JOURNAL OF BACTERIOLOGY, Nov. 1997, p. 7118–7128 Vol. 179, No. 22
0021-9193/97/$04.00+0
Copyright © 1997, American Society for Microbiology

pCal, a Highly Unusual Ty1/*copia* Retrotransposon from the Pathogenic Yeast *Candida albicans*

GENEVIEVE D. MATTHEWS, TIMOTHY J. D. GOODWIN, MARGARET I. BUTLER, THOMAS A. BERRYMAN,† AND RUSSELL T. M. POULTER*

Department of Biochemistry, University of Otago, Dunedin, New Zealand

Received 27 June 1997/Accepted 8 September 1997

Retrotransposons are mobile genetic elements. They can transpose via the reverse transcription of mRNA into double-stranded DNA (dsDNA) followed by the insertion of this dsDNA into new sites within the host genome. The unintegrated, linear, dsDNA form of retrotransposons is usually very rare. We report here the isolation of a retrotransposon from *Candida albicans* which is unusual in this respect. This element, which we have named pCal, was first identified as a distinct band when uncut *C. albicans* DNA was examined on an agarose gel. Sequence analysis of the cloned element revealed that it is a retrotransposon belonging to the Ty1/*copia* group. It is estimated that pCal produces 50 to 100 free, linear, dsDNA copies of itself per cell. This is a much higher level of expression than even that of the system in which Ty1 is expressed behind the highly active *GAL1* promoter on a high-copy-number plasmid (about 10 copies per cell). Another unusual feature of pCal is that its Pol enzymes are likely to be expressed via the pseudoknot-assisted suppression of an upstream, in-phase stop codon, as has been shown for Moloney murine leukemia virus.

Candida albicans is an asexual yeast species which is the major fungal pathogen of humans. Although it is commonly found as a harmless commensal organism, inhabiting mucosal membranes and the digestive tract, it can cause superficial infections, such as oral thrush, in otherwise healthy people, and can cause severe, often fatal, systemic infections in immunocompromised patients (49). The recent increased use of immunosuppressive treatments and the increased incidence of immunosuppressive diseases, such as human immunodeficiency virus (HIV) infections, have meant that *C. albicans* infections are of increasing medical significance. There is significant strain variation within this species, potentially affecting virulence, and mobile retroelements have been suggested as one source of this (11).

Retroelements are a widespread family of sequences that can replicate via the reverse transcription of single-stranded RNA into double-stranded DNA (dsDNA) or are assumed to have arisen in this way. Two major types of retroelement are the retroviruses, such as HIV-1 (71) and Moloney murine leukemia virus (MMLV) (61), and the retrotransposons, such as Ty1 and Ty3 from *Saccharomyces cerevisiae* (3, 12, 30). The structures and life cycles of retrotransposons and retroviruses are very similar. The major difference between the two groups is that the retroviruses can form infectious virus particles which can be transmitted between cells and between individuals. Retrotransposons can form intracellular virus-like particles (VLPs), but they lack the genes coding for the viral envelope, so the VLPs are usually confined to the one cell.

Similarly to retroviruses, retrotransposons consist of an internal domain flanked by long terminal direct repeats (LTRs). In Ty1, for example, the LTRs are about 335 bp in length and the internal domain is about 5.3 kb long (3, 12). The internal region has two long open reading frames (ORFs) homologous to the *gag* and *pol* ORFs of retroviruses. The *gag* gene encodes the structural proteins which make up the VLP, while downstream, the *pol* gene encodes the enzymes required for reverse transcription and integration—protease, integrase, reverse transcriptase, and RNase H. The LTRs contain the promoter and the transcription termination signals and are functionally divided into three regions—U_3, R, and U_5. Transcription proceeds from the U_3/R boundary in the left LTR to the R/U_5 boundary in the right LTR to produce an RNA molecule which has the R region repeated at each end. Translation of this terminally redundant mRNA is usually regulated to ensure that the structural proteins of the VLP (Gag) are produced in much higher quantities than the enzymes (Pol). This is because large quantities of the Gag proteins are required for the assembly of the VLP, but only catalytic quantities of the Pol enzymes are required. In Ty1, for example, the Pol enzymes are produced at about 3% of the level of the Gag proteins (36).

The most common method of down-regulating the translation of the *pol* ORF is to have it out of frame relative to the upstream *gag* ORF. A rare, programmed ribosomal frameshift is thus required for translation of the *pol* ORF. A number of retrotransposons employ a +1 frameshift. Ty1 achieves this by tRNA slippage, while the Ty3 mechanism involves the skipping of a base. The Ty1-slippage mechanism involves a seven-base sequence, CUU AGG C. It is thought that a $tRNA^{LeuUAG}$, which can recognize all six leucine codons, slips forward one base from CUU-Leu to UUA-Leu during a translational pause caused by a rare $tRNA^{ArgCCU}$ (2). The Ty3 +1 frameshift also involves a seven-base sequence, GCG AGU U. An alanine-valine sequence (encoded by GCG-GUU) is produced, but tRNA slippage is not involved. It is thought that out-of-frame aminoacyl-tRNA binding or four-base decoding is responsible. Frameshifting is stimulated by the low availability of the tRNA decoding the AGU-Ser codon and also by the 12 nucleotides downstream of the AGU codon (22). Retrotransposons have also been found to use a −1 frameshift; an example is CfT-I of *Cladosporium fulvum*. Here the ribosome is thought to slip back one base on the sequence AAAA slightly upstream of the *gag* termination codon (45). Most retroviruses use a −1 frameshift to down-regulate *pol* (for example, Rous sarcoma virus

* Corresponding author. Mailing address: Department of Biochemistry, University of Otago, P.O. Box 56, Dunedin, New Zealand. Phone: 64 3 479 7856. Fax: 64 3 479 7866. E-mail: russell@sanger.otago.ac.nz.

† Present address: AgResearch, Grasslands, Palmerston North, New Zealand.

[RSV]) (60). Some, however, use read-through suppression of an in-frame UAG termination codon (for example, MMLV) (25, 77). In both cases this involves the ribosome stalling at a stem-loop structure, often a pseudoknot, in the mRNA and then, in the case of the −1 frameshift, slipping back one base on a slippery sequence (32, 33). In the MMLV read-through mechanism (which is not fully understood), the UAG stop codon is translated as glutamine (77).

An alternative method of down-regulation has been found in the *copia* retrotransposon (47, 78). Here the *gag* and *pol* ORFs are fused into one long continuous ORF, but a splicing reaction usually occurs prior to translation to excise most of the *pol* region from the mRNA. Only occasionally is a full-length RNA translated with the concomitant production of the Pol enzymes. A further form of *pol* down-regulation has recently been described for the Tf1 retrotransposon of *Schizosaccharomyces pombe* (1, 43). Here the *gag* and *pol* ORFs are also fused into one long continuous ORF, but no evidence for splicing of the mRNA has been found. It is believed that the Gag and Pol proteins are produced in equimolar amounts but that an enhanced rate of degradation of the Pol enzymes ensures that the Gag and Pol proteins end up in the right relative abundance.

Following translation, the retrotransposon proteins and RNA can form into a VLP. This consists of a shell of Gag proteins with the Pol enzymes and genomic RNA packaged inside (5, 21). The VLP is the site of reverse transcription. In general, the process of reverse transcription in retrotransposons is very similar to the well-characterized process of retroviral reverse transcription (3, 35). Two important steps in the reverse transcription process are the priming of minus- and plus-strand DNA synthesis. Minus-strand synthesis is most commonly primed by a cytoplasmic tRNA (often initiator methionine tRNA) which is packaged within the VLP along with the mRNA of the retrotransposon. The retrotransposon has a region adjacent to the left LTR, known as the minus-strand primer binding site [(−)PBS], which is complementary to the 3′ end of this tRNA. The tRNA binds to the retrotransposon RNA at the (−)PBS and can then be used by reverse transcriptase as a primer for the synthesis of minus-strand DNA. Plus-strand synthesis is primed by a short purine-rich sequence, known as a polypurine tract (PPT), located just upstream of the right LTR. After minus-strand DNA synthesis has passed this sequence, the RNA is nicked between the PPT and the LTR. The PPT RNA can then be used as a primer for the synthesis of the plus strand. Reverse transcription is generally very inefficient; greater than 10% of cellular mRNA can be retrotransposon RNA (17), yet the dsDNA form is not usually detectable by Southern blotting.

Following its synthesis, the dsDNA form of the retrotransposon may integrate at a new site within the host genome. This process is likely to involve a complex of the integrase enzyme associated with the two ends of the retrotransposon DNA (20, 46). In a process which is not well understood, the integrase complex must be released from the VLP, move into the nucleus, and then insert the DNA into a new genomic site. Studies with Ty1 and Ty3 have shown that the integration site selection mechanisms of these retrotransposons are nonrandom and appear to be specifically adapted to avoid causing disruption to the host genome (8, 9, 15, 19, 34).

Retrotransposons can be divided into three major groups based on their reverse transcriptase sequences and the order of the genes within their *pol* ORFs. Members of the Ty3/*gypsy* group are the most closely related to the retroviruses and share a Pol protein order: protease, reverse transcriptase, RNase H, and integrase. Examples of these elements are Ty3 of *S. cerevisiae* (13, 30), *gypsy* of *Drosophila melanogaster* (44), Tf1 of *Schizosaccharomyces pombe* (42), and *del* of *Lilium henryi* (62). Members of the *Pao* group, for example *Pao* of *Bombyx mori* (74) and Tas of *Ascaris lumbricoides* (23), have a *pol* gene order similar to that of Ty3/*gypsy* retrotransposons but can be distinguished from them by their reverse transcriptase sequence. Ty1/*copia* elements are most easily distinguished from Ty3/*gypsy* and *Pao* retrotransposons and retroviruses by the domain order of the Pol protein—protease, integrase, reverse transcriptase, RNase H. This group includes Ty1 and Ty2 of *S. cerevisiae* (3), *copia* (47) and 1731 (26) of *D. melanogaster*, Tst1 of *Solanum tuberosum* (7), and Tnt1 of *Nicotiana tabacum* (29).

