J. Cell Sci. Suppl. 7, 155–167 (1987)
Printed in Great Britain © *The Company of Biologists Limited 1987*

THE GENETIC ORGANIZATION OF THE YEAST Ty ELEMENT

ALAN J. KINGSMAN, JANE MELLOR, SALLY ADAMS, PETER D. RATHJEN, MICHAEL H. MALIM, SANDRA M. FULTON, WILMA WILSON AND SUSAN M. KINGSMAN

Department of Biochemistry, South Parks Road, Oxford, England

SUMMARY

The genetic organization of the yeast transposon Ty resembles that of higher eukaryotic retroviruses and other elements such as the *copia*-like sequences of *Drosophila*. The Ty genome is 5·9 kb (10^3 bases) long. It has 340 bp (base pairs) terminal repeats known as delta sequences and it produces a terminally redundant 5·7 kb RNA that starts in the 5′ delta and ends in the 3′ delta. Ty transcription is directed by signals upstream and downstream of the major RNA start site and is regulated by the mating-type configuration of the cell. The 5·7 kb transcriptional unit is divided into two overlapping open reading frames, *TYA* and *TYB*. *TYA* occupies approximately the first quarter of the transcriptional unit while *TYB* occupies the rest. *TYB* overlaps *TYA* by either 38 or 44 nucleotides, depending on the element, and is in the plus one reading frame with respect to *TYA*. *TYA* is expressed to produce protein p1 ($50\times10^3 M_r$) and *TYB* is expressed as a *TYA*:*TYB* fusion protein, p3 ($190\times10^3 M_r$). Both of these proteins are subsequently cleaved to produce proteins p2, p4, p5, p6, reverse transcriptase and a protease that is responsible for some of these cleavage events. These proteins are assembled into virus-like particles (Ty-VLPs) that contain Ty RNA and reverse transcriptase activity. It is likely that the Ty-VLPs are units of transposition as Ty transposes *via* an RNA intermediate.

INTRODUCTION

Ever since the general structure of the Ty elements of yeast and the similar *copia*-like sequences of *Drosophila* were determined and partial sequence data were available, molecular biologists have commented on the similarities of these elements to retroviral proviruses. The logical consequences of these comparisons has been a long standing expectation that Ty and *copia*-like sequences would transpose *via* an RNA intermediate. This would act as a template for a reverse transcription reaction prior to integration of a newly synthesized, double-stranded DNA molecule at a new site. Recent data, largely obtained with the yeast Ty element, have shown that these early hypotheses were correct and that Ty not only has remarkable structural but also functional similarities to other retroelements such as retroviruses. However, as the analyses and comparisons are extended it is becoming clear that, although generally true, these conclusions require qualification.

In this review we will describe the organization and activities of the yeast Ty element and then relate these to other retroelements.

GENERAL STRUCTURE

Almost all of the 30–35 copies of Ty in laboratory strains of yeast are about 5·9 kb in length (Roeder & Fink, 1983; Williamson, 1983). They are composed of a 5·2 kb

unique region, called epsilon, flanked by direct repeats of about 340 bp, called delta sequences. The total genomic complement of Ty elements constitutes a somewhat heterogeneous family of sequences (Kingsman *et al.* 1981; Williamson, 1983) dispersed throughout the genome (Kingsman *et al.* 1981; Cameron *et al.* 1979; Klein & Petes, 1984). This family can be divided into two broad classes, I and II, that differ by two large substitutions (Fig. 1) (Kingsman *et al.* 1981; Williamson *et al.* 1983). Within each class there are minor variations recognized initially as differences in restriction site patterns. The major transcription product is a 'full-length', 5·7 kb species that starts in the left or 5′ delta and ends in the right or 3′ delta such that the RNA has terminal repeats of about 50 nucleotides (Elder *et al.* 1983) (Fig. 1). This RNA is therefore directly analogous to the full-length, terminally redundant genomic RNA of retroviruses (Varmus, 1983).

Complete nucleotide sequence information has recently become available for several elements (Hauber *et al.* 1985; Clare & Farabaugh, 1985; Warmington *et al.* 1985; Fulton *et al.* unpublished). This has shown that Ty is divided into two open reading frames, *TYA* and *TYB*, that occupy almost all of the 5·7 kb transcriptional unit (Fig. 2). In the first element to be sequenced, Ty902, *TYA* begins at residue 291, within the 5′ delta, and ends at residue 1610. *TYB* starts at residue 1572 with an ACA (threonine) codon, in the plus one reading phase with respect to *TYA*, and ends at residue 5556 close to the end of the epsilon region. *TYA* and *TYB* therefore overlap by 38 bp (Fig. 2). The ATG at the start of *TYA* is the first ATG in the major 5·7 kb RNA. The organization of the class II element, Ty1-17, is essentially the same

