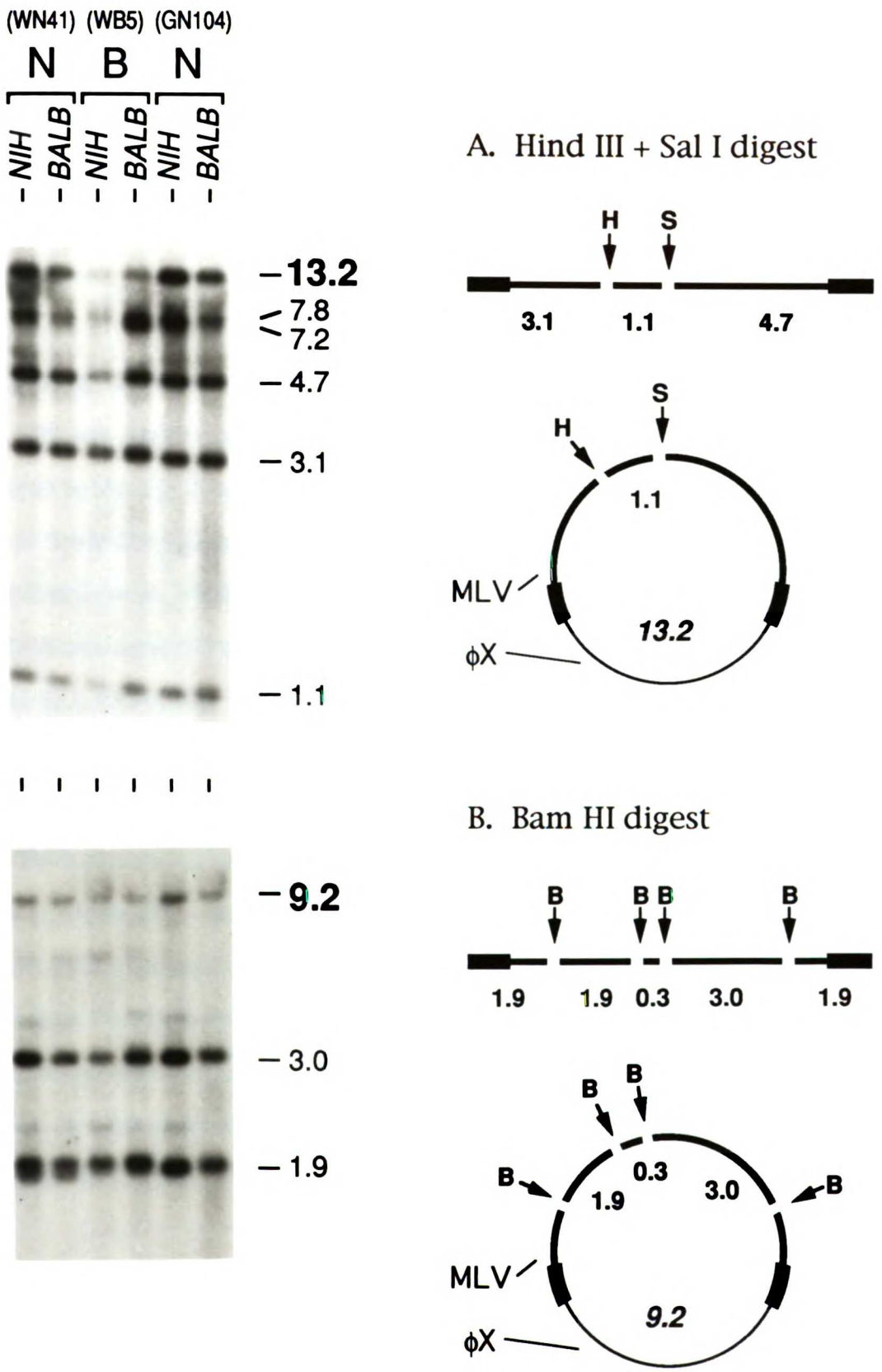
Integration in vitro

The above experiments show that levels of linear viral DNA in the nuclear fractions of infected restrictive cells are nearly normal, and yet integration is inhibited. Therefore, we compared the ability of nuclear extracts from permissive and restrictive infections to integrate their DNA in vitro into an exogenously provided target. In the in vitro integration assay, DNA-containing nucleoprotein complexes present in cell extracts insert linear viral DNA into an added circular DNA target (ϕX174 DNA). Upon restriction enzyme digestion, the integration products give rise to a large new cleavage product, which is visualized by hybridization with an MLV probe after electrophoresis and blotting (Brown et al., 1989; Pryciak et al., 1992). We observed that nuclear extracts from restrictive infections were as competent to integrate their DNA as those from permissive infections (Figure 5), as indicated by the efficiency of generating a 13.2 kb product (Hind III + Sal I digest) or a 9.2 kb product (Bam HI digest). The mild variability in the final amount of product is reflective of the initial amount of viral DNA (as indicated by the 1.1 kb or 3.0 kb bands in the Hind III + Sal I or Bam HI digests, respectively); thus, the specific activities were roughly equivalent for complexes present after infection of permissive and restrictive cells. Control experiments demonstrate that the integration product bands were not present before incubation, or when incubation was in the absence of target DNA or in the presence of EDTA (not shown).

Therefore, the preintegration complexes present in restrictively-infected cells are not irreversibly disabled; instead they can be extracted from the cells in functional form. Furthermore, linear DNA in extracts from restrictive infections is found associated with nucleoprotein complexes of normal size, as measured by sucrose gradient

Figure 2-5. Restriction does not operate during integration in vitro. Nuclear extracts from infections of permissive and restrictive cells, containing similar levels of linear viral DNA (uncleaved DNAs in these same extracts are shown in Figure 2A), were tested for integration activity in vitro. Extracts were incubated with ϕX174 DNA as an integration target, and then analyzed by digestion with Hind III + Sal I (top) or with Bam HI (bottom). Integration of MLV DNA into ϕX174 DNA gives unique 13.2 kb (top) or 9.2 kb (bottom) digestion products (as diagrammed at right), which are seen to be approximately equal for extracts from infections of permissive and restrictive cells. The relative amounts of viral DNA in the extracts are indicated by internal digestion products of 1.1 kb (top) or 3.0 kb (bottom) in size, which are generated from both unintegrated and integrated DNAs. The bands at approximately 7.8 and 7.2 kb in the Hind III + Sal I digest (top) correspond in size to either 1- or 2-LTR circles or intramolecular recombinants ("autointegrants"; Shoemaker et al., 1981; Lee and Coffin, 1990; Farnet and Haseltine, 1991); the absence of a significant 2-LTR circle digestion product in the Bam HI digest (expected size = 3.8 kb) suggests that they are primarily autointegrants.

Figure 2-5


(WN41) (WB5) (GN104)
N B N
- NIH - BALB - NIH - BALB - NIH - BALB
-13.2
7.8
7.2
-4.7
-3.1
-1.1
-9.2
-3.0
-1.9
A. Hind III + Sal I digest
H S
3.1 1.1 4.7
S
H
1.1
MLV
13.2
ϕX
B. Bam HI digest
B B B B
1.9 1.9 0.3 3.0 1.9
B B
B
0.3
1.9
3.0
B
MLV
9.2
ϕX

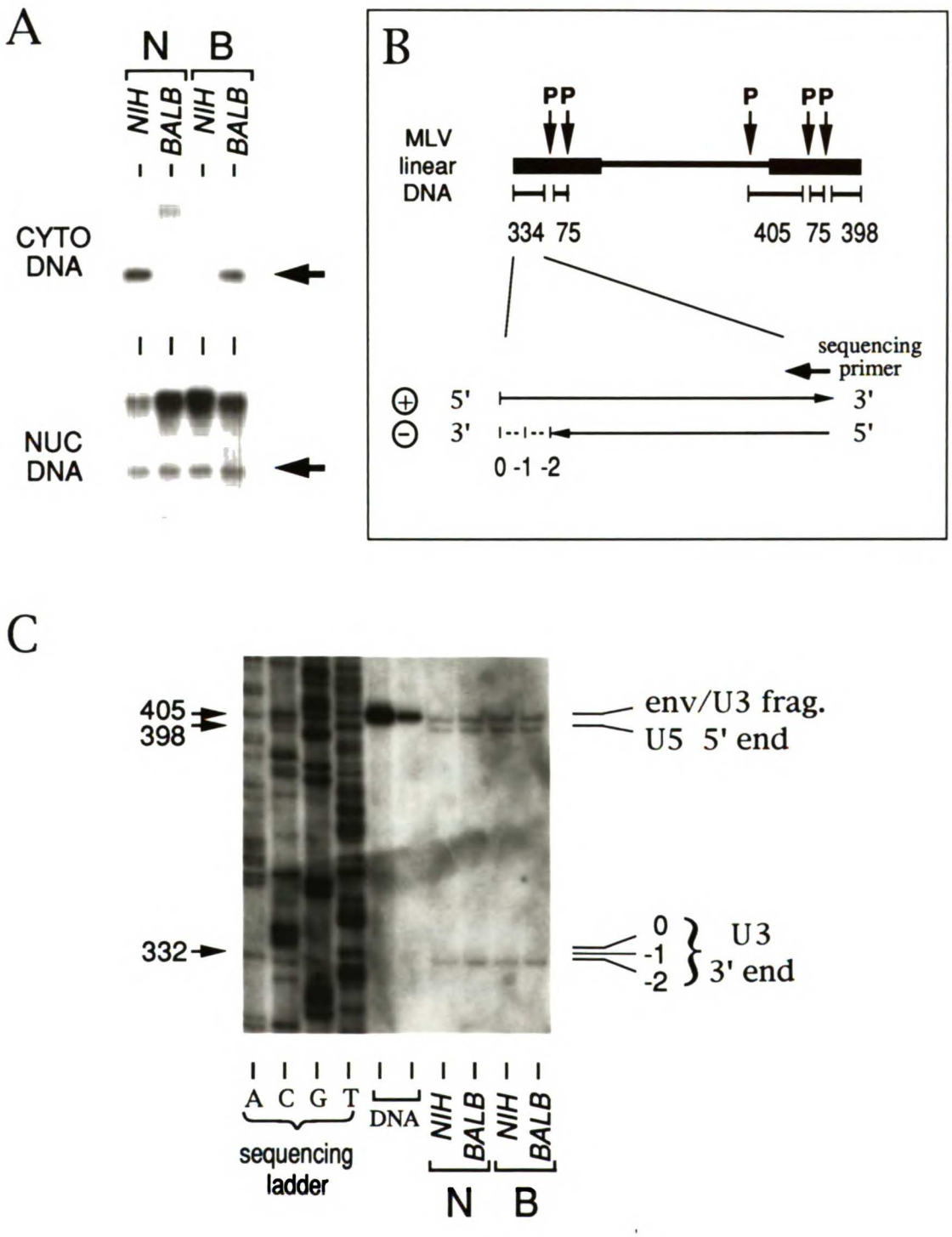
sedimentation (not shown), and normal amounts of CA protein appear in cytoplasmic and nuclear fractions in restrictive infections (not shown). It is possible that some restricting component is normally associated with the integration machinery in vivo but becomes disassociated or inactivated upon preparation of the extracts for in vitro analysis. Alternatively, restriction may not operate by directly inhibiting integration activity.

Processing of linear viral DNA in permissive and restrictive infections

To ask whether other activities ascribed to the viral integrase (IN) are restricted by *Fv-1* in vivo, we tested the ability of IN to remove 2 bases from the 3' ends of linear DNA (Fujiwara and Mizuuchi, 1988; Brown et al., 1989; Roth et al., 1989; Craigie et al., 1990; Katz et al., 1990). The exact 3' terminus of the U3 end of viral DNA was determined by restriction enzyme cleavage at a point close to the end (Figure 6B), followed by denaturing polyacrylamide gel electrophoresis and hybridization with an LTR probe equivalent to plus strand of viral DNA (Fujiwara and Mizuuchi, 1988; Brown et al., 1989). We found that the U3 nuclear DNA in restrictive infections had a normally-processed 3' end, recessed 2 bases from the 5' end (332 base fragment; Figure 6C). Note that this blot also documents the integrity of the 5' end of U5 (398 base fragment). Other experiments indicated that the 3' end was also properly recessed at the U5 end in restrictive infections (not shown). Thus, although integration is blocked in restrictive infections, not all functions of the viral IN product are inhibited--the IN-mediated processing of linear viral DNA occurs normally in restrictive infections.

Figure 2-6. Integrase-dependent processing of linear viral DNA occurs normally in vivo during restrictive infections. **A.** Cytoplasmic and nuclear extracts were prepared 18 hr after infection of NIH and BALB cells with N-ASMF or B-ASMF virus. Virus harvested from parallel plates 36 hr after infection showed the following relative amounts of virus: 100, 10, 4, and 100 for N on NIH, N on BALB, B on NIH, and B on BALB, respectively, as measured by dot-blot hybridization of virion RNAs (not shown) in a manner similar to that shown for Figures 2A and 4. Linear DNA present in cytoplasmic and nuclear fractions is indicated (arrows); hybridization signal above the indicated position of linear DNA is due to fragmented chromosomal DNA contaminating the fractions--this region did not hybridize upon reprobing of the blot with a SupF probe, as did the indicated linear species (not shown). **B.** The inset diagrams the products of Pvu II digestion of linear DNA that are detectable with the probe used, and an enlarged view of the U3, or left, end is shown. Proper cleavage of the 3' end at position "-2" (Fujiwara and Mizuuchi, 1988; Brown et al., 1989; Roth et al., 1989; Craigie et al., 1990) reduces the 334 base fragment of minus strand to 332 bases in length. **C.** Full length linear viral DNA in the nuclear fractions shown in (A) was agarose gel-purified from approximately 10^7 cell equivalents each, digested with Pvu II, denatured and run on a 6% polyacrylamide/7M urea gel (Brown et al., 1989). The gel was electroblotted and the blot hybridized with a plus-strand specific M13 clone containing an Xba I-Xba I circle junction MoMLV LTR fragment as described (Brown et al., 1989). Also run in the same gel were samples of pMoMLV-SupF plasmid DNA that had been digested with Pvu II and Sac I (DNA) and products of sequencing reactions using the M13 DNA described above as template and a primer that has a 5' end equivalent to that formed by the Pvu II digestion (primer "D" in Brown et al., 1989). The blot shows that in all nuclear extracts the U3 3' ends have been properly processed to the 332 base "-2" position; also indicated above are the 398 and 405 base Pvu II digestion products representing the U5 5' end and env/U3 fragments which come from the right LTR, as diagrammed in (B).

Figure 2-6


A
N
B
NIH
BALB
NIH
BALB
CYTO DNA
NUC DNA
B
PP
P
PP
MLV linear DNA
334 75
405 75 398
sequencing primer
5'
3'
3'
5'
0 -1 -2
C
405
398
332
env/U3 frag.
U5 5' end
0
-1
-2
U3 3' end
A C G T
sequencing ladder
DNA
NIH
BALB
NIH
BALB
N
B

DISCUSSION

In this study, we investigate in molecular detail the mechanism of *Fv-1* restriction. We report here: (i) a novel affect of *Fv-1* on the appearance of linear viral DNA in subcellular fractions; (ii) that viral DNA-containing nucleoprotein complexes in extracts from infections of restrictive cells are competent to integrate their DNA in vitro; and (iii) that integrase-mediated processing of the 3' ends of linear viral DNA occurs normally upon infections of restrictive cells.

The effect of *Fv-1* on accumulation of viral DNA

In our experiments we examined DNA levels in cell extracts that had been separated into cytoplasmic and nuclear fractions. The observation that cytoplasmic DNA levels were decreased but nuclear DNA levels were nearly normal in restrictive infections was unexpected, but was extremely reproducible. This pattern of DNA accumulation has not been reported in previous studies of *Fv-1* restriction. Most earlier studies analyzed total DNA; we see only a small inhibitory effect on total DNA in restrictive infections (as measured by summing the cytoplasmic and nuclear DNAs), but a large effect on cytoplasmic DNA that is well-correlated with the degree of restriction. The only study that did fractionate infected cells into cytoplasmic and nuclear fractions found 3- to 6-fold effects in cytoplasmic DNA levels in cases where restriction was 60- to 70-fold (Jolicoeur and Rassart, 1981); differences in fractionation procedures may account for the more drastic effects seen here.

What accounts for the differential effects on cytoplasmic and nuclear DNAs, and how do these effects relate to the restriction mechanism? There are several possible explanations. First, it should be noted that reduction in total viral DNA is insufficient to account for restriction in these experiments, since total DNA (as measured by summing together cytoplasmic and nuclear DNAs) is usually decreased by only 2- to 5-fold, while restriction is usually ≥ 20-fold. The degree of effect on cytoplasmic DNA (usually 10-

to 30-fold), on the other hand, correlates well with the degree of restriction (see Figures 2, 3, 4, 6).

Perhaps the simplest explanation for the DNA phenotype is that DNA synthesis is slowed, but not stopped, upon entry into a restrictive cell. If migration to the nucleus were to occur at a normal rate independent of the completion of DNA synthesis, then the slower DNA synthesis would appear to occur primarily in the later compartment, the nucleus. The linear DNA eventually appearing in the nuclear fraction would nevertheless be blocked from integrating by *Fv-1*. This model might predict a noticeable difference in nuclear DNA amounts between permissive and restrictive infections at the earliest of appearance in that fraction and a compensatory equalization at later times (which was not observed; see Figure 3), but such a prediction is dependent upon the exact relative rates of slowed DNA synthesis and nuclear migration. A potentially related observation was made in quiescent cells infected with spleen necrosis virus (SNV), where, upon release from quiescence, DNA appeared in the nuclear fraction without a prior appearance in the cytoplasm (in contrast to when growing cells were infected with SNV; Fritsch and Temin, 1977).

An alternative, and perhaps extreme, model to explain the DNA phenotype would posit that the DNA in the nuclear fraction does not participate in integration inside the cell and, instead, the DNA in the cytoplasmic fraction is the relevant precursor to integrated proviruses. This would explain the correlation between cytoplasmic DNA levels, and not nuclear DNA levels, with integration. If integration is rapid upon entry into the nuclear compartment, those integration complexes that successfully integrate their DNA might never be detected as unintegrated DNA in the nucleus, in contrast to those that fail to integrate. While it is true that complexes in nuclear extracts are competent to integrate their DNA in vitro this assay may bypass a condition normally prohibiting the unintegrated DNA detected in the nuclear fraction from integrating in

vivo. Little information is available on the precursor-product relationship between linear DNA specifically in the nuclear fraction and integrated proviruses.

The effect of *Fv-1* on integration

Assuming that viral DNA in the nuclear fraction can participate in integration in permissive cells, there must be some mechanism for prevention of integration in vivo. Integration in vitro, however, is not inhibited using extracts of restrictively infected cells. The most trivial resolution to this apparent paradox is that there is an *Fv-1*-dependent apparatus that directly inhibits integration activity in vivo, but is inactivated or is disassociated from the integration complexes upon preparation of the in vitro extracts. Such an explanation would be difficult to prove wrong. An exhaustive survey of extract preparation conditions may uncover a method that preserves restriction activity.

Nevertheless, our observations do limit the potential mechanisms by which *Fv-1* prevents integration in vivo. The ability to isolate fully functional integration complexes from restrictively infected cells (Figure 5) shows that they are not irreversibly or drastically damaged in the cell. Thus, mechanisms that involve large-scale proteolytic cleavage or disassembly of the integration complex components can be disregarded.

In addition, potential mechanisms are further limited by the important observation that the 3' end processing reaction occurs normally during infection of restrictive cells (Figure 6). Thus, the *Fv-1*-dependent host cell apparatus cannot simply be inhibiting all activities of the IN protein in vivo. Instead, IN-mediated 3' end processing occurs normally, and yet integration is prevented. It could potentially be argued the early stage at which this processing occurs (Brown et al., 1989; Roth et al., 1989) precedes the action of *Fv-1* against IN activity at later stages. However, the

cytoplasmic DNA phenotype suggests that the *Fv-1*-dependent action is very early, and that it in fact precedes completion of full-length DNA synthesis.

Thus, in vivo, *Fv-1* may not directly inhibit integration activity. An alternative model is that integration in vivo has more stringent requirements than are present in the in vitro assay, and that it is one of these requirements that is inhibited by *Fv-1* in vivo. One version of this type of model would postulate that the integration machinery in vivo must transit to particular locations within the nucleus in order to come into contact with chromosomal DNA, and that this proper intranuclear localization is prevented in restrictive infections. Possibly relevant in this regard is the phenotype of decreased circle formation in restrictive infections. Some circular DNAs are in fact intramolecular recombinants, or "autointegrants" (Shoemaker et al., 1981), but a significant fraction represent the products of host cell activities (such as ligation and homologous recombination) and are not dependent upon viral integration activities (Shoemaker et al., 1981; Donehower and Varmus, 1984; Schwartzberg et al., 1984; Donehower, 1988; Quinn and Grandgenett, 1988). These host activities could be located in the same intranuclear compartment as chromosomal target DNA, so that mislocalization prevents both integration and host-mediated circle formation. Indeed, fractionation studies support a differential localization or association of circular DNA than linear DNA, and therefore perhaps also of DNA that can and cannot be integrated: in infected cells lysed with non-ionic detergents, linear viral DNA initially present in the nuclear fraction will leak over time into the cytoplasmic fraction, wheras circular DNA is completely resistant to such leakage, and requires physical disruption of the nuclei in order to leak out of the nuclear fraction (unpublished observations).

A more extreme model would propose that the linear DNA seen in our nuclear fraction in restrictive infections is not actually in the nucleus, but is instead attached to the outside of the nucleus or perhaps associated with the cytoskeleton. Here, the inability to migrate into the nucleus would prevent both integration and circle

formation. Biochemical fractionation has been previously observed in some cases to give misleading information about subcellular localization (Herrick et al., 1976; Ellman et al., 1984). In addition, large cytoplasmic structures, such as the cytoskeleton, most likely partition with the nuclear fraction (Cervera et al., 1981; Katze et al., 1989). An attempt to measure appearance of MLV DNA in the nucleus by in situ hybridization (Pinkel et al., 1988) rather than by biochemical fractionation would be helpful in resolving whether DNA ever appears in the nucleus in restrictive infections. Such in situ analyses have recently been successful for studies of MoMLV infection in rat and mouse cells (T. Reynolds, T. Roe, and P. Brown, personal communication).

Along these lines, it is relevant to note that MLV (and other retroviral) infection of permissive cells in the presence of either of the drugs aphidicolin (a cellular DNA synthesis inhibitor) or cycloheximide (a protein synthesis inhibitor) results in a phenotype similar to that of *Fv-1* restriction--namely, normal levels of linear viral DNA are synthesized, but circle formation and integration are inhibited (Yang et al., 1980b; Hagino-Yamagishi et al., 1981; Chinsky and Soeiro, 1982; Hsu and Taylor, 1982). Recent experiments demonstrate that the aphidicolin affect (and by analogy probably also the cycloheximide effect) can be explained by a requirement for the cell to undergo a round of nuclear envelope breakdown and reformation during mitosis in order for nuclear localization (as measured by in situ hybridization) and integration of MLV DNA to take place (T. Roe, T. Reynolds, and P. Brown, personal communication). We have found that in extracts of cells infected with MLV in the presence of aphidicolin or cycloheximide, linear viral DNA appears normally in the nuclear fraction (unpublished observations), even though true nuclear localization is most likely blocked in these infections; thus, the biochemical fractionation may indeed be misleading. Furthermore, integration of DNA synthesized in such drug-treated cells occurs normally in vitro using extracts prepared from these cells, even though circle formation and integration in vivo

are prevented (unpublished observations); these observations mirror those of *Fv-1* restriction.

Thus, the exact molecular mechanisms responsible for *Fv-1* restriction remain unidentified. The present report, however, offers important clues which help provide a framework for the design of further experiments.

MATERIALS AND METHODS

Plasmids, viruses, and cells

Plasmids pN20-7 and pB16-5 (DesGroseillers and Jolicoeur, 1983) were obtained from P. Jolicoeur (Montreal, Canada), and pWN41, pWB-5, and pGN104 (Boone et al., 1983) were obtained from W. K. Yang (Oak Ridge, Tennessee). To construct pN-SupF and pB-SupF, the viral Hind III fragments from pN20-7 and pB16-5 were first cloned into the Hind III site of a Pst I site-deleted derivative of the vector pSP65. A SupF-containing Pvu II fragment of the plasmid pVSU-II (Lobel and Goff, 1984), which contains the SupF gene flanked successively by Eco RI and Pvu II sites, was then cloned into the unique Pst I sites in the viral LTRs after first blunting those sites using T4 polymerase; the products were pN-SupF and pB-SupF. The plasmids pN-MoF and pB-MoF were constructed by site-directed mutagenesis of the parental plasmid pMoMLV-SupF (Lobel et al., 1985; Brown et al., 1987), changing the codons at positions 109, 110 of CA in *gag* (Gln, Ala) to Gln, Arg (pN-MoF) or Thr, Glu (pB-MoF), which are the tropism-determining amino acids in naturally-occurring N- and B-tropic viruses (DesGroseillers and Jolicoeur, 1983). The plasmids pN-ASMF and pB-ASMF were constructed by replacing the Aat II to Sal I fragment in pMoMLV-SupF with the analogous fragment from pN20-7 and pB16-5, respectively.

Virus producing cell lines were established by first transfecting SC-1 cells with these plasmid DNAs encoding the viral genomes (which for all except the non-MoMLV-SupF derivatives was first cleaved with Hind III and self-ligated), and then cocultivating pooled transfected cells with fresh SC-1 cells. Virus used for infections was always harvested fresh from these producer lines.

The cell lines NIH-3T3 (*Fv-1*$^{n/n}$), BALB/3T3 (*Fv-1*$^{b/b}$), and SC-1 (phenotypically Fv-1^{-}), and the virus-producing SC-1 derivatives, were all maintained in DME-H16 medium supplemented with 10 % fetal calf serum.

Virus Infections

Cells were plated 48 hr before infections at densities of $2x10^6$ per 100mm dish for target cells and $1x10^5$ per 100mm dish for virus producer cells. At 12 hr before infection, growth medium was replaced on the virus producer cells. Cells were infected in the presence of 8 μg/ml polybrene in a volume of 4 ml of virus plus growth medium, with each plate of target cells receiving virus from one-fourth plate of virus producer cells (for naturally-occurring AKV N- and B-tropic viruses and their derivatives; indicated at top of Figure 1 in black or white) or from one-tenth plate of virus producer cells (for MoMLV-SupF derivatives; indicated at bottom of Figure 1 as shaded). When virus production from the newly-infected cells was to be assayed, medium was removed from the cells 12 hr after infection, cells were rinsed once with phosphate buffered saline solution, and fresh growth medium was added; virus was usually harvested 24 to 36 hr after infection, except in one case (63 hr; Figure 4).

Extract preparations, integration reactions, and analysis of nucleic acids

Infected cells were harvested by trypsinization, and cytoplasmic and nuclear extracts were prepared by digitonin lysis and ball-bearing homogenization as described (Brown et al., 1987), except that 125 μl of extract was prepared for each plate of cells rather than 250 μl. Separation of the cytoplasmic and nuclear fractions was performed no longer than five minutes after digitonin lysis. Extracts were frozen in liquid nitrogen and stored at -80° C, and were thawed immediately before integration assays or preparation of nucleic acids.

Integration reactions were performed essentially as described (Brown et al., 1987; Brown et al., 1989; Pryciak et al., 1992) using 500 ng of ϕX174 DNA as an integration target for 50 μl of integration extract. Nucleic acids were prepared from extracts or integration reactions as described (Brown et al., 1987).

DNAs were analyzed either with or without restriction enzyme digestion by electrophoresis in 0.7 % agarose gel, blotting to a Hybond-N nylon membrane (Amersham), UV cross-linking, and hybridization with a nick-translated probe. The probe used was either a mixture of AKV viral plasmid clones (pN20-7 and pB16-5, or pWN41, pWB5, and pGN104) or of MoMLV-derived plasmid clones (pN-ASMF and pB-ASMF), or a SupF Pvu II fragment from the plasmid pVSU-II.

Virion RNAs were analyzed by pelleting 450 µl virus for 5 min. at 30 p.s.i. in a Beckman Airfuge. The supernatant was removed by aspiration, and the pellet resuspended in 100 µl of 10 mM TRIS, pH 7.4, 10 mM EDTA, 200 mM NaCl, 0.5 % SDS, followed by extraction twice with phenol/chloroform and then twice with ether. The sample was then heated to 65° C for 2 min., and mixed with 100 µl of 20x SSC. This mixture was subjected to serial five-fold dilutions into 10x SSC, and the dilutions were then blotted onto Hybond-N nylon membrane by use of a dot-blot apparatus; the blot UV crosslinked and hybridized as for DNA analysis.

REFERENCES

Boone, L. R., Myer, F. E., Yang, D. M., Ou, C.-Y., Koh, C. K., Roberson, L. E., Tennant, R. W. and Yang, W. K. (1983). Reversal of *Fv-1* host range by in vitro restriction endonuclease fragment exchange between molecular clones of N-tropic and B-tropic murine leukemia virus genomes. J. Virol. *48*, 110-119.

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

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

Bushman, F. D., Fujiwara, T. and Craigie, R. (1990). Retroviral DNA integration directed by HIV integration protein in vitro. Science *249*, 1555-1558.

Cervera, M., Dreyfuss, G. and Penman, S. (1981). Messenger RNA is translated when associated with the cytoskeletal framework in normal and VSV-infected HeLa cells. Cell *23*, 113-120.

Chinsky, J. and Soeiro, R. (1981). *Fv-1* host restriction of Friend leukemia virus: analysis of unintegrated proviral DNA. J. Virol. *40*, 45-55.

Chinsky, J. and Soeiro, R. (1982). Studies with aphidicolin on the *Fv-1* host restriction of Friend murine leukemia virus. J. Virol. *43*, 182-190.

Chinsky, J., Soeiro, R. and Kopchick, J. (1984). *Fv-1* host cell restriction of Friend leukemia virus: microinjection of unintegrated viral DNA. J. Virol. *50*, 271-274.

Craigie, R., Fujiwara, T. and Bushman, F. (1990). The IN protein of moloney murine leukemia virus processes the viral DNA ends and accomplishes their integration in vitro. Cell *62*, 829-837.

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

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.

Donehower, L. A. 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 *81*, 6461-6465.

Duttagupta, S. and Soeiro, R. (1981). Host restriction of Friend leukemia virus: gag proteins of host range variants. Proc. Natl. Acad. Sci. USA *78*, 2320-2324.

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

Ellman, M., Bikel, I., Figge, J., Roberts, T., Schlossman, R. and Livingston, D. M. (1984). Localization of the simian virus 40 small t antigen in the nucleus and cytoplasm of monkey and mouse cells. J. Virol. *50*, 623-628.

Farnet, C. M. and Haseltine, W. A. (1991). Circularization of human immunodeficiency virus type 1 DNA in vitro. J. Virol. *65*, 6942-6952.

Fritsch, E. F. and Temin, H. (1977). Inhibition of viral DNA synthesis in stationary chicken embryo fibroblasts infected with avian retroviruses. J. Virol. *24*, 461-469.

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

Hagino-Yamagishi, K., Kano, K. and Mano, Y. (1981). Effects of aphidicolin on retrovirus DNA synthesis in vivo. Biochem. Biophys. Res. Comm. *102*, 1372-1378.