The first *Candida* retroelement, Tca1, was identified through the discovery of multiple-copy isolated LTRs dispersed around the genome (11). These LTRs were discovered fortuitously in an analysis of moderate repeat elements. Subsequently, composite elements, named Tca1, consisting of two LTRs flanking a 5.5-kb internal domain, were also found. In the *C. albicans* strains tested, one to two Tca1 loci were found, indicating between one and four copies of Tca1, depending on whether the loci were homozygous or not. Tca1 has many features of a typical retrotransposon, including 388-bp LTRs, beginning with TG and ending with CA, with six-nucleotide inverted repeats, TGTTCG. . .CGAACA, at either end. The element is flanked by 5-bp duplications of the host DNA and is transcribed to give an approximately unit-length mRNA. Within the 5.5-kb internal domain, a (−)PBS and a plus-strand priming site are evident. The (−)PBS was not immediately obvious: no complementarity to $tRNA^{iMet}$ (as used by Ty1 and Ty3) could be found. Bases 31 to 39 of $tRNA^{Arg3}$ of *S. cerevisiae* (37), however, perfectly complemented the nine bases immediately adjacent to the left LTR (GATTAGAAG). There is, for some tRNA, a high degree of conservation between *S. cerevisiae* and *C. albicans*, leading to the suggestion that a cleavage product of a *C. albicans* $tRNA^{Arg}$ might serve as the primer. This suggestion is supported by the knowledge that the primer used by the *copia* retrotransposon is a cleavage product of $tRNA^{iMet}$ containing only the first 39 nucleotides (38).

Tca1 has been shown to be transcriptionally active, but an analysis of 1,200 bp of its internal sequence has indicated that it is defective, there being multiple stop codons in all three reading frames (6). It is remarkable, given the clearly nonfunctional nature of this element, that the LTRs remain identical and that the plus- and minus-strand priming sites remain in apparently functional form. It is possible that the defective Tca1 retrotransposon has been maintained via the passive reverse transcription of its RNA by the products of a functional *C. albicans* retrotransposon. This passive replication would require that the element has identical LTRs and functional plus- and minus-strand priming sites but would be independent of the element's internal sequence.

We report here the isolation and sequencing of pCal, an unusual Ty1/*copia* retrotransposon from *C. albicans*. The free, linear, double-stranded DNA form of this element is so highly expressed that it can be seen as a distinct band when uncut genomic *C. albicans* DNA is simply analyzed on an agarose gel. It contains features conserved in Tca1 and other retrotransposons and has additional features previously unreported in the retrotransposon family.

MATERIALS AND METHODS

Strains and culture conditions. The isolate iB65, a precursor to the *C. albicans* strain currently under investigation (hOG1042), was isolated as a *met2* heterozygote from an Otago University intermediate biology student in 1983. It was subsequently mutagenized with UV radiation (53) and *N*-methyl-*N*-nitro-*N*-nitrosoguanidine (53) to produce five strains—hOG758, hOG759, hOG760, hOG761, and hOG762—which are all *met2* homozygotes and also auxotrophic for adenine. hOG1042 is an *ade2*/*ade2 MET2*/*met2* revertant of hOG762. The

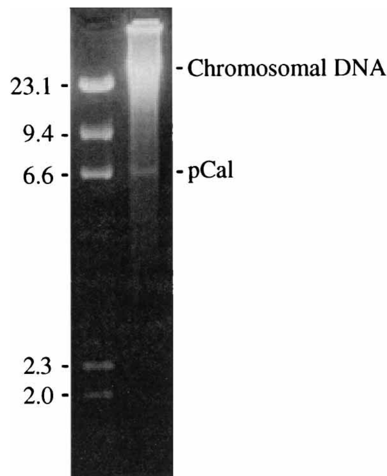

FIG. 1. Presence of a high-copy-number, extrachromosomal element in *C. albicans* hOG1042. An uncut sample of hOG1042 DNA was electrophoresed on a 1% agarose gel alongside some marker DNA (sizes are indicated to the left in kilobases). A distinct band of about 6.5 kb running ahead of the bulk of the chromosomal DNA (>20 kb) indicates the presence of an extrachromosomal element in this strain. The relative intensity of the band suggests that the element exists at about 50 to 100 copies per cell (see text). The gel photo was scanned with a Bio-Rad GS-670 imaging densitometer and annotated with Adobe Photoshop 4.0.

strains were grown at 37°C in YPD medium (1% yeast extract, 2% peptone, 2% glucose).

Enzymes. Agarase (GELase) and phosphatase (HK phosphatase) were purchased from Epicentre Technologies. T4 DNA ligase, RNase A, and restriction endonucleases were purchased from Boehringer Mannheim GmbH, Biochemica, Mannheim, Germany. Vent polymerase was purchased from New England Biolabs.

Nucleic acid manipulations. *C. albicans* genomic DNA was prepared essentially by the method of Cryer et al. (16). DNA was separated on 1% agarose gels with TAE buffer (56). Gel purification of DNA was from low-melting-point agarose with agarase. Bacterial plasmids were prepared by an alkaline lysis-polyethylene glycol precipitation method from Applied Biosystems, Inc. PCRs were performed with an Autogene II programmable cycling water bath from Grant Instruments (Cambridge), Ltd. Temperature cycling consisted of 35 cycles of 95°C for 1 min, 45°C for 1 min, and 72°C for 1 min. PCR products were purified for sequencing with the QIAquick PCR purification kit from Qiagen GmbH, Hilden, Germany.

Sequencing and nucleotide analysis. Sequencing was performed with a combination of subcloning and specifically designed oligonucleotide primers. The sequences were determined on an automated DNA sequencer (Applied Biosystems 373A DNA sequencer). Oligonucleotides were purchased from Macromolecular Resources, Fort Collins, Colo., or synthesized on an ABI 380B oligonucleotide synthesizer. Sequences were edited with SeqEd 1.0.3 (Applied Biosystems). Sequence contigs were assembled with VTUTIN 5.21 (63) and HOMED 5.14 (64). Other sequence analysis was carried out with version 8 of the University of Wisconsin Genetics Computer Group sequence analysis package (18). The ORFs were translated according to the nonstandard *C. albicans* genetic code (CUG codes for serine instead of leucine [59, 72]). Sequences for the alignments in Fig. 4 and for the phylogenetic analysis were obtained from the GenBank database under the following accession numbers: 17.6, A03971; 1731, S00954; CfT-I, Z11866; *copia*, A03324; dong, L08889; *gypsy*, B25666; HIV-1, K02013; Hopscotch, U12626; jockey, JT0396; MMLV, A03956, Osser, S32437; RSV, S26418; Ta1, S05465; Tf1, A36373; Tnt1, S04273; Tom, S34639; Tst1, X52387; Tx1, B32494; Ty1, B28097; Ty2, S45842; Ty3, S53577; Ty4, P47024; and Ty5, U19263. The trees were constructed by the unweighted pair group method using arithmetic averages (UPGMA), neighbor-joining, and parsimony methods available in the PHYLIP package (24). Bootstrapping was performed with SEQBOOT, and consensus trees were derived with CONSENSE (both programs also from PHYLIP).

Nucleotide sequence accession number. The nucleotide sequence of pCal has been submitted to GenBank and has been assigned accession no. AF007776.

RESULTS

Cloning and mapping. The work we describe here began when some uncut genomic DNA prepared from *C. albicans* hOG1042 was analyzed on an agarose gel and a distinct band running at about 6.5 kb was found (Fig. 1). Such a band had never previously been reported from any *Candida* strain or species. To analyze this feature, the band was extracted from an agarose gel and tested to see if it could be cut with restriction enzymes. A number of enzymes cut the band into smaller fragments which indicated that it was made up of dsDNA. At this point, we named the band pCal (for plasmid of *C. albicans*). The restriction digests allowed the construction of a simple restriction map of pCal. This work revealed that pCal was linear, with a *Pst*I site about 1 kb from one end, an *Eco*RI site about 1 kb from the opposite end, and an *Asp*718 site near the middle. To permit further analysis, the fragments of pCal produced with *Asp*718 were cloned into the *Asp*718 site of pUC19. Five clones were isolated, and each was found to contain just a single *Asp*718 site, the other apparently being destroyed during the cloning, as expected. Three of the clones contained a *Pst*I site, and two contained an *Eco*RI site.

Nucleotide sequence of pCal. The five plasmids containing the pCal fragments were all sequenced from both ends in the hope of finding an identifiable feature which would provide an insight into the nature of pCal. The first remarkable features to be found were 280-bp direct repeats. The existence of these direct repeats suggested that pCal was likely to be a retrotransposon. Because no other retrotransposon had ever been found existing at a high copy number in a free, linear, dsDNA form, we decided that it would be of interest to determine the complete sequence of pCal. Therefore, the three clones of pCal carrying the *Pst*I site and one of the two clones carrying the *Eco*RI site were completely sequenced. In addition, a region of pCal spanning the central *Asp*718 site used in the cloning was amplified by PCR, and each strand was sequenced. This analysis confirmed that there was only one *Asp*718 site and that therefore the clones that we had of each half of pCal truly represented adjacent fragments.

Assembly of the 6,426-bp pCal sequence revealed many characteristics typical of a retrotransposon. An obvious feature was the presence of identical 280-bp LTRs. The borders of these LTRs are short, imperfect, inverted repeats 6 bp long (5′-TGTTGG. . .CCATCA-3′). This repeat is very similar to that found in the LTRs of Tca1 (TGTTCG) (11), Ty3 (TGT TGTAT) (30), 1731 (TGTTG) (26), and copia (TGTTGGA AT) (47). Within the LTRs, putative TATA boxes and a polyadenylation signal were identified. These and other features are highlighted on the sequence of pCal in Fig. 2.

The (−)PBS was found adjacent to the left LTR and consists

FIG. 2. (A) General structure of pCal. The boxed triangles represent the LTRs. The long boxes represent the internal region. The arrows below the boxes indicate the extent of the two long ORFs. The positions of the encoded products are indicated: *GAG*, structural protein of the virus-like particle; PR, protease; INT, integrase; RT, reverse transcriptase; RNH, RNase H. The termination codon at the end of each ORF is indicated by a vertical line. Selected restriction sites are shown above the diagram: B, *Bgl*II; P, *Pst*I; Sac, *Sac*I; A, *Asp*718; Sal, *Sal*I; E, *Eco*RI. (B) Complete nucleotide sequence of pCal and deduced amino acid sequence of the two long ORFs (translated according to the nonstandard *C. albicans* genetic code). Every 10th nucleotide is indicated by a dot above the sequence. The terminal inverted repeats of the LTRs are underlined. The putative poly(A) signal and TATA boxes are highlighted in boldface and labelled above the sequence. The (−)PBS and the additional region complementary to the $tRNA^{Arg}$ fragment are in italics. The stop codon at the end of the *gag* ORF, the adjacent purine-rich tract (PRT), and the stems of the pseudoknot are highlighted in boldface. The purine-rich tract is also in italics. The 5′ and 3′ limits of the pseudoknot are indicated by < and >, respectively. The 3′ PPT (PPT1) and internal PPT (PPT2) are highlighted in boldface.

