

UC Irvine

UC Irvine Previously Published Works

Title

Integration Specificity of Retrotransposons and Retroviruses

Permalink

<https://escholarship.org/uc/item/12m864qp>

Journal

Annual Review of Genetics, 24(1)

ISSN

0066-4197

Authors

Sandmeyer, Suzanne B

Hansen, Lori J

Chalker, Douglas L

Publication Date

1990-12-01

DOI

10.1146/annurev.ge.24.120190.002423

Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at <https://creativecommons.org/licenses/by/4.0/>

Peer reviewed

INTEGRATION SPECIFICITY OF RETROTRANSPOSONS AND RETROVIRUSES

Suzanne B. Sandmeyer, Lori J. Hansen, and Douglas L. Chalker

Department of Microbiology and Molecular Genetics, College of Medicine, University of California, Irvine, California 92717

KEY WORDS: Ty elements, retroelements, transposable elements, polymerase III, recombination

CONTENTS

INTRODUCTION	491
<i>Classes of Retroelements</i>	491
<i>Retroviruses and Retrotransposons</i>	492
RETROVIRUS INTEGRATION	493
<i>IN Function Defi</i>	493
<i>Nonrandom Retroviral Integration</i>	495
NONRANDOM RETROTRANSPOSON INTEGRATION	498
<i>Drosophila Retrotransposons</i>	498
<i>Yeast Retrotransposons</i>	499
<i>Dictyostelium Retrotransposons</i>	507
<i>Polymerase III-Transcribed Genes as Genomic Integration Targets</i>	508
SUMMARY	509
PERSPECTIVES	501

INTRODUCTION

Classes of Retroelements

As we learn more about the integration patterns of retroelements, nonrandom integration has come to be regarded as a more general property of their lifestyle. This is perhaps not surprising, as successful propagation of these

elements is contingent on genomic cycling. The position and orientation of a given retroelement insertion can determine both its expression and retention in the genome. For some retrotransposons, sequence-specific integration is dependent upon the activity of the retroelement-encoded endonuclease (6, 140a). Retrovirus and retrotransposon integration specificity, the subject of this review, is, for the most part, not well defined. Integration of these elements depends upon the encoded integrase. Data from multiple systems, however, suggest that host chromosomal DNA, proteins, and nuclear functions are intrinsic to the targeting process.

Retrotransposons are retroviruslike elements found in various eukaryotic cells. Retrotransposons differ from retroviruses in that they do not have an obligatory extracellular phase. Retroviruses and retrotransposons have long terminal repeats (LTRs) of several hundred basepairs (bp) which flank an internal coding domain. The proteins encoded by the internal domains of retrotransposons and retroviruses are analogous in function and some are also homologous based on sequence comparison. Integration patterns of these elements vary greatly, from the apparent sequence specificity of the *Drosophila* gypsylike elements, to the position specificity of the yeast element Ty3, to the more degenerate patterns of other *Drosophila* elements, the yeast Ty1 element and animal retroviruses. Despite diverse patterns of integration, the integrases encoded by these elements have distinct similarities. In addition, we argue that many—and possibly all—retroviruses and retrotransposons exhibit some degree of integration site preference.

In vitro integration systems for animal retroviruses (18, 42) and for Ty1 (37) are yielding considerable information about the requirements of the integration reaction. The recent identification of elements with novel integration properties, together with in vitro systems with which to examine molecular mechanisms of integration, will facilitate future efforts to understand the target site selection process.

Retroviruses and Retrotransposons

Retroviruses (131) and retrotransposons (7, 9, 40) are similar in genomic organization and encoded proteins. These elements are typically between 5 and 10 kilobase pairs (kbp) in length. They are composed of an internal domain flanked by LTRs of several hundred bp in length. With a few exceptions among the retrotransposons, these elements begin and end with short inverted repeats. These short, usually imperfect, inverted repeats typically terminate in the dinucleotides TG/CA. Integrated retroviruses and retrotransposons are flanked by direct repeats of 4 to 6 bp.

Simple retroviruses are generally composed of three open reading frames (ORFs): *gag*, *pol*, and *env*. These ORFs typically encode the structural proteins of the nucleocapsid, catalytic proteins required for replication and

integration, and envelope proteins of the virus particle, respectively. Major structural proteins include matrix (MA), capsid (CA), and nucleocapsid (NC) proteins. The catalytic proteins are protease (PR), reverse transcriptase (RT), RNaseH (RN), and integrase (IN). Retrotransposons typically have one or two ORFs encoding the structural and catalytic proteins. The upstream coding region is analogous to the *gag* gene of retroviruses and encodes structural proteins of the retrotransposon nucleocapsid. The downstream region is analogous to the retroviral *pol* gene and encodes catalytic functions. The order in which catalytic functions are encoded can differ from retroviruses, as in the arrangement PR, IN, RT, and RN, found in copia of *Drosophila* (86), *Tal* of *Arabidopsis* (134), *Tnt1* of *Nicotianae* (51), and *Ty1* (24) and *Ty2* (60, 137) of *Saccharomyces cerevisiae* or can be generally similar, as in PR, RT, RN, and IN found in gypsylike elements of *Drosophila* (69, 83, 105, 125, 129), *Ty3* of *S. cerevisiae* (58), *del* of *Lilium henryi* (117), and *Tf1* of *Schizosaccharomyces pombe* (H. Levin & J. Boeke, personal communication).

Reverse transcription of the retroviral genome takes place inside the virion core after viral entry into the host cell. Retroviral replication is reviewed in detail by Varmus & Brown (131). The products of replication are full-length, linear molecules, and circularized forms of this product, containing one or two LTRs. For some retroviruses the linear form has been shown to be 2 bp longer on each end than the integrated provirus. Relatively little is known about the process by which replicated retroviral DNA arrives in the nucleus, or with what proteins, in addition to the integrase, it is likely to be associated at this stage. Similarities of IN structure, replication intermediates, and in vitro requirements suggest that retrotransposon integration will be fundamentally similar to that of retroviruses.

RETROVIRUS INTEGRATION

IN Function Defined In Vivo and In Vitro

Mutations in the downstream region of the *pol* genes of retroviruses have been shown to block viral infectivity at a point past replication (31, 32, 93, 99, 111). The pp32 protein in avian leukosis and sarcoma viruses (ALSV), p46 in MoMLV, and p34 of the human immunodeficiency virus (HIV) (131) have been identified as IN proteins. Alignment of the predicted protein sequences from the regions encoding these proteins has shown that they are homologous, and motifs conserved among these proteins are present in other retroviruses and retrotransposons (33, 71, 129, 140). Mutation analysis has shown that regions which encode proteins with similarity to the retrovirus IN, p90 (10, 37) and p61/58 (59) in *Ty1* and *Ty3*, respectively, are also required for integration.

The time course of processing of full-length, extrachromosomal retroviral

DNA has been determined in MoMLV-infected rat cells. Over the initial ten hours after infection, the proportion of blunt-ended, linear species having two extra bp at each end decreases, compared to the increased amount of a species in which the 3' end is recessed by 2 nt (103). The intermediate with 3'-recessed ends was not detectable in an experiment performed with a MoMLV IN mutant, consistent with an *in vivo* role for IN in producing this species. While a circular two LTR species is also present in cells and can apparently act as an integration donor in some systems (94), the *in vitro* data support a linear integration precursor.

The development of *in vitro* integration systems has greatly facilitated experiments to identify proteins directly involved in the reaction. The original *in vitro* systems measured the ability of nucleocapsids produced from genetically tagged MoMLV (18) or Tyl (37) to produce integrated DNA when incubated together with a naked λ DNA target. The genetic marker, a bacterial suppressor *supF*, allowed the target λ gtWES, which contained three amber mutations, to replicate in a wild-type host. In these assays, activity was associated with nucleocapsids and was proportional to the concentration of linear DNA (18). In addition, the reaction was dependent on wild-type IN activity (19, 37). Products and intermediates of the *in vitro* MoMLV reaction were consistent with ligation of 3'-recessed ends of linear donor DNA to 5'-extended ends of target DNA (19, 43). Insertions propagated by *in vitro* integration and recovered by biological selection were flanked by the characteristic short direct repeat. The reaction required divalent cations, but did not display dependence on dNTPs, ATP, or supercoiling of the target DNA.

More recently, efforts have focused on defining activity associated with IN (52). Characterized IN proteins are DNA-binding proteins (63, 79, 85, 104). Efforts to demonstrate endonucleolytic activity with the purified IN protein *in vitro* were first successful with the avian retrovirus IN. The purified AMV pp32 was shown to specifically protect the ends of the LTRs *in vitro* against DNase I degradation (79, 85). In addition, this IN protein and recombinant avian retroviral IN protein produced in *E. coli* were shown to have endonucleolytic activity, first on substrates containing the LTR/LTR junction present on circularized viral genomes (34, 35, 53–55, 127) and then on dimeric oligonucleotides representing the individual LTR termini (75). IN specifically removes the terminal dinucleotide representing the 3' end of the LTR from the dimeric oligonucleotide.

Demonstration that donor DNA in the MoMLV reaction could be separated from the nucleocapsid was achieved *in vitro* using a linearized plasmid with partial LTR sequences at its termini as an exogenous donor, rather than replicated DNA within the virion core (42). In this reaction, the exogenous DNA was integrated into a λ DNA target. Recently, partially purified,

recombinant MoMLV IN protein produced in baculovirus has been shown to catalyze incorporation of dimeric oligonucleotides representing the ends of the MoMLV LTRs into naked target DNA (R. Craigie, personal communication). As in the original *in vitro* reactions, neither exogenous energy cofactors nor supercoiling of donor or target DNAs is essential. The recovered products of these reactions are consistent with insertion into a 4-bp staggered cut in the DNA target. These results are thus consistent with IN catalysis of both the donor and target cleavage and ligation of the 3' ends of donor molecules to the 5' ends of the target.

To what extent independence of the latter *in vitro* reaction of nucleocapsid and target proteins reflects requirements of the reaction *in vivo* has not yet been addressed. If the entire nucleocapsid is required for integration *in vivo*, it could significantly influence the logistics of association between viral DNA and chromatin. In the original retroviral and Ty1 *in vitro* systems, integration was assayed biologically, and insertions that interrupted essential genes in the target were not recovered. This restriction notwithstanding, integration appeared to be relatively random (18, 37, 42). Nevertheless, the intergenic regions are somewhat more AT-rich than are coding regions. It will be of interest to determine whether AT bias is a feature of target sites when products are not recovered via biological selection. As described below, the apparent randomness of *in vitro* insertions is at variance with *in vivo* observations.

