MOLECULAR AND CELLULAR BIOLOGY, June 1988, p. 2572–2580
0270-7306/88/062572-09$02.00/0

Vol. 8, No. 6

Transposable Element-Mediated Enhancement of Gene Expression in *Saccharomyces cerevisiae* Involves Sequence-Specific Binding of a *trans*-Acting Factor

ANURAG GOEL AND RONALD E. PEARLMAN*

Department of Biology, York University, 4700 Keele Street, Toronto, Ontario, Canada M3J 1P3

Received 21 January 1988/Accepted 22 March 1988

In our studies on the regulation of adjacent-gene expression by Ty sequences, we demonstrated that a single-base-pair change (T-A→C-G) in the epsilon sequence of Ty917-derived elements is primarily responsible for enhancement of β-galactosidase expression from *lacZ* fusion plasmids. Using an electrophoretic gel mobility assay, we showed that the same base pair transition is required for binding of a *trans*-acting factor, TyBF, from crude cell extracts in vitro. We identified the site of TyBF binding and determined the guanine nucleotide contact sites required for TyBF interaction. We propose that TyBF binding to *cis*-acting Ty2 sequences activates adjacent-gene transcription.

Saccharomyces cerevisiae transposable (Ty) elements make up a family of genetically mobile, repetitive DNA sequences. Each element consists of an approximately 5.3-kilobase-pair internal region called epsilon flanked by 330-base-pair (bp) direct repeats called delta. Although these elements cross-hybridize because of various degrees of sequence homology, they differ from one another by point mutations, insertions, deletions, and substitutions (for a review, see reference 31). Ty elements can be divided into two main classes: Ty1 and Ty2 elements (23). Each haploid genome of *S. cerevisiae* harbors about 30 Ty1 elements and approximately 5 elements of the Ty2 class.

Ty elements share structural similarities with copia elements of *Drosophila melanogaster* (33) and the proviral form of retroviruses (38). Boeke et al. (4) have shown that Ty elements transpose via a retroviruslike mechanism, i.e., through an RNA intermediate. Sequence analysis of Ty1 elements (5, 25) provides evidence that they encode their own reverse transcriptase. One of the most interesting properties of Ty elements, which is shared with some other transposable elements, including certain retroviruses, is their ability to regulate the expression of genes adjacent to an insertion site. The mechanism of adjacent-gene regulation remains to be elucidated and is the focus of this paper.

Ty elements often transpose into the 5′ regulatory sequences of genes (9). Insertion of a Ty into a 5′ upstream sequence often results in overproduction of the product of the adjacent gene and renders expression of that gene subject to control by genes at the mating type (*MAT*) locus. In a class of mutants originally designated ROAM (regulated overproducing alleles responding to mating type), both expression of the adjacent gene and Ty transcription are governed by the *MAT* locus of the cell (11). In haploid *MAT* (**a** or α) cells or diploids homozygous for *MAT* (**a**/**a** or α/α), the adjacent gene is expressed at an approximately 5- to 20-fold greater level than in diploids heterozygous for *MAT* (**a**/α). In all ROAM mutants, Ty elements are oriented such that their transcription is divergent from that of the adjacent gene. Ty transcripts initiate in the delta sequence proximal to the adjacent gene and terminate in the distal delta sequence (10). The size of the adjacent-gene transcript remains unaltered in the insertion mutant, suggesting that Ty insertions do not affect the transcription initiation site. Insertion of a Ty element does not necessarily lead to an overproducing phenotype. Ty insertions at some loci result in auxotrophic mutants (29, 36).

The *his4-917* mutation in *S. cerevisiae* is the result of the insertion of a Ty2 element, Ty917, 9 bp upstream of the mRNA start site of the *HIS4* gene (8, 29). This mutation renders the cells histidine requiring and therefore differs from the mutations described above. The element is oriented such that *HIS4* and Ty are divergently transcribed. Several Ty917 derivatives, which arose by gene conversion between a genetically marked Ty917 at *HIS4* and other chromosomal Ty elements, confer a variety of His phenotypes: His^-, weak His^+, strong His^+, and His^+ under *MAT* control (30). These derivatives differ only by the sequence of the Ty element; the orientation and position of the elements remain unaltered, as do flanking *HIS4* sequences. Silverman and Fink (35) have shown by Northern (RNA) blot analysis that the Ty insertions do not affect the size of the *HIS4* transcript. We have previously demonstrated that sequences in a 730-bp region of Ty proximal to *HIS4* are responsible for generating the His^+ phenotypes (32). Sequence analysis of this region of a weak His^+ derivative, Ty917(467), revealed only a single-base-pair change, T-A→C-G, in the epsilon sequence compared with the sequence of the Ty917 element. This transition occurs 612 bp from the 5′ end of the Ty element. Another derivative, Ty917(480), displays a strong His^+ phenotype and possesses the same T-A→C-G transition. There are 10 additional base pair changes in the delta sequence of Ty917(480) not present in either Ty917 or Ty917(467). Here we present additional evidence that the single-base-pair change found at position +612 in both Ty917(467) and Ty917(480) is necessary for activation of adjacent-gene expression. Furthermore, we propose that this regulation is mediated through the sequence-specific interaction of a *trans*-acting factor, which we designate as Ty-binding factor or TyBF.

MATERIALS AND METHODS

Plasmids. Standard procedures described by Maniatis et al. (24) were used for plasmid constructions. The cloning strategy for the various constructions is described in Fig. 1A

* Corresponding author.