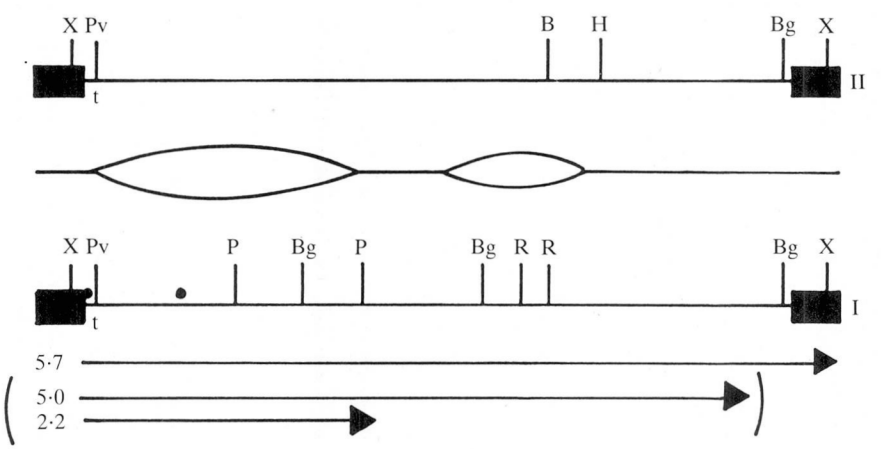

Fig. 1. A summary of the organization of Ty elements. Partial restriction maps of the class I and II elements, Ty1-15 and Ty1-17, are shown. Class II is above class I. Between them is a schematic diagram of a heteroduplex formed between the two elements (Kingsman *et al.* 1981). Size classes of Ty RNA are shown as long arrows. They are drawn against the class I element as they have been defined best for these elements. The 5·0 kb and 2·2 kb RNAs are in parentheses because their coordinates are unclear (see text). t, region homologous to the 3′ end of $tRNA^{Met}$ (Eibel *et al.* 1980); ●, sequences responsible for the ROAM phenomenon (see text); X, *Xho*I; Pv, *Pvu*II; B, *Bam*HI; H, *Hin*dIII; Bg, *Bgl*II; P, *Pst*I; R, *Eco*RI.

(Warmington *et al.* 1985), minor differences being that *TYA* and *TYB* overlap by 44 bp and *TYB* starts with a GCG (alanine) codon.

TRANSCRIPTION SIGNALS

Ty transcription

By analogy with retroviral proviruses (Varmus, 1983), it has been expected that at least some signals for transcription of the full-length, 5·7 kb RNA would be found within the 5′ LTR (delta). There is some early evidence for this (Bowen *et al.* 1984) and in all delta sequences there are conserved regions, including a TATA box, that have been regarded as good candidates for promoter components (Williamson *et al.* 1983; Elder *et al.* 1983; Bowen *et al.* 1984). There are also conserved transcription termination signals within delta sequences (Williamson *et al.* 1983; Elder *et al.* 1983; Bowen *et al.* 1984). As is the case for retroviruses the production of a terminally

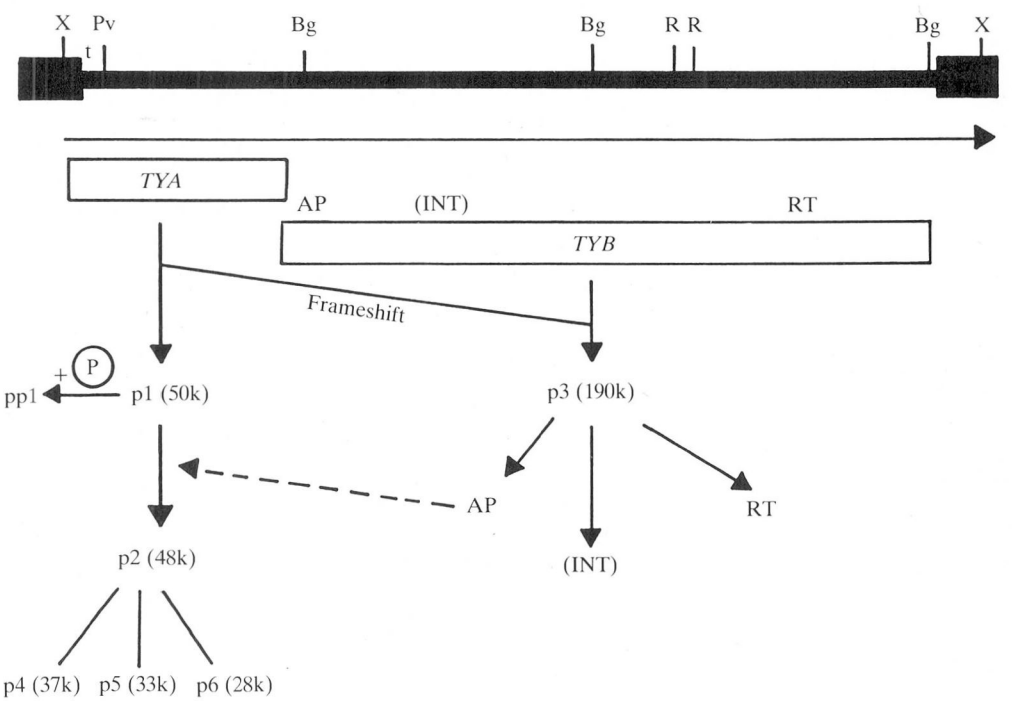

Fig. 2. Organization and expression of the 5·7 kb transcript. A class I element is shown with only the 5·7 kb RNA marked. The Ty genes, *TYA* and *TYB*, are represented by open boxes with regions of homology to retroviral proteins marked above the boxes such that: AP, region homologous to the active sites of acid proteases (Toh *et al.* 1985); INT, region homologous to retroviral integrases (Mount & Rubin, 1985); RT, region homologous to reverse transcriptases (Hauber *et al.* 1985; Clare & Farabaugh, 1985; Warmington *et al.* 1985). Below the *TYA* and *TYB* genes are the corresponding Ty proteins (see text) such that: AP, Ty processing protease; RT, reverse transcriptase; INT, integrase. INT is given in parentheses as its existence has not been proven. t, region that is homologous to the 3′ end of $tRNA^{Met}$ (36); ●, sequences responsible for the ROAM phenomenon; X, *Xho*I; Pv, *Pvu*II; Bg, *Bgl*II; P, *Pst*I; R, *Eco*RI. K = $10^3 M_r$.

redundant 'full-length' transcript requires that only the promoter components are active in the 5′ delta and only the termination signals are active in the 3′ delta. It is not clear how the termination signals are suppressed in the 5′ delta but it seems likely that the promoter activity of the 3′ delta is suppressed by a '5′ promoter-dominance' phenomenon similar to that described for ALV (Cullen *et al.* 1984).

