Dodging the genes

A recent study shows that yeast *Ty1* retrotransposons integrate selectively into certain 'hotspot' regions upstream of tRNA genes.

How do retroviruses and retrotransposons select host DNA sites to use as integration targets? This is an important issue, as a poor target choice can have disastrous consequences. If the insertion of retroelement DNA disrupts a required cellular gene, then integration can be lethal. In vertebrates, insertions of retroelements have been found to cause mutations that, when present with a second recessive allele, lead to diseases such as hemophilia. Dominant mutations as a consequence of retrovirus insertion are also well documented in vertebrates, leading for example to proto-oncogene activation and cancer. Organisms that spend part of their lives in the haploid state, such as yeast, are even more vulnerable to mutations caused by retroelement insertion. A recent pioneering study of target-site selection by the yeast retrovirus-like transposon *Ty1* [1] has revealed that this retroelement selects its target very carefully: integration sites cluster outside of the coding regions of cellular genes, in DNA that can apparently tolerate insertions readily.

Most earlier studies of target-site selection *in vivo* by retroviruses and retrotransposons have been limited by the fact that genetic selections were used to recover integration events, thereby biasing the population obtained. Early studies examined insertions of retroviruses that activated oncogenes, and insertions of *Ty* elements into genes in yeast for which convenient selections were available. These studies did not provide a clear picture of site selection in a natural setting, in which experimental selective pressures are absent.

Studies of unselected retrovirus insertions have not yet provided a simple picture of how target sites are chosen. An early study of unselected insertions of Moloney murine leukemia virus suggested that integration sites were not chosen randomly. Although site selection showed no obvious specificity with respect to target DNA sequence, integrations were found to cluster near DNase I hypersensitive sites [2]. This finding supported the idea that potential target sites may be blocked *in vivo* by packaging into chromatin. This study is not easy to reconcile with more recent *in vitro* experiments, which indicate that certain target sites are actually used more efficiently when wrapped in nucleosomes [3]. A study of unselected integrations of Rous sarcoma virus revealed a class of integration sites that are used much more frequently than expected if site selection is random [4]. The mechanism of this effect is unclear. The interpretation of all these studies is limited by the fact that new integrations generally fall in previously uncharacterized DNA, and so the relation of insertion sites to chromosomal features could not easily be determined.

Ji and coworkers [1] recently advance the study of target-site selection by developing a method of studying integrations of *Ty1* into yeast chromosome III; as chromosome III is now fully sequenced [5], this represents the first opportunity to map unselected integration sites relative to chromosomal sequence features. *Ty1*, like retroviruses, replicates by: first, transcription of retroelement DNA; second, packaging of the element RNA with element proteins in a virus-like particle; third, reverse transcription to yield a double-stranded DNA copy of the element RNA; and fourth, integration of the DNA copy into a site in the host chromosome. Many of the element-encoded proteins that carry out these steps are similar in sequence between retroviruses and *Ty1*, including the protease, reverse transcriptase, RNase H and integrase. Unlike retroviruses, *Ty1* particles never acquire an envelope and leave the parental cells. *Ty* elements instead carry out all steps of the life cycle inside a single cell, resulting in transposition rather than infection.

High levels of *Ty1* transposition were induced by placing *Ty1* transcription under the control of the strong Gal1 promoter, the Gal1–*Ty1* construct being carried on a plasmid in transformed yeast. When transcription of the Gal1–*Ty1* fusion gene was activated by growing the yeast in the presence of galactose, transposition was induced to such a high rate that on average nine new insertions appeared per genome. A selectable marker, *TRP1*, was inserted into the *Ty1* transposon used, so that Trp1$^-$ cells containing the modified *Ty1* element could grow in the absence of exogenous tryptophan. Because the plasmid vector carrying the Gal1–*Ty1* construct could be eliminated by selecting against a *URA3* gene also carried on the plasmid, cells containing new transpositions into the yeast genome could be obtained by selecting for a Trp1$^+$ and a Ura3$^-$ phenotype. To avoid selective pressures that might bias the population of integration sites recovered, Ji *et al.* [1] maintained chromosome III in the diploid state throughout the course of the analysis. To facilitate further analysis, however, it was necessary to separate integrations into chromosome III from insertions into other chromosomes.

Ji and collegues [1] cleverly exploited the unusual features of the *kar1-1* mutation to transfer copies of chromosome III containing new insertions into a 'clean' genetic background. In matings between a wild-type cell and a cell containing the *kar1-1* mutation, the cytoplasmic membranes of the two cells fuse but the two parental nuclei fail to fuse. In the next round of cell division, two daughter cells are produced with mixed cytoplasms but unmixed parental nuclei. Single chromosomes can, however, transfer at a low frequency from one nucleus to the other, a process known as 'chromoduction'. Chromosome III is chromoduced at a frequency of about one in 10^3 matings.