Herrick, G., Spear, B. B. and Veomett, G. (1976). Intracellular localization of mouse DNA polymerase α. Proc. Natl. Acad. Sci. USA *73*, 1136-1139.

Hirt, B. (1967). Selective extraction of polyomavirus DNA from infected mouse cell cultures. J. Mol. Biol. *26*, 365-369.

Hopkins, N., Schindler, J. and Hynes, R. (1977). Six NB-tropic murine leukemia viruses derived from a B-tropic virus of BALB/c have altered p30. J. Virol. *21*, 309-318.

Hsu, I.-C., Yang, W. K., Tennant, R. W. and Brown, A. (1978). Transfection of *Fv-1* permissive and restrictive mouse cells with integrated DNA of murine leukemia viruses. Proc. Natl. Acad. Sci. USA *75*, 1451-1455.

Hsu, T. W. and Taylor, J. M. (1982). Effect of aphidicolin on avian sarcoma virus replication. J. Virol. *44*, 493-498.

Huang, A. S., Besmer, P., Chu, L. and Baltimore, D. (1973). Growth of pseudotypes of vesicular stomatitis virus with N-tropic murine leukemia virus coats in cells resistant to N-tropic viruses. J. Virol. *12*, 659-662.

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

Jolicoeur, P. and Baltimore, D. (1976). Effect of *Fv-1* gene product on proviral DNA formation and integration in cells infected with murine leukemia viruses. Proc. Natl. Acad. Sci. USA *73*, 2236-2240.

Jolicoeur, P. and Rassart, E. (1980). Effect of Fv-1 gene product on synthesis of linear and supercoiled viral DNA in cells infected with murine leukemia virus. J. Virol. *33*, 183-195.

Jolicoeur, P. and Rassart, E. (1981). Fate of unintegrated viral DNA in *Fv-1* permissive and resistant mouse cells infected with murine leukemia virus. J. Virol. *37*, 609-619.

Kashmiri, S. V. S., Rein, A., Bassin, R. H., Gerwin, B. I. and Gisselbrecht, S. (1977). Donation of N- or B-tropic phenotype to NB-tropic murine leukemia virus during mixed infections. J. Virol. *22*, 626-633.

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

Katze, M. G., Lara, J. and Wambach, M. (1989). Nontranslated cellular mRNAs are associated with the cytoskeletal framework in influenza virus or adenovirus infected cells. Virology *169*, 312-322.

Kozak, C. (1985). Analysis of wild-derived mice for *Fv-1* and *Fv-2* murine leukemia virus restriction loci: a novel wild mouse *Fv-1* allele responsible for lack of host range restriction. J. Virol. *55*, 281-285.

Lee, Y. M. H. and Coffin, J. M. (1990). Efficient autointegration of avian retrovirus DNA in vitro. J. Virol. *64*, 5958-5965.

Lieber, M., Sherr, C., Potter, M. and Todaro, G. (1975). Isolation of type-C viruses from the Asian feral mouse *Mus musculus molossinus*. Int. J. Cancer *15*, 211-220.

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

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

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

Pinkel, D., Landegent, J., Collins, C., Fuscoe, J., Segraves, R., Lucas, J. and Gray, J. (1988). Fluorescence in situ hybridization with human chromosome-specific libraries: detection of trisomy 21 and translocations of chromosome 4. Proc. Natl. Acad. Sci. USA *85*, 9138-9142.

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

Quinn, T. P. and Grandgenett, D. P. (1988). Genetic evidence that the avian retrovirus DNA endonuclease domain of *pol* is necessary for viral integration. J. Virol. *62*, 2307-2312.

Rein, A., Kashmiri, S. V. S., Bassin, R. H., Gerwin, B. I. and Duran-Troise, G. (1976). Phenotypic mixing between N- and B-tropic murine leukemia viruses: infections particles with dual sensitivity to *Fv-1* restriction. Cell *7*, 373-379.

Roth, M. J., Schwartzberg, P. L. and Goff, S. P. (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.

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.

Shoemaker, C., Hoffmann, J., Goff, S. P. and Baltimore, D. (1981). Intramolecular integration within moloney murine leukemia virus DNA. J. Virol. *40*, 164-172.

Sveda, M. M. and Soeiro, R. (1976). Host restriction of Friend leukemia virus: synthesis and integration of the provirus. Proc. Natl. Acad. Sci. USA *73*, 2356-2360.

Voytek, P. and Kozak, C. (1988). HoMuLV: a novel pathogenic ecotropic virus isolated from the European mouse, *Mus hortulanus*. Virology *165*, 469-475.

Yang, W. K., Boone, L. R., Tennant, R. W. and Brown, A. (1983). Restriction of murine leukemia viruses by *Fv-1*: a model for studying host genetic control of retroviral gene movement and leukemogenesis. Prog. Nuc. Acid Res. *29*, 175-192.

Yang, W. K., Kiggans, J. O., Yang, D.-M., Ou, C.-y., Tennant, R. W., Brown, A. and Bassin, R. H. (1980a). Synthesis and circularization of N- and B-tropic retroviral DNA in *Fv-1* permissive and restrictive mouse cells. Proc. Natl. Acad. Sci. USA *77*, 2994-2998.

Yang, W. K., Yang, D.-M. and Kiggans, J. O. (1980b). Covalently closed circular DNAs of murine type C retrovirus: depressed formation in cells treated with cycloheximide early after infection. J. Virol. *36*, 181-186.

CHAPTER THREE

RETROVIRAL INTEGRATION INTO MINICHROMOSOMES IN VITRO

The EMBO Journal vol.11 no.1 pp.291 - 303, 1992

Retroviral integration into minichromosomes *in vitro*

Peter M.Pryciak[1], Anita Sil[1] and Harold E.Varmus[1,2]

Departments of [1]Biochemistry & Biophysics and [2]Microbiology & Immunology, University of California San Francisco, San Francisco, CA 94143-0502, USA

Communicated by P.Berg

We describe here the use of chromatin as a target for retroviral integration *in vitro*. Extracts of cells newly infected with murine leukemia virus (MLV) provided the source of integration activity, and yeast TRP1ARS1 and SV40 minichromosomes served as simple models for chromatin. Both minichromosomes were used as targets for integration, with efficiencies comparable with that of naked DNA. In addition, under some reaction conditions the minichromosomes behaved as if they were used preferentially over naked DNAs in the same reaction. Mapping of integration sites by cloning and sequencing recombinants revealed that the integration machinery does not display a preference for nucleosome-free, nuclease-sensitive regions. The distributions of integration sites in TRP1ARS1 minichromosomes and a naked DNA counterpart were grossly similar, but in a detailed analysis the distribution in minichromosomes was found to be significantly more ordered: the sites displayed a periodic spacing of ~10 bp, many sites sustained multiple insertions and there was sequence bias at the target sites. These results are in accord with a model in which the integration machinery has preferential access to the exposed face of the nucleosomal DNA helix. The population of potential sites in chromatin therefore becomes more limited, in a manner dictated by the rotational orientation of the DNA sequence around the nucleosome core, and those sites are used more frequently than in naked DNA.

Key words: chromatin/minichromosomes/MLV/nucleosome position/retroviral integration

Introduction

Not all DNA sequences in the cellular genome are in the same state or environment. For example, it is well established that DNA in eukaryotic cells is organized into nucleosomes and various higher degrees of chromatin packaging (for review, see Pederson *et al.*, 1986a). Such differential packaging of chromatin has often been postulated to serve both organizational and regulatory roles. Evidence suggests that different chromatin structures correlate with different physiological states, such as transcriptionally active versus inactive DNA (for review, see Gross and Garrard, 1988). What is the direct effect of chromatin on proteins that need to have access to cellular DNA? Particularly in the case of transcriptional initiation, it has been suggested that nucleosomes can participate in inhibitory regulation, by sterically restricting access of control sequences to transcription factors (for review, see Grunstein, 1990). Whether all DNA based activities are necessarily inhibited by chromatin is less clear; processive phenomena like DNA replication and transcriptional elongation appear not to be inhibited by nucleosomes (Lorch *et al.*, 1987; Bonne-Andrea *et al.*, 1990), although the function of a replication origin can be inhibited by nucleosomes *in vivo* (Simpson, 1990).

Retroviruses integrate a DNA copy of their genome into cellular DNA as an essential part of their life cycle (for review, see Varmus and Brown, 1989). Thus, retroviral integration can potentially provide a paradigm for understanding how various cellular functions are affected by the packaging of DNA into chromatin. But the role of chromatin structure in the choice of integration site has seldom been studied. While it is clear that integration can occur at many positions in the cellular genome and that little sequence preference is displayed (Varmus and Brown, 1989; Sandmeyer *et al.*, 1990), some studies suggest that the choice of integration site is not completely random. It has been observed that integration sites tend to map in or near transcriptionally active regions and nuclease-sensitive regions of chromatin (Vijaya *et al.*, 1986; Rohdewohld, 1987; Scherdin *et al.*, 1990; Mooslehner *et al.*, 1990). In addition, the frequency at which gene expression can be interrupted by retroviral integration can differ markedly from expectation based on random insertion (King *et al.*, 1985). In perhaps the most compelling example, Rous sarcoma virus (RSV) DNA has been observed to integrate at an unusually high frequency into certain preferred regions of the chicken genome, and within those regions insertions tend to occur into the exact same site (Shih *et al.*, 1988). Other studies have argued that most of the genome is available for integration (Reddy *et al.*, 1991).

Few attempts have been made to study these issues *in vitro*. *In vitro* integration assays generally use simple, naked DNA targets (Brown *et al.*, 1987; Craigie *et al.*, 1990; Katz *et al.*, 1990) and are therefore poorly suited to address the effects of complex physiological changes such as chromatin packaging, replication, or transcription on integration. We have begun to address such issues by modifying our previous *in vitro* integration assay to include chromatin targets. As simple chromatin models to serve as targets for integration *in vitro* we chose minichromosomes based upon a yeast plasmid and a mammalian viral genome. Because previous *in vivo* studies suggested that integration sites correlate with nuclease-hypersensitive regions, we used minichromosomes with both nucleosomal and nuclease-sensitive, nucleosome-free regions; this allowed us to design the *in vitro* experiments to test the effect of structural accessibility on choice of integration sites. Using such a strategy, we have been able to show that chromatin can serve as an integration target *in vitro*, that integration is not grossly inhibited by the presence of nucleosomes on the target DNA, and that

the availability of target sites is restricted, favoring some sites over others, when DNA is wrapped around nucleosomal cores.

Results

Characterization of minichromosome targets

We chose two types of well-characterized, relatively small minichromosomes (MCs) as simple models to serve as targets for retroviral integration into chromatin: the TRP1ARS1 plasmid (TA) from yeast, and the SV40 genome (SV) from acutely infected monkey cells (Figure 1). When isolated from yeast cells, the TA MC contains seven nucleosomes precisely arranged on structurally and functionally distinct regions of the 1.5 kb DNA molecule (Thoma *et al.*, 1984; Pederson *et al.*, 1986b; Thoma and Simpson, 1985). Two nucleosomal regions, three and four nucleosomes in size, correspond to untranscribed and transcribed regions, respectively, and they are separated by two nucleosome-free regions which map with the transcriptional control region and the origin of replication. The more heterogeneous SV MCs contain 20–27 nucleosomes distributed on 5.2 kb DNA molecules (Shelton *et al.*, 1980; Sogo *et al.*, 1986; Ambrose *et al.*, 1990). A nuclease-sensitive, nucleosome-free region is often present in the region encompassing the origin of replication and transcriptional control region for the two divergent transcription units (Varshavsky *et al.*, 1978; Saragosti *et al.*, 1980; Ambrose *et al.*, 1986).

These MCs were purified to varying extents, as described in Materials and methods. Because of the nature of the integration assays (described below), the MCs do not have to be highly purified. However, the MC preparation should be free of significant amounts of other DNA, such as genomic or mitochondrial DNA, since these will compete with MCs as integration targets: the MCs should be sufficiently concentrated to serve as targets for the integration machinery; and the vast bulk of the MCs should be present as chromatin, rather than naked DNA from disassembled MCs, to ensure that any integration events will occur into MCs *per se*.

Sedimentation in sucrose gradients demonstrated that the bulk of the DNA in the preparations was present as MCs, clearly sedimenting more rapidly than naked DNAs in the same gradients (Figure 2A and B). Digestion of TA MCs with micrococcal nuclease revealed the expected nucleosomal ladder pattern, with protected regions of ~150–200 bp (Figure 2C), whereas naked DNA was digested into randomly sized fragments and rendered undetectable at relatively low concentrations of nuclease. Hybridization of the digestion products of TA MCs with a probe corresponding to only a small region of the DNA molecule (*Eco*RI–*Xba*I, shown; others, not shown) revealed an unequal pattern consistent with the arrangement of nucleosomes determined by the more extensive analysis of Thoma *et al.* (1984; see Figure 1). Digestion of SV MCs with DNase I showed a peak of cleavage at the expected hypersensitive region located near the origin (Figure 2D), generating fragments of 2.5–2.7 kb after digestion with *Bam*HI, in a manner consistent with many earlier studies (Varshavsky *et al.*, 1978; Saragosti *et al.*, 1980; Ambrose *et al.*, 1986).

Minichromosomes as integration targets

In the experiments described here, the source of viral integration activity was extracts of cells newly infected with MLV; these extracts contain a large viral nucleoprotein complex ready to integrate linear viral DNA into an exogenous target (Brown *et al.*, 1987; Bowerman *et al.*, 1989). The product of the integration reaction can be detected by restriction enzyme digestion and Southern blot analysis (Figure 3A). Enzymes that cleave the viral DNA but not the target DNAs (e.g. *Bst*EII) generate molecules in which full-length target DNA is attached to the ends of viral DNA. Since the size of the crucial digestion product depends upon the size of the target, targets of different sizes can be assayed in the same reaction to assess relative use.

An example of this assay, using MCs and naked, heterologous DNAs as targets present separately or together in the same reaction, demonstrates integration into the TA or SV MCs (Figure 3B): the large fragments of different sizes (5.6, 7.0, 9.3 or 9.5 kb), indicative of integration, reflect the presence of differently sized targets. However, this experiment does not rigorously address how well the MCs were used compared with the naked DNA targets, or whether the MCs *per se*, rather than naked DNA from disassembled MCs, were serving as the integration target. To address these issues, we first examined the effect of target concentration upon integration efficiency (Figure 3C). Integration was sensitive to the amount of target, whether the target was naked φX174 DNA or SV40 MCs. More importantly, a comparison of the two targets shows that the amount of integration product is similar at similar concentrations of target. Thus, we cannot ascribe the MC integration products to use of a small fraction of naked DNA in the MC preparation.

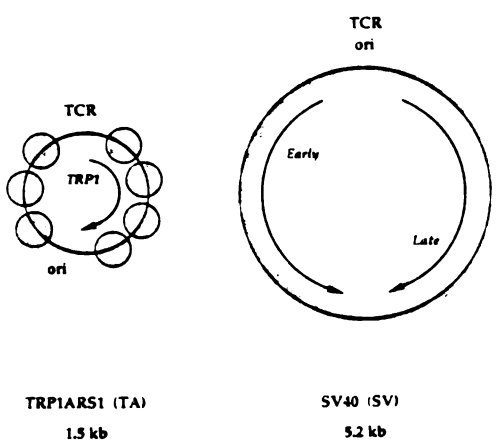


Fig. 1. Schematic representation of minichromosome targets. Nucleosomal and nucleosome-free regions, corresponding to transcriptional control regions (TCR) and origins of replication (ori), are indicated along with transcription units (curved arrows). The 1453 bp TRP1ARS1 (TA) minichromosome (left) is drawn with well defined, solid bordered nucleosomes to indicate their precise positioning into two regions separated by nuclease-sensitive, nucleosome-free regions (Thoma *et al.*, 1984; Pederson *et al.*, 1986). The 5243 bp SV40 (SV) minichromosome (right) is drawn with non-bordered nucleosomes to indicate their imprecise positioning (Ambrose *et al.*, 1990), in an array that is often punctuated by a nuclease-sensitive, nucleosome-free region near ori (Varshavsky *et al.*, 1978; Saragosti *et al.*, 1980; see discussion in Ambrose *et al.*, 1986).

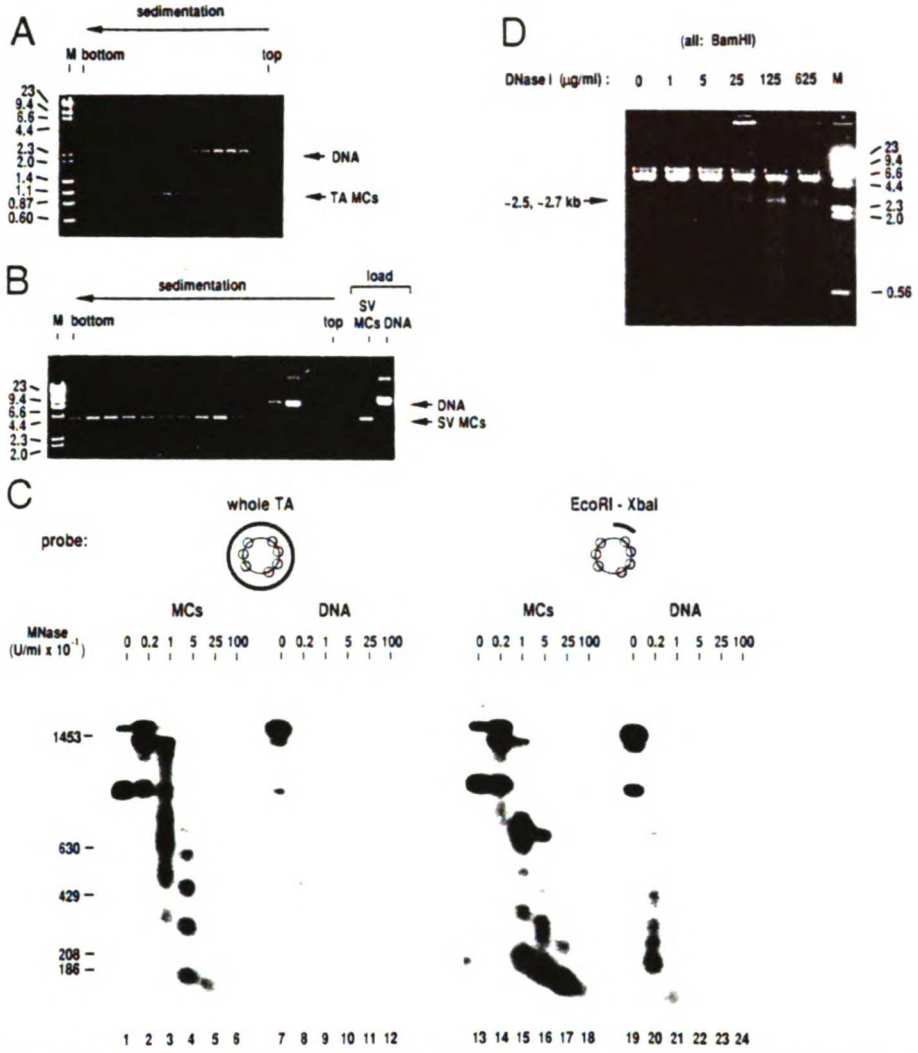


Fig. 2. Characterization of minichromosome targets. All analyses were performed with the same MC preparations used for the subsequent integration assays. **A.** Sedimentation of TA MCs mixed with a naked DNA plasmid (pTA-R; 4.2 kb) in a 10–30% sucrose gradient, for 3 h at 40 000 r.p.m. (Beckman SW50.1 rotor). After sedimentation, 16 equal volume fractions were collected from the bottom of the gradient and DNA prepared from these fractions was then separated on an 0.8% agarose gel and visualized by ethidium bromide staining. Lane M contains molecular weight standards of λ/*Hin*dIII and φX174/*Hae*III digests. **B.** Sedimentation of SV MCs mixed with a naked DNA plasmid (pBS-SVR; 8.2 kb) in a 5–30% sucrose gradient for 2 h at 40 000 r.p.m. (Beckman SW41 rotor). After sedimentation, 15 equal volume fractions were collected from the bottom of the gradient and DNA prepared from these fractions was then separated on an 0.8% agarose gel and visualized by ethidium bromide staining. Also included in the gel are samples of the SV MCs and plasmid DNA before loading on the gradient (load), and molecular weight standards of a λ/*Hin*dIII digest (M). The SV MCs show two peaks, probably corresponding to a 75S MC peak and a 180S 'previrion' peak (Boyce *et al.*, 1982). **C.** Mapping chromatin structure of TA MCs. Micrococcal nuclease (MNase) digestion of TA MCs and DNA (from deproteinized MCs) with the indicated MNase concentration was for 5 min at 37°C in the presence of 5 mM $CaCl_2$. Digestion products were separated on parallel 0.8% gels which were then blotted and hybridized with probes corresponding to the whole TA molecule (left, lanes 1–12) or to a small region between the *Eco*RI and *Xba*I sites (right, lanes 13–24), as schematically depicted above. Molecular sizes are indicated on the left, as determined by the products of restriction enzyme digestion of TA DNA run in the same gel (not shown). The greater protection from nuclease and the 150–200 bp ladder of protected fragments observed in the MC lanes are indicative of nucleosomal DNA. The strong enrichment for the tetranucleosome sized product when probing with the *Eco*RI–*Xba*I fragment is indicative of preferential cleavage at the two nucleosome-free regions, releasing products of three and four nucleosomes in size, of which only the latter will hybridize with the probe used. This experiment, and analysis with other probes and cleavage of MNase digestion products with restriction enzymes (not shown), is consistent with the nucleosome structure shown in Figure 1, as determined by Thoma *et al.* (1984). **D.** Mapping nuclease-sensitive region of SV MCs. DNase I digestion of SV MCs was performed with the indicated concentrations for 5 min at 37°C in the presence of 5 mM $MgCl_2$. DNase I digestion products were digested with *Bam*HI and then separated on a 0.8% agarose gel, along with λ/*Hin*dIII size markers (M), and visualized by ethidium bromide staining. *Bam*HI cleaves SV40 DNA (5.2 kb) at position 2533, and hence the appearance of DNase I-dependent products in the 2.5–2.7 kb size range (arrow) is indicative of preferential cleavage of the SV MCs near position 0, corresponding to the ori region indicated as often being nuclease-sensitive and nucleosome-free in Figure 1.

We also used direct competition assays to compare the behavior of MCs and naked DNA as integration targets (Figure 4). In these assays as many as three different targets of three different sizes were present in the same reaction, and their relative use was again measured by the different sizes of their products. Increasing amounts of a naked DNA (#2) were added as competitor to reactions containing another naked DNA (#1) and TA MCs; the relative use of #1 DNA and TA MCs matched very closely throughout the range of competition (Figure 4A). This provides additional evidence that the primary relevant target in the MC sample was the MCs *per se* since, in the presence of competitor, the level of integration into a small fraction of contaminating naked TA DNA from disassembled MCs would be expected to reflect its relative proportion of the total available DNA target.

A surprising but useful difference between MC and naked DNA targets was revealed when the integration reactions were carried out in the presence of polyethylene glycol (PEG; Figure 4B). Under these conditions, excess naked DNA (#2) did not interfere with the use of TA MCs as targets, even though it competed effectively with another naked DNA (#1) in the same reaction. Thus, in the presence of PEG, MCs behaved differently from naked DNA, again arguing that the vast majority of integration events into the MC target must have been into the MCs *per se*. Analogous experiments revealed similar differences in behavior between SV MCs and naked DNA in the presence of PEG (Figure 4C). The differences were dependent upon maintenance of the MC target as chromatin, since deproteinized SV MCs (SV DNA) behaved like other naked DNAs in response to increasing amounts of competitor DNA (Figure 4D). While we have made use of these differences, we have not elucidated their underlying mechanisms; nor have we attempted to assign values for absolute or relative target efficiencies, since they depend upon reaction

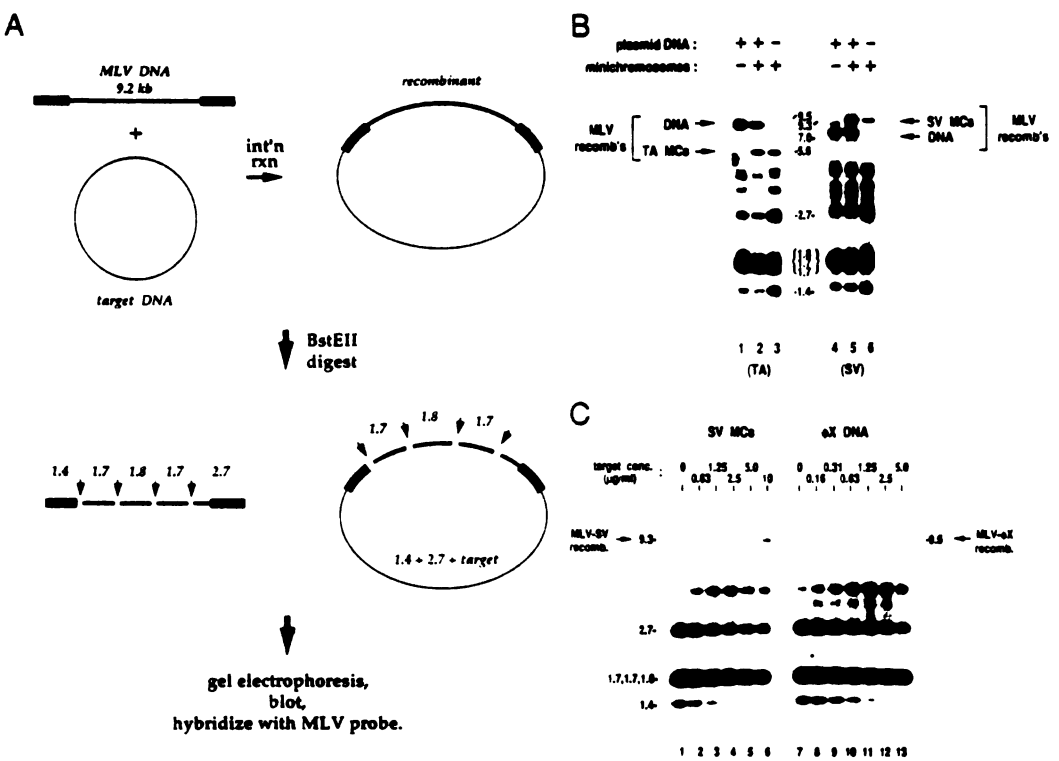


Fig. 3. Integration assays. A. Schematic depiction of the integration reaction and assay. The integration reaction gives a product with the MLV DNA (MoMLV-SupF, 9.2 kb) inserted into the target DNA (whose size is variable). Rarely does all the MLV DNA participate, so a subsequent *Bst*EII digest of the integration reaction products yields fragments originating from both unintegrated and integrated DNAs. However, a recombinant-specific fragment is generated from attachment of the ends of MLV DNA (1.4 and 2.7 kb in size) to the target DNA [and will be of a size equal to the sum of 1.4 + 2.7 + target size (in kb)]. B. Example of integration into MCs. Reactions contained either 1 μg/ml TA MCs, 10 μg/ml øX174 DNA or both as targets (left), or 10 μg/ml SV MCs, 10 μg/ml pGEM-2 DNA or both as targets (right). All reactions were carried out in parallel. Recombinants resulting from integration into TA MCs (1.5 kb), øX174 DNA (5.4 kb), SV MCs (5.2 kb) and pGEM-2 DNA (2.9 kb) are visualized by the appearance of 5.6, 9.5, 9.3 and 7.0 kb bands, respectively. Hybridizing fragments of sizes between those from recombinants and those from unintegrated DNA are usually observed, resulting (data not shown) from '1-LTR' and '2-LTR' circles and intramolecular 'autointegrants' (Shoemaker *et al.*, 1980, 1981; Varmus and Brown, 1989) that are not important for the experiments presented in this study. For a recent discussion of autointegrant formation *in vitro*, see Lee and Coffin (1990). C. Response of integration efficiency to target concentration. Reactions contained the indicated concentration of either SV MCs (left) or øX174 DNA (right), increasing in two-fold steps and recombinants (arrows) are visualized as described for (B). All reactions were carried out in parallel. The reactions in (B) and (C) were carried out in the presence of PEG (see Figure 4).

conditions. Also notable in the presence of PEG was the distinct boost in use of low amounts of MC targets upon addition of naked DNA (Figure 4B, lanes 2 and 3; Figure 4C, lanes 2 and 3); this is in accord with a consistently observed increase in all integration activity upon addition of DNA to these MLV integration extracts (P.M.P. and H.E.V., unpublished observations; Patrick Brown, personal communication), and is not pursued further here.

Because the addition of PEG strongly affected the relative use of MCs and DNA as targets, we tested other additives, including some non-ionic polymers that might be expected to change the apparent concentration of solutes by 'volume exclusion' (Tanford, 1961) and the polyamine spermidine (Figure 5A). Parallel reactions were performed using either SV MCs alone or SV MCs and naked plasmid DNA as targets. Of several reaction conditions examined, addition of spermidine was the most effective in stimulating use of both MCs and DNA (Figure 5A). Moreover, in the presence

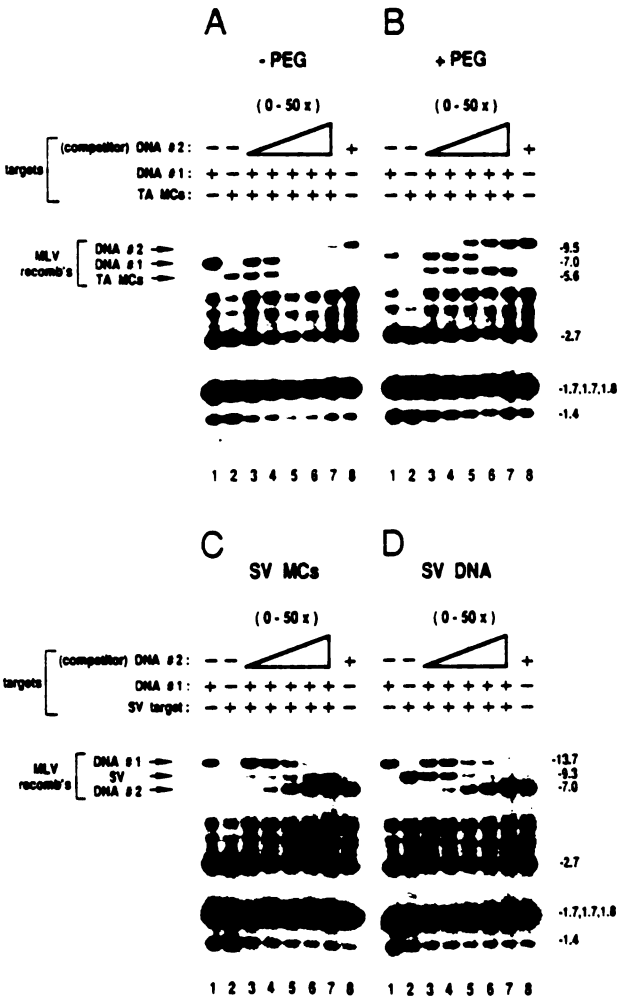


Fig. 4. Comparison of MCs with DNA as integration targets in response to excess competitor DNA. Integration reactions contained one, two, or three potential targets in the same mixture. In lanes 3–7, an MC target and a naked DNA (#1) target were present in constant amounts, and a varying amount of a third target (competitor; naked DNA #2) was present. A and B, TA MCs were absent (−) or present (+) at 0.8 μg/ml; pGEM-2 DNA (DNA #1) was absent (−) or present (+) at 1 μg/ml; φX174 DNA (DNA #2) was absent (−), present at 10 μg/ml (+) or present at 0, 0.4, 2, 10 or 50 μg/ml in lanes 3, 4, 5, 6 and 7, respectively, representing an increase from 0 to 50-fold excess over each of the other two targets. Recombinants (arrows) into TA MCs (1.5 kb), DNA #1 (pGEM-2; 2.9 kb), and DNA #2 (φX174 DNA; 5.4 kb) were identified by the appearance of 5.6, 7.0 and 9.5 kb bands, respectively. Reactions were carried out in the absence (A) or presence (B) of PEG. C and D, SV target was absent (−) or present (+) at 1 μg/ml; pSV-RI DNA (DNA #1) was absent (−) or present (+) at 1 μg/ml; pGEM-2 DNA (DNA #2) was absent (−), present at 10 μg/ml (+) or present at 0, 0.4, 2, 10 or 50 μg/ml in lanes 3, 4, 5, 6 and 7, respectively, representing an increase from 0 to 50-fold excess over each of the other two targets. Recombinants (arrows) into DNA #2 (pGEM-2; 2.9 kb), SV (5.2 kb) or DNA #1 (pSV-RI; 9.6 kb) were identified by the appearance of 7.0, 9.3 and 13.7 kb bands, respectively. Reactions were carried out in the presence of PEG, and the SV target used was either MCs (C) or DNA from deproteinized MCs (D). All reactions in A–D were performed in parallel.

of spermidine, as with PEG, DNA competed poorly with MCs as integration targets (Figure 5B). As a result, we added spermidine to all subsequent reactions.

