Structure of the Human Gene Encoding the Protein Repair L-Isoaspartyl (D-Aspartyl) O-Methyltransferase¹

Christopher G. DeVry, William Tsai, and Steven Clarke²

Department of Chemistry and Biochemistry and Molecular Biology Institute, University of California, Los Angeles, California 90095-1569

Received July 16, 1996, and in revised form August 29, 1996

The protein L-isoaspartyl/D-aspartyl O-methyltransferase (EC 2.1.1.77) catalyzes the first step in the repair of proteins damaged in the aging process by isomerization or racemization reactions at aspartyl and asparaginyl residues. A single gene has been localized to human chromosome 6 and multiple transcripts arising through alternative splicing have been identified. Restriction enzyme mapping, subcloning, and DNA sequence analysis of three overlapping clones from a human genomic library in bacteriophage P1 indicate that the gene spans approximately 60 kb and is composed of 8 exons interrupted by 7 introns. Analysis of intron/exon splice junctions reveals that all of the donor and acceptor splice sites are in agreement with the mammalian consensus splicing sequence. Determination of transcription initiation sites by primer extension analysis of poly(A)⁺ mRNA from human brain identifies multiple start sites, with a major site 159 nucleotides upstream from the ATG start codon. Sequence analysis of the 5'-untranslated region demonstrates several potential *cis*-acting DNA elements including SP1, ETF, AP1, AP2, ARE, XRE, CREB, MED-1, and half-palindromic ERE motifs. The promoter of this methyltransferase gene lacks an identifiable TATA box but is characterized by a CpG island which begins approximately 723 nucleotides upstream of the major transcriptional start site and extends through exon 1 and into the first intron. These features are characteristic of housekeeping genes and are consistent with the wide tissue distribution observed for this methyltransferase activity.

© 1996 Academic Press, Inc.

Key Words: methyltransferases; protein aging; protein repair.

Precisely functioning proteins are generated with a high degree of fidelity by the cellular transcriptional and translational machinery. Once synthesized, however, the proteins are immediately subjected to various spontaneous chemical degradative reactions that can affect their biological activity (1, 2). One such degradative process results in the deamidation of asparaginyl residues and the isomerization and racemization of aspartyl residues, giving rise to D- and L-isoaspartyl and D-aspartyl derivatives (3, 4). The presence of these altered aspartyl residues can lead to protein inactivation (5–7). Thus, the ability of cells to recognize and repair or proteolytically degrade such damaged proteins may represent an evolutionary adaptation for long-term survival.

One enzyme that appears to play a crucial role in the repair of damaged aspartyl and asparaginyl residues is the protein L-isoaspartyl/D-aspartyl O-methyltransferase (EC 2.1.1.77). This enzyme catalyzes the transfer of a methyl group from S-adenosylmethionine to the α -carboxyl group of L-isoaspartyl residues in a variety of prokaryotic and eukaryotic organisms and the β -carboxyl group of D-aspartyl residues in mammals (for a review see Ref. 8). Methylation of an L-isoaspartyl residue in synthetic peptides leads to the reformation of a succinimidyl intermediate and has been shown to result in the net reconversion to the L-aspartyl residue (9, 10; Fig. 1). The role of this enzyme in protein repair is supported by recent studies showing that the recombinant human methyltransferase can convert the deamidated, impaired form of the bacterial phosphocarrier protein HPr, containing isoaspartyl residues, to a form with normal aspartyl residues at these sites and enhanced phosphotransferase activity (11). Earlier studies had demonstrated the regeneration of active

¹ This work was supported by Grant GM-26020 from the National Institutes of Health and the philanthropy of Kay Kimberly Siegel and the Siegel Life Project of the UCLA Center on Aging. C.G.D. is supported by United States Public Health Service Training Grant GM-07185.

² To whom correspondence should be addressed at the Department of Chemistry and Biochemistry, UCLA, Los Angeles, CA 90095-1569. Fax: (310) 825-1968. E-mail: clarke@ewald.mbi.ucla.edu.

FIG. 1. Role of the protein L-isoaspartate (D-aspartate) *O*-methyltransferase (PCMT1 gene product) in the conversion of abnormal L-isoaspartyl residues to L-aspartyl residues. L-Isoaspartyl residues, the predominant product of the spontaneous degradation of L-aspartyl and L-asparaginyl residues, are recognized at high affinity by this methyltransferase to initiate the repair process (see text).

calmodulin by incubation of age-damaged protein with the bovine methyltransferase (12). Deletion of the gene encoding L-isoaspartyl methyltransferase in *Escherichia coli* results in methyltransferase-deficient mutants which are abnormally sensitive to heat shock and survive poorly in stationary phase, a period with little or no new protein synthesis (13). This suggests that an accumulation of proteins with altered aspartyl residues may be detrimental, and that the methyltransferase may normally limit the buildup of these proteins.

We have thus been interested in the role of this enzyme in the pathophysiology of the human aging process. Previously, a single gene encoding the human methyltransferase (PCMT1)³ was mapped to the q22.3-q24 region of human chromosome 6 (14). Recently, we have found at least three polymorphic sites within the gene that result in amino acid changes and may be important in the function or stability of the enzyme (15). In this study, we have obtained and analyzed three genomic clones encoding the human methyltransferase and have characterized the structure of the gene by restriction mapping the region, establishing exon/intron splice junctions, and analyzing the 5'-untranslated region for transcriptional start sites and promoter elements.

MATERIALS AND METHODS

Oligonucleotide primer synthesis and purification. Primers were synthesized using β -cyanoethyl *N,N*-diisopropylphosphoramidite chemistry in a Gene Assembler Plus DNA synthesizer (Pharmacia LKB Biotechnology). After synthesis, the DNA was hydrolyzed from the solid support by incubation in 1 ml of 14.8 M ammonium hydroxide for 16 h at 55°C (16) and precipitated from the solution using sodium acetate (17). The sequences of the primers used are: E1F (5'-CTCCGAGTGTGCTTAGCGATGGCCTGGAAATCCGGCGGCGC-3'), E1F' (5'-TGTGCTTAGCGATGGCCTGGAAATC-3'), E1R (5'-TACAGAACCACCTTCAGCGCGACGA-3'), E2F (5'-CCGCTCCACTATGCAAAATGTAA-3'), E2R (5'-CTGTAGCCAGCATCACTTCAAATAC-3'), E3F (5'-GTTTCCAAGCAACAATCAGTGCTCC-3'), E3F' (5'-TCAGTGCTCCACACATG-3'), E3R (5'-GATTGTGCTTGGAAC-3'), I3F (5'-TATTGTGAAAAATAACGTATGGA-3'), I3R (5'-ACATAACTTATAGATGATAAA-3'), E4F (5'-GCGCTAGAACTTCTATTGATCAGTTG-3'), E4F' (5'-TCTTGATGTAGGATCTGGAAGTGG-3'), E4R (5'-GCTTTAGCTCCTTCATGCAACTG-3'), E5F (5'-GTTGGATGTACTGGAAAAGTCATAGG-3'), E5F' (5'-AATGTCAGGAAGGACGATCCAACAC-3'), E5R (5'-CTGAGTCATCTAGCTCTTTAATG-3'), E5R' (5'-CAAGCTGTACTCTCCCTGAAGACAG-3'), E6F (5'-GGGATGGAAGAATGGGATATGCTGA3'), E6F' (5'-ATGCCATTATGTGGGAGCTGCA-3'), E6R (5'-CATAAGGGGCTTCTTCAGCATATC-3'), E6R' (5'-GGTACACAGGGGCTGCAGC-3'), E7aF (5'-GTTAAAGCCCGGAGGAAGATTGATATG-3'), E7aF' (5'-AAATGAAGCCTCTGATGGG-3'), E7aR (5'-TCCAACATTTGGTTTCCGCC-3'), E7aR' (5'-GGACCACTGCTTTTCTTATCTG-3'), E8F (5'-ATTACTTTAACATGCCCATATT-3'), E8F' (5'-GATGGCAGGTGATGTCTGTAA-3'), E8R (5'-GCTGTGATGGTGTGTTTTC-3'), E8R' (5'-CAATGCACA-AAGCAATCTGAT-3').