Ji *et al.* mated the cells containing new *Ty1* insertions to a *kar1-1* strain, selected for chromosome III chromoductants and then screened among them for chromoductants containing new *Ty1* insertions. Thus, a set of strains containing many unselected integrations into chromosome III was constructed.

The locations of 116 independent *Ty1* integration sites in chromosome III were determined by separating genomic DNA on contour clamped heterogeneous electric field (CHEF) gels, and analysis by ethydium bromide staining and Southern blotting. As the *Ty1* element used in the experiment contained a single *Not*I site, integration could be readily mapped after *Not*I cleavage of the genomic DNA. Unexpectedly, 99% of the unselected insertions were found to occur in four major 'hotspots'. The hottest hotspot accounted for 48% of the insertions studied. Even more surprising was the location of the insertions with respect to the cellular genes. Thirty insertion sites were amplified by the polymerase chain reaction (PCR), sequenced and then located on the chromosome III sequence. Not one of the sequenced insertion sites lay in a cellular protein coding region, despite the fact that chromosome III is in fact 70% coding sequence. The only open reading frame interrupted by integration was part of a pre-existing *Ty2* element.

Intriguingly, more than half the sequenced insertions lay within 400 base pairs (bp) upstream of a tRNA gene (Fig. 1). At the hottest hotspot, all the integration sites lay upstream of a $tRNA^{Gly}$ or $tRNA^{Asn}$ gene, with one curious exception. A single integration mapped to this region by CHEF gel analysis lay in a $tRNA^{Asp}$ gene that was not present in the chromosome III sequence. This region had previously been reported to be particularly polymorphic among yeast strains, possibly because of insertion and recombination events involving *Ty* elements. Evidently the $tRNA^{Asp}$ gene was not present in the strain used for sequencing, but was present in the variant chromosome III used by Ji *et al.*. Not all the reported tRNA genes on chromosome III were transposition hotspots, suggesting that the presence of a tRNA gene alone may not be sufficient to make a genomic site a hotspot. An important caveat to this conclusion is that the chromosome III used as a target has not yet been checked for the presence of the 'cold' tRNA genes.

The hotspot regions also frequently contained single *Ty* long terminal repeat (LTR) sequences. It seems likely that single LTRs are the product of homologous recombination between two LTRs that once flanked an integrated *Ty* element. Thus, these sequences probably represent molecular 'fossils' of earlier integration events. The presence of pre-existing *Ty* sequences near the hotspots might simply reflect their prior insertion into the hotspot regions, though the possibility that such pre-existing *Ty* sequences are themselves important to generate a hotspot has not been ruled out.

Might the *Ty1* integration apparatus simply recognize specific sequences in the target DNA, thereby directing integration to certain sites? That this is not the case is suggested by an analysis of target sequences from all available *Ty1* insertion sites, which indicates that target sequences

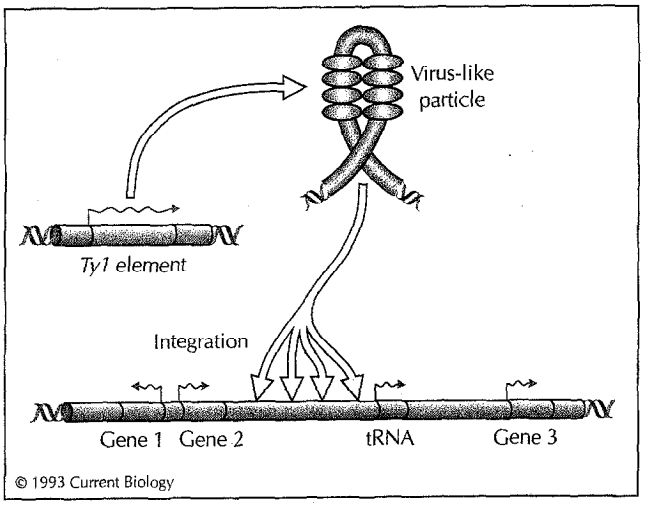

Fig. 1. Integration-target selection by the yeast retrotransposon *Ty1*. Transcription from a pre-existing *Ty1* element, followed by complex formation with viral proteins and reverse transcription, yields a virus-like particle. Integration occurs preferentially in non protein-coding sequences upstream of tRNA genes. Integration events are rare in the coding regions of cellular genes, despite the fact that 70% of the target chromosome is coding region.

bear only very weak sequence similarity to one another. In fact, the sequence similarity among the sites used was more evident as an 'anti-consensus' sequence, a sequence least likely to be used as an integration target, than as a consensus sequence. Furthermore, the hotspots studied *in vivo* are not hotspots *in vitro*, even though the recombinationally active virus-like particles used *in vitro* behave like *Ty1 in vivo* by a variety of criteria. Thus, sequence-specific recognition by the integration apparatus does not seem to explain the observed site selection.