There are two other Ty homologous RNAs that have been described, these are 5·0 kb and 2·2 kb (Mellor *et al.* 1985*c*; Elder *et al.* 1980)(Fig. 1). There is some confusion over the nature of the 5·0 kb transcript. An RNA species of this size has been known since 1980 and it has been suggested that it is 5′ coterminal with the 5·7 kb species (Roeder & Fink, 1983), although these data have never been independently confirmed. Recently, Winston *et al.* (1984) have shown that mutations in the *SPT3* gene that suppress the deleterious effects of Ty elements on the expression of flanking genes also result in the major Ty transcript being a 5·0 kb rather than 5·7 kb RNA. In fact the 5·7 kb species is almost completely absent in these mutants. This 5·0 kb RNA is not 5′ coterminal with the 5·7 kb species but starts about 800 bp into the element. It is not clear whether the two 5·0 kb species are in fact the same or different but with coincidental sizes. Nevertheless, the 5′-terminus of the 5·0 kb '*spt3* transcript' is sufficiently removed from the transcription start site for the 5·7 kb species to suggest that it may be the product of a post-transcriptional processing event, or that its synthesis may be directed by a second set of transcription signals downstream from those within the 5′ delta. The origin of the 2·2 kb RNA is unclear at present.

Recently we have shown that Ty transcription signals are extraordinarily complex and somewhat unlike their retroviral counterparts. The most significant finding is that the major determinant of transcription of the 5·7 kb RNA is downstream from the RNA start site. Deletion of the 5′ delta from its 5′ extremity to the RNA start site reduces the level of this transcript only marginally whereas deletion of sequences within the *TYA* gene dramatically reduces transcription (Fulton *et al.* manuscript submitted).

Regulation of Ty expression

It is remarkable that many retroelements have a close regulatory relationship with their 'host' genome. The ETn and IAP sequences of *Mus* are transcribed only during specific developmental windows (Brulet *et al.* 1983; Piko *et al.* 1984), VL30 expression is dependent on proliferative growth (Singh *et al.* 1985), MMTV transcription is regulated by glucocorticoid hormones (Strand & McDonald, 1985) and expression of *Drosophila copia* sequences is stimulated under a variety of stress conditions as well as being developmentally controlled (Flavell *et al.* 1980). It is conceivable, but by no means clear at present, that these regulatory circuits may serve to ensure a stable association between host and retroelement by restricting transposition or viral replication to certain cell states. Ty transcription is cell type specific being about 20-fold higher in *MAT***a** or *MAT*α haploids than in *MAT***a**/α diploids (Elder *et al.* 1983). Two sequences are required for this control to be exerted on the 5·7 kb RNA. The first is upstream of the start site, within the 5′ delta, and the

other is downstream just beyond the 3′ end of the 5′ delta. It is not yet known how the mating type configuration of the cell mediates control through these two sequences but deletion of either one abolishes control, resulting in high levels of Ty RNA in *MAT***a**/α diploids (Fulton *et al.* unpublished).

The ROAM phenomenon

Like other transposons Ty can have a profound affect on 'host' gene expression as it transposes around the genome. Clearly integration of 5·9 kb of DNA can disrupt a gene simply by acting as a sort of 'molecular brick'. However, more interesting than this insertional inactivation is the ability of Ty to activate 'host' gene expression and to bring the activated gene under the control of new regulatory circuits.

When a Ty element integrates into the promoter region of, for example, the *ADR2* gene, the gene is no longer under glucose/ethanol control but is expressed constitutively (with respect to glucose/ethanol) at relatively high levels (Williamson *et al.* 1983). However, like Ty transcription, expression of *ADR2* becomes cell-type specific. In haploid cells expression levels are high whereas in *MAT***a**/α diploid cells levels are reduced by 5–20 fold. The mutants that express this complex phenotype are known as ROAM (Regulated Overproducing Alleles responding to Mating-type) mutants (Errede *et al.* 1980). ROAM mutant derivatives of other genes, notably CYC7 (Errede *et al.* 1980), have been isolated or, in the case of the *PGK* gene, created *in vitro* (Rathjen *et al.* unpublished). The ROAM effect is not due to the provision of a promoter reading out of a delta sequence, because in all cases the Ty element concerned is integrated in the orientation in which Ty transcription is directed away from the mutant gene (Williamson *et al.* 1983; Elder *et al.* 1983; Errede *et al.* 1980). The relationship of Ty and the mutant gene is analogous to the category III mode of activation of c-*myc* by ALV (Payne *et al.* 1982) and it has been assumed that some sort of enhancer-like phenomenon is involved. Recently we have defined the sequences within Ty that are required for a ROAM mutation of the *PGK* gene (Fig. 1) (Rathjen *et al.* unpublished). These sequences have been designated Ty Activator Sequences (TASs). There are two TASs, TAS1 and TAS2. TAS1 is located between nucleotides 90 and 184 and TAS2 is between 512 and 599. Both are able to activate a gene independently although together they act additively. Interestingly, unlike the 'classical' enhancers of higher eukaryotes their ability to activate is strictly orientation dependent. TAS1 and TAS2 differ in their efficiencies and in that TAS1 is closely linked to a determinant for cell type control. At present it is not clear whether the ROAM phenomenon is mediated by the same sequences that control Ty transcription.

