PCR-aided Amplification: A Tool for the Study of *In Vivo* Protein-DNA Interactions in the Yeast U6 Promoter

Uma Shah
University Undergraduate Fellow, 1990-91
Texas A&M University
Department of Biochemistry/Biophysics

APPROVED:

Fellows Advisor _____

Honors Program Director _____

ACKNOWLEDGEMENTS

In the course of my research my most significant findings were about science itself. One of my most valuable lessons was the necessity for patience in research. Often it appears that progress is very slow, however it is at those times that it is most crucial to follow the scientific method of logically working through obstacles. I have also seen the value of integrity in the reporting of findings because the goal of basic science is not to produce for personal benefit but to contribute to a much larger picture in a search for the truth.

I thank members of the Kunkel laboratory, especially Randall Goomer and Gary R. Kunkel for technical assistance. I acknowledge the gift of the yeast strain from Michael Cusick. I am also extremely grateful to Gary R. Kunkel , Edward Funkhouser, and Robert Greve for comments that have greatly improved this thesis.

ABSTRACT

The classification of RNA polymerase III transcribed genes into genes with internal promoters (within the coding region) and those with external promoters is no longer a clear one. The *S. cerevisiae* U6 promoter is comprised of control elements outside the coding region (on the 5' and 3' sides) and a possible internal control element. In the following study I demonstrate the optimization of ligation-mediated PCR for the *S. cerevisiae* system as a technique that is instrumental in genomic footprinting for determination of protein-DNA interactions in the yeast U6 promoter.

INTRODUCTION

The process of interpreting the genetic code is truly a remarkable one. It involves a series of complex interactions many of which may serve as key regulatory steps in gene expression. A primary point of regulation of gene expression is transcription initiation. It is the first step in decoding a gene and is mediated differently in prokaryotic and eukaryotic cells. In prokaryotic (bacterial) transcription, one RNA polymerase exists and bears the principle responsibility for transcription. In eukaryotes, it appears that transcription factors are the primary recognizors of the promoter and mediators of RNA polymerase binding. Eukaryotic transcription takes on further dimensions due to the existence of several "classes" of genes differentiated by the use of one of three possible eukaryotic RNA polymerases.

Initially, a clear distinction could be made between RNA pol II promoters (found in protein-coding genes and some snRNAs) and RNA pol III promoters (found in 5S RNA and tRNA genes). Pol II promoters were characterized as residing in the 5' flanking region while pol III promoters were localized downstream of the transcription start site. More specifically, most protein-coding genes have a module with a T/A-rich sequence known as the TATA box (Fig. 1A). The TATA box is the only promoter element with a relatively constant position (roughly 25 bp upstream) with regard to the start site and functions in selection of the start site [1,2]. Some pol II promoters, such as those for most small nuclear RNA (snRNA genes), lack TATA boxes. These snRNA genes contain a proximal sequence

element or PSE which is the functional equivalent of the TATA box but is located farther upstream than the normal TATA box location [3] (Fig. 1B).

Transcription of RNA pol III genes is proving to be more complicated; many modules and their corresponding transcription factors are involved. The internal promoter or internal control region (ICR) of these genes consists of discontinuous segments of DNA comprised of consensus sequences that bind particular pol III transcription factors [1,4-6]. Class III_{ICR} promoters are divided into Type I and Type II [4]. Type I is responsible for the transcription of 5S RNA genes and contains two functional modules, *box A* and *box C* (Fig. 2A). Type II promoters are found mostly in tRNA genes and consist of a *box A* (similar in sequence to the A box of Type I promoters) and a B box (Fig. 2B). The B box is a conserved sequence among Type II promoters [1,7].

The discovery of Class III_{EXT} genes makes difficult the characterization of RNA pol II and RNA pol III promoters into two distinct categories based on the location of promoter elements with respect to the start site (Fig. 2C). Class III_{EXT} genes produce RNA pol III transcripts yet have no ICR [8-10]. Furthermore, the 5' flanking promoter elements resemble those of typical pol II promoters. To date, all Class III_{EXT} promoters that have been characterized contain a TATA-like element around -30 and a PSE around -60 [2]. The first examples of these genes were the vertebrate small nuclear RNA U6 gene and the 7SK RNA gene [11-16]. More recent research involving the gene for snRNA U6 has further blurred this boundary.

Small nuclear RNA U6 is a factor involved in the organization of a complex known as the spliceosome. In eukaryotes, a trademark in the progression of DNA to RNA to gene product is the presence of noncoding regions of RNA (transcribed directly from DNA) that need to be spliced out to yield a messenger RNA containing only a transcript of the coding sequence. This process of RNA maturation is directed by the spliceosome in which snRNA U6 is an integral part [17].

What is most unique about the snRNA U6 gene is that this gene in yeast exemplifies characteristics of Class III_{ICR} genes and Class III_{EXT} genes! The *S. cerevisiase* U6 gene is similar to Class III_{EXT} genes in that it contains a TATA-like element at -30 and a perfect eleven nucleotide match in the PSE region [17]. However, the yeast U6 gene also appears to have a consensus A block sequence within the coding region whose role is yet uncertain [18]. Furthermore, the most distinguishing feature of the yeast U6 gene is that a B block consensus sequence has also been identified, and it provides the first example of a pol III promoter element located downstream of the coding region. It is situated 120-130 bp beyond the actual U6 coding region. Even more intriguing is that the sequence of this 3' B block is nearly a perfect match with the Class III_{ICR} B block found in tRNA genes. In the case of the yeast U6 gene, the B block has been deemed essential for promoter function *in vivo* and *in vitro,* yet this sequence is not found in vertebrate U6 genes [18]. Finally, to confuse the boundary between classes of genes even further is recent research that has suggested TFIIIB (a transcription factor suspected in binding of the yeast U6 promoter) can be split into two components one of which may actually be **TFIID**, a previously characterized pol II transcription

factor. This TFIID is thought to bind the TATA box---the same association found in pol II protein-coding genes [19]!

The identification of Class II-type promoter elements in the 5' flanking region, the internal consensus A box, and the unique downstream B box of U6 strongly suggest the need for additional investigation into the potential of these elements to serve as sites of protein-DNA interactions. Ligation-Mediated PCR is an ideal method for observing these interactions *in vivo* due to its specific yet versatile nature. In the experiments described here, I use this technique in the study of protein-DNA interactions by manipulating the procedure to examine such interactions in the *S. cerevisiae* U6 promoter.

Ligation-Mediated PCR is used to amplify particular sites of protection or enhanced cleavage on the U6 DNA. The LMPCR technique was chosen over traditional *in vivo* footprinting techniques for several reasons. The originators of this method, Mueller and Wold [20], report that the LMPCR technique overcomes problems such as the difficulty of footprinting on single-copy sequences of DNA due to need for large numbers of cells. In addition, LMPCR also avoids the high signal-to-noise ratio often characteristic of *in vivo* footprinting. The need for LMPCR rather than standard PCR amplification arises because cleaved genomic DNA is used. The technique is centered around a key ligation step which provides uniformity to the target DNA prior to amplification [20,21]. Optimization of the LMPCR procedure with regard to the yeast system will enable the positive determination of whether suspected promoter elements do in fact bind proteins.

MATERIALS AND METHODS

Yeast strain. *Saccharomyces cerevisiae* EJ101 was used for these studies; obtained from M. Cusick (Department of Medical Biochemistry and Genetics, Texas A&M). YPD medium containing 2% dextrose, 2% peptone, and 1% yeast extract was used for culturing cells.