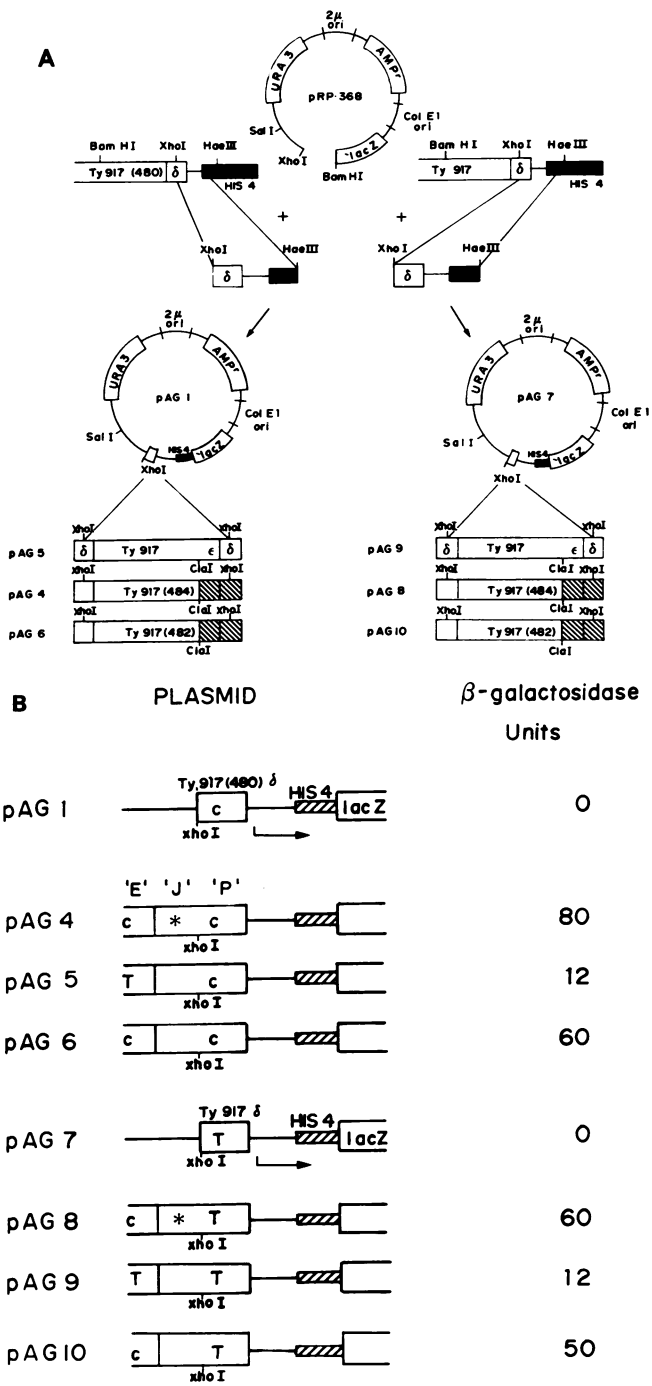

FIG. 1. Effect of Ty sequences on β-galactosidase expression. (A) Construction of *HIS4-lacZ* fusion plasmids. The Ty-*HIS4* junctions contained on *Xho*I-*Hae*III fragments from Ty917 and Ty917(480) were cloned into the pRP368 vector to construct pAG1 and pAG7. This created an in-frame fusion of the first 36 bp of *HIS4* coding sequences (▬) and the *E. coli lacZ* gene. pAG7 differs from pAG1 by a single-base-pair change (C-G→T-A) in the Ty delta sequence. Ty elements were cloned as *Xho*I fragments into the unique *Xho*I site of these plasmids. Ty917(484) and Ty917(482) are hybrid Ty elements reconstructed in vitro (32) that contain the 730-bp regulatory region (▨) from Ty917(480) and Ty917(467), respectively (see Materials and Methods). ori, Origin of replication. (B) β-Galactosidase activity of *S. cerevisiae* transformed with *lacZ* fusion plasmids. Yeast strain YeRP7 (*MAT***a** *his4-912 ura3-52*) was transformed with each plasmid, and β-galactosidase activity was measured. The numbers represent units of β-galactosidase activity, normalized to total cellular protein, from each transformed yeast strain. These values are averages of at least four independent enzyme assays with at least two independent transformants. The standard error was less than 15%. Other yeast strains, when transformed with the *lacZ* fusion plasmids, produced the same results. Also shown is a description of the Ty-*HIS4* region from each plasmid. Sequence differences in the Ty element of each of the plasmids fell into three regions designated E, J, and P. The * in region J of some constructs represents nine clustered base-pair changes found in the delta of Ty917(480) but not present on Ty917(467) and Ty917 (32). ↳, Normal *HIS4* transcription initiation site (35). ▨, *HIS4* coding sequence.

and B. Plasmid pRP368 (obtained from R. Yocum) is a derivative of pLG670 (16). It contains a *Bam*HI linker at the unique *Bam*HI site that allows correct in-frame *lacZ* fusions with our *HIS4*-containing fragment. pRP368 was digested with *Bam*HI, the 5′ protrusions were filled in by standard procedures, and the plasmid was subsequently digested with *Xho*I. This removes the *CYC1* coding and 5′ upstream sequences. A 360-bp *Xho*I-*Hae*III fragment that spans the Ty917(480)-*HIS4* junction was ligated into the pRP368 *Bam*HI-*Xho*I vector. The resultant plasmid is pAG1. pAG7 is the same as pAG1, except that the *Xho*I-*Hae*III insert contains the Ty917-*HIS4* junction. There is only a single-base-pair difference, a C-G→T-A change, between pAG1 and pAG7. This occurs in the delta of the Ty (region P, Fig. 1B). Into the unique *Xho*I site of pAG1 were cloned *Xho*I fragments containing the following Ty element sequences: Ty917(484), a strong on hybrid resulting in pAG4; Ty917, off resulting in pAG5; and Ty917(482), a weak on hybrid producing pAG6. The hybrid Ty elements are those used to construct the yeast strains S823 and S825 described by Roeder et al. (32), except that the *URA3*-containing fragment was removed by *Hin*dIII digestion and religation. *Xho*I fragments carrying Ty917(484) strong on, Ty917 off, and Ty917(482) weak on were also cloned into the *Xho*I site of pAG7. The resultant plasmids were pAG8, pAG9, and pAG10, respectively.

For protein-binding assays, 230-bp *Cla*I-*Pvu*II fragments from region E of pAG4 and pAG5 (Fig. 1B) were subcloned into the polylinker of pUC18, producing pRP632 and pRP630. Fragments were excised with *Eco*RI and *Hin*dIII for binding and methylation interference analysis.

Yeast strains. The *S. cerevisiae* strain used for transformation of the *lacZ* fusion plasmids was YeRP7 (*MAT***a** *his4-912 ura3-52*). Ura$^+$ transformants were selected on minimal medium. The resultant strains and the plasmids they contained were as follows: YeRP43(pAG1), YeRP45(pAG4), YeRP47(pAG5), YeRP49(pAG6), YeRP51(pAG7), YeRP53 (pAG8), and YeRP55(pAG9). YeRP7 was the principal strain from which crude extracts were prepared. Other strains used for preparation of extracts were YP3 (*MAT***a**/α *ura3-52/ura3-52 lys2-801/lys2-801 ade2-101/ade2-101*) (18), FW102 (*MAT*α *his4-917 leu2-1 ura3-52*), and FW494 (*MAT***a** *his4-917 leu2-1 ura3-52 spt3-101*) (40).

β-Galactosidase assay. Measurement of β-galactosidase activity in *S. cerevisiae* was as described by Guarente and Ptashne (16). Cells were disrupted by treatment with 0.05 ml of chloroform and 0.02 ml of 0.1% sodium dodecyl sulfate and vortexing. β-Galactosidase activity was determined as described by Miller (26).

Yeast transformation. Yeast strains were transformed by the spheroplast method (20).

DNA-protein-binding assays. Yeast extracts were prepared

by the method of Berman et al. (3) and stored at −70°C, and 630 off and 632 on *Eco*RI-*Hin*dIII fragments were end labeled with [α-^{32}P]dATP by standard methods (24). A typical binding reaction contained in 20 μl approximately 5 ng of end-labeled fragment (10,000 to 20,000 cpm), 70 mM KCl, 15 mM *N*-2-hydroxyethylpiperazine-*N'*-2-ethanesulfonic acid (HEPES; pH 7.5), 15% (vol/vol) glycerol, 15 mM $MgCl_2$, and extracts and competitor DNA as described in Fig. 2, 3, and 6. The mixture was incubated at 20°C for 25 min before electrophoresis through 0.8-mm-thick native 5% polyacrylamide gels (80:1 acrylamide-bis). Electrophoresis was performed at 180 V for approximately 2 h in a cold room at 5°C. Gels were wrapped in Saran Wrap and exposed to Kodak XAR film overnight.

