Nucleotide sequence of a yeast Ty element: Evidence for an unusual mechanism of gene expression

(DNA sequencing/protein homology/S1 nuclease mapping/β-galactosidase fusion/translational frameshifting)

JEFFREY CLARE AND PHILIP FARABAUGH

Department of Microbiology, University of Connecticut Health Center, Farmington, CT 06032

Communicated by Gerald R. Fink, December 10, 1984

ABSTRACT We have determined the DNA sequence of the transposable element Ty912 of yeast. The 5918-base-pair element encodes two genes, tya912 and tyb912, which specify proteins similar to sequence-specific DNA-binding proteins of Escherichia coli and retroviral reverse transcriptases, respectively. The tya912 gene is atypical of eukaryotic genes since (i) it begins 1336 nucleotides into the Ty912 mRNA (i.e., downstream of the tya912 gene) and (ii) the first in-frame AUG is 921 nucleotides into the coding frame. Protein blot analysis of Ty--lacZ fusions shows that the tya912 gene is translated starting at the 5' end of the tya912 gene and that the primary translational product is a tya912::tyb912 fusion protein. We have shown that synthesis of this fusion protein probably does not occur by RNA splicing. The data are consistent with a mechanism of translational frameshifting occurring within the region of overlap between the 3' end of tya912 and the 5' end of tyb912.

Ty elements are a family of approximately 30 transposable elements, which are dispersed throughout the yeast genome (1–4). Details of the structure of Ty elements suggest that they are more closely related to the retroviruses of higher cells than to bacterial transposons. They consist of a central region (termed epsilon) approximately 5300 base pairs (bp) in length surrounded by 330-bp delta elements. Elder et al (5) showed that the major Ty-encoded mRNA has a terminal repeat similar to the R region of retroviral provirus transcripts. At the 5' epsilon boundary of several Ty elements there is a region complementary to the 3' end of the initiator methionine tRNA of yeast (4), a structure analogous to the negative strand primer binding site of retroviruses. Varmus (6) has pointed out that Ty elements also include an oligopurine tract adjacent to the 3' delta that is similar to the binding site for the retrovirus primer of positive strand synthesis. These factors, all of which are required for the retroviral life cycle, imply that Ty elements are related to metazoan retroviruses. In this report we present further evidence to support this idea, both from the complete DNA sequence of a Ty/l element and from studies of the unusual organization and expression of its gene products.

MATERIALS AND METHODS

DNA Manipulations. To construct the tyb912::lacZ fusion plasmid, pPF350, a 2.64-kilobase-pair (kb) Sal I/Bgl II fragment containing the 5' end of Ty912 was ligated into Xho I/BamHI-digested pLGS5, a high copy number lacZ fusion vector based on the 2-μm element (7). This procedure gives an in-frame fusion of lacZ to the Bgl II site of tyb912 at nucleotide (nt) 1707. Plasmid pJC8 was made by opening pPF350 at the unique Sph I site at nt 1218 and removing the cohesive ends with T4 DNA polymerase, followed by ligating the blunt ends. Plasmids used for sequencing by the Maxam and Gilbert method (8) were Sal I and Pst I subclones of YIp312, previously described (3).

S1 Nuclease Analysis. To map Ty transcripts with S1 nuclease, total RNA was prepared as described (5) from a pPF350 yeast transformant and enriched for poly(A)+ RNA by using oligo(dT)-cellulose chromatography. The DNA-RNA hybrids protected from S1 nuclease digestion were separated on neutral and alkaline agarose gels (9) and visualized by Southern hybridization. Under these conditions a splice that removes as little as 11 nt should be detected (9). The lacZ-specific Southern blot probe used was a 2.7-kb Bgl II fragment isolated from a Ylp312 subclone. This fragment contains the complete 5' region of the element up to the Bgl II site at nt 1702.

Isolation of tyb912::lacZ Fusion Proteins. Fusion proteins were isolated from a yeast transformant of pPF350 by affinity chromatography using p-aminobenzyl-thio-β-galactosamine-agarose (10). Proteins retained by the column were analyzed by electrophoresis in 9% NaDodSO4 polyacrylamide gels and blotting (11), using monoclonal mouse IgG anti-β-galactosidase (a gift from J. Partaledis). Bands were visualized using goat anti-mouse IgG antibody conjugated with horseradish peroxidase and Bio-Rad HRP color development reagent.

β-Galactosidase Assays. Protein extracts were assayed for β-galactosidase activity as described by Miller (12). β-Galactosidase activities of growing yeast cultures were determined by the rapid chloroform/NaDodSO4 permeabilization procedure (7).