Oligonucleotide probes. The following DNA sequences were synthesized for use in LMPCR and various control experiments (numbers in parentheses describe the location relative to the initiation site of the *S. cerevisiae* gene):

primer1	CGT ACC ATT GCA TAG CTG TA (+362-+343)
primer 2	CTA ATA GCA TTC TTA CGC AC (+311-+292)
primer 3	CGG TTC ATC CTT ATG CAG GG (+86-+67)
primer 4	GTT TCA ACA CAG CCT GGC AT (-90- -70)
TOP	GAC CCG GGA GAT CTG AAT TC (top strand of linker)
BOT	GAA TTC AGA TC (bottom strand of linker)
Msp	GCG TAT AAA CGT GGT GTA AA

PREPARATION OF DNA FOR USE IN LMPCR

In vivo DNase treatment. *In vivo* DNase digestion was performed according to the protocol of Huibregtse and Engelke [23] with minor variations incorporated from a similar protocol described previously [23]. *S. cerevisiae* EJ101 cultures of 250 ml were grown in YPD medium to an optical density of .5 at 600 nm. The cells were

harvested and then washed in 15 ml of 40mM EDTA-90mM mercaptoethanol. Cells were resuspended in 7 ml of 1M sorbitol-1mM EDTA-3mM DTT and 2 mg/ml Zymolyase (ICN Biochemicals, 100,000 U/G) in order to prepare spheroplasts. A 25 min incubation at 30^0 C was performed with occasional swirling followed by pelleting of spheroplasts by centrifugation (3000 x g, 5 min). Spheroplasts were resuspended in 1.2 ml of lysis buffer containing 50 mM Tris (pH 7.4), 75 mM KCl, 6 mM $MgCl_2$, .5 mM $CaCl_2$, 2.5 mM 2-mercaptoethanol, and .075% Nonidet P-40. Spheroplasts were disrupted with 15 strokes of a Dounce homogenizer using a loose pestle. Aliquots of .3 ml of cell lysate were then added to pre-prepared tubes containing 30 ul DNase I at various concentrations. Final DNase I concentrations of 0, .2, .4, and .8 mg/ml (diluted from 1mg/ml in lysis buffer) were used. DNase digestion was allowed to continue for 5 min at room temperature and was then terminated by addition of .33 ml stop solution containing 2% sodium dodecyl sulfate, 1M NaCl, 50 mM EDTA, 50 mM Tris-Cl(pH 7.4), and .2mg/ml proteinase K (U.S.B., 33 Anson units/g). Reactions were incubated at 50^0 C for 30 min and extracted twice with an equal volume of phenol:chloroform (1:1). Extraction was followed by precipitation with one volume isopropanol and RNase treatment as described previously [23].

DNase treatment of purified DNA. For the sake of uniformity, purified DNA was obtained by performing the protocol just stated except that DNase digestion was delayed until the end. DNA was purified and DNase I digestion was performed on individual aliquots of DNA (22.5 ug) with .0001 mg/ml DNase I for various times,

ranging from 1 to 5 min The extent of DNase digestion was qualitatively assayed on alkaline denaturing agarose gels [24].

In vivo methylation and cleavage with piperidine. The *in vivo* methylation of yeast DNA followed by piperidine cleavage was performed according to the protocol of Giniger, et al. [25] with slight modifications. The protocol was followed exactly up to the centrifugation of nucleic acid after isopropanol precipitation. In addition to air drying nucleic acid pellets overnight, pellets were dried in a SpeedVac concentrator the following day. This further drying step resulted in easier resuspension of the relatively large pellets. RNase incubation was conducted overnight followed by a standard ethanol precipitation instead of the suggested spermidine precipitation. (Spermidine is known to inhibit blunt-ended ligations [24], a step involved in LMPCR.) Methylated DNA was then cut with Xba I in order to reduce viscosity. Typically about 10 ug of methylated DNA was used. The restriction enzyme Xba I was chosen because of the absence of Xba I sites within the coding and near flanking regions of the yeast U6 gene. DNA was then reprecipitated and cleaved with piperidine as described previously [25].

Methylation and piperidine cleavage of purified yeast DNA. Pure genomic DNA was methylated and then cleaved by the traditional Maxim-Gilbert "G reaction" for chemical sequencing as described earlier [24].

Ligation Mediated Polymerase Chain Reaction. (a complete protocol is available from Gerd P. Pfeifer et al.; Molecular Biology Section, Beckman Research Institute of the City of Hope, Duarte, CA)

The following provides a stepwise outline of the LMPCR procedure. A more detailed description has been given earlier [20].

LMPCR procedure (Fig. 4):

a. The target material of LMPCR is genomic DNA which has been cleaved enzymatically or chemically.

b. This double stranded DNA is denatured and one end of the target molecules is fixed by annealing a gene specific primer (primer 1, specific to U6).

c. Primer extension results in a population of double stranded, blunt-ended DNA molecules all with one fixed end specified by primer 1.

d. A common linker of known sequence is ligated to these blunt-ended primer extension products, thus providing DNA with known sequences at both ends.

e. The DNA is again denatured and a second gene specific primer (primer 2) is annealed. The position of this primer is slightly 5' to primer 1 to increase specificity. Primer extension is carried out resulting in a population of double stranded molecules with the linker attached.

f. Amplification by the polymerase chain reaction (PCR) is performed using the top strand of the linker and primer 2 as the DNA probes (PCR conditions: 94^0 C, 1 min; 66^0 C, 2 min; 76^0 C, 3 min + 5 sec extension per cycle; 18 cycles)

g. The final visualization of the sequence ladder is achieved by hybridization with primer 3 and autoradiography. Primer 3 is end-labelled with ^{32}P γ–ATP and placed in a position internal to the second primer on the same strand.

Preliminary test for primer binding. A Perkin Elmer Cetus DNA Thermal Cycler was used to perform PCR and LMPCR. Verification of primer binding was established by standard PCR with 50 ng of genomic DNA and 100 pmol of each primer. A fourth primer was selected to hybridize further upstream and on the opposite strand from primers 1, 2, and 3. Standard PCR was conducted with primers 1&4, 2&4, and 3&4. PCR products of expected sizes (452, 401, and 176 bp) confirmed binding capability of primers 1, 2, and 3. Primer binding tests were also performed under the PCR conditions outlined for LMPCR as well as at less stringent conditions by reducing the hybridization temperature to 50^0C and the primer extension temperature to 72^0 C.

Restriction enzyme control. To verify that the LMPCR method was working properly, a restriction enzyme digest of the purified genomic DNA was used to provide a set of control target molecules of known size. Therefore, upon conducting LMPCR, products of a particular size could be expected. Aliquots of genomic DNA were digested with the restriction endonucleases Cla I and Alu I (relevant restriction sites at -119 and -211; Fig. 3). This DNA was then pooled and used as a substrate for LMPCR.

Blocking of 3' OH groups. To improve the efficiency of ligation 3' OH groups of genomic DNA fragments were blocked. Dideoxynucleotides were added to 3' OH's by terminal dideoxynucleotidyl transferase (Pharmacia LKB Biotechnology, 21000 u/ul) and Sequenase 2.0 (U.S.B., 13u/ul) as described earlier [26].

Ligation Assay. In order to test for successful ligation, pGEM/U6 DNA (a cloned human U6 snRNA gene; obtained from G. Kunkel, Department of Biochemistry/Biophysics, Texas A&M) was cut with the blunt cutting restriction enzyme Rsa I. This digest yielded four double stranded, blunt-ended fragments of DNA which were then used in the T4-DNA ligase-catalyzed ligation to the common linker. Of the four fragments, one contains a short DNA sequence which hybridizes to the Msp oligo. This test involved one sample in which equimolar amounts of TOP and Msp probes were added to the PCR reaction mixture. The second sample served as a control in which only TOP oligo was added (the same amount of TOP oligo added in the actual "test" sample). Therefore, in the event of successful ligation, PCR using the probes TOP (top strand of the linker) and the Msp oligo should yield a PCR product of expected size (352 base pairs) in the test sample which is absent in the control.