Isolation and characterization of the human PCMT1 genomic clones. Using exon 7a-specific primers E7aF and E7aR' (see above) supplied by our laboratory, Genome Systems Inc. (St. Louis, MO) screened the DuPont-Merck Pharmaceutical Company human foreskin fibroblast genome bacteriophage P1 library by PCR analysis (18). These primers would give a PCR product corresponding to nucleotides 515-671 of a full-length cDNA clone (pDM2) of the isoaspartyl methyltransferase previously obtained from a human

³ Abbreviations used: PCMT1, gene designation for the protein L-isoaspartate (D-aspartate) *O*-methyltransferase; bp, base pairs; kb, kilobase pairs; dNTPs, deoxyribonucleotide triphosphates; PCR, polymerase chain reaction; EST, expressed sequence tag; ARE, antioxidant response element; CREB, cyclic AMP response element binding protein; ERE, estrogen receptor response element; MED-1, multiple start site element downstream; XRE, xenobiotic response element.

brain library (19). Three P1 clones were obtained and were designated as clone 658 (DMPC-HFF 1-0662A9), clone 659 (DMPC-HFF 1-1015G6), and clone 660 (DMPC-HFF 1-1382E9).

P1 DNA was isolated using a modified Qiagen plasmid preparation procedure. Bacteria containing these clones were grown overnight in 10 ml of Luria-Bertani (LB) broth containing 25 μ g/ml kanamycin monosulfate (Sigma). This culture was then transferred to 1 liter of LB broth with 25 μ g/ml kanamycin monosulfate; after 1 h of growth at 37°C, 100 μ l of 1 M isopropyl- β -D-thiogalactopyranoside (Sigma) was added in order to induce multiple copies of the P1 plasmid. Growth at 37°C was continued for approximately 5–6 h, to an OD₆₀₀ of 0.8–0.9. Cells were subsequently pelleted and resuspended in 13.5 ml of Qiagen buffer P1. Chicken egg white lysozyme (1.5 ml of a 10 mg/ml solution dissolved in P1 buffer) (Sigma) was then added and incubated 5 min at room temperature. The remainder of the procedure followed the Qiagen Handbook for the Plasmid Mega Kit.

The three P1 clones were characterized by PCR and Southern analysis. The 50- μ l PCR reaction mixture contained P1 plasmid (100–200 ng), 40 pmol of each of primer, 0.2 mM of each of the four deoxynucleotide triphosphates (dNTPs), 2 mM of MgCl₂, 1 $\times$ reaction buffer (Promega), and 2.5 U *Taq* DNA polymerase (Promega). The PCR cycling conditions for each amplification were 95°C for 2 min (without enzyme) and then 30 cycles at 95°C for 32 s and 58°C for 32 s. Southern analysis was performed according to the procedure of Sambrook *et al.* (17). P1 plasmid template DNA (2 μ g) was digested with both the *Eco*RI (50 U) and *Pst*II (50 U) restriction endonucleases at 37°C for 5 h. The digested fragments were subsequently separated on a 0.6% agarose gel and transferred onto an Immobilon-NC membrane (Millipore). Oligonucleotide probes were prepared from primers E1F, E2R', E3F, and E4F by end-labeling to a specific activity of 4–5 $\times 10^8$ cpm/mg with [γ -³²P]ATP (ICN Biomedicals). Membranes were prehybridized at 65°C for 3 h in 5 $\times$ SSPE, 5 $\times$ Denhardt's solution, 0.5% SDS, and 100 μ g/ml denatured salmon sperm DNA. Hybridization was performed at 65°C overnight with denatured probe and fresh solution. Membranes were washed with 6 $\times$ SSPE and 0.1% SDS at room temperature for 15 min, followed by 4 $\times$, 2 $\times$, and 1 $\times$ SSPE concentrations as required.

Inverse polymerase chain reaction amplification. To determine both the flanking splice-site sequences and intron sequences, we employed the inverse polymerase chain reaction method as described (20) with the following modifications. During the intramolecular ligation of the restriction enzyme-digested plasmid fragments, the reaction was diluted 1:10 and incubated at 37°C for 1.5 h. After intramolecular ligation, direct PCR amplification using exon-specific primers was performed without any additional cleavage. Subsequently, the resulting linear template flanked by the exon region was amplified. DNA sequencing of the template, as described below, was performed using the same primers used to generate the inverse PCR products, with the exception of the exon 3 intron boundaries, where intron-specific primers I3F and I3R were used for sequencing.

Restriction site mapping of the PCMT1 gene. P1 clones 658 and 659 were digested with *Bam*HI, *Eco*RI, *Hind*III, and *Pst*II; random fragments were subcloned into pBluescript SK⁺ II (Stratagene) and then transformed into competent *Escherichia coli* DH5 α cells (Gibco-BRL). The subclones were screened for specific exons in one of two ways. The first method involved PCR amplification with exon-specific primer sets. Positive subclones (20 from each digest) were isolated, the plasmids were prepared using the Wizard minipreps DNA purification system (Promega) and then used as templates for the PCR reaction. The reaction mixture contained 1 $\times$ reaction buffer (Promega), 2 mM MgCl₂, 0.2 mM dNTPs, 20 pmol of each primer, 100–200 ng DNA, and 2.5 U *Taq* DNA polymerase (Promega). The amplification procedure began with a "hot start" at 95°C for 2 min (without enzyme), and the addition of the polymerase was followed by 25 cycles of 95°C for 30 s, 50°C for 30 s, and 72°C for 1 min. The products were then analyzed by agarose gel electrophoresis. The second method utilized a colony hybridization screening technique (17). Positive subclones (50 from each digest) were isolated and patched onto a

new plate in a grid pattern. After lysis and transfer to a nitrocellulose membrane (17), screening was carried out by hybridization to a ³²P-end-labeled oligonucleotide probe specific for each exon. The membranes were hybridized in 5 $\times$ SSPE, 5 $\times$ Denhardt's solution, 0.5% SDS, and 100 μ g/ml denatured salmon sperm DNA overnight at 55°C (17) and then subjected to a washing regimen that began with 2 $\times$ SSC containing 0.1% SDS at room temperature and was repeated with stepwise lower salt concentrations as needed. Positives were confirmed by digesting the DNA with the restriction enzyme used to generate the subclone, allowing for the identification of the pBluescript vector and the subcloned insert.