Nonrandom Retroviral Integration

Although retroviruses can insert in many different sites (64), much evidence suggests that integration of retroviruses is not random *in vivo*. Different approaches have been taken, first to search for conserved insertion sequence motifs and second, to determine functional correlates of regions that are integration targets. The results of these analyses of *in vivo* integration patterns are consistent with a combination of restricted access and more specific targeting interactions.

Characteristic short direct repeats of target DNA flank retroviral insertions (29, 114, 115). The sequences of short direct repeats and flanking DNA have been inspected for common primary sequence features or potential common secondary structures for several insertions of spleen necrosis virus (114). This study showed that the direct repeats were 77% AT rich, compared to flanking cellular DNA at about 58%, but did not reveal other distinctive features. Extensive comparisons of junction sequences generated by *in vivo* integration of other retroviruses have not been published.

Several experiments performed in the mid-eighties with MoMLV infection of murine cells suggested that integration was nonrandom, and that DNA integrated by retroviral infection, compared with DNA introduced by transfection, tended to be in transcribed regions of the genome. Insertion of

retroviruses into selectable genes allows rough determination of the frequencies of insertion into a given portion of the genome. King et al (76) measured the incidence of spontaneously occurring hypoxanthine phosphoribosyltransferase deficiency (hprt-) in F9 embryonal carcinoma cells, compared to the frequency in cells with 20 to 50 integrated MoMLV proviruses. Infection resulted in approximately a tenfold increase in the frequency of hprt- cells. Nevertheless, in a screen of 10^8 cells, if insertions within coding and intron regions disrupted hprt function, about 10^4 mutants would be expected, compared to the 70 actually recovered. These data establish that integration did not occur at this transcribed locus at the frequency predicted by random insertion throughout the genome, but they do not show where insertions did occur.

Despite the low frequency of hprt- cells generated by MoMLV insertion mutagenesis, results of other experiments have suggested that, overall, insertions are associated with transcribed regions of the genome. Expression of genes at foreign chromosomal locations reflects local transcriptional activity. These effects are attributed to the degree of condensation of chromatin (56). The expression of integrants can then be used to gauge whether insertion occurred into a region with potential for transcriptional activity. Hwang & Gilboa (66) compared the expression of integrated *neo*^R sequences introduced into NIH 3T3 cells by calcium phosphate transfection and by retroviral infection. The amount of transcripts generated from viral insertions did not show much variation among 20 strains and was 10- to 50-fold higher than the amount in transfected cells. These workers concluded that the results were consistent with integration into transcribed regions, while cautioning that negative effects of pBR vector DNA on expression of the transfected *neo*^R gene could not be ruled out. More recently, Scherdin et al (108) screened a recombinant integration library generated from MoMLV-infected NIH 3T3 cells and showed that about half of the clones contained sequences that hybridized to radioactive probes generated in nuclear run-on reactions. Although these results cannot be extrapolated to quantitate a preference, they are consistent with a significant proportion of insertions into transcriptionally active regions.

Efforts have been made to further define the association between transcribed regions and retroviral insertion sites. Transcribed or potentially transcribed regions have been shown to correspond to regions of heightened nuclease sensitivity. In addition, they contain nuclease hypersensitive regions, so-called because they are about two orders of magnitude more sensitive to nuclease digestion. Where studied, these regions of nuclease hypersensitivity are nucleosome-free and allow access of proteins to specific DNA sequences. These proteins include topoisomerases, RNA polymerases, and transcription factors. Both polymerase II- and polymerase III-transcribed genes have been associated with nuclease hypersensitive sites. Several studies

have attempted to localize the positions of multiple insertions of ASLV or MoMLV within transcribed regions relative to DNase I hypersensitive sites. In a study by Robinson & Gagnon (100), 66 insertions of ALV upstream of the second exon of *c-myc* were isolated from chicken lymphoblastomas. Twenty-four of these were mapped in some detail and were positioned within 150 bp of one of five DNase I hypersensitive sites in a 3-kbp region 5' to the protein coding sequences of *c-myc*. Because these insertions were selected on the basis of the associated effect on *c-myc* expression, the conclusion that insertions are associated with DNase I hypersensitive sites is qualified. Nevertheless, these experiments suggested that insertions might be close to nucleosome-free regions with *cis*-acting function.

Two later studies addressed the problem of selection to some extent. In a study by Vijaya et al (132) in which seven insertion sites of MoMLV were mapped, three had been selected as tumor derivatives and four were considered unselected, although two of these were derived from MoMLV-induced rat thymomas and two from a rat fibroblast cell line. Each target site mapped within 500 bp of a hypersensitive site. Overall, these workers estimated a density of DNase I hypersensitive sites of about one every 6.5 kbp and the probability of this association occurring randomly at approximately 9×10^{-3} . In a related study by Rohdewohld et al (101), six MoMLV insertions into mouse early embryonic cells or into NIH 3T3 fibroblasts were analyzed for an association with DNase I hypersensitive sites. At a similar estimate of DNase I hypersensitive site density, the probability of all six sites occurring within the observed range of less than 0.7 kbp was estimated at less than 2×10^{-3} .

Although selected integrants represented a large proportion of the total sample analyzed, the uniform conclusion in the three studies was that a nonrandom association exists between insertion sites and DNase I hypersensitive sites. If this association represents actual bias in the insertion site distribution, as seems likely, it poses another, more intriguing question. Namely, which genomic feature(s) is common among regions containing DNase I hypersensitive sites and targets of integration? Clearly, even given the vagaries of restriction map analysis, these insertions are neither extremely close (within one nucleosome), nor at a uniform distance from DNase I hypersensitive sites. Thus, whether integration and these sites are specifically related or whether both associated with some common feature is not yet resolved. Results of integration studies in the yeast systems described below suggest that one such feature could be the 5' ends of transcriptionally active genes.

To directly screen for sequences common to multiple retroviral integration sites, Shih et al (113) constructed integration libraries and control libraries from RSV-infected turkey embryo fibroblasts. RSV marked with the bacterial *supF* gene was used to infect turkey embryo fibroblasts. Libraries of genomic

DNA were constructed in a derivative of λ gtWES. The integration library was recovered on a wild-type host and the control library on a suppressor-containing bacterial strain. Screening the integration library with probes derived from integration sites showed that 20% of the tested probes were represented in the independently constructed integration libraries and were therefore highly represented in these libraries, relative to control libraries. Sequence analysis of four of the highly used sites showed that they were not identical in sequence and that the two or three insertions recovered at each were into the same position at the respective loci. These insertion sites were estimated to be used once in 4,000 integrations, resulting in an overrepresentation in the integration libraries, compared to the control library, of about one-millionfold. The number of the sites present in the tested collection and the frequency of usage suggested that there are about 800 of these sites for an overall frequency of every 2.5×10^6 bp. They are less frequent than DNase I hypersensitive sites, which occur about every 6 kbp and they do not appear to be transcribed. These regions are therefore not equivalent to DNase I hypersensitive sites, although they could represent a subset of those sites. Preliminary results based on fractionation of genomic DNA suggest that the sites identified in this study cofractionate with regions of nuclear matrix attachment (J. Coffin, personal communication). Determination of the basis of preferential usage of these targets should offer interesting insights into the process of retroviral integration in vivo.

NONRANDOM RETROTRANSPOSON INTEGRATION

Drosophila Retrotransposons

Approximately ten well-characterized families of *Drosophila* retrotransposons are present in 30 to 50 copies in the *Drosophila* genome (7). The gypsylike family is a superfamily with retroviruslike organization of its coding domains and similarities among *pol* coding domains. It includes the elements gypsy (83), 17.6 (105, 129), 297 (69), tom (125), and possibly H. M. S. Beagle (118). These elements are most similar in the protein sequence of the polymerase domain of reverse transcriptase to caulimoviruses (33, 140), Ty3 (33), Tf1 (H. Levin & J. Boeke, personal communication), and del (117). These elements are distinct from some other retrotransposons and retroviruses in that the primer region is apparently not separated from the LTR by two bases (and therefore the extrachromosomal form, if it has extra bp, could not derive them from this region of the template), and the LTRs do not terminate in the TG/CA inverted repeat. Examination of sequences flanking a large number of endogenous gypsylike elements has led to the conclusion that these elements insert with sequence specificity. Although the DNA sequence at the insertion site, as well as across LTR-genomic junctions, has been

determined in several cases, similarity between the presumed ends of the elements and target DNA has made unequivocal definition of the termini difficult. Despite this ambiguity, similarities exist between the short flanking direct repeats of gypsy and H. M. S. Beagle and among the flanking repeats of 17.6, 297, and tom. Analysis of several gypsy (41, 83) and one H. M. S. Beagle (118) insertion sites in variant and wild-type (uninterrupted) configurations is consistent with the target sequence 5'TAYATA3', where Y=pyrimidine, and the 4-bp direct repeat is TAYA. Analysis of insertions of 297 (67), 17.6 (68), and tom (125), suggest a target of TATATA and a repeat of (T)ATAT, depending on assignment of the element termini.

Despite the apparent sequence specificity, host proteins may also play a role in targeting insertions to the observed sites. A functional correlate of the consensus target TATATA is the *Drosophila* TATA box. In fact, several gypsy-group insertions were into the TATA boxes of known genes (67, 68, 118). Moreover, some of these elements appear to display additional preferences. The element tom is linked to a hypermutability phenotype that is most apparent for genes involved in optic morphology (62). The common feature of these genes or tissues contributing to hypermutability has not been identified, but a specific target for integration could exist.