RESULTS

Initial LMPCR experiments. Initial experiments were conducted using cell lysate and pure DNA cleaved enzymatically (see Materials and Methods). The detailed LMPCR protocol provided by Pfeifer, et al. was followed as written. An attempt was made to visualize a genomic sequence ladder by the primer extension with labelled primer 3 followed by analysis of the samples on an 8% polyacrylamide gel and autoradiography. Examination of the autoradiograph revealed no sequence ladder. Unincorporated radioactivity had migrated to its expected position, but no other bands were seen on the autoradiograph (data not shown).

Preliminary tests for primer binding. Upon obtaining the same result with a second trial, it was suspected that the DNA was not being amplified sufficiently. Preliminary tests to confirm primer binding of primers 1, 2, and 3 were performed using a fourth primer (primer 4) which hybridizes further upstream and on the opposite strand (see Materials and Methods; Fig. 3) in PCR. Standard PCR experiments were first tried using primers 2&4, since the binding of primer 2 is essential for DNA amplification. The target DNA was genomic DNA treated *in vivo* with DNase I. PCR conditions identical to those described in the LMPCR protocol were used with an increase in the number of PCR cycles (see Materials and Methods). No PCR products were seen. The size of the expected product was 401 base pairs, yet the only DNA visualized appeared to be the unincorporated oligonucleotide primers 2&4 (Fig. 5A). A similar experiment involving primer sets 1&4, 2&4, and 3&4 and the use of less stringent PCR

conditions (see Materials and Methods) yielded PCR products of expected size for primers 1&4 and 3&4 (452 and 176 base pairs). Primers 2&4 produced no PCR product (Fig. 5B). Primer sets 1&4, 2&4, and 3&4 were then tested at PCR conditions used in LMPCR. Results were exactly those of the test in which less stringent PCR conditions were used (data not shown). Inability of primer 2 to bind DNA would prevent amplification of the DNA and offers a plausible hypothesis to explain the results of initial LMPCR experiments. To correct for the inadequacy of primer 2, the PCR amplification of the DNA in the LMPCR procedure was conducted by using primers TOP and 1 in all subsequent experiments (Fig. 4).

Restriction enzyme control. Purified yeast DNA digested with Alu I and Cla I was the template and served as a positive control for LMPCR. The location of these restriction sites in relation to the position of primers 1 and 3 permits only two bands of known size to be the expected LMPCR products. A complete LMPCR experiment was carried out but failed to give expected products (317 bp for Alu I and 225 bp for Cla I; data not shown).

Blocking of 3' OH groups. Based on these preliminary results, the focus was switched from the actual amplification of the DNA to the ligation of the common linker. It was suspected that perhaps the ligation failed to occur or occurred with low efficiency. The ligation is likely to cause problems due to the nature of the reaction itself (see Discussion). Also, ligation among the DNA fragments generated from DNase I cleavage is likely because of the property of DNase I to leave a 5' phosphate group on the DNA it cuts. The presence of this phosphate group presents a problem because it is

essential for ligation to the common linker, but it also allows for ligation of DNA fragments to each other. To correct this problem, the 3' OH groups were blocked as described (see Materials and Methods) and LMPCR was then performed on restriction enzyme control DNA described above. Note that DNA extended from primer 1 will have an unblocked 3' OH. The autoradiograph revealed two bands of expected size although intensity of the Alu I band was considerably less than that of the Cla I band (data not shown). This procedure was then applied to purified yeast DNA treated with DNase I. A genomic sequence ladder was not produced (data not shown).

Ligation assay. At this point it was still thought that defective ligation was preventing amplification of the sequence ladder. A concern was the possibility that the ligation of the linker was inefficient due to incomplete annealing of the two strands of the common linker since a double stranded linker is essential for the ligation. Linker hybridization was optimized by adjusting the DNA concentrations to yield a suggested concentration of 20 pmol/ul of annealed linker (personal communication, Riggs' laboratory, Molecular Biology Section, Beckman Research Institute of the City of Hope, Duarte Cal.) and carrying out the annealing reaction for a longer time with gradual cooling. The ligation assay using pGEM/U6 plasmid DNA (see Materials and Methods) was performed to confirm successful ligation. Successful ligation was indicated by a PCR product of 352 base pairs (Fig. 6).

LMPCR of yeast DNA cleaved with DNase I or DMS/piperidine. Upon establishing optimal ligation conditions an LMPCR experiment was performed on several samples, including a restriction enzyme control (Cla I only), purified DNA treated with

DNase I or DMS/piperidine, and DNA treated *in vivo* with DNase I or DMS followed by piperidine cleavage. LMPCR was performed using a newly annealed stock of common linker and an increase of PCR cycles to 30. A sequence ladder was visualized with all samples of purified DNA. Samples of DNA treated *in vivo* either enzymatically or chemically failed to produce sequence ladder (Fig. 7). An unexpected ladder of bands appeared in the Cla I samples and matched the pattern of bands seen in puified DNA treated with DNase I (Fig. 7). In restriction enzyme control samples primer extension with labelled primer 3 was attempted with 1x and 2x concentrations of primer to determine if the labelled primer was suffering from competition from other primers. The expected band in the Cla I sample was visible and more intense in the sample visualized with 2x primer (Fig. 7).

LMPCR of yeast DNA treated with DMS/piperidine. Because the methylation pattern of the purified DNA from the previous experiment was clear, this experiment was conducted using only DNA treated with DMS/piperidine. It was suspected that absence of DNA from samples treated *in vivo* was due to minor technical problems in preparation of DNA for analysis. Minor modifications in the procedure for in vivo methylation (see Materials and Methods) resulted in a much higher yield of DNA. LMPCR revealed a sequence ladder in the case of DNA treated *in vivo* as well as purified DNA treated with DMS/piperidine. No protection pattern was seen in the region of U6 visualized in this experiment upon comparison of the band patterns of the control (pure DNA) and DNA treated *in vivo* (Fig. 8).

DISCUSSION

The process of transcription is a primary focus when discussing the regulation of gene expression. Many promoter regions have been identified due to their conserved sequences and ability to function as protein binding sites for transcription factors. The focus of this study was to use Ligation-Mediated Polymerase Chain Reaction as a technique to amplify a particular gene to aid in footprinting. The LMPCR technique is superior to conventional footprinting techniques which involve cleavage of genomic DNA and direct visualization of a sequence ladder by primer extension. LMPCR is ideal for footprinting studies on single-copy genes and is therefore the method of choice for footprinting in yeast. The LMPCR technique also offers the advantages of increased specificity conferred by the choice of primers (Fig.4) while maintaining high integrity. Integrity is not compromised because the differential starting concentrations of parts of the sequence ladder that result in a footprint are maintained throughout the amplification [20]. Specific amplification of the single-copy target DNA overcomes the technical difficulties characteristic of conventional footprinting such as poor signal-to-noise ratio and the requirement of a large number of cells.