Once a subclone was identified as having an exon-containing insert, further analysis was carried out with the restriction endonucleases *Bam*HI, *Eco*RI, *Eco*RV, *Hind*III, and *Pst*II. The location of the exon(s) within the subclones was subsequently determined by Southern analysis using exon-specific oligonucleotide primers (17). The subclones were used to establish a map of the PCMT1 gene based on overlapping restriction patterns. The gaps that remained were characterized by PCR amplification and subcloning of the adjoining region between two subclones (for instance, clones G2H and 3A2). The appropriate ends of the two flanking subclones were sequenced using the Δ *Taq* cycle-sequencing kit (United States Biochemical). Primers were designed from these sequences in the proper orientation to span the gap region. Extended PCR parameters were used to optimize the amplification over potentially large distances. In a 50- μ l reaction volume, components included 1 $\times$ reaction buffer (Promega), 3.5 μ M MgCl₂ (Invitrogen Hot Wax Mg²⁺ beads), 0.4 mM dNTPs, 20 pmol of each primer, 100–200 ng of DNA, 2 U *Taq* DNA polymerase (Promega), and 2 U *Taq* extender (Stratagene). The cycling parameters consisted of a 95°C "hot start" for 1 min, followed by 30 cycles of 95°C for 30 s, 50°C for 30 s, and 72°C for 7 min. The PCR product was then either blunt-end ligated into pBluescript II SK (Stratagene) or directly ligated into a vector with T-overhangs using a TA Cloning Kit (Invitrogen). These gap subclones were also restriction mapped, and the ends were sequenced to confirm the overlap with the flanking subclones.

Sequencing the upstream/promoter region of the PCMT1 gene. The DNA sequence of the upstream/promoter region was obtained using the 9E1B subclone as a template in a dideoxy chain-termination reaction with the Sequenase Version 2.0 DNA sequencing kit (United States Biochemical) and [α -³⁵S]dATP (New England Nuclear Research Products). Successive oligonucleotide primers were designed using the sequence data obtained from the previous primers.

Primer extension analysis. Primer extension was performed using the synthetic oligonucleotide primers E1R 5'-TACAGAACCACC-TTCAGCGCGACGA-3' (corresponding to the antisense sequence from -23 to -47) and CD1R 5'-GGTGGCACTTACTGCGGAGAT-TGTTG-3' (+41 to +66). These primers were end-labeled with [γ -³²P]ATP (ICN Biomedicals), and approximately 0.1 pmol of each primer was annealed to 500 ng of poly(A)⁺ mRNA from human brain (Clontech Laboratories Inc.). The mixture was incubated at 80°C for 10 min, hybridized at 50°C for 1 h, and then extended at 42°C for 1 h in a reaction mixture containing 30 mM Tris-Cl, pH 8.3, 15 mM MgCl₂, 8 mM DTT, 220 μ g/ml actinomycin D, 0.2 mM or 2 mM dNTPs, and 5 U avian myeloblastosis virus reverse transcriptase (United States Biochemical) (21). The samples were then ethanol precipitated and analyzed on a 7 M urea/6% acrylamide gel adjacent to dideoxynucleotide chain-termination sequencing ladders derived from the genomic subclone 9E1B using the same primers.

RESULTS

Isolation of the human PCMT1 gene. Based on the exon structure of the mouse protein L-isoaspartyl methyltransferase (22), primers specific for the putative exon 7a sequence of a human cDNA clone (19) were synthesized and used by Genome Systems Inc. (St.

Louis, MO) to screen a human bacteriophage P1 genomic library by PCR analysis (18). Three clones were obtained and used as templates for PCR reactions using exon-specific primer sets E5F/E5R, E6F/E6R', E7aF/E7aR', and E8F/E8R', corresponding to exons 5–8, respectively. We found that we could generate the appropriate-sized PCR fragments from P1 clone 658 using primers for the putative exons 5–8. Similarly, P1 clone 659 gave products with primers for exons 6 and 7 but not 8, while P1 clone 660 gave products with primers for exons 7 and 8 but not 6. The presence of putative exons 1–4 in clone 659 was demonstrated by Southern analysis. Since the average insert size in the library used is about 80 kb (18), it was possible that these clones collectively contained the entire methyltransferase gene as well as large amounts of the 5' and 3' flanking regions. These results suggested that clone 659 contained the 5' end of the methyltransferase gene, while clones 658 and 660 contained the 3' end of the methyltransferase gene. Thus, the entire methyltransferase gene is contained within these three overlapping clones (Fig. 2A).

Structural organization of the human PCMT1 gene. A restriction map of the human PCMT1 gene was constructed (Fig. 2C) by analysis of the three overlapping bacteriophage P1 clones and 10 overlapping subclones generated from these genomic clones (Figs. 2A and 2D). Restriction maps of clones 658 and 659 for *EcoRI* and *PstI* were initially determined by Southern blotting with exon-specific oligonucleotide probes. The subclones made from these genomic fragments were then analyzed at greater resolution by restriction mapping with *BamHI*, *EcoRI*, *EcoRV*, *HindIII*, and *PstI*. The sizes of the subcloned restriction fragments corresponded with the sizes of the hybridizing bands observed in Southern blots of both the P1 clones (data not shown) and human placental genomic DNA (23). As apparent in Fig. 2D, a gap remained in the subclone map due to the large intronic region separating exons 1 and 2. In this region, therefore, the indicated distances and restriction pattern were determined by Southern analysis.

We obtained the DNA sequences of the exon/intron junctions using an inverse PCR protocol. As described under Materials and Methods, DNA from the P1 genomic clones was digested with various restriction endonucleases and the resulting fragments were religated under conditions favoring intramolecular ligation. Then, using primers specific for the ends of a putative exon and with 3' ends oriented outward toward the flanking introns, products containing the ends of the exons and the flanking intron DNA up to the restriction site on either side were amplified. Specifically, we obtained the following inverse PCR products using the indicated primers and enzymes: exon 1, *PstI* (550 bp; E1F'/E1R); exon 2, *HaeIII* (650 bp; E2F/E2R); exon 3,

Sau3AI (700 bp; E3F'/E3R); exon 4, *EcoRI* (1.2 kb; E4F'/E4R); exon 5, *Sau3AI* (600 bp; E5F'/E5R); exon 6, *HaeIII* (1.6 kb; E6F'/E6R); exon 7, *HaeIII* (750 bp; E7aF'/E7aR); and exon 8, *Sau3AI* (450 bp; E8F'/E8R). The PCR products were then sequenced using the same oligonucleotide primers to provide information on the splice sites and the upstream and downstream intron regions. The complete DNA sequences of these junction fragments are shown in Fig. 3, and the sequences of the splice junctions are shown in Fig. 4. All of these splice sites correspond closely to the mammalian consensus sequences (24). The putative branch sites for splicing (Fig. 4) also correspond closely to the mammalian consensus branch site sequences (24). A potential 3'-splice site was identified between exons 7b and 7c. However, no corresponding cDNA has yet been isolated that joins exon 7a and exon 7c. The alternative splicing pattern of two previously identified cDNAs that result in 3' ends of exon 7a–8 (pDM2) and exon 7a–7b–8 (pRK1) (19) are also shown in Fig. 4. These alternatively spliced human cDNAs were recently confirmed by Takeda *et al.* (25), who, in addition, isolated a cDNA with an exon 7a–7b–7c pattern.