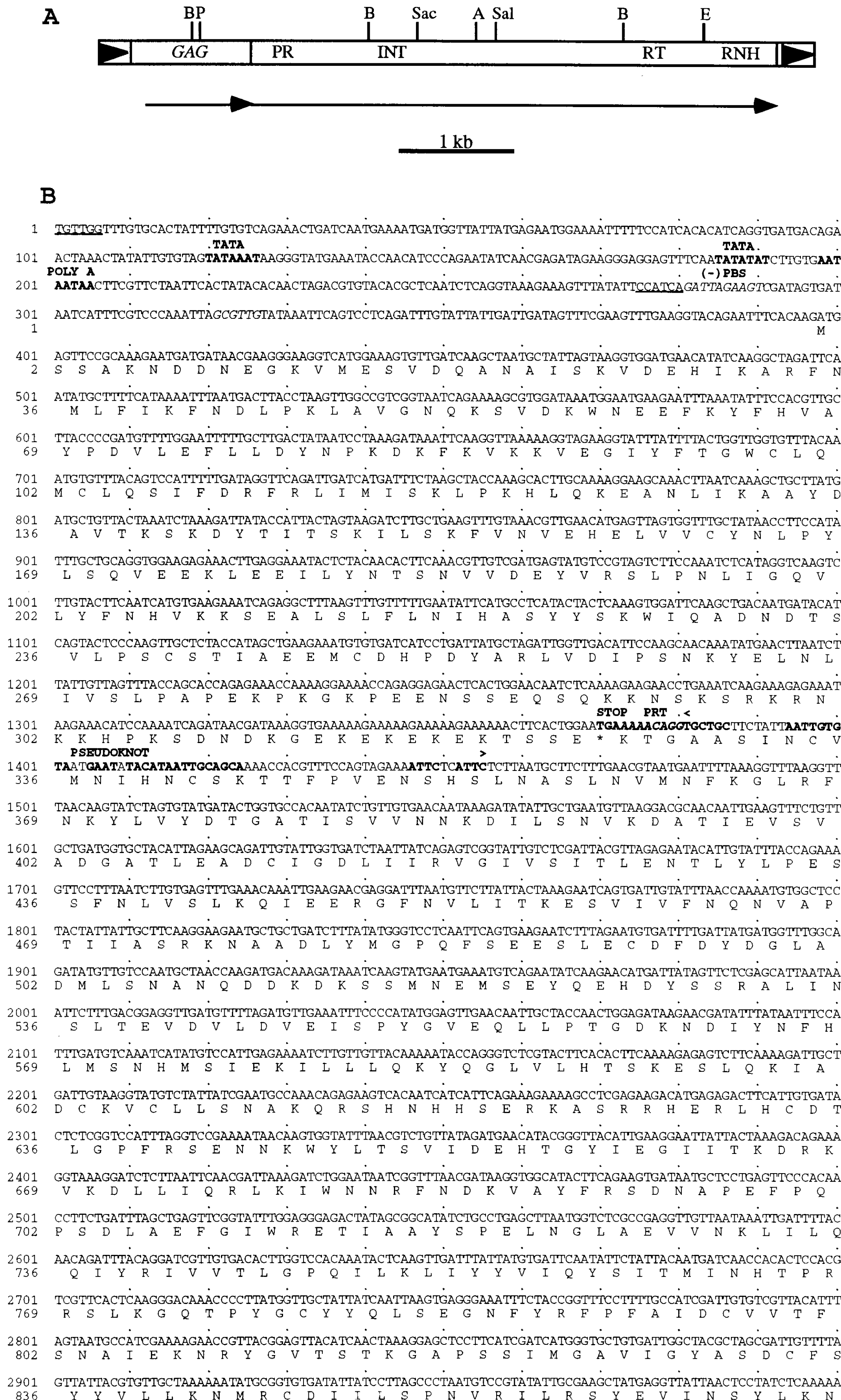

FIG. 2.

3001 CTTATCCACTACACCTATGTCACACATTGTTCCTATGGCTGAAGGTATCCAGGGAAGGCAACTGGGCGCTCAGTACGAGGTACGCGGAACATATGTGGAA
869 L S T T P M S H I V P M A E G I Q G R Q S G A Q Y E V R G T Y V E

3101 AGTGAATATGACAATACAAATGACGTGATGCACATGCCCAAAGAGTCATATTCAGTTCAGCCAGCATCGTTTACTTTAACTACGGGTAACAGTTCTAACG
902 S E Y D N T N D V M H M P K E S Y S V Q P A S F T L T T G N S S N E

3201 AATATGTTATAAATGATGATCCAGTACAGATTACCATTGAGAATCCCGATGATTTTTCTAACCCTCTTCAACTAACTGAAGAATCACACGATATGGTATC
936 Y V I N D D P V Q I T I E N P D D F S N P L Q L T E E S H D M V S

3301 CGAAGTAAAATCGGATGAGAATCCTAAACCCAGTCTCCACGAGCTAACACCTGGGGATAATCCGGTGTCTAAACCTCCTCAACTTGGTACCGAGACTTCA
969 E V K S D E N P K P S L H E L T P G D N P V S K P P Q L G T E T S

PPT2
3401 GTAATAGGGAAGTCTAAAGAGCCTATTACAAACCACACAAAGGACGCCCCTTCC**ATCCAGGGGAG**GGACCATAAACGCCTGGAATCTACTGCTCAGGTTG
1002 V I G K S K E P I T N H T K D A P S I Q G R D H K R S E S T A Q V G

3501 GACTATCACACCAACCCCAGACTGGTACTCCCGCTTCGGAGGAGTCAAAATTGTCAGGAACAGATCATTTCGGTGTCGACGTTGTTAAAGAAACAGTCTC
1036 L S H Q P Q T G T P A S E E S K L S G T D H F G V D V V K E T V S

3601 AGAAGATTGGCATACTTCTGACTACCCAGAAACTAGTGCTGAAGATGAACAGCAAAATCCCTCGTTACTGGCTAATAAGAATCGGGTAACTGAAAAAATA
1069 E D W H T S D Y P E T S A E D E Q Q N P S L S A N K N R V T E K I

3701 GATGAGGGAGAAAATATTTCATTTCCGGGGGGTGATGATGATTCTGTCGTGATCAACTCAAATGTTGAGCAATCTAATGTTGAAACAGAGGATGCTGGTA
1102 D E G E N I S F P G G D D D S V V I N S N V E Q S N V E T E D A G N

3801 ACAGTCCAATTCAAGACGAAGTTTCTCAAGAGGGAAGAATACTTAATGAACAAACTGATATAGTTGATACTGTTGCTAAAGTTATTGAGAATGAAAAAAT
1136 S P I Q D E V S Q E G R I L N E Q T D I V D T V A K V I E N E K I

3901 CTCTCCTATTAATTCATTAGATGATCATACTGAACTTGCTACAGACTCGGGAAATGATAGCAATTCAACAGAATCCGACATTCAATCGAAAAATGAAATA
1169 S P I N S L D D H T E L A T D S G N D S N S T E S D I Q S K N E I

4001 TCACCAGTGATTAATGAGAAAAATACTGAAATAATCCAAAAACACATTGAAAGTATCCTTGCTGATAAGAGATTGGATGAATTTGAAACGTATAATGTTG
1202 S P V I N E K N T E I I Q K H I E S I L A D K R L D E F E T Y N V D

4101 ATGAAATTGAGAATGTGATTAATGACGATGACATTGCTGAAGCTAATCCACTACCAGATGAAAATAATGATGTTCAGATGAATGAGAGTTTTGATAATAA
1236 E I E N V I N D D D I A E A N P L P D E N N D V Q M N E S F D N N

4201 TCATAGCATGTCACGAGCAAAGAAGAAATACACATTTGAGAAAGAAGTTAACGAAAAAATTGCTGGTACTAAACATTCACTTGATACAACTGATCCAAGA
1269 H S M S R A K K K Y T F E K E V N E K I A G T K H S L D T T D P R

4301 GAAGCAATCAGAGTGTTAAATACTGGTGAAACCAAGAGAATCGAACCCAAGAAAAGAGAGGTGCCTATCACTGTGAAATTAAACAAAAGATCGCAATACA
1302 E A I R V L N T G E T K R I E P K K R E V P I T V K L N K R S Q Y K

4401 AGTCACCATATGTTACAAGAAGTGGTAGAACGGTTATAAACCCCAAGAGGTATTTACATGCGGTCGTCAACAAAATCGACTATAATGATCCGGGATGGAT
1336 S P Y V T R S G R T V I N P K R Y L H A V V N K I D Y N D P G W I

4501 AAAGTCAATGAATGCTGAACTAGAGAAATTTAGATCAAAAGATGTTTACGAAGAAGTTCCAATTCCCACCGGTGTGAAGCCTATATCTATGGGTTGGGTA
1369 K S M N A E L E K F R S K D V Y E E V P I P T G V K P I S M G W V

4601 CATACTGAGAAAATTGATTCTCTCAAAGGTGTTGTTCGGAAATCACGTTGTGTTGTCCATGGCAACAGACAAAAGGAAAAATTGGATTATGACCCTTTTA
1402 H T E K I D S L K G V V R K S R C V V H G N R Q K E K L D Y D P F S

4701 GTGTTAGTTCACCTGTTATAGATCTTGTGACTATAAGATTATTGACAATAATAGGTTGTGAATTAGGAATGACAATTCAACATTTAGACGTCGAGTCGGC
1436 V S S P V I D L V T I R L L T I I G C E L G M T I Q H L D V E S A