Primer 2 aberration. Based on the nature of the results obtained after initial attempts to use LMPCR, a trouble-shooting scheme was chosen in which the LMPCR technique was divided into individual steps to be checked sequentially. As described earlier the first set of tests revealed the inability of primer 2 to hybridize to the

yeast DNA. Two possibilities were considered for the inadequacy of primer 2 to function in PCR. My first consideration was the possibility that primer 2 could form stem-loop or hairpin structures as a consequence of intramolecular base pairing. However this possibility is unlikely due to the absence of any extensive complementary sequences leading to a higher order structure of primer 2. A more likely possibility may be the existence of polymorphisms within the sequence of the U6 gene between strains of *S. cerevisiae*. The experiments described here were conducted using *S. cerevisiae* EJ101, a strain different from that on which the sequences of primers 1,2, or 3 are based. The inability of primer 2 to bind the DNA may be a consequence of DNA sequence polymorphisms in that region of the U6 gene. Supporting this hypothesis is the fact that today no true wild-type *Saccharomyces* strains are used in genetic studies. Many strains in use today are actually products of pedigrees involving mutagenic strains and therefore may contain "mutant characters" [27].

Ligation of the common linker. In the use of PCR amplification for genomic footprinting, ligation of the common linker is vital. It is the addition of the linker that provides unity among a population of DNA fragments in which one end is fixed by a gene specific primer while the other end is determined solely by enzymatic or chemical cleavage. Addition of the common linker provides each DNA segment with two ends of known sequence--the foundation on which PCR amplification is based. The crucial implications of the failure of linker ligation established this step as a focus for optimization of the LMPCR procedure. The ligation is a blunt-ended

ligation, thereby rendering this reaction more susceptible to problems concerning reaction kinetics, concentrations of components in the reaction mixture, and the DNA ends themselves. A blunt-ended ligation reaction, though catalyzed by T4 DNA ligase, occurs significantly more slowly (100x) than a nick-sealing reaction due to the absence of cohesive ends. The slower reaction kinetics have been explained by the inability of blunt-ended DNA fragments to anneal, making the reaction dependent on the amount of time the 5' phosphate and the 3' OH of the fragments to be joined are in close proximity and in the correct orientation for ligation. The problem of slower reaction kinetics is compounded by the requirement of relatively specific concentrations of ligase, ATP, and a critical concentration of the ligating termini [28]. The concentration of ligating termini (DNA ends available for ligation to the linker) was a particular concern because of the likelihood of DNA fragments ligating to themselves (see Results).

As explained earlier, a likelihood exists for the ligation to be very inefficient due to the presence of 5' phosphate groups left by DNase I cleavage. A method described for blocking 3' OH groups prior to LMPCR (see Materials and Methods) initially appeared to solve the problem of inefficient ligation at least in the case of the restriction enzyme control (see Results). However, addition of the blocking step did not result in amplification of a sequence ladder with yeast DNA treated with DNase (see Results). To explain these results I offer two hypotheses. It is possible that the two bands expected in the restriction enzyme control can appear without sufficient amplification of the DNA. The small genome size of yeast adds merit to this

hypothesis because although U6 is a single-copy gene, it makes up a larger fraction of the yeast genome than that of a single-copy gene in the genome of a higher eukaryote. This explanation is unlikely because a control experiment was carried out in which the same DNA was used in a primer extension with labelled primer 3, and no bands were visible. A second hypothesis is that ligation was occurring with very low efficiency, resulting in a reduced number of target molecules for PCR. Low efficiency could explain why when only two bands are expected, they are visible, but when an entire ladder is expected, there is not enough DNA per band to be visible.

Failure of LMPCR to produce a sequence ladder focussed attention on the DNA end structure to promote proper annealing of the linker. Proper annealing is essential for providing a double stranded common linker substrate for ligation. The annealing reaction is dependent on the concentration of the single stranded DNA fragments, the sequence complementarity between them, and a minimum salt concentration [1,24]. Because of the specific requirements of this reaction it became the focus of my troubleshooting approach. Optimization of the annealing procedure in terms of the concentrations of each strand in the reaction mix as well as other minor modifications resulted in successful ligation (see Results).

LMPCR using DNA cleaved enzymatically (DNase I) or Chemically (DMS/piperidine). Upon confirming successful ligation an LMPCR experiment involving enzymatically and chemically cleaved DNA was performed. As mentioned earlier, this experiment also included two Cla I control DNA samples---one using 1x primer 3 and

one using 2x primer 3 in the final primer extension. Results of that part of the study indicate that primer 3 does suffer somewhat from competition from other primers left over in the reaction mixture; this is shown by the increase in intensity of the expected Cla I band in the sample with 2x primer 3. However, this level of competition is low enough to not affect visualization of the sequence ladder significantly. The sequence ladders produced from DNA cleaved chemically and enzymatically contained slightly different band patterns, as expected, but it is unclear why extra bands appearing in the Cla I lanes seem to align perfectly with the bands seen in the sequence ladder of pure DNA treated with DNase I. Possible explanations for the extra bands in the Cla I sample include DNA degradation or incomplete polymerization during the last primer extension step.

LMPCR using chemically cleaved DNA. Because LMPCR with the DMS-treated DNA appeared to give a pattern of bands distinct from the DNase-treated or Cla I-digested DNA, the final LMPCR experiment was performed only on chemically cleaved DNA. Optimization of the *in vivo* methylation procedure resulted in the success of LMPCR in producing a sequence ladder for *in vivo* and pure DNA treated chemically. Comparison of the band pattern with the published yeast U6 sequence [18] revealed that methylation and cleavage had been specific for G residues. Identification of each band allowed for delineation of putative transcriptional control elements. For example, an A/T rich region in the area of the TATA box was confirmed. The data displayed no differences between the pure DNA sequence ladder and the *in vivo* methylated samples. Two points are

worthy of consideration to explain this result. First, in order for this result to be accepted the experiment must be repeated more times and both strands of the DNA must be probed for footprints, not only the non-coding strand. It is quite possible that footprints are not visible on the non-coding strands because the majority of the protein-DNA interactions occur on the opposite strand. Secondly, it is important to note that optimization of footprinting using DNA treated with DNase I is absolutely necessary. Methylation of DNA followed by piperidine cleavage is not advantageous for identifying footprints over the TATA-like sequence due to the absence of G residues. A/T-rich regions will appear identical in control lanes (no protein) and *in vivo* methylated DNA.

Significance of PCR-aided Genomic Footprinting in Yeast to Higher Eukaryotes. In the studies described here, I have optimized the procedure of LMPCR for the study of protein-DNA interactions on the yeast U6 promoter and given evidence of the utility of LMPCR in such studies. As I started my study it appeared that the path to identifying protein-DNA interactions on the *S. cerevisiae* U6 promoter by this method was a clear one. However, as I began my research I learned that as important as individual scientific discoveries themselves is the virtue that circumstances are not always the way they seem, thus providing the foundation of science. Applying this valuable lesson, the focus of this project changed from the identification of protein-DNA interactions using LMPCR to actually dissecting the procedure itself and customizing it for the system of interest.

Studies in the regulation of yeast U6 transcription have left many questions unanswered. Do the PSE and the TATA-like motifs in the U6 5' flanking regions serve as protein binding sites? Does the internal A box consensus sequence play an active role in transcription regulation by also binding protein(s)? Is the B box involved in similar regulatory activities though its location is unique? Further optimization of Ligation-Mediated PCR will be advantageous in elucidating transcriptional control elements in the yeast U6 promoter. Identification of these elements will lead to knowledge of this control mechanism and may build on the already existing evidence for the evolution of regulatory mechanisms in transcription.