From these results, we were able to determine the genomic organization of the human PCMT1 gene. The gene spans about 60 kb and consists of 8 exons interrupted by 7 introns. The location and size of the exons/introns are illustrated in Figs. 2 and 3, and listed in Table I. Exons range in size from 32 bp (exon 3) to 784 bp (exon 8), while intron sizes range from about 1.8 kb (intron 2) to about 20.4 kb (intron 1). We found that three of the introns were in phase 0, while three were in phase 1 and one was in phase 2, consistent with the overall distribution noted by Long *et al.* (26). The translational start site is located in exon 1, and three putative translational termination sites have been identified in exons 7b, 7c, and 8. In addition, the *S*-adenosylmethionine-binding motifs, as determined by sequence homology with other methyltransferases of known structure, are found in exons 4, 6, and 7a (Table I). The motifs thought to be responsible for the specific recognition of L-isoaspartyl and D-aspartyl residues, identified as pre-region I and post-region III, are found in exons 3 and 7a, respectively (Table I). Finally, the three polymorphisms located at amino acid positions 22, 119, and 205 (15) correspond to exons 2, 5, and 7a, respectively (Table I). Our results show a similar organization of the PCMT1 gene to that seen in the corresponding mouse gene (22, 27) with the conservation of exons 1–7 between mice and humans.

Transcription initiation and termination sites. The transcriptional start sites of the human methyltransferase gene were mapped by primer extension analysis. A representative primer extension result is shown in Fig. 5 using human brain poly(A)⁺ mRNA and the CD1R primer (+41 to +66). Adjacent lanes display a

FIG. 2. Structural organization of the human protein L-isoaspartyl (D-aspartyl) O-methyltransferase (PCMT1) gene. A, three overlapping P1 genomic clones (clone 658, 659, and 660) span the entire gene as well as large amounts of the 5'- and 3'-flanking regions. The region of the PCMT1 gene contained within each genomic clone, as confirmed by Southern blot or PCR screening, is represented by a solid bold line. The dashed portion of the line indicates the region that may also be contained within the clone but has not been confirmed. B, a diagram of the PCMT1 gene. Exons are represented by vertical lines and numbered. The intron and exon sizes are given above and below the horizontal line, respectively. Putative Alu repetitive elements are indicated with asterisks (*). C, the restriction map of the PCMT1 gene. The restriction sites (B, BamHI; E, EcoRI; EV, EcoRV; H, HindIII; P, PstI) are indicated by vertical lines. Below the complete restriction map are the restriction patterns for each individual enzyme. D, the subclone map of the PCMT1 gene. Subclones were generated from clones 658 and 659 by restriction digest or PCR amplification for more accurate mapping of the PCMT1 gene.