YEAST RETROTRANSPOSONS

Four classes of retrotransposons have been described in *S. cerevisiae*. Ty1-3 (9, 12) are mentioned above and Ty4 has been described more recently (123). Insertions of isolated LTR sequences, as well as complete Ty elements, are found in the genomes of common laboratory strains. In strains expressing high levels of Ty1 (1, 45, 87) or Ty3 (L. Hansen, D. Chalker, & S. B. S., unpublished results) RNA, a nucleocapsid particle is formed that is largely composed of structural proteins encoded by the first ORF. In addition, these particles contain Ty genomic RNA, full-length DNA, and reverse transcriptase activity. These nucleocapsid particles are analogous to the virion core particle within which reverse transcription of the retroviral genome takes place. Genomic Ty elements are flanked by direct repeats of 5 bp that are generated upon integration. An in vitro integration system has been developed for Ty1 and displays requirements for activity similar to those displayed by the MoMLV in vitro integration system (37). Thus, many aspects of the transposition cycles of these elements resemble retroviral replication and integration. Of the yeast elements, the integration patterns of Ty1 and Ty3 have been examined most extensively and are consistent with the suggestion from retroviral integration that insertion is related in some way to transcription.

TY1 Ty1 is 5.9 kbp in length and is composed of an internal domain flanked by direct repeats of delta, which is 334–338 bp in length. Ty2 is closely related to Ty1, based on gene organization and predicted protein sequence, and also has delta LTRs. The minus strand primer binding site of these elements is adjacent to the LTR with no intervening bp. There are from 30 to 35 Ty1 elements in typical haploid strains of *S. cerevisiae* and over 100 delta elements (20). Ty2 is present in 5 to 15 copies (77).

Ty1 integration has been monitored in several different ways. There is a large collection of data representing analysis of endogenous Ty1 elements. Ty1 insertions are often detectable by effects on target gene expression. Insertions can inactivate, activate, or confer constitutive expression or mating-type regulation on neighboring transcriptional units (12). Recombinant techniques have facilitated the recovery of de novo insertions of the Ty1 element. Boeke et al (11) fused the inducible *GAL1* promoter to the Ty1 transcription start site and achieved about 100-fold increase in transposition frequency. In addition, Boeke et al (14) and Garfinkel et al (46) have used genetically marked elements to identify and clone insertions, making efficient recovery of otherwise unselectable insertions straightforward.

Mapping and cloning of a large number of endogenous Ty1 and Ty2 elements have allowed some conclusions to be drawn concerning their genomic distribution. First, they tend to be associated with AT-rich regions (92); second, several examples of clustered elements exist that could be partially explained by a high frequency of insertion of one element within another, or by the presence of some common target (135, 136); and third, Ty1 and Ty2, as well as Ty4, are frequently found in association with tRNA genes (38, 44, 61, 77, 90, 122). These observations, although interesting, are of limited quantitative value because of their anecdotal nature. The association of endogenous Ty1 elements with AT-rich regions could reflect the high AT-base composition of the yeast genome, coupled with the greater likelihood of retaining Ty elements in relatively AT-rich intergenic regions. Clusters of Ty1 elements that are in the vicinity of tRNA genes have been described on chromosomes X (112) and III (135). Ty1 and Ty2 (20, 77, 91), Ty3 (25), Ty4 (123), delta (20, 91) tau (23) and sigma (27, 107) elements were each discovered as polymorphisms in the vicinity of tRNA genes, and many representatives of each class are within 500 bp of tRNA genes (30, 38, 44, 90, 122). That the association is nonrandom can be inferred, because yeast has an estimated 360 tRNA genes and, for the most part, these are dispersed in the genome (57). Thus tRNA genes occur about every 40 kbp while Ty1 would occur roughly every 400 kbp and Ty2 elements even less frequently (see below). The extent of a general association between Ty1 elements and genes transcribed by polymerase III is not clear. Ty1 sequences have also been associated with 5S sequences (97, 133) and a delta element has

been identified about 100 bp upstream of the single-copy U6 gene in yeast (D. Brow & C. Guthrie, personal communication). The association between tRNA genes and Ty3 is more defined and is discussed below. This association will likely differ qualitatively from the association observed between the other elements and tRNA genes.

Multiple spontaneous de novo integrations of Ty1 have been characterized in the vicinity of *HIS3* (11, 13), *ADH2* (139), *LYS2* (36, 89, 116), *URA3* (89), *CAN1* (138), and *SUP4* (48) on the basis of disruption or activation of the target gene. Interesting patterns of integration have emerged from these studies. Nevertheless, the profile of insertions in most of these systems is necessarily influenced by the nature of the selectable phenotype. This complication is particularly severe for insertions that increase expression, because of the requirement that Ty1 be oriented so that the promoter-proximal enhancer is close to the target gene and that Ty1 insertion does not directly or indirectly block transcription of the target gene.

Four insertions of Ty1 and one of Ty2 which conferred constitutive expression on *ADH2*, the gene for the glucose-repressible alcohol dehydrogenase, were mapped. These insertions occurred from -125 to -210 upstream of the translation initiation site. None of the five insertions examined was close to the characterized UAS, but two of the insertions were within 10 bp of the TATA box.

A *HIS3* gene in which λ DNA replaces the promoter region, carried on a centromere-bearing plasmid, can be activated by Ty1 insertion as a convenient method of scoring transposition. Yeast plasmids are found in nucleosome structures (95) and so should resemble chromosomal targets in at least some respects. A collection of 40 integration sites were mapped in a strain in which expression of Ty1 on a high-copy plasmid was under control of the *GAL1* promoter (13). In this study, a hotspot at which insertion of more than half of the elements occurred was located at position -60, relative to the *HIS3* ATG. Although this target is located on a plasmid and transcription in the vicinity of the *HIS3* gene has not been mapped, the gene is not expressed sufficiently to rescue a His⁻ phenotype. This existence of such an integration hotspot argues that although the 5' flank of some target genes is used preferentially for insertion, specific initiation of transcription in the upstream region may not be absolutely essential for use of that gene as a target, or for observation of insertion hotspots.

Selection schemes based on loss of activity of the integration target extend the potential of the Ty1 system for detection of insertions within a gene, but conversely may limit recovery of insertions into the region upstream of transcription initiation. Mutants that have lost function of *LYS2*, *CAN1* and *URA3* can be directly selected in yeast. Eight of ten Ty1 or Ty2 insertions which caused α -aminoadipate resistance by disruption of *LYS2* were shown

to be insertions in the 5'-noncoding region of that gene (36). In a more extensive study, using the galactose-inducible Ty1 element on a high-copy plasmid, 31 out of 78 resistant colonies could be attributed to Ty1 insertions in the *LYS2* gene (89). These insertions were distinctly nonrandom. Of the ten insertions for which junction sequences were determined, half were in the 5'-transcribed, noncoding region (four at position -42, relative to the coding sequence and one further upstream) and the other half were at position +226. Although none of these insertions were upstream of the transcription start site at -82, out of a coding region of about 4 kbp, this insertion distribution was heavily biased toward the 5' region of the gene.

Ty1 insertions that disrupt the *URA3* gene and confer resistance to 5-fluoroorotic acid (5FOA) have also been analyzed to examine positions of integration (89). Sequence analysis of the positions of 82 independent insertions showed that, as in the *LYS2* locus, insertion was biased toward the 5' end and insertion hotspots occurred. The upstream one-fourth of the gene contained more insertions than the downstream three-fourths. The pattern of insertion of at *URA3* differed from that at *LYS2*, however, in that very few insertions (2/82) were recovered in the noncoding region, probably reflecting the requirement for complete loss of activity to establish 5FOA resistance. The 82 insertions occurred at 42 positions in the 800-bp gene and one position, bp 341, was used 12 times. No particular bias in the orientation was apparent inside the coding region.

Wilke et al (138) characterized 55 spontaneous insertions of Ty1 or Ty2 elements that disrupted expression of the gene for arginine permease, *CAN1*, and thereby conferred resistance to the toxic analogue canavanine. In one strain, 10/21 of the events mapped in the 5'-noncoding region. Most other insertions were mapped in two *rad6* mutant strains. In these strains, no insertions were recovered in the 5'-noncoding region and insertions were distributed throughout the coding region. The distribution of insertions at this locus in the wild-type strain are consistent with observations at *LYS2* and *URA3* in that a bias toward insertions in the 5' end was observed. The absence of a bias in the *rad6* strain is potentially interesting, as the *RAD6* gene encodes an enzyme that polyubiquitinates histones H2A and H2B in vitro (70, 124). Therefore an effect of a mutation in *RAD6* on the pattern of Ty1 insertion could implicate nucleosomes in defining regions of retrotransposon insertion. It is also possible that *rad6* mutants simply have more tolerance for insertions in the 5'-noncoding flanking sequence of the *CAN1* gene. This could be resolved by determining the canavanine resistance phenotype of 5' Ty1 insertions crossed into the *rad6* background.

Consensus sequences summarized by Wilke et al (138) and Natsoulis et al (89) for Ty1 insertion at a number of loci suggest that the immediate insertion site sequence is barely constrained. Both studies suggested that the 5-bp

repeat sequences are three-way degenerate on the ends and AT-rich in the center.

Targeting of Ty1 and Ty2 to tRNA genes was examined directly by Giroux et al (48). Mutants that had lost activity of a gene encoding a tRNA^{Tyr} ochre suppressor, *SUP4-o*, were selected on canavanine-containing medium in a genetic background (*can1-100*) where expression of arginine permease depends on an ochre suppressor. Among the 196 spontaneous mutations recovered were 12 insertions of Ty1 DNA into *SUP4*: one was at position +17, ten at position +37, and one at position +43. The hotspot and the insertion at position +43 were within a run of ten consecutive A/T bp. This region was accessible to DNase I in stable transcription complexes (74). Because of the requirement for loss of *SUP4-o* activity in this selection, it is not likely that Ty insertions upstream of *SUP4* would have been identified.

TY3 Ty3 is 5.4 kbp in length and is composed of a 4.7-kbp internal domain flanked by direct repeats of the 340-bp sigma element (25). From one to four copies of Ty3 (25) and about 30 copies of the sigma element (27, 107) are found in the genomes of typical haploid laboratory strains of *S. cerevisiae*.