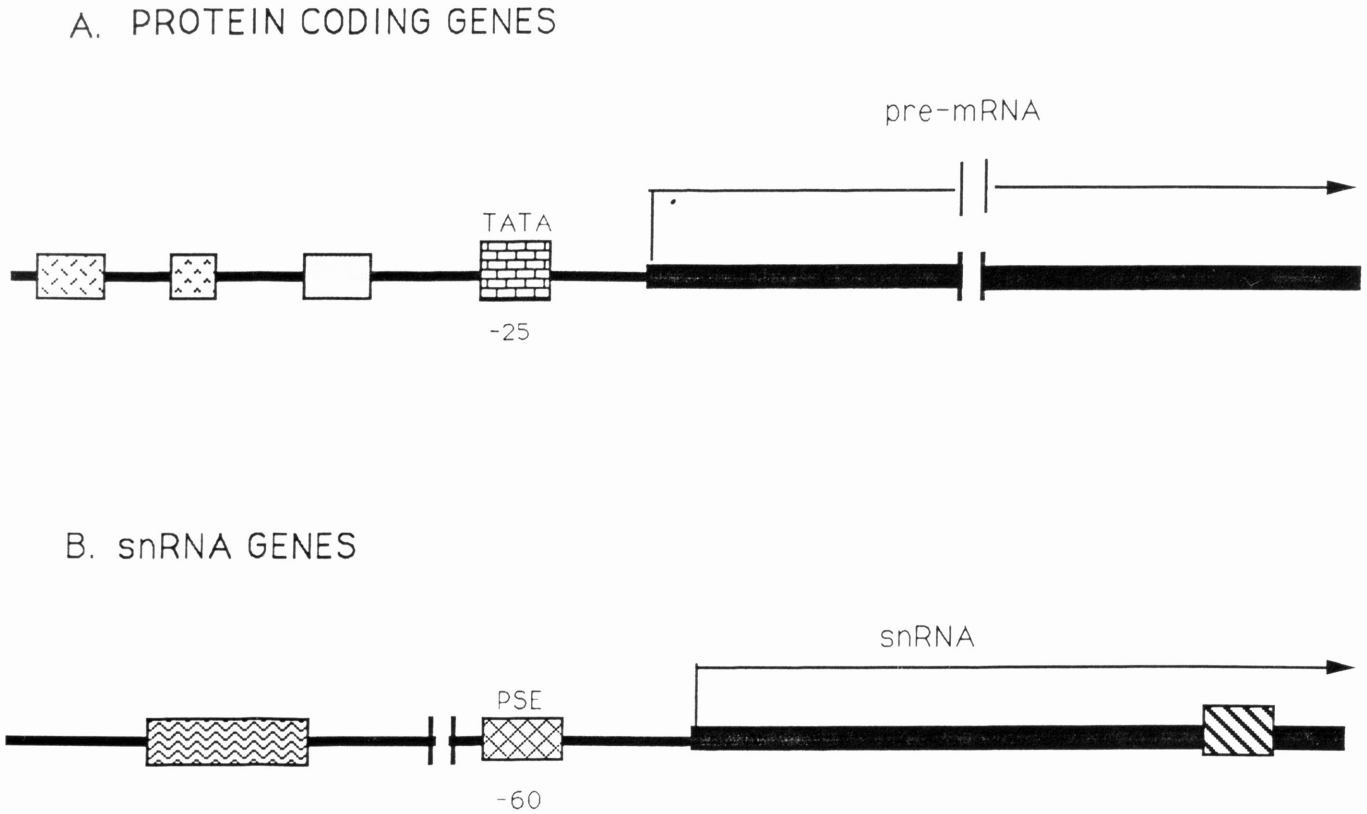
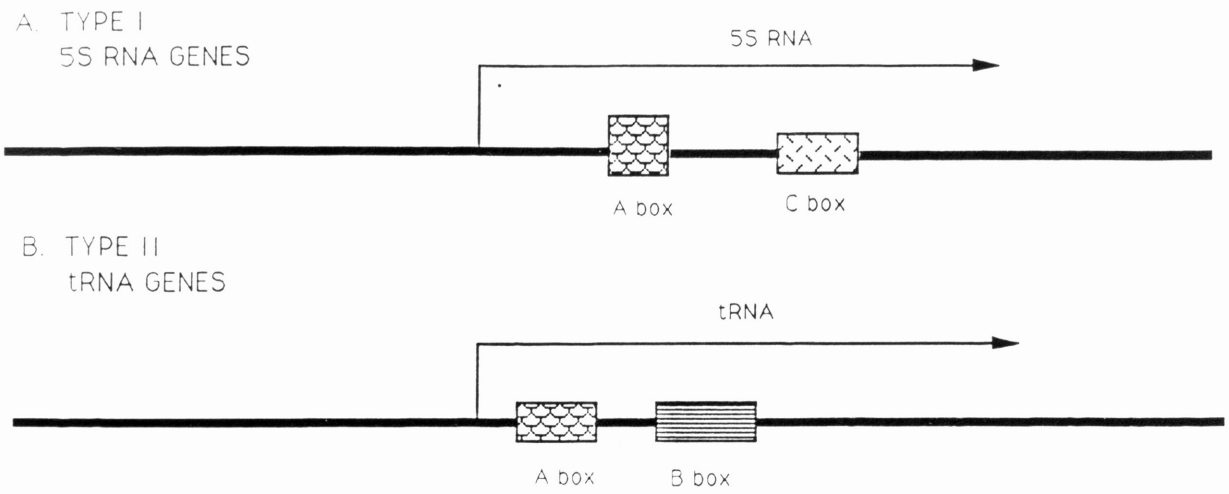


Figure 1 Typical promoters of Class II genes

A Promoter region of protein coding genes, displays the TATA box--a module with a T/A -rich sequence. The TATA box represents the only promoter element with a relatively fixed position.

B Promoter region of snRNAs (U1 and U2)--appearance of a new module, the proximal sequence element (PSE). The PSE functionally replaces the TATA box which is absent from these promoters.

Class III-ICR GENES



Class III-EXT GENES

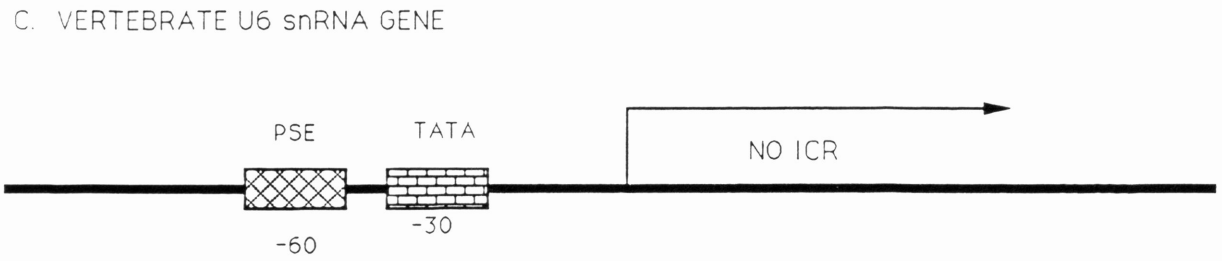


Figure 2 Promoters of Class III genes

A. Type I promoters are found in 5S RNA genes. They have an internal control region (ICR) comprised of two consensus sequences, box A and box C.

B. Type II promoters are found in tRNA genes. Box A sequence is conserved between Type I and Type II promoters. Also a box B is present.

C. Class III-ext genes---newly characterized Class III genes having no ICR. Typical Class II gene promoter elements are seen.