RESULTS

The Ty912 Element Includes Two Open Reading Frames. We have sequenced the Ty912 element, which inserted into the promoter of the his4 gene, causing the his4-912 mutation (3, 13). The element is 5918 bp in length and consists of a 5250-bp region flanked by a 334-bp perfect direct repeat called delta912. The sequence in Fig. 1 is the coding strand of Ty912. Inspection of the Ty912 sequence reveals two open reading frames covering 89% of the element (5269 bp). The 5'-proximal frame, tya912, which begins at an ATG codon at nt 294 and continues to a TGA codon at nt 1614, could encode a protein of 440 amino acids with a predicted molecular weight of 48,951. The AUG nearest the 5' end of the Ty mRNA is the first AUG in ty912, so we would expect the gene to be translated efficiently in vivo. Furthermore, the context around this AUG, A-C-A-A-U-G-G-G, favors initiation (16). A second frame, tyb912 overlaps the last 38 bp of tya912 and extends from an ACAA codon at nt 1576 to a TAG codon at nt 5560. The first ATG codon in tyb912 is 921 bp into the gene at nt 2497. It is likely, given its length, that the

Abbreviations: bp, base pair(s); kb, kilobase pair(s); nt, nucleotide(s); RSV, Rous sarcoma virus.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.
FIG. 1. (Figure continues on the next page.)
region upstream of the AUG is translated, suggesting that the AUG is not the site of initiation of tyb912. From inspection of the DNA sequence we could not assign a site for initiation. The tyb912 open reading frame could encode a protein of as much as 151,200 daltons. There are no other reading frames longer than 250 bp in either strand of the element.

The tyb912 Protein Includes Homology to DNA Binding Proteins. The tya912 protein includes two regions of unusual primary sequence. A region of 125 amino acids, beginning at amino acid 58, is 21.6% proline. The carboxyl-terminal region of 90 amino acids, composed of 20.2% basic residues, is reminiscent of a DNA-binding protein. Upstream of this basic region, from residue 313 to residue 327 (Fig. 1), is a sequence similar to the consensus for DNA-binding proteins from bacteria (14). The homology is found within a region predicted by the Chou and Fasman method (17) to include two potential α-helical regions separated by a β-turn, with the homology occurring at the correct position with respect to these putative secondary structural features. Of the 10 amino acids that have been shown or predicted to interact with DNA (14, 18, 19) 9 occur within the putative two α-helices of the tya912 protein (Fig. 1). The primary and secondary structural features suggest that the protein may be a DNA-binding protein.

The tyb912 Protein Shows Homology to Retroviral Reverse Transcriptase. Given the similarity between Ty elements and retroviruses, we hypothesized that the tyb912 protein might be homologous to retroviral reverse transcriptase. Toh et al. (20) and Patarca and Haselstein (15) have found scattered primary sequence similarities among both retroviral reverse transcriptases and DNA polymerases of double-stranded DNA viruses. We find that the region from amino acid 835 to amino acid 1046 of tyb912 includes sequences similar to four homology regions referred to as R2, I, II, and III (15) and two additional regions (M3 and M4; ref. 15) from amino acid 306 to amino acid 359 that we find are similar to reverse transcriptases but not DNA polymerases (see Fig. 1). This suggests that the product of the tyb912 gene may be a reverse transcriptase.

An Unusual Mode of Expression for the tyb912 Gene. The tyb912 gene is unusual in two respects: first, the 5' end of the tyb912 open reading frame overlaps the 3' end of the tya912 open reading frame for a distance of 38 bp; second, the first in-frame ATG codon occurs 921 bp from the beginning of the open reading frame. It is doubtful that this is the initiator codon for the tyb912 gene since it is known that eukaryotes normally initiate translation within 40–80 nt of the 5' end of the transcript (16).

To examine the expression of the tyb912 gene we constructed a gene fusion with the lacZ gene of Escherichia coli. In this construction, pPF350, the lacZ gene is fused to the tyb912 coding frame at a Bgl II site (nt 1702), upstream of the initial ATG codon but downstream from the 3' end of the tya912 gene (Fig. 1). When pPF350 is used to transform yeast, significant levels of β-galactosidase (1004 units) are detected, showing that the tyb912 coding frame is expressed and that translation of the gene product must initiate upstream of the first in-frame ATG codon. We have sequenced the fusion junction of pPF350 to confirm that lacZ expression is not a cloning artifact.

The simplest explanation for the expression of the tyb912 coding frame would be that the gene is spliced to an upstream initiator, most likely that used by the tya912 gene. To investigate this possibility S1 nuclease mapping was performed, using the tyb912: lacZ gene fusion of pPF350 (Fig. 2). DNA-RNA hybrids protected by S1 nuclease degradation were analyzed in neutral and alkaline agarose gels followed by Southern blotting. A lacZ-specific Southern probe reveals a single band in both gels of the size expected for an unspliced transcript. When a Southern probe specific for the 5' end of Ty912 was used, this band, plus another, was observed in both gels. The second band is due to protection of the S1 nuclease probe by wild-type Ty mRNA. No other protected species were observed. S1 nuclease analysis was also performed by using a 5'-end-labeled restriction fragment extending from the Sph I site in the tya912 gene to the Pvu II site in lacZ (data not shown). The sequence ladder of this portion of DNA was used as size markers for the protected hybrids. No evidence of spliced RNA was observed.