Methylation interference analysis. The 632 on DNA fragment labeled at the 5′ end of one strand was partially methylated as described by Siebenlist and Gilbert (34). The methylated DNA was used in a 10-fold-scaled-up binding reaction, and the mixture was electrophoresed as described above. The gel was exposed to film for 3 h to identify bound and unbound fractions. Gel slices containing bound and unbound fractions were excised with a razor. DNA from each fraction was electroeluted onto ion-exchange paper (NA45; Schleicher & Schuell, Inc.) directly by using an agarose gel with preformed slots to support the paper and polyacrylamide gel fragment. Each fraction was eluted with 1 M NaCl and extracted with phenol-chloroform and chloroform and precipitated with 95% ethanol. Each sample was then treated with piperidine as described by Siebenlist and Gilbert (34) before electrophoresis through 8% polyacrylamide–7 M urea sequencing gels. The control G-ladder contains 632 on methylated DNA cleaved with 1 M piperidine.

RESULTS

A single-base-pair change in Ty affects adjacent-gene expression. To characterize the effect of Ty sequences on *HIS4* transcription further, we constructed a series of *HIS4-lacZ* fusion plasmids. Two plasmids, pAG1 and pAG7, in which nucleotides encoding the first 12 amino acids of the *HIS4* protein are fused in frame to the *Escherichia coli* β-galactosidase (*lacZ*) gene were initially constructed. These *lacZ* fusions contain the *HIS4*-Ty junction, including approximately 240 bp of the Ty delta sequence. Different Ty elements were cloned as *Xho*I fragments into the unique *Xho*I site in each plasmid. β-Galactosidase activity was measured in *S. cerevisiae* transformed with these plasmids. Figure 1A illustrates the cloning strategy used. pAG1 and pAG7 differ only by the T→C change present in the delta of Ty917(480) but not in Ty917 or Ty917(467). Insertion of sequences from Ty917 (off element) and its two on derivatives into each of these plasmids allows assessment of the contribution to β-galactosidase expression made by each of the changes in the critical 730-bp region. The remainder of the Ty in each construction is from the Ty917 off element. Thus, all Ty elements in the resultant plasmid constructions are identical in sequence except for those changes indicated in Fig. 1B. The copy numbers of plasmids in all transformed strains were essentially the same. The regions containing sequence changes were designated E, J, and P. Region E is in the epsilon of the Ty bounded by *Cla*I and *Pvu*II restriction sites. Adjacent to this is the epsilon-delta junction or J region, which lies between the *Pvu*II restriction site in the epsilon and the *Xho*I restriction site in the delta. The remaining sequences from the *Xho*I site to the end of the delta are designated P because this region contains Ty promoter elements. The level of β-galactosidase activity from a yeast strain transformed with each of the plasmids is also shown in Fig. 1B. In these constructions, normal *HIS4* promoter and upstream regulatory sequences are absent but transcription of the *HIS4-lacZ* fusion initiates at the normal position (W. J. Feaver and R. E. Pearlman, unpublished data). The A/T-rich Ty delta sequence likely provides the TATA promoter element required for *HIS4* gene expression. Constructs pAG1 and pAG7 contain this delta region but no additional Ty sequences and fail to demonstrate any detectable β-galactosidase activity. Therefore, additional upstream epsilon sequences are required for even the basal level of gene expression observed from constructs containing the Ty917 off elements (pAG5 and pAG9 in Fig. 1B). When the T in region E was replaced by a C, β-galactosidase activity increased dramatically (pAG5 versus pAG6 and pAG9 versus pAG10 in Fig. 1B). The additional changes in region J have only a minor effect on β-galactosidase activity (pAG6 versus pAG4 and pAG10 versus pAG8 in Fig. 1B) as does the T→C change in region P (pAG8 versus pAG4 and pAG10 versus pAG6 in Fig. 1B). This T→C transition in region P alone has no effect on β-galactosidase expression (pAG9 versus pAG5). These results illustrate that the sequence in region E is the major contributor to regulation of adjacent-gene expression.

The Ty regulatory sequence binds a *trans*-acting factor in vitro. Because the T→C change in region E (Fig. 1B) is contiguous with an enhancer core sequence (32), we tested the ability of the approximately 240-bp *Cla*I-*Pvu*II E region fragment to bind proteins in vitro from a crude yeast extract. A gel electrophoresis mobility shift assay (13) was used. When the DNA fragments were electrophoresed through native polyacrylamide gels, the mobility of DNA fragments bound to proteins was retarded relative to the mobility of naked or unbound DNA. In this manner, less-than-femtomole quantities of sequence-specific DNA-binding proteins can be detected (37). Appropriate fragments from region E of off and on Ty elements were subcloned as *Cla*I-*Pvu*II fragments into pUC18 and excised with *Eco*RI and *Hin*dIII. The off and on fragments differ only by a single T→C change and are designated 630 (from plasmid pAG5) and 632 (from plasmid pAG4), respectively. Approximately 5 ng (20,000 cpm) of end-labeled fragments was incubated with yeast whole-cell extract for 25 min at room temperature and subsequently electrophoresed through native 5% polyacrylamide gels as described in Materials and Methods. The gels were autoradiographed without drying. Figure 2 shows the results of such an experiment. Each fragment was incubated with 1 μl of crude extract in which the protein concentration was approximately 15 mg/ml. Upon incubation with extract, the 632 on fragment produces a relatively sharp band of reduced mobility that is indicative of a specific DNA-protein complex (Fig. 2, lanes 1 and 2). The same experiments, performed with an end-labeled 630 off fragment, produced very different results (Fig. 2, lanes 3 and 4). No distinct band of reduced mobility was visible when extract was added. Clearly, there was a striking difference between the protein-binding abilities of these two fragments that differ only by a single-base-pair change. The on fragment specifically bound a factor from crude yeast extract, whereas the off fragment failed to exhibit specific binding. These results are particularly interesting in light of the information obtained from the *lacZ* fusion constructions. The single-base-pair change that enhances adjacent-gene

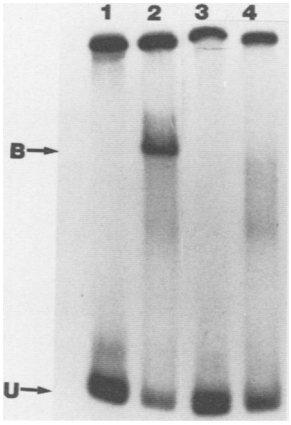

FIG. 2. Factor binding to region E of Ty2 elements. The 240-bp *Eco*RI-*Hin*dIII fragments, 632 (from region E of pAG4) and 630 (from region E of pAG5), were end labeled and mixed with crude whole-cell extract from YeRP7. Reactions were incubated at 20°C for 25 min before electrophoresis through 0.8-mm-thick 5% native polyacrylamide gels (80:1 acrylamide-bis ratio) in TBE (89 mM Tris hydrochloride, 89 mM boric acid, 20 mM EDTA; pH 8.0). Each reaction (20 μl) contained approximately 5 ng of end-labeled DNA (20,000 cpm) in 70 mM KCl, 15 mM $MgCl_2$, 15 mM HEPES (7.4), 15% glycerol, and different amounts of crude extract (protein concentration, 15 mg/ml). Lanes: 1, 632 DNA fragment without extract; 2, 632 fragment with 1 μl of extract; 3, 630 DNA fragment without extract; 4, 630 fragment with 1 μl of extract. B, Protein-bound DNA; U, unbound DNA.

expression in vivo is also required for a sequence-specific protein-DNA interaction in vitro.