4801 GTATCTAAATGCCTCTATTACTCATTCAAATCCAATTTATGTCTTTCCTCCTAAATCAGTACCTTTGAAGAAAAACCATTGTTGGTTATTGAAACGTTCT
1469 Y L N A S I T H S N P I Y V F P P K S V P L K K N H C W L L K R S

4901 GTCTATGGGTTAAAACAGTCGGGTTTGGAATGGTATCACACTATCAAAAGAGTATTGGAAGACATTGGTTTTACTCAAGTTTTACACAATGATGGTTTAT
1502 V Y G L K Q S G L E W Y H T I K R V L E D I G F T Q V L H N D G L F

5001 TTCACATTGAATATGAAGAGGGATCAGTAATATATTTAGGTTTATATGTTGATGATATTCTTATGGTTGGAAGTTCACAAAAAGTTATTGATAATTTTGT
1536 H I E Y E E G S V I Y L G L Y V D D I L M V G S S Q K V I D N F V

5101 GGATCAATTGAGAGATCATTTTGAAGTTAAAGTGTTTGGTGAAATATCAAATTATCTTGGTATTGAATTTCGTAAAACCGAATCTGGTTATATTTTATCT
1569 D Q L R D H F E V K V F G E I S N Y L G I E F R K T E S G Y I L S

5201 CAAGAAAAATTTCTCAAGAAATTACTTAAGGATTTCAAACTAGATGACTCATATGGAAAAACATACCCTGGATTCCGAATGACAAATATGAAAAGGTTG
1602 Q E K F L K K L L K D F K L D D S Y G K N I P W I P N D K Y E K V A

5301 CAATAATTCGTGAAAACGTTAATCCAGAGAATGATTTTGAAAAGGTTCCGAATGAGACATTGCTTGACCCTGATGCTAAAAAACTATACCAAAGTGGTGT
1636 I I R E N V N P E N D F E K V P N E T L L D P D A K K L Y Q S G V

5401 TGGCCTGCTTTTATGGGCTGCCACAAACACACGTCCAGATATATCGGTCGTAGTGAATTCGTTGGGTTCTAAATCTGCAAATCCAAATGTCCATGATTAT
1669 G S L L W A A T N T R P D I S V V V N S L G S K S A N P N V H D Y

5501 GAGAAATTGATTTATTGTCTTAGGTATATCAAAAATAGCATGGGATATCACATTGAGTACAAAAGAAACAGATTGAATATACCACCAAAATCATTTGTTA
1702 E K L I Y C L R Y I K N S M G Y H I E Y K R N R L N I P P K S F V I

5601 TCGAATGTTTCAGTGATGCGTCATTTGCACCAGGATTGGATAGAAAATCTATTAGTGGAACTTTGATTTATGTGAATGGAAATTTGGTGCAATGGGCGAC
1736 E C F S D A S F A P G L D R K S I S G T L I Y V N G N L V Q W A T

5701 CAAAAAACAAACGGTCATAGCACAAAGCTCAGCAGCTTGTGAAATGTTGGCTCTAAATTATACAATGTTGAAAGCTATCGAAATAAAAAACCATTTAATG
1769 K K Q T V I A Q S S A A C E M L A L N Y T M L K A I E I K N H L M

5801 GATTTGGGTTTTGAAGTAGGTAAGATACATTGTCATCAAGACAACCAAGCTGTGATTAAAGTTTTGAGAAATAACTATTGTCACCCACATCGACCAATAG
1802 D L G F E V G K I H C H Q D N Q A V I K V L R N N Y C H P H R P I D

5901 ATATCTGCTATAAGTTTCTACGCCAATTGATCAATGATAAAGTATTTTCAATATCCTATGTGAAGACAAATGATAATTACGCCGATTGTATGACTAAGTG
1836 I C Y K F L R Q L I N D K V F S I S Y V K T N D N Y A D C M T K C

6001 TCTAAGTCGTGCTAAATTCAAAGCATTCGTTGAGGGTATGATAAAACGGTTAGACCTAGAAGATAATCAAACACTGATACAAAATGCAATAACGGCAGAA
1869 L S R A K F K A F V E G M I K R L D L E D N Q T S I Q N A I T A E

PPT1
6101 TAAGTGGATTTATCATTACTATTATCGTAATGCTC**AATCAGGGGAG**TGTTGGTTTGTGCACTATTTTGTGTCAGAAACTGATCAATGAAAATGATGGTTA
STOP *

TATA
6201 TTATGAGAATGGAAAATTTTTCCATCACACATCAGGTGATGACAGAACTAAACTATATTGTGTAG**TATAAAT**AAGGGTATGAAATACCAACATCCCAGAA

TATA **POLY A**
6301 TATCAACGAGATAGAAGGGAGGAGTTTCAA**TATATAT**CTTGTG**AATAATAA**CTTCGTTCTAATTCACTATACACAACTAGACGTGTACACGCTCAATCTC

6401 AGGTAAAGAAAGTTTATATTCCATCA 6426

FIG. 2.

A

S. cerevisiae tRNA^{Arg3}

```
                                                               1
                        39          UCUG      GGUAAUGCGGUGCGCUCG-5'
                      3'-CUAAUCUUCAG    CGCAAC
                         |||||||||||    ||||||
pCal RNA  5'...UAUUCCAUCAGAUUAGAAGUC    GCGUUGUAUAAAUUCAGUCCU...3'
                                  G    A
                               UGAUA    UUAAA
                              G              C
                               AUAAUCAUUUCGUCC
```

B

```
Tca1 PPT:            GAATC-AGGG-AG
                      |||| |||| ||
pCal 3' PPT:          AATC-AGGGGAG
                       ||| |||||||
pCal internal PPT:    ATCCAGGGGAG
```

FIG. 3. Plus- and minus-strand priming sites of pCal. (A) (−)PBS. The region of pCal around the (−)PBS (bottom) is compared to the first 39 bases of tRNA^{Arg3} of *S. cerevisiae* (top) (37). The region of pCal shown here extends from base 271 to base 341. The bases of pCal within the LTR are underlined. For clarity, the bases of the tRNA molecule are shown in their unmodified form. (B) Comparison of the PPTs of pCal and Tca1. The Tca1 and pCal 3′ PPTs are adjacent to the right LTRs. The pCal internal PPT (bases 3455 to 3465) is from within the presumed integrase-coding region.

Protease

```
1731 (268)    TQWCLDSGATSHMC
copia (287)   CGFVLDSGASDHLI
Tnt1 (292)    SEWVVDTAASHHAT
Ty1 (456)     GHLLLDSGASRTLI
Ty4 (410)     KLVIIDTGSGVNIT
pCal (370)    KYLVYDTGATISVV
                   *^^^
```

Integrase (zinc finger)

```
1731 (397)    HKRNGH -28- CKTC
copia (419)   HERFGH -30- CEPC
Tnt1 (426)    HKRMGH -25- CDYC
Ty1 (599)     HRMLAH -32- CPDC
Ty4 (562)     HKRMGH -29- CQTC
pCal (568)    HLMSNH -29- CKVC
              * ^ ^*      *  *
```

Integrase (continued)

```
1731 (518)    KIKCIRSDNGGEFVNNVFDDYLKAHGIARQLTIPHTPQQNGVAERANRTLVEM
copia (543)   KVVYLYIDNGREYLSNEMRQFCVKKGISYHLTVPHTPQLNGVSERMIRTITEK
Tnt1 (543)    KLKRLRSDNGGEYTSREFEEYCSSHGIRHEKTVPGTPQHNGVAERMNRTIVEK
Ty1 (729)     SVLVIQMDRGSEYTNRTLHKFLEKNGITPCYTTTADSRAHGVAERLNRTLLDD
Ty4 (689)     KVREINSDRGTEFTNDQIEEYFISKGIHHILTSTQDHAANGRAERYIRTIITD
pCal (687)    KVAYFRSDNAPEFPQPSDLAEF...GIWRETIAAYSPELNGLAEVVNKLILQQ
              ^^    ^*^^ *^            **    ^    ^  ^*^^*^ ^^^^
```

Reverse Transcriptase

```
1731 (880)    HHMDVCTAYLNSEL..KDTVYMKQPQGFTDAANPDQVLLLRKAIYGLKQSGREWN -32- ILVYVDDLIL
copia (999)   HQMDVKTAFLNGTL..KEEIYMRLPQGISCNS..DNVCKLNKAIYGLKQAARCWF -34- VLLYVDDVVI
Tnt1 (919)    EQLDVKTAFLHGDL..EEEIYMEQPEGFEVAGKKHMVCKLNKSLYGLKQAPRQWY -33- LLLYVDDMLI
Ty1 (1343)    TQLDISSAYLYADI..KEELYIRPPPHLGM...NDKLIRLKKSLYGLKQSGANWY -29- ICLFVDDMVL
Ty4 (1381)    KTLDINHAFLYAKL..EEEIYIPHPHD......RRCVVKLNKALYGLKQSPKEWN -30- IAVYVDDCVI
pCal (1461)   QHLDVESAYLNASITHSNPIYVFPPKSVPL..KKNHCWLLKRSVYGLKQSGLEWY -33- LGLYVDDILM
                ^*^  *^*   ^   ^^^*   *           ^  * ^^ *****^   *         ^^***
```

RNase H

```
1731 (1129)   AFTGFVDADWGGDRLDRKSYTGYV
copia (1247)  KIIGYVDSDWAGSEIDRKSTTGYL
Tnt1 (1174)   ILKGYTDADMAGDIDNRKSSTGYL
Ty1 (1604)    KLVAISDASY.GNQPYYKSQIGNI
Ty4 (1639)    KVIAITDASV.GSEYDAQSRIGVI
pCal (1734)   VIECFSDASFAPG.LDRKSISGTL
                    *^^  ^   ^^^*  *
```

FIG. 4. Conserved motifs in the *pol* ORF of pCal compared to those of other Ty1/*copia* retrotransposons. Absolutely conserved amino acids are indicated by asterisks. Positions containing four or five identical amino acids or in which there are only two types of amino acids present are indicated by a caret (∧). The numbers in parentheses indicate the positions of the motifs from the start sites of the Gag-Pol fusion proteins.

of the sequence GATTAGAAGTC. This is very similar to the (−)PBS of Tca1, GATTAGAAG, but complements 11 bases (bases 29 to 39) rather than 9 bases of a possible tRNA^{Arg} cleavage product (37). The *S. cerevisiae* retrotransposons Ty1, Ty2, and Ty3 have been found to contain additional sequences 3′ to the (−)PBS which complement additional regions of the primer tRNA (73). These additional sequences are likely to be involved in the packaging of the primer tRNA within the VLP. An additional region of complementarity is also apparent in pCal; the sequence GCGTTG, approximately 30 nucleotides 3′ of the (−)PBS, perfectly complements the sequence CAACGC (bases 19 to 24) in the primer tRNA^{Arg} fragment (Fig. 3).