Distribution of integration sites in MCs and DNA

To examine the distribution of integration sites within MCs, we cloned a large number of independent recombinants and sequenced the viral DNA–target DNA junctions, as described in Materials and methods. Most clones were sequenced at only one of the two junctions. Seven randomly picked MLV-TA MC recombinant clones were sequenced at both junctions, and all seven showed the expected loss of 2 bp from both ends of viral DNA and duplication of 4 bp of target DNA (Varmus and Brown, 1989), indicating that

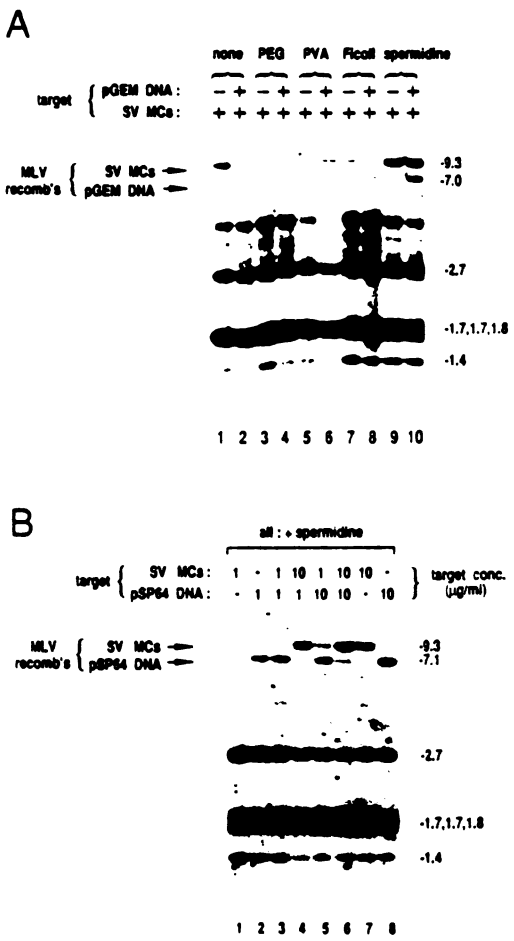


Fig. 5. Effect of reaction conditions on use of MC and DNA targets. **A.** All reactions contained 2 μg/ml SV MCs and either 0 (–) or 10 mg/ml (+) pGEM-2 DNA. Also, reactions were supplemented with additives as indicated (at concentrations given in Materials and methods). Recombinants (arrows) were identified as described in Figures 3 and 4. **B.** All reactions were supplemented with spermidine and were performed in the absence or presence of the indicated concentrations of SV MCs or pSP64 DNA. Recombinants (arrows) were identified as described in Figures 3 and 4 (pSP64 is 3.0 kb).

the insertions into MCs represent the products of legitimate MLV integration events (data not shown). The positions of 89 independent insertions into TA MCs, compared with 77 into TA DNA, reveal no strong preference for the nucleosome-free, nuclease-sensitive regions of the TA MCs (Figure 6A). This observation was supported by mapping 30 insertions into SV MCs (Figure 6B). In general, the distributions of insertions into TA MCs and DNA were grossly similar, with no major clustering or excluded regions. However, there was slightly more clustering of integration sites in MCs than in DNA, and there seemed to be a mild preference for, rather than a bias against, the nucleosomal regions in the MCs, but we have not attempted to analyze these features more rigorously here.

Because we mapped insertions by cloning and sequencing, allowing precise identification of insertion sites, we were able to recognize several non-random aspects of the distribution of integration events into MCs. All pairs of insertions that mapped within 25 bp of each other were identified, and the separation distance for each pair was measured (Figure 7A). Insertions into the TA MCs tended to be spaced in a regular fashion, mapping either very close (within 2 bp), or ~10 or 20 bp from each other, following a 10 bp periodic distribution. Such periodicity was not observed when naked TA DNA was used as the target. Because of its larger size and fewer insertions mapped, there were too few proximal pairs of insertions into SV MCs to analyze by these methods. A χ^2 goodness-of-fit test applied to these spacing data (see Materials and methods) suggested that the spacings of insertions into the TA MC were inconsistent with a random distribution ($P<0.001$), unlike the spacings of insertions into TA DNA ($P>0.10$). By summing the values in the histograms, the degree of bias towards this periodicity in TA MCs was estimated to be nearly 3-fold (Figure 7A). The insertion site period of ~10 bp correlates with the period of a DNA double helix, suggesting that the integration machinery displays a preference for one face of the DNA helix over the other in MCs. A model that can explain these observations (Figure 7B) is considered at length in the Discussion.

Two additional non-random features were observed in the distribution of integration sites in the TA MCs. First we compiled the coincident insertions, or independent integration events that map to the exact same position (Table I). In general, there were more frequent coincident insertions in MCs than in DNA: 34 of 89 in MCs versus eight of 77 in DNA. The fact that many coincident pairs were of different orientation and the careful steps taken in the cloning (see Materials and methods) argue for the independent origin of the coincident insertions. A statistical analysis (see Materials and methods) suggests that the probability that the number of coincident events observed with TA MCs would result from a random (Poisson) distribution is <0.001, in contrast to when TA DNA was the target ($P>0.50$).

Second, we looked for sequence bias among the sites used for insertions into TA MCs and DNA (Figure 8). The sequences of the target sites were aligned according to the position at which the four-base staggered break was made during the integration reaction. We then looked for bias among the aligned bases; specifically, we looked for frequent presence at any position of particular bases, A/T- or G/C-richness, or A/T-rich di- and trinucleotides that are known to show the strongest preferred positions in nucleosomal

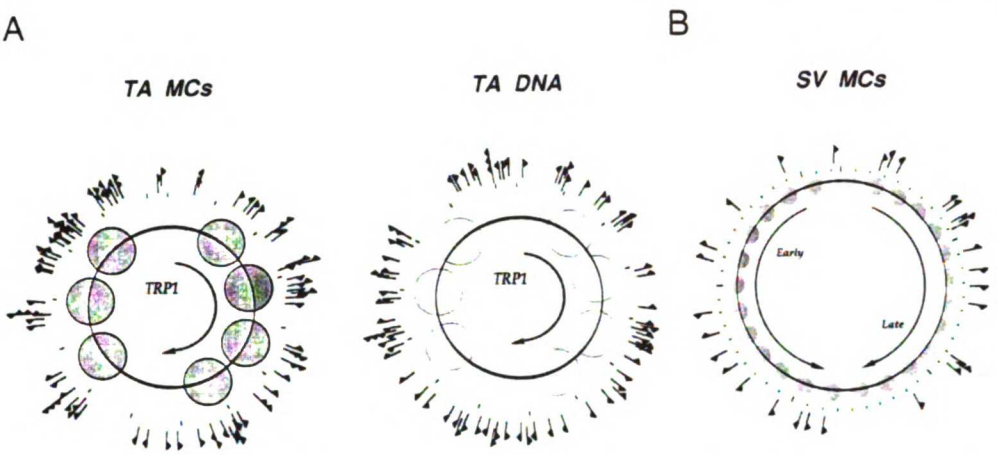


Fig. 6. Distributions of integration sites. Schematic representations of targets are similar to those in Figure 1. Flags designate mapped positions of independent insertion events (determined by cloning and sequencing as described in Materials and methods) and the orientation of each insertion. **A.** Distribution of 89 insertions into TA MCs (left) and 77 into TA DNA (right). Hash marks designate the progression along the TA sequence in 100 bp intervals, with position 0/1453 at top. In the TA DNA diagram, empty dashed-border circles indicate the positions of nucleosomes before deproteinization. **B.** Distribution of 30 insertions into SV MCs. Hash marks designate the progression along the SV sequence in 100 and 1000 bp intervals, with position 0/5243 at top.

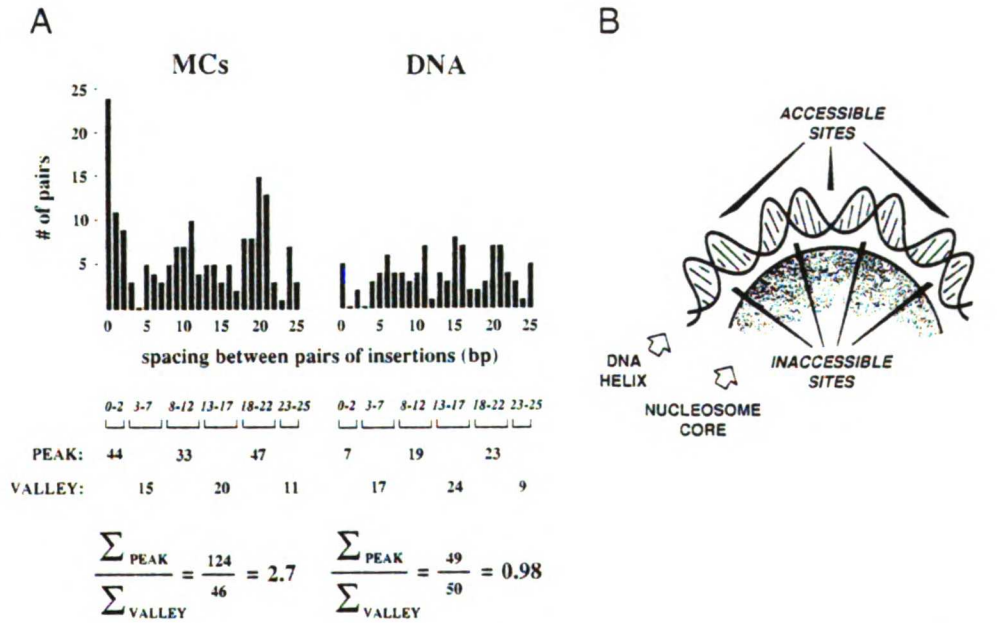


Fig. 7. A. Periodic distribution of integration sites in TA MCs. The distances between all pairs of independent insertions that map within 25 bp of each other were measured, and the histograms (above) show the number of pairs with each spacing value. In TA MCs (left), an ~10 bp periodic distribution of spacings is observed that is not observed in TA DNA (right). The MC case is inconsistent with a random distribution ($P<0.001$), in contrast to the DNA case ($P>0.10$; see text and Materials and methods for calculations). In addition, the numbers were combined in groups indicated by the brackets, corresponding to the 'peak' and 'valley' regions of the MC distribution, and the ratio of these sums was determined (below). In the DNA case, the ratio of these sums was close to 1 (0.98), as expected for a random distribution. In the MC case, however, the ratio of sums was 2.7, providing an estimate of the degree of bias (~3-fold) for the periodic distribution. **B.** Model for restriction of integration sites in nucleosomal DNA. The diagram depicts a segment of a DNA helix wrapped around a nucleosome core. The outside, or exposed, face of the helix is suggested to be preferentially accessible to the integration machinery. The implications of the model and its utility in explaining the experimental results are discussed in the text.

Table I. Higher frequency of coincident insertions in MCs than in DNA

	TA MCs		TA DNA	
	No. of inserts	No. of sites (orientations)	No. of inserts	No. of sites (orientations)
Sites not used	0	1364	0	1381
Sites used once	55	55	69	69
Sites used twice	24	12 (7+,−; 2+,+; 3−,−)	2	1 (−,−)
Sites used three times	6	2 (1+,−,−; 1+,+,−)	6	2 (2−,−,−)
Sites used four times	4	1 (+,−,−,−)	0	0
Total:	89	1453	77	1453
Frequency of coincident inserts	34/89 = 38%		8/77 = 10%	
Probability from random[a]		<0.001		>0.50

[a]Calculated as a χ^2 goodness-of-fit to a random (Poisson) distribution (see Materials and methods).

DNA (Satchwell *et al.*, 1986). As expected, insertion sites in naked DNA showed little preference at most positions, although a T was strikingly frequent at the second base from the site of cleavage on each strand. In contrast, insertion sites in TA MCs revealed additional base bias at several positions. Specifically, A/T-rich regions were favored in small clusters, following a roughly symmetrical organization about the target site: at the immediate center of the integration site and 10−12 bases to either side of the center. Such a bias pattern was independently revealed by noting positions where A/T-rich mono-, di- or trinucleotides were frequently present.

Thus, three distinct non-random features arose when comparing the integration sites in MCs with those in DNA: a periodic spacing of integration sites, an increased frequency of coincident insertions and an increased sequence bias at the target sites.

Discussion

Minichromosomes are used as targets for integration in vitro

We have developed a system for studying the integration of retroviral DNA into chromatin *in vitro*. Using viral nucleoprotein integration complexes from infected cells and two kinds of minichromosomes (MCs), we have found that integration occurs in both nucleosomal and nucleosome-free regions of chromatin. Moreover, the distribution of insertion sites in nucleosomal regions suggests that the integration machinery shows a preference for the exposed face of the DNA helix.

Several observations argue that the MCs *per se*, rather than naked DNA from disassembled MCs, were used as integration targets: (i) integration efficiency responded to similar concentrations of MC and DNA targets (Figure 3C); (ii) MC and DNA targets were sensitive to similar ratios of competitor target DNA (Figure 4A); (iii) MCs differed from DNA as targets under certain reaction conditions (Figures 4 and 5); and (iv) the distribution of insertion sites into TA MCs was less random than the distribution of those into TA DNA (Figures 7 and 8; Table I). The sequences of the virus−target junctions in the integration products confirm that the insertions into MCs resulted from legitimate integration events (not shown).

We were initially concerned that it might be difficult to observe integration events into MCs because much of the DNA would be unavailable when incorporated into nucleosomes. Instead, the MCs were used as integration targets with efficiencies similar to those of naked DNA targets. There were, however, complex effects of reaction conditions on the utilization of MCs and DNA as integration targets (Figures 4 and 5) that, although experimentally useful, are difficult to understand. It is possible that higher apparent concentrations of all components in the reaction mixture (e.g. in the presence of PEG or spermidine) favors integration into MCs. Alternatively, some partitioning (e.g. Yamamoto *et al.*, 1970; Hoopes and McClure, 1981) of integration machinery from target into different phases might occur, such that a greater proportion of MCs than DNA partitions with the integration machinery. Finally, it is possible that distinct classes of integration complexes exist, each specific for either MCs or DNA, although we disfavor this idea because of the observation that MCs can compete with DNA (e.g. see Figure 3B, and Figure 5B, lanes 3−8).

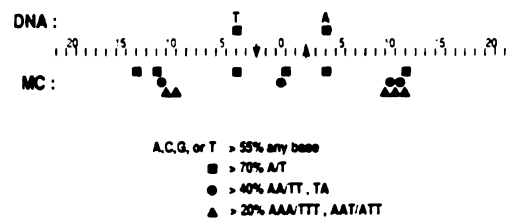


Fig. 8. Insertions into MCs show increased sequence bias. Insertion sites in TA MCs (86 total) and TA DNA (70 total) were separately aligned according to the position of the 4 bp staggered cut made in the target DNA (arrows), such that the orientation of MLV DNA was always in the same direction. At each position within the 4 base stagger, and 20 bases to each side, the number of occurrences of each base was counted (presented only for the top strand for simplicity). Also counted was the number of occurrences of A or T (squares), of the dinucleotides AA/TT or TA (circles), and of the trinucleotides AAA/TTT or AAT/ATT (triangles), and the frequencies of each occurrence calculated by dividing by the total number of sequences analyzed. In order to summarize the measured frequency distributions concisely, the positions at which these frequencies exceeded selected threshold values are shown. The probabilities of exceeding these thresholds at any one position are $P<0.001$ for the '>55% any one base' example and $P<0.05$ for the other examples (see Materials and methods for statistical calculations). Position 0 represents the exact center of the integration site, and positions 5, 10, 15 and 20 bonds from the center in each direction are shown.

Retroviral integration: a function that does not prefer nucleosome-free DNA

The distributions of integration sites into MCs and DNA were determined by sequencing individual clones of many integration products. The results show there is no strong preference for integration into nucleosome-free regions in either SV40 (SV) or TRP1ARS1 (TA) MCs; instead, nucleosomal DNA can be used as frequently as, and perhaps even more frequently than, nucleosome-free DNA in the same molecule. In contrast, nucleases such as micrococcal nuclease, DNase I, and restriction enzymes tend to show a strong preference for nucleosome-free regions, and, therefore, are used to determine nucleosome placement. However, such enzymes may not be able to interact productively with nucleosomal DNA. Retroviral integration, while a viral activity, may provide a useful paradigm for classes of cellular functions, such as general recombination or replication, that are active on (and may even prefer) nucleosomal DNA. Because retroviral integration can be studied both *in vivo* and *in vitro*, with many different DNA regions serving as targets, it is especially well suited to serve as such a paradigm.

A model for integration into nucleosomal DNA: the nucleosome influences choice of target sites

Although integration is not grossly inhibited by the presence of nucleosomes on the target DNA, a detailed analysis of the insertion sites suggests that the population of potential sites becomes more limited in MCs. Three distinct consequences of site limitations were noticed in our comparisons of integration sites in TA MCs and TA DNA: in MCs there was a periodic spacing of insertion sites at ~10 bp intervals, an increase in the frequency of coincident insertions and an increase in the apparent sequence bias. All three of these independent observations can be explained by the same model (Figure 7B). In this model, the orientation of the DNA sequences about a nucleosome core limits potential integration sites. We suggest that the face of the DNA helix against the nucleosome core is poorly accessible to the integration machinery, and the face away from the nucleosome core remains accessible. Recognizable features of the DNA helix, such as the major groove, would then be available only according to the period of the helix itself—roughly 10 to 10.5 bp—and would give rise to the observed periodic spacing of insertion sites (see Figure 7A). Alternatively, rather than inhibition of the inside face, enhancement of the outside face of the nucleosomal DNA helix, e.g. through bend-induced perturbations of DNA structure, could also result in its preferential use. One prediction from our model might be that the periodic spacing would not be observed in the nucleosome-free regions of the TA MCs. Unfortunately our collection contains too few proximal insertions in these regions (only four pairs in total) to address this point.

Periodic modulation of accessible sites in nucleosomal DNA has previously been observed with DNase I. This enzyme cuts nucleosomal DNA only at sites where the minor groove of the DNA helix is exposed on the face away from the nucleosome core, producing a 10–10.5 bp periodic pattern (Lutter, 1978; Drew, 1984). In its limitation to one face of the nucleosomal DNA helix, DNase I resembles the retroviral integration machinery, but it differs by strongly preferring nucleosome-free regions. Perlmann and Wrange (1988) and Pina *et al.* (1990) have shown that a transcription factor, the glucocorticoid receptor, can bind its sites in the MMTV LTR even when the DNA is incorporated into a nucleosome. Furthermore, it was suggested that sites in nucleosomal DNA were only bound if the major groove of the helix at that site was facing away from the nucleosome core. This scenario is highly analogous to that which we are proposing for retroviral integration.

The model proposed to explain our findings with retroviral integration (or results of others with glucocorticoid receptor binding) requires that the nucleosomal DNA sequences have a consistent rotational orientation. In other words, the same bases must consistently face either toward or away from the nucleosome core in most or all nucleosomes. Is this a reasonable and usual condition? The answer appears to be yes, at least for the cases which have been directly analyzed. In general there is strong evidence that the primary determinant of the rotational orientation of the nucleosomal DNA helix is the sequence of the DNA segment itself (Linxweiler and Horz, 1985; Drew and Travers, 1985; Satchwell *et al.*, 1986). Particular short DNA sequences display strong preferences for placement either toward or away from the nucleosome core, because of differences in the physical constraints of the compressed (inside) versus the expanded (outside) portions of the wrapped DNA. The incorporation of DNA into the nucleosome appears to occur in a manner which maximizes the placement of as many sequences as possible at their preferred positions (Satchwell *et al.*, 1986). These general principles have even been used to design artificial DNA segments that are incorporated into nucleosomes more favorably than natural sequences (Shrader and Crothers, 1989). In several closely examined cases, such as the MMTV LTR fragment mentioned earlier, DNA sequences are placed into nucleosomes precisely, such that the vast majority of nucleosomes in a population show the same rotational orientation of the nucleosomal DNA (Simpson and Stafford, 1983; Linxweiler and Horz, 1985; Pina *et al.*, 1990). However, the rotational orientations of the seven nucleosomes in the TA MCs used in our studies have not been determined, and the nucleosome positions along the DNA have been determined only at low resolution (±10–20 bp; Thoma *et al.*, 1984). One prediction of our work, therefore, is that the TA MC nucleosomes will show a consistent or major rotational positioning.

Increased frequency of coincident insertions in minichromosome targets

The above sort of specific rotational positioning of nucleosomal DNA sequences, which we hypothesize exists in the TA MCs, could explain the increased frequency of insertions into the exact same site (coincident insertions). A specific or major rotational positioning would make a particular subset of all possible sites, clustered at 10 bp intervals, preferentially available. In the simplest case, fewer actual sites would be available in the MC than in DNA, because many of them would be made unavailable due to their position on the core-proximal face of nucleosomal DNA (see Figure 7B). Thus, with fewer sites and a similar number of events analyzed, more events would occur in the same site in MCs than in DNA, in accord with our observations (Table I). Again, it is also possible that the creation of exceptionally good sites on the outer face of nucleosomal DNA could contribute to the increased frequency of

coincident insertions. We presently lack any compelling evidence to distinguish between these two ideas, although the lack of bias against nucleosomal DNA (and for nucleosome-free DNA) may favor the second explanation.

Some information is available about the frequency of coincident insertions *in vivo*. There are several loci in which numerous proviral insertions have been mapped following selection for insertional activation of proto-oncogenes (Selten *et al.*, 1984; Shih *et al.*, 1984; Raines *et al.*, 1985). In all of these cases, it is clear that many different positions have served as integration sites. However, the locations of some proviruses are indistinguishable at the resolution level of restriction mapping, but the actual viral–host junctions have not yet been sequenced in these cases to determine whether any represent coincident insertions. Shih *et al.* (1988) showed that a disproportionate number of RSV-mediated integration events occur into a subset of chromosomal regions, and that all of the insertions within those regions were at exactly the same base. This represents a degree of bias not easily explained by our model, and may depend on more complicated phenomena. Importantly, when these high frequency sites were used as naked DNA targets in an *in vitro* integration reaction, they were not used preferentially, and insertions occurred at multiple sites (J.Coffin, personal communication). In a large number of insertions of the yeast retrotransposon Ty1 selected on the ability to impair function of the *URA3* and *LYS2* genes, an exceptionally high proportion of sites sustained multiple insertions (Natsoulis *et al.*, 1989). While these observations may suffer from the usual caveats of selection bias, the use of disruption rather than activation insertions would seem the most permissive strategy. In any case, it should now be possible to compare unselected insertions into the same loci *in vitro*, in a manner analogous to the present study, by use of an *in vitro* Ty1 transposition system (Eichinger and Boeke, 1988). Integration of the yeast retrotransposon Ty3 occurs virtually exclusively at the transcriptional start sites of tRNA genes (Chalker and Sandmeyer, 1990). This extreme degree of target site preference is not explainable by target sequence alone, since the tRNA gene must be transcriptionally competent to serve as a high-frequency target (D.Chalker and S.Sandmeyer, personal communication). This may be the clearest example of how the same DNA sequence can be used differently as an integration target depending on its physiological state.

Insertions into minichromosomes show target sequence bias

As expected from previous attempts to define target consensus sequences (Shimotohno and Temin, 1980; Shoemaker *et al.*, 1981), the sites used for insertion into naked DNA bear little resemblance to each other, although there was a striking preference for a T base on each strand at the second position 5′ from the target cleavage site (see Figure 8). More importantly, the insertion sites into MCs showed additional sequence bias. This presents another non-random feature of integration into chromatin that can be explained by the model presented above, without a requirement for specific sequence recognition by the integration machinery. As discussed above, the DNA sequence determines the rotational positioning of the DNA helix around the nucleosome core, i.e. which bases face towards and which away from the core. The most critical features of the DNA sequence are the distribution of A/T- and G/C-rich regions. These will become arranged to maximize the placement of A/T-rich regions at positions where the minor groove of the DNA helix faces inward and G/C-rich regions where it faces outward, with particular di- and trinucleotides showing the strongest periodic modulation of position (Drew and Travers, 1985; Satchwell *et al.*, 1986). The sequence bias of the MC insertion sites is predominantly towards a symmetrical pattern of A/T-rich mono-, di- and trinucleotides. We suggest that this results from integration into the major groove of the DNA helix on the face away from the nucleosome core, such that A/T-rich regions will tend to be present at the center and at ~10 bp intervals to either side of the center of the integration site. Insertion by attack into the major groove is consistent with the fact that the 4–6 bp 5′ staggered cuts of target DNA made during retroviral integration (Varmus and Brown, 1989) can be accomplished by cleavage of phosphodiester bonds that face each other directly across the major groove. Importantly, we argue that the insertion site sequence bias observed for the MC case would not be due to a true bias by the integration machinery for those sequences; instead it would be due to the presentation by the nucleosome of a limited subset of the original potential sites, as determined by their rotational arrangement. We are currently subjecting these ideas to a more thorough test by using as an integration target nucleosomal DNA for which the rotational orientation is known.

The cloning and sequencing method of mapping has allowed us to draw interesting conclusions from a higher density of retroviral integration sites than previously obtained. Nevertheless, the data obtained can be regarded as rather sparse, in the sense that far fewer insertions were mapped than there are potential sites in the targets. We have therefore recently developed a general technique for mapping integration site distributions in large populations of recombinants rather than individual clones. Preliminary results relevant to the findings presented here suggest that the reaction conditions observed to affect the efficiency of integration into MCs and DNA (see Figures 4 and 5) do not affect the distribution of integration sites in these targets or the chromatin structures of the MC targets (P.M.P. and H.E.V., unpublished observations).

Chromatin structure and integration in vivo

Among the motivations for this work were previous suggestions that integration *in vivo* occurs preferentially in nuclease-sensitive and/or transcriptionally active regions. Do our observations conflict with those claims? We think not. In prior efforts to correlate integration sites with nuclease-sensitive sites *in vivo* (Vijaya *et al.*, 1986; Rohdewohld *et al.*, 1987), the criteria for proximity of integration sites to nuclease-sensitive sites allowed distances of up to 500 bp. By this measure, most of the target DNA in our MCs can be considered to be 'near' a nuclease-sensitive region; in the case of the TA MCs, all bases are within 500 bp of a nuclease-sensitive, nucleosome-free region. Most importantly, our results imply that the previous *in vivo* data cannot be explained by preferential use of nucleosome-free DNA over nucleosomal DNA. It seems plausible, however, that some preference could operate by discrimination between uncondensed and more highly condensed levels of chromatin packaging; in our *in vitro* studies we have only

utilized chromatin targets with a relatively low condensation level. Thus, the use of more complex chromatin models as *in vitro* integration targets would seem warranted.

Materials and methods

Strains, cells and plasmids

MoMLV-SupF (see Brown *et al.*, 1987) was propagated on SC-1 cells (courtesy of J.Levy, University of California San Francisco) and used to infect NIH-3T3 cells. The triple protease-deficient yeast strain BJ2168 (*MATa pep4-3, prc1-407, prb1-1122, ura3-52, trp1, leu2, gal2*; Jones, 1991), made cir^0 by selecting for loss of endogenous 2μ plasmid (by J.Thomas and K.Yamamoto, University of California San Francisco), served as the host for the TRP1ARS1 plasmid. This plasmid was introduced by transformation of a self-ligated *Eco*RI fragment cut from the plasmid pTA-R (from J Thomas), which contains the TRP1ARS1 *Eco*RI fragment cloned into the *Eco*RI site of the pUC18 vector. Following transformation, a clone (BJ2168cir^0:TA-9) that contained monomer-size plasmid (as determined by gel and Southern blot analysis) was picked for further use and was grown in standard synthetic medium without tryptophan (Dean *et al.*, 1989). Wild-type SV40 strain #777 (from E.Shekhtman and N.Cozzarelli, University of California Berkeley) was propogated on CV-1 monkey cells by standard methods (e.g. Oudet *et al.*, 1989). The plasmid pSV-RI (from M.Verderame) contains an *Eco*RI-linearized SV40 genome cloned into the vector pBR322; subsequently, the *Eco*RI SV40 fragment was cut from pSV-RI and cloned into the pBS-KS(+) vector to obtain the pBS-SVR plasmid. The plasmids pGEM-2 and pSP64 (Promega) and ϕX174 (New England Biolabs) were used as integration targets. A plasmid pUC8.2, constructed by cloning the 8.2 kb *Hin*dIII fragment of permuted circular MoMLV DNA from p8.2 (Schwartzberg *et al.*, 1983) into the *Hin*dIII site of pUC18, was used as a probe for Southern blot hybridization. The bacterial strain XAC-1 (F' $lacI_{171}$ $lacZ_{alfam}$ *proB*+/ F$^-$ Δ(lacpro)$_{111}$ nalA, rif, $argE_{am}$, ara; Normanly *et al.*, 1986) was obtained from J.Miller (University of California Los Angeles).

Integration extract preparation and reactions

MLV-SupF integration extracts were prepared as described (Brown *et al.*, 1987), 12–16 h after infection of NIH-3T3 cells with fresh 12–24 h harvests of virus from chronically infected SC-1 cells and were stored frozen in aliquots at −80°C. Only cytoplasmic extracts were used for the experiments described here. Integration reactions were performed as described (Brown *et al.*, 1987). Briefly, 50 μl of cold extract was combined with target DNA, minichromosome preparation or both (usually at 1–10 μg/ml; for specific concentrations see text) and brought to a final reaction volume of 75 μl with water. Some reactions (see text) also contained 5% polyethylene glycol 8000 (PEG), 3.3% polyvinyl alcohol (PVA), 4.2% Ficoll 400 or 15 mM spermidine as additives; these were always added last. Reactions were then incubated at 37°C for 30–60 min and were stopped; nucleic acids were prepared as described (Brown *et al.*, 1987). To assay the extent of integration reactions (Brown *et al.*, 1989), the purified nucleic acids were digested with *Bst*EII and run on 0.7% agarose gels, which were then blotted onto Hybond-N nylon membranes (Amersham), hybridized (Church and Gilbert, 1984) to an MLV probe [pUC8.2 labelled with ^{32}P by the hexamer-priming method (Feinberg and Vogelstein, 1983)] and analyzed by autoradiography.