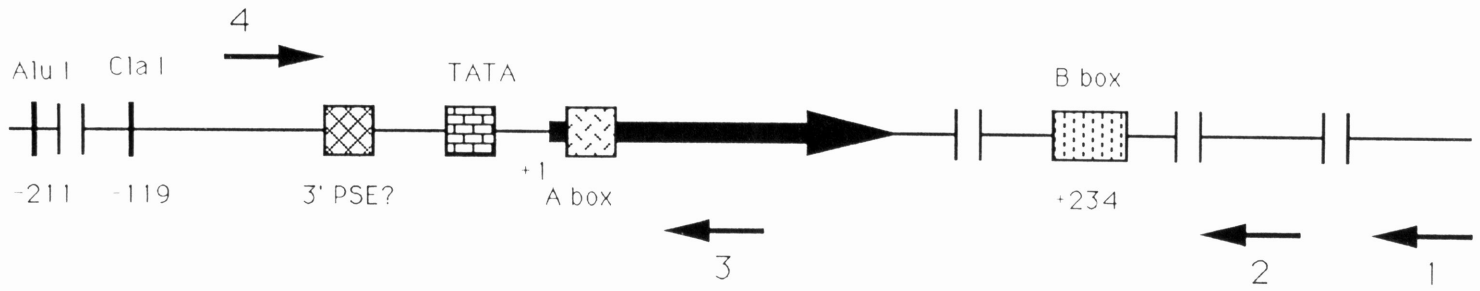
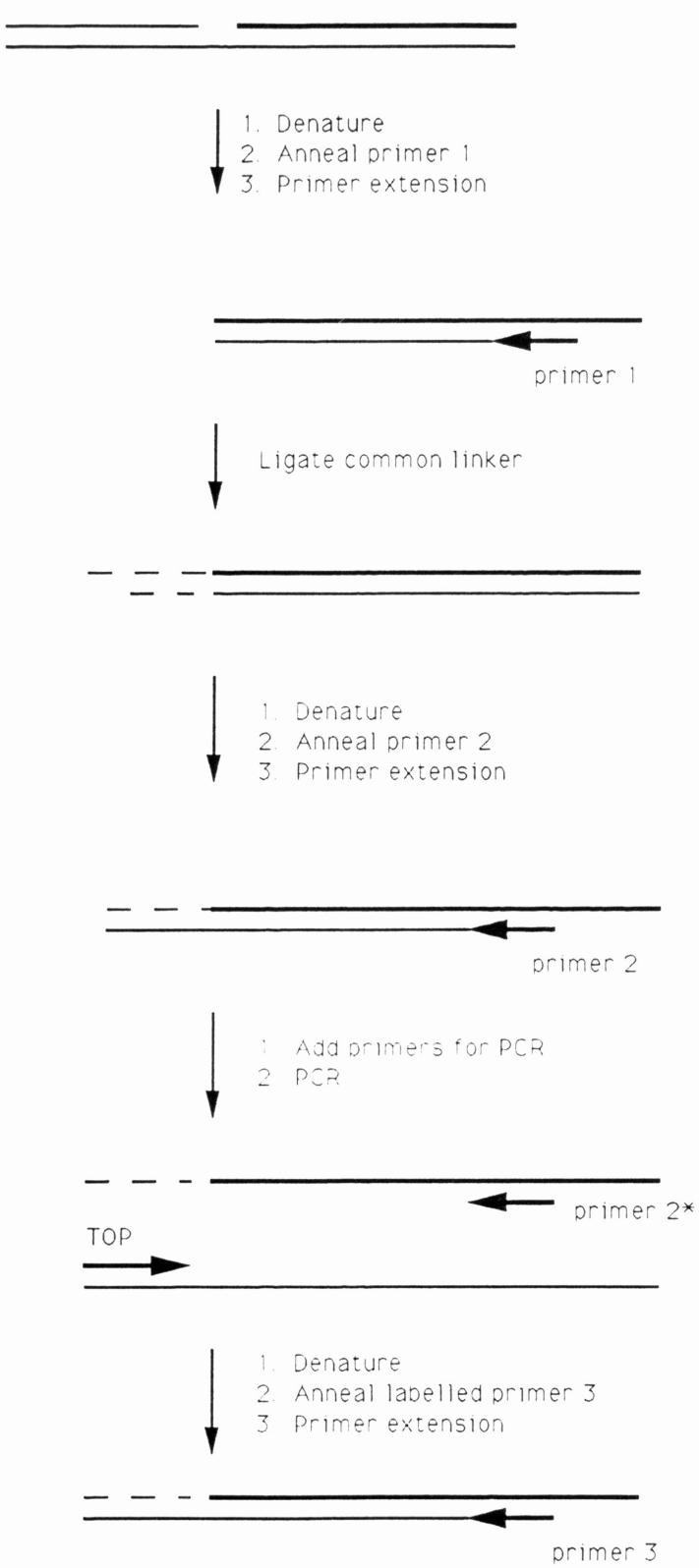


Figure 3. *S. cerevisiae* U6 gene

This gene contains examples of promoter elements typical of Class IIIext genes, such as a possible PSE and a TATA box. Also present is a putative A box as well as the first downstream promoter element, the B box. Arrows 1 (+343-+362), 2 (+292-+311), and 3 (+67-+86) indicate primers used in LMPCR. Arrow 4 (-90--70) is the additional primer used in the preliminary test for primer binding described in Figure 5. The Cla I restriction site at position -119 and the Alu I site at -211 are those sites described in the restriction enzyme control explained earlier.



*replaced by primer 1 (see Results)

Figure 4. Ligation-Mediated Polymerase Chain Reaction

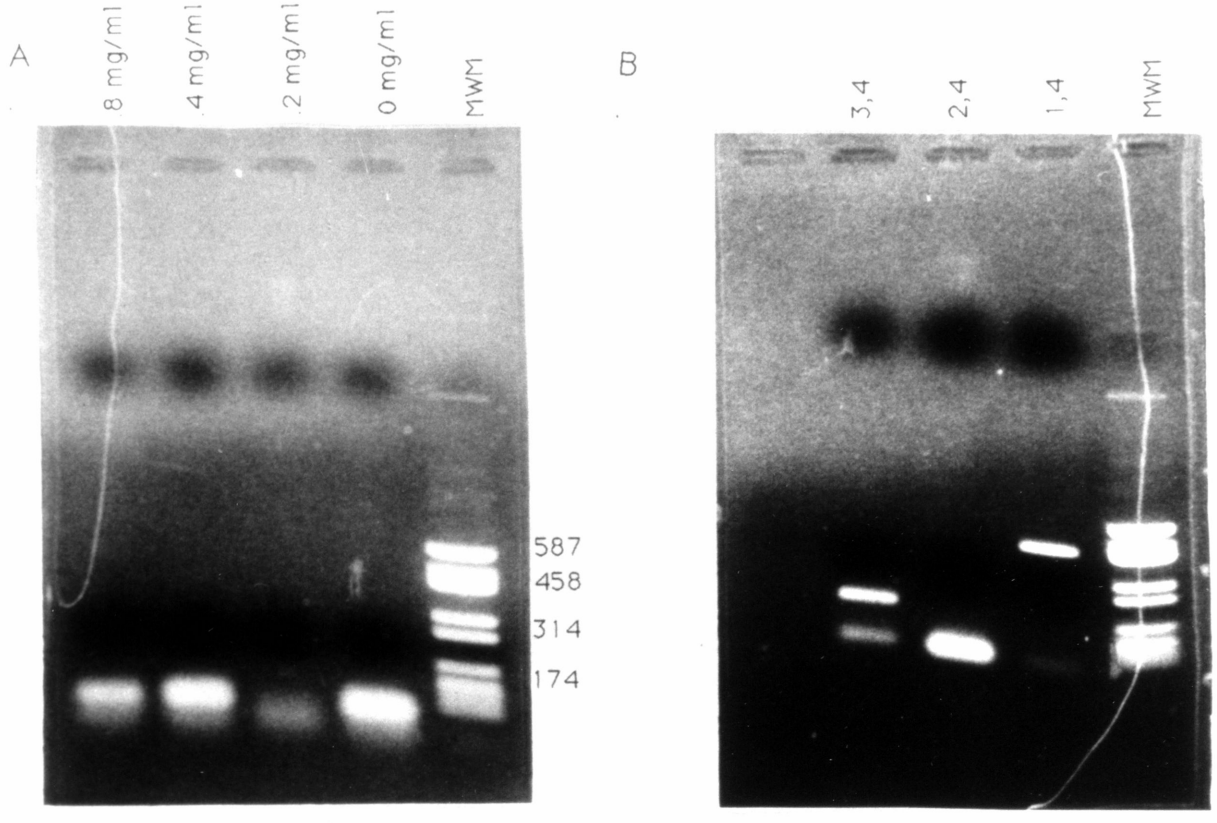


Figure 5. Preliminary test for primer binding

A. Standard PCR was conducted on genomic DNA treated with various concentrations of DNase I. No PCR products were observed.