Ty PROTEINS

Ty proteins have been identified by massively over-expressing all or part of the 5·7 kb transcriptional unit (Dobson *et al.* 1984; Mellor *et al.* 1985*a*). In particular either the *TYA* region alone or the complete transcriptional unit (*TYA* + *TYB*) of the class I element, Ty1-15, has been over-expressed to the extent that Ty proteins

become sufficiently abundant that they can be easily seen by SDS–PAGE. Six Ty proteins have been identified to date (Fig. 2). Protein p1 is encoded by *TYA*. It is slightly basic with a pI of 7·8, it is rich in proline residues, has a molecular weight of about $50\times10^3 M_r$, is phosphorylated and binds DNA in a crude binding assay. The sequence of the *TYA* gene reveals amino acid homology with Tn*3* resolvase and bacterial DNA binding proteins (Clare & Farabaugh, 1985; Warmington *et al.* 1985). A large ($190\times10^3 M_r$) protein, p3, is encoded by both *TYA* and *TYB*. The two genes are expressed as a TYA:TYB fusion protein *via* a frameshift event that brings them in phase (Clare & Farabaugh, 1985; Mellor *et al.* 1985*b* and Wilson *et al.* 1986). Preliminary data suggest that the frameshift does not result from a splicing event but that it is achieved at the ribosome (Clare & Farabaugh, 1985; Mellor *et al.* 1985*b*; Wilson *et al.* 1986). This phenomenon is directly analogous to the mode of expression of the *gag* and *pol* genes in a retrovirus such as RSV where *gag* is expressed either as a primary translation product of the full-length RNA or as a *gag-pol* fusion of about $180\times10^3 M_r$ (Varmus, 1983; Schwartz *et al.* 1983). Because *gag* and *pol* are out of phase the production of this $180\times10^3 M_r$ species also requires a frameshift event (Jacks & Varmus, 1985). The frameshift mechanisms in both Ty and retroviruses are unknown but it is likely that they will have features in common.

Like the primary translation products of retroviral *gag* and *pol* genes, p1 and p3 are precursor proteins that are proteolytically cleaved to produce other Ty protein species (Mellor *et al.* 1985*a*). p2 ($48\times10^3 M_r$) is the major product of p1 and p4 ($37\times10^3 M_r$), p5 ($33\times10^3 M_r$) and p6 ($28\times10^3 M_r$) are produced by further cleavage of p2. The protease responsible for the cleavage of p1 to p2 is encoded by the 5′ end of the *TYB* gene and has homology with the active sites of acid proteases (Adams *et al.* unpublished). Retroviruses encode a similar activity either at the 3′ ends of their *gag* genes or the 5′ ends of their *pol* genes, the retroviral activity being responsible for the maturation of viral proteins (Toh *et al.* 1985; Crawford & Goff, 1985; Katoh *et al.* 1985; Yoshinaka *et al.* 1985).

Ty TRANSPOSITION AND VIRUS LIKE PARTICLES

At least some of the Ty proteins are assembled into Ty virus like particles (Ty-VLPs). These have been observed in thin sections of yeast transformants that are over-expressing the entire 5·7 kb transcriptional unit (Mellor *et al.* 1985*c*; Garfinkel *et al.* 1985) and as purified particles (Mellor *et al.* 1985*c*)(Fig. 3). The Ty-VLPs are roughly spherical, 60 nm particles with a central core and what appears, in the electron microscope, to be a bipartite 'shell'. They resemble murine intracisternal A-type particles to some extent (Kuff *et al.* 1968) but are substantially smaller than authentic retroviral particles (80–120 nm) (Teich, 1982). The Ty particles contain 5·7, 5·0 and 2·2 kb RNA species and reverse transcriptase activity (Mellor *et al.* 1985c). It is also likely that they contain a primer:template substrate for the reverse transcriptase as labelled nucleotides can be incorporated into Ty DNA by the reverse transcriptase in the absence of any exogenous primer:template (Mellor *et al.* 1985*c*). By analogy with the retroviral systems it is likely that the full-length RNA is the

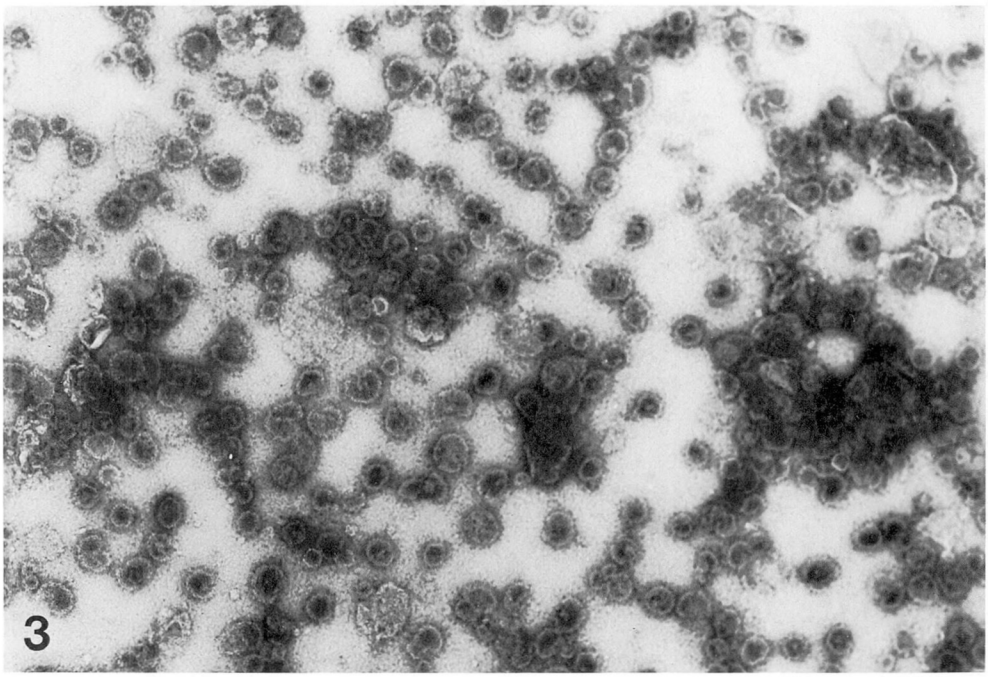

Fig. 3. An electronmicrograph of purified 60 nm Ty-VLPs. This electronmicrograph was produced by Dr Keith Gull, University of Kent.

template and we would suggest that the primer is a $tRNA^{Met}$ molecule that could anneal with the first 10 residues of the epsilon region in the full-length RNA (Eibel *et al.* 1980). We have shown that the reverse transcriptase activity is encoded by the *TYB* gene (Adams *et al.* unpublished) and there is substantial amino acid sequence homology with retroviral and other reverse transciptases in this region (Hauber *et al.* 1985; Clare & Farabaugh, 1985; Warmington *et al.* 1985) (Fig. 1). The major protein component within the particles is p2 with p1, p4, p5 and p6 present at lower amounts (Mellor *et al.* 1985*c*; Adams *et al.* unpublished).