A plus-strand priming site, or PPT, was found immediately upstream of the right LTR. It is very similar to the PPT described for Tca1. Interestingly, a second sequence very similar to the 3′ PPT was found near the middle of pCal (bases 3455 to 3465). Internal PPTs which function as plus-strand priming sites have been identified in Ty1 (31, 52) and HIV-1 (10) and may serve to speed up the reverse transcription process. The two pCal PPTs and that of Tca1 are compared in Fig. 3. We believe that the internal PPT of pCal may also be serving as a site for plus-strand initiation during the reverse transcription process (28).

Tca1 and pCal have very similar (−)PBSs and PPTs and very similar borders to their LTRs. A comparison of the remainder of the LTRs, however, revealed that the similarity did not extend beyond these regions (data not shown).

The ORFs. Two long ORFs were found in pCal; the first was 972 bp long (324 amino acids), and the second was 4,728 bp long (1,576 amino acids). Conserved motifs from the four *pol*-coded proteins—protease, integrase, reverse transcriptase, and RNase H—were identified in the second ORF. The order of these motifs (as listed above) places pCal within the Ty1/*copia* group of retrotransposons. The pCal motifs are compared to those of other Ty1/*copia* elements in Fig. 4. No conserved motifs were found in the first ORF, but it is similar in size and position to the *gag* genes of other retroelements. Retroelement *gag* genes in general are known to be extremely variable, and it is not uncommon for no identifiable conserved features to be present.

Unlike other retrotransposons, the *gag* and *pol* ORFs of pCal are in the same phase separated only by a UGA termination codon. This arrangement is similar to what has been found for mammalian type C retroviruses, such as MMLV (25, 77). In MMLV, a UAG termination codon separates the *gag* and *pol* ORFs. Translation of the *pol* ORF occurs via the occasional read-through suppression of the UAG codon. This suppression requires an 8-bp purine-rich sequence immediately downstream of the stop codon and an adjacent pseudoknot (a structural element of RNA formed upon the annealing of the nucleotides of a loop region with nucleotides outside of that loop [67]). In pCal, an 8-bp purine-rich sequence, AA AACAGG, lies immediately downstream of the UGA codon, and this is followed immediately by a potential pseudoknot. These features are illustrated in Fig. 5. A further unusual feature is apparent slightly upstream of the UGA codon. It consists of four tandem repeats of the sequence GAAAAA. The role, if any, of this distinctive sequence in the ribosomal *gag-pol* transition is unclear.

Copy number of pCal. The copy numbers of other extrachromosomal elements from lower eukaryotes have been determined. For instance, the 2μm circle plasmids of *Saccharomyces* species exist at 50 to 100 copies per cell (14, 68), and the Ddp elements of *Dictyostelium discoideum* exist at 50 to 300 copies per cell (48). When uncut genomic DNA from the *Saccharomyces* and *Dictyostelium* species containing these elements is run out on agarose gels, the extrachromosomal elements appear as distinct bands running ahead of the chromosomal DNA (48, 68). The intensity of the bands relative to that of the chromosomal DNA is indicative of the elements' copy numbers. These elements are comparable in size to pCal, and the host genomes are similar in size to that of *C. albicans*. Therefore, using the relative intensity of extrachromosomal and chromosomal DNA in *Saccharomyces* and *Dictyostelium* as a guide, we estimated, from the relative intensity of pCal and hOG1042 chromosomal DNA, that pCal exists at 50 to 100 copies per cell. This estimate, however, will require accurate experimental confirmation.

Phylogenetic analysis. In an attempt to gain a better understanding of the relationship of pCal to other retroelements, a phylogenetic tree of a number of retrotransposons and retro-

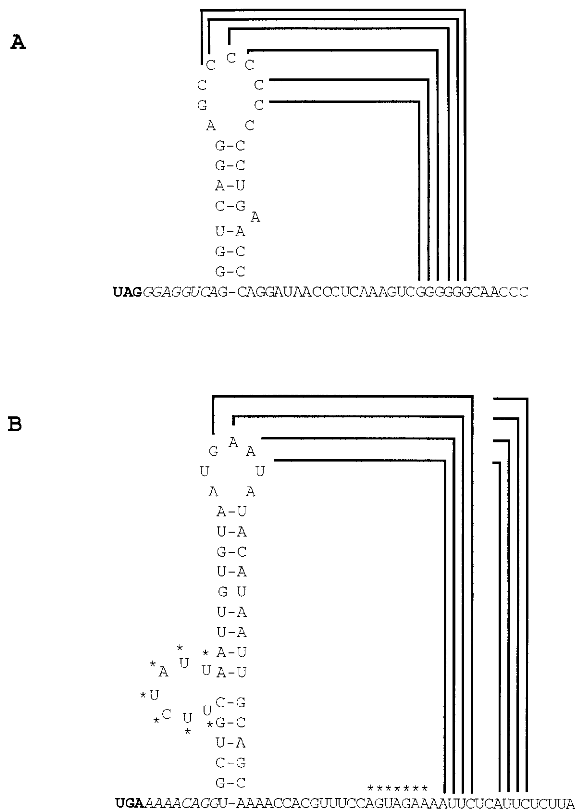

FIG. 5. Comparison of the putative pseudoknot structures of MMLV (A) and pCal (B) at the boundary of their *gag* and *pol* ORFs. The stop codons are shown in boldface, and the 8-bp purine-rich tracts are shown in italics. The long lines represent the base pairings in the second stems. Note that in pCal, there are two downstream regions to which the first loop of the pseudoknot can anneal. The nucleotides in the bulge of the first stem of pCal also have a downstream region to which they can potentially anneal (bases marked with asterisks). Base pairing between these sequences could lead to the formation of an alternative pseudoknot.

viruses was constructed. The data used in the analysis were the predicted amino acids of the seven conserved domains of reverse transcriptase identified by Xiong and Eickbush (75). The tree was constructed by the UPGMA method within the PHYLIP package (24) and is shown in Fig. 6. It is generally consistent with the trees constructed earlier by Xiong and Eickbush (75) and Xiong et al. (74). For instance, the retroviruses and the *gypsy*-type retrotransposons are closer to each other than to the Ty1/*copia* retrotransposons. Within the retroviral group, HIV-1 and RSV are closer to each other than to MMLV, and within the Ty3/*gypsy* group, CfT-I and Tf1 form a group, as do the *Drosophila* elements 17.6, Tom, and *gypsy*. The tree placed pCal with the Ty1/*copia* elements. This placement of pCal is in agreement with the fact that pCal has the Pol protein order protease, integrase, reverse transcriptase, RNase H. Such an order is diagnostic for Ty1/*copia* elements. Within the Ty1/*copia* division, two broad groups are apparent. One group contains the *Saccharomyces* elements Ty1, Ty2, and Ty4, and the other contains *copia* and 1731 of *Drosophila*; Ty5 of *Saccharomyces*; the plant elements Hopscotch, Tst1, Ta1, and Tnt1; Osser from the green alga *Volvox carteri*; and pCal. Within the second group, pCal is the most divergent element. Similar results were obtained by the neighbor-joining and parsimony methods of tree construction (not shown).

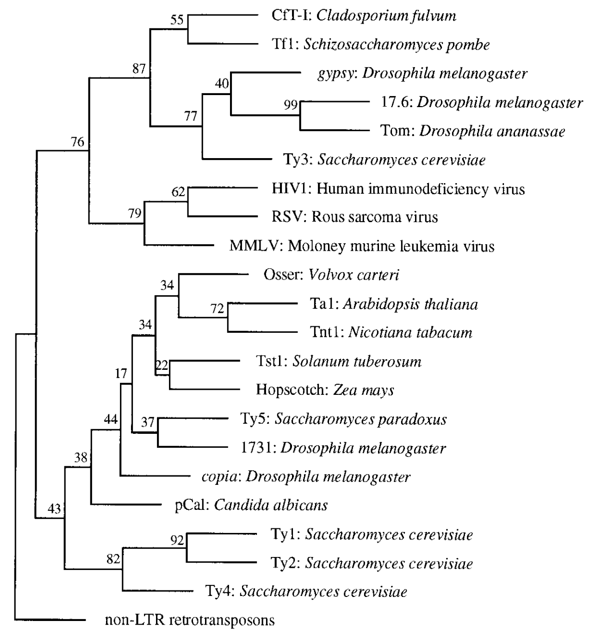

FIG. 6. Phylogenetic tree of some LTR retroelements. The data used in the tree construction were the predicted amino acids of the seven conserved domains of reverse transcriptase identified by Xiong and Eickbush (75). The tree was constructed by the UPGMA method available within the PHYLIP package (24). The percentages of trees (from 500 bootstrap replications) supporting each branch are indicated. Non-LTR retrotransposons were used as an outgroup to root the tree. The accession numbers for the sequences of the elements can be found in Materials and Methods.

Partial sequencing of additional clones of pCal. At the start of this work, all five of the clones of pCal were partially sequenced. When the partial sequences of the three clones carrying the *Pst*I site, which represent the left half of pCal, were compared, it was found that one clone differed from the other two at a small number of sites. To determine the full extent of these differences, we decided to completely sequence each of these three clones. When the sequences were compared, we found that two of the clones were identical, but they differed from the third clone at 12 sites. The differences were all base substitutions. This finding suggested the possibility that the total population of pCal within a cell might be made up of a number of subpopulations with different sequences. Such a situation could arise in a number ways. For instance, there could be a number of integrated retrotransposons varying in sequence, each contributing to the pCal population. Alternatively, pCal could be a self-sustaining molecule (i.e., independent of any integrated copies), and the inherent inaccuracy of reverse transcriptase could be introducing variation into the system. To investigate this idea further, we obtained four additional clones of pCal from a region which differed among the original clones (from the 5′ border of the 5′ LTR to the *Pst*I site at position 905). The region of greatest variability was then sequenced in each of these new clones. Analysis of the sequences revealed that the four new clones were identical in sequence to each other and to the two original clones which had been found to be identical. This result suggests that the majority of the pCal molecules in the total pCal population are likely to be very similar, if not identical, in sequence. One cannot, however, rule out the possibility that more than one integrated retrotransposon is contributing to the pCal population or that pCal is a self-sustaining system.