Ty3 and sigma elements are closely associated with the 5' ends of tRNA genes. Southern blot analysis of 25 sigma elements from one strain showed that all were on the same restriction fragments as tRNA genes (16). This implies that about 10% of the estimated 360 tRNA genes are associated with Ty3 or sigma elements. Analysis of genomic DNA flanking two Ty3 and over thirty sigma elements from a combination of several strains, reviewed in Sandmeyer et al (106), showed that each is at position -17, plus or minus one or two bases, relative to the 5' end of a mature tRNA-coding sequence. A comparison of 20 of these sequences revealed only limited similarity among the insertion sites, however. In 11 of 20 genes, it could be inferred that the tRNA gene-proximal nick for Ty3 integration would have occurred on the nontranscribed strand within the dinucleotide sequence CA. This coincides with the presence of a common context for tRNA gene transcription initiation at the first A in the nontranscribed strand sequence PyCAACA (47). Insertions of Ty3 are unevenly distributed between the two possible orientations relative to the tRNA gene. Two-thirds of the sigma and Ty3 insertions are oriented so that transcription of the element is divergent from that of the tRNA gene. Such elements are said to be in the major orientation.

Despite the fact that Ty3 has a distinctive pattern of integration, its mode of integration is likely to be fundamentally similar to that used by retroviruses. Similarities of the yeast retrotransposons to animal retroviruses were mentioned previously. The Ty3 IN protein(s), p61/58 (59), has been identified by Western blot analysis and shown to be required for integration. The predicted protein sequence of Ty3 IN is more similar to MoMLV IN than to the other

Ty IN sequences. Ty3 IN has amino- and carboxyl-terminal extensions in addition to a large central region of similarity with characterized integrases.

To investigate the basis of the position-specific integration displayed by Ty3, a study was undertaken to identify genomic sites of *de novo* integration (D. Chalker & S. B. S., submitted). As in the Ty1 studies, transcription of the element was artificially elevated to enhance transposition. In this case, the *GAL1-10* upstream activating sequence (UAS) was fused upstream of the Ty3 TATA sequence (58). Transposition of a Ty3 element genetically marked with the yeast *HIS3* gene and the bacterial *neo^R* gene was monitored. The bacterial *neo^R* gene confers resistance to kanamycin in bacteria and to G418 in yeast. A total of 91 strains that had acquired one additional Ty3 element during induction of transposition were identified by Southern blot analysis. Seven genomic loci were the targets of two insertions each. This is close to the number of double insertions predicted by a Poisson distribution, if the actual number of targets is roughly equivalent to the estimated number of tRNA genes. The junctions of twelve of these insertions with host DNA were also characterized by sequence analysis, which showed that the seven loci corresponded to seven tRNA genes. Interestingly, in each case, the Ty3 insertion at a given locus was in one orientation only, although the minor orientation was represented at three loci. Therefore, some feature of the genomic target, rather than an inherent asymmetry in the transposition apparatus, determines the orientation of the insertion. The position of insertion was not absolutely conserved with respect to a given tRNA gene sequence, however, as two insertions upstream of a tRNA^{Met} gene occurred with one insertion at position -16 and one at position -17.

The 5S genes, which are also transcribed by polymerase III, have some features in common with tRNA genes. Yeast has about 150 5S genes, almost all of them included within a large tandem rDNA repeat on chromosome XII (96). These were examined to determine whether common features among polymerase III-transcribed genes were sufficient to constitute a genomic Ty3 target. Southern blot analysis showed that none of the 91 strains which acquired a single Ty3 element in the previous experiment had undergone an integration of Ty3 into a 5S gene.

To investigate a role for homology or tendency of Ty3 elements to cluster at preferred sites, Southern blot analysis was also performed using a sigma-specific probe. The 30 tRNA genes already associated with sigma elements were targets of only four insertions, which is slightly less than would be predicted if these were used randomly. Limited homology between the donor and target, in this case, therefore, did not confer preferential usage on the target. Thus, the most frequent targets of Ty3 genomic integration are tRNA genes, but there does not appear to be a highly preferred subset of those genes.

More recently, an assay was devised to allow recovery of Ty3 insertions into a plasmid-borne target (D. L. C. & S. B. S., unpublished results). The target assay has been used to identify and dissect the features of the target for Ty3 integration. Eight genes encode a single tRNA^{Tyr} isoacceptor species in yeast (49, 82, 91). The *SUP2* gene for tRNA^{Tyr} was chosen for the test target, because it is associated with sigma in some strains (107) and because the promoter sequences are relatively well-characterized (2,4). The major transcription initiation site for *SUP2* transcription by polymerase III is at position -10 (P. Kinsey & S. B. S., unpublished results). Ty3 insertions upstream of *SUP2* carried on the plasmid occurred in both orientations at positions -16 and -17. The staggered nicks inferred to accompany Ty3 insertion would have occurred between bp -15 and -16 and between -10 and -11 for a Ty3 insertion beginning at -16. Thus, the tRNA gene-proximal nick is in the vicinity of the transcription initiation site. The genomic integration pattern together with these data suggested that the common denominator marking Ty3 integration sites is related to target gene transcription.

The plasmid target assay was used to more directly investigate the relationship of Ty3 transposition to transcription of the tRNA gene target. The essential features of the tRNA gene internal control region (ICR) or promoter are the conserved box A, positions 8 to 19, and box B, positions 52 to 62, relative to the mature tRNA-coding sequence (reviewed in 47). Two transcription factors are required, in addition to polymerase III, for in vitro transcription of tRNA genes, TFIIC and TFIIB. TFIIC contacts box B and to a lesser extent, box A. Transcription initiation is typically at a purine in the nontranscribed strand, the selection of which is influenced by the position of box A sequences (47). TFIIC binding is required to load TFIIB onto the DNA upstream of the tRNA gene, completing stable complex formation (74). Although no recognition sequence has been identified for TFIIB, this factor footprints in DNase I protection assays over the region from -16 to -54 upstream of the tRNA^{Tyr} gene *SUP4* in the presence (74) and absence (73) of other transcription factors and is the only factor required for repeated rounds of polymerase III transcription (73). In vivo footprint analysis shows that this upstream region is protected in cells in logarithmic growth and in stationary phase (65).

Mutations were introduced to a *SUP2* gene to assess the target requirement for ability to form stable complex and to test for linkage between the transcription initiation site and the position of integration (D. L. C. & S. B. S., unpublished results). The first mutation in a target tRNA gene disrupted the ability of the tRNA gene to form a stable complex by changing a conserved base in box B (2, 4, 80, P. Kinsey & S. B. S., unpublished results). This mutation also made the tRNA gene unable to act as an integration target. To investigate the functional linkage between the position of Ty3

integration and the initiation site of tRNA gene transcription mutations that changed the context of initiation were also examined. Ty3 integration into these target plasmids occurred as into the wild-type target, near the transcription initiation site. Together the data from these mutant studies argue that the integration site is fixed by a protein-DNA complex, rather than by a specific DNA sequence.

The results of the tRNA gene target experiments implicated some feature of polymerase III transcription in Ty3 target formation. Two other yeast genes transcribed by polymerase III, the 5S gene (47) and the U6 gene (17, 85a), were therefore tested using this assay (D. L. C. & S. B. S., unpublished results). Integration of Ty3 occurred near the transcription initiation site of the plasmid-borne 5S gene. The 5S gene in yeast is dependent for expression on two regions, one bounded by positions -14 to +8 and box C, positions +81 to +94 (22). A conserved box A is also present, but is apparently dispensable in yeast for 5S expression. Transcription of the 5S gene is dependent on TFIIA, in addition to TFIIC and TFIIB, the factors required for tRNA gene transcription, and on polymerase III (126). In DNase I protection studies, TFIIB extends the protection of TFIIC and TFIIA upstream to bp 45 with partial protection at position -4 (15). When the protection of TFIIB alone is examined after removal of other factors, it protects positions -8 through -43 (73). Thus for both the tRNA gene and for the 5S gene, the position of insertion corresponds closely to the downstream border of TFIIB protection and is near the transcription initiation site.

Why the genomic 5S genes were not a target for transposition is not known, although there are several interesting possibilities. The integration complex may be simply excluded from the nucleolus, the structure of which is not clearly understood. The particular 5S gene used as a target was subcloned from rDNA repeat and is transcriptionally active (15). Alternatively, because the 5S genes are in the rDNA spacers, the flanking configurations or activities of promoter, enhancer, and terminator sequences (72) may somehow preclude integration. Topological constraints which operate to repress recombination within rDNA may also make rDNA refractory to integration (39). Whether or not Ty1 integration into rDNA occurs readily is not known. Ty1 sequences are underrepresented in the rDNA (92) and the delta and Ty1 sequences associated with rDNA that have been characterized are either associated with variant rDNA flanking regions and not located within the major repeat (97) or do not exhibit the flanking direct repeats characteristic of Ty1 insertions by retrotransposition (133). The R1 and R2 elements of insects which insert specifically into rDNA are also in contrast to Ty3 exclusion from this genomic target.

The yeast U6 gene is a member of a class of polymerase III-transcribed genes that share some promoter features with polymerase II-transcribed genes, including an upstream TATA box consensus (17). The required promoter elements for expression of animal cell 7SK (88) and U6 (26) genes are

contained in the upstream sequences. In the yeast U6 gene, the upstream sequence, which contains a TATA consensus, is required for full activity in vivo and in vitro as observed for animal cell genes (D. Brow & C. Guthrie, personal communication). However, an internal box A consensus also exists and a box B sequence is present 120 bp downstream of the gene. The box B sequence has been shown to be required for full expression in vivo and in vitro with complete transcription extracts (D. Brow & C. Guthrie, personal communication). Moenne et al (85a) have shown that in vitro expression of the U6 gene depends on the TFIIB transcription factor. Despite potential differences between promoter elements of the U6 and tRNA and 5S genes, a plasmid-borne copy of the U6 gene functioned as a target for insertions of Ty3 into the vicinity of the transcription initiation site. The results of experiments with the target plasmids are consistent with several potential target requirements, including features of the stable complex or even the open transcription complex.

Dictyostelium Retrotransposons

Several mobile elements have been characterized in the slime mold *Dictyostelium discoideum* (40). These include Tdd1-3 (Tdd1 = DIRS) and DRE1/2 (now DRE). The element Tdd-1 is 4.8 kbp in length and has inverted LTR sequences of 313 bp in length (102, 141). Insertions of this element are not flanked by direct repeats of the target sequence. Despite these differences with characterized retrotransposons, Tdd-1 encodes a reverse transcriptase that is likely to be involved in its replication. Analysis of a number of Tdd-1 genomic insertions shows that they are often located within Tdd-1 elements (21, 40). The pattern of these insertions would be difficult to completely reconcile with an homologous recombination model. An alternate possibility would be that Tdd-1 elements are themselves preferential targets for Tdd-1 insertions. Because the position of insertion is not conserved, either regional hotspots, acting as foci of multiple insertions or a specific sequence target within Tdd-1, but not necessarily at the integration site, would be consistent with an integration specificity model.