As it is now known that Ty transposes *via* an RNA intermediate (Boeke *et al.* 1985), and that at least some of the protein and nucleic acid components required for an RNA mediated transposition event are assembled into a particle, it is tempting to suggest that the Ty-VLP is a transposition unit or 'transposisome'. The particles would be directly analogous to retroviral particles with the most significant difference being that Ty probably remains intracellular, and therefore can only perform intragenomic transposition events, whereas a retrovirus can transpose between genomes. The model for transposition would be similar to the replication cycle of a retrovirus (Varmus & Swanstrom, 1982). Following transcription from a member of the Ty family the 5·7 kb RNA would be packaged into a Ty-VLP containing a primer, possibly $tRNA^{Met}$, and reverse transcriptase. The particle may then remain inert for sometime. Eventually the full-length Ty RNA would be copied into a double-stranded DNA molecule. If this process is the same as the retroviral systems

the resulting DNA molecule would be circular and contain two complete copies of the delta sequences (Grandgenett & Vora, 1985). This circular derivative would then be the substrate for an integration reaction that would complete the transposition event. In retroviruses this integration step is catalysed by an integrase that is produced as a cleavage product of a *pol* encoded precursor. The enzyme recognizes the junction between the LTRs in the circular retroviral DNA and apparently random target sequences as components in the recombination event. No such integrase has been identified in Ty-VLPs although the predicted translation product of the *TYB* gene has partial amino acid homology with retroviral integrases (Mount & Rubin, 1985). Clearly the idea that Ty-VLPs are transposition units is easy to test. One flaw in the argument is that most if not all of the Ty-VLPs observed to date are cytoplasmic. This is not the expected location for an intermediate in transposition and might suggest that the particles are irrelevant to transposition and merely represent an obsolete viral packaging pathway that has been retained, perhaps, because it removes the potentially chaotic reverse transcriptase from the cell's RNA population and so prevents massive proliferation of retroposons. This question will be resolved shortly.

Ty VARIABILITY

An additional, intriguing property of yeast Ty sequences is their variability. This phenomenon can be broadly divided into major and minor differences. The major differences are those between class I and class II elements (Kingsman *et al.* 1981; Williamson *et al.* 1983). The minor differences are small sequence variations within the two classes. One possible explanation for this variability is that most Ty elements are inactive and therefore drift. In fact it has been suggested that the reason for the low frequency of Ty transposition is that most elements are inactive (Boeke *et al.* 1985). These proposals are based on the observation that transposition frequencies can be dramatically increased as a result of a minor increase in expression of a single active element. These arguments are simplistic. Many elements that can be distinguished on the basis of restriction pattern have been shown to transpose suggesting that many elements are at least *cis* active. Furthermore the five Ty elements sequenced to date all show the same genetic organization with only minor sequence variations. This is not compatible with most elements being free to drift but rather suggests that, at least as far as this sample is concerned, all elements have been subject to selection for their organization and integrity of the *TYA* and *TYB* genes. It has been claimed that inactive Ty variants are produced because transposition *via* a reverse transcriptase step is mutagenic yet the changes that were seen in recently transposed Ty elements were highly localized within the reverse transcriptase encoding region of the element (Boeke *et al.* 1985). An error-prone reaction would be expected to generate random mutations along the length of the element. These arguments are also based on the lack of fidelity of reverse transcriptase *in vitro* but there is no evidence that this is true *in vivo*. We would argue therefore that most Ty elements are active. The reason that transposition is rare yet Ty RNA is relatively

abundant might be that there is host control of transposition that is circumvented when Ty transcription is driven by a high efficiency non-Ty promoter.

The other perplexing feature of Ty variability is the maintenance of the two classes of elements within the same cell. This is best illustrated by comparison of the *TYA* regions (Fulton *et al.* 1985). The *TYA* regions roughly correspond to the region of the first substitution loop that is present in heteroduplexes between class I and class II elements (Fig. 1) (Kingsman *et al.* 1981; Williamson *et al.* 1983). The expectation was that these two regions would be quite different at the DNA level. Sequence information confirms this but rather than the difference being due to an apparent substitution the two *TYA* regions appear to differ by a large number of point mutations such that the aligned dispersed homology is about 64 %. This means that even though Ty elements are known to interact by gene conversion (Roeder & Fink, 1982), a process that leads to homogeneity in repetitive sequences, it would seem that the two classes have evolved by many small mutations rather than a single, perhaps recombinational, event. This problem is compounded when the protein sequences of the two *TYA* regions are compared. They share about 48 % homology at the protein level with 52 % of the differences being conservative suggesting that even though substantial variation has occurred at the DNA level that variation has been constrained by selection for function, presumably of p1. From the conservation of protein structure and from evidence of function it is quite clear that we do not have a situation where one class is junk while the other has function. It is possible that the two classes fulfil subtly different complementing functions although these functions have not been identified. Both classes transpose and both form Ty-VLPs (Fulton & Malim, unpublished). Alternatively they may have evolved separately and only recently come together within the same strain.