ATGGCTCTGGAATCCCGGCCGCCGCCGCCAGGCTCGAGGCTAATCCCAATCATCTCCGCAAGtaagtgccacttccaaacccgt
M A W K S G G A S H S E L I H N L R
tgtaaggagactggcgagcgtgctgtgacgggtccctgtggcagctcggaagctatctatctgtgccgcgcgcctgt
tggagggttcggcgacagagttgcccgcgtgtgtcccggaacggcgctgcgttgcagtcgcgcgcctcggaacctntcac
tcgtcgacgacgtggactgggaaggagaggaagagccagggcc
..... (intron 1:total of 28.4 kb)
agtagagacggttcagcgtgtgtgccaagctggcttcggaactcctgagctcaggccaaacaccgcgtgttgcccaattac
acaacactccaactctctctctctcannnttcagtcctctatagcagactcgggtgtaattattctcaaatatctctgtac
atgtctctctctctcgaataatttgcctcacttgtttcagAAATGGAATCATCAGACAGATAAAGTATTGTAAGTGA
K N G I I K T D K V F E V
TGCTGGCTACAGACCGCTCCCACTATGCAAAATGTAAACCATACATGGATTCTCCACAAATCAATAGtaagctcttagat
M L A A T D R S H Y A K C N P Y M D S P Q S I
ttctgaaataatcatagtttttcaacttacttattgtatttttgggaagtgaagaaacgctgggacacagatatattatc
tattattaccttcttaatttgaattttatgactcagcaagtttcataatcccaacatacctatttttaataaaattat
gtatgtttttctgttttctcgacatatata
..... (intron 2:total of 1.8 kb)
aaaaaaaaaaaaaaaaaagaattatttgtgaaaaataacgtatggtatggaagaaagaattattgtgaaatgactatatt
itttacagctgactgtcatcagtagttctctctcttcttcttggcagGTTTCCAAGCAACATCAGTGCCTCCACACATGgt
G F Q A A T I S A P H M
aagttaagtttgttcacttggtttacaataaggattacatttttctctataatcattttcttactaaaaaagattttaa
gaagttgttatcaatttactact.
..... (intron 3:total of 16.2 kb)
agagtgcagtaagccttgattgacactgcactcagcctgggcaacaacaacaacaaaaacactactgaggcacagt
gtgggggtggggttagctagtagagaggggggtcagcttagcctcaacaactcagaatttctaataaacagtttacttaag
taatacctggcagatagcaagataacttggttatcttggtagtgacaatagttagaaaaatttgtagaaaagttgncaa
ctttatttatataccatgatgtgtgacttttagtcttaatgaattttgttttcttggcagCATGCTCCATGTGCGCT
H A Y A L
AGAAGTCTTATGTATCAGTGTGCATGAAGGACATAAGCTCTTGTATGTAGGATCTGGAAGTGAATCTCTTACTGCATG
E L L F D Q L H E G A K A L D V G S G S G I L T A C
TTTGTGACGATGTGtaagaggttttatttttctcactctggcccaacagaagaattattatgatttggctgtatgttt
F A R M
tttttaactagtgttttcatgttttatttgtttgttaattctatttaggatagcagtttgaatgagcatcacttaggat
ttgtgtgtgtgtgcaagtgcacagaacatgcacccaaactggcagatnaa
..... (intron 4:total of 3.5 kb)
aaagtgtatataatgataaataatttataaactgaggttttagcaactctatagaagccttaagctattttgtttct
ctttgtagaattagactttttagaaggaataacttcaataactcttagaataaactttagccaatgagcGact
gaattgttttctcttttcagGTTGGATGTACTGGAAGTATAGGAATTTGATCACAATTAAGAGCTAGTAGATGAC
V G C T G K V I D H I K E L V D D
TCAGTAAATAATGTCAGGAAGGACATCCAACACTCTGTCTTCAGGGAGATGACAGCTGTTGtaaaatcatcaga
S V N N V R K D D P T L L S S G R V Q L V
acttttaaaaaatttcttcagagatttttattttcagaagatccaatgcaagctaaataactcagagggtaattattg
tattttttttacaatcacttttaagtaactaaaaa
..... (intron 5:total of 3.8 kb)
aaaaagagtgtgtgagatagaaatgaagcgctttcattactaaatgatgtttaactccagtttttcatcaaaataa
aattccatctgttttaagtcttctgcagcaacaaaaataaaaaataaagtgttaaaacagcaaaaaagtgaaatcc
agttgtcccgagatcatattttagaaaaactatctcttctccattgttaacacctctgactagaatttctcaatcaa
gtttgaGctaaacaggtttttaagcatagctgttttttnnncagTGGGGGATGGAAGAAATGGGATATGCTGAAGAGC
V G D G R M G Y A E E A
CCCTTATGATGCCATTATGTTGGGAGCTGCAGCCCTGTTGTACCCAGCGCGtaagtcgggatttttctgtttgtg
P Y D A I H V G A A A P V V P Q A
tgtttttatccaactaaaactctacaagactttaaataagaaaactcttagatgatgttcaaaataataattagacttt
tatatcatacatttttggttaacatttagctgttactctttcttttcatatgcaaaccaattgtataaatgacatgaa
aattactaaaaccccgagtttggcgctcatatttgatccaggtaagaactaggacaacagtagaattgggaaaaatccc
ctggaggtttatgacctgataactcaggtttttt
..... (intron 6:total of 5.7 kb)
gagactgaggcaggagaataccttgaacctgggaggcagaggttgcagtgagctgagatcggtgtactgcaactccagc
ctgggcgacagagtgagacacctgtctcaaaaataaaaataaaaataaaatGctaaactgtgtgtgatgtgtg
cttttttttttttcttttagCTAATAGTACAGTTAAAGCCCGAGGAAAGATTGATATGTCCTGTTGCTCTCGAGCGC
L I D Q L K P G G R L I L P V G P A G
GAAACCAATTTGGAGCAGTATGCAAGACTACAAGTGGCAGCATCAAAATGAAGCCTCTGATGGGGTGATATACG
G N Q M L E Q Y D K L Q D S I K M K P L M G V I Y
TGCTTTAACAGATAAAGAAAAGCAGTGGTCCAG
V P L T D K E K Q W S R
GTGGAAGTGATTTTATCTCTGCTCTTCTCTCTCCACACATGCAAG
W K stop (pRK1)
GTGAAGGGGTGTGATTTTAAGACATTAGACTACAAGAGTGATTTTGTGTTGTACCTTTATGCTCTCCATTATAA
stop
CGTCAGAAATTCATTACATTAATAATGTGAAAAATGTtgcagtcctccttatttggtttaacatatatcttgggaagg
ggttagactttgtgcacaagcttt
..... (intron 7:total of 8.1 kb)
gatcagcgcaatttttattgacttaatagttgtgtctaaactattataaagccaattgggtagtgtgttaactccGact
aagtattattctctcttttttctcagcGGAATGAATTTGTAAGGCAACATCAGCTTGACCGATGATTAATACCTAGC
D E L stop (pDM2)
ATTGCTCATCTCAGTCCCTCAAAGCTTTTGAACCAACACCATCAGCAGCTGTTTGTGCACTTTGTTACACTGTTATT
TCAGCATGAAAATGTGTGTTTTTTAGGGTTCTGATCTTCAAGAGGCAAGCAGAGCAAAATTTGTTAGAGGAAGGAT
GCAAGATATAAATTTGTGTAAATTTACTTTAACTGCCCATATTTTACTTTGGAATAATTAAGAAAGAGGGTTCTGTAA
AATTGGAATACTTAGTTTGTGAATGATTGTTGAGGAGTGGTTTTCTCTTTCTTGGACACTTAATCTGTTCTGATATTA
ATTATTCAGATGCTTTTGTGTCATGGATAACACACCATTCACAAATTAAGATTCTGGTATTTGGATATCTGTGTAG
ATGCTACTAAGAAATAAGATAGACTGCTTTCTTTTAAAGCTTTTGTAGTGGTGTCATAGATAGCATGTTGTAGATAC
AATCAGCTGCTTTGTTACCTTAAACTAGGCATTTGTAATATTTAAACCTTAAAGTGGCAGGTGATGTCCTGTAAACA
CTCAGCTGTTTCAGATGGACATAACTGATTAGTTTCTCCCTTCTCTCTCTCAAAATATATAGGACATTTGTAGTTTAA
GTGTGCTTTTCTGTTTCTAATTTGCTGCTGATAATGATATGAAGACTTCAACATGCTGTGTGAAGCTGTGTCCTTAGTCTT
GAACAGTTATTCAGCATGCTGCTTTGCCAAATAAAGTTAAAGTTAAAGCAGGCATtggttggtttttgaatagaagg
aagaacaca

5'Splice Site	Branch Site	3' Splice Site
(A/C)AG gtragt	tnc <tr>trac</tr>	(y)nnyag G
----- (38-18 bp) -----		
TCCGCA gtaagt (Exon 1)	ttctgaa	atttgccctcacttgtttcag AAAATG (Exon 2)
CAATAG gtaagc (Exon 2)	agctgac	gctctcttcttcttcttttcag GTTTCC (Exon 3)
CACATG gtaagt (Exon 3)	tgtgtac	atcttggtttcttcttttcag CATGCA (Exon 4)
CGTATG gtaaga (Exon 4)	tactgaa	aattggtttcttcttttcag GTTGGA (Exon 5)
TTGTTG gtaagt (Exon 5)	tctaaac	gcatagctgttttttnncag TGGGGG (Exon 6)
CAGGCG gtgagt (Exon 6)	ctgtaac	cttttttttttttcttttag CTAATG (Exon 7a)
GTCAG GTGGAA (Exon 7a) (Exon 7b)	ttctgct	TCTTCTTCCACACATGCAAG GTGAAA (Exon 7b) (Exon 7c) *
GTCAG GTGGAA (Exon 7a) (Exon 7b)	tactaat	tctctcttttttctctcacag GGATGA (Exon 8) pDM2
TGCAAG GTGAAA (Exon 7b) (Exon 7c)	tactaat	tctctcttttttctctcacag GGATGA (Exon 8) pRK1

FIG. 4. Alignment of exon/intron boundaries and comparison with splice-site and branch-site consensus sequences. Fragments containing the flanking intron were isolated using an inverse PCR protocol and the exon/intron boundaries were sequenced with exon-specific primers. Capital letters represent exons and lower case letters represent introns. Underlined sequences represent identity to the mammalian consensus sequence illustrated at the top of the column. The asterisk (*) identifies a hypothetical splice site that has not yet been observed in any cDNA clones. pDM2 and pRK1 are previously isolated cDNAs with 3' ends containing exons 7a–8 and exons 7a–7b–8, respectively. r, purine; y, pyrimidine; n, any base.

dideoxynucleotide chain-termination sequencing ladder derived from the same primer using a plasmid subclone (9E1B) of this genomic region as a template. The experiment was repeated with a second primer (E1R, –23 to –47) that resulted in the identification of nearly identical sites (data not shown). Each primer gave a major signal corresponding to an A residue 159 bp upstream of the translational start codon. Transcripts initiated at minor start sites were also apparent, including sites at –123, –131, –132, –135, –137, –161, –162, and –174, producing additional faint bands visible on longer autoradiographic exposures (data not shown).