Another yeast retrotransposon, *Ty3*, also integrates near tRNA genes, but in this case the integration specificity is much more extreme. *Ty3* inserts only within one to four nucleotides of the start of transcription of tRNA genes and other RNA polymerase III (Pol III) transcribed genes. Mutations in the Pol III transcription signals within the gene inactivate it as a transposition target, but mutations at the point of insertion have little effect. From these observations it seems likely that some component of the *Ty3* integration machinery interacts specifically with a component of the Pol III transcription complex, thereby directing *Ty3* integration to its site [6].

What advantage does this target specificity have for *Ty* elements? In the case of *Ty3*, insertions around the transcription start site do not disrupt expression of the tRNA gene target. *Ty1*, which integrates further upstream, is also likely to have little effect on tRNA transcription. Possibly *Ty1* and *Ty3* have evolved to integrate preferentially into such sites in order to minimize the impact on the host and so promote their own survival.

The target sites used frequently by *Ty1* are spread over several hundred bp upstream of the target tRNA gene, so a *Ty3*-like mechanism for site selection based on interactions with the Pol III transcriptional machinery cannot explain the results as easily. Given that the hotspots are not hot for integration as naked DNA, some aspect of the *in vivo* environment is evidently essential for sites to be

hotspots. One simple model might be that cellular factors bind the hotspot regions *in vivo* and stimulate integration. Alternatively, perhaps packing of DNA into chromatin suppresses integration except upstream of tRNA genes, though, by itself, this explanation seems somewhat unsatisfactory: many sequence-specific DNA-binding proteins can obviously access sites at many places in the yeast genome besides the regions upstream of tRNA genes.

If assembly into chromatin does contribute to suppressing integration outside the hotspots, the effect may be due to binding of the basic amino-terminal histone arms in the major groove of the target DNA, as integration in the retroviral system apparently takes place on the major groove side of the target DNA [3]. If so, histone acetylation, which apparently disrupts the binding of the arms [7], may promote integration. The higher order organization of chromosomes within the nucleus may also affect target-site selection.

Given the tools available in the yeast system, it should be possible to test many of the models for hotspot generation in a straightforward manner. Simple studies of mutant hotspots *in vivo* should allow the determinants of hotspots to be identified. The *in vitro* integration system will allow candidate functions to be tested in reconstructed reactions. Genetic approaches offer the potential to identify *trans*-acting factors, and to test the importance, *in vivo*, of candidate factors identified in biochemical experiments.

What about retroviruses? At present there is no evidence that retroviruses avoid integrating into coding regions or prefer to integrate upstream of tRNA genes, but as almost all unselected insertions studied took place in 'anonymous' DNA, as yet, insufficient data exist to rule out these possibilities. In our recent study in which unselected *in vivo* retrovirus integrations were characterized, SV40 was used as a target [8]. In this case, no bias away from integration into coding sequences was reported; whether this lack of bias will also hold with cellular chromosomal targets remains to be seen. Studies of integration site selection in chromosomal DNA are possible with retroviruses if a defined target DNA is chosen for study in advance. The locations of insertion sites in this region can then be determined using PCR with one primer in the target DNA and the other in the retroviral DNA. A thorough examination of diverse targets with this method should reveal whether retroviruses display *Ty*-like biases in target-site selection.

References

1. JI H, MOORE DP, BLOMBERG MA, BRAITERMAN T, VOYTAS DF, NATSOULIS G, BOEKE JD: **Hotspots for unselected Ty1 transposition events on yeast chromosome III are near tRNA genes and LTR sequences.** *Cell* 1993, 73:1–20.
2. VIJAYA S, STEFFAN DL, ROBINSON HL: **Acceptor sites for retroviral integrations map near DNaseI-hypersensitive sites in chromatin.** *J Viro* 1986, 60:683–692.
3. PRYCIAK PM, VARMUS HE: **Nucleosomes, DNA-binding proteins, and DNA sequences modulate retroviral integration target selection.** *Cell* 1992, 69:769–780.
4. SHIH C-C, STOYE JP, COFFIN JM: **Highly-preferred targets for retrovirus integration.** *Cell* 1988, 53:531–537.
5. OLIVER SG, VAN DER AART OJM, AGOSTONI-CARBONE ML, AIGLE M, ALBERGHINA M, ALEXANDARAKI D, ANTOINE G, ANWAR R, BALLESTA JPG *ET AL.*: **The complete DNA sequence of yeast chromosome III.** *Nature* 1992, 357:38–46.
6. CHALKER DL, SANDMEYE SB: **Ty3 integrates within the region of RNA polymerase III transcription initiation.** *Genes Dev* 1992, 6:117–128.
7. LEE DY, HAYES JJ, PRUSS D, WOLFFE AP: **A positive role for histone acetylation in transcription factor access to nucleosomal DNA.** *Cell* 1993, 72:73–84.
8. PRYCIAK PM, MULLER H-P, VARMUS HE: **Simian virus 40 minichromosomes as targets for retroviral integration *in vivo*.** *Proc Natl Acad Sci USA* 1992, 89:9237–9241.

Frederic Bushman, Infectious Disease Laboratory, Salk Institute, La Jolla, California 92014, USA.