THE RELATIONSHIP OF Ty TO OTHER RETRO-ELEMENTS

It is clear that Ty has a very close structural and functional similarity to retroviruses and so might be considered as an endogenous retrovirus of yeast. However, this may be too narrow a view. Fig. 4 shows a comparison of the genetic organization of Ty and a number of other elements that use reverse transcriptase in their replication or transposition cycle (See also chapter by Hull *et al.* in this volume). Comparing Ty and the retroviruses first, it is clear that while *TYA* and *TYB* would correspond to *gag* and *pol* respectively Ty lacks an *env* region. In retroviruses the *env* region encodes proteins that are associated with the membrane envelope formed as the virus particles bud from their host cells. The absence of *env* in Ty is therefore compatible with its apparent restriction to a totally intracellular replication cycle (transposition) and its relatively small size. In this respect Ty is similar in organization to *copia* which also lacks an *env*. It is also interesting that a recently sequenced intracisternal A-type particle element (IAP-H18) possesses an *env* region but the gene is riddled with stop codons (Ono *et al.* 1985). Therefore Ty, *copia* and IAP-H18 are defective in their *env* regions. The second difference between Ty and the retrovirus genome is in the organization of their *pol* (*TYB*) regions. The regions

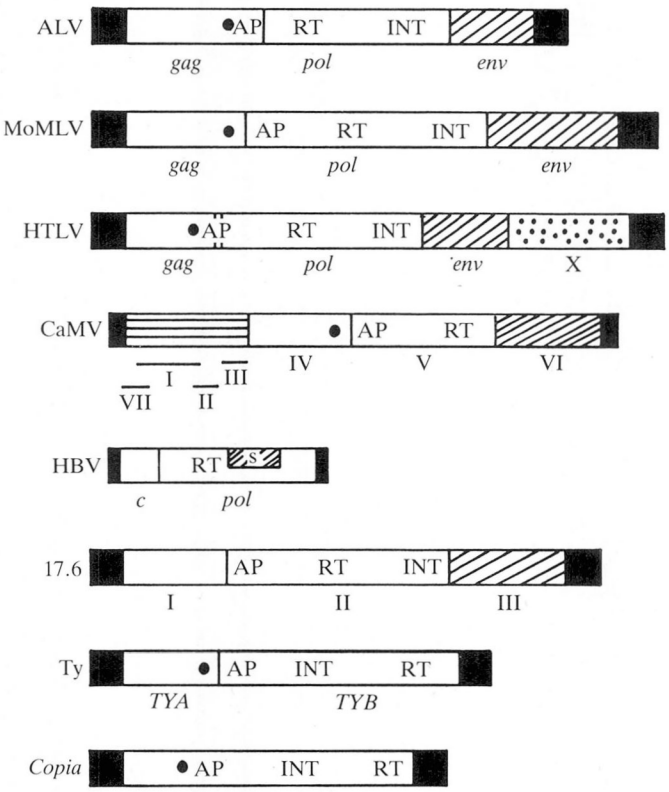

Fig. 4. A comparison of retroelements. The basic unit of a retroelement (see text) is shown as two open boxes. Filled box, terminal repeats; diagonal hatching, regions encoding structures that envelop the basic unit; horizontal hatching (CaMV), specialized accessories for vector and cell-to-cell transmission; stippled region, HTLV *trans* activator; ●, DNA binding sequence; AP, INT and RT, acid protease, integrase and reverse transcriptase homologies. Designations of coding regions are shown below each element. Sources of information are: avian leukosis virus (ALV) (Schwartz *et al.* 1983); Moloney murine leukaemia virus (MoMLV) (Shinnick *et al.* 1981); human T-cell leukaemia virus I (HTLVI) (Seiki *et al.* 1983); cauliflower mosaic virus (CaMV) (Hull & Covey, 1983); hepatitis B virus (HBV) (Ono *et al.* 1983); 17·6 and *copia* (Mount & Rubin, 1985).

within *pol* that are known to encode the integrase and the reverse transcriptase in retroviruses and the corresponding regions based on amino acid sequence homology in Ty appear to be interchanged. The organization of the *Drosophila* element *copia* resembles Ty whereas another *Drosophila* transposon, 17·6 (Saigo *et al.* 1984), resembles a retrovirus to such an extent that it might be prudent to ask if 17·6 forms infective particles. It would seem that Ty and *copia* should be grouped together as representatives of a class of elements that are distinct from retroviruses but which have a clear evolutionary relationship with them.

A broader view of retroelements might lead us to make an even more general statement about the significance of Ty and *copia*. It may be reasonable to regard them as basic units of a retro-system. This unit (shown as the open boxes in Fig. 4)

comprises a structural component (*gag* or *TYA*), that may be important in stabilizing the RNA 'genome' and/or presenting it for reverse transcription, and the enzymic functions (*pol* or *TYB*), the protease, integrase and reverse transcriptase system. All other retroelements could be regarded as comprising this unit with 'bolt-on' accessories needed to cope with specific hosts (Fig. 4). For example, retroviruses add an *env* region that controls cell-to-cell transmission and hepatitis B virus adds a surface protein gene. Similarly cauliflower mosaic virus adds the components of the syncytial inclusion body that contains the replicating genome, and a set of coding sequences, required for cell-to-cell infection *via* plasmodesmata and attachment to aphid mouthparts for vector transmission (see Hull & Covey, 1985). It is, therefore, too narrow a view to regard Ty specifically as an endogenous retrovirus. Instead it should be seen as a member of a more or less continuous family of retroelements within which, with *copia*, it represents a simple class. In the future other retroelements with other accessories attached to the 'Ty/*copia*' unit may be found. The nature of the accessories will determine the biological niche occupied by the basic unit.

We thank Mick Tuite and Keith Gull for generating the electron micrograph and Frank Caddick for preparing the Figures. P.D.R. is a Rhodes Scholar. J.M. is supported by an SERC:Celltech cooperative award. Work on Ty in this laboratory is supported by MRC and SERC grants to A.J.K. and S.M.K.

REFERENCES

BOEKE, J. D., GARFINKEL, D. J., STYLES, C. A. & FINK, G. R. (1985). Ty elements transpose through an RNA intermediate. *Cell* **70**, 491.

BOWEN, B. A., FULTON, A. M., TUITE, M. F., KINGSMAN, S. M. & KINGSMAN, A. J. (1984). Expression of Ty-lacZ fusions in *Saccharomyces cerevisiae*. *Nucl. Acids Res.* **12**, 1627.