Preparation of minichromosomes

TRP1ARS1 minichromosomes (TA MCs) were purified from yeast cells by a method based upon that of Dean *et al.* (1989). Yeast cultures (OD_{600} = 1) were collected, spheroplasts were prepared, collected and lysed, and nuclei were collected according to Dean *et al.* (1989). Nuclear pellets were washed once in 80 mM KCl, 5 mM $MgCl_2$, 10 mM PIPES pH 6.3, 1 mM EGTA, 0.5 mM spermidine, 1% aprotinin and 18% Ficoll 400, and were then resuspended in (1 ml per liter original culture) 200 mM NaCl, 5 mM $MgCl_2$, 10 mM MOPS pH 7.4, 0.5 mM EGTA, 1% aprotinin and 0.1% 2-mercaptoethanol [nuclear elution buffer (NEB)]. MCs were then allowed to elute from these nuclei on ice for 1.5 h, followed by centrifugation in a GSA rotor at 7000 r.p.m. for 5 min at 4°C, collection of the supernatant and then one repeat of resuspension, elution, centrifugation and supernatant collection. The combined supernatants were mixed with an equal volume of 80% Nycodenz (Accurate Chemical) in NEB, and centrifuged for 36 h at 45 000 r.p.m., 4°C, in a Beckman VTi50 rotor. Fractions containing MCs were pooled and re-centrifuged for 18 h at 55 000 r.p.m., 4°C, in a Beckman VTi65 rotor. Peak fractions (250 μl per liter original culture, in ~40% Nycodenz) were pooled, made 8% in sucrose and frozen in aliquots at −80°C. These preparations contained ~8 ng/μl TRP1ARS1 DNA, as estimated by comparison with known standards on agarose–ethidium bromide gels and confirmed by fluorometric measurement.

SV40 minichromosomes (SV MCs) were purified 68 h after infection of CV-1 cells at a multiplicity of infection of 10 by a method based upon Luchnik *et al.* (1982) and Oudet *et al.* (1989). Cells were scraped from culture plates, pelleted and washed once in 10 mM PIPES pH 6.8, 150 mM NaCl and 1 mM EDTA [wash buffer (WB)], then lysed in WB plus 0.05% digitonin. Nuclei were harvested by centrifugation, washed twice in WB plus digitonin, then resuspended in (100 μl/10^7 cells) 10 mM MOPS pH 8.0, 200 mM NaCl and 1 mM EDTA. MCs were allowed to elute for 4 h on ice, then nuclei and debris were pelleted at 10 000 *g* for 15 min. The supernatant was made 8% in sucrose and frozen in aliquots at −80°C. Such (68 h post-infection) preparations contained significantly more DNA than others at shorter times post-infection, although much of the DNA was associated with previrion (180S) structures (Boyce *et al.*, 1982). The concentration of SV40 DNA was determined to be ~100 ng/μl by methods described above for the TA MCs.

Cloning and sequencing of integration products

To prepare integration products for cloning and sequencing, integration reactions were carried out as described above (with spermidine present in the reaction), in two separate experiments: experiment #1 involved separate parallel reactions into SV and TA MCs, and experiment #2 involved separate parallel reactions into TA MCs and TA DNA (which was obtained by deproteinization, phenol–chloroform extraction and ethanol precipitation). Data for TA MCs from the two experiments were combined for all analyses in this study. Integration products were cloned by digestion with *Sac*I (TA) or *Xba*I (SV) and ligation into the corresponding site in λ ZAP DNA (Stratagene), which contains a debilitating amber termination codon (Sam100) that can be suppressed by the suppressor tRNA gene resident in the MLV-SupF genome, allowing growth on a Sup$^+$ bacterial host. Ligated DNA was packaged into phage, which were then plated onto XAC-1 cells in the presence of X-gal and IPTG. Phage containing MLV-SupF sequences gave rise to blue plaques, and those blue plaques that contained recombinants (usually 5–10%) were identified by hybridization of plaque lifts (Benton and Davis, 1977) to SV or TA probes (these target DNAs could only be cloned as MLV recombinants, since they do not contain sites for the restriction enzymes used for the cloning). Recombinant plaques were picked from several independent platings of the same packaged phage, and were stored as individual clones. The inserts in these phage were subsequently excised as phagemids (according to the manufacturer's instructions), were grown in cultures, and plasmids were prepared by alkaline lysis miniprep procedures. Lambda phage clones were individually manipulated by this procedure in small groups (no more than 24 clones picked per plate) on several different days (no more than 24 clones per day) in order to ensure the independent origin of each final miniprep DNA sample. Thus, of the 11 cases of same-orientation same-site inserts (see Table I), 10 were from different experiments (#1 or #2 TA MC) or from different plates; only one case was from the same plate of lambda phage plaques.

Double-stranded miniprep DNAs were sequenced using an oligonucleotide primer (SupF-17) complementary to SupF gene sequences in the MLV-SupF LTR sequence and directed toward the viral DNA–target DNA junction. Some reactions gave unreadable sequence and were not resequenced; in total, 30 SV MC recombinants, 89 TA MC recombinants and 77 TA DNA recombinants (196 total) gave readable sequence. The insertion sites were located by matching the sequence immediately past the junction with the known target sequence [using the computer program DNA Strider, (Marck, 1988)]. These were used to generate the insertion site distribution maps. All clones were sequenced at one end of the viral DNA (using the SupF-17 primer). In addition, seven TA MC recombinant clones were picked at random and were sequenced at the other end (using the MoU5L17 primer) to verify that the insertions into MCs were legitimate products of retroviral integration. All others, except eight of 196, displayed a normal junction ending in the viral CA dinucleotide at the one end sequenced; but one of 30 SV MC, two of 89 TA MC and five of 77 TA DNA recombinants had junctions ending in CATT, suggesting that the proximal TT dinucleotide was not removed during the integration process. Seven of these eight clones were then sequenced at the other end, and all showed proper ending of the viral sequence with CA but aberrantly sized duplication of target sequences. These structures are consistent with normal integration, followed by misrepair of the gapped intermediate without removal of the unpaired AA dinucleotide of viral DNA at one end. These clones were not used in the detailed analyses of insertion site periodicity and sequence bias (none were members of coincident insertion sets), because of the ambiguity of the exact site of integration. Also not included in these detailed analyses were three other TA recombinants: two (one MC and one DNA) for which the sequencing gel was slightly obscured in the region

of the target junction so that the insertion site could only be assigned ± 1 bp, and 1 (DNA) that was found to have duplicated 5 bp of target DNA after being randomly picked for sequencing of both junctions. The remaining 86 TA MC and 70 TA DNA recombinants were then analyzed further. The exact insertion sites are not presented here, but can be obtained by written request to the authors.

Statistical methods for analysis of insertion site distributions

The insertion site distributions were analyzed by χ^2 tests of goodness-of-fit to a random distribution (Zar, 1984). Such tests calculated $\chi^2 = \Sigma(f_o - f_e)^2/f_e$, where f_o is the observed frequency and f_e is the expected frequency of an occurrence. The calculated χ^2 was then compared with a table of χ^2 critical values for the appropopriate degrees of freedom, υ. For the analysis of the spacings between insertion sites: f_e is equal to the total number of pairs of insertion sites with spacings of 25 bp or less (170 for TA MCs, 99 for TA DNA) divided by the total number of potential spacing values (=26; thus, f_e is 170/26 for TA MCs and 99/26 for TA DNA); f_o for each spacing value (0 to 26) is equal to the actual observed number of pairs with that spacing value (see Figure7A). Thus, χ^2 = 98.3 for the TA MC distribution and 32.6 for the TA DNA distribution; in both cases υ = 25; comparison with a table of critical values reveals that the TA MC distribution is inconsistent with a random distribution (χ^2 = 98.3, $P<0.001$) (i.e. the null hypothesis, that the observed data resulted from a random distribution, should be rejected), but the TA DNA distribution (χ^2 = 32.6; $0.25>P>0.1$) is consistent with a random distribution (i.e. the null hypothesis cannot be rejected).

For the analysis of coincident insertions, f_e was calculated according to a Poisson distribution, such that $f_e(X) = N\mu^X e^{-\mu}/X!$ where X is the number of inserts at the same site (0 to 4), N is the total number of potential sites (1453) and μ is the population mean number of occurrence per site (equal to the number of inserts sequenced divided by N; so μ = 99/1453 for TA MCs and 77/1453 for TA DNA); f_o for each value of X is equal to the number of sites that were observed to receive X inserts at the same site (see Table I); χ^2 was then calculated as above, with pooling of the frequencies in the tails of the distribution such that f_e was never <1.0 (Cochran, 1954; υ was 2 after such pooling); the TA MC distribution (χ^2 = 68.3) was inconsistent with a random (Poisson) distribution ($P<0.001$), while the TA DNA distribution was consistent with a random distribution (χ^2 = 0.767; $0.75>P>0.50$).

For the analysis of sequence bias, f_e was directly measured by counting the frequencies of bases, dinucleotides and trinucleotides in the TRP1ARS1 sequence. This sequence contains 29.8% A, 29.8% T, 20.2% G, 20.2% C, 59.6% A/T, 40.4% G/C, 27.1% AA/TT+TA and 12.3% AAA/TTT+AAT/ATT. χ^2 was calculated for exceeding the selected threshold frequencies that are shown in Figure 8: for >55% any one base, χ^2 = 26.2 (MC) or 21.1 (DNA), giving $P<0.001$; for >70% A/T, χ^2 = 3.86 (MC), giving $P<0.05$; for >40% AA/TT+TA, χ^2 = 7.24 (MC), giving $P<0.01$; and for >20% AAA/TTT+AAT/ATT, χ^2 = 4.72 (MC), giving $P<0.05$.

Acknowledgements

The authors thank Jay Thomas for generous gifts of reagents and advice in minichromosome preparations. We are especially grateful to Pat Brown for advice, discussion and assistance during many stages of this work. We also thank Julien Hoffman, Stanton Glantz, Joan Hilton and Vojtech Licko for advice on statistical analysis, and Hans-Peter Muller for comments on the manuscript. This work was supported by a grant from the NIH to H.E.V., and P.M.P. was supported by a UCSF Chancellor's Fellowship and an NIH Genetics Training Grant. H.E.V. is an American Cancer Society Research Professor.

References

Ambrose,C., Blasquez,V. and Bina,M. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 3287–3291.

Ambrose,C., Rajadhyaksha,A., Lowman,H. and Bina,M. (1989) *J. Mol. Biol.*, **209**, 255–263.

Ambrose,C., Lowman,H., Rajadhyaksha,A., Blasquez,V. and Bina,M. (1990) *J. Mol. Biol.*, **214**, 875–884.

Benton,W.D. and Davis,R.W. (1977) *Science*, **196**, 180–182.

Bonne-Andrea,C., Wong,M.L. and Alberts,B.M. (1990) *Nature*, **343**, 719–726.

Bowerman,B., Brown,P.O., Bishop,J.M. and Varmus,H.E. (1989) *Genes Dev.*, **3**, 469–478.

Boyce,F.M., Sundin,O., Barsoum,J. and Varshavsky,A. (1982) *J. Virol.*, **42**, 292–296.

Brown,P.O., Bowerman,B., Varmus,H.E. and Bishop,J.M. (1987) *Cell*, **49**, 347–356.

Brown,P.O., Bowerman,B., Varmus,H.E. and Bishop,J.M. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 2525–2529.

Chalker,D.L. and Sandmeyer,S.B. (1990) *Genetics*, **126**, 837–850.

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

Cochran,W.G. (1954) *Biometrics*, **10**, 417–451.

Craigie,R., Fujiwara,T. and Bushman,F. (1990) *Cell*, **62**, 829–837.

Dean,A., Pederson,D.S. and Simpson,R.T. (1989) *Methods Enzymol.*, **170**, 26–41.

Drew,H.R. (1984) *J. Mol. Biol.*, **176**, 535–557.

Drew,H.R. and Travers,A.A. (1985) *J. Mol. Biol.*, **186**, 773–790.

Eichinger,D. and Boeke,J.D. (1988) *Cell*, **54**, 955–966.

Feinberg,A.P. and Vogelstein,B. (1983) *Anal. Biochem.*, **132**, 6–13.

Gross,D.S. and Garrard,W.T. (1988) *Annu. Rev. Biochem.*, **57**, 159–197.

Grunstein,M. (1990) *Annu. Rev. Cell Biol.*, **6**, 643–678.

Hoopes,B.C. and McClure,W.R. (1981) *Nucleic Acids Res.*, **9**, 5493–5505.

Jones,E.W. (1991) *Methods Enzymol.*, **194**, 428–453.

Katz,R.A., Merkel,G., Kulkosky,J., Leis,J. and Skalka,A.M. (1990) *Cell*, **63**, 87–95.

King,W., Patel,M.D., Lobel,L.I., Goff,S.P. and Nguyen-Huu,M.C. (1985) *Science*, **228**, 554–558.

Lee,Y.M.H. and Coffin,J.M. (1990) *J. Virol.*, **64**, 5958–5965.

Linxweiler,W. and Horz,W. (1985) *Cell*, **42**, 281–290.

Lorch,Y., LaPointe,J.W. and Kornberg,R.D. (1987) *Cell*, **49**, 203–210.

Luchnik,A.N., Bakayev,V.V., Zbarsky,I.B. and Georgiev,G.P. (1982) *EMBO J.*, **1**, 1353–1358.

Lutter,L.C. (1978) *J. Mol. Biol.*, **124**, 391–420.

Marck,C. (1988) *Nucleic Acids Res.*, **16**, 1829–1836.

Mooslehner,K., Karls,U. and Harbers,K. (1990) *J. Virol.*, **64**, 3056–3058.

Natsoulis,G., Thomas,N., Roghmann,M.-C., Winston,F. and Boeke,J.D. (1989) *Genetics*, **123**, 269–279.

Normanly,J., Masson,J.-M., Kleina,L.G., Abelson,J. and Miller,J.H. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 6548–6552.

Oudet,P., Weiss,E. and Regnier,E. (1989) *Methods Enzymol.*, **170**, 14–25.

Pederson,D.S., Thoma,F. and Simpson,R.T. (1986a) *Annu. Rev. Cell Biol.*, **2**, 117–147.

Pederson,D.S., Venkatesan,M., Thoma,F. and Simpson,R.T. (1986b) *Proc. Natl. Acad. Sci. USA*, **83**, 7206–7210.

Perlmann,T. and Wrange,O. (1988) *EMBO J.*, **7**, 3073–3079.

Pina,B., Bruggemeier,U. and Beato,M. (1990) *Cell*, **60**, 719–731.

Raines,M.A., Lewis,W.G., Crittenden,L.B. and Kung,H.-J. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 2287–2291.

Reddy,S., DeGregori,J.V., von Melchner,H. and Ruley,H.E. (1991) *J. Virol.*, **65**, 1507–1515.

Rohdewohld,H., Weiher,H., Reik,W., Jaenisch,R. and Breindl,M. (1987) *J. Virol.*, **61**, 336–343.

Sandmeyer,S.B., Hansen,L.J. and Chalker,D.L. (1990) *Annu. Rev. Genet.*, **24**, 491–518.

Saragosti,S., Moyne,G. and Yaniv,M. (1980) *Cell*, **20**, 65–73.

Satchwell,S.C., Drew,H.R. and Travers,A.A. (1986) *J. Mol. Biol.*, **191**, 659–675.

Scherdin,U., Rhodes,K. and Breindl,M. (1990) *J. Virol.*, **64**, 907–912.

Schwartzberg,P., Colicelli,J. and Goff,S.P. (1983) *J. Virol.*, **46**, 538–546.

Selten,G., Cuypers,H.T., Zijlstra,M., Melief,C. and Berns,A. (1984) *EMBO J.*, **3**, 3215–3222.

Shelton,E.R., Wassarman,P.M. and DePamphilis,M.L. (1980) *J. Biol. Chem.*, **255**, 771–782.

Shih,C.-K., Linial,M., Goodenow,M.M. and Hayward,W.S. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 4697–4701.

Shih,C.-C., Stoye,J.P. and Coffin,J.M. (1988) *Cell*, **53**, 531–537.

Shimotohno,K. and Temin,H.M. (1980) *Proc. Natl. Acad. Sci. USA*, **77**, 7357–7361.

Shoemaker,C.S., Goff,S., Gilboa,E., Paskind,M., Mitra,S.W. and Baltimore,D. (1980) *Proc. Natl. Acad. Sci. USA*, **77**, 3932–3936.

Shoemaker,C., Hoffmann,J., Goff,S.P. and Baltimore,D. (1981) *J. Virol.*, **40**, 164–172.

Shrader,T.E. and Crothers,D.M. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 7418–7422.

Simpson,R.T. (1990) *Nature*, **343**, 387–389.

Simpson,R.T. and Stafford,D.W. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 51–55.

Sogo,J.M., Stahl,H., Koller,Th. and Knippers,R. (1986) *J. Mol. Biol.*, **189**, 189–204.

Tanford,C. (1961) *Physical Chemistry of Macromolecules*. Wiley, New York.
Thoma,F. and Simpson,R.T. (1985) *Nature*, **315**, 250–252.
Thoma,F., Bergman,L.W. and Simpson,R.T. (1984) *J. Mol. Biol.*, **177**, 715–733.
Varmus,H.E. and Brown,P.O. (1989) In Berg,D.E. and Howe,M.M. (eds), *Mobile DNA*. American Society of Microbiology, Washington, DC pp. 53–108.
Varshavsky,A.J., Sundin,O.H. and Bohn,M.J. (1978) *Nucleic Acids Res.*, **5**, 3469–3478.
Vijaya,S., Steffen,D.L. and Robinson,H.L. (1986) *J. Virol.*, **60**, 683–692.
Yamamoto,K.R., Alberts,B.M., Benzinger,R., Lawhorne,L. and Treiber,G. (1970) *Virology*, **40**,734–744.
Zar,J.H. (1984) *Biostatistical Analysis*. Prentice-Hall, Inc., New Jersey.

Received on August 27, 1991; revised on October 7, 1991

CHAPTER FOUR

NUCLEOSOMES, DNA-BINDING PROTEINS, AND DNA SEQUENCE MODULATE RETROVIRAL INTEGRATION TARGET SITE SELECTION

SUMMARY

Integration of retroviral DNA can serve as a paradigm for cellular functions that act upon DNA and are affected by the packaging of DNA into chromatin. We have used a novel PCR-based assay to survey DNA and chromatin for the precise distribution of many integration sites. Integration into naked DNA targets is shown to be non-uniform, implying a nucleotide sequence bias. In mononucleosomes assembled in vitro, a direct comparison with sites cut by DNase I demonstrates that integration occurs into the major groove on the exposed face of the nucleosomal DNA helix, resulting in a 10-bp periodic spacing of preferred sites. Integration into minichromosomes containing both nucleosomal and nucleosome-free DNA occurs most frequently at sites created in nucleosomal regions, demonstrating that chromatin assembly can enhance the reactivity of many sites. In contrast, integration is prevented in a region occupied by a site-specific DNA binding protein. Comparisons of integration events mediated by viral nucleoprotein complexes or by two different retroviral integrases show that the integration machinery also affects target site selection.

INTRODUCTION

DNA in the nucleus of eukaryotic cells is organized and condensed by its association with histones and other proteins to form chromatin (Kornberg, 1977; Pederson et al., 1986). Assembly of DNA into nucleosomes and higher-order forms of chromatin is assumed to alter its accessibility to factors that mediate many important physiological events, including replication, transcription, recombination, and repair. In general, transcriptional control studies and experimental probing with chemicals, enzymes, or DNA binding proteins suggest that nucleosomes restrict access to DNA and rarely, if ever, promote it (reviewed in Gross and Garrard, 1988; Grunstein, 1990; Wolffe, 1991; Kornberg and Lorch, 1991; and Felsenfeld, 1992). However, efforts to understand how packaging of DNA into chromatin influences cellular functions have been compromised by the limited number of functions and well-defined targets that can be manipulated in vitro or in vivo. We have, therefore, been studying how chromatin assembly affects the use of DNA as a target for retroviral integration.

As a normal part of their life cycle, retroviruses integrate a DNA copy of their genome into the chromosomal DNA of the host cell (reviewed in Varmus and Brown, 1989). The state of the DNA may influence its accessibility to retroviral integration machinery, potentially providing a model for how some processes, such as recombination reactions in general, are affected by chromatin structure. Retroviral integration has several potential advantages as a model probe for the effects of chromatin packaging: numerous sites, not just single sites, can be surveyed, and each insertion of retroviral DNA permanently marks the interacting site in target DNA. Retroviral integration is clearly not sequence-specific, but in vivo studies have suggested some non-random preferences for certain structural or functional states (King et al., 1985; Vijaya et al., 1986; Rohdewohld et al., 1987; Shih et al., 1988; Scherdin et al., 1990; Mooslehner et al., 1990), perhaps reflecting the influence of chromatin assembly.

Retroviral integration can be studied in vitro using high molecular weight nucleoprotein complexes from infected cell extracts (Brown et al., 1987; Bowerman et al., 1989; Lee and Coffin, 1990; Ellison et al., 1990; Farnet and Haseltine, 1990) or purified integrase protein (Craigie et al., 1990; Katz et al., 1990; Bushman et al., 1990) as the source of integration activity. We recently reported that minichromosomes (MCs) can be correctly and efficiently used as integration targets in an in vitro integration reaction mediated by viral nucleoprotein complexes (Pryciak et al., 1992). By mapping insertions in a limited number of cloned recombinants, we showed that there is no preference for integration into nucleosome-free regions of MCs, but a periodic distribution of integration sites in MCs suggested preferential use of the exposed face of the nucleosomal DNA helix. The approach used in those studies could not, however, address the fraction of sites affected by such mechanisms, the degree to which individual sites were affected, the exact organization of nucleosomal DNA at integration sites, or whether periodicity was limited only to nucleosomal regions.

Here we have developed an entirely different and more powerful approach to measure the frequency of integration at individual sites in both naked DNA and nucleosomal targets in vitro, using a novel polymerase chain reaction (PCR)-based assay. Neither nucleoprotein complexes nor purified integrases display a preference for nucleosome-free DNA over neighboring nucleosomal DNA. Instead, we find that the retroviral integration machinery sees many sites in nucleosomal DNA as more attractive than counterparts in nucleosome-free DNA or than the same sites in naked DNA. This represents the first clear example in which chromatin assembly creates preferential sites in DNA by increasing their reactivity.

RESULTS

Integration site distribution assay

In the in vitro integration reactions described here, the retroviral integration machinery inserts linear viral DNA into target DNA at many different positions. For a complete understanding of the factors that dictate the distribution of integration sites, it is desirable to analyze more insertions than there are potential sites in the target, in order to eliminate statistical fluctuations and allow comparison of even small differences in frequencies between different sites. We therefore developed a PCR-based assay that allows for a rapid and quantitative measurement of the distribution of thousands of integration events at once and with single nucleotide resolution. In this assay, amplification of recombinants between a primer in viral DNA and one in target DNA generates products with lengths dependent on the distances between those primer sites, and therefore on the sites of integration (Figure 1). The sizes and abundance of PCR products, as detected after electrophoresis in polyacrylamide gels, provide a measure of the distribution of integration sites. By varying the target-specific primer, even relatively large targets can be completely surveyed for use of integration sites.

To be useful, the PCR-based assay for integration products must be insensitive to unreacted viral and target DNAs, and each species of recombinant must give rise only to a single species of amplification product. A first example of the assay presents some important tests of these criteria (Figure 2A). Integration reactions were performed using murine leukemia virus (MLV) nucleoprotein complexes as the source of integration activity and either TRP1ARS1 (TA) minichromosomes (MCs) or naked TA DNA as the target (which are described more thoroughly in a subsequent section). PCR reactions were then performed on the integration products using an end-labelled MLV primer and either of two TA primers. It is evident that PCR products were generated only when a genuine integration reaction was allowed to occur, and not when integration reactions were carried

Figure 4-1. Schematic representation of the PCR-based insertion site distribution assay. During an integration reaction, viral DNA is inserted at many positions into an integration target (naked DNA, minichromosomes, etc.), generating a large population of recombinant products. The integration target is always present in large molar excess (1000 - 10,000 fold) over the viral DNA, so that integration is unlikely to occur more than once per target molecule; thus, sites are not effectively depleted during the course of the reaction. Two oligonucleotide primers (arrows), one complementary to viral DNA sequences and one to target DNA sequences, are used to amplify the recombinants. The lengths of the PCR products are determined by the distances between the two primers, determined in turn by the sites of integration. The asterisks indicate that the viral DNA primer is 5' end-labelled with ^{32}P, and the amplification products will therefore also be labelled at that end. Usually a low number of amplification cycles is used to promote quantitative amplification. The PCR products are loaded onto a gel, usually a denaturing polyacrylamide gel, and are visualized following electrophoresis by autoradiography. The distribution of sizes and intensities of PCR products provide a measure of the distribution of integration events.

Figure 4-1

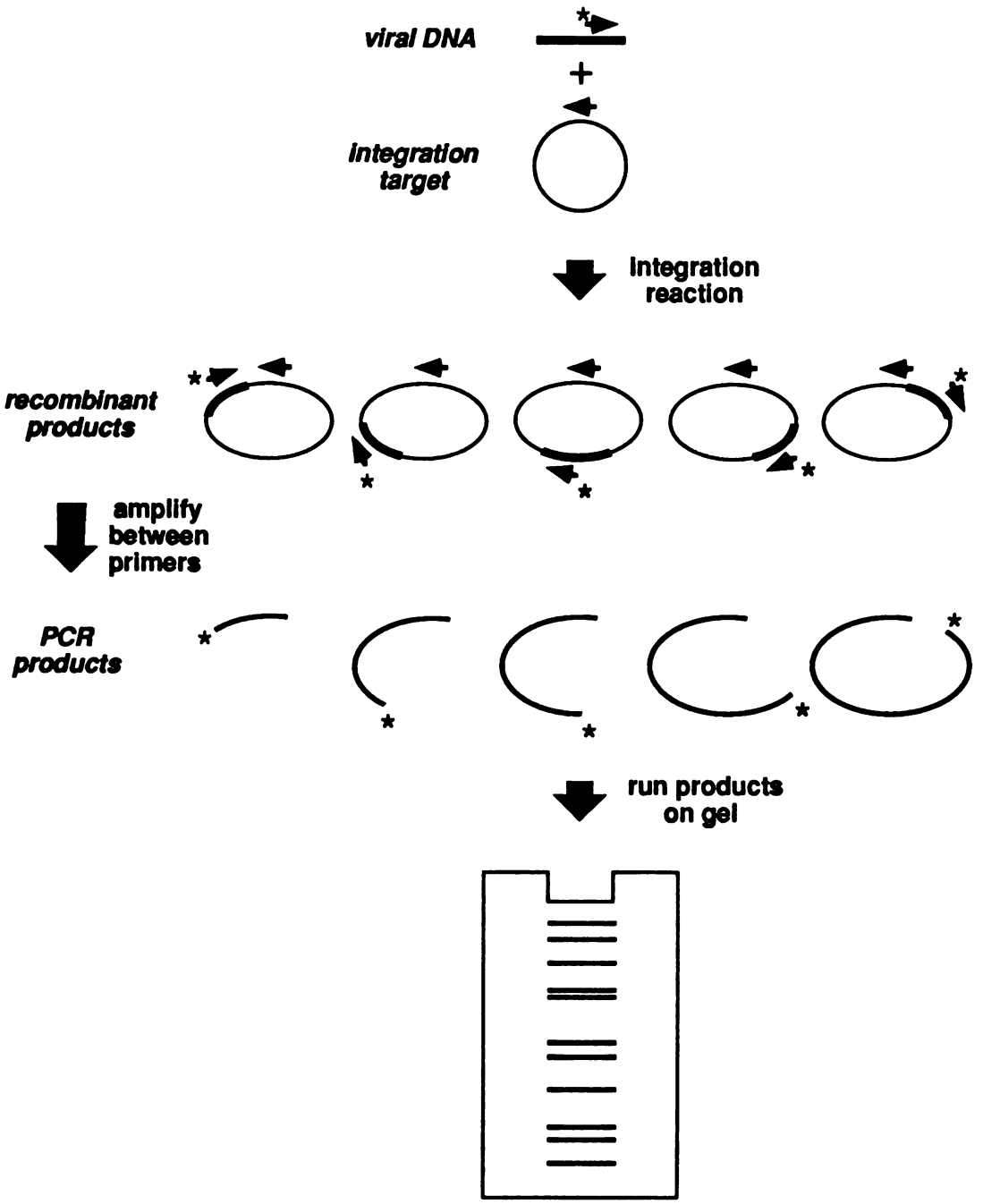
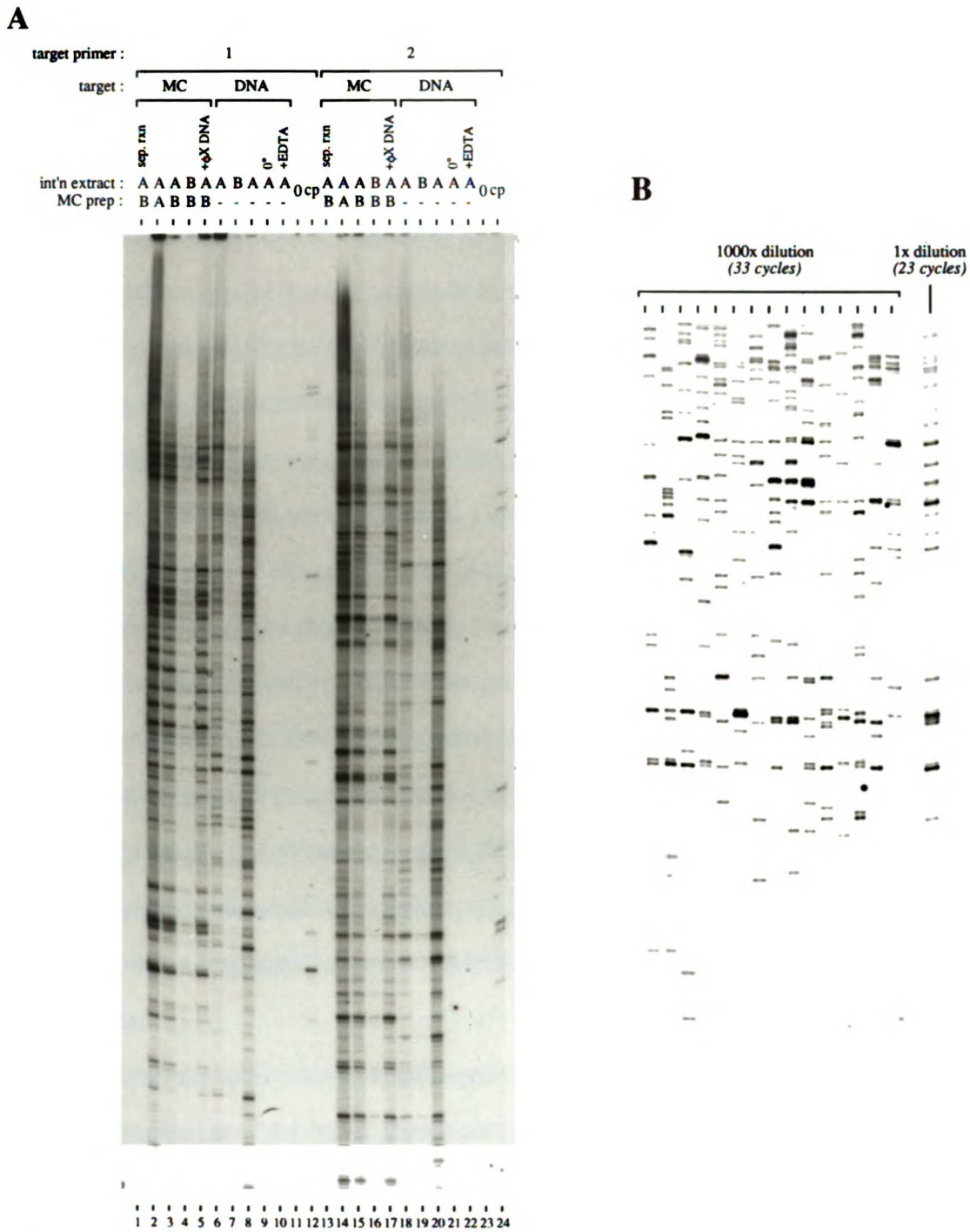

viral DNA
integration target
integration reaction
recombinant products
amplify between primers
PCR products
run products on gel

Figure 4-2. Tests of the PCR-based assay. **A.** Integration reactions were performed using two different preparations (A and B) of MLV-infected cell extract supplying the integration activity. Targets for the integration reactions were either naked TA DNA or two different preparations of TA MCs (designated A and B). In one case the integration extract and the target were incubated separately and then mixed together after stopping the reaction with SDS and EDTA (lanes 1, 13). Another integration reaction was performed with a second target present, naked ϕX174 DNA (500ng; lanes 5, 17). Integration reactions were also performed at 0°C (lanes 9, 21) or in the presence of 10 mM EDTA (lanes 10, 22). All integration reactions were carried out in the presence of spermidine except for those represented in lanes 8-10 and 20-22. The products of the integration reactions were subjected to PCR reactions using a 5' end ^{32}P-labelled viral DNA primer (MoU5L26) and either of two target primers: target primer 1 was TA1119+ (lanes 1-12), and target primer 2 was TA1273- (lanes 13-24). (The generation of PCR products was completely dependent upon the presence of a target DNA primer; not shown.) In addition, PCR reactions were performed using no template (indicated as 0; lanes 11, 23), or a pool of cloned recombinants (Pryciak et al., 1992; indicated as cp [clone pool]; lanes 12, 24). This clone pool consists of a mixture of cloned recombinant DNAs present in unequal amounts (see also legend to Figure 4). **B.** Fifteen parallel 1000-fold dilutions of the products of a single integration reaction with TA MCs as target were analyzed by 33 cycles of PCR; an undiluted sample was analyzed by 23 cycles (far right lane). Again, the labelled primer was MoU5L26, and the target primer was TA1119+. For the undiluted case, the expected number of integration events per site can be estimated by the following calculation: in a typical integration reaction, ~10^6 MLV DNA molecules are present and give rise to ~10^5 recombinants that, for a target size of ~10^3 bp, are distributed on average at 10^2 recombinants per site; typically 1/10 of an integration reaction is used in a PCR reaction, so there will be ~10 recombinants per site present on average; non-uniform integration could increase some to 100-1000 per site, depending on the degree of bias between sites.