Competition analysis of 632 fragment binding. The mobility shift assay was used in conjunction with various unlabeled competitor DNAs to examine further the specificity of the 632 DNA-protein interaction in vitro. An end-labeled 632 on fragment was incubated with 2 μl of crude extract in the presence of several unlabeled competitor DNAs before electrophoresis through a native 5% polyacrylamide gel (Fig. 3). One-hundred-fold molar excesses of poly(dI-dC), calf thymus DNA, and pUC18 DNA failed to compete for binding of the 632 on fragment (Fig. 3, lanes 1 to 4). However, the binding was essentially abolished by addition of 1 μg of a pUC18 plasmid containing the 632 Ty fragment (Fig. 3, lane 5). This competition was due to the presence of Ty DNA sequences, since pUC18 failed to compete for binding. The unlabeled 632 fragment sequences were present in approximately 20-fold molar excess. An equivalent amount of pUC18 carrying the 630 off DNA fragment was unable to compete for the sequence-specific binding (Fig. 3, lane 6). This is consistent with data from previous experiments in which the 630 off fragment did not display sequence-specific binding. In fact, even a 100-fold molar excess of the 630 off sequence failed to compete for binding of this factor (data not shown). To demonstrate further that the observed binding was specific for the 632 on fragment, we used additional plasmid DNAs containing either on or off Ty917 elements for competition analysis. pRP204 and pRP206 contain complete Ty917(480) and Ty917(467) sequences, respectively. Both of these on Ty element sequences contain the crucial T→C change in region E and were able to compete for binding of the factor (Fig. 3, lanes 7 and 8). Incubation with an equivalent amount of unlabeled pRP200, a plasmid carrying the Ty917 off element, did not result in competition for factor binding (Fig. 3, lane 9). The single-base-pair change in region E that defines the site of *trans*-factor interaction lies adjacent to an enhancer core-like sequence. We also used both linear and supercoiled simian virus 40 enhancer-containing plasmids for competition analysis (Fig. 3, lanes 10 and 11). pSV2CAT contains the simian virus 40 early promoter region but failed to compete for binding with the 632 on fragment, even at 100-fold molar excess. The enhancer core sequence present on both 630 and 632 fragments, therefore, appears not to contribute to binding in vitro. Finally, we treated the crude cell extract with proteinase K both before and after incubation with the 632 fragment (Fig. 3, lanes 12 and 13). All binding was eliminated, establishing that the retarded mobility of the labeled fragment was indeed due to interaction with protein. There appeared to be no requirement for topology of the competing molecules, since supercoiled and linear self DNAs exhibited equal abilities to compete with labeled 632 DNA (data not shown).

The gel electrophoretic mobility shift assay was successfully used to establish that the 632 on fragment is capable of binding a protein from crude yeast extracts. We designated this factor TyBF. TyBF binding is specific for the 632 on fragment. A single-base-pair change in the 632 fragment abolished this binding.

High-resolution mapping of the TyBF-binding site. The T→C change in region E of on and off elements is most likely in the region of TyBF interaction. To analyze the nucleotides required for binding further, we performed methylation interference experiments. Guanine residues in the major

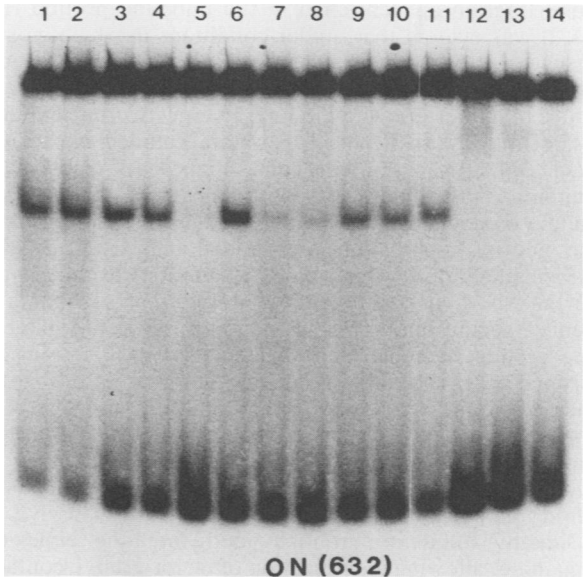

FIG. 3. Competition analysis of 632 fragment binding was performed as described in the legend to Fig. 2, except that each reaction contained end-labeled 632 DNA (20,000 cpm; 5 ng), 2 μl of extract, and the following amounts of unlabeled competitor DNAs. Lanes: 1, no competitor DNA; 2, 2 μg of poly(dI-dC) copolymer; 3, 10 μg of sonicated calf thymus DNA; 4, 1 μg of pUC18 DNA; 5, 1 μg of pRP632; 6, 1 μg of pRP630; 7, 1 μg of pRP204 [pBR322 containing Ty917(480)]; 8, 1 μg of pRP206 [pBR322 containing Ty917(467)]; 9, pRP200 (pBR322 containing Ty917); 10, 1 μg of linear pSV2CAT; 11, 1 μg of supercoiled pSV2CAT; 12, extract treated with 1 μg of proteinase K before binding reaction; 13, binding reaction treated with 1 μg of proteinase K (5 min) before electrophoresis; 14, no extract. All competitor DNAs were added as linear DNAs unless otherwise indicated.

groove making direct contact with the protein are identified by this technique. End-labeled 632 on fragments were partially methylated by treatment with dimethyl sulfate under conditions such that approximately one guanine per fragment was modified (34). These fragments were then used in scaled-up binding assays. The reaction mixtures were electrophoresed through 5% polyacrylamide gels and briefly autoradiographed to identify bound and free DNA fractions. DNA from each fraction was eluted separately and subjected to piperidine treatment, which cleaves the phosphate backbone adjacent to modified guanine residues. Finally, piperidine-treated bound and free DNAs were electrophoresed through an 8% polyacrylamide–7 M urea sequencing gel. Fragments methylated at the binding site should fail to bind protein and therefore be excluded from the bound fraction. Bands missing from the guanine sequence ladder of the bound fraction correspond to sites of methylation that eliminate protein binding. These same bands may be enhanced in the unbound fraction, depending on the proportion of DNA bound to protein.

The results of the methylation interference analysis with the 632 on fragment are shown in Fig. 4A. The upper and lower strands represent, respectively, coding and noncoding strands with respect to Ty transcription. The upper-strand sequence ladder of the bound fraction disclosed the absence of two guanine cleaves (Fig. 4A, arrows). The decreased intensity of the sequence ladder for the unbound fraction was indicative of the small proportion of unbound DNA in this experiment, yet the intensity of the same two guanine cleavages was greatly enhanced. The lower-strand sequence ladder revealed three guanine cleavages missing from the bound fraction and enriched in the unbound portion (Fig. 4A). It was evident that, under conditions in which most of the methylated fragment was in the bound state, the predominant bands in the unbound fraction were those methylated at sites that interfere with binding. Even with crude extracts, this technique can be used to obtain detailed information about DNA-protein interactions. Figure 4B displays the minimum TyBF-binding site as defined by major-groove guanine contacts and its flanking sequences. The guanines identified by methylation interference analysis are marked with filled circles. Also shown is the C→T change that prevents both adjacent-gene expression in vivo and sequence-specific binding in vitro. The methylation interference results demonstrated that the C-G→T-A transition disrupts a critical protein-guanine contact in the TyBF-binding site. It was also evident that the enhancer core sequence adjacent to this site was not involved in protein-guanine contacts made by TyBF in vitro. This does not preclude possibilities that the enhancerlike sequence is required for recognition of TyBF or other factors in vivo.