Our genomic sequence at the 3' end of exon 8 (Fig. 3) revealed an AATAAA polyadenylation sequence 20 bases upstream of a TTGTTTGTGTTT sequence. This may correspond to the G/T cluster often found about 30 bp downstream of the consensus AATAAA motif (28), suggesting that this is indeed a polyadenylation site for this gene. If the poly(A) endonuclease cuts as expected at a CA dinucleotide pair 11–20 residues downstream of the AATAAA sequence (28), this would result in a AATAAAGTTAAAGTAAAGCAGGCA(A)_n sequence at the 3' end of the PCMT1 mRNA. Such a sequence has been observed in a human cDNA from an

expressed sequence tag (EST) library (GenBank Accession No. H16778, NCBI ID 276716). It is also possible, however, that cleavage occurs after the CA dinucleotide four nucleotides upstream (25). Interestingly, both human cDNAs previously characterized in our laboratory appear to have truncated 3' ends, where the poly(T) primer used in making the cDNA appears to have initiated reverse transcription on either this A-rich segment, to yield the AAT(A)_n 3' tail in pRK1, or at an internal A-rich segment in exon 8 (bold/italicized in Fig. 3), to generate the TACT(A)_n 3' tail of pDM2 (19).

Analysis of the 5'-flanking sequence. By "primer walking," we characterized the nucleotide sequence up to 2714 bp upstream of the major transcriptional start site at –159. The most striking feature of this sequence was a CpG island beginning approximately 723 bp upstream of this transcriptional start site and extending through the first exon and 221 bp into the first intron (Fig. 6). These islands are stretches of DNA 1–2 kb in length with an average GC content of 60–70% and with a frequency of CpG dinucleotides near that predicted from the nucleotide composition (29), and are often found upstream of vertebrate housekeeping genes. Abnormal methylation of CpG islands in aging and cell transformation can cause changes in gene expression

TABLE I
Location and Size of Exons/Introns in the Human L-Isoaspartyl/D-Aspartyl Methyltransferase Gene

Exon	Size (bp)	Amino acids	Features	Introns	Size (kbp)	
					Human	Mouse (Ref. 27)
1	213	18	5'-UTR	1	20.4	13.7
2	105	35	Ile/Leu ₂₂ polymorphism	2	1.8	1.6
3	32	11	Pre-region I	3	16.2	3.2
4	105	35	SAM-binding motif I	4	3.5	3.5
5	121	40	Ile/Val ₁₁₉ polymorphism	5	3.0	0.9
6	86	29	SAM-binding motif II	6	5.7	1.9
7a	170	57	SAM-binding motif III; post-region III; Lys/Arg ₂₀₅ polymorphism	7	8.1	5.2
7b	47	3	Stop codon (pRK1); polyadenylation signal			
7c	115	—	Stop codon (predicted)			
8	784	4	Stop codon (pDM2); polyadenylation signal			

(30). The CpG island found upstream of the PCMT1 gene is approximately 1 kb in length and has a GC content of 65%. Based on the nucleotide composition, this segment of genomic DNA would be expected to have 105 occurrences of the dinucleotide pair CpG; in fact, 98 CpG dinucleotides were observed.

The PCMT1 5' sequence also contained matches to a variety of nucleotide consensus sequences that have been implicated as binding sites important in the transcriptional regulation of many genes (Fig. 6). Although no upstream TATA box was found, several potential binding sites for the transcription factor ETF (−787, −781, −775, −455, −241, and −230) were identified, which recognizes various GC-rich sequences and stimulates transcription *in vitro* from TATA-less promoters (31). A number of putative SP1 sites were also found on both the template and coding strands at −454, −446, −267, −263, and −101 upstream from the translational start site (32). The 5'-flanking region also contained a CREB-like binding site at −701 (33), an AP1 binding site at −318 (34), five AP2 binding sites at −683, −381, −274, −184, and +215 (35), two xenobiotic-like response elements (XREs) at −339 and −1704 (36), and an antioxidant response element (ARE) at −206 (37). We were also interested in the possibility of steroid hormone-regulated gene expression, which might explain the observed age- and sex-specific differences in methyltransferase activity (38). While no complete sites were observed, we found several half-sites for the estrogen-responsive element (ERE) (39): five 5' half-palindromic 5'-GGTCA-3' motifs (−2025, −1788, −1765, −1516, −1294) and three 3' half-palindromic 5'-TGACC-3' motifs (−2373, −503, −429). Such widely spaced half-palindromic ERE motifs have been shown

to act synergistically in the ovalbumin gene (40). Similarly, half-palindromic TRE motifs function in the promoter of the growth hormone gene (41). We also found a MED-1 element (*m*ultiple start site *e*lement *d*ownstream) at −19; this sequence is thought to function in multiple start site utilization for many TATA-less promoters (42), and may account for our observation of several start sites.

Members of the *Alu* family of short interspersed repetitive sequences were found in the intragenic regions of the human methyltransferase gene, and are indicated by asterisks in Fig. 2. *Alu* sequences represent about 5–6% of the total human genome, appearing approximately every 3–6 kb (43). Due to the limited amount of intronic sequence data, we cannot yet establish the frequency of *Alu* repetitive elements within the methyltransferase gene. However, the intronic sequence data obtained from the ends of the subclones revealed a number of potential *Alu* repeat elements which matched sequences in the GenBank *Alu* database and correlated with the *Alu* consensus sequence established by Jurka and Smith (44). The 5'-untranslated region contains three *Alu* elements within the 3 kb of sequence that was analyzed (−2315 to −2604, −1854 to −1543, −1360 to −1061). This appears to be a much higher frequency than would be predicted, but the relevance or potential effects on regulation of the methyltransferase gene remain unknown.

DISCUSSION

Nucleotide sequencing, Southern analysis, and restriction mapping revealed that the human protein L-isoaspartyl (D-aspartyl) *O*-methyltransferase gene

FIG. 5. Mapping of the human PCMT1 gene transcriptional start site. The transcriptional start site was determined by primer extension. An end-labeled primer, CD1R, was hybridized to whole brain poly(A) RNA and extended with AMV reverse transcriptase. Extension products were separated on a denaturing polyacrylamide gel alongside a dideoxynucleotide sequencing ladder primed on a genomic DNA plasmid subclone with the same oligonucleotide primers. The signal at 159 upstream of the ATG codon represents the major transcriptional start site and the fainter bands at -174, -137, and -135 identify the presence of minor alternative start sites, indicated by arrows. The sequencing ladders were loaded from left to right with G, A, T, and C. The bottom panel is a schematic representation of the results for this assay. BT1R is a second primer used in this analysis and it gave similar results (data not shown). Sizes are in base pairs, and the asterisks (*) represent SP1 sites.