Comparison with Kiwi, a potential *C. albicans* retroelement. The sequence of another *C. albicans* element, potentially retrotransposon-like in nature, has recently been submitted to the databases by a group in the United Kingdom (accession no. Y08494). This element has been named Kiwi and is defined as an LTR. It consists of a repeated sequence about 400 bp in length, flanked by 5-bp direct repeats of the host DNA and associated with tRNA genes. The borders of the element consist of short, imperfect, inverted repeats: 5′-TAATGTATA . . .TATACAACA-3′. Such an element is reminiscent of the isolated LTRs of other retrotransposons which are the result of homologous recombination between the ends of a retrotransposon with the concomitant deletion of the internal region (54). No significant similarity is detectable between the Kiwi sequence and the LTRs of Tca1 or pCal.

DISCUSSION

Analysis of the complete 6,426-bp sequence of pCal revealed that it is a free (i.e., unintegrated), dsDNA form of a new retrotransposon belonging to the Ty1/*copia* group. Initially, no significant similarity at the nucleotide level was found between pCal and any other sequence in the databases. This was not considered surprising, however, because reverse transcriptase has no editing function, so reverse transcriptase-based elements have a higher mutation rate than those utilizing other polymerases (27). A more appropriate and useful analysis was to look for the conserved functional motifs expected to be present. Such areas have tight evolutionary constraints and are often similar, even in highly divergent elements such as *copia* and *gypsy*. A close examination of the sequence revealed that pCal has many of the features commonly found in retrotransposons. Such features include the 280-bp LTRs with short inverted repeats and putative transcriptional initiation and termination signals, a (−)PBS adjacent to the left LTR, a PPT adjacent to the right LTR, and two long ORFs, the first similar in size and position to the *gag* ORFs of other retroelements and the second containing motifs homologous to *pol* ORFs. Within the *gag* ORF of pCal, no nucleic acid binding motif could be identified. A $CX_2CX_4HX_4C$ nucleic acid binding motif is found within the *gag* ORF of some retrotransposons of the Ty1/*copia* group (for example, Ta1, *copia*, 1731, and Tp1) (55). However, this motif is not found in the functional retrotransposon Ty1. Taken together, all of the features required for retrotransposition appear to be intact in pCal, suggesting that it is likely to be a functional retrotransposon.

The order of the motifs within the *pol* gene of pCal (protease, integrase, reverse transcriptase, RNase H) suggests that pCal is a member of the Ty1/*copia* group. In agreement with this observation, a phylogenetic analysis based on the reverse transcriptase genes of a diverse range of retroelements also placed pCal within the Ty1/*copia* group (Fig. 6). This analysis, however, also revealed that pCal has no close relatives within the known set of Ty1/*copia* retrotransposons: pCal was placed as the most divergent element in a large group of retrotransposons containing representatives from plants (Ta1, Tnt1, Hopscotch, and Tst1), insects (*copia* and 1731), a green alga (Osser), and yeast (Ty5). It is probable that the reverse transcriptase of pCal is functional, and so, therefore, this placement of pCal is probably a genuine reflection of the divergent nature of this element, rather than being the result of the unselected accumulation of random mutations.

Within the LTRs of pCal, there was no extended DNA sequence homology to the other *C. albicans* retroelements, Tca1 and Kiwi. Tca1 and pCal do, however, have features such as similar inverted terminal repeats on their LTRs and a very similar PPT sequence, and they potentially utilize the same $tRNA^{Arg}$ fragment as a primer. The Tca1 (−)PBS complements nine nucleotides at the 3′ end of the $tRNA^{Arg}$ fragment (bases 31 to 39). The pCal (−)PBS complements 11 nucleotides of the $tRNA^{Arg}$ fragment (bases 29 to 39) and, similarly to what has been found in Ty1, Ty2, and Ty3, pCal has an additional sequence downstream of the (−)PBS which complements a further 6 bases (bases 19 to 24) of the $tRNA^{Arg}$ fragment.

Given that pCal and Tca1 are believed to use an internal fragment of the $tRNA^{Arg}$ (nucleotides 1 to 39), it is of great interest that the retrotransposon *copia* uses the first 39 nucleotides of $tRNA^{iMet}$ as a primer (38). It is not clear if the fragment is the result of normal tRNA degradation. Kikuchi et al. (39, 40) have suggested that the *copia* primer is a product of hyperprocessing of $tRNA^{iMet}$ by *Drosophila* RNase P. Hyperprocessing was defined as the processing of a mature tRNA to produce another functional RNA molecule, although, to date, the only assigned function of these tRNA fragments is as primers for retrotransposons. The RNA component of *Escherichia coli* RNase P was shown to cleave a number of sites in the $tRNA^{iMet}$, one of these being between nucleotides 39 and 40 (39, 40). The *Drosophila* $tRNA^{iMet}$ and yeast $tRNA^{Arg3}$ have very similar physical structures in terms of the numbers and positions of loops and stems, the residues in each loop, the number of base pairs in each stem, and the total number of nucleotides in the tRNA (37, 40). It is therefore possible that a similar hyperprocessing reaction is occurring with a $tRNA^{Arg}$ in *C. albicans* to produce the primers for pCal and Tca1.

If pCal is using a tRNA fragment for priming, there are implications for control of replication. An element using a whole tRNA as a primer has a pool of normal, functional tRNAs to draw on, even if the tRNA in question is a rare one. Elements using a fragment, however, have to contend with the stability of tRNAs and the possibility that once a tRNA starts degrading, it may be rapidly further degraded. The elements using a fragment as a primer will have to bind the tRNA after only partial degradation. This process could be a limiting step in the reverse transcription process and consequently could control the copy number of pCal.

Most retrotransposons and retroviruses have been found to have their *gag* and *pol* ORFs lying in different phases on the mRNA. The necessary down-regulation of the *pol* gene with respect to the *gag* gene is thus brought about by the fairly low frequency of ribosomal frameshifting from the *gag* reading frame to the *pol* reading frame. There are, however, exceptions to this finding. For instance, the *gypsy*-type retrotransposon Tf1 from *Schizosaccharomyces pombe* has its *gag* and *pol* ORFs fused into one long ORF (42, 43). The *gag* and *pol* gene products are thus produced in equal amounts. The required excess of Gag protein to Pol enzyme is produced posttranslationally, via an enhanced rate of degradation of the Pol enzymes (1). Some insect and plant retrotransposons of the Ty1/*copia* group, for example, *copia* (47), Ta1 (70), and Tnt1 (29), also have their *gag* and *pol* ORFs fused into one long ORF. In *copia*, at least, the down-regulation of *pol* occurs by the frequent splicing of the mRNA to remove most of the *pol* ORF (78). The fact that the *gag* and *pol* ORFs of pCal are in the same phase implies that pCal is another retrotransposon that doesn't use frameshifting to down-regulate *pol*. Instead it seems likely that some form of stop codon suppression is required for translation of the *pol* ORF, and this would also be likely to result in the down-regulation of *pol* relative to *gag*. It is therefore interesting that pCal has some structural similarities to mammalian type C retroviruses, such as MMLV, in the vicinity of the *gag*/*pol* boundary (25). In MMLV, a UAG stop

codon which separates the *gag* and *pol* ORFs is suppressed with an efficiency of about 5%, being translated as glutamine. An 8-bp purine-rich sequence immediately 3′ to the stop codon and an adjacent pseudoknot structure are both necessary and sufficient for stop codon suppression. Mutations disrupting the stems of the pseudoknot impaired suppression, and compensatory mutations restored suppression. Also, the sequence of the purine-rich tract between the stop codon and the pseudoknot was found to be critical, and it is likely that the length of this sequence is important. The MMLV read-through mechanism is not yet fully understood, but a pseudoknot-induced ribosomal pause at the suppressed UAG codon is likely to be involved (25). Similarly to MMLV, pCal has an 8-bp purine-rich sequence immediately 3′ to the UGA stop codon, although not the same sequence as in MMLV, and it has a putative pseudoknot (Fig. 5). There is only the 8-bp purine-rich sequence between the termination signal and the start of the putative pseudoknot. It is therefore likely that a similar form of read-through suppression is occurring in pCal and MMLV.

It has been reported that *C. albicans* and some other closely related *Candida* species contain a tRNA capable of suppressing UAG and UGA stop codons. This tRNA, $tRNA^{SerCAG}$, was originally identified as being responsible for the translation of the universal CUG-leucine codon as serine in certain *Candida* species (50, 76). This phenomenon has been well documented now (41, 59, 65, 72), and while a number of deviations from the "universal" genetic code have been reported in nuclear and mitochondrial genomes (51), this is the only example of the reassignment of a nuclear sense codon. The $tRNA^{SerCAG}$ has some unusual structural features (76), and a recent report has even shown that $tRNA^{SerCAG}$ can be charged to a low degree (about 3%) with leucine and can incorporate this leucine into proteins during translation (66). This is one of the first examples of the assignment of a single tRNA species to two amino acids. This strange tRNA was also implicated in some aberrant translational events reported by Tuite and coworkers (57, 69). This group found that when *C. albicans* tRNAs were added to in vitro translation systems, proteins which migrated more slowly than expected on sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) gels were produced. They initially interpreted these results as evidence that *C. albicans* contains a tRNA capable of suppressing UAG and UGA stop codons. In a later paper, the tRNA responsible for the unusual translational events was identified as $tRNA^{SerCAG}$ (58). In the same paper, however, they showed that their earlier results could not be simply explained by $tRNA^{SerCAG}$ being an omnipotent nonsense suppressor; they found that the amino-terminal regions of proteins synthesized in the presence of $tRNA^{SerCAG}$ also migrated more slowly than expected by SDS-PAGE. At present, it is unclear what the actual effects of $tRNA^{SerCAG}$ are, aside from incorporation of serine at CUG codons. This leaves open the question of what molecule it is that mediates the suppression of the UGA termination codon at the *gag/pol* boundary of pCal. Sequencing of the Gag- and the Gag-Pol fusion proteins and mutational analyses of the regions surrounding the stop codon could be used to determine the mechanism by which the *pol* genes of pCal are translated.