B. Standard PCR was conducted with primer sets 1&4, 2&4, and 3&4 on genomic DNA treated with .4 mg/ml DNase I. PCR products of expected size were seen with primer sets 1&4 3&4. No product was seen with primers 2&4.

1% agarose gels; DNA visualized by ethidium bromide staining.

Molecular weight marker (MWM): pGEM 3ZF(-)/Hae III: 587, 487, 458, 314, 267, 174, 142, 125.

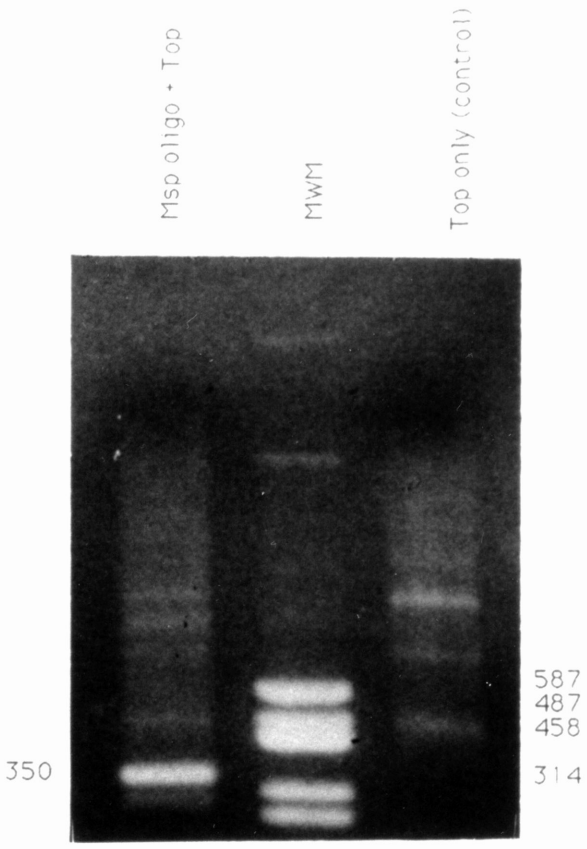


Figure 6. Confirmation of Ligation

1% agarose gel; DNA visualized by ethidium bromide staining.

pGEM U6 plasmid DNA was cut using the blunt cutter Rsa I to yield 4 restriction fragments. One of these fragments contained a site which would hybridize to the Msp oligo. Ligation of the DNA and common linker was carried out as described in the LMPCR procedure. PCR was conducted using oligos Top and Msp I to test for ligation. Only Top was added to the control. Successful ligation is shown by the presence of a PCR product of 352 base pairs.

Molecular weight marker (MWM): pGEM 3ZF(-)/Hae III: 587, 487, 458, 434, 314, 267.

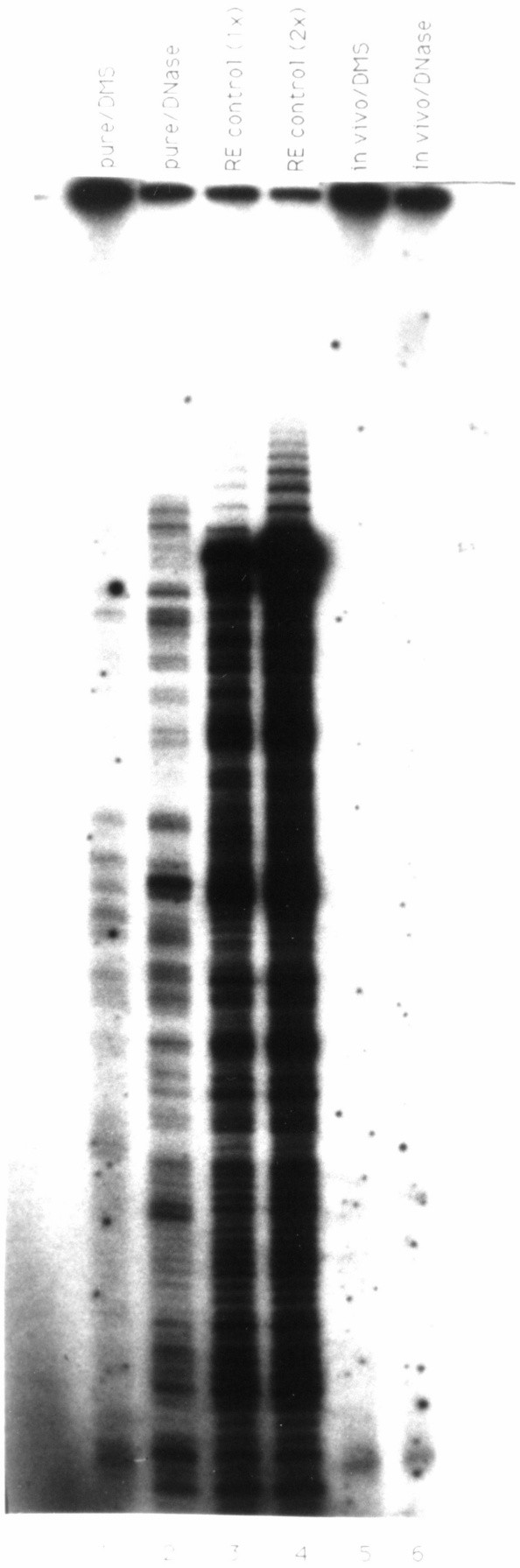


Figure 7. LMPCR of Yeast Genomic DNA

Autoradiograph of polyacrylamide gel

Lane 1. Purified genomic DNA was methylated using the "G reaction" of the Maxam and Gilbert chemical sequencing method and cleaved with piperdine prior to LMPCR

Lane 2. Purified genomic DNA was treated with .0001 mg/ml DNase for 2.5 min. prior to LMPCR.

Lane 3. Before LMPCR, pure genomic DNA was cut with Cla I (during the final hybridization in LMPCR, 1x labelled primer was used).

Lane 4. Pure DNA cut with Cla I (2x labelled primer was used in LMPCR).

Lane 5. Genomic DNA was methylated in vivo, digested with Xba I, and cleaved with piperdine prior to LMPCR.

Lane 6. Spheroplasts were lysed and treated with .4 mg/ml DNase I for 5 min prior to LMPCR.