It appears from these experiments that splicing of the tya912 and tyb912 genes does not occur. Nevertheless, we have shown that tyb912 expression is dependent upon translation of the tya912 gene by constructing a frameshift within the open reading frame. This alteration in the tya912 reading frame of pPF350 introduces an in-frame TAA stop codon 4 bp downstream of the Sph I site at nt 1218 (plasmid pJCB). Virtually no β-galactosidase activity (<1 unit) is observed when this plasmid is present in yeast. Thus tyb912 gene expression is eliminated by early termination of the tya912 gene and cannot be due to translation initiating upstream of the Bgl II site within the tyb912 open reading frame itself (e.g., at a codon other than ATG).
The nucleotide sequence presented here reveals the unusual genetic organization of the yeast transposon Ty912. We have identified two overlapping reading frames, tya912 and tyb912, a structure similar to that found in the retrovirus Rous sarcoma virus (RSV) (21). The analogy with metazoan retroviruses is further strengthened by our demonstration of a tya::tyb912 fusion protein, which is reminiscent of the retroviral gag–pol proteins. The predicted tya912 protein exhibits several structural features characteristic of DNA-bind-
ing proteins. Examination of the tyb912 sequence reveals scattered homology to several reverse transcriptases and DNA polymerases. The degree of similarity found is comparable to that displayed by any of these enzymes when compared to the rest of the group, suggesting that this observation is significant. Thus sequence analysis of the Ty912-encoded proteins suggest a possible role in the transposition of the element itself.

We have also demonstrated an unusual mode of expression for the tyb912 gene. We have shown that the product of this gene is a tya912:tyb912 fusion protein, yet we observed no evidence of spliced RNA. It is possible that the protein results from the excision of an intron too short to be detected by S1 endonuclease. However, we consider this unlikely due to the high sensitivity of the technique (9) and because of the absence of sites known to be required for splicing of yeast nuclear mRNA (i.e., donor, acceptor, and T-A-C-T-A-A-C sites). Furthermore, the efficiency of mRNA splicing in yeast appears to be greatly reduced with a diminishing size of intron (22). Our alternative explanation is that the ty912:tyb912 fusion protein is produced by translational frameshifting within the region of overlap between the two genes.

Several other examples of translational frameshifting are known. The lysis genes of the RNA phages f2 and MS2 (23, 24) are expressed by frameshifting within the corresponding upstream coat gene. Spontaneous readthrough of leaky frameshifting mutations occurs within the octl gene of yeast mitochondria (25). The addition of a carboxyl-terminal extension onto the phage T7 major coat protein also occurs by a frameshift during translation (26). The level of frameshifting observed in each case is due to translational misincorporation (27).

Another system in which translational frameshifting allows the expression of alternative proteins is the gag–pol genes of RSV (21). The primary translation product of pol is gag–pol fusion protein. The expression of the fusion is usually attributed to a splicing event, although there is no direct proof of this (21). The RSV genome includes a 58-bp overlap between the gag and pol genes similar to that found in Ty912. Thus it is also possible that expression of the pol gene occurs by frameshifting within the overlap, as was proposed by Schwartz et al. (28). A model of translational misincorporation may also account for the expression of the pol gene of Moloney murine leukemia virus. Here the gag and pol genes are in the same frame but separated by a single UAG nonsense codon. Translation of the latter gene may occur by readthrough of this codon. It is possible that mechanisms involving translational infidelity have a general role in the expression of proteins required in low amounts.

We have also sequenced a Ty2 element, Ty917, and found that it too contains overlapping reading frames (unpublished data). When this sequence is compared to that of Ty912 the homology found in the overlap is no greater than that of the adjacent 800-bp region. We therefore believe that the primary structure of the overlap region is not important in promoting frameshifting. Rather, our model involves translational misincorporation at the UGA terminator codon of the tya912 gene. Weiss (29) has described a general mechanism to explain how certain normal tRNA molecules are able to promote ribosomal frameshifting. This is based upon studies that identify these so-called “shifty” tRNAs, and their corresponding “shifty” codons (30, 31). The conserved U33 nucleotide immediately 5′ to the normal anticodon in these tRNAs is proposed to base pair with the “shifty” codon. By applying the “reciprocating ratchet” model, in which translational transition is caused by an allosteric transition of the anticodon loop (32), Weiss has predicted that this offset anticodon–codon pairing would cause a −1 or +1 frameshift. The direction of the frameshift produced is determined by which of two possible allosteric conformations is energetically more favorable, and this in turn is dependent on the particular tRNA and codon involved.

It is possible that tyb912 gene expression occurs by such a mechanism. Binding of a leucyl-, methionyl-, or valyl-tRNA, each of which contains an offset 3′-ACU-5′ anticodon, to the 5′-UGA-3′ terminator codon of the tya912 gene could occur. This interaction would presumably be in competition with the binding of termination factor to the ribosome such that translation of tyb912 is dependent on the relative binding rates of tRNA and factor. Further experiments are necessary to define the actual mechanism of tyb912 gene expression.

We thank Lee Scionti and Celeste Sticca for their capable technical assistance and Janice Seagren for typing the manuscript. This work was supported by Grant GM 29480 from the Public Health Service and Grant 6-30151 from the American Heart Association.