The Tca1 element described by Chen and Fonzi (11) appears to be defective, having multiple stop codons in all three reading frames within the internal region. However, it has identical LTRs, a feature suggesting that it has transposed recently. A possible explanation for this unusual occurrence is that the transposition of Tca1 is being supported by a functional *C. albicans* retrotransposon. This would require that Tca1 produce a full-length RNA, that this RNA be packaged within the VLPs of the functional retrotransposon, and that the enzymes within these VLPs would be able to recognize and process the Tca1 RNA. Tca1 has been shown to be transcriptionally active, and the finding that pCal and Tca1 have very similar (−)PBSs, PPTs, and borders to their LTRs suggests that pCal could well support the retrotransposition of Tca1.

The pCal system produces much more free dsDNA—estimated at 50 to 100 copies per cell—than any other reported retrotransposon system. This is true even of the system in which Ty1 of *S. cerevisiae* is expressed off a high-copy-number plasmid under the control of the highly inducible *GAL1* promoter (4). Such a *GAL* promoter system is capable of producing about 10 dsDNA copies per cell, and the DNA requires Southern blotting before it can be detected (20). It has been suggested that the cause of this paucity of dsDNA in this system is inefficient reverse transcription (20). Given that a retrotransposon that produces vast amounts of its free, linear, dsDNA form has not previously been reported, it would be intriguing to determine what is different about pCal. In this respect, it is of interest that we have detected integrated retrotransposons, similar in sequence to pCal, which we have named Tca2 (28). This integrated form has been detected in a diverse range of *C. albicans* strains. Extremely high levels of the free, linear, dsDNA form (pCal), however, have only been detected in hOG1042 and its close relatives (descendants of iB65) (28). This suggests that hOG1042 and its relatives are carrying some sort of mutation. However, it is not obvious whether it is a mutation within some cellular function that Tca2-type retrotransposons use for their regulation or whether pCal is being produced from a mutant Tca2 retrotransposon in an otherwise normal cell. We are presently trying to determine the cause of the appearance of pCal. Possibilities being considered include an increased rate of transcription with a concomitant increase in reverse transcription, increased production of the primer tRNA fragment, a hyperactive reverse transcriptase, and altered read-through of the stop codon at the *gag/pol* boundary.

Overall, pCal presents itself as a highly unusual retrotransposon. While having many of the features conserved among retrotransposons, it has a number of features which set it apart from other elements of its class. For instance, the translation of the *pol* ORF seems to be dependent upon the pseudoknot-assisted read-through of a UGA stop codon. This is similar to the mechanism used by mammalian type C retroviruses but has not been previously reported in retrotransposons. A phylogenetic analysis of the reverse transcriptase sequences of a number of LTR retroelements showed that while pCal lies within the Ty1/*copia* group of retrotransposons, it is one of the most divergent elements within this group. The most distinctive feature of pCal, however, is that it exists at a high copy number as a free, linear, dsDNA molecule. The reasons behind this and its implications are as yet unclear.

REFERENCES

1. **Atwood, A., J.-H. Lin, and H. L. Levin.** 1996. The retrotransposon Tf1 assembles virus-like particles that contain excess GAG relative to integrase because of a regulated degradation process. Mol. Cell. Biol. **16:**338–346.
2. **Belcourt, M. F., and P. J. Farabaugh.** 1990. Ribosomal frameshifting in the yeast retrotransposon Ty: tRNAs induce slippage on a 7 nucleotide minimal site. Cell **62:**339–352.
3. **Boeke, J. D.** 1989. Transposable elements in *Saccharomyces cerevisiae*, p. 335–374. *In* D. E. Berg and M. M. Howe (ed.), Mobile DNA. American Society for Microbiology, Washington, D.C.
4. **Boeke, J. D., D. J. Garfinkel, C. A. Styles, and G. R. Fink.** 1985. Ty elements transpose through an RNA intermediate. Cell **40:**491–500.
5. **Brookman, J. L., A. J. Stott, P. J. Cheeseman, N. R. Burns, S. E. Adams, A. J. Kingsman, and K. Gull.** 1995. An immunological analysis of Ty1 virus-like particle structure. Virology **207:**59–67.