The elements Tdd-2 and Tdd-3 are about 4.5 kbp in length, but do not have the LTRs characteristic of retrotransposons. Insertions of Tdd-3 are flanked by 9–10 bp repeats of extremely AT-rich DNA. Analysis of a number of Tdd-3 insertions reveals that a number are located within Tdd-2 elements, close to a 5'22-bp region of homology between the two elements (98). Several other genomic Tdd-3 elements have been identified about 100 bp downstream of tRNA genes (84) or inserted into the 3' ends of other Tdd-3 elements. An intriguing feature shared by these hotspots for Tdd-3 insertion is that they contain consensus box B sequences (T. Dingermann & R. Marschalek, personal communication).

The elements DRE1 and DRE2, previously described by Marschalek et al (84) are the LTR and internal domain, respectively, of a retrotransposon of about 6.6 kbp in length, which is now named DRE. The internal domain of this element is composed of two ORFs which encode protein sequences with similarity to retroviral NC, PR, RT, and IN proteins. A survey of 50 out of about 200 DRE elements in the *Dictyostelium* genome shows that they are all associated with tRNA genes. Sequence analysis of the regions flanking 12 elements has shown that, as in the case of the original isolates, they occur consistently, about 50 bp upstream of the 5' ends of different tRNA genes (T. Dingermann & R. Marschalek, personal communication). These observations are particularly intriguing in light of the work from the Geiduschek laboratory showing that TFIIB protects a region in the DNA about this far upstream of the tRNA gene. If this is also the case for an analogous factor in *Dictyostelium*, then this factor could be involved in target formation.

Polymerase III-Transcribed Genes as Genomic Integration Targets

The observations of a frequent association of Ty elements with tRNA genes and the relatively specific association of Ty3, Tdd-3, and DRE with tRNA genes would suggest that mechanisms have evolved to use polymerase III promoters as targets of integration events. although it is not known whether recombination hotspots are correlated with preferred integration targets, tRNA genes are also active in ectopic meiotic gene conversion (3). Whether the integration targeting functions have been adapted from cellular transcription and recombination functions can be better addressed when additional participants in these processes have been characterized at the molecular level. Recurring motifs in protein-protein and DNA-protein interactions suggest that integrases and transcription factors could interact directly or could have common domains for contacting other proteins or DNA. Versions of the metal finger motif, for instance, are common to the integrases and also occur in TFIIA, the 5S transcription factor.

tRNA genes may be present in distinctive regions of chromatin, such as are postulated for retroviral insertions, or may even be distinguished topologically. Nuclease analysis of *Drosophila* chromatin showed that tRNA and 5S genes are associated with positions of nuclease hypersensitivity (28, 81). Microscopic analysis of tRNA and 5S genes associated with transcription proteins has produced results consistent with protein-induced changes in the structure of the DNA (5, 109, 121). DNA bends have been demonstrated to participate in integration complex formation (50, 119) and even to promote cleavage in some nonretroid transposition systems (110). Electrophoretic analysis of permuted fragments containing a yeast tRNA gene complexed with TFIIC or TFIIB showed evidence of an induced bend in the vicinity of

the transcription initiation site (T. L  veillard, G. Kassavetis, & E. P. Geiduschek, personal communication). One intriguing possibility is that such a conformational marker could play a role in targeting integration to genes transcribed by polymerase III.

SUMMARY

Analysis of in vivo integration patterns has provided no data to support the notion that retroelement integration is random. Rather, the diversity of insertion patterns of retroelements suggests numerous ways in which genomic DNA is identified for preferential targeting. These range from specific to general and include sequence content, removal of chromatin proteins, nuclear localization, distinctive topology, and association with particular *trans*-acting factors. Many are similar to mechanisms already demonstrated to affect activities of previously described recombinases. Moreover, such proposed targeting mechanisms could act directly or indirectly to influence integration site selection.

A variety of observations are consistent with the preferential use of open chromatin for retroelement insertion. The site-specific retroelements insert into transcribed regions. In vitro studies with retroviruses and Ty1 have shown that naked DNA can function, at least under some conditions, as a target. The association of integration sites of retroviruses and regions in which DNase I hypersensitive sites exist and the preferential integration of Ty1 at the 5' ends of some genes might suggest that regions which do not have phased nucleosomes are targets for integration. Is targeting to transcriptionally active regions essentially passive, because they are not densely associated with chromatin proteins, or is targeting active in the sense of being a more specific process?

Specific targets could be generated from DNA or protein motifs. Nucleosome-free regions are associated with a variety of nonnucleosome proteins, including topoisomerases, nuclear matrix proteins, transcription factors, or replication proteins. These then are candidates for proteins which target integration directly, by associating with the transposition complex or, indirectly, by inducing changes in the DNA. Polymerase III-transcribed genes, which are relatively defined targets of integration for some retrotransposon systems, probably exemplify several of these mechanisms. Promoter sequences may be directly involved in targeting some elements and positioning of the transcription complex may fix the integration sites of others.

The most common sequence feature of characterized in vivo insertion sites is that they are AT-rich. This may reflect specificity of the IN protein, particularly the gypsylike elements, increased nicking of DNA, which is relatively weakly base-paired, as appears to be the case for the FLP recom-

binase (130), or simply the AT content of accessible regions in chromatin. Some of these questions will be resolved by the characterization of in vitro integration sites that have been recovered by physical means, rather than by biological assay. The insertion patterns of a number of retroposons suggest that retroelements can insert with a high degree of sequence specificity. Furthermore, expression of recombinant endonucleases has shown that in a heterologous host transcription is not required for site-specific cleavage of intragenic sequences (140a).

A relatively unexplored issue is the role of nuclear localization in targeting. Three kinds of effects have been considered: First, the possibility that the position of DNA relative to the nuclear matrix might make it more or less accessible to the integration apparatus is quite intriguing. It would allow specific predictions to be made concerning what constitutes the integration complex and what physical constraints operate on its delivery to the integration site. Second, the genomic context of the target may play a role in determining the orientation of insertions. In the case of Ty3, genomic targets, but not 2 micron-based, plasmid targets, appear to specify insertion orientation. This suggests that position of the target relative to some chromosomal feature, rather than the target or immediate flanking DNA, is the orientation determinant. Third, position effects play a role in determining the frequency with which some targets are used. The observation that genomic integration of Ty3 upstream of 5S genes was not observed, yet integration upstream of a 5S gene by itself or in tandem with a tRNA gene on a plasmid, suggests that the use of 5S sequences as a target is restricted by positioning within the rDNA repeat or by the nucleolar localization of these sequences.

The extent to which recombinogenic regions are subject to elevated retroelement insertion is largely unexplored. Recombination has been correlated with transcriptional activity in several systems including immunoglobulin genes in animal cells (8), and *GAL1* (128), *HOT1* (120), and *MAT* (78) in yeast. Like the hotspots for integration, the 5' ends of several fungal genes have been shown to have hotspots for recombination, at least in meiotic cells. If transcription is a general mechanism by which regions are made accessible for integration, then parallels between recombination and integration frequencies should be observed. A study currently underway to map genomic integrations of a marked Ty1 into chromosome III will address whether broad correlations exist between transcribed regions or recombinogenic regions and integration activity. Preliminary data from this study are consistent with nonrandom integration (D. Moore, G. Natsoulis, M. Nacionales, & J. Boeke, personal communication).

Insertion specificity has been shown for members of the retroposon and retrotransposon classes of retroelements, making it clear that retroelements as a class can have highly specialized integration functions. Given the demon-

strated potential of the RNA genome for diversity, it would be rather surprising if some forms of specialized integration were not identified in additional classes of retroelements.

PERSPECTIVES

Elucidation of the mechanism of retrotransposon and retrovirus integration and target site selection should proceed rapidly in the next few years. Questions to be addressed *in vitro* include: (a) identification of IN domains required for recognition and nicking of donor and target DNAs; (b) testing of potential roles for proteins in addition to IN, e.g. CA, NC, histones, and even transcription factors, in integration; and (c) identification of the energy source that drives the ligation of donor ends to host DNA.

Our increasing understanding of the integration reaction based on *in vitro* studies should benefit the design of experiments to more methodically explore *in vivo* integration. Central issues are (a) the effect of replication on integration; (b) the role of nuclear localization in target site discrimination; (c) the role of transcription and chromatin organization in activation or exclusion of potential integration targets; and (d) definition of the *in vivo* specificities of retroelements.

New tools are available in many systems for these *in vivo* studies. Yeast strains in which the availability of histones and transcription factors can be controlled will be useful in testing hypothetical roles of chromatin structure and transcription. Plasmid target assays using highly regulated promoters and marked yeast elements would be explicitly informative about the role of transcription and UAS sequences in target determination without requiring specific integration phenotypes. In yeast and animal cells, studies with marked elements to reduce experimentally imposed constraints on target site usage and application of the polymerase chain reaction to characterization of individual insertions should accelerate the collection of data. Integration libraries can be specifically screened with probes for some candidate high-frequency targets such as tRNA genes, regions of topoisomerase binding or nuclear matrix attachment, highly transcribed regions and oncogenes. Similarly, analysis of integration libraries could be informative about the unexplored possibility that integration patterns differ among retroviruses. A more refined profile of retrovirus insertion would facilitate their use as mutagens and genetic tags for cloning target genes. Results from *in vitro* and *in vivo* studies can be applied to the design of vectors with controlled patterns of integration; these would have applications in gene therapy and basic research. Improved understanding of retroelement integration will greatly enhance its use as a probe of the eukaryotic nucleus.

ACKNOWLEDGMENTS

We thank V. Bilanchone, J. Boeke, D. Brow, J. Coffin, R. Craigie, T. Dingeramn, E. P. Geiduschek, G. W. Hatfield, P. Kinsey, T. Léveillard, H. Levin, and M. McClure for helpful discussions and for making results available prior to publication. We thank E. P. Geiduschek and D. Brow and C. Guthrie for gifts of clones containing the yeast 5S and SNR6 genes. Work in this laboratory is supported by a Public Health Service grant from the National Institutes of Health (GM33281) and by a contract from Enterprise Partners to S. B. S.