BRULET, P., KAGHAD, M., XU, Y., CROISSANT, O. & JACOB, F. (1983). Early differential tissue expression of transposon-like repetitive DNA sequences of the mouse. *Proc. natn. Acad. Sci. U.S.A.* **80**, 5641.

CAMERON, J. R., LOH, E. Y. & DAVIS, R. W. (1979). Evidence for transposition of dispersed repetitive DNA families in yeast. *Cell* **16**, 739–751.

CLARE, J. & FARABAUGH, P. (1985). Nucleotide sequence of a yeast Ty element: evidence for an unusual mechanism of gene expression. *Proc. natn. Acad. Sci. U.S.A.* **82**, 2829.

CRAWFORD, S. & GOFF, S. P. (1985). A deletion mutation in the 5′ part of the *pol* gene of MoMuLV blocks proteolytic processing of the *gag* and *pol* polypeptides. *J. Virol.* **53**, 899.

CULLEN, R., LOMEDICO, P. T. & TU, G. (1984). Transcriptional interference in avian retroviruses – implications for the promoter insertion model of leukaemogenesis. *Nature, Lond.* **307**, 241–245.

DOBSON, M. J., MELLOR, J., FULTON, A. M., ROBERTS, N. A., BOWEN, B. A., KINGSMAN, S. M. & KINGSMAN, A. J. (1984). The identification and high level expression of a protein encoded by the yeast Ty element. *EMBO. J.* **3**, 1115.

EIBEL, H., GAFNER, J., STOTZ, A. & PHILIPPSEN, P. (1980). Characterisation of the yeast mobile element Tyl. *Cold Spring Harb. Symp. quant. Biol.* **46**, 609.

ELDER, R. T., LOH, E. Y. & DAVIS, R. W. (1983). RNA from the yeast transposable element Tyl has both ends in the direct repeats, a structure similar to retroviruses. *Proc. natn. Acad. Sci. U.S.A.* **80**, 2432.

ELDER, R. T., ST JOHN, T. P., STINCHCOMB, D. T. & DAVIS, R. W. (1980). Studies on the transposable element Tyl of yeast I. RNA homologous to Tyl. *Cold Spring Harb. Symp. quant. Biol.* **45**, 581.

ERREDE, B., CARDILLO, T. S., WEVER, G. & SHERMAN, F. (1980). ROAM mutations causing increased expression of yeast genes: their activation by signals directed toward conjugation functions and their formation by insertion of Tyl repetitive elements. *Cold Spring Harb. Symp. quant. Biol.* **45**, 961.

FLAVELL, A. J., RUBY, S. W., TOOLE, J. J., ROBERTS, B. E. & RUBIN, G. M. (1980). Translation and developmental regulation of RNA encoded by the eukaryotic transposable element copia. *Proc. natn. Acad. Sci. U.S.A.* **77**, 7107–7111.

FULTON, A. M., MELLOR, J., DOBSON, M. J., CHESTER, J., WARMINGTON, J. R., INDGE, K. J., OLIVER, S. G., DE LA PAZ, P., WILSON, W., KINGSMAN, A. J. & KINGSMAN, S. M. (1985). Variants within the yeast Ty sequence family encode a class of structurally conserved proteins. *Nucl. Acids Res.* **13**, 4097.

GARFINKEL, D. J., BOEKE, J. D. & FINK, G. R. (1985). Ty element transposition: reverse transcriptase and virus-like particles. *Cell* **42**, 507–517.

GRANDGENETT, D. P. & VORA, A. C. (1985). Site specific nicking at the avian retrovirus LTR circle junction by the viral pp32 DNA endonuclease. *Nucl. Acids Res.* **13**, 6205.

HAUBER, J., NELBOCK-HOCHSTETHER, P. & FELDMAN, H. (1985). Nucleotide sequence and characteristics of a Ty element from yeast. *Nucl. Acids Res.* **13**, 2745.

HULL, R. & COVEY, S. N. (1983). Replication of cauliflower mosaic virus DNA. *Sci. Prog., Oxf.* **68**, 403–422.

HULL, R. & COVEY, S. N. (1985). Cauliflower mosaic virus: pathways to infection. *Bio Essays* **3**, 160–163.

JACKS, T. & VARMUS, H. E. (1985). Expression of the Rous sarcoma virus *pol* gene by ribosomal frame shifting. *Science* **230**, 1237–1242.

KATOH, I., YOSHINAKA, Y., REIN, A., SHIBUYA, M., ODAKA, T. & OROSZLAN, S. (1985). Murine leukaemia virus maturation: protease region required for conversion from 'immature' to 'mature' core form and for virus infectivity. *Virology* **145**, 280.

KINGSMAN, A. J., GIMLICH, R. L., CLARKE, L., CHINAULT, A. C. & CARBON, J. (1981). Sequence variation in dispersed repetitive sequences in *Saccharomyces cerevisiae*. *J. molec. Biol.* **145**, 619.

KLEIN, H. L. & PETES, T. D. (1984). Genetic mapping of Ty elements in *Saccharomyces cerevisiae*. *Molec. Cell Biol.* **4**, 329–339.

KUFF, E. L., WIVEL, N. A. & LEUDERS, K. K. (1968). The extraction of intracisternal A-type particles from a mouse plasma-cell tumour. *Cancer Res.* **28**, 2137.

MELLOR, J., FULTON, A. M., DOBSON, M. J., ROBERTS, N. A., WILSON, W., KINGSMAN, A. J. & KINGSMAN, S. M. (1985*a*). The Ty transposon of *Saccharomyces cerevisiae* determines the synthesis of at least three proteins. *Nucl. Acids Res.* **13**, 6249.