6. **Butler, M. I.** Unpublished results.
7. **Camirand, A., and N. Brisson.** 1990. The complete nucleotide sequence of the Tst1 retrotransposon of potato. Nucleic Acids Res. **18:**4929.
8. **Chalker, D. L., and S. B. Sandmeyer.** 1990. Transfer RNA genes are genomic targets for de novo transposition of the yeast retrotransposon Ty3. Genetics **126:**837–850.
9. **Chalker, D. L., and S. B. Sandmeyer.** 1992. Ty3 integrates within the region of RNA polymerase III transcription initiation. Genes Dev. **6:**117–128.
10. **Charneau, P., M. Alizon, and F. Clavel.** 1992. A second origin of DNA plus-strand synthesis is required for optimal human immunodeficiency virus replication. J. Virol. **66:**2814–2820.
11. **Chen, J.-Y., and W. A. Fonzi.** 1992. A temperature-regulated, retrotransposon-like element from *Candida albicans*. J. Bacteriol. **174:**5624–5632.
12. **Clare, J., and P. Farabaugh.** 1985. Nucleotide sequence of a yeast Ty element: evidence for an unusual mechanism of gene expression. Proc. Natl. Acad. Sci. USA **82:**2829–2833.
13. **Clark, D. J., V. W. Bilanchone, L. J. Haywood, S. L. Dildine, and S. B. Sandmeyer.** 1988. A yeast sigma composite element, Ty3, has properties of a retrotransposon. J. Biol. Chem. **263:**1413–1423.
14. **Clark-Walker, G. D., and G. L. G. Miklos.** 1974. Localization and quantification of circular DNA in yeast. Eur. J. Biochem. **41:**359–365.
15. **Craigie, R.** 1992. Hotspots and warm spots: integration specificity of retroelements. Trends Genet. **8:**187–190.
16. **Cryer, D. R., R. Eccleshall, and J. Marmur.** 1975. Isolation of yeast DNA. Methods Cell Biol. **12:**39–44.
17. **Curcio, M. J., A.-M. Hedge, J. D. Boeke, and D. J. Garfinkel.** 1990. Ty RNA levels determine the spectrum of retrotransposition events that activate gene expression in *Saccharomyces cerevisiae*. Mol. Gen. Genet. **220:**213–221.
18. **Devereux, J., P. Haeberli, and O. Smithies.** 1984. A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. **12:**387–395.
19. **Devine, S. E., and J. D. Boeke.** 1996. Integration of the yeast retrotransposon Ty1 is targeted to regions upstream of genes transcribed by RNA polymerase III. Genes Dev. **10:**620–633.
20. **Eichinger, D. J., and J. D. Boeke.** 1990. A specific terminal structure is required for Ty1 transposition. Genes Dev. **4:**324–330.
21. **Eichinger, D. J., and J. D. Boeke.** 1988. The DNA intermediate in yeast Ty1 element transposition copurifies with virus-like particles: cell-free Ty1 transposition. Cell **54:**955–966.
22. **Farabaugh, P. J., H. Zhao, and A. Vimaladithan.** 1993. A novel programmed frameshift expresses the *POL3* gene of retrotransposon Ty3 of yeast: frameshifting without tRNA slippage. Cell **74:**93–103.
23. **Felder, H., A. Herzceg, Y. de Chastonay, P. Aeby, H. Tobler, and F. Muller.** 1994. *Tas*, a retrotransposon from the parasitic nematode *Ascaris lumbricoides*. Gene **149:**219–225.
24. **Felsenstein, J.** 1989. PHYLIP—Phylogeny Inference Package (version 3.2). Cladistics **5:**164–166.
25. **Feng, Y.-X., H. Yuan, A. Rein, and J. G. Levin.** 1992. Bipartite signal for read-through suppression in murine leukemia virus mRNA: an eight-nucleotide purine-rich sequence immediately downstream of the *gag* termination codon followed by an RNA pseudoknot. J. Virol. **66:**5127–5132.
26. **Fourcade-Peronnet, F., L. d'Auriol, J. Becker, F. Galibert, and M. Best-Belpomme.** 1988. Primary structure and functional organization of *Drosophila* 1731 retrotransposon. Nucleic Acids Res. **16:**6113–6125.
27. **Gabriel, A., M. Willems, E. H. Mules, and J. D. Boeke.** 1996. Replication infidelity during a single cycle of Ty*1* retrotransposition. Proc. Natl. Acad. Sci. USA **93:**7767–7771.
28. **Goodwin, T. J. D.** Unpublished results.
29. **Grandbastien, M.-A., A. Spielman, and M. Caboche.** 1989. Tnt1, a mobile retroviral-like transposable element of tobacco isolated by plant cell genetics. Nature **337:**376–380.
30. **Hansen, L. J., D. L. Chalker, and S. B. Sandmeyer.** 1988. Ty3, a yeast retrotransposon associated with tRNA genes, has homology to animal retroviruses. Mol. Cell. Biol. **8:**5245–5256.
31. **Heyman, T., B. Agoutin, S. Friant, F. X. Wilhelm, and M. L. Wilhelm.** 1995. Plus-strand DNA synthesis of the yeast retrotransposon Ty1 is initiated at two sites, PPT1 next to the 3′ LTR and PPT2 within the *pol* gene. PPT1 is sufficient for Ty1 transposition. J. Mol. Biol. **253:**291–303.
32. **Jacks, T., H. D. Madhani, F. R. Masiarz, and H. E. Varmus.** 1988. Signals for ribosomal frameshifting in the Rous sarcoma virus *gag-pol* region. Cell **55:**447–458.
33. **Jacks, T., M. D. Power, F. R. Masiarz, P. A. Luciw, P. J. Barr, and H. E. Varmus.** 1988. Characterization of ribosomal frameshifting in HIV-1 *gag-pol* expression. Nature **331:**280–283.
34. **Ji, H., D. P. Moore, M. A. Blomberg, L. T. Braiterman, D. F. Voytas, G. Natsoulis, and J. D. Boeke.** 1993. Hotspots for unselected Ty1 transposition events on yeast chromosome III are near tRNA genes and LTR sequences. Cell **73:**1007–1018.
35. **Katz, R. A., and A. M. Skalka.** 1994. The retroviral enzymes. Annu. Rev. Biochem. **63:**133–173.
36. **Kawakami, K., S. Pande, B. Faiola, D. P. Moore, J. D. Boeke, P. J. Farabaugh, J. N. Strathen, Y. Nakamura, and D. J. Garfinkel.** 1993. A rare tRNA-Arg(CCU) that regulates Ty*1* element ribosomal frameshifting is essential for Ty*1* retrotransposition in *Saccharomyces cerevisiae*. Genetics **135:**309–320.
37. **Keith, G., and G. Dirheimer.** 1980. Reinvestigation of the primary structure of brewer's yeast $tRNA_3^{Arg}$. Biochem. Biophys. Res. Commun. **92:**116–119.
38. **Kikuchi, Y., Y. Ando, and T. Shiba.** 1986. Unusual priming mechanism of RNA-directed DNA synthesis in *copia* retrovirus-like particles of *Drosophila*. Nature **323:**824–826.
39. **Kikuchi, Y., N. Sasaki, and Y. Ando-Yamagami.** 1990. Cleavage of tRNA within the mature tRNA sequence by the catalytic RNA of RNase P: implication for the formation of the primer tRNA fragment for reverse transcription in *copia* retrovirus-like particles. Proc. Natl. Acad. Sci. USA **87:**8105–8109.
40. **Kikuchi, Y., and N. Sasaki.** 1992. Hyperprocessing of tRNA by the catalytic RNA of RNase P. J. Biol. Chem. **267:**11972–11976.
41. **Leuker, C. E., and J. F. Ernst.** 1994. Toxicity of a heterologous leucyl-tRNA (anticodon CAG) in the pathogen *Candida albicans*: in vivo evidence for non-standard decoding of CUG codons. Mol. Gen. Genet. **245:**212–217.
42. **Levin, H. L., D. C. Weaver, and J. D. Boeke.** 1990. Two related families of retrotransposons from *Schizosaccharomyces pombe*. Mol. Cell. Biol. **10:**6791–6798.
43. **Levin, H. L., D. C. Weaver, and J. D. Boeke.** 1993. Novel gene expression mechanism in a fission yeast retroelement: Tf1 proteins are derived from a single primary translation product. EMBO J. **12:**4885–4895.
44. **Marlor, R. L., S. M. Parkhurst, and V. G. Corces.** 1986. The *Drosophila melanogaster* gypsy transposable element encodes putative gene products homologous to retroviral proteins. Mol. Cell. Biol. **6:**1129–1134.
45. **McHale, M. T., I. N. Roberts, S. M. Noble, C. Beaumont, M. P. Whitehead, D. Seth, and R. P. Oliver.** 1992. CfT-I: an LTR-retrotransposon in *Cladosporium fulvum*, a fungal pathogen of tomato. Mol. Gen. Genet. **233:**337–347.
46. **Moore, S. P., M. Powers, and D. J. Garfinkel.** 1995. Substrate specificity of Ty1 integrase. J. Virol. **69:**4683–4692.
47. **Mount, S. M., and G. M. Rubin.** 1985. Complete nucleotide sequence of the *Drosophila* transposable element copia: homology between copia and retroviral proteins. Mol. Cell. Biol. **5:**1630–1638.
48. **Noegel, A., D. L. Welker, B. A. Metz, and K. L. Williams.** 1985. Presence of nuclear associated plasmids in the lower eukaryote *Dictyostelium discoideum*. J. Mol. Biol. **185:**447–450.
49. **Odds, F. C.** 1988. Candida and candidosis. A review and bibliography. Bailliere Tindal, London, United Kingdom.
50. **Ohama, T., T. Suzuki, M. Mori, S. Osawa, T. Ueda, K. Watanabe, and T. Nakase.** 1993. Non-universal decoding of the leucine codon CUG in several *Candida* species. Nucleic Acids Res. **21:**4039–4045.
51. **Osawa, S.** 1995. Evolution of the genetic code. Oxford University Press, Oxford, United Kingdom.
52. **Pochart, P., B. Agoutin, S. Rousset, R. Chanet, V. Doroszkiewicz, and T. Heyman.** 1993. Biochemical and electron microscope analyses of the DNA reverse transcripts present in the virus-like particles of the yeast transposon Ty1. Identification of a second origin of Ty1DNA plus strand synthesis. Nucleic Acids Res. **21:**3513–3520.
53. **Poulter, R., K. Jeffrey, M. J. Hubbard, M. G. Shepherd, and P. A. Sullivan.** 1981. Parasexual genetic analysis of *Candida albicans* by spheroplast fusion. J. Bacteriol. **146:**833–840.
54. **Roeder, G. S., and G. R. Fink.** 1980. DNA rearrangements associated with a transposable element in yeast. Cell **21:**239–249.
55. **Rothnie, H. M., K. J. McCurrach, L. A. Glover, and N. Hardman.** 1991. Retrotransposon-like nature of Tp1 elements: implications for the organisation of highly repetitive, hypermethylated DNA in the genome of *Physarum polycephalum*. Nucleic Acids Res. **19:**279–286.
56. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
57. **Santos, M., D. R. Colthurst, N. Wills, C. S. McLaughlin, and M. F. Tuite.** 1990. Efficient translation of the UAG termination codon in *Candida* species. Curr. Genet. **17:**487–491.
58. **Santos, M. A. S., G. Keith, and M. F. Tuite.** 1993. Non-standard translational events in *Candida albicans* mediated by an unusual seryl-tRNA with a 5′-CAG-3′ (leucine) anticodon. EMBO J. **12:**607–616.
59. **Santos, M. A. S., and M. F. Tuite.** 1995. The CUG codon is decoded *in vivo* as serine and not leucine in *Candida albicans*. Nucleic Acids Res. **23:**1481–1486.
60. **Schwartz, D. E., R. Tizard, and W. Gilbert.** 1983. Nucleotide sequence of Rous sarcoma virus. Cell **32:**853–869.
61. **Shinnick, T. M., R. L. Lerner, and G. Sutcliffe.** 1981. Nucleotide sequence of Moloney murine leukaemia virus. Nature **293:**543–548.
62. **Smyth, D. R., P. Kalitsis, J. L. Joseph, and J. W. Sentry.** 1989. Plant retrotransposon from *Lilium henryi* is related to *Ty3* of yeast and the gypsy group of *Drosophila*. Proc. Natl. Acad. Sci. USA **86:**5015–5019.
63. **Stockwell, P. A.** 1985. VTUTIN: a full screen gel management editor. Comp. Appl. Biosci. **1:**253–259.
64. **Stockwell, P. A., and G. B. Petersen.** 1987. HOMED: a homologous sequence editor. Comp. Appl. Biosci. **3:**37–43.

65. **Sugiyama, H., M. Ohkuma, Y. Masuda, S.-M. Park, A. Ohta, and M. Takagi.** 1995. *In vivo* evidence for non-universal usage of the codon CUG in *Candida maltosa*. Yeast **11:**43–52.
66. **Suzuki, T., T. Ueda, and K. Watanabe.** 1997. The 'polysemous' codon—a codon with multiple amino acid assignment caused by dual specificity of tRNA identity. EMBO J. **16:**1122–1134.
67. **ten Dam, E., K. Pleij, and D. Draper.** 1992. Structural and functional aspects of RNA pseudoknots. Biochemistry **31:**11665–11676.
68. **Toh-E, A., S. Tada, and Y. Oshima.** 1982. 2-μm DNA-like plasmids in the osmophilic haploid yeast *Saccharomyces rouxii*. J. Bacteriol. **151:**1380–1390.
69. **Tuite, M. F., P. A. Bower, and C. S. McLaughlin.** 1986. A novel suppressor tRNA from the dimorphic fungus *Candida albicans*. Biochim. Biophys. Acta **866:**26–31.
70. **Voytas, D. F., and F. M. Ausubel.** 1988. A copia-like transposable element family in *Arabidopsis thaliana*. Nature **336:**242–244.
71. **Wain-Hobson, S., P. Sonigo, O. Danos, S. Cole, and M. Alizon.** 1985. Nucleotide sequence of the AIDS virus, LAV. Cell **40:**9–17.
72. **White, T. C., L. E. Andrews, D. Maltby, and N. Agabian.** 1995. The "universal" leucine codon CTG in the secreted aspartyl proteinase 1 (*SAP1*) gene of *Candida albicans* encodes a serine in vivo. J. Bacteriol. **177:**2953–2955.
73. **Wilhelm, M., F. X. Wilhelm, G. Keith, B. Agoutin, and T. Heyman.** 1994. Yeast Ty1 retrotransposon: the minus-strand primer binding site and a *cis*-acting domain of the Ty1 RNA are both important for packaging of primer tRNA inside virus-like particles. Nucleic Acids Res. **22:**4560–4565.
74. **Xiong, Y., W. D. Burke, and T. H. Eickbush.** 1993. *Pao*, a highly divergent retrotransposable element from *Bombyx mori* containing long terminal repeats with tandem copies of the putative R region. Nucleic Acids Res. **21:**2117–2123.
75. **Xiong, Y., and T. H. Eickbush.** 1990. Origin and evolution of retroelements based upon their reverse transcriptase sequences. EMBO J. **9:**3353–3362.
76. **Yokogawa, T., T. Suzuki, T. Ueda, M. Mori, T. Ohama, Y. Kuchino, S. Yoshinara, I. Motoki, K. Nishikawa, S. Osawa, and K. Watanabe.** 1992. Serine tRNA complementary to the nonuniversal serine codon CUG in *Candida cylindracea*: evolutionary implications. Proc. Natl. Acad. Sci. USA **89:**7408–7411.
77. **Yoshinaka, Y., I. Katoh, T. D. Copeland, and S. Oroszlan.** 1985. Murine leukemia virus protease is encoded by the *gag-pol* gene and is synthesized through suppression of an amber termination codon. Proc. Natl. Acad. Sci. USA **82:**1618–1622.
78. **Yoshioka, K., H. Honma, M. Zushi, S. Kondo, S. Togashi, T. Miyake, and T. Shiba.** 1990. Virus-like particle formation of *Drosophila copia* through autocatalytic processing. EMBO J. **9:**535–541.