Literature Cited

- Adams, S. E., Mellor, J., Gull, K., Sim, R. B., Tuite, M. F., et al. 1987. The functions and relationships of Ty-VLP proteins in yeast reflect those of mammalian retroviral proteins. *Cell* 49:111-19
- Allison, D. S., Goh, S. H., Hall, B. D. 1983. The promoter sequence of a yeast tRNA^{Tyr} gene. *Cell* 34:655-64
- Amstutz, H., Munz, P., Heyer, W.-D., Leupold, U., Kohli, J. 1985. Concerted evolution of tRNA genes: intergenic conversion among three unlinked serine tRNA genes in *S. pombe*. *Cell* 40:879-86
- Baker, R. E., Gabrielsen, O., Hall, B. D. 1986. Effects of tRNA^{Tyr} point mutations on the binding of yeast RNA polymerase III transcription factor C. *J. Biol. Chem.* 261:5275-82
- Bazett-Jones, D. P., Brown, M. L. 1989. Electron microscopy reveals that transcription factor TFIIIA bends 5S DNA. *Mol. Cell. Biol.* 9:336-41
- Berg, D. E., Howe, M. M. 1989. *Mobile DNA*. Washington, DC: Am. Soc. Microbiol.
- Bingham, P. M., Zachar, Z. 1989. Retrotransposons and the FB transposon from *Drosophila melanogaster*. See Ref. 6, pp. 485-502
- Blackwell, T. K., Moore, M. W., Yancopoulos, G. D., Suh, H., Lutzker, S., et al. 1986. Recombination between immunoglobulin variable region gene segments is enhanced by transcription. *Nature* 324:585-89
- Boeke, J. D. 1989. Transposable elements in *Saccharomyces cerevisiae*. See Ref. 6, pp. 335-74
- Boeke, J. D., Eichinger, D., Castrillon, D., Fink, G. R. 1988. The *Saccharomyces cerevisiae* genome contains functional and nonfunctional copies of transposon Ty1. *Mol. Cell. Biol.* 8:1432-42
- Boeke, J. D., Garfinkel, D. J., Styles, C. A., Fink, G. R. 1985. Ty elements transpose through an RNA intermediate. *Cell* 40:491-500
- Boeke, J. D., Sandmeyer, S. B. 1990. Yeast transposable elements. In *The Molecular and Cellular Biology of the Yeast Saccharomyces cerevisiae*, ed. J. R. Broach, E. W. Jones, J. Pringle. Cold Spring Harbor, NY: Cold Spring Harbor Lab.
- Boeke, J. D., Styles, C. A., Fink, G. R. 1986. *Saccharomyces cerevisiae* SPT3 gene is required for transposition and transpositional recombination of chromosomal Ty elements. *Mol. Cell. Biol.* 6:3575-81
- Boeke, J. D., Xu, H., Fink, G. R. 1988. A general method for the chromosomal amplification of genes in yeast. *Science* 239:280-82
- Braun, B. R., Riggs, D. L., Kassavetis, G. A., Geiduschek, E. P. 1989. Multiple states of protein-DNA interaction in the assembly of transcription complexes on *Saccharomyces cerevisiae* 5S ribosomal RNA genes. *Proc. Natl. Acad. Sci. USA* 86:2530-34
- Brodeur, G. M., Sandmeyer, S. B., Olson, M. V. 1983. Consistent association between sigma elements and tRNA genes in yeast. *Proc. Natl. Acad. Sci. USA* 80:3292-96
- Brow, D. A., Guthrie, C. 1988. Spliceosomal RNA U6 is remarkably conserved from yeast to mammals. *Nature* 334:213-18
- Brown, P. O., Bowerman, B., Varmus, H. E., Bishop, J. M. 1987. Correct integration of retroviral DNA in vitro. *Cell* 49:347-56
- Brown, P. O., Bowerman, B., Varmus, H. E., 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* 86:2525-29
20. Cameron, J. R., Loh, E. Y., Davis, R. W. 1979. Evidence for transposition of dispersed repetitive DNA families in yeast. *Cell* 16:739-51
21. Cappello, J., Cohen, S., Lodish, H. F. 1984. *Dictyostelium* transposable element DIRS-1 preferentially inserts into DIRS-1 sequences. *Mol. Cell. Biol.* 4:2207-13
22. Challice, J. M., Segall, J. 1989. Transcription of the 5S rRNA gene of *Saccharomyces cerevisiae* requires a promoter element at +1 and a 14-base pair internal control region. *J. Biol. Chem.* 264:20060-67
23. Chisholm, G. E., Genbauffe, F. S., Cooper, T. G. 1984. tau, a repeated DNA sequence in yeast. *Proc. Natl. Acad. Sci. USA* 81:2965-69
24. Clare, J., Farabaugh, P. 1985. Nucleotide sequence of a yeast Ty element: evidence for an unusual mechanism of gene expression. *Proc. Natl. Acad. Sci. USA* 82:2829-33
25. Clark, D. J., Bilanchone, V. W., Haywood, L. J., Dildine, S. L., Sandmeyer, S. B. 1988. A yeast sigma composite element, TY3, has properties of a retrotransposon. *J. Biol. Chem.* 263: 1413-23
26. Das, G., Henning, D., Wright, D., Reddy, R. 1988. Upstream regulatory elements are necessary and sufficient for transcription of a U6 RNA gene by RNA polymerase III. *EMBO J.* 7:503-12
27. Del Rey, F. J., Donahue, T. F., Fink, G. R. 1982. Sigma, a repetitive element found adjacent to tRNA genes of yeast. *Proc. Natl. Acad. Sci. USA* 79:4138-42
28. DeLotto, R., Schedl, P. 1984. Internal promoter elements of transfer RNA genes are preferentially exposed in chromatin. *J. Biol. Chem.* 179:607-28
29. Dhar, R., McClements, W. L., Enquist, L. W., Vande Woude, G. F. 1980. Nucleotide sequences of integrated Moloney sarcoma provirus long terminal repeats and their host and viral junctions. *Proc. Natl. Acad. Sci. USA* 77:3937-41
30. Dobson, M. J., Kingsman, S. M., Kingsman, A. J. 1981. Sequence variation in the *LEU2* region of the *Saccharomyces cerevisiae* genome. *Gene* 16:133-39
31. Donehower, L. A. 1988. Analysis of mutant Moloney murine leukemia viruses containing linker insertion mutations in the 3' region of pol. *J. Virol.* 62:3958-64
32. Donehower, L. A., Varmus, H. E. 1984. A mutant murine leukemia virus with a single missense codon in *pol* is defective in a function affecting integration. *Proc. Natl. Acad. Sci. USA* 80:6461-65
33. Doolittle, R. F., Feng, D.-F., Johnson, M. S., McClure, M. A. 1989. Origins and evolutionary relationships of retroviruses. *Q. Rev. Biol.* 64:1-30
34. Duyk, G., Leis, J., Longiaru, M., Skalka, A. M. 1983. Selective cleavage in the avian retroviral long terminal repeat sequence by the endonuclease associated with the alpha-beta form of avian reverse transcriptase. *Proc. Natl. Acad. Sci. USA* 80:6745-49
35. Duyk, G., Longiaru, M., Cobrinik, D., Kowal, R., DeHaseth, P., et al. 1985. Circles with two tandem long terminal repeats are specifically cleaved by *pol* gene-associated endonuclease from avian sarcoma and leukosis viruses: nucleotide sequences required for site-specific cleavage. *J. Virol.* 56:589-99
36. Eibel, H., Philippsen, P. 1984. Preferential integration of yeast transposable element Ty into a promoter region. *Nature* 307:386-88
37. Eichinger, D. J., Boeke, J. D. 1988. The DNA intermediate in yeast Ty1 element transposition copurifies with virus-like particles: cell-free Ty1 transposition. *Cell* 54:955-66
38. Eigel, A., Feldmann, H. 1982. Ty1 and delta elements occur adjacent to several tRNA genes in yeast. *EMBO J.* 1:1245-50
39. Fink, G. R. 1989. A new twist to the topoisomerase problem. *Cell* 58:225-26
40. Firtel, R. A. 1989. Mobile genetic elements in the cellular slime mold *Dictyostelium discoideum*. See Ref. 6, pp. 557-66
41. Freund, R., Meselson, M. 1984. Long terminal repeat nucleotide sequence and specific insertion of the gypsy transposon. *Proc. Natl. Acad. Sci. USA* 81:4462-64
42. Fujiwara, T., Craigie, R. 1989. Integration of mini-retroviral DNA: a cell-free reaction for biochemical analysis of retroviral integration. *Proc. Natl. Acad. Sci. USA* 86:3065-69
43. Fujiwara, T., Mizuuchi, K. 1988. Retroviral DNA integration: structure of an integration intermediate. *Cell* 54:497-504
44. Gafner, J., De Robertis, E. M., Philippsen, P. 1983. Delta sequences in the 5' non-coding region of yeast tRNA genes. *EMBO J.* 2:583-91
45. Garfinkel, D. J., Boeke, J. D., Fink, G.