Previous dilution experiments (not shown) showed that recombinants were lost upon diluting 10- to 1000-fold, corresponding to a range of average to 100-fold above the expected average frequency of ~10 recombinants per site. Thus, for the 1000-fold diluted samples analyzed here, most recombinants are likely to be present in single copy, with a chance that some will be present as a few copies.

Figure 4-2


A
target primer :
1
2
target :
MC
DNA
MC
DNA
sep. rxn
+ϕX DNA
0"
+EDTA
sep. rxn
+ϕX DNA
0"
+EDTA
int'n extract : A A A B A A B A A A A 0 cp A A A B A A B A A A A 0 cp
MC prep : B A B B B - - - - - B A B B B - - - - -
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24
B
1000x dilution
(33 cycles)
1x dilution
(23 cycles)

out on ice (lanes 9, 21) or in the presence of EDTA (lanes 10, 22), or if the integration extract and the target were incubated separately and then the DNAs mixed after the incubation (lanes 1, 13). Different target DNA primers, which should monitor integration over different regions of the target, gave different patterns of PCR products (compare lanes 1-12 with 13-24). The pattern of PCR products was reproducible using different preparations of either integration extract or MCs (lanes 2-4, 14-16), but was dependent upon whether the target was MC or naked DNA (compare lanes 2-5 with 6-8, and 14-17 with 18-20); this suggests that the pattern is dictated by events during the integration reaction and not simply by the nucleic acids present during the PCR reaction. Inclusion of a second target in the integration reaction (ϕX174 DNA) to generate additional unrelated recombinants did not alter the pattern of PCR products when monitoring only TA recombinants with the TA primer (lanes 5, 17). Finally, when PCR reactions were carried out using a small pool of cloned MLV-TA recombinants for which the exact integration site positions are known (Pryciak et al., 1992), bands only appeared at the expected positions for those recombinants, and no others were generated (lanes 12, 24). This argues that the multiple bands generated from the integration reaction products are due to integration at multiple positions, rather than production of multiple amplification products from each integration product. The clone pool and its PCR products serve in subsequent experiments as both positive controls and size markers, whose exact molecular correspondance to the products of the experimental integration reactions allows precise assignment of integration site positions.

Several control experiments indicated that the band intensity of each PCR product is a direct measurement of the initial abundance of the corresponding integration product. In one such experiment, we analyzed 15 parallel highly diluted samples of recombinants (Figure 2B), and observed two important points. First, there is a direct correlation between the strength of a band in the undiluted case and its likelihood of appearing in the diluted case. This argues that the intensity differences in the undiluted case are due to differences

in abundance, which arise from different frequencies of insertion at different sites. Second, in the highly diluted cases, the recombinants remaining after dilution are mostly present as single copies (see Figure 2B legend), and they give rise to amplification products of approximately equal intensities. This observation suggests that equal copies of different recombinants give equal signals and are therefore equally amplifiable.

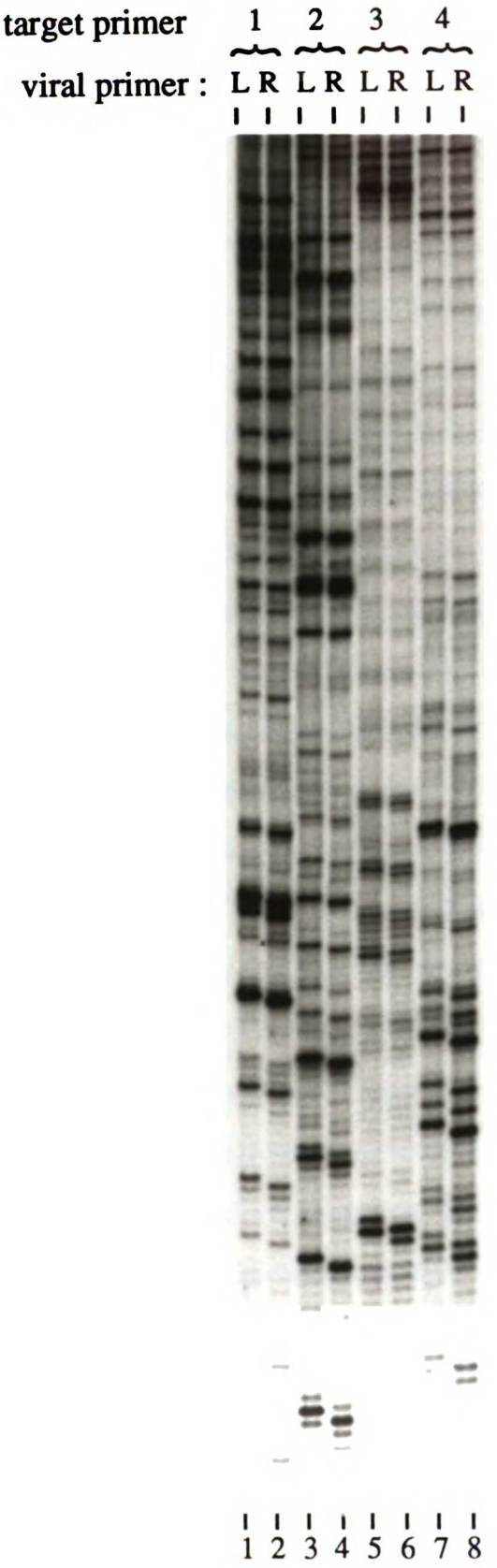
Changes in the integration reaction conditions (e.g. the presence or absence of spermidine, which is present in our standard reaction mixtures), or in the PCR conditions (e.g. annealing temperature or magnesium concentration) did not affect the distribution of PCR amplification products (data not shown). Other experiments (see Figures 4 and 5) demonstrated that strong and weak insertion sites retain their relative insertion frequencies when assayed using target primers on different sides of the site. Since all the evidence suggests that the PCR-based assay accurately measures the distribution of integration events, we will therefore refer to the patterns of PCR products as patterns of integration events.

MLV integration is orientation-independent

In the assay as described, integration recombinants are detected only in one of two possible orientations--that in which the viral DNA primer is pointed toward the chosen target DNA primer. In order to examine whether the non-uniform pattern of PCR products is due to non-uniform integration or to uniform integration but preferential orientation at different sites, we directly compared insertions in the two possible orientations into the TA MC target (Figure 3). This comparison was achieved by performing PCR reactions with four different TA primers and either of two viral DNA primers--one that extended from the "left" or U3 end (primer "L"), and the other from the "right" or U5 end (primer "R"). These primers were chosen such that their 5' ends were at the same distance from the end of viral DNA, so that amplification products of insertions at the same site would give products of the same size. The results (Figure 3) show that the distributions of insertions

Figure 4-3. MLV integration is orientation-independent. Integration products generated from a reaction using TA MCs as target were analyzed using several different combinations of primers. The labelled, viral DNA primer was either MoU5L26, which is equivalent to the plus strand of viral DNA and reads off of the U5 or "right" end of viral DNA (primer R), or SupF28, which is equivalent to the minus strand of viral DNA and reads off of the U3 or "left" end of viral DNA (primer L). The unlabelled, target DNA primer was TA1119+ (primer 1), TA1273- (primer 2), TA940- (primer 3), or TA1296+ (primer 4). The one-half to one nucleotide sized differences in mobility between the products generated by the L and R primers are most likely due to the different sequences present in the 73 nucleotides between the 5' ends of the primers and the recombinant junctions.

Figure 4-3



were identical in the two different orientations over a wide range of target sequences. Therefore, insertions are completely orientation-independent during MLV integration in vitro.

Distribution of integration sites in naked DNA and minichromosomes demonstrates sequence bias and exceptionally reactive nucleosomal sites

We next examined the distribution of integration sites into a minichromosome target, as compared to a naked DNA counterpart, in order to determine the effect of nucleosomes on the use of DNA as a target for integration (Figure 4). For this we used a homogeneous population of TA MCs purified from yeast cells and containing both nucleosomal and nucleosome-free regions; their nucleosome structure, their use as an integration target, and some limited mapping of insertion sites have been described (Pryciak et al., 1992). Integration reactions were performed in duplicate using either TA MCs or naked TA DNA as the target. We then performed PCR reactions, separately using six different target DNA primers to look over six regions of the TA target; with some overlap, these reactions measure the relative insertion frequency at every position in the 1453 bp target. The distributions of integration sites were compared to the known positions of nucleosomes on the TA MC target (Thoma et al., 1984; Pryciak et al., 1992).

These analyses, illustrated in Figure 4, reveal several interesting features. First, integration is non-uniform in naked DNA; the insertion frequencies of strong and weak sites can differ by as much as an order of magnitude (as estimated from differential exposure of autoradiograms and densitometry; data not shown). Thus the integration reaction is biased towards some sequences over others in target DNA, which could reflect direct interaction with bases or sensitivity to sequence-dependent structural fluctuations. A superficial comparison of the sequences present at numerous sites did not reveal any features obviously common to either strong or weak sites. Further experimental

investigations of this sequence dependence are not presented here, but some potential explanations are offered in the Discussion.

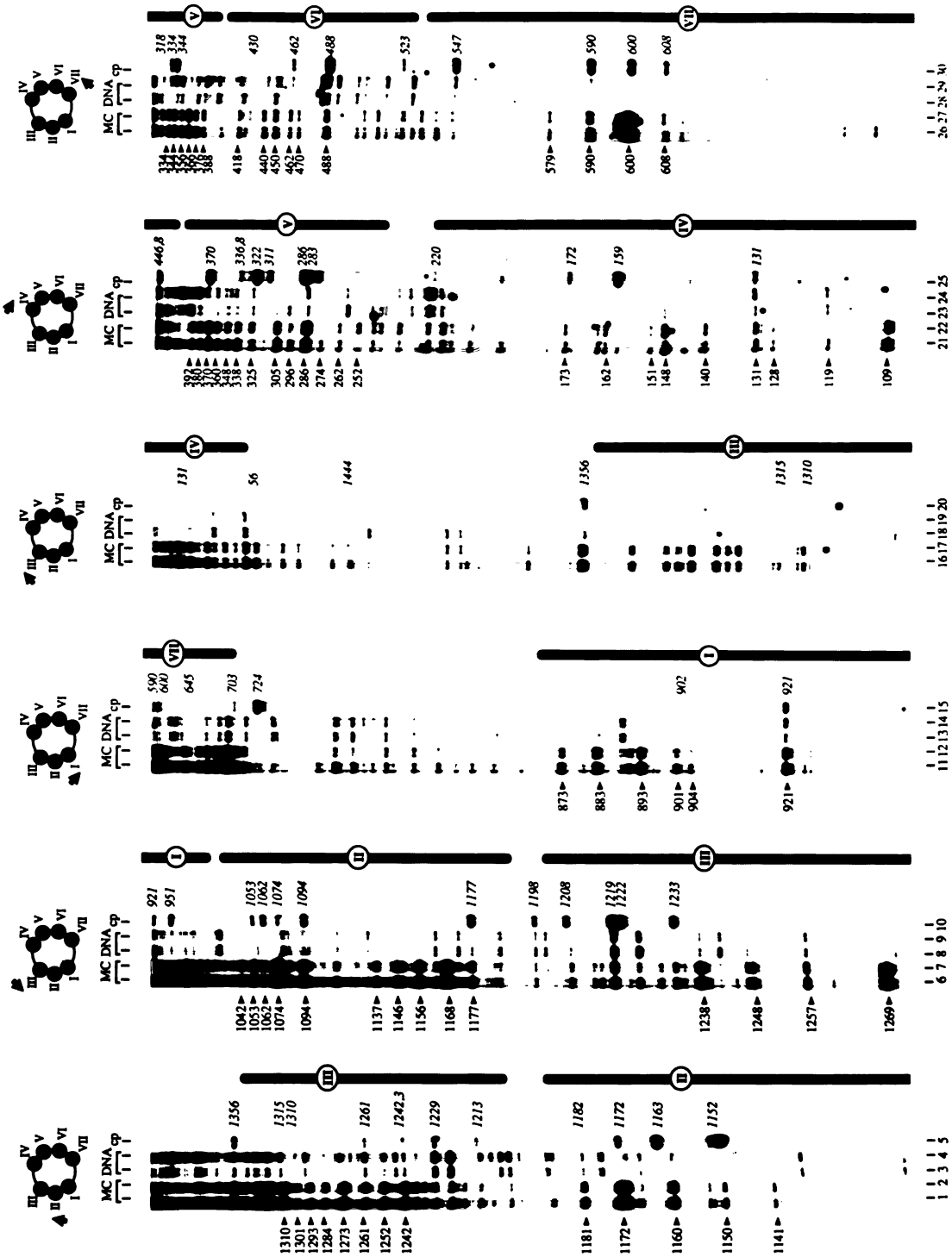
Integration into MCs is also non-uniform, but the distribution of integration sites differs markedly between MCs and DNA. The most apparent differences are at sites that are very strongly used in the nucleosomal regions of the MC; most of these sites are grouped in sets and spaced with a regular period of approximately 10 bp. For example, in nucleosome II of the TA MC target, a set of five periodically-spaced sites from positions 1141 to 1181 were observed to be preferentially used within that local region, and each site was used more frequently than when the target was naked DNA (lanes 1-4). The same set of preferred sites was seen when analyzed using a target primer on the other side of the region, a primer which monitors insertions into the complementary DNA strand. These sites now show up as high-frequency attacks at positions from 1177 to 1137 (lanes 6-9), with the apparent four nucleotide shift in position reflecting the four-base staggered cleavage of target DNA that occurs during the MLV integration reaction. The 10-bp periodic pattern of hyperreactive sites appears again further along nucleosome II in a set of preferential insertion sites from 1094 to 1042 (lanes 6-9). Similar sets were observed in nucleosome III from positions 1242 to 1310 (lanes 1-4); some of these, at positions 1238 to 1269, were confirmed for the opposite DNA strand (lanes 6-9). Such sets of MC-specific, regularly spaced, high frequency sites appear in each of the seven nucleosomal regions of the TA MC, with perhaps the strongest sets in nucleosomes II and III. The different frequencies of use among these preferential MC sites did not correlate with the relative frequency of use of the same sites in naked DNA (compare, for example, the relative frequencies in MCs of sites in the set from 1238 to 1269 with those of the same sites in naked DNA [lanes 6-9].)

Between these preferential sites, there occasionally appeared examples of sites inhibited in MCs--for example, between 1172 and 1181 (lanes 1-4) or between 1248 and 1238 (lanes 6-9). The preferential sites are not, however, sites that simply remained

Figure 4-4. Comparison of MLV integration site distributions in minichromosome and naked DNA targets demonstrates enhanced use of nucleosomal sites and sequence bias. Parallel, duplicate integration reactions were performed using either TA MCs or TA DNA (from deproteinized MCs) as the target. Equal amounts of the integration reaction products were analyzed by PCR, as was a clone pool (cp)--a pool of 89 recombinant clones (from integration into TA MCs) whose exact positions of insertion had been previously determined by sequencing (Pryciak et al., 1992; see also legend to Figure 2A). Note that each clone in the pool consists of an insertion in only one of two possible orientations, and therefore will only be detected by one of the two different target primer orientations (+ or -). The labelled viral DNA primer in all reactions was MoU5L26. The target primers were TA1119+ (lanes 1-5), TA1273- (lanes 6-10), TA940- (lanes 11-15), TA1296+ (lanes 16-20), TA95+ (lanes 21-24), and TA661- (lanes 26-30). The position of the target primer is indicated (arrow) at top according to a schematic depiction of the nucleosome organization of the TA MC (as determined by Thoma et al., 1984). In the schematic pictures, position 1 of the TA sequence is at top, and map positions increase to position 1453 in a clockwise direction. The nucleosomes are designated with roman numerals I through VII, based upon an extension of the nomenclature previously used (Thoma and Simpson, 1985). The PCR reactions represented in lanes 1-20 were run in parallel and their products separated on the same gel; those represented in lanes 21-30 were run in parallel with each other and their products separated on another gel. To the right of each set of PCR products, the positions of products of the clone pool are indicated adjacent to the appropriate band; included are some positions whose bands were faint in the original autoradiogram and may not be visible in the reproduction. Next to these clone pool positions are diagrammed the nucleosome positions (thin black ovals with roman numerals); the nucleosome borders are approximate, determined to roughly +/- 10-20 bp resolution by micrococcal nuclease mapping (Thoma et al., 1984). The positions of periodically spaced integration sites that

are highly preferred in the MC integrations are indicated to the left of each set of PCR products (arrowheads). The exact positions of these sites were assigned by comparison to both the clone pool and a sequencing ladder (not shown) run in the same gels; resolution necessary to assign the exact positions of those in the higher regions of these gels was achieved by running all products shown here for greater distances on other gels (not shown). Note that the assay detects recombinant junctions made to the strand of target DNA that is complementary to the target primer, not identical to it.

Figure 4-4



[redacted]

accessible in nucleosomal DNA; instead, they were used with greater frequency than the same sites in naked DNA. In most cases these sites were used between 2 and 10 times more frequently than the same sites in naked DNA, with the most extreme enhancements being 20- to 50-fold (as estimated from differential exposure of autoradiograms and densitometry; data not shown). Thus, the integration machinery does not prefer the nucleosome-free regions over nucleosomal regions. In fact, the reverse seems to be true: when the MC is the target, the most frequently used sites are those in nucleosomal regions. These high frequency sites were used more often than the same sites in naked DNA and more often than any sites in the nucleosome-free region of the MC.

In contrast to their distribution in nucleosomal regions, the distributions of integration sites were very similar between MCs and DNA in the two nucleosome-free regions (between nucleosomes I and VII, lanes 11-14, and between nucleosomes III and IV, lanes 16-19). Not only were the overall distributions similar, but the actual frequency of use at individual sites was nearly equal between MCs and DNA for most positions in the nucleosome-free regions. Thus, the nucleosome-free regions of the MCs are nearly indistinguishable from naked DNA to the integration machinery, and insertion site differences between MCs and DNA are mostly confined to those regions occupied by nucleosomes. In the MCs, some of the regions lying approximately between nucleosomes also behaved similarly to naked DNA, with decreased integration frequency and greater similarity between MCs and DNA. This was particularly evident in the region between nucleosomes II and III (lanes 1-4 and 6-9, in the vicinity of position 1200). Thus, the patterns of insertion site differences between nucleosomal and nucleosome-free (or internucleosomal) DNA help to confirm the expected MC chromatin structure.

All of the PCR reaction products shown in Figure 4 were also run in non-denaturing polyacrylamide gels, which show the insertion site distribution over the entire TA target for each target primer but at lower resolution. The results from these gels confirm that the strongest peaks were in nucleosomal regions of the MCs and that the band

intensities obtained using different target primers are directly and quantitatively comparable (data not shown).

Integration in nucleosomal DNA favors positions where the major groove is exposed

In the preceding experiments, the choice of target site by the retroviral integration machinery was found to be strongly influenced by the nucleosome. The 10-bp periodic distribution of nucleosomal integration sites is compatible with a model in which integration occurs preferentially at positions where both strands of target DNA face away from the nucleosome core. We subjected this model to a direct test by using a nucleosome for which the rotational orientation of the DNA helix was known.

For this test we used a well-characterized mononucleosome, which can be assembled in vitro by histone transfer onto a DNA fragment from the MMTV LTR. The rotational orientation of the DNA helix on the assembled nucleosome has been determined by both DNase I and hydroxyl radical cleavage (Perlmann and Wrange, 1988; Piña et al., 1990). We similarly assembled this mononucleosome and confirmed the rotational orientation by DNase I digestion (not shown; see Figure 5 legend and Experimental procedures). We then assayed integrations into the assembled and unassembled DNA, using target primers for both strands of the short target DNA (Figure 5A). The results show that, as in the TA MC, integration in nucleosomal DNA shows preferential use of sites that are distributed with a regular period of roughly ten base pairs. Some of the preferred mononucleosome sites (e.g. -185/-181, -165/-161, -136/-132, and -50/-46) are used relatively infrequently in naked DNA, whereas others (e.g. -155/-151 and -113/-109) are used as often in naked DNA. In addition, between these preferential sites are others that are inhibited in the mononucleosome (which, for simplicity, are not indicated). By comparing the positions of the integration sites with those cut by DNase I in the mononucleosome, we see that most of the preferred integration sites lie between pairs of

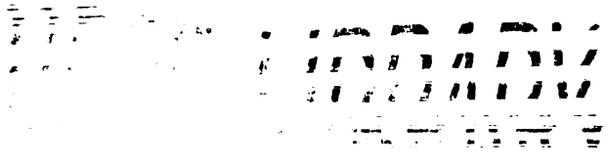
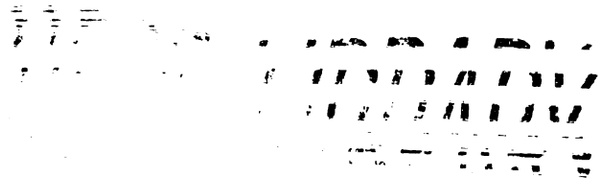
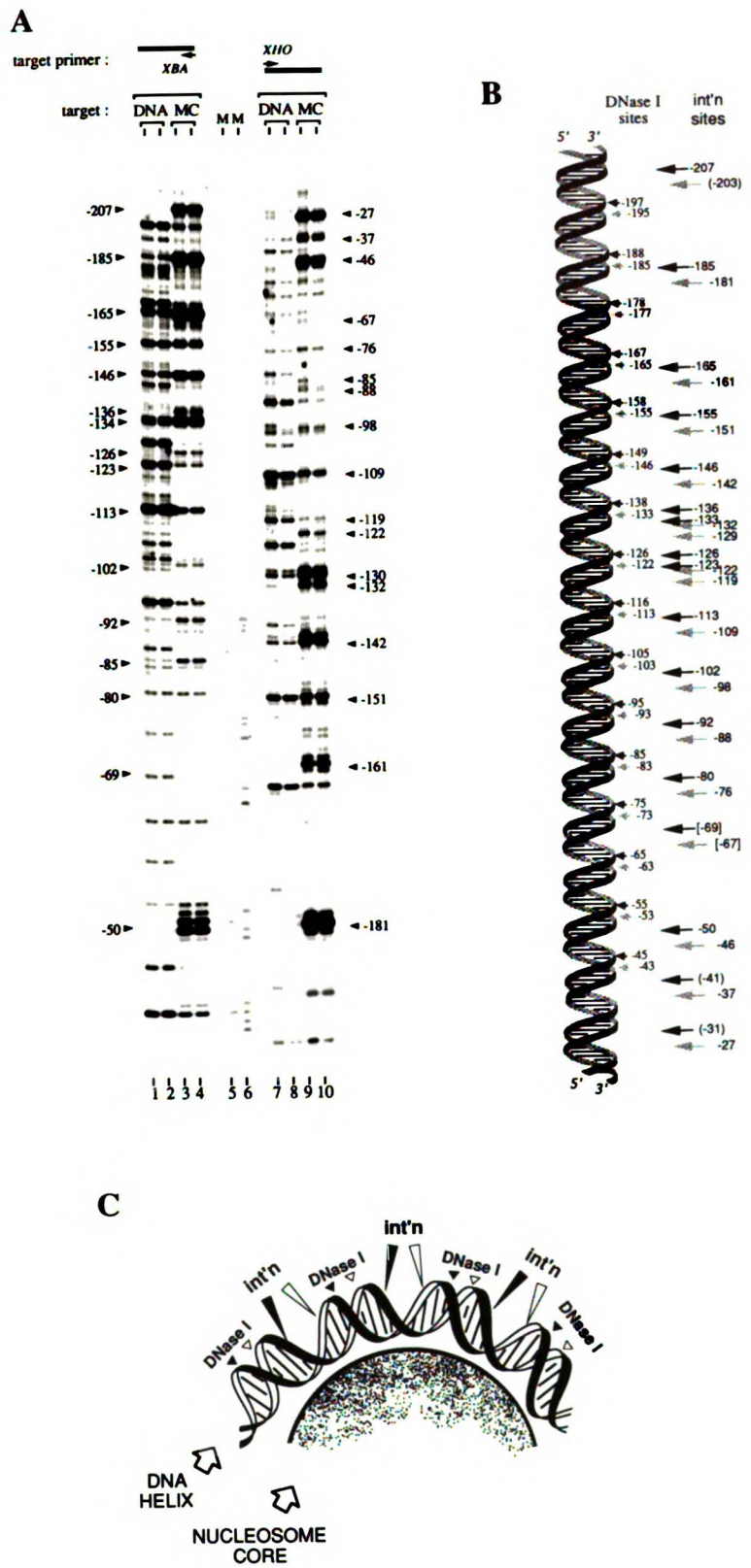


Figure 4-5. Comparison of MLV integration sites with DNase I cleavage sites in an MMTV mononucleosome indicates that favored insertion sites are found where the major groove faces outward from the nucleosome core. **A.** Distribution of integration sites in naked and mononucleosome DNA. Integration reactions were performed in duplicate using either unassembled DNA (DNA) or assembled and purified mononucleosomes (MC) as the target. PCR reactions were performed using a labelled viral DNA primer, MoU5L26. The target primers are schematically indicated at top (arrows) according to their positions on the linear target DNA; they were MTVXBAI (XBA; lanes 1-4) and MTVXHOI (XHO; lanes 7-10). Lanes 5 and 6 contain two different sequencing ladders that were used as size markers (M) to assign positions of integration sites. The positions of the sites used preferentially or remaining accessible in the mononucleosome are indicated (arrowheads); the numbers refer to positions relative to the transcriptional start site in the MMTV LTR. Note again that the assay detects recombinant junctions made to the strand of target DNA that is complementary to the target primer, so that the XBA primer detects junctions made to the "top" strand, and the XHO primer detects junctions made to the "bottom" strand. The results shown here were reproduced using two additional separate preparations of the mononucleosome (not shown). **B.** Direct comparison of DNase I cleavage sites and integration sites in the mononucleosome. The positions of integration sites (large arrows) were taken from panel A; junctions assayed using the XBA primer (black large arrows) and the XHO primer (speckled large arrows) are compiled. Note that each pair of black and speckled large arrows separated by four bases results from a single integration event, in which a four-base staggered cleavage is made in target DNA. Integration sites indicated in parantheses represent those that could not be directly measured for one strand (because they lie within one primer binding site) and instead have been extrapolated from the other strand. Positions indicated in brackets ([-69] and [-67]) represent those where sites on opposite strands are not separated by the expected 4 nucleotides. All positions indicated in panel A are shown here except for two--at position -85 on each strand; there were no corresponding



sites on the opposite strand for these and they were not well reproduced in other experiments. The positions in nucleosomal DNA of integration sites are compared to DNase I cleavage sites (small arrows) by placement on a schematic depiction of a DNA helix. The cut sites for DNase I shown here are taken directly from Piña et al., 1990 (see also Perlmann and Wrange, 1988). These sites were confirmed for one strand (those indicated by black small arrows) for the preparation of mononucleosomes used in the experiments shown in panel A. C. Model for integration into nucleosomal DNA. The comparison of positions shown in panel B is interpreted here in generalized form. Since DNase I cleaves across the minor groove of the exposed face of nucleosomal DNA and most of the integration sites (large arrows) lie between pairs of DNase I sites (small arrows), we conclude that integration is favored at positions where the major groove is on the exposed face of nucleosomal DNA; this orientation places both of the strands that are to be attacked at near-maximal distance from the nucleosome core. Again, as in panel B, each pair of (black and white) large arrows represents a single integration event.