MELLOR, J., FULTON, S. M., DOBSON, M. J., WILSON, W., KINGSMAN, S. M. & KINGSMAN, A. J. (1985*b*). A retrovirus-like strategy for expression of a fusion protein encoded by yeast transposon Tyl. *Nature, Lond.* **313**, 243.

MELLOR, J., MALIM, M. H., GULL, K., TUITE, M. F., MCCREADY, S., DIBBAYAWAN, T., KINGSMAN, S. M. & KINGSMAN, A. J. (1985*c*). Reverse transcriptase activity and Ty RNA are associated with virus-like particles in yeast. *Nature, Lond.* **318**, 583–586.

MOUNT, S. M. & RUBIN, G. M. (1985). Complete nucleotide sequence of the *Drosophila* transposable element *copia*: homology between *copia* and retroviral proteins. *Molec. Cell Biol.* **5**, 1630.

ONO, M., TOH, H., MIYATA, T. & AWAYA, T. (1985). Nucleotide sequence of the Syrian hamster intracysternal A-particle gene: close evolutionary relationship of type A particle gene to types B and D oncovirus genes. *J. Virol.* **55**, 387–394.

ONO, Y., ONDA, H., SASADA, R., IGARISHI, K., SUGINO, Y. & NISHIOKA, K. (1983). The complete nucleotide sequence of cloned hepatitis B virus DNA; subtype adr and adw. *Nucl. Acids Res.* **11**, 1747–1757.

PAYNE, G. S., BISHOP, J. M. & VARMUS, H. E. (1982). Multiple arrangements of viral DNA and an activated host oncogene in bursal lymphomas. *Nature, Lond.* **295**, 209.

PIKO, L., HAMMONS, M. D. & TAYLOR, K. D. (1984). Amounts, synthesis and some properties of intracisternal A particle-related RNA in early mouse embryos. *Proc. natn. Acad. Sci. U.S.A.* **81**, 488–492.

ROEDER, G. S. & FINK, G. R. (1982). Movement of yeast transposable elements by gene conversion. *Proc. natn. Acad. Sci. U.S.A.* **79**, 5621.

ROEDER, G. S. & FINK, G. R. (1983). Transposable elements in yeast. In *Mobile Genetic Elements* (ed. J. A. Shapiro), p. 299. New York: Academic Press.

ROEDER, G. S., ROSE, A. B. & PEARLMAN, R. E. (1985). Transposable element sequences involved in the enhancement of yeast gene expression. *Proc. natn. Acad. Sci. U.S.A.* **82**, 5428.

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, Lond.* **312**, 659.

SCHWARTZ, D. E., TIZARD, R. & GILBERT, W. (1983). Nucleotide sequence of Rous Sarcoma virus. *Cell* **32**, 853.

SEIKI, M., HATTORI, S., HIRAYAMA, Y. & YOSHIDA, M. (1983). Human adult T-cell leukaemia virus: complete nucleotide sequence of the provirus integrated in leukaemia cell DNA. *Proc. natn. Acad. Sci.* **80**, 3618–3622.

SHINNICK, T. M., LERNER, R. A. & SUTCLIFFE, J. G. (1981). Nucleotide sequence of Moloney murine leukaemia virus. *Nature, Lond.* **293**, 543–548.

SINGH, K., SARAGOSTI, S. & BOTCHAN, M. (1985). Isolation of cellular genes differentially expressed in mouse NIH 3T3 cells and a Simian Virus-40-transformed derivative: growth-specific expression of VL30 genes. *Molec. Cell Biol.* **5**, 2590–2598.

STRAND, D. J. & MCDONALD, J. F. (1985). *Copia* is transcriptionally responsive to environmental stress. *Nucl. Acids Res.* **13**, 4401–4410.

TEICH, N. (1982). Taxonomy of retroviruses. In *RNA Tumor Viruses* (ed. R. Weiss, N. Teich, H. Varmus & J. Coffin), p 25. Cold Spring Harbor Laboratory.

TOH, H., ONO, M., SAIGO, K. & MIYATA, T. (1985). Retroviral protease-like sequence in the yeast transposon Tyl. *Nature, Lond.* **315**, 691.

VARMUS, H. E. (1983). Retroviruses. In *Mobile Genetic Elements* (ed. J. A. Shapiro), p. 411. New York: Academic Press.

VARMUS, H. & SWANSTROM, R. (1982). Replication of retroviruses. In *RNA Tumor Viruses* (ed. R. Weiss, N. Teich, H. Varmus & J. Coffin), p. 368. Cold Harbor Laboratory.

WARMINGTON, J. R., WARING, R. B., NEWLON, C. S., INDGE, K. J. & OLIVER, S. G. (1985). Nucleotide sequence characterisation of Tyl-17, a class II transposon from yeast. *Nucl. Acids Res.* **13**, 6679.

WILLIAMSON, W. M. (1983). Transposable elements in yeast. Int. Rev. Cytol. **83**, 1.

WILLIAMSON, V. M., COX, D., YOUNG, E. T., RUSSELL, D. W. & SMITH, M. (1983). Characterisation of transposable element-associated mutations that alter yeast alcohol dehydrogenase II expression. *Molec. Cell Biol.* **3**, 20–31.

WILSON, W., MALIM, M. H., MELLOR, J., KINGSMAN, S. M. & KINGSMAN, A. J. (1986). Expression strategies of the yeast Ty element: a short sequence determines ribosomal frameshifting. *Nucl. Acids Res.* **14**, 7001–7016.

WINSTON, F., DURBIN, K. J. & FINK, G. R. (1984). The *SPT3* gene is required for normal transcription of Ty elements in *Saccharomyces cerevisiae*. *Cell* **39**, 675.

YOSHINAKA, Y., KATOH, I., COPELAND, T. D. & OROSZLAN, S. (1985). Murine leukaemia virus protease is encoded by the *gag-pol* gene and is synthesised through suppression of an amber termination codon. *Proc. natn. Acad. Sci. U.S.A.* **82**, 1618.