- R. 1985. Ty element transposition: reverse transcriptase and virus-like particles. *Cell* 42:507-17
46. Garfinkel, D. J., Mastrangelo, M. F., Sanders, N. J., Shafer, B. K., Strathern, J. N. 1988. Transposon tagging using Ty elements in yeast. *Genetics* 120:95-108
47. Geiduschek, E. P., Tocchini-Valentini, G. P. 1988. Transcription by RNA polymerase III. *Annu. Rev. Biochem.* 57:873-914
48. Giroux, C. N., Mis, J. R. A., Pierce, M. K., Kohalmi, S. E., Kunz, B. A. 1988. DNA sequence analysis of spontaneous mutations in the *SUP4-o* gene of *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* 8:978-81
49. Goodman, H. M., Olson, M. V., Hall, B. D. 1977. Nucleotide sequence of a mutant eukaryotic gene: the yeast tyrosine-inserting ochre suppressor *SUP4-o*. *Proc. Natl. Acad. Sci. USA* 74:5453-57
50. Goodman, S. D., Nash, H. A. 1989. Functional replacement of a protein-induced bend in a DNA recombination site. *Nature* 341:251-54
51. Grandbastien, M.-A., Spielmann, A., Caboche, M. 1989. Tn1, a mobile retroviral-like transposable element of tobacco isolated by plant cell genetics. *Nature* 337:376-80
52. Grandgenett, D. P., Mumm, S. R. 1990. Unraveling retrovirus integration. *Cell* 60:3-4
53. Grandgenett, D. P., Vora, A. C. 1985. Site specific nicking at the avian retrovirus LTR circle junction by the viral pp32 endonuclease. *Nucleic Acids Res.* 13:6205-21
54. Grandgenett, D. P., Vora, A. C., Schiff, R. D. 1978. A 32,000-Dalton nucleic acid-binding protein from avian retrovirus cores possesses DNA endonuclease activity. *Virology* 89:119-32
55. Grandgenett, D. P., Vora, A. C., Swanson, R., Olsen, J. 1986. Nuclease mechanism of the avian retrovirus pp32 endonuclease. *J. Virol.* 58:970-74
56. Gross, D. S., Garrard, W. T. 1988. Nuclease hypersensitive sites in chromatin. *Annu. Rev. Biochem.* 57:159-97
57. Guthrie, C., Abelson, J. 1982. Organization and expression of tRNA genes in *Saccharomyces cerevisiae*. In *The Molecular Biology of the Yeast Saccharomyces*, ed. J. N. Strathern, E. W. Jones, J. R., Broach, pp. 487-28. Cold Spring Harbor, NY: Cold Spring Harbor Lab.
58. Hansen, L. J., Chalker, D. L., Sandmeyer, S. B. 1988. Ty3, a yeast retrotransposon associated with tRNA genes, has homology to animal retroviruses. *Mol. Cell. Biol.* 8:5245-56
59. Hansen, L. J., Sandmeyer, S. B. 1990. Characterization of a transcriptionally active Ty3 element and identification of the Ty3 IN protein. *J. Virol.* 64:2599-607
60. Hauber, J., Nelbock-Hochstetter, P., Feldmann, H. 1985. Nucleotide sequence and characteristics of a Ty element from yeast. *Nucleic Acids Res.* 13:2745-58
61. Hauber, J., Stucka, R., Krieg, R., Feldmann, H. 1988. Analysis of yeast chromosomal regions carrying members of the glutamate tRNA gene family: various transposable elements are associated with them. *Nucleic Acids. Res.* 16:10623-34
62. Hinton, C. W. 1984. Morphogenetically specific mutability in *Drosophila ananassae*. *Genetics* 106:631-53
63. Hizi, A., Hughes, S. H. 1988. Expression of the Moloney murine leukemia virus and human immunodeficiency virus integration proteins in *Escherichia coli*. *Virology* 167:634-38
64. Hughes, S. H., Shank, P. R., Spector, D. H., Kung, H.-J., Bishop, J. M., et al. 1978. Proviruses of avian sarcoma virus are terminally redundant, co-extensive with unintegrated linear DNA and integrated at many sites. *Cell* 15:1397-410
65. Huibregtse, J. R., Engelke, D. R. 1989. Genomic footprinting of a yeast tRNA gene reveals stable complexes over the 5'-flanking region. *Mol. Cell. Biol.* 9:3244-52
66. Hwang, S. L.-H., Gilboa, E. 1984. Expression of genes introduced into cells by retroviral infection is more efficient than that of genes introduced into cells by DNA transfection. *J. Virol.* 50:417-24
67. Ikenaga, H., Saigo, K. 1982. Insertion of a movable genetic element, 297, into the T-A-T-A box for the H3 histone gene in *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. USA* 79:4143-47
68. Inouye, S., Yuki, S., Saigo, K. 1984. Sequence-specific insertion of the *Drosophila* transposable genetic element 17.6. *Nature* 310:332-33
69. Inouye, S., Yuki, S., Saigo, K. 1986. Complete nucleotide sequence and genome organization of a *Drosophila* transposable genetic element, 297. *Eur. J. Biochem.* 154:417-25
70. Jentsch, S., McGrath, J. P., Var-

- shavsky, A. 1987. The yeast DNA repair gene *RAD6* encodes a ubiquitin-conjugating enzyme. *Nature* 329:131-34
71. Johnson, M. S., McClure, M. A., Feng, D.-F., Gray, J., Doolittle, R. F. 1986. Computer analysis of retroviral *pol* genes: assignment of enzymatic functions to specific sequences and homologies with non-viral enzymes. *Proc. Natl. Acad. Sci. USA* 83:7648-52
72. Johnson, S. P., Warner, J. R. 1989. Unusual enhancer function in yeast rRNA transcription. *Mol. Cell. Biol.* 9:4986-93
73. Kassavetis, G. A., Braun, B. R., Nguyen, L. H., Geiduschek, E. P. 1990. *S. cerevisiae* TFIIIB is the transcription initiation factor proper of RNA polymerase III, while TFIIIA and TFIIIC are assembly factors. *Cell* 60:235-45
74. Kassavetis, G. A., Riggs, D. L., Negri, R., Nguyen, L. H., Geiduschek, E. P. 1989. Transcription factor IIIB generates extended DNA interactions in RNA polymerase III transcription complexes on tRNA genes. *Mol. Cell. Biol.* 9:2551-66
75. Katzman, M., Katz, R. A., Skalka, A. M., 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.* 63:5319-27
76. King, W., Patel, M. D., Lobel, L. I., Goff, S. P., Nguyen-Huu, M. C. 1985. Insertion mutagenesis of embryonal carcinoma cells by retroviruses. *Science* 228:554-58
77. Kingsman, A. J., Gimlich, R. L., Clarke, L., Chinault, A. C., Carbon, J. 1981. Sequence variation in dispersed repetitive sequences in *Saccharomyces cerevisiae*. *J. Mol. Biol.* 145:619-32
78. Klar, A. J. S., Strathern, J. N., Abraham, J. A. 1984. Involvement of double-strand chromosomal breaks for mating-type switching in *Saccharomyces cerevisiae*. *Cold Spring Harbor Symp. Quant. Biol.* 49:77-88
79. Knaus, R. J., Hippenmeyer, P. J., Misra, T. K., Grandgenett, D. P., Müller, U. R., Fitch, W. M. 1984. Avian retrovirus pp32 DNA binding protein. Preferential binding to the promoter region of long terminal repeat DNA. *Biochemistry* 23:350-59
80. Kurjan, J., Hall, B. D., Gillam, S., Smith, M. 1980. Mutations at the yeast *SUP4* locus: DNA sequence changes in mutants lacking suppressor activity. *Cell* 20:701-6
81. Louis, C., Schedl, P., Samal, B., Worcel, A. 1980. Chromatin structure of the 5S RNA genes of *D. melanogaster*. *Cell* 22:387-92
82. Madison, J. T., Kung, H. K. 1967. Large oligonucleotides isolated from yeast tyrosine transfer ribonucleic acid after partial digestion with ribonuclease. *T1. J. Biol. Chem.* 242:1324-30
83. Marlor, R. L., Parkhurst, S. M., Corces, V. G. 1986. The *Drosophila melanogaster* gypsy transposable element encodes putative gene products homologous to retroviral proteins. *Mol. Cell. Biol.* 6:1129-34
84. Marschalek, R., Brechner, T., Amon-Bohm, E., Dinger, T. 1989. Transfer RNA genes: landmarks for integration of mobile genetic elements in *Dictyostelium discoideum*. *Science* 244:1493-96
85. Misra, T. K., Grandgenett, D. P., Parson, J. T. 1982. Avian retrovirus pp32 DNA-binding protein I. Recognition of specific sequences on retrovirus DNA terminal repeats. *J. Virol.* 44:330-43
- 85a. Moenne, A., Camier, S., Anderson, G., Margottin, F., Beggs, J., Sentenac, A. 1990. The U6 gene of *Saccharomyces cerevisiae* is transcribed by RNA polymerase C (III) in vivo and in vitro. *EMBO J.* 9:271-77
86. Mount, S. M., Rubin, G. M. 1985. Complete nucleotide sequence of the *Drosophila* transposable element copia: homology between copia and retroviral proteins. *Mol. Cell. Biol.* 5:1630-38
87. Müller, F., Brühl, K. H., Friedel, K., Kowallik, K. V., Ciriacy, M. 1987. Processing of TY1 proteins and formation of Ty1 virus-like particles in *Saccharomyces cerevisiae*. *Mol. Gen. Genet.* 207:421-29
88. Murphy, S., Di Liegro, C., Melli, M. 1987. The in vitro transcription of the 7SK RNA gene by RNA polymerase III is dependent only on the presence of an upstream promoter. *Cell* 51:81-87
89. Natsoulis, G., Thomas, W., Roghmann, M.-C., Winston, F., Boeke, J. D. 1989. Ty1 transposition in *Saccharomyces cerevisiae* is nonrandom. *Genetics* 123:269-79
90. Nelböck, P., Stucka, R., Feldmann, H. 1985. Different patterns of transposable elements in the vicinity of tRNA genes in yeast: a possible clue to transcriptional modulation. *Biol. Chem. Hoppe-Seyler* 366:1041-51
91. Olson, M. V., Loughney, K., Hall, B. N. 1979. Identification of the yeast DNA sequences that correspond to specific