Dimethyl sulfate preferentially methylated the ^{7}N position of guanine, allowing identification of major-groove contacts. This experiment also provided information about the geometry of the binding site. The positions of the five guanine contacts on the DNA molecule are such that four of five contacts are on one face of the helix, while the remaining position is on the opposite face. The binding sequence, as defined by methylation interference, spans at least 13 bp of DNA. This *cis*-acting binding sequence is not homologous to any known protein-binding or regulatory sequence.

Other TyBF-like recognition sequences in Ty elements. Ty2 elements possess another region with limited homology to the TyBF-binding sequence approximately 50 bp downstream of the TyBF-binding site (Fig. 5). This downstream sequence, present on both 632 on and 630 off fragments, was not protected in the methylation interference experiment (Fig. 4A); therefore, it is not a functional binding site for TyBF. It lacks two major-groove guanine contacts, including the nucleotide which is also altered in the upstream site of the Ty917 off element (Fig. 5). This sequence is denoted by brackets in Fig. 5 to emphasize that it does not bind TyBF in vitro. We also examined other published Ty sequences for

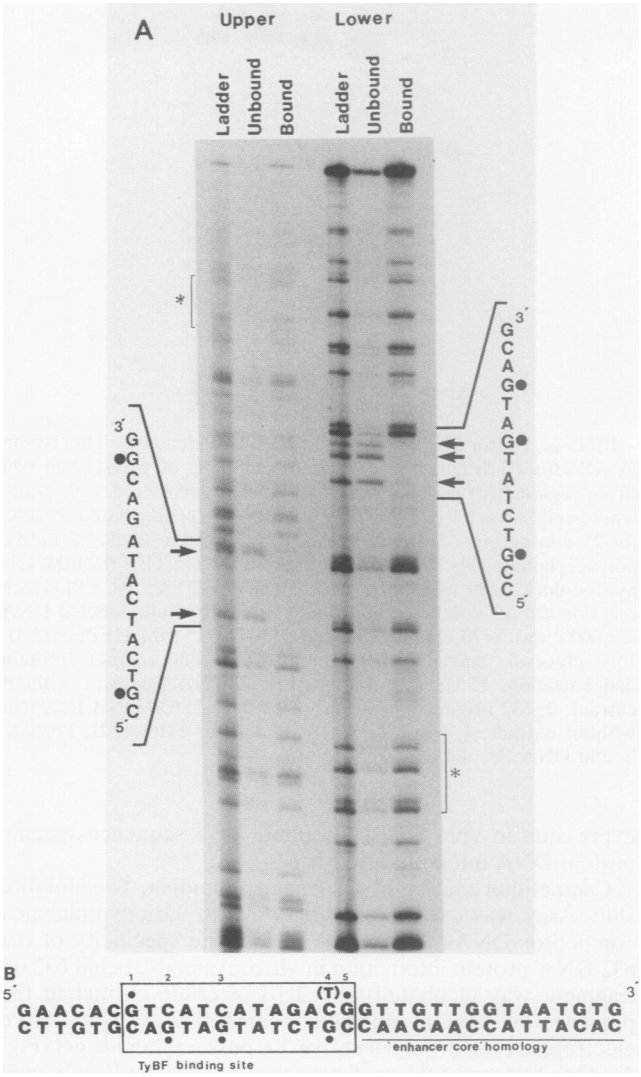

FIG. 4. TyBF-binding sequence. (A) Methylation interference analysis of the 632 fragment. Bound and unbound DNA fractions were excised from the gel, and the DNA was electroeluted. After treatment with 1 M piperidine, each sample was electrophoresed through an 8% polyacrylamide–7 M urea sequencing gel. The G-ladder shown is a portion of the methylated, end-labeled 632 DNA directly treated with piperidine. Sequences for the upper and lower strands are written 5′→3′ in the direction of *HIS4* transcription, which is opposite to that of Ty transcription. Symbols: ●, guanines absent from the bound fraction and enriched in the unbound fraction; *, a downstream TyBF-like site to which there was no binding (see Fig. 6); ⟶, enhanced or missing guanine cleavage sites. (B) TyBF-binding sequence. Methylation interference experiments defined a minimum 13-bp sequence (boxed sequence) with five protein-guanine contacts (●) numbered 1 to 5. The C→T change at G4 that eliminated binding is shown in parentheses. Also shown is the region of homology to the simian virus 40 enhancer core sequence (32).

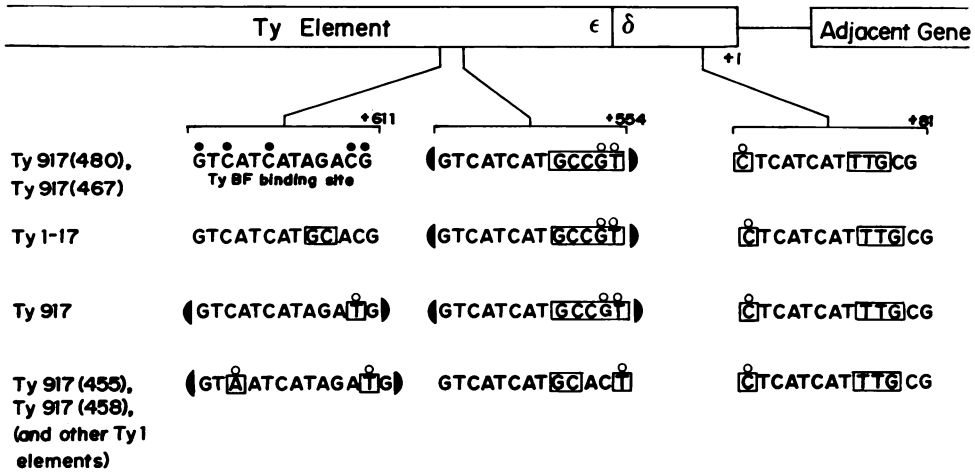

FIG. 5. Presence of TyBF-like binding sequences on other Ty2 and Ty1 elements. Shown are sequences in the 5′ (adjacent-gene-proximal) end of other Ty elements that share homology to the TyBF-binding site present on the His$^+$ or on derivatives of Ty917, Ty917(480), and Ty917(467). Ty1-17 is another Ty2 element described by Fulton et al. (14). Ty917(455) and Ty917(458) are Ty1 elements. The weak TyBF-binding sequence from Ty917(455) and Ty917(458) at +554 is present on all of the other Ty1 elements examined (5, 12, 14, 17). Symbols: ●, protein-guanine contacts of the TyBF-binding site; □, nucleotides that differ from the 13-bp TyBF-binding sequence; ○ changes that disrupt any of the five protein-guanine contacts; ( ), sequence that is not a functional TyBF-binding site (on the basis of sequence homologies and methylation interference experiments). Numbering starts (+1) at the end of the Ty delta and increases in the direction of Ty transcription toward the epsilon sequences.

the presence of the TyBF-binding sequence. Other Ty2 elements and Ty1 elements also contain TyBF-like recognition sequences. Ty1-17 is a Ty2 element whose 5′ sequence has been reported previously (14). A nearly perfect TyBF-binding site, including all five major-groove guanine contacts, was present in the same position and orientation as in Ty917(467) and Ty917(480) (Fig. 5). On the basis of sequence homology, we predict that TyBF binds to this site.