(PCMT1) consists of 8 exons interrupted by 7 introns, spanning a genomic region of approximately 60 kb (Fig. 2). By sequencing the exon/intron splice junctions, we were able to determine that all of the splice sites were in agreement with the mammalian consensus sequence (24) (Fig. 4). These human splice sites were identical in position and very similar in DNA sequence to the splice sites of the murine methyltransferase gene (22). Although the similarity of both the protein-coding and 5'-noncoding regions between the human and murine

methyltransferase cDNAs is very high (>90%), the genomic structures have a few significant differences. Comparison of the intron sizes between the human and murine methyltransferase genes has revealed that the human gene generally contains much larger intronic regions (22, 27). However, there does seem to be some conservation in the relative size of certain introns, such as introns 1, 2, and 4 (Table I).

Overlapping sequences of the pRK1 and pDM2 cDNA clones (19) predict that a major PCMT1 mRNA species

FIG. 6. Sequence analysis of the human methyltransferase gene promoter and exon 1. The major transcriptional start site is indicated by an asterisk (*), and the oligonucleotides used in the primer extension reaction are identified with arrows. All putative *cis*-acting elements, including potential SP1, ETF, AP1, AP2, XRE, ARE, CREB, and half-palindromic ERE motifs, are enclosed in boxes. The protein coding sequence of exon 1 is shaded. The CpG island is boxed in bold. Nucleotides are numbered sequentially from the first nucleotide of the ATG codon, designated +1. The nucleotide sequence reported in this figure has been submitted to the GenBank/EMBL Data Bank with Accession No. U49740.

in the human brain is about 1.6 kb. Northern analysis of human HeLa cell RNA revealed a major 1.6-kb species, as well as a significant amount of a 2.6-kb species and a minor amount of a 4.5-kb species (45). Similar analysis of human lens epithelium demonstrated a major 2.5-kb transcript as well as a less prevalent 1.7-kb transcript (46), while analysis of a human erythroid leukemic cell line showed a major 1.0-kb transcript, as well as a minor 1.6-kb transcript (25). The 1.0-kb transcript would be consistent with poly(A) addition at an ATTAAG sequence upstream of a "G/T" cluster in exon 7c, while the 1.6- to 1.7-kb transcripts are consistent with poly(A) addition at the AATAAG site at the 3' end of exon 8 (Fig. 3). The structure of the larger mRNA species (>1.6–1.7 kb) has yet to be determined. It seems possible that the termination site identified at the end of exon 8 does not result in termination of all RNA species and that additional AATAAG sites are also positioned 1 and 2.9 kb downstream to generate the 2.6- and 4.5-kb species. Alternatively, they could result from alternative splicing using as yet undiscovered exons.

Multiple mRNA transcripts have also been detected in mouse at ~1.0, 1.4, 1.7, 2.8, and 3.9 kb (22) and in rat at 1.1, 1.7, 2.5, and 4.0 kb (47). The smallest transcripts (~1.0 and 1.1) appear to be testis-specific and may result from poly(A) addition at sites homologous to the ATTAAG site in the human exon 7c (22, 47) (Fig. 3). The rodent 1.7-kb species, probably reflecting a 1580-bp cDNA in mouse (22, 48) and a 1598-bp cDNA in rat (49), may result from poly(A) addition at a site homologous to the AATAAG site at exon 8 in human genomic DNA, while the larger transcripts may reflect longer exon 8 transcripts (see above).

Determination of transcription initiation sites by primer extension of brain poly(A)⁺ mRNA revealed multiple start sites, with a major site 159 bp upstream from the translational start. The multiple start sites found in this region are very similar to those observed in the murine methyltransferase gene (48). A number of sequenced human EST clones also appeared to begin within this region, with the longest 5' sequence extending to nucleotide -177 in a cDNA from infant brain (GenBank Accession No. HO7963, NCBI ID 267843; Washington University/Merck EST Project). Beyond this point, the upstream sequences between human and mouse begin to diverge (59% sequence identity), and beyond approximately 800 bp upstream there is little to no similarity between the two sequences (32%). This may suggest the possibility of differential regulatory mechanisms involving specific enhancer-binding factors between the mouse and human genes or may reflect that all critical elements are within this region.

Sequence inspection revealed a number of potential *cis*-acting DNA elements, including matches to SP1, ATF, AP1 and AP2, XRE, ARE, CREB, MED-1, and

ERE half-palindromic consensus motifs. PCMT1 gene expression, while perhaps constitutive in general, may be more important in certain tissues or under stress conditions. For example, under stress conditions there is a potential for increased rates of protein damage caused by heat and oxidative stress. The elements identified in the promoter region of the PCMT1 gene may help to elucidate how expression is affected by these conditions. Clearly, the presence of antioxidant (ARE) and xenobiotic (XRE) response elements would suggest the ability of PCMT1 to be upregulated in the presence of such damage-causing agents. Moreover, the observed age- and sex-specific differences in methyltransferase activity (38) may be partially accounted for by the presence of multiple estrogen-responsive elements which together have been shown to effect gene expression (40). It should be emphasized that none of the above sequence motifs have yet been shown to function in the regulation of human PCMT1 gene expression. However, we are currently attempting to identify the contribution of the core regulatory elements by constructing a series of reporter genes with fragments of the methyltransferase promoter.

Examination of the nucleotide composition of the proximal 5'-flanking region and exon 1 revealed them to be enriched in both GC content and CpG dinucleotides. The promoter region of this gene, which also lacks an identifiable TATA box, is therefore characteristic of that found in housekeeping genes, consistent with the broad distribution of the methyltransferase. However, this does not necessarily establish the methyltransferase as being constitutively expressed. These genes are often highly and variously regulated at the level of transcription (50), and some regulation has been observed involving age- and gender-specific differences in expression (38).

ACKNOWLEDGMENTS

We thank Duncan MacLaren for his helpful advice concerning P1 plasmid isolation and various mapping techniques and Dr. Mary Beth Mudgett for her help with the transcriptional start site mapping analysis.

REFERENCES

1. Stadtman, E. R. (1988) *J. Gerontol.* **43**, B112–B120.
2. Harding, J. J., Beswick, H. T., Ajiboye, R., Huby, R., Blakytyn, R., and Rixon, K. C. (1989) *Mech. Ageing Dev.* **50**, 7–16.
3. Geiger, T., and Clarke, S. (1987) *J. Biol. Chem.* **262**, 785–794.
4. Patel, K., and Borchardt, R. T. (1990) *Pharmaceut. Res.* **7**, 703–711.
5. Chazin, W. J., Kordel, J., Thulin, E., Hofmann, T., Drakenberg, T., and Forsen, S. (1989) *Biochemistry* **28**, 8646–8653.
6. Friedman, A. R., Ichihpruani, A. K., Brown, D. M., Hillman, R. M., Krabill, L. F., Martin, R. A., Zurcher-Neely, H. A., and Guido, D. M. (1991) *Int. J. Peptide Protein Res.* **37**, 14–20.
7. Sharma, S., Hammen, P. K., Anderson, J. W., Leung, A., Georges, F., Hengstenberg, W., Klevit, R. E., and Waygood, E. B. (1993) *J. Biol. Chem.* **268**, 17695–17704.