- tyrosine-inserting nonsense suppressor loci. *J. Mol. Biol.* 132:387-410
92. Oyen, T. B., Gabrielson, O. S. 1983. Non-random distribution of the TY1 elements within nuclear DNA of *Saccharomyces cerevisiae*. *FEBS Lett.* 161:201-6
 93. Panganiban, A. T., Temin, H. M. 1984. The retrovirus *pol* gene encodes a product required for DNA integration: Identification of a retrovirus *int* locus. *Proc. Natl. Acad. Sci. USA* 81:7885-89
 94. Panganiban, A. T., Temin, H. M. 1984. Circles with two tandem LTRs are precursors to integrated retrovirus DNA. *Cell* 36:673-79
 95. Pérez-Ortín, J. E., Matallana, E., Franco, L. 1989. Chromatin structure of yeast genes. *Yeast* 5:219-38
 96. Petes, T. D. 1979. Yeast ribosomal DNA genes are located on chromosome XII. *Proc. Natl. Acad. Sci. USA* 76:410-14
 97. Piper, P. W., Lockheart, A., Patel, N. 1984. A minor class of 5S rRNA genes in *Saccharomyces cerevisiae* X2180-1B, one member of which lies adjacent to a Ty transposable element. *Nucleic Acids Res.* 12:4083-96
 98. Poole, S. J., Firtel, R. A. 1984. Genomic instability and mobile genetic elements in regions surrounding two discoidin I genes of *Dictyostelium discoideum*. *Mol. Cell. Biol.* 4:671-80
 99. Quinn, T. P., Grandgenett, D. P. 1988. Genetic evidence that the avian retrovirus DNA endonuclease domain of *pol* is necessary for viral integration. *J. Virol.* 62:2307-12
 100. Robinson, H. L., Gagnon, G. C. 1986. Patterns of proviral insertion and deletion in avian leukosis virus-induced lymphomas. *J. Virol.* 57:28-36
 101. Rohdewohld, H., Weiher, H., Reik, W., Jaenisch, R., Breindl, M. 1987. Retrovirus integration and chromatin structure: Moloney murine leukemia proviral integration sites map near DNase I-hypersensitive sites. *J. Virol.* 61:336-43
 102. Rosen, E., Sivertsen, A., Firtel, R. A. 1983. An unusual transposon encoding heat shock inducible and developmentally regulated transcripts in *Dictyostelium*. *Cell* 35:243-51
 103. Roth, M. J., Schwartzberg, P. L., 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
 104. Roth, M. J., Tanese, N., Goff, S. P. 1988. Gene product of Moloney murine leukemia virus required for proviral integration is a DNA-binding protein. *J. Mol. Biol.* 203:131-39
 105. Saigo, K., Kugimiya, W., Matsuo, Y., Inouye, S., Yoshioka, K., Yuki, S. 1984. Identification of the coding sequence for a reverse transcriptase-like enzyme in a transposable genetic element in *Drosophila melanogaster*. *Nature* 312:659-61
 106. Sandmeyer, S. B., Bilanchone, V. W., Clark, D. J., Morcos, P., Carle, G. F., Brodeur, G. M. 1988. Sigma elements are position-specific for many different yeast tRNA genes. *Nucleic Acids Res.* 16:1499-515
 107. Sandmeyer, S. B., Olson, M. V. 1982. Insertion of a repetitive element at the same position in the 5'-flanking regions of two dissimilar yeast tRNA genes. *Proc. Natl. Acad. Sci. USA* 79:7674-78
 108. Scherдин, U., Rhodes, K., Breindl, M. 1990. Transcriptionally active genome regions are preferred targets for retrovirus integration. *J. Virol.* 64:907-12
 109. Schultz, P., Marzouki, N., Marck, C., Ruet, A., Oudet, P., Sentenac, A. 1989. The two DNA-binding domains of yeast transcription factor tau as observed by scanning transmission electron microscopy. *EMBO J.* 8:3815-24
 110. Schwartz, C. J., Sadowski, P. D. 1989. FLP recombinase of the 2 micron circle plasmid of *Saccharomyces cerevisiae* bends its DNA target. Isolation of FLP mutants defective in DNA bending. *J. Mol. Biol.* 205:647-58
 111. Schwartzberg, P., Colicelli, J., Goff, S. P. 1984. Construction and analysis of deletion mutations in the *pol* gene of murine leukemia virus: a new viral function of required for productive infection. *Cell* 37:1043-52
 112. Shalit, P., Loughney, K., Olson, M. V., Hall, B. D. 1981. Physical analysis of the *CYC1-SUP4* interval in *Saccharomyces cerevisiae*. *Mol. Cell Biol.* 1:228-36
 113. Shih, C.-C., Stoye, J. P., Coffin, J. M. 1988. Highly preferred targets for retrovirus integration. *Cell* 53:531-37
 114. Shimotohno, K., Temin, H. 1980. No apparent nucleotide sequence specificity in cellular DNA juxtaposed to retrovirus proviruses. *Proc. Natl. Acad. Sci. USA* 77:7357-61
 115. Shoemaker, C., Goff, S., Gilboa, E., Paskind, M., Mitra, S. W., Baltimore, D. 1980. Structure of a cloned circular Moloney murine leukemia virus DNA molecule containing an inverted seg-

- ment: Implications for retrovirus integration. *Proc. Natl. Acad. Sci. USA* 77:3932-36
116. Simchen, G., Winston, F., Styles, C. A., Fink, G. R. 1984. Ty-mediated gene expression of the *LYS2* and *HIS4* genes of *Saccharomyces cerevisiae* is controlled by the same *SPT* genes. *Proc. Natl. Acad. Sci. USA* 81:2431-34
117. Smyth, D. R., Kalitsis, P., Joseph, J. L., Sentry, J. W. 1989. Plant retrotransposon from *Lilium henryi* is related to Ty3 of yeast and the gypsy group of *Drosophila*. *Proc. Natl. Acad. Sci. USA* 86:5015-19
118. Snyder, M. P., Kimbrell, D., Hunkapiller, M., Hill, R., Fristrom, J., Davidson, N. 1982. A transposable element that splits the promoter region inactivates a *Drosophila* cuticle gene. *Proc. Natl. Acad. Sci. USA* 79:7430-34
119. Snyder, U. K., Thompson, J. F., Landy, A. 1989. Phasing of protein-induced DNA bends in a recombination complex. *Nature* 341:255-57
120. Stewart, S. E., Roeder, G. S. 1989. Transcription by RNA polymerase I stimulates mitotic recombination in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* 9:3464-72
121. Stillman, D. J., Better, M., Geiduschek, E. P. 1985. Electron-microscopic examination of the binding of a large RNA polymerase III transcription factor to a tRNA gene. *J. Mol. Biol.* 185:451-55
122. Stucka, R., Hauber, J., Feldmann, H. 1987. One member of the tRNA(Glu) gene family in yeast codes for a minor GAGtRNA(Glu) species and is associated with several short transposable elements. *Curr. Genet.* 12:323-28
123. Stucka, R., Lochmuller, H., Feldmann, H. 1989. Ty4, a novel low-copy number element in *Saccharomyces cerevisiae*: one copy is located in a cluster of Ty elements and tRNA genes. *Nucleic Acids Res.* 17:4993-5001
124. Sung, P., Prakash, S., Prakash, L. 1988. *RAD6* protein of *Saccharomyces cerevisiae* polyubiquitinates histones and its acidic domain mediates this activity. *Genes Dev.* 2:1476-85
125. Tanda, S., Shrimpton, A. E., Ling-Ling, C., Itayama, H., Matsubayashi, H., Saigo, K., Tobari, Y. N., Langley, C. H. 1988. Retrovirus-like features and site specific insertions of a transposable element, *tom* in *Drosophila ananassae*. *Mol. Gen. Genet.* 214:405-11
126. Taylor, M. J., Segall, J. 1985. Characterization of factors and DNA sequences required for accurate transcription of the *Saccharomyces cerevisiae* 5S RNA gene. *J. Biol. Chem.* 260:4531-40
127. Terry, R., Soltis, D. A., Katzman, M., Cobrinik, D., Leis, J., Skalka, A. M. 1988. Properties of avian sarcoma-leukosis virus pp32-related polendonucleases produced in *Escherichia coli*. *J. Virol.* 62:2358-65
128. Thomas, B. J., Rothstein, R. 1989. Elevated recombination rates in transcriptionally active DNA. *Cell* 56:619-30
129. Toh, H., Kikuno, R., Hayashida, H., Miyata, T., Kugimiya, W., et al. 1985. Close structural resemblance between putative polymerase of a *Drosophila* transposable genetic element 17.6 and *pol* gene product of Moloney murine leukemia virus. *EMBO J.* 4:1267-72
130. Umlauf, S. W., Cox, M. M. 1988. The functional significance of DNA sequence structure in a site-specific genetic recombination reaction. *EMBO J.* 7:1845-52
131. Varmus, H., Brown, P. 1989. Retroviruses. See Ref. 6, pp. 53-108
132. Vijaya, S., Steffen, D. L., Robinson, H. L. 1986. Acceptor sites for retroviral integrations map near DNase I-hypersensitive sites in chromatin. *J. Virol.* 60:683-92
133. Vincent, A., Petes, T. D. 1986. Isolation and characterization of a Ty element inserted into the ribosomal DNA of the yeast *Saccharomyces cerevisiae*. *Nucleic Acids Res.* 14:2939-49
134. Voytas, D. F., Ausubel, F. M. 1988. A copia-like transposable element family in *Arabidopsis thaliana*. *Nature* 336:242-44
135. Warmington, J. R., Anwar, R., Newlon, C. S., Waring, R. B., Davies, R. W., et al. 1986. A 'hot-spot' for Ty transposition on the left arm of yeast chromosome III. *Nucleic Acids Res.* 14:3475-85
136. Warmington, J. R., Green, R. P., Newlon, C. S., Oliver, S. G. 1987. Polymorphisms on the right arm of yeast chromosome III associated with Ty transposition and recombination events. *Nucleic Acids Res.* 15:8963-82
137. Warmington, J. R., Waring, R. B., Newlon, C. S., Indge, K. J., Oliver, S. G. 1985. Nucleotide sequence characterization of Ty 1-17, a class II transposon from yeast. *Nucleic Acids Res.* 13:6679-93
138. Wilke, C. M., Heidler, S. H., Brown, N., Liebman, S. W. 1989. Analysis of

- yeast retrotransposon Ty insertions at the *CAN1* locus. *Genetics* 123:655-65
139. Williamson, V. M., Cox, D., Young, E. T., Russell, D. W., Smith, M. 1983. Characterization of transposable element-associated mutations that alter yeast alcohol dehydrogenase II expression. *Mol. Cell. Biol.* 3:20-31
140. Xiong, Y., Eickbush, T. H. 1988. Similarity of reverse transcriptase-like sequences of viruses, transposable elements, and mitochondrial introns. *Mol. Biol. Evol.* 5:675-90
- 140a. Xiong, Y., Eickbush, T. H. 1988. Functional expression of a sequence-specific endonuclease encoded by the retrotransposon R2Bm. *Cell* 55:235-46
141. Zuker, C., Lodish, H. F. 1981. Repetitive DNA sequences cotranscribed with developmentally regulated *Dictyostelium discoideum* mRNAs. *Proc. Natl. Acad. Sci. USA* 78:5386-90