Figure 4-5


A
target primer :
XBA
XHO
target :
DNA MC
M M
DNA MC
-207
-185
-165
-155
-146
-136
-134
-126
-123
-113
-102
-92
-85
-80
-69
-50
-27
-37
-46
-67
-76
-85
-88
-98
-109
-119
-122
-130
-132
-142
-151
-161
-181
1 2 3 4 5 6 7 8 9 10
B
DNase I sites
int'n sites
5' 3'
-207
(-203)
-197
-195
-188
-185
185
-181
-178
-177
-167
-165
165
-161
-158
-155
155
-151
-149
-146
146
-142
-138
-133
136
133
132
129
-126
-122
126
123
122
119
-116
-113
113
-109
-105
-103
-102
-98
-95
-93
-92
-88
-85
-83
-80
-76
-75
-73
[-69]
[-67]
-65
-63
-55
-53
-50
-46
-45
-43
(-41)
-37
(-31)
-27
5' 3'
C
int'n
DNase I
int'n
DNase I
int'n
DNase I
DNase I
DNA HELIX
NUCLEOSOME CORE

DNase I sites (Figure 5B). Since it is known that DNase I prefers to cut across the minor groove on the exposed face of the nucleosomal DNA helix (Drew, 1984; Drew and Travers, 1985; Suck et al., 1988), this comparison demonstrates that integration occurs preferentially at sites where the major groove of the DNA helix is on the exposed face of nucleosomal DNA (Figure 5C).

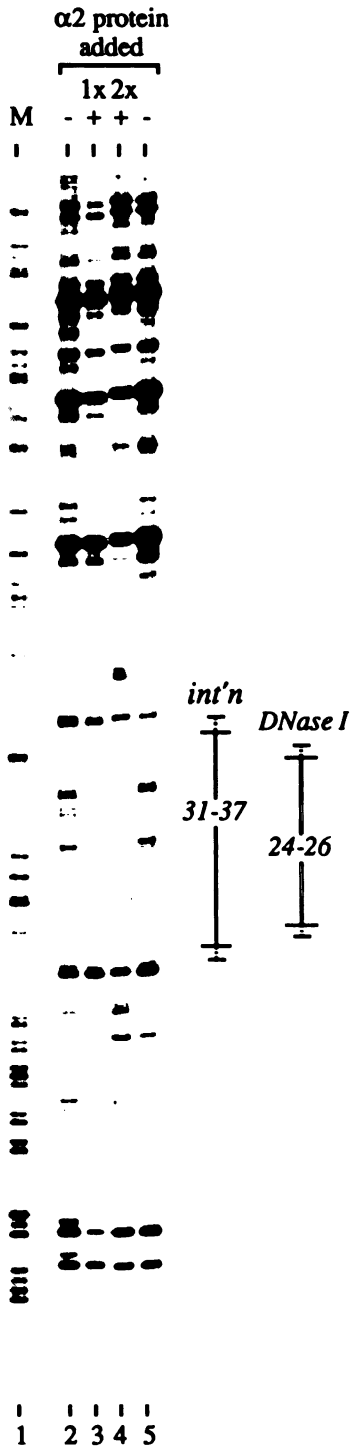
A site-specific DNA binding protein blocks integration into its binding site

Since the assembly of DNA around a histone core to form a nucleosome has a clear effect upon the use of that DNA as an integration target, we asked about the effects of non-histone DNA binding proteins on integration. Specifically we tested the effect of a site-specific DNA binding protein, the yeast transcriptional repressor α2 (Johnson and Herskowitz, 1985), on integration into the binding site for that protein. We used as an integration target a plasmid containing a 23-bp minimal binding site for the α2 protein. When α2 was pre-bound to target DNA, integration was severely blocked in the region corresponding to the α2 binding site (Figure 6). Because integration was non-uniform in the unbound target, the size of the blocked region could not be precisely measured; at least 31 bp and no more than 37 bp were blocked. Since the DNA was masked by the presence of bound protein, the integration machinery does not appear to have a displacement or clearing activity competent to remove the α2 protein. Remarkably, the size of the masked region, or integration "footprint", was similar in size (roughly 31 bp) to that of the DNase I footprint (roughly 24 bp; Sauer et al., 1988), even though the integration machinery used in these experiments consists of a very high molecular weight, multi-component nucleoprotein complex (Bowerman et al., 1989), while DNase I is a monomeric protein (Suck et al., 1984).

Figure 4-6. Integration is blocked in a region occupied by a site-specific DNA binding protein. The target DNA used for integration was a plasmid (pUC-Δ13) that contains a 23-bp insert encompassing a minimal binding site for a dimer of the α2 protein (see Experimental Procedures). This DNA was left unbound or was prebound with either of two concentrations of α2, and was then used as a target in an integration reaction (lanes 2-5). PCR reactions used a labelled viral DNA primer (MoU5L26) and an unlabelled target DNA primer (M13[-47]24). A sequencing ladder (M) was run as a size marker in the same gel (lane 1) as the PCR products to facilitate the assignments of integration site positions. Integration was blocked in the presence of bound protein in a region corresponding to the binding site. The size of the masked region is indicated in nucleotides at right; it is slightly ambiguous because some of the peripheral sites were so infrequently used in the unbound DNA. A similar ambiguity exists for the size of the DNase I footprint of the same protein-DNA complex (Sauer et al., 1988), which is indicated for comparison.

Figure 4-6

YEAST ALPHA-2 REPRESSOR BLOCKS MLV INTEGRATION AT ITS BINDING SITE



The source of integration machinery affects target site selection

We next conducted experiments to determine whether the integration site preferences observed in naked and nucleosomal DNA are influenced by the integration machinery, as well as by the target. Integration site distributions into either TA MCs or DNA were compared for the MLV nucleoprotein complex (npc), which was used in all previous experiments, and purified integrase protein (IN) from two different species of retrovirus--MLV and HIV (Figure 7). It is important to note that for these experiments, in contrast to the previous experiments, relative efficiencies of site selection between different integration reactions cannot be directly compared, since the amounts of integration products amplified were chosen to give similar overall levels of signal in all lanes of the gel.

An example of such comparisons (Figure 7) shows, first of all, that the MLV npc insertion site distributions in both MCs and DNA can be recapitulated by MLV IN (compare lanes 1,2 and 7,8 with 3,4 and 9,10, respectively). The few differences observed could reflect either the additional proteins present in the npc or the nonviral sequences attached to the MLV IN, which is a fusion protein. Aside from these occasional differences, the MLV npc and IN distributions were very similar over either the entire TA MC or DNA target (Figure 7, and additional data not shown). Second, it is apparent that nucleosomes create high frequency sites for the purified IN proteins of both viral species, just as they do for the MLV npc. Many of these high-frequency sites, or regions, are shared among all three machineries (open arrowheads, lanes 1-6). Third, despite many similarities, the distributions of insertion sites are not identical in either DNA or MCs when MLV IN and HIV IN are compared as the source of integration activity (lanes 3-6, 9-12, 15-18, and 21-24). This shows that site selection is not solely dictated by the target and that the integration machinery must also play a role. The different species of IN can even respond differently to the influence of nucleosomes, as evidenced by some MC sites very strongly preferred by HIV IN (closed arrowheads, lanes 1-6) that are not especially

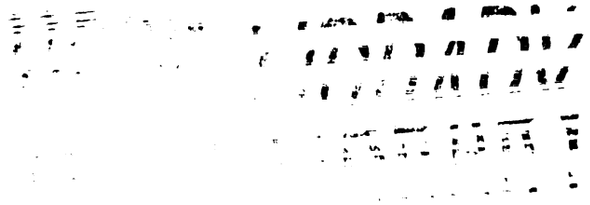
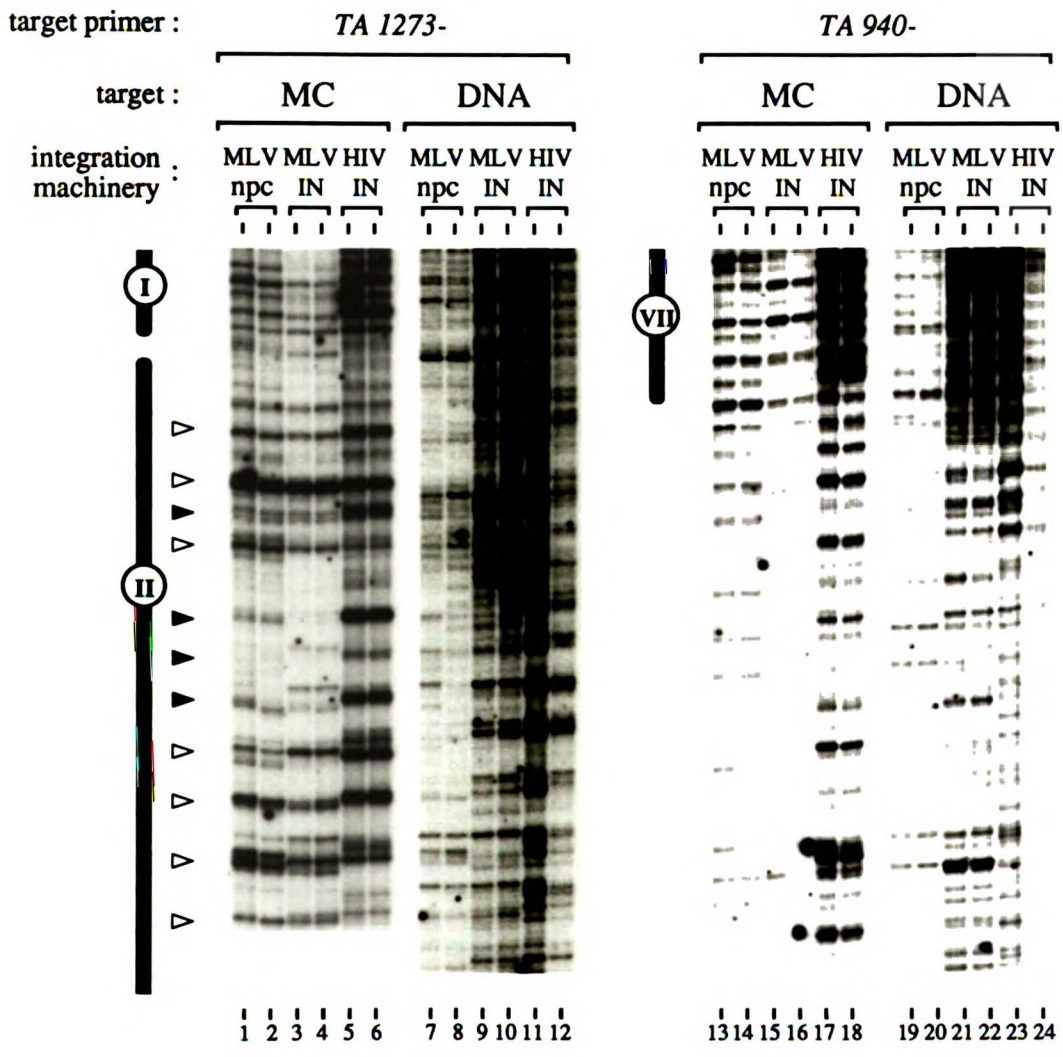


Figure 4-7. Comparison of insertion site distributions using different sources of integration activity. Integration reactions were performed using either TA MCs or TA DNA as the target. The integration machinery consisted of MLV nucleoprotein complexes (MLV npc) in infected cell extracts, purified E. coli-produced MLV IN-GST fusion protein (MLV IN), or purified yeast-produced HIV IN (HIV IN). The MLV npc integrated its endogenous viral DNA, whereas the purified INs integrated exogenously-provided double-stranded oligonucleotides corresponding to the right ends of viral DNA (see Experimental Procedures for details). For the PCR reactions, the labelled viral DNA primer was U5MLVA27 for MLV npc and MLV IN reactions, and HIVU527+ for HIV IN reactions; these were chosen so that the distance from the 5' ends of the primers to the recombinant junctions was the same for the products of all three machineries. The target DNA primers are indicated at top. All PCR reactions were carried out in parallel and run on the same gel. Open arrowheads indicate preferred sites in MC targets shared among all three activities, whereas closed arrowheads denote sites in MC targets that are especially preferred by HIV IN and much less so by the MLV activities. The positions of nucleosomes are shown as thin black ovals as in Figure 4.

Figure 4-7

PATTERN OF INTEGRATION INTO DNA AND MINICHROMOSOMES IS AFFECTED BY SOURCE OF INTEGRATION ACTIVITY



[redacted]

preferred over other sites by MLV IN. Finally, both MLV IN and HIV IN, like the MLV npc, show no preference for the nucleosome-free regions in the TA MC (lanes 13-18, and additional data not shown). Thus, the ability to integrate into nucleosomal DNA at least as efficiently as into nucleosome-free DNA is a property inherent to the IN protein and not dependent upon the non-IN components of the npc. In addition, this important property is not affected by differences in the reaction conditions used for the three sources of integration activity.

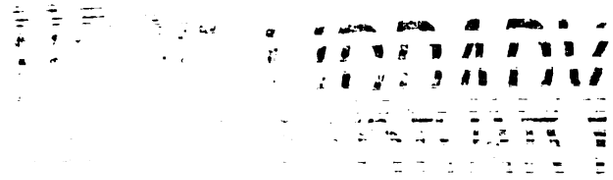
DISCUSSION

In this paper we describe a novel assay for rapidly measuring the distribution of retroviral integration sites, and we use it to study retroviral target site selection and the effects of chromatin packaging on DNA accessibility. This analysis can be performed over any chosen stretch of target DNA, with single-base resolution if desired, and in situations where the number of integration events is greater than the number of different target sites, thus eliminating statistical fluctuation. The results bear upon the accessibility of DNA in chromatin, the factors that contribute to retroviral target site selection, and the utility of retroviral integration as a probe for changes in the physiological state of DNA.

Integration into chromatin

Our results unambiguously demonstrate that nucleosome-free regions of chromatin are not preferred by the integration machinery, consistent with our earlier findings based on limited insertion site mapping (Pryciak et al., 1992). Nucleosomal DNA is indeed used as an integration target, and the distribution of integration sites is dramatically different in nucleosomal DNA than in naked DNA. Nucleosome-free regions, on the other hand, show similar integration patterns to naked DNA. Importantly, the current experiments show that integration machineries display a marked preference for nucleosomal regions (or at least for particular sites in those regions) over nucleosome-free regions. These properties are true even for purified integrase (IN) molecules from two species of retrovirus, as well as for the nucleoprotein complex from infected cells. Such properties contrast sharply with those of enzymes such as micrococcal nuclease and DNase I, which are commonly-used tools for studying chromatin structure, showing that different classes of behavior can be exhibited by different machineries.

Choice of integration sites is affected by wrapping of DNA into nucleosomes, in a manner that frequently results in a 10-bp periodic pattern of preferred sites, caused by both



decreased use of some positions and increased use of others. This pattern is observed in seven different nucleosomes in the TA MC (Figure 4) and in an unrelated mononucleosome (from MMTV sequences; Figure 5), but not in nucleosome-free regions of chromatin. Integration frequency varies among these preferred sites, and is therefore likely to be modulated by the actual sequences at the integration sites. We were able to directly test and prove the model that this 10-bp periodicity results from favored integration at positions where the major groove is on the exposed face of the nucleosomal DNA helix, by comparing DNase I cleavage and integration patterns into the MMTV mononucleosome (Figure 5). There are at least two classes of model that could account for access of the integration machinery to the strands facing each other across the major groove: one in which the integration machinery descends into the groove and attacks outwards at the neighboring strands, and another in which two opposed halves of the integration machinery lie upon the surface of the helix on either side of the groove and attack the strands that lie across from each other. We have no compelling evidence to distinguish between these models, but a study of the relative role of sequence at the center versus the periphery of the insertion site in determining site preferences in naked DNA may assist in resolving this issue.

Perhaps the most striking feature seen upon integration into nucleosomal DNA is the strength of the preferred, periodically-spaced sites. In most cases these sites do not simply remain accessible, but instead are actually used more frequently in nucleosomal DNA than the same sites in naked DNA. Such a true hypersensitivity--i.e. more sensitive than in naked DNA--is rarely seen in studies of chromatin accessibility. Nuclease cleavage sites, for example, usually become insensitive or, at best, remain almost as sensitive in chromatin as in naked DNA. In the most directly applicable case, DNase I digestion of nucleosomal DNA reveals that periodically-spaced sites remain sensitive to cleavage; but cleavage at such sites often requires roughly ten-fold more concentrated nuclease than for the same site in naked DNA (e.g. Simpson and Stafford, 1983; Drew and Travers, 1985;

Piña et al., 1990; Taylor et al., 1991). Similarly, site-specific DNA binding proteins able to bind nucleosomal DNA usually do so with reduced affinity (Piña et al., 1990; Taylor et al., 1991). In contrast, we observe hypersensitive integration sites without the need to analyze more events into MCs than into DNA, by comparing parallel reactions in which integration occurs with nearly equal efficiency into MCs and DNA. While such hypersensitive sites are not present at every turn of the nucleosomal DNA helix, they are a predominant feature in both of the chromatin targets tested.

Using a minichromosome in which we could compare insertions into seven nucleosomal and two nucleosome-free regions, we found that the most frequently used sites are in nucleosomal regions. These findings have important ramifications for retroviral integration in vivo, where the integration machinery may be faced with a choice between nucleosomal and nucleosome-free DNA. The different regions can be regarded as competing with each other for the integration machinery. Our results show that many sites can compete better when incorporated into nucleosomes than do their counterparts that remain nucleosome-free. To our knowledge, this is the first reported example of general increased reactivity or attractiveness of DNA sites resulting from chromatin assembly. This implies that chromatin assembly can not only block some sites but also enhance others, leading to a larger differential of accessibility than achievable by either inhibition or stimulation alone.

Factors contributing to integration site selection

What accounts for the increased frequency of integration, relative to naked DNA, at the periodically-spaced sites in nucleosomal DNA? An MC-specific increase in the number of same-site insertions, observed in our previous work, was suggested to result from a redistribution of integration events among the fewer remaining accessible sites in the MC, leading to more events per site (Pryciak et al., 1992). This scheme predicts that all remaining accessible sites, both nucleosomal and nucleosome-free, would receive an

increased number of events. We now show this simple model to be incorrect, since integration frequencies are increased at many sites in nucleosomal DNA but are virtually unchanged in nucleosome-free DNA. This restriction of high-frequency sites to nucleosomal regions could result either from a more local redistribution, such that attacks intended for inaccessible positions get redirected to the nearest accessible position, or from a mechanism by which sites away from the nucleosome core are made even better sites than they were in naked DNA. The "local redistribution" model would seem limited to a 10-fold increase at most in insertion frequency. Since some positions sustain a 10- to 50-fold increase, sites may indeed be made better in the nucleosome than in naked DNA.

How could nucleosomal sites be made better than naked DNA sites? It is possible that a direct interaction between the integration machinery and histones is involved; if so, IN protein would appear to be sufficient to mediate such an interaction. Alternatively, the explanation may lie solely in the structural consequences of wrapping DNA around the nucleosome core. For instance, it has been proposed that expansion of the minor groove during DNA bending can explain the increased frequency of DNase I cleavage at sites on the outside face of the bend, and that sequence-dependent variations in minor groove width underlies site preferences during DNase I cleavage of naked DNA (Drew and Travers, 1984; Drew and Travers, 1985; Suck et al., 1988; Drew et al., 1990; Travers and Klug, 1990). Integration sites could analogously appear more attractive to the integration machinery in nucleosomal DNA than in naked DNA because of bend-induced expansion of the major groove on the exposed face of the helix. Sequence-dependent variation in major groove width may similarly govern non-uniformity of integration sites in naked DNA. In any case, these mechanisms could presumably operate by promoting either an initial interaction step or the reaction step; none of our current data distinguishes between these possibilities. In contrast to histone-bound DNA, DNA bound by the α2 repressor did not show hypersensitive integration sites in the region bound. It seems most likely that more extensive α2-DNA contacts accounts for this difference, but it is also possible that an

absence of nucleosome-like DNA bending or of specific protein-protein interactions with histones is involved.

While much of our evidence shows that the state of the target can dramatically influence integration site choice, we also find that the integration machinery plays a role. In a direct comparison, purified MLV IN showed a distribution of integration sites very similar to the MLV npc, demonstrating that the general features of site selection described for the npc are not dependent upon components other than IN protein itself. On the other hand, comparing integrases from different species of retrovirus, MLV and HIV, showed many differences in their insertion site distributions, demonstrating that the integration machinery can clearly influence the choice of integration sites. Curiously, the existence of strong nucleosomal sites specific for HIV IN (Figure 7) suggests a cooperative interplay between the nucleosome and the integration machinery during the integration at high-frequency sites. Other experiments showing that HIV IN integrates two divergent sequences (HIV U5 and U3 ends) with exactly the same distribution of target sites argue that it is the integrating protein, rather than the integrating nucleic acid, that determines site preferences (Leavitt et al., 1992).

Retroviral integration as a probe for DNA accessibility

The data presented here show that retroviral integration is sensitive to the organizational state of the target DNA. In both the TA MC and the MMTV mononucleosome, the rotational orientation of nucleosomal DNA strongly influences target site choice. In fact, the integration site distribution data for the TA MCs (Figure 4) argue that all seven of the in vivo-assembled TA nucleosomes have preferred rotational orientations, which has not been previously investigated. (Rotational orientations of some nucleosomes in a derivative of the TA plasmid have recently been examined by Shimizu et al., 1991.) In addition, the positions of nucleosomal and nucleosome-free regions of the TA MC are evident in the integration patterns, when comparing MC and DNA targets. We

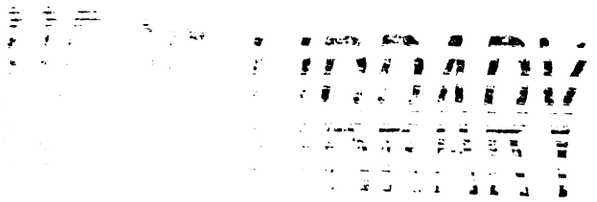
also show that the integration site distribution is sensitive to the presence of non-histone bound factors, such as the α2 repressor (Figure 6). These observations together argue that retroviral integration can actually be used as a probe for chromatin organization.

Importantly, retroviral integration is a process that can be studied in vivo as well as in vitro. It is rare to be able to study effects of chromatin organization on DNA accessibility using a probe that conducts functional interactions with chromatin in the cell. Integration may be sensitive to additional physiological processes, such as replication and transcription, in ways that could be revealed by comparison between in vitro and in vivo integration site distributions. Work in progress shows that the assay presented here can be used to study chromatin accessibility during retroviral integration in vivo into episomal minichromosomes. The results are consistent with our in vitro studies, but they also suggest that greater degrees of bias can exist between sites in vivo than between the same sites in vitro (unpublished observations). These methods may eventually be extendable to chromosomal targets. We have not yet examined the effects of histone H1-promoted condensation of nucleosomal DNA on integration, but assembly of higher-order chromatin in vitro (Hirano and Mitchison, 1991) may be of great value in this regard.

Nucleosomes and chromatin packaging in general are usually viewed as inhibitory to DNA access. Considering the likelihood that the majority of DNA in the nucleus is assembled into nucleosomes, it is important to know how prevalent such inhibition is. While repressive effects have some clear examples (in situations such as transcriptional initiation and factor binding), the present results and recent studies of some site-specific DNA binding proteins (NF-1 and glucocorticoid receptor {Piña et al., 1990}, HSF and GAL4 {Taylor et al., 1991}) suggest that different classes of eukaryotic DNA binding proteins exist: those that are, and those that are not inhibited by nucleosomes. The latter class may even include some that prefer nucleosomal DNA, or at least some sites in nucleosomal DNA, as evidenced by the present studies on retroviral integration. Given the wide variety of functions likely to operate on eukaryotic DNA, it seems reasonable to

[redacted]

suppose that some can utilize, and perhaps take advantage of, the more highly assembled states of chromosomal DNA.



Experimental Procedures

Integration reactions

All integration reactions (except for some of those shown in Figure 7) were performed using MLV-SupF infected cell extracts as described (Pryciak et al., 1992; Brown et al., 1987). Briefly, 20 µl of cold extract was combined with the appropriate integration target (40 ng TA MCs or DNA; or 1 ng MMTV mononucleosome or DNA), brought to a final reaction volume of 30 µl, and incubated for 1 hr at 37°C. Nucleic acids purified from such a reaction were resuspended in 20 µl of TE, and usually 0.5 to 1.0 µl of this was used for a single PCR reaction. All extract-mediated integration reactions except for those shown in Figure 6 and those indicated in the legend to Figure 2 were performed in the presence of 15 mM spermidine (Pryciak et al., 1992). For most experiments, parallel duplicate integration reactions were performed to determine the reproducibility of any observation.

Integration reactions in the presence of bound α2 protein (described in Figure 6) were performed as follows: 300, 150, or 0 ng of purified α2 protein (a gift from C. Goutte and A.D. Johnson, UCSF) was incubated with 25 ng of pUC-Δ13 DNA (a plasmid containing a 23bp minimal α2 binding site insert; Smith and Johnson, 1992) in the presence of 150 mM NaCl, 10 mM HEPES, pH 7.5, and 2.5 mM $MgCl_2$ in a volume of 10 µl for 0.5 hr at 25°C. Then 20 µl of integration extract was added, and incubation proceeded for 1 hr at 30°C.

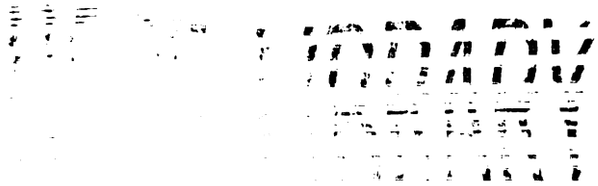
Integration reactions using MLV IN were carried out using an uncleaved MLV IN-Glutathione-S-transferase (GST) fusion protein purified from E. coli (a gift from I. Dotan and P. Brown, Stanford University). Roughly 0.3µg of the MLV IN-GST fusion was added to the integration target in the presence of 15 nM double stranded MLV att-site oligonucleotides (annealed U5MLVA and U5MLVB) and 20 mM MOPS, pH 7.2, 10 mM DTT, 20% glycerol, 4 mM $MnCl_2$, and 65 mM NaCl, in a volume of 10 µl; these reactions contained as integration targets 20 ng TA MCs or 2 ng TA DNA. (Less DNA than MCs

was used because excess DNA inhibited the reaction, presumably by competing with the oligonucleotides for MLV IN.)

Integration reactions using HIV IN were carried out using HIV-1 IN produced in yeast (Leavitt et al., 1992). Roughly 0.2 μg of the HIV IN protein was added to integration targets (in amounts identical to the MLV IN-GST reactions above) in the presence of 15 nM double stranded HIV att-site oligonucleotides (annealed HIVU529+ and HIVU529-) and 50 mM MOPS, pH 7.0, 1 mM DTT, 15 mM $MnCl_2$, 10 μg/ml BSA, and 10 mM NaCl, in a volume of 10 μl. Incubations of both MLV IN and HIV IN reactions were for 1 hr at 30°C; nucleic acids were prepared as for extract-mediated reactions (Pryciak et al., 1992) and resuspended in 20 μl of TE. Typically 0.02 to 0.1 μl of these resuspended products was used per PCR reaction.

PCR assay for integration site distribution

Ice-cold integration reaction products were added to an ice-cold PCR reaction cocktail, giving final reaction concentrations of 50 mM KCl, 10 mM TRIS, pH 8.8, 1.5 mM $MgCl_2$, 0.1 mg/ml gelatin, 0.2 mM dNTPs, 1 U Taq polymerase (AmpliTaq; Cetus-Perkin Elmer), 0.25 μM target primer, 0.2 μM unlabelled viral DNA primer, 0.05 μM 5' end ^{32}P-labelled viral DNA primer, in a final reaction volume of 20 μl; the mixture was overlaid with 20 μl of mineral oil and kept on ice until ready to begin thermal cycling. The reaction mixtures were amplified for 25 cycles (except for those in Figure 2B) by thermal cycling at 94°C for 1 min., 55°C for 1 min., and 72°C for 2 min. (except for those in Figure 5, which used 50°C rather than 55°C). Tubes containing reaction mixtures were transferred directly from ice to the 94°C block in the first cycle. Following amplification, the samples were extracted once with 60 μl $CHCl_3$, and the 20 μl aqueous phase was collected and stored frozen. For analysis, 6 μl PCR products was mixed with 4 μl of 95% formamide/10 mM EDTA/bromophenol blue/xylene cyanol, heated at 80°C for 15 min., and then loaded onto a 6% polyacrylamide (19:1 acryl:bis)/urea/1xTBE denaturing gel for



high resolution, short range analysis. Alternatively, the PCR products can be made 20% in sucrose and loaded onto a 5% polyacrylamide (30:0.8 acryl:bis)/1xTBE nondenaturing gel for lower resolution, but longer range analysis.

The product of the integration reaction, which serves as the template in the PCR reaction, is actually an intermediate in which the 3' ends of viral DNA have been joined to the 5' ends of the cleaved target DNA, but the other strands remain unjoined (Fujiwara and Mizuuchi, 1988; Brown et al., 1989). Extension across this gapped intermediate, if left unrepaired, in the first cycle is possible from the target DNA primer but not by the viral DNA primer; subsequent cycles would then allow extension from the viral DNA primer. In practice, however, in the brief amount of time spent at extension temperature during warming of the reaction mixture from 0°C to 94°C, the Taq polymerase repairs the gap by strand displacement synthesis using the 3' end of the cleaved target DNA as a primer (unpublished observations).

The positions of integration sites in target DNA were deduced from the sizes of PCR products (measured by comparison to size markers in gels) by calculating the distance from the 3' end of the target primer as follows: (distance from 3' end of target primer) = (PCR product size) - [(distance from 5' end of viral DNA primer to recombinant junction) + (length of target primer)]. Insertion is said to be at position *n* of target DNA for recombinant junctions formed by cleavage of target DNA between bases *n* and *n+1* (applied separately for each strand).