Ty917(455) and Ty917(458) are additional His$^+$ gene conversion derivatives of the *his4-917* mutation (32). They are Ty1 elements present in the same position and orientation as Ty917. Each of these elements contains TyBF-binding sequence homologies in the same location as in the Ty2 elements (Fig. 5). A 350-bp fragment from this region of Ty917(455) was assessed for its ability to compete with the labeled 632 fragment for TyBF binding (Fig. 6). The results illustrate that the Ty1 fragment competes for TyBF binding more efficiently than does the 630 off fragment but much less efficiently than does the cold 632 self competitor. A 10-fold molar excess of the Ty1 fragment was required to achieve a level of competition equivalent to that obtained with the 632 DNA fragment (Fig. 6, lane 3 versus lane 6). Similar results were obtained when the corresponding 350-bp fragment from Ty917(458) was used as a competitor. This suggests that TyBF binds to Ty1 elements but with a greatly reduced affinity compared with binding to Ty2 elements. Although Ty917(455) and Ty917(458) insertions at *HIS4* regulate adjacent-gene expression and both contain simian virus 40 enhancer core similarities (32), specific sequences responsible for the His phenotype have not been identified in these elements. The significance of weak TyBF binding to these elements is undetermined.

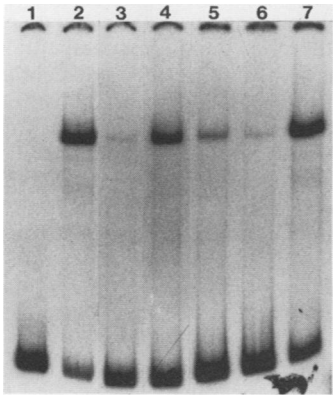

FIG. 6. Ty1 fragment competition for TyBF binding. Each binding reaction contained 5 ng of 632 DNA (25,000 cpm), 2 μl of YeRP7 extract, and 2 μg of poly(dI-dC) copolymer. Other conditions were as described in the legend to Fig. 2, except that unlabeled competitor DNAs were added where indicated. Lanes: 1, no competitor or extract; 2, no competitor; 3, 10-fold molar excess of unlabeled 632 DNA; 4, 200-fold molar excess of 630 DNA fragment; 5, 50-fold molar excess of 350-bp Ty917(455) fragment; 6, 100-fold molar excess of Ty917(455) fragment; 7, 1,000-fold molar excess of sonicated calf thymus DNA. The gel was dried before autoradiography.

In addition to having a TyBF-like binding site in the epsilon sequences, each Ty examined also contained a TyBF-like recognition sequence in the delta (Fig. 5). Four of five guanine contacts identified in the 632 TyBF-binding site were conserved, and this sequence also had a low affinity for TyBF binding in vitro (data not shown). The functional relevance of weak binding at this site is unclear, but it suggests that multiple complex interactions may govern Ty-mediated adjacent-gene expression.

TyBF binding in vitro independent of *MAT* and *SPT3*. The mating type (*MAT*) locus of *S. cerevisiae* often regulates Ty-mediated adjacent-gene expression (11). We tested the ability of crude extracts from cells differing at *MAT* to bind the 632 on fragment by using the electrophoretic gel mobility assay. Extracts from haploid *MAT***a** (Fig. 7, lane 2) and *MAT*α (lane 3) cells and *MAT* **a**/α diploids (lane 5) all bound the 632 on fragment. Although regulation of *HIS4* gene expression in Ty917(467) and Ty917(480) is under mating

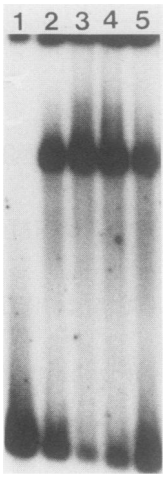

FIG. 7. Effect of *MAT* and *SPT3* on TyBF binding in vitro. The end-labeled 632 DNA fragment (20,000 cpm; 5 ng) was used in the electrophoretic binding assay under conditions described in the legend to Fig. 2, except that crude extracts from different strains were used. Lanes: 1, no extract; 2, 2 μl of YeRP7 extract (15 mg/ml); 3, 2 μl of FW102 (*MATα his4-917 leu2-1 ura3-52*) extract (12 mg/ml); 4, 2 μl of FW494 (*MATa his4-917 leu2-1 ura3-52 spt3-101*) extract (12 mg/ml); 5, 2 μl of YP3 (*MATa/α ura3-52/ura3-52 lys2-801/lys2-801 ade2-101/ade2-101*) extract (10 mg/ml). All reactions contained 1 μg of poly(dI-dC).

type control (32), binding of TyBF in vitro is not affected by the *MAT* locus. These data suggest that the cell type control of Ty transcription and Ty-mediated adjacent-gene expression may be exerted through *cis*-acting sequences and *trans*-acting factors different from those responsible for adjacent-gene activation. Another locus known to regulate both Ty transcription and Ty-mediated adjacent-gene expression is *SPT3* (39, 40). The *spt3-101* mutation supresses the effect of Ty insertions and restores *HIS4* gene expression to normal levels. We tested crude extracts from isogenic yeast strains differing only at *MAT* and *SPT3* for their ability to bind the 632 on fragment. Extracts from strains FW102 (*SPT3*) and FW494 (*spt3-101*) bind the 632 on fragment equivalently (Fig. 7, lanes 3 and 4). Thus, *SPT3* control of adjacent-gene expression appears to be independent of TyBF binding. This is not surprising, since the *SPT* mutations described by Winston et al. (39) were selected by their ability to suppress the effects of a Ty912 delta insertion. The primary target sequence for TyBF appears in the epsilon sequences of Ty2 elements. On the basis of this observation, we do not expect TyBF to be a product of one of the *SPT* genes; however, its production or activity may be regulated by one of the other *SPT* genes.

DISCUSSION

Ty elements can be considered as endogenous, noninfective retroviruses of *S. cerevisiae*. They may simply be parasitic or selfish DNA, but the ability of Ty elements to regulate the expression of genes adjacent to an insertion site may also confer selective advantages to a cell. We investigated the mechanism of adjacent-gene regulation by Ty2 elements. Our results with the *lacZ* fusion plasmids confirmed previous reports that sequences in the 5′ end of a Ty element are responsible for regulation of adjacent-gene expression (12, 32). In these plasmid constructions, the Ty917-derived elements are positioned, with respect to *HIS4* sequences, identically to the chromosomal Ty element in the *his4-917* mutation. The cloning strategy we used ensured that almost-full-length elements (each Ty lacks only about 240 bp of DNA outside the *Xho*I site in the 3′ delta sequence) were inserted in the correct position. This eliminated possible effects on β-galactosidase expression due to flanking plasmid sequences. We also took the precaution of not deleting Ty sequences or introducing additional sequences into the element, ensuring that any differences in β-galactosidase levels could be directly attributed to specific nucleotide changes in the Ty. The sequence changes we introduced were the result of reconstructing naturally occurring Ty elements. The most important regulatory sequence is that defined by the T→C change found in region E of Ty917 on derivatives. The level of β-galactosidase activity increased slightly when the additional T→C change in region P was present. We attributed this effect to the fact that this particular change altered a region in the Ty promoter from TATAA to TATGA. There is evidence (L. Coney and S. Roeder, personal communication) that this change markedly decreased Ty transcription (as measured by β-galactosidase production in *lacZ* fusion plasmids), which might otherwise have competed with transcription of the adjacent gene for limiting factors.