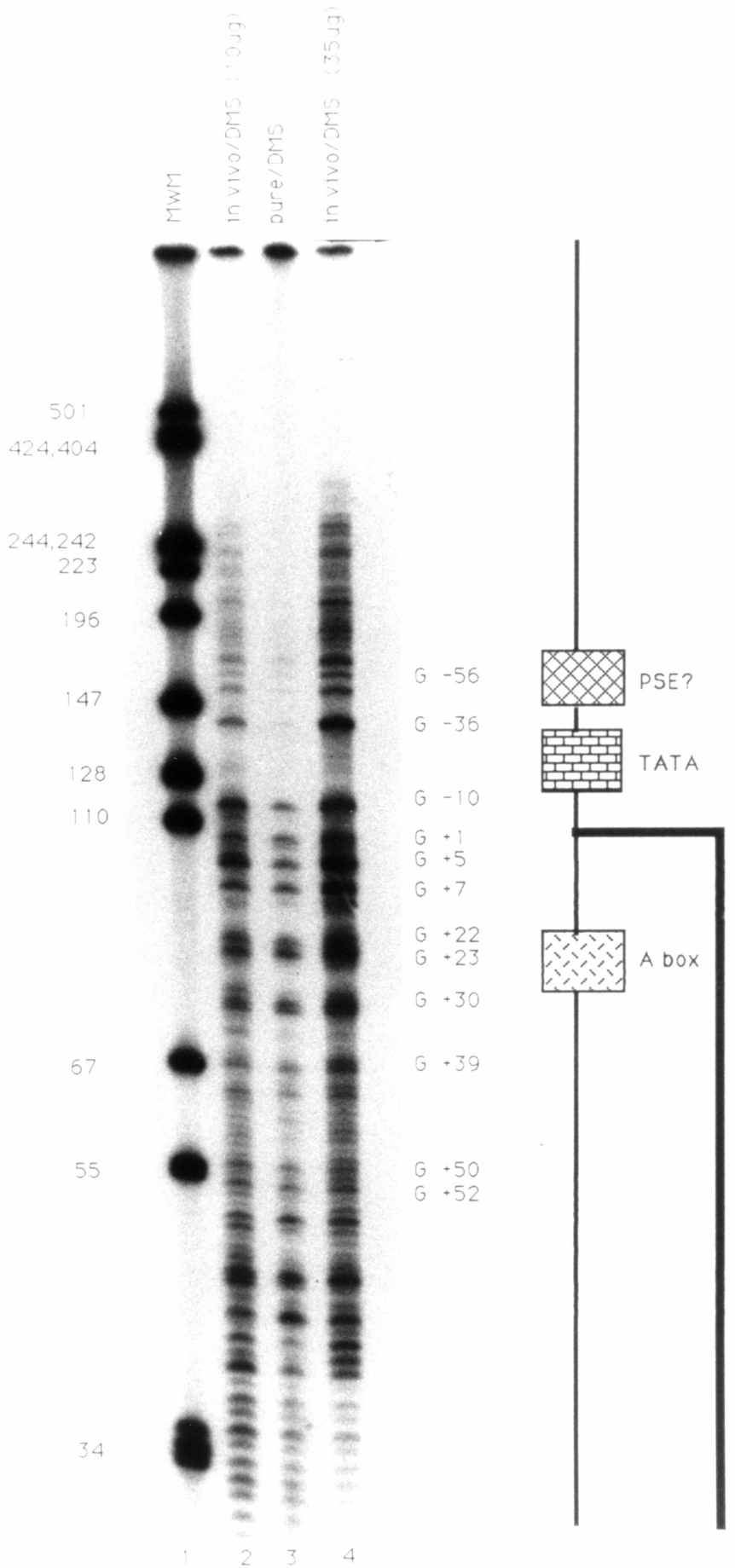


Figure 8 LMPCR of Yeast Genomic DNA

Autoradiograph of polyacrylamide gel

Lane 1. Molecular weight marker pGEM/ Msp I

Lane 2. Genomic DNA (10 ug) methylated in vivo, digested with Xba I, and cleaved with piperdine prior to LMPCR.

Lane 3. Purified genomic DNA was methylated using the "G reaction" of the Maxam and Gilbert chemical sequencing method and cleaved with piperdine prior to LMPCR.

Lane 4. Genomic DNA (35 ug) methylated in vivo, digested with Xba I, and cleaved with piperdine prior to LMPCR.

REFERENCES

1 Lewin, B. (1990) *Genes IV*, Ch. 29, Oxford University Press, New York and Cell Press, Cambridge, Mass.

2 Kunkel, Gary R. (1991) *Biochimica et Biophphysica Acta* 1088, 1-9.

3 Skuzeski, J.M., Lund, E., Murphy, J.T., Steinberg, T.H., Burgess, R.R. and Dahlberg, J.E. (1984) *J. Biol. Chem.* 259, 8345-8352.

4 Hall, B.D., Clarkson, S.G. and Tocchini-Valentini, G. (1982) *Cell* 29, 3-5.

5 Ciliberto, G., Castagnoli, L. and Cortese, R. (1983) *Curr. Top. Dev. Biol.* 18, 59-88.

6 Sharp, S.J., Schaak, J., Cooley, L., Burke, D.J. and Soll, D. (1985) *CRC Crit. Rev. Biochem.* 19, 107-44.

7 Geiduschek, E.P. and Tocchini-Valentini, G.P. (1988) Ann. Rev. *Biochem.* 57, 873-914.

8 Sollner-Webb, B. (1988) *Cell* 52, 153-154.

9 Folk, W.R. (1988) *Genes Dev.* 2, 373-375.

10 Murphy, S., Moorefield, B. and Pieler, T. (1989) *Trends Genet.* 5, 122-126.

11 Krol, A., Carbon, P., Ebel, J.-P. and Appel, B. (1987) *Nucleic Acids Res.* 15, 2463-2478.

12 Murphy, S., Di Liegro, C. and Melli, M. (1987) *Cell* 51, 81-87.

13 Carbon, P., Murgo, S., Ebel, J.-P., Krol, A., Tebb, G. and Mattaj, I.W. (1987) *Cell* 51, 71-79.

14 Kleinhart, H. and Benecke, B.J. (1988) *Nucleic Acids Res.* 16, 1319-1331.

15 Das, G., Henning, D., Wright, D. and Reddy, R. (1988) *EMBO J.* 7, 503-512.

16 Kunkel, G.R. and Pederson, T. (1988) *Genes Dev.* 2, 196-204.

17 Brow, D.A. and Guthrie, C. (1988) *Nature* 334, 213-218.

18 Brow, D.A. and Guthrie, C. (1990) *Genes Dev.* 4, 1345-1356.

19 Margottin, F., Dujardin, G., Gerard, M., Egly, J.-M., Huet, J. and Senentac, A. (1991) *Science* 251, 424-426.

20 Mueller, P.R. and Wold, B. (1989) *Science* 246, 780-786.

21 Pfeifer, G.P., Steigerwald, S.D., Mueller, P.R., Wold, B. and Riggs, A.D. (1989) *Science* 246, 810-813.

22 Huibregtse, J.M. and Engelke, D.R. (1989) *Mol. Cell. Biol.* 9, 3244-3252.

23 Huibregtse, J.M., Evans, C.F. and Engelke, D.R. (1987) *Mol. Cell. Biol.* 7, 3212-3220.

24 Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989) *Molecular Cloning,* 2nd ed., Cold Spring Harbor Laboratory Press, New York.

25 Giniger, E., Varnum, S.M., and Ptashne, M. (1985) *Cell* 40, 767-774.

26 Tanguay, R.L., Pfeifer, G.P., and Riggs, A.D. (1990) *Nucleic Acids Res.* 18, 5902.

27 Sherman, F. (1991) in *Guide to Yeast Genetics and Molecular Biology, Methods in Enzymology*, vol. 194 (Guthrie, C. and Fink, G.R., eds.), pp. 3-21, Academic Press, Inc., San Diego, Cal.

28 Rodriguez, R.L. and Tait, R.C. (1983) *Recombinant DNA Techniques: An Introduction,* pp. 86-87, Addison-Wesley Publishing Co., Reading, Mass.