Oligonucleotides

The following oligonucleotides (prepared by the Biomolecular Resource Center, UCSF), indicated by their names, sequences, and map locations (in parantheses), were used in this study:

MoU5L26 5' CGACTTGTGGTCTCGCTGTTCCTTGG (8335-8360 MoMLV)

SupF28 5' CCCCCACCACCATCACTTTCAAAAGTCC (171-198 SupF)

TA1119+ 5' GTACATACCTCTCTCCGTATCCTCG (1095-1119 TRP1ARS1)

TA1273- 5' ACTTGACGACTTGAGGCTGATGGTG (1297-1273 TRP1ARS1)

TA940- 5' TGGTAAAAGTCAACCCCCTGCGATG (964-940 TRP1ARS1)

TA1296+ 5' ACACCATCAGCCTCAAGTCGTCAAG (1272-1296 TRP1ARS1)

TA95+ 5' ACGTGATTAAGCACACAAAGGCAGC (71-95 TRP1ARS1)

TA661- 5' CGCCATTTAATCTAAGCGCATCACC (685-661 TRP1ARS1)

M13(-41)24 5' CGCCAGGGTTTTCCCAGTCACGAC (6343-6320 M13mp18)

MTVXBAI 5' AATCAGCACTCTAGAATATCTTG (-14 to -36 MMTV LTR)

MTVXHOI 5' CCTTGCGGTCTCGAGGGCTT (-217 to -198 MMTV LTR)

U5MLVA 5' TTGACTACCCGTCAGCGGGGGTCTTTCATT (8380-8409 MoMLV)

U5MLVB 5' AATGAAAGACCCCCGCTGACGGGTAGTCAA (8409-8380 MoMLV)

U5MLVA27 5' TGACTACCCGTCAGCGGGGGTCTTTCA (8381-8407 MoMLV)

HIVU529+ 5' TTTAGTCAGTGTGGAAAATCTCTAGCAGT (9709-9738 HIV-1 SF-2)

HIVU529- 5' ACTGCTAGAGATTTTCCACACTGACTAAA (9738-9709 HIV-1 SF-2)

HIVU527+ 5' TTTAGTCAGTGTGGAAAATCTCTAGCA (9709-9736 HIV-1 SF-2)

Minichromosome preparation and mononucleosome assembly

TRP1ARS1 (TA) minichromosomes were prepared exactly as described (Pryciak et al., 1992), and naked TA DNA was prepared from these minichromosomes by deproteinization. MMTV mononucleosomes were assembled by nucleosome transfer using chromatin fragments prepared as described (Piña et al., 1990) except that mouse liver was used instead of rat liver. The DNA fragment to be assembled contained MMTV LTR sequences from position -217 to -14 with respect to the transcriptional start site. This

fragment was prepared by PCR between the MTVXBAI primer (5' end ^{32}P-labelled) and the MTVXHOI primer using the plasmid pGMCO (containing the entire MMTV LTR; from J. Thomas and K. Yamamoto, UCSF) as a template. Using this fragment, mononucleosomes were assembled by salt-gradient dialysis and purified by sedimentation in glycerol gradients as described (Piña et al., 1990). Assembly and purification were followed by electrophoresis in 5% polyacrylamide/1xTBE nondenaturing gels. The rotational positioning of the assembled mononucleosomes was checked for one strand (corresponding to the ^{32}P-labelled MTVXBAI primer) by parallel digestion of unassembled and assembled DNA with DNase I in a manner similar to that previously described (Piña et al., 1990).

Acknowledgements

We thank the following people for generous gifts of reagents: Caroline Goutte and Sandy Johnson for α2 protein; Iris Dotan and Pat Brown for MLV IN protein; and Andy Leavitt for HIV IN protein. We also thank Pat Brown, Sandy Johnson, Titia de Lange, Hans-Peter Muller, Michael Redd, Ian Taylor, and Keith Yamamoto for useful comments on the manuscript. This work was supported by grants from the NIH. H.E.V. is an American Cancer Society Research Professor.

[redacted]

References

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

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

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

Bushman, F. D., Fujiwara, T. and Craigie, R. (1990). Retroviral DNA integration directed by HIV integration protein in vitro. Science *249*, 1555-1558.

Craigie, R., Fujiwara, T. and Bushman, F. (1990). The IN protein of moloney murine leukemia virus processes the viral DNA ends and accomplishes their integration in vitro. Cell *62*, 829-837.

Drew, H. R. (1984). Structural specificities of five commonly used DNA nucleases. J. Mol. Biol. *176*, 535-557.

Drew, H. R., McCall, M. J. and Calladine, C. R. (1990). New approaches to DNA in the crystal and in solution. In DNA topology and its biological effects, N. R. Cozzarelli and J. C. Wang, eds. (Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press), pp. 1-56.

Drew, H. R. and Travers, A. A. (1984). DNA structural variations in the E. coli *tyrT* promoter. Cell *37*, 491-502.

Drew, H. R. and Travers, A. A. (1985). DNA bending and its relation to nucleosome positioning. J. Mol. Biol. *186*, 773-790.

Ellison, V., Abrams, H., Roe, T., Lifson, J. and Brown, P. (1990). Human immunodeficiency virus integration in a cell-free system. J. Virol. *64*, 2711-2715.

Farnet, C. M. and Haseltine, W. A. (1990). Integration of human immunodeficiency virus type 1 DNA in vitro. Proc. Natl. Acad. Sci. USA *87*, 4164-4168.

Felsenfeld, G. (1992). Chromatin as an essential part of the transcriptional mechanism. Nature *355*, 219-224.

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

Gross, D. S. and Garrard, W. T. (1988). Nuclease hypersensitive sites in chromatin. Annu. Rev. Biochem. *57*, 159-197.

Grunstein, M. (1990). Histone function in transcription. Annu. Rev. Cell Biol. *6*, 643-678.

Hirano, T. and Mitchison, T. J. (1991). Cell cycle control of higher-order chromatin assembly around naked DNA in vitro. J. Cell Biol. *115*, 1479-1489.

Johnson, A. D. and Herskowitz, I. (1985). A repressor (*MATα2* product) and its operator control expression of a set of cell type specific genes in yeast. Cell *42*, 237-247.

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

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.

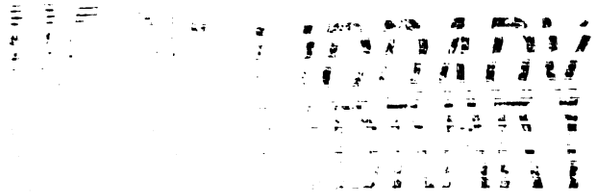
Kornberg, R. D. (1977). Structure of chromatin. Annu. Rev. Biochem. *46*, 931-954.

Kornberg, R. D. and Lorch, Y. (1991). Irresistible force meets immovable object: transcription and the nucleosome. Cell *67*, 833-836.

Leavitt, A. D., Rose, R. B. and Varmus, H. E. (1992). Both substrate and target oligonucleotide sequences affect in vitro integration mediated by HIV-1 integrase protein (IN) produced in Saccharomyces cerevisiae. J. Virol. in press.

Lee, Y. M. H. and Coffin, J. M. (1990). Efficient autointegration of avian retrovirus DNA in vitro. J. Virol. *64*, 5958-5965.

Mooslehner, K., Karls, U. and Harbers, K. (1990). Retroviral integration sites in transgenic mov mice frequently map in the vicinity of transcribed DNA regions. J. Virol. *64*, 3056-3058.



Pederson, D. S., Thoma, F. and Simpson, R. T. (1986). Core particle, fiber, and transcriptionally active chromatin structure. Annu. Rev. Cell Biol. *2*, 117-147.

Perlmann, T. and Wrange, O. (1988). Specific glucocorticoid receptor binding to DNA reconstituted in a nucleosome. EMBO J. *7*, 3073-3079.

Piña, B., Bruggemeier, U. and Beato, M. (1990). Nucleosome positioning modulates accessibility of regulatory proteins to the mouse mammary tumor virus promoter. Cell *60*, 719-731.

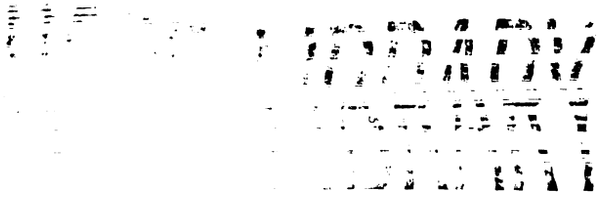
Pryciak, P. M., Sil, A. and Varmus, H. E. (1992). Retroviral integration into minichromosomes in vitro. EMBO J. *11*, 291-303.

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

Sauer, R. T., Smith, D. L. and Johnson, A. D. (1988). Flexibility of the yeast α2 repressor enables it to occupy the ends of its operator, leaving the center free. Genes Dev. *2*, 807-816.

Scherdin, U., Rhodes, K. and Breindl, M. (1990). Transcriptionally active genome regions are preferred targets for retrovirus integration. J. Virol. *64*, 907-912.

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



Shimizu, M., Roth, S. Y., Szent-Gyorgyi, C. and Simpson, R. T. (1991). Nucleosomes are positioned with base pair precision adjacent to the α2 operator in *Saccharomyces cerevisiae*. EMBO J. *10*, 3033-3041.

Simpson, R. T. and Stafford, D. W. (1983). Structural features of a phased nucleosome core particle. Proc. Natl. Acad. Sci. USA *80*, 51-55.

Smith, D. L. and Johnson, A. D. (1992). A molecular mechanism for combinatorial control in yeast: MCM1 protein sets the spacing and orientation of the homeodomains of an α2 dimer. Cell *68*, 133-142.

Suck, D., Lahm, A. and Oefner, C. (1988). Structure refined to 2Å of a nicked DNA octanucleotide complex with DNase I. Nature *332*, 464-468.

Suck, D., Oefner, C. and Kabsch, W. (1984). Three-dimensional structure of bovine pancreatic DNase I at 2.5 Å resolution. EMBO J. *3*, 2423-2430.

Taylor, I. C. A., Workman, J. L., Schuetz, T. J. and Kingston, R. E. (1991). Facilitated binding of GAL4 and heat shock factor to nucleosomal templates: differential function of DNA-binding domains. Genes Dev. *5*, 1285-1298.

Thoma, F., Bergman, L. W. and Simpson, R. T. (1984). Nuclease digestion of circular TRP1ARS1 chromatin reveals positioned nucleosomes separated by nuclease-sensitive regions. J. Mol. Biol. *177*, 715-733.

Thoma, F. and Simpson, R. T. (1985). Local protein-DNA interactions may determine nucleosome positions on yeast plasmids. Nature *315*, 250-252.

Travers, A. A. and Klug, A. (1990). Bending of DNA in nucleoprotein complexes. In DNA topology and its biological effects, N. R. Cozzarelli and J. C. Wang, eds. (Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press), pp. 57-106.

Varmus, H. E. and Brown, P. (1989). Retroviruses. In Mobile DNA, D. E. Berg and M. M. Howe, eds. (Washington D.C.: American Society for Microbiology), pp. 53-108.

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

Wolffe, A. P. (1991). Activating chromatin. Current Biology *1*, 366-368.

CHAPTER FIVE

SV40 MINICHROMOSOMES AS TARGETS FOR RETROVIRAL INTEGRATION IN VIVO

ABSTRACT

We present a method for studying multiple retroviral integration events into a small DNA target in vivo. Episomal SV40 genomes established by infection of CV-1 cells served as integration targets upon subsequent infection with murine leukemia virus (MLV). A PCR-based assay for the abundance and distribution of integration events was used to detect recombinants shortly after MLV infection. Integration of MLV into SV40 DNA was detectable as early as 4 hours and reached a maximum level by 8 hours after MLV infection. The level of integration but not the distribution of integration sites was sensitive to the stage in the SV40 life cycle at which MLV infection was performed, in a manner reflective of the accumulated copy number of SV40. Using a temperature-sensitive T antigen mutant SV40 strain we observed that active replication of the target DNA is not required for efficient integration. The distribution of integration sites in vivo was closely approximated by in vitro reactions using isolated SV40 minichromosomes as integration targets. However, the degree of bias between strong and weak sites is greater in vivo than in vitro, such that some sites become especially favored over others in vivo.

INTRODUCTION

As an essential part of their life cycle, retroviruses insert a DNA copy of their genome into the DNA of the host cell, in a process termed integration (for review, see Varmus and Brown, 1989). The development of cell-free systems for studying retroviral integration has speeded the dissection of steps in the process and of factors influencing the course of the reaction. These factors include the nature of the integration machinery, the viral protein (integrase, or IN) responsible for catalyzing the reaction, viral DNA sequences required, and the chemistry of the viral DNA cleavage and strand transfer steps (see Whitcomb and Hughes, 1992 for review). We are interested in how the retroviral integration machinery chooses a target site in which to insert its DNA, and how that choice is affected by the physiological state of the target DNA.

Retroviral integration is site-specific with regard to the retroviral DNA--insertion occurs such that the terminal bases of retroviral DNA are invariantly joined to the target DNA--but it is relatively non-specific with regard to the target DNA (for reviews, see Varmus and Brown, 1989; Sandmeyer et al., 1990). There is little preference for particular sequences at sites of integration, and many sites are available to serve as integration targets both in vivo and in vitro (Shimotohno and Temin, 1980; Shoemaker et al., 1981; Brown et al., 1987; Pryciak et al., 1992). Nevertheless, integration sites are not chosen randomly; the distribution of sites is clearly non-uniform even in naked DNA in vitro (Pryciak and Varmus, submitted). The choice of integration sites can be further influenced by more complex targets, such as DNA assembled into nucleosomes (Pryciak et al., 1992; Pryciak and Varmus, submitted). In addition, integration in vivo shows a non-random tendency to occur near DNase I hypersensitive sites and in transcriptionally active regions (Vijaya et al., 1986; Rohdewohld et al., 1987; Mooslehner et al., 1990; Scherdin et al., 1990). Also in vivo, a small population of sites can be used as integration sites for Rous sarcoma virus

DNA up to a million-fold more frequently than expected by random choice (Shih et al., 1988).

Such observations have suggested that target site selection during retroviral integration in vivo is sensitive to complex changes in the physiological state of DNA--such as transcription and replication, or different degrees of chromatin condensation--many of which are not immediately amenable to in vitro analysis. For instance, to assess the role that transcriptional activity has on use of the affected DNA as an integration target, it would be best to compare the extent of integration into that DNA in transcriptionally active and inactive states. However, it is currently difficult to conduct comparisons of this sort in vivo since, in most typical experiments, it is rare for even a single integration event to occur into a locus of interest, due to the enormous number of potential sites in animal cell genomes. In vitro reactions can limit the number of potential integration sites by providing a small target, but it is difficult to impose complex physiological states on the majority of target molecules in a population in vitro.

We have therefore developed a system with which to study multiple integration events into a single relatively small target in vivo, where the target can also be studied using an in vitro integration reaction. Our strategy was to coinfect cells with a DNA virus (SV40) and a retrovirus (murine leukemia virus, MLV), generating multiple episomal copies of SV40 DNA as potential integration targets for MLV in vivo. Experiments presented here demonstrate the integration of MLV DNA into SV40 minichromosomes in vivo, which occurs with a similar non-random distribution of integration sites as events into isolated SV40 minichromosomes in vitro. The efficiency and distribution of integration events is not appreciably affected by replication of the SV40 target in vivo, but undefined factors in cells accentuate the degree of bias between preferred and underutilized integration sites.

RESULTS

To conduct retroviral integrations into SV40 minichromosomes in vivo, susceptible cells were first infected with SV40, and then with MLV after variable lengths of time (Figure 1A). Since the cells must be permissive for both SV40 and MLV infection, we used CV-1 monkey cells, which are permissive for SV40, and an amphotropic strain of MLV (see Materials and Methods), which can infect non-murine cells. After a time sufficient for MLV to synthesize and integrate its DNA (4-24 hrs), extrachromosomal DNA was prepared and MLV-SV40 recombinants were analyzed by the following methods: (i) cloning of individual recombinants and sequencing of MLV-SV40 junctions; (ii) restriction enzyme cleavage and Southern blotting; or (iii) polymerase chain reaction (PCR)-based amplification of recombinant junctions.

We first cloned 12 independent in vivo recombinants in lambda-based vectors as previously described for in vitro integration products (Pryciak et al., 1992). The cloned DNAs appeared to represent bona fide retroviral integration products because sequencing from a primer in the MLV LTR revealed that the junctions between MLV and SV40 DNA immediately followed the terminal CA dinucleotide of MLV DNA (not shown). The positions in SV40 DNA of the 12 insertions mapped in this way were each different and were distributed without apparent preference for any particular region of the target, including the nucleosome-free, nuclease-sensitive region near the origin of replication (not shown).

Analysis of MLV integration into SV40 DNA in vivo by restriction enzyme cleavage and Southern blotting (as previously described for in vitro reactions; Pryciak et al., 1992) at various times after MLV infection (from 12 to 48 hrs) suggested that recombinants were readily detectable 12 hrs after MLV infection (not shown). Harvest times exceeding 18 to 24 hrs after MLV infection showed the accumulation of additional products that were consistent with oligomeric MLV-SV40 products that might result from

Figure 5-1. Strategy and demonstration of MLV integration into SV40 DNA in coinfected cells. **A.** Schematic depiction of experimental strategy. Cells which have been infected with SV40 are subsequently infected with MLV. At top is indicated that, after entry and DNA synthesis by MLV, linear double-stranded DNA containing nucleoprotein complexes can potentially integrate into either SV40 DNA (circular double helix) or cellular chromosomal DNA (linear double helix). The presence of MLV-SV40 recombinants is measured in preparations of extrachromosomal DNA prepared at variable times after infection. **B.** Time course of integration of MLV DNA into SV40 DNA in vivo. Duplicate plates of cells were infected with MLV 20 hr after infection with SV40, and the coinfected cells were then harvested at 2 to 16 hr after MLV infection as indicated. Extrachromosomal DNA was then prepared and MLV-SV40 recombinants were detected by PCR amplification between an end-labelled primer in the MLV LTR (MoU5L26) and an unlabelled primer in SV40 DNA (SV273+). In addition, extrachromosomal DNAs harvested at the last time point from parallel plates infected with one of the two viruses and mock-infected with the other, either MLV (-M) or SV40 (-S), were mixed together before the PCR analysis to demonstrate that PCR products are not generated unless the cells were coinfected (lane 14). Also, PCR reactions were performed using a pool of 30 cloned in vitro MLV-SV40 recombinants, or clone pool (cp), for which the exact positions of insertion are known (Pryciak et al., 1992), which serve as positive controls and size markers (lane 15). An additional size marker, an end-labelled 123-bp ladder (MW), was also included (lane1).

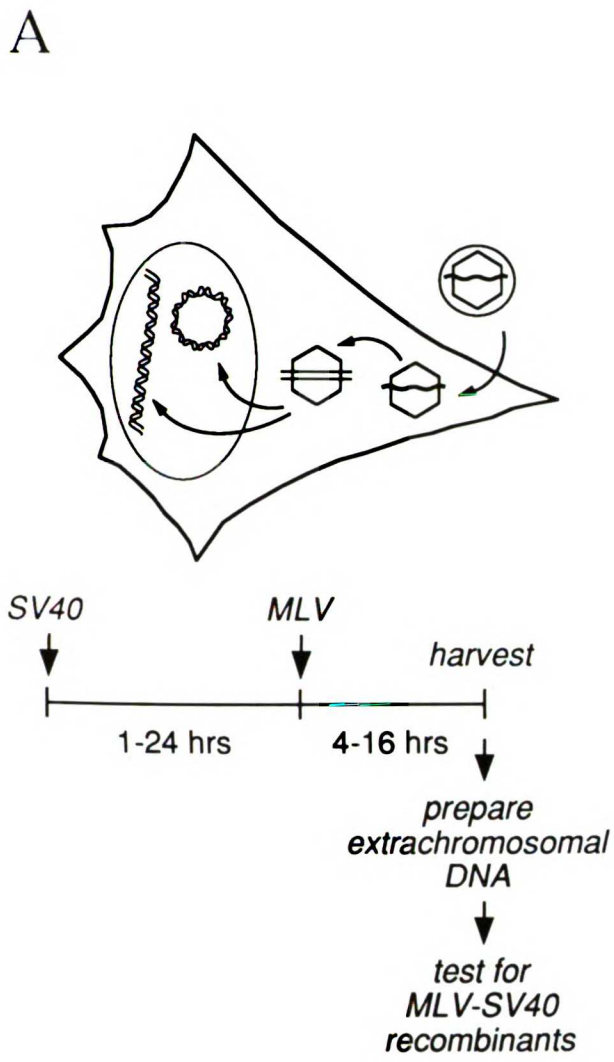

A
SV40
MLV
harvest
1-24 hrs
4-16 hrs
prepare
extrachromosomal
DNA
test for
MLV-SV40
recombinants
B
MW
2hr
4hr
6hr
8hr
12hr
16hr
(-S)
+
(-M)
cp
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15

Figure 5-1

replication of the recombinants (not shown). Thus, to promote the analysis of initial rather than replicated integration products, we restricted harvest times to 16 hr or less after MLV infection.

We found the PCR-based assay (Pryciak and Varmus, submitted) to be the most useful for analysis of integration into SV40 in vivo because of its ability to measure both the abundance of total recombinants and the distribution of insertion sites. In this assay, recombinants are amplified between a 5' end ^{32}P-labelled MLV primer and an unlabelled SV40 primer; the sizes and abundance of the PCR products are measures of the positions and frequencies of integration events, respectively. The size of the SV40 genome (5.2 kb) dictates that a single target primer cannot be used to analyze insertions over its entire length, because of low amplification efficiency of products greater than 1-2 kb in length. We therefore used four different target primers, distributed so as to monitor insertions over the entire SV40 target (see Materials and Methods, and Figure 3A).

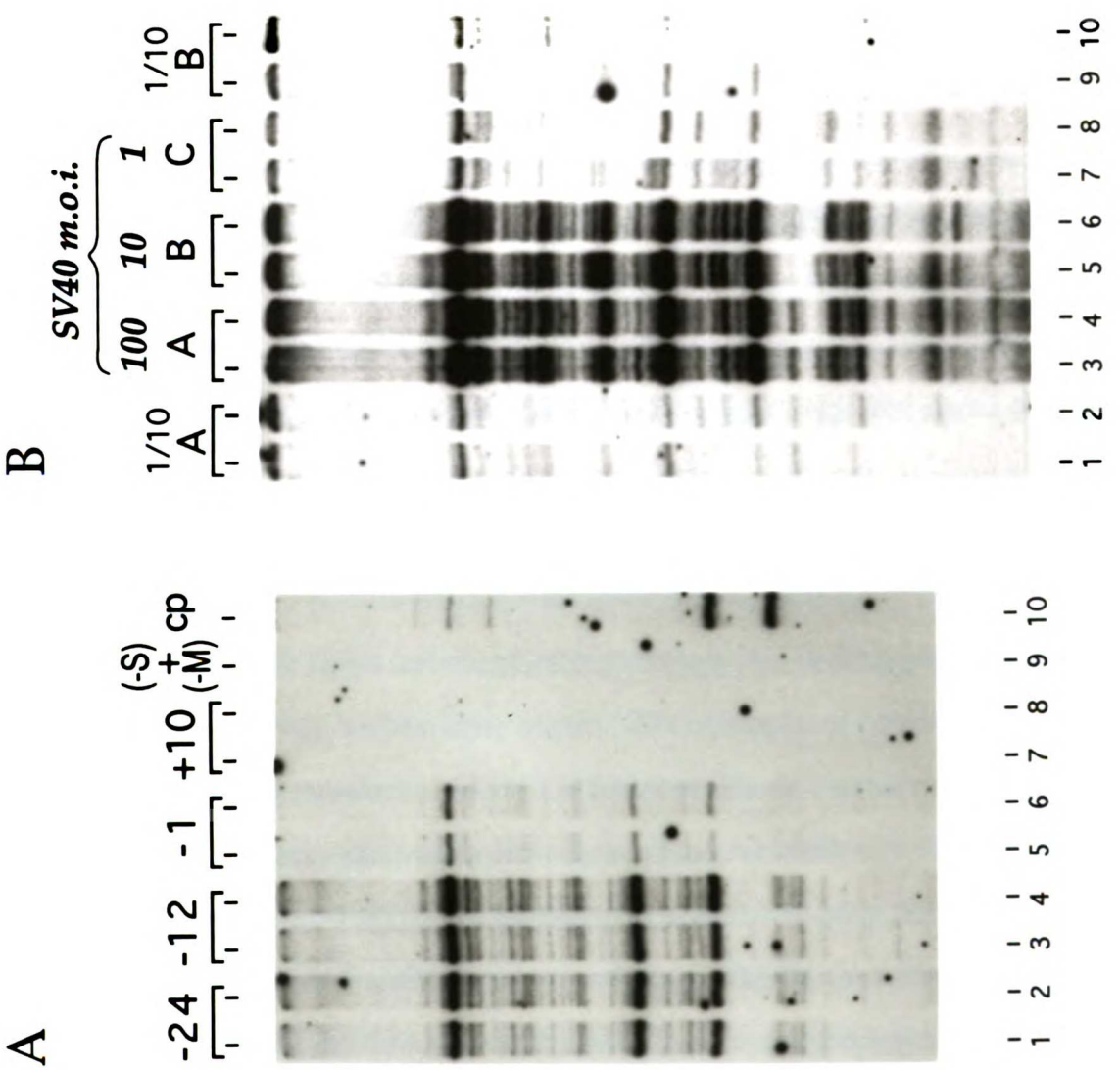
Using this assay, we monitored the appearance of MLV-SV40 recombinants at various times after MLV infection with one of the four primers (Figure 1B). Integration was readily detectable as early as 4 hr after MLV infection, and it reached an apparent plateau by 8 hr. The distribution of insertions was strikingly non-uniform, with a discrete population of sites used far more frequently than others, including one particularly favored site; this site, which maps to the middle of the SV40 late region, will be further discussed below. The accumulation of recombinants at increasing times after MLV infection was generally parallel for all sites; in the time course presented, there was no evidence of preferential accumulation of some recombinants over others that might result from preferential replication following integration. Other experiments demonstrate that the distribution observed is not dependent upon SV40-mediated replication of recombinants (see Figures 3, 5); however, at extended times after MLV infection (24 to 48 hr), preferential replication of some recombinants may alter the apparent distribution (not shown).

We next compared the amount and distribution of integration events as a function of the time after SV40 infection (Figure 2A). The level of integration into SV40 increased with longer intervals between SV40 and MLV infection, with no increase beyond a 12 hour delay; no integration into was observed in the example where SV40 infection followed MLV infection. In this experiment, MLV DNA could conceivably integrate into SV40 DNA as early as 0-6 hrs and as late as 40 hrs after SV40 infection. During this period, changes occur both in the stage of the SV40 life cycle and in the total copy number of SV40 DNA per cell. Over the time course of the experiment, no significant changes in the integration site distribution were apparent in any region of the SV40 target (shown for positions 1990 to ~600; other regions not shown); only the level of integration changed, which is most simply explained by the increases in SV40 copy number in the cell. If increased copy number of SV40 molecules underlies the observed increase in recombinant products, then it appears that the amount of integration target can become saturating, since recombinants were not more numerous when MLV was added 24 rather than 12 hrs after SV40, even though the level of SV40 DNA was still increasing between these times (see Figure 2A legend). The findings in Figure 2A are supported by an experiment in which SV40 infection was performed at a single time before MLV infection (-12 hr) but at three different multiplicities of infection (1, 10, and 100; Figure 2B). We found that the number of recombinants increased with increased multiplicity, but was again eventually saturable (between m.o.i. of 10 and 100; Figure 2B), even though the SV40 DNA copy number continued to increase (see Figure 2B legend).

The distribution of integration sites observed during integration into SV40 in vivo was next compared with that observed during in vitro reactions (Pryciak et al., 1992) using either SV40 minichromosomes (purified from infected cells) or SV40 DNA (from deproteinized minichromosomes) as targets for the MLV nucleoprotein complex (Figure 3A). It should be noted that the in vivo and in vitro reactions cannot be accurately compared with respect to absolute frequencies of integration events per site. Instead, the

Figure 5-2. Integration of MLV DNA into SV40 DNA in vivo is dependent upon the accumulated copy number of SV40. **A.** Duplicated plates were infected with SV40 (multiplicity of infection = 10) at times 24, 12 , or 1 hr before MLV infection or 10 hr after MLV infection, as indicated. Cells were harvested at 16 hr after MLV infection; recombinants and controls (lanes 9, 10) were analyzed as in Figure 1B, except that the target primer was SV1990-. The copy number of SV40 DNA at the time of harvest was estimated by ethidium bromide staining of samples in an agarose gel to be ~10^5, ~$2x10^4$, ~$5x10^2$, and <10^2 per cell for -24, -12, -1, and +10 experiments, respectively (not shown). **B.** Duplicate plates were infected with SV40 12 hr before MLV infection at mulitiplicities of infection (m.o.i.) of 100, 10, or 1, as indicated for infections A, B, and C, respectively. Cells were harvested 12 hr after MLV infection. Recombinants were analyzed as in (A) above. In addition, PCR reactions were performed using ten-fold dilutions of some of the recombinants (set A, lanes 1, 2, and set B, lanes 9, 10) in order to quantitatively compare the number of integration events in the different experiments. The copy number of SV40 DNA at the time of harvest was estimated as in (A) to be ~$1.5x10^5$, ~$3x10^4$, and ~$3x10^3$ per cell for experiments A, B, and C, respectively (not shown).

Figure 5-2


A
-24
-12
-1
+10
(-S) + (-M) cp
1 2 3 4 5 6 7 8 9 10
B
1/10 A
SV40 m.o.i.
100 A
10 B
1 C
1/10 B
1 2 3 4 5 6 7 8 9 10

amounts of in vivo recombinants analyzed by PCR was chosen so as to give a similar signal strength at most positions as for the in vitro recombinants, and the relative distributions of sites were compared.

We find that the in vivo distribution of integration sites is remarkably well approximated by in vitro reactions using SV40 minichromosomes as a target (Figure 3A). This supports the contention that the in vivo distribution is reflective of the initial integration events and not of preferential replication of some recombinants. Reactions in vitro using minichromosomes as target showed more similar distributions to in vivo integrations than did those using naked SV40 DNA as the target (Figure 3A). Despite the overall similarities, however, there are some differences between the in vitro and in vivo distributions. In particular, the degree of bias between sites, measured as the differences in frequency of integration between the strong and weak positions, is greater in vivo than in vitro.

It is apparent that some positions become especially favored in vivo, relative to their neighboring positions. The most dramatic example is seen at position ~710 of SV40 (Figure 3A, lanes 13,14; single arrowhead); this position is also strongly preferred in minichromosomes in vitro, but to a lesser degree. Other examples of particularly strong sites in vivo are seen at positions ~1640 and ~1802 (lanes 20, 21), and at ~3705 (lanes 27, 28); two of these sites are also strongly preferred in minichromosomes in vitro, but to a lesser degree (~1640 and ~3705, single arrowheads), whereas the other site is not especially preferred in vitro (~1802, double arrowhead). Other features are also notable: (i) some sites are specifically used when naked DNA is an integration target in vitro and are inhibited in minichromosomes in vitro and in vivo (asterisks, lanes 9-14 and 23-28); (ii) some are specific to minichromosome reactions in vivo and in vitro, and are used relatively less frequently when naked DNA is the target (filled circle, lanes 9-14); and (iii) the nuclease-sensitive region near the origin of replication is not particularly favored in vivo or in vitro (lanes 1-7; c.f. ref.Pryciak et al., 1992).

Figure 5-3. Comparison of distribution of integration sites during in vivo and in vitro integration reactions. **A.** Distribution of integration sites was compared by PCR analysis of the products of in vitro integration reactions using naked SV40 DNA (DNA) or SV40 minichromosomes (MC) as targets, or of products of integration of MLV into SV40 in coinfected cells (in vivo). PCR reactions were also performed using the clone pool (cp) as in Figure 1. At top are indicated the name of the unlabelled target DNA primer used in the PCR reactions, along with a schematic depiction of the location and orientation of the primer; in the schematic, regions of SV40 such as the origin of replication (o), and early (E) and late (L) transcription units, are indicated, and map positions in SV40 DNA run clockwise from the top. Each target primer allows measurement of integration products from the position adjacent to its terminal base (given in the primer name) to 1-2 kb away, in the direction indicated. To the left of each panel of PCR products are given the corresponding map positions in SV40 DNA, as deduced from the products of the clone pool. Notable positions are indicated with symbols (single and double arrowheads, circle, asterisks) as described in the text. Some of the particularly strong sites in vivo (single and double black arrowheads) were more accurately located on sequencing gels to be at the following positions: 710 (+/- 5 bp, lanes 13, 14); 1640 (+/- 5 bp, single arrowhead, lanes 20, 21); 1802 (+/- 1 bp, double arrowhead, lanes 20, 21); and 3705 (+/- 1 bp, lanes 27, 28). Speckle-filled single and double arrowheads indicate some of these strong sites that are visible using another target primer (positions 1640 and 1802, lanes 13, 14, and position 710, lanes 20, 21), but which are fainter because of decreasing efficiency of amplification of products longer than ~1 kb. In vitro integration reactions were all performed in parallel in the presence of spermidine (see B), and all PCR reactions were performed in parallel and were run in the same gel; products of in vitro MC reactions were underloaded by two-fold (compared to DNA reactions) in order to facilitate comparison of distributions. **B.** The effect of spermidine on the ability of in vitro reactions to approximate the in vivo distribution of integration sites. Duplicate in vitro integration reactions were performed in

the presence or absence of 15 mM spermidine using either naked SV40 DNA (DNA) or SV40 minichromosomes (MC) as target. PCR analyses, using the SV1990- target primer, compared the distributions of integration events in these reactions with those obtained by coinfection (in vivo). The products represented in lanes 3 and 4 were underloaded by two-fold in order to facilitate this comparison.