In eucaryotes, transcriptional regulation of gene expression is dependent on the sequence-specific interaction of *trans*-acting factors at multiple *cis*-acting regulatory sites. These regulatory sequences are found at various distances both upstream and downstream of genes and, in some cases, within genes. Normal expression of the *HIS4* gene of *S. cerevisiae* also requires multiple upstream sequences: the TATA box at approximately −60 relative to the mRNA start site (27) and several copies of the *GCN4*-binding sequence (2, 19). These sequences are displaced approximately 6 kilobases by Ty sequences in the *his4-917* mutation. The sequence defined by the single-base-pair change in region E of Ty2 elements that activates adjacent-gene transcription is a good candidate for *trans*-factor interaction. The results of the electrophoretic mobility shift assay confirmed this postulate. The 632 on fragment specifically bound a factor from crude yeast extract, while a fragment, 630 off, differing only by a single-base-pair transition, failed to bind. It is conceivable that TyBF binding to a Ty element may repress Ty transcription, thereby eliminating competition for transcription of the divergently transcribed adjacent gene. However, the T → C change in region E that activates adjacent-gene expression and allows TyBF binding does not decrease Ty transcription (Coney and Roeder, personal communication). We therefore believe TyBF to be a transcriptional activator of adjacent-gene expression. A role for TyBF binding in vivo in the activation of adjacent-gene expression is supported by the presence of a block at the TyBF-binding site to exonuclease III digestion of nuclear chromatin in on, but not off, yeast strains (Feaver and Pearlman, manuscript in preparation).

High-resolution mapping of the TyBF-binding site by methylation interference analysis revealed five guanines in direct contact with protein. The minimum binding site defined by these guanine residues is 13 bp long. Four of five major-groove contacts made by TyBF lie on one surface of the DNA helix, while the other (G3) is on the opposite face. Conceivably, TyBF has arms that wrap around the DNA molecule. The 13-bp binding site does not display dyad symmetry, although modest symmetry can be observed if sequences flanking the binding site are included in the analysis. The guanine contacts lack the symmetry usually

associated with procaryotic homo-oligomeric DNA-binding proteins. TyBF binding to its target sequence does not resemble that of other well-characterized eucaryotic DNA-binding proteins. SP1 (15), T antigen (21), and NF1 (7) all display major-groove contacts on only one face of the helix. In this respect, TyBF is more similar to yeast *HAP1* protein, which binds both *CYC7*- and *UAS1*-binding sites via a single major-groove back-face contact (28). Unlike the situation with *HAP1*, we failed to detect any adenine contacts in the minor groove by methylation interference analysis. Dimethyl sulfate also methylated the ^{3}N position of adenine, but these residues were cleaved less efficiently with piperidine than were guanine residues. Further analysis is required to establish whether TyBF binds via major-groove contacts alone.

Examination of sequences in corresponding regions from six other Ty elements revealed that a TyBF-like recognition sequence is present at least twice in each element. Although Ty1 epsilon fragments containing these potential binding sites competed for TyBF binding and themselves bound proteins weakly from crude extracts, we were unable to detect protection or enhancement at these Ty1 sequences by using methylation interference analysis (data not shown). Several *cis*-acting regulatory sequences in a Ty1 element in a ROAM mutation have been identified (6, 12). The organization of these putative regulatory regions in the 5′ end of the element is complex, and only one of these regions (block I, reference 12) includes TyBF-binding site similarities. It is possible that Ty1 elements utilize additional, if not different, *cis* sequences and *trans*-acting factors to regulate adjacent-gene expression. These apparent differences in regulatory mechanisms between Ty1 and Ty2 elements may be reflected in the fundamental differences between the off-on regulation at *HIS4* presented here and the overproducing phenotype of the constitutively expressed ROAM mutations described by others (31). The TyBF-like binding sequence found in the Ty2 deltas cannot alone activate adjacent-gene transcription. pAG1 and pAG7 contain this Ty delta sequence; however, there was no detectable expression of β-galactosidase from these constructs. Therefore, the presence of a single, weak binding site is insufficient for adjacent-gene activation. Possibly, a protein-protein interaction involving binding at the epsilon and delta TyBF sites is required for adjacent-gene expression. There is precedence for a model in which interaction between two proteins bound to DNA is required for transcriptional activation in *S. cerevisiae* (22).

In this report, we have demonstrated the presence of a protein, TyBF, which specifically interacts with yeast transposable elements. We determined the precise site of TyBF binding and provided evidence that the sequence-specific interaction of TyBF with a *cis*-acting element in Ty2 activates adjacent-gene expression. Recent characterization of Ty-encoded proteins (1) suggests that TyBF is unlikely to be encoded by the Ty element. If TyBF is host encoded, then what is its role in normal cellular metabolism? We are presently purifying TyBF and attempting to clone the gene that encodes this protein as a prelude to addressing further questions about the normal function of TyBF and about its role in Ty-mediated enhancement of adjacent-gene expression.

ACKNOWLEDGMENTS

This work was supported by grants from the Natural Sciences and Engineering Research Council of Canada and the Medical Research Council of Canada. Anurag Goel was supported by a Province of Ontario Graduate Fellowship.

We thank William J. Feaver for constructing pRP630 and pRP632 and for advice and discussion throughout the course of this work, Anita Samardzic for technical assistance, Colleen Kelly for secretarial assistance, and Gordon Temple for photography. We also thank Leslie Coney and Shirleen Roeder for communicating data prior to publication and for critical reading of the manuscript and Fred Winston for supplying *SPT3* mutant strains.