8. Lowenson, J. D., and Clarke, S. (1995) in *Deamidation and Isoaspartate Formation in Peptides and Proteins* (Aswad, D. W., Ed.), pp. 47–64, CRC Press, Boca Raton, FL.
9. McFadden, P. N., and Clarke, S. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 2595–2599.
10. Johnson, B. A., Murray, E. D. Jr., Clarke, S., Glass, D. B., and Aswad, D. W. (1987) *J. Biol. Chem.* **262**, 855–866.
11. Brennan, T. V., Anderson, W. J., Jia, Z., Waygood, E. B., and Clarke, S. (1994) *J. Biol. Chem.* **269**, 24586–24595.
12. Johnson, B. A., Langmack, E. L., and Aswad, D. W. (1987) *J. Biol. Chem.* **262**, 12283–12287.
13. Li, C., and Clarke, S. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 9885–9889.
14. MacLaren, D. C., O'Connor, C. M., Xia, Y.-R., Mehrabian, M., Klisak, I., Sparkes, R. S., Clarke, S., and Lusk, A. J. (1992) *Genomics* **14**, 852–856.
15. Tsai, W., and Clarke, S. (1994) *Biochem. Biophys. Res. Commun.* **203**, 491–497.
16. Reynolds, T. R., and Buck, G. A. (1992) *Biotechniques* **12**, 518–521.
17. Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd ed., Cold Spring Harbor Press, Cold Spring Harbor, NY.
18. Shepherd, N. S., Pfrogner, B. D., Coulby, J. N., Ackerman, S. L., Vaidyanathan, G., Sauer, R. H., Balkenhol, T. C., and Sternberg, N. (1994) *Proc. Natl. Acad. Sci. USA* **91**, 2629–2633.
19. MacLaren, D. C., Kagan, R. M., and Clarke, S. (1992) *Biochem. Biophys. Res. Commun.* **185**, 277–283.
20. Triglia, T., Peterson, M. G., and Kemp, D. J. (1988) *Nucleic Acids Res.* **16**, 8186.
21. Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A., and Struhl, K. (1994) *Current Protocols in Molecular Biology*, Wiley, New York.
22. Romanik, E. A., Ladino, C. A., Killoy, L. C., D'Ardenne, S. C., and O'Connor, C. M. (1992) *Gene* **118**, 217–222.
23. Ingrosso, D., Kagan, R. M., and Clarke, S. (1991) *Biochem. Biophys. Res. Commun.* **175**, 351–358.
24. Green, M. R. (1991) *Annu. Rev. Cell Biol.* **7**, 559–599.
25. Takeda, R., Mizobuchi, M., Murao, K., Sato, M., and Takahara, J. (1995) *J. Biochem.* **117**, 683–685.
26. Long, M., Rosenberg, C., and Gilbert, W. (1995) *Proc. Natl. Acad. Sci. USA* **92**, 12495–12499.
27. MacLaren, D. C., and Clarke, S. (1996) *Genomics* **35**, 299–307.
28. Birnstiel, M. L., Busslinger, M., and Strub, K. (1985) *Cell* **41**, 349–359.
29. Cross, S. H., and Bird, A. P. (1995) *Curr. Opin. Genet. Dev.* **5**, 309–314.
30. Vertino, P. M., Spillare, E. A., Harris, C. C., and Baylin, S. B. (1993) *Cancer Res.* **53**, 1684–1689.
31. Kageyama, R., Merlino, G. T., and Pastan, I. (1989) *J. Biol. Chem.* **264**, 15508–15514.
32. Kadonaga, J. T., Jones, K. A., and Tjian, R. (1986) *Trends Biochem. Sci.* **11**, 20–23.
33. Fink, J. S., Verhave, M., Kasper, S., Tsukada, T., Mandel, G., and Goodman, R. H. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 6662–6666.
34. Lee, W., Mitchell, P., and Tjian, R. (1987) *Cell* **49**, 741–752.
35. Williams, T., and Tjian, R. (1991) *Genes Dev.* **5**, 670–682.
36. Denison, M. S., Fisher, J. M., and Whitlock, J. P. Jr. (1988) *J. Biol. Chem.* **263**, 17221–17224.
37. Rushmore, T. H., Morton, M. R., and Pickett, C. B. (1991) *J. Biol. Chem.* **266**, 11632–11639.
38. Johnson, B. A., Shirokawa, J. M., Geddes, J. W., Choi, B. H., Kim, R. C., and Aswad, D. W. (1991) *Neurobiol. Aging* **12**, 19–24.
39. Klein-Hitpass, L., Ryffel, G. U., Heitlinger, E., and Cato, A. C. B. (1988) *Nucleic Acids Res.* **16**, 647–663.
40. Kato, S., Tora, L., Yamuchi, J., Masushige, S., Bellard, M., and Chambon, P. (1992) *Cell* **68**, 731–742.
41. Kim, H. S., Crone, D. E., Sprung, C. N., Tillman, J. B., Force, W. R., Crew, M. D., Mote, P. L., and Spindler, S. R. (1992) *Mol. Endocrinol.* **6**, 1489–1501.
42. Ince, T. A., and Scotto, K. W. (1995) *J. Biol. Chem.* **270**, 30249–30252.
43. Moyzis, R. K., Torney, D. C., Meyne, J., Buckingham, J. M., Wu, J.-R., Burks, C., Sirotkin, K. M., and Goad, W. B. (1989) *Genomics* **4**, 273–289.
44. Jurka, J., and Smith, T. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 4775–4778.
45. Ladino, C. A., and O'Connor, C. M. (1992) *J. Cell. Physiol.* **153**, 297–304.
46. Kodama, T., Mizobuchi, M., Takeda, R., Torikai, H., Shinomiya, H., and Ohashi, Y. (1995) *Biochim. Biophys. Acta* **1245**, 269–272.
47. Mizobuchi, M., Murao, K., Takeda, R., and Kakimoto, Y. (1994) *J. Neurochem.* **62**, 322–328.
48. Galus, A., Lagos, A., Romanik, E., and O'Connor, C. M. (1994) *Arch. Biochem. Biophys.* **312**, 524–533.
49. Sato, M., Yoshida, T., and Tuboi, S. (1989) *Biochem. Biophys. Res. Commun.* **161**, 342–347.
50. Azizkhan, J. C., Jensen, D. E., Pierce, A. J., and Wade, M. (1993) *Crit. Rev. Eukaryot. Gene Expression* **3**, 229–254.