Figure 5-3

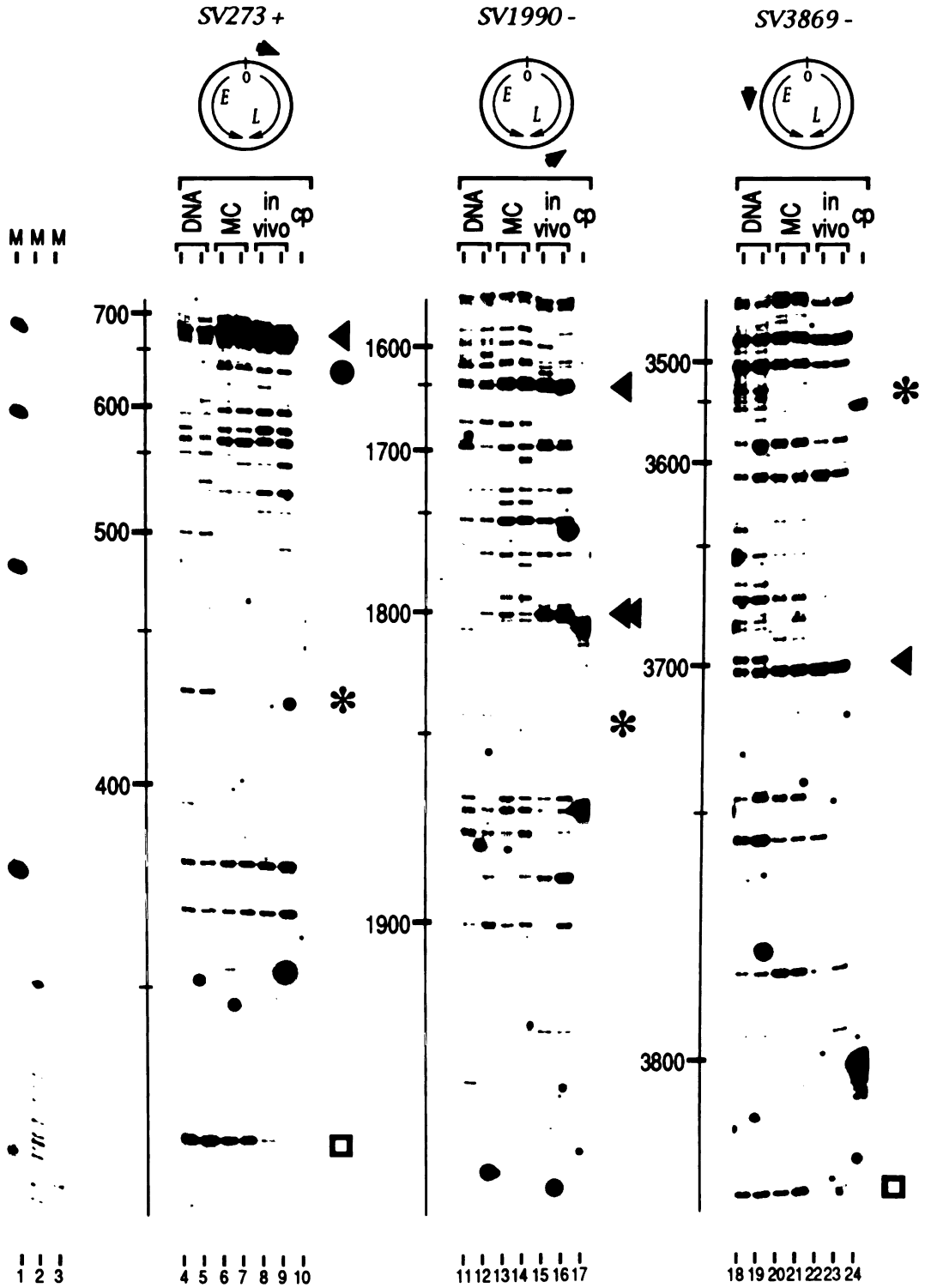
A
SV272 -
SV273 +
SV1990 -
SV3869 -
cp DNA MC in vivo
B
+spermidine -spermidine
DNA MC in vivo DNA MC cp
4000
4500
5000
5200
5243
200
1500
1000
500
300
2500
3000
3500
3800
1900

We found that the in vivo integration site distribution was most closely approximated by in vitro reactions with minichromosome targets when performed in the presence of spermidine (Figure 3B; additional data not shown). This dependence of integration site choice on reaction conditions--specifically, the presence or absence of spermidine--differs from our observations using other minichromosome and naked DNA targets (discussed in Pryciak and Varmus, submitted; other observations unpublished). In contrast to reactions with SV40 minichromosome targets, integration into naked SV40 DNA was relatively independent of the presence of spermidine (Figure 3B). It is possible that the SV40 minichromosomes tend to dissassemble or randomize their structure in the absence but not the presence of spermidine. Nucleosome core particles have been reported to become stabilized by spermidine and other polyamines (Morgan et al., 1987). We have not yet tested other alterations in reaction conditions for their ability to simulate in vivo choice of integration sites. We have, however, compared eight different protocols for preparing SV40 minichromosomes, including different salt concentrations, presence or absence of divalent cations, lysis procedure, time after SV40 infection, and sucrose gradient fractionation (see ref. Oudet et al., 1989); all showed indistinguishable integration site distributions which were equivalently affected by the presence or absence of spermidine (not shown).

The comparisons of in vitro and in vivo integration site distributions were also analyzed at high resolution by running the PCR products in denaturing polyacrylamide gels (Figure 4). The features noted above in the lower resolution analysis (Figure 3A) are confirmed here, and additional points can be made from the higher resolution view. First, each of the specific examples noted in Figure 3A are indicated here using the same symbols (arrowheads, circle, asterisks), as are some additional notable sites (squares). The higher-resolution analysis shows that most of the preferred sites (arrowheads, circle) consist of single positions, rather than regions; on the other hand, the sites inhibited in

Figure 5-4. Comparison of in vivo and in vitro integration site distributions at high resolution. The distribution of integration sites during in vitro reaction into naked or minichromosomal SV40 DNA were compared to those during in vivo coinfections. The same PCR products shown in Figure 3A were run here on a nondenaturing polyacrylamide gel. Notable positions are indicated with the symbols (single and double arrowheads, circle, and asterisks) as in Figure 3A, and as described in the text, with the addition of some (squares) that are not noted in Figure 3A. Size markers consisted of the clone pool (cp) reactions as well as a 100 bp ladder and two sequencing ladders (M, lanes 1-3).

Figure 5-4


SV273+
SV1990-
SV3869-
E
0
L
DNA
MC
in vivo
cp
M M M
700
600
500
400
1600
1700
1800
1900
3500
3600
3700
3800
1 2 3
4 5 6 7 8 9 10
11 12 13 14 15 16 17
18 19 20 21 22 23 24

minichromosomes in vitro and in vivo (asterisks) can correspond to either single positions (lanes 4-10) or larger regions (lanes 11-17 and 18-24). Second, in contrast to our previous observations using purified yeast TRP1ARS1 minichromosomes or in vitro assembled mononucleosomes as integration targets, we cannot detect a ~10 bp periodic distribution of preferred integration sites in the minichromosome reactions. The absence of this feature is probably due to a paucity of rotationally-positioned nucleosomes in the SV40 minichromomes (see Discussion).

Because SV40 DNA is replicated during the course of the co-infection experiments in vivo, we wanted to know whether replication affected the frequency or distribution of integration events. This issue was addressed using a temperature-sensitive (ts) mutant of SV40 (tsA28) which is unable to initiate DNA replication at the non-permissive temperature (Loeber et al., 1989). Cells were infected at the permissive temperature (32°C) to allow accumulation of multiple copies of SV40 DNA per cell, and then were either left at the permissive temperature or shifted to the restrictive temperature (40°C) concurrent with MLV infection. At the restrictive temperature, all subsequent initiation of SV40 DNA replication is inhibited, and already initiated molecules will elongate and resolve into daughter molecules within minutes (Chou et al., 1974). We observed that integration into the SV40 genome does not require active replication of SV40 DNA, since it occurred as efficiently at 40°C as at 32°C (Figure 5A). In addition, this experiment shows that the distribution of integration events observed with wild-type SV40, particularly the pattern of exceptionally favored sites, does not reflect preferential replication of some of the initial integration recombinants. Finally, there is no indication that the choice of integration sites is sensitive to the replicative status of the target, but since the fraction of total SV40 DNA that is replicating (at permissive temperature, or in wild-type infections) may be small, then any replication-induced changes in integration site preference could be obscured by an excess of integration events into the non-replicating majority of DNA.

Figure 5-5. Lack of dependence of integration efficiency or site distribution on replication of or packaging of SV40 DNA. **A.** Infection of cells with the tsA28 strain of SV40 proceeded at 32° C for 24 hr, and then MLV infection proceeded at either 32° C or 40° C for 12 hr. Recombinants were analyzed by PCR using either the SV1990- or SV273+ target DNA primer. **B.** Infection of cells with the tsC219 strain of SV40 (tsC) proceeded at 40° C for 24 hr, and then MLV infection proceeded at 40° C for 12 hr. Wild-type SV40 (wt) infection was at 37° C for 12 hr and followed by 12 hr MLV infection at the same temperature, but was not performed in parallel with the tsC experiments. PCR products using either SV1990- or SV273+ primers were analyzed for tsC and wt experiments on different gels.

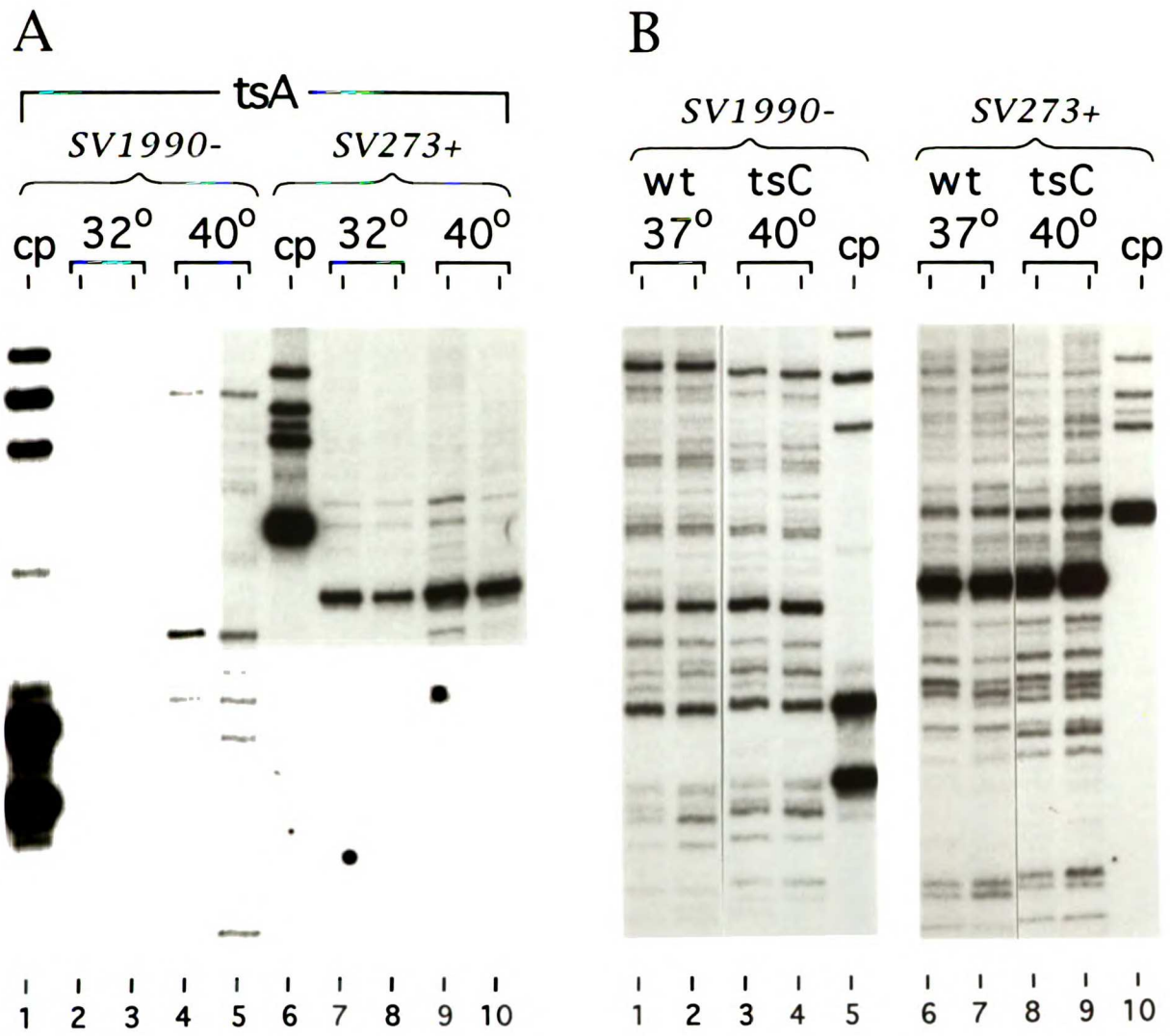

A
tsA
SV1990-
SV273+
cp
32°
40°
cp
32°
40°
1 2 3 4 5 6 7 8 9 10
B
SV1990-
SV273+
wt
tsC
37°
40°
cp
wt
tsC
37°
40°
cp
1 2 3 4 5 6 7 8 9 10

Figure 5-5

During the late phase of the SV40 life cycle, some viral DNA is packaged into capsid structures, which may alter its accessibility to retroviral integration machinery. To ask whether packaging of SV40 DNA affects the distribution of MLV integration sites, we performed additional experiments using a ts mutant of SV40 that affects the major virion structural protein VP1 (tsC219), which cannot assemble virions at the nonpermissive temperature (Bina et al., 1983). We found that the distribution of integration sites into SV40 DNA was unaltered at the restrictive temperature (Figure 5B), demonstrating that the pattern is not dependent upon virion assembly. However, since the proportion of SV40 DNA undergoing assembly may be small, we cannot say whether or not the assembly process has an effect upon the use of SV40 DNA as an integration target.

DISCUSSION

Using SV40 DNA in infected cells as a target for retroviral integration, we have developed a simple and efficient system for analyzing the selection of integration sites within a relatively small target in vivo. The results confirm the applicability of our previous in vitro work to in vivo situations, provide a starting point for further investigations of the sensitivity of integration to the physiological state of the target, and establish the ability to use retroviral integration as an in vivo probe for chromatin structure.

The success of the present experiments is probably attributable to several features. First, the high copy number of SV40 DNA may be important in order to compete with genomic DNA for a substantial proportion of the total integration events. Maximal integration levels were reached at 10^3 to 10^4 copies of SV40 per cell (representing ~0.1 and ~1 % of total cell DNA, respectively; see Figure 2). Second, it is possible that SV40 DNA competes with genomic DNA disproportionately for its fraction of total DNA mass, either because it is an episome or perhaps because of special properties inherent to SV40. In this regard it is notable that other episomal elements, bovine papilloma virus (BPV)-based vectors, have successfully served as in vivo integration targets in recent studies (H.-P. Muller and H.E.V., unpublished observations). We do not yet have an accurate measurement of the relative efficiency of integration into SV40 and chromosomal DNAs in vivo. Conceivably, the saturation phenomenon observed (Figure 2) could result from nearly all integration events occurring into SV40. Considering previous suggestions that retroviral integration occurs preferentially in transcriptionally active or more uncondensed regions of chromatin, the bulk of genomic DNA may compete poorly with SV40 DNA. Finally, the use of an extrachromosomal DNA fraction as the source of PCR templates offers a technical advantage of lowering non-specific amplification that can result from the high concentration and complexity of genomic DNA (unpublished observations).

The ability to study numerous retroviral integration events into a single locus in vivo is essential to deciphering the contribution of various physiological parameters to the choice of integration target sites. Unfortunately, the large number of sites available in the genome makes it difficult in practice to isolate multiple insertions into single-copy regions, which has been achieved in only a few cases (e.g. Shih et al., 1988; Natsoulis et al., 1989). In contrast, the experimental system described here allows for an entirely different degree of analysis: (i) very rapid reactions do not require replication of the recombinants; (ii) large numbers of events approach saturation of the potential target sites (as shown by the reproducibility of integration site distribution); and (iii) hundreds to thousands of integration events can be analyzed in a single lane of a gel. A recent report (Isfort et al., 1992) of integration of the retrovirus reticuloendotheliosis virus (REV) into the genome of the herpesvirus Marek disease virus (MDV) is conceptually similar to the events studied here, and in fact shows that such events have occurred in the past in animals coinfected with both viruses; however, the low frequency of those events required weeks of passage of coinfected cells for their detection, such that the recombinants identified probably probably have been biased by replicative selection.

In addition, this study presents the first comparison of in vivo and in vitro insertion site distributions for retroviral integration. We previously observed that integration in vitro did not show a strong preference for nucleosome-free and/or nuclease sensitive regions in minichromosome targets (Pryciak et al., 1992; Pryciak and Varmus, submitted). Here we observe that integration in vivo also does not show such a preference. Instead, a small number of cloned recombinants and a PCR-based survey of insertions in a large fraction of the SV40 target show that many sites are available for integration throughout the SV40 DNA in vivo. In fact, the most frequently used sites were within the transcribed, coding regions (see Figure 3A). The ability of the in vitro reactions to closely mimic the in vivo distributions supports the use of the in vitro reactions as an appropriate system for dissecting the mechanisms of integration target site choice.

Importantly, however, the in vivo integration site distributions did show some differences from those of the in vitro reactions. We showed previously that assembly of DNA into nucleosomes increased bias between sites, caused by both inhibition and enhancement of site reactivities (Pryciak and Varmus, submitted). Here we find that the degree of bias between strong and weak sites is further accentuated in vivo, such that some sites become exceedingly favored over their neighbors, and others become less favored. Thus, the target is used differently in vivo than in vitro. One interpretation of this observation is that the target is simply more homogeneous or ordered in vivo, and that some randomization of the minichromosome structure occurs during their isolation for in vitro reactions. Alternatively, the target may be qualitatively in a different state or environment in vivo--e.g. transcribed, replicated, or attached the nuclear matrix--in a manner that is not duplicated in the in vitro reaction mixture. The in vitro reaction can now be further modified to try to closer approximate the in vivo observations, thereby providing a type of in vitro assay for the native in vivo state of a piece of chromatin.

We observed previously that integration in vitro into TRP1ARS1 minichromosomes or a MMTV-based mononucleosome showed a ~10 bp periodic distribution of preferred sites, resulting from preferential use of the exposed face of the nucleosomal DNA helix (Pryciak and Varmus, submitted). In contrast, integration into SV40 minichromosomes in vitro or in vivo did not show a this periodic distribution (Figure 4). It seems most likely that the nucleosomes in SV40 minichromosomes do not exhibit the strict rotational positioning necessary to observe periodic usage. The large number of neighboring nucleosomes in SV40 (20-27 nucleosomes total), which are poorly phased translationally (Ambrose et al., 1990, and references therein), may compromise the ability of individual nucleosomes to assume their preferred rotational orientations (Drew and Calladine, 1987; Drew and McCall, 1987). It remains to be determined how frequently DNA segments manage to secure their preferred rotational phasing in the context of neighboring influences in vivo, since rotational orientations of DNA segments assembled into nucleosomes in vivo

have been measured in only a few cases (Thomas and Elgin, 1988; Zhang and Gralla, 1989; Pfeifer and Riggs, 1991; Shimizu et al., 1991; Pryciak and Varmus, submitted).

By using a temperature sensitive T antigen mutant strain of SV40 (tsA28) we were able to demonstrate that, as during in vitro reactions, active replication is not required for DNA to serve as a good integration target in vivo. This observation helps to distinguish among explanations for several earlier experiments in which integration occured principally into recently replicated DNA or was inhibited in cells in which cellular DNA synthesis was blocked (Varmus et al., 1977; Chinsky and Soeiro, 1982; Hsu and Taylor, 1982). The lack of a direct requirement for active replication of an integration target in vivo is supportive of recent experiments that demonstrate a requirement for the host cell to pass through the mitotic phase of the cell cycle in order for nuclear entry and integration of MLV DNA to occur (T. Roe, T. Reynolds, and P. Brown, personal communication), a prerequisite to which the earlier observations can now be attributed.

The present experiments provide a starting point for further manipulations of the in vivo physiological state of the target DNA and study of the resulting effects on integration. While we did not observe any effects of shutting down SV40 replication or virion assembly on the pattern of target site preferences, it seems likely that heterogeneity of the SV40 target in vivo could obscure such effects. The heterogeneity of SV40 genomes in cells counters the advantages of their very high copy number. Recent work, however, shows that the strategy used here can be extended to lower copy number episomal targets, such as BPV-based vectors, by relatively mild scaling up of the number of cells analyzed (H.-P. Muller and H.E.V, unpublished observations). Such alternative targets offer potential advantages of homogeneity in chromatin structure, synchronous replication of the target with cellular DNA, and the absence the eventual killing of the cell that occurs in SV40 infections. Finally, such methods may be extendable to chromosomal targets, although there will almost certainly be some decrease in integration frequency when analyzing single copy loci.

These experiments document the use of retroviral integration as an in vivo probe for chromatin structure and DNA reactivity. Since retroviral integration has evolved to conduct functional interactions with cellular DNA, it may show sensitivities to parameters other than just accessibility--such as functional states or intranuclear localization of target DNA--that may be ignored by probes such as chemicals, radiation, and nucleases. Indeed, our previous in vitro work has already highlighted the fact that retroviral integration can respond differently to the incorporation of DNA into nucleosomes than do most other previously used probes. Additional experiments of the type presented here may shed new light on how the functional and organizational state of cellular DNA can influence cellular activities.

Materials and Methods

Cells and viruses

All cells were grown in DME H-21 medium supplemented with 10% Fetal Calf Serum. The monkey cell line CV-1 served as the host for SV40 and MLV coinfections. Wild type SV40 was strain #777 ; the mutant SV40 strains tsA28 and tsC219 were gifts from P.Tegtmeyer (State University of New York, Stony Brook) and M. Bina (Purdue University), respectively. Amphotropic MLV capable of infecting CV-1 cells was harvested from the cell line PA317: MoMLV-SupF, which was created by infecting the amphotropic packaging cell line PA317 (Miller and Buttimore, 1986) with replication competent ecotropic MoMLV-SupF (Brown et al., 1987) harvested from the cell line SC-1: MoMLV-SupF. The resulting cell line (PA317:MoMLV-SupF) produces a mixed stock of MLV, carrying both ecotropic and amphotropic envelope proteins.

Virus infections and harvest of recombinants

Duplicate plates of CV-1 cells (1 x 10^6 per 100mm culture dish) were infected with wild type SV40 (m.o.i. = 10, except as indicated in Figure 2B) for 1hr at 37°C in 1 ml medium, and then 5 ml of fresh growth medium was added and cells returned to 37°C. SV40 infection usually proceeded for 12 hr before MLV infection, but actual times are given in the Figure legends. Temperature-sensitive mutant (tsA28 and tsC219) SV40 infections were at m.o.i. = 1 (see Figure 5 legend). MLV infections used 24 hr freshly harvested virus from 5 x 10^6 PA317:MoMLV-SupF cells (in 4 ml) for each plate of (10^6) CV-1 cells, and were performed in the presence of 8 µg/ml polybrene usually for 12 hours, but actual times are given in the Figure legends. Coinfected cells were harvested by trypsinization and collected by pelleting through ice-cold growth medium at 1000 g for 5 min. The cell pellets were lysed by gentle resuspension in 150 µl of cold extrachromosomal extraction buffer (10 mM TRIS, pH 8.0, 250 mM KCl, 5 mM $MgCl_2$, 0.5 % NP-40). The lysates were left on ice for 2 hrs, and then centrifuged at 15,000 g for 10 min. at 4°C. Nucleic

acids were prepared from the supernatants by brief treatment with 0.25 mg/ml proteinase K in the presence of 8 mM EDTA and 0.5 % SDS, followed by two rounds of phenol/chloroform extraction and one round of chloroform extraction, and finally two rounds of ethanol precipitation. The final nucleic acid pellet was resuspended in 40 µl of TE, and 0.5 to 1 µl was used for the PCR analysis.

In vitro integration reactions

Integration in vitro was mediated by viral nucleoprotein complexed using naked or minichromosomal SV40 DNA as the target as previously described (Pryciak et al., 1992). All reactions were carried out in the presence of 15 mM spermidine except for those indicated in Figure 3B. The products of reactions mediated by 20 µl of integration extract were resuspended in 20 µl of TE, and 0.5 µl was used for the PCR analysis.

PCR-based analysis of MLV-SV40 recombinants

In vivo or in vitro integration products were analyzed by PCR as previously described (Pryciak and Varmus, submitted). In all cases, the 5' end ^{32}P-labelled primer was the viral DNA primer MoU5L26, and the target DNA primer was unlabelled. The following oligonucleotides (synthesized by the Biomolecular Resource Center, UCSF), indicated by name, sequence, and map location (in parantheses) were used in this study:

MoU5L26 5' CGACTTGTGGTCTCGCTGTTCCTTGG (8335-8360 MoMLV)

SV272- 5' ACCTTCTGAGGCGGAAAGAACCAGC (296-272 SV40)

SV273+ 5' CCCTAACTGACACACATTCCACAGC (249-273 SV40)

SV1990- 5' CCTGTAGTTTGCTAACACACCCTGC (2014-1990 SV40)

SV3869- 5' GAACAGCCCAGCCACTATAAGTACC (3893-3869 SV40)

The products of the PCR reactions were separated in 5 % nondenaturing acrylamide or 6 % denaturing acrylamide gels, and exposed to X-ray film.

REFERENCES

Ambrose, C., Lowman, H., Rajadhyaksha, A., Blasquez, V. and Bina, M. (1990). Location of nucleosomes in simian virus 40 chromatin. J. Mol. Biol. *214*, 875-884.

Bina, M., Blasquez, V., Ng, S.-C. and Beecher, S. (1983). SV40 morphogenesis. Cold Spr. Harb. Symp. Quant. Biol. *47*, 565-569.

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

Chinsky, J. and Soeiro, R. (1982). Studies with aphidicolin on the *Fv-1* host restriction of Friend murine leukemia virus. J. Virol. *43*, 182-190.

Chou, J. Y., Avila, J. and Martin, R. G. (1974). Viral DNA synthesis in cells infected by temperature-sensitive mutants of simian virus 40. J. Virol. *14*, 116-124.

Drew, H. R. and Calladine, C. R. (1987). Sequence-specific positioning of core histones on an 860 base-pair DNA. J. Mol. Biol. *195*, 143-173.

Drew, H. R. and McCall, M. J. (1987). Structural analysis of a reconstituted DNA containing three histone octamers and histone H5. J. Mol. Biol. *197*, 485-511.

Hsu, T. W. and Taylor, J. M. (1982). Effect of aphidicolin on avian sarcoma virus replication. J. Virol. *44*, 493-498.

Isfort, R., Jones, D., Kost, R., Witter, R. and Kung, H.-J. (1992). Retrovirus insertion into herpesvirus in vitro and in vivo. Proc. Natl. Acad. Sci. USA *89*, 991-995.

Loeber, G., Tevethia, M. J., Schwedes, J. F. and Tegtmeyer, P. (1989). Temperature-sensitive mutants identify crucial regions of simian virus 40 large T antigen. J. Virol. *63*, 4426-4430.

Miller, A. D. and Buttimore, C. (1986). Redesign of retrovirus packaging cell lines to avoid recombination leading to helper virus production. Mol. Cell. Biol. *6*, 2895-2902.

Mooslehner, K., Karls, U. and Harbers, K. (1990). Retroviral integration sites in transgenic mov mice frequently map in the vicinity of transcribed DNA regions. J. Virol. *64*, 3056-3058.

Morgan, J. E., Blankenship, J. W. and Matthews, H. R. (1987). Polyamines and acetylpolyamines increase the stability and alter the conformation of nucleosome core particles. Biochemistry *26*, 3643-3649.

Natsoulis, G., Thomas, W., Roghmann, M.-C., Winston, F. and Boeke, J. D. (1989). Ty1 transposition in Saccharomyces cerevisiae is nonrandom. Genetics *123*, 269-279.

Oudet, P., Weiss, E. and Regnier, E. (1989). Preparation of simian virus 40 minichromosomes. Methods in Enzymology *170*, 14-25.

Pfeifer, G. P. and Riggs, A. D. (1991). Chromatin differences between active and inactive X chromosomes revealed by genomic footprinting of permeabilized cells using DNase I and ligation-mediated PCR. Genes Dev. *5*, 1102-1113.

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

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

Sandmeyer, S. B., Hansen, L. J. and Chalker, D. L. (1990). Integration specificity of retrotransposons and retroviruses. Annu. Rev. Genet. *24*, 491-518.

Scherdin, U., Rhodes, K. and Breindl, M. (1990). Transcriptionally active genome regions are preferred targets for retrovirus integration. J. Virol. *64*, 907-912.

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

Shimizu, M., Roth, S. Y., Szent-Gyorgyi, C. and Simpson, R. T. (1991). Nucleosomes are positioned with base pair precision adjacent to the α2 operator in *Saccharomyces cerevisiae*. EMBO J. *10*, 3033-3041.

Shimotohno, K. and Temin, H. M. (1980). No apparent nucleotide sequence specificity in cellular DNA juxtaposed to retrovirus proviruses. Proc. Natl. Acad. Sci. USA *77*, 7357-7361.

Shoemaker, C., Hoffmann, J., Goff, S. P. and Baltimore, D. (1981). Intramolecular integration within moloney murine leukemia virus DNA. J. Virol. *40*, 164-172.

Thomas, G. H. and Elgin, S. C. R. (1988). Protein/DNA architecture of the DNase I hypersensitive region of the *Drosophila hsp26* promoter. EMBO J. *7*, 2191-2201.

Varmus, H. E. and Brown, P. (1989). Retroviruses. In Mobile DNA, D. E. Berg and M. M. Howe, eds. (Washington D.C.: American Society for Microbiology), pp. 53-108.

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.

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

Whitcomb, J. M. and Hughes, S. H. (1992). Retroviral reverse transcription and integration: progress and problems. Annu. Rev. Cell Biol. in press.

Zhang, L. and Gralla, J. D. (1989). In situ nucleoprotein structure at the SV40 major late promoter: melted and wrapped DNA flank the start site. Genes Dev. *3*, 1814-1822.