LITERATURE CITED

1. **Adams, S. E., J. Mellor, K. Gull, R. B. Sim, M. F. Tuite, S. M. Kingsman, and A. J. Kingsman.** 1987. The functions and relationships of Ty-VLP proteins in yeast reflect those of mammalian retroviral proteins. Cell **49**:111–119.
2. **Arndt, Z. K., and G. R. Fink.** 1986. GCN4 protein, a positive transcription factor in yeast, binds general control promoters to all 5′ TGACTC 3′ sequences. Proc. Natl. Acad. Sci. USA **83**:8516–8520.
3. **Berman, J., C. Y. Tachibana, and B.-K. Tye.** 1986. Identification of a telomere-binding activity from yeast. Proc. Natl. Acad. Sci. USA **83**:3713–3717.
4. **Boeke, J. D., D. J. Garfinkle, C. A. Styles, and G. R. Fink.** 1985. Ty elements transpose through an RNA intermediate. Cell **40**: 491–500.
5. **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.
6. **Company, M., and B. Errede.** 1987. Cell-type-dependent gene activation by yeast transposon Ty1 involves multiple regulatory determinants. Mol. Cell. Biol. **7**:3205–3211.
7. **deVries, E., W. van Driel, S. J. L. van den Heuvel, and P. C. van der Vliet.** 1987. Contact point analysis of the HeLa nuclear factor I recognition site reveals symmetrical binding at one side of the DNA helix. EMBO J. **6**:161–168.
8. **Donahue, T. F., P. J. Farabaugh, and G. R. Fink.** 1982. The nucleotide sequence of the HIS4 region of yeast. Gene **18**:47–59.
9. **Eibel, H., and P. Philippsen.** 1984. Preferential integration of yeast transposable element Ty into a promoter region. Nature (London) **307**:386–388.
10. **Elder, R. T., E. Y. Loh, and R. W. Davis.** 1983. RNA from the yeast transposable element Ty1 has both ends in the direct repeats, a structure similar to retrovirus RNA. Proc. Natl. Acad. Sci. USA **80**:2432–2436.
11. **Errede, B., T. S. Cardillo, F. Sherman, E. Dubois, J. Deschamps, and J. M. Waime.** 1980. Mating signals control expression of mutations resulting from insertion of transposable repetitive element adjacent to diverse yeast genes. Cell **25**:427–436.
12. **Errede, B., C. Mahshid, J. D. Ferchak, C. A. Hutchison, and W. S. Yarnell.** 1985. Activation regions in a yeast transposon have homology to mating type control sequences and to mammalian enhancers. Proc. Natl. Acad. Sci. USA **82**:5423–5427.
13. **Fried, M., and D. M. Crothers.** 1981. Equilibria and kinetics of Lac repressor operator interactions by polyacrylamide gel electrophoresis. Nucleic Acids Res. **9**:6505–6525.
14. **Fulton, A. M., J. Mellor, M. J. Dobson, J. Chester, J. R. Warmington, K. J. Indge, S. G. Oliver, P. de la Paz, W. Wilson, A. J. Kingsman, and S. M. Kingsman.** 1985. Variants within the yeast Ty sequence family encode a class of structurally conserved proteins. Nucleic Acids Res. **13**:4097–4112.
15. **Gidoni, D., W. S. Dynan, and R. Tjian.** 1984. Multiple specific contacts between a mammalian transcription factor and its cognate promoters. Nature (London) **312**:409–413.
16. **Guarente, L., and M. Ptashne.** 1981. Fusion of *Escherichia coli lacZ* to the cytochrome *c* gene of *Saccharomyces cerevisiae*. Proc. Natl. Acad. Sci. USA **78**:2199–2203.
17. **Hauber, J., P. Nelbock-Hochstetter, and H. Feldman.** 1985. Nucleotide sequence and characteristics of a Ty element from yeast. Nucleic Acids Res. **13**:2745–2759.
18. **Hieter, P., D. Pridmore, J. H. Hegemann, M. Thomas, R. W. Davis, and P. Philippsen.** 1985. Functional selection and analysis of yeast centromeric DNA. Cell **42**:913–921.
19. **Hill, D. E., I. A. Hope, J. P. Macke, and K. Struhl.** 1986.

Saturation mutagenesis of the yeast *HIS3* regulatory site: requirements for transcriptional induction and for binding by GCN4 activator protein. Science **234**:451–457.
20. **Hinnen, A., J. B. Hicks, and G. R. Fink.** 1978. Transformation of yeast. Proc. Natl. Acad. Sci. USA **75**:1929–1933.
21. **Jones, K. A., and R. Tjian.** 1984. Essential contact residues within SV40 large T antigen binding sites I and II identified by alkylation-interference. Cell **36**:155–162.
22. **Keegan, L., G. Grace, and M. Ptashne.** 1986. Separation of DNA binding from the transcription-activating function of a eukaryotic regulatory protein. Science **231**:699–703.
23. **Kingsman, A. J., R. Gimlich, L. Clarke, C. Chinault, and J. Carbon.** 1981. Sequence variation in dispersed repetitive sequences in *Saccharomyces cerevisiae*. J. Mol. Biol. **145**:619–632.
24. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
25. **Mellor, J., S. M. Fulton, J. M. Dobson, W. Wilson, S. M. Kingsman, and A. J. Kingsman.** 1985. A retrovirus-like strategy for expression of a fusion protein encoded by yeast transposon Ty1. Nature (London) **313**:243–246.
26. **Miller, J. H.** 1972. Assay of β-galactosidase, p. 352–355. *In* J. H. Miller (ed.), Experiments in molecular genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
27. **Nagawa, F., and G. R. Fink.** 1985. The relationship between the "TATA" sequence and transcription initiation sites at the *HIS4* gene of *Saccharomyces cerevisiae*. Proc. Natl. Acad. Sci. USA **82**:8557–8561.
28. **Pfeifer, K., T. Prezant, and L. Guarente.** 1987. Yeast HAP1 activator binds two upstream activation sites of different sequence. Cell **49**:19–27.
29. **Roeder, G. S., P. J. Farabaugh, D. T. Chaleff, and G. R. Fink.** 1980. The origins of gene instability in yeast. Science **209**:1375–1380.
30. **Roeder, G. S., and G. R. Fink.** 1982. Movement of yeast transposable elements by gene conversion. Proc. Natl. Acad. Sci. USA **79**:5621–5625.
31. **Roeder, G. S., and G. R. Fink.** 1983. Transposable elements in yeast, p. 299–328. *In* J. A. Shapiro (ed.), Mobile genetic elements. Academic Press, Inc., New York.
32. **Roeder, G. S., A. B. Rose, and R. E. Pearlman.** 1985. Transposable element sequences involved in enhancement of yeast gene expression. Proc. Natl. Acad. Sci. USA **82**:5428–5432.
33. **Rubin, G. M.** 1983. Dispersed repetitive DNAs in drosophila, p. 329–361. *In* J. A. Shapiro (ed.), Mobile genetic elements. Academic Press, Inc., New York.
34. **Siebenlist, U., and W. Gilbert.** 1980. Contacts between *E. coli* RNA polymerase and an early promoter of phage T7. Proc. Natl. Acad. Sci. USA **77**:122–126.
35. **Silverman, S. J., and G. R. Fink.** 1984. Effects of Ty insertions on *HIS4* transcription in *Saccharomyces cerevisiae*. Mol. Cell. Biol. **4**:1246–1251.
36. **Simchen, G., F. Winston, C. A. Styles, and G. R. Fink.** 1984. Ty-mediated gene expression of the *LYS2* and *HIS4* genes of *Saccharomyces cerevisiae* is controlled by the same SPT genes. Proc. Natl. Acad. Sci. USA **81**:2431–2434.
37. **Singh, H., R. Sen, D. Baltimore, and P. A. Sharp.** 1986. A nuclear factor that binds to a conserved sequence motif in transcriptional control elements of immunoglobulin genes. Nature (London) **319**:154–158.
38. **Varmus, H. E.** 1983. Retroviruses, p. 411–503. *In* J. A. Shapiro (ed.), Mobile genetic elements. Academic Press, Inc., New York.
39. **Winston, F., D. T. Chaleff, B. Valent, and G. R. Fink.** 1984. Mutations affecting Ty-mediated expression of the *HIS4* gene of *Saccharomyces cerevisiae*. Genetics **107**:179–197.
40. **Winston, F., K. J. Durbin, and G. R. Fink.** 1984. The *SPT3* gene is required for normal transcription of Ty elements in *S. cerevisiae*. Cell **39**:675–682.