Weiner, J. H., Dettman, H. D., Henry, G. D., O'Neil, J., & Sykes, B. D. (1987) *Biochem. Soc. Trans. 15*, 87.
Wüthrich, K. (1986a) *NMR in Biological Research: Peptides and Proteins*, North-Holland, Amsterdam.
Wüthrich, K. (1986b) *NMR of Proteins and Nucleic Acids*, Wiley, New York.
Wüthrich, K., Billeter, M., & Braun, W. (1984) *J. Mol. Biol. 180*, 715.
Zetta, L., De Marco, A., & Zannoni, G. (1986) *Biopolymers 25*, 2315.
Zuiderweg, E. R. P., Hallenga, K., & Olejniczak, E. T. (1986) *J. Magn. Reson. 70*, 207.

Complete Sequence and Organization of the Murine β-Glucuronidase Gene[†,‡]

Melanie A. D'Amore,[§,∥] Patricia M. Gallagher,[§] Thomas R. Korfhagen,[⊥] and Roger E. Ganschow[*,§]
Division of Basic Science Research and Division of Pulmonary Biology, Children's Hospital Research Foundation, Cincinnati, Ohio 45229, and Graduate Program in Developmental Biology, University of Cincinnati, Cincinnati, Ohio 45221
Received March 3, 1988; Revised Manuscript Received May 25, 1988

ABSTRACT: The murine β-glucuronidase structural gene (*Gus-s*) has been isolated from a BALB/cJ sperm DNA bacteriophage library and its nucleotide sequence established. The gene is organized into 12 exons comprising 17.5% of the 14 009 base pair (bp) region spanning the interval between transcription initiation and the putative site of polyadenylation. A TATA box sequence, embedded within a GC-rich region, is found 28 bp upstream from the transcription initiation site. Eleven members of the B1 family and eight members of the B2 family of murine repetitive elements were identified within *Gus-s* and 2440 bp of flanking sequence. Other novel sequences found within *Gus-s*, including a $(AC)_{19}$ homocopolymer tract within intron 3 and a 23 base pair complex direct repeat within intron 9, are presented and discussed.

The murine β-glucuronidase gene complex, designated [*Gus*], provides an excellent model system for examining the structure and function of mammalian regulatory elements which serve to control and modify gene expression. The GUS structural gene, *Gus-s*, and three GUS-specific regulatory elements, identified through characterization of natural variants of GUS expression, define the GUS gene complex on chromosome 5 of the mouse [for a review, see Paigen (1979)]. Three common alleles of *Gus-s* ($-s^a$, $-s^b$, $-s^h$) specify allozymes which differ in electrophoretic mobility, heat stability, or both (Paigen, 1961; Swank et al., 1973; Lalley & Shows, 1974). Specific alleles of each of three GUS regulatory elements (*Gus-r*, *Gus-t*, and *Gus-u*) are associated with specific alleles of *Gus-s*, and these associations define three common haplotypes, $[Gus]^a$, $[Gus]^b$, and $[Gus]^h$, and several rare haplotypes. The effects of each regulatory variant on the expression of GUS have been examined in considerable detail at the cellular and biochemical levels.

Gus-r, a cis-active regulatory element tightly linked to *Gus-s*, controls the androgen responsiveness of kidney GUS mRNA (Palmer et al., 1983). Recently, a rare GUS haplotype, $[Gus]^{or}$, was described in which kidney GUS does not respond to androgen (Lund et al., 1988). $[Gus]^{or}$ represents a cis-active, "null" variant of the androgen response of kidney GUS which may represent either a mutant form of a novel, GUS-specific response element or an allele of *Gus-r*. To date, no androgen-responsive element has been identified in an androgen-responsive gene. Structural and functional comparisons of DNAs within and surrounding the GUS structural gene between this variant and haplotypes which respond normally could provide identification of this element.

A second element, *Gus-u*, is tightly linked to *Gus-s* and controls in cis the levels of GUS synthesis in all tissues at all times (Lusis et al., 1983). In addition, a third regulatory element, designated *Gus-t*, is tightly linked to *Gus-s* and exerts an additional control in trans over GUS synthesis in certain tissues (Meredith & Ganschow, 1978; Lusis et al., 1983). Recent preliminary studies from our laboratory strongly infer that the control of the rates of GUS synthesis by *Gus-u* and *Gus-t* is not exerted through control of GUS mRNA levels (Wawrzyniak and Ganschow, unpublished experiments). Systematic comparisons and genetic manipulations of the DNA within the GUS gene complex could reveal the DNA determinants of these regulatory elements.

Clearly, efforts to identify the DNA determinants of the GUS-specific regulatory elements require the characterization of the GUS structural gene (*Gus-s*). To this end, we have used a GUS cDNA,[1] pGUS-1 (Palmer et al., 1983), to isolate a series of overlapping bacteriophage clones which span $Gus\text{-}s^a$. Determination of the entire nucleotide sequence from one haplotype provides molecular access into the GUS gene complex and allows the design of experiments to identify the differences in DNA structure within [*Gus*] which are causally related to differences in regulation among the GUS haplotypes.

Materials and Methods

Materials. Restriction enzymes, T4 DNA ligase, and M13 sequencing primer (17-mer) were purchased from New England BioLabs, Inc. T4 polynucleotide kinase, RNase T1, and T3 polymerase were obtained from Bethesda Research Laboratories. Sodium 2′,3′-dideoxynucleoside 5′-triphosphates, sodium 2′-deoxynucleoside 5′-triphosphates, and M13mp18

[†] Supported by Grants DK14770 and GM36266 to R.E.G. and Grant GM10799 to T.R.K. from the National Institutes of Health.
[‡] The nucleic acid sequence in this paper has been submitted to GenBank under Accession Number J02836.
[*] To whom correspondence should be addressed at Children's Hospital Research Foundation, IDR Room 720, Elland and Bethesda Avenues, Cincinnati, OH 45229-2899.
[§] Division of Basic Science Research.
[∥] Graduate Program in Developmental Biology.
[⊥] Division of Pulmonary Biology.

[1] Abbreviations: bp, base pair(s); cDNA, complementary DNA; RF, replicative form.

0006-2960/88/0427-7131$01.50/0

and M13mp19 phage RF DNAs were purchased from Pharmacia P-L Biochemicals, Inc. Bluescribe vector was obtained from Stratagene. Avian myeloblastosis virus reverse transcriptase was obtained from Life Sciences, Inc. The Rapid Deletion Subcloning Kit was purchased from International Biotechnologies, Inc. Bacterial alkaline phosphatase and RNase A were obtained from Sigma. Low-melting agarose was purchased from Seakem. NA45 paper was purchased from Schleicher & Schuell. Radionucleotides were obtained from New England Nuclear Products/Dupont.

Identification of Cloned Genomic DNAs. A murine BALB/cJ sperm DNA library, constructed in Charon 4A in the laboratory of Leroy Hood and kindly provided by Jerry B. Lingrel, was screened for GUS genomic clones. A nick-translated (Rigby et al., 1977) GUS cDNA clone, pGUS-1 (Palmer et al., 1983; Gallagher et al., 1987), was used as a hybridization probe to screen approximately 10^6 recombinant phage in this library by in situ hybridization (Benton & Davis, 1977; Maniatis et al., 1978). Cloned genomic DNAs identified in this manner were purified (Maniatis et al., 1982) and characterized by restriction mapping with *Eco*RI, *Bam*HI, and *Hin*dIII.

DNA Sequence Analysis. Restriction fragments from λAGus-1, λAGus-4, and λAGus-5 (see Figure 1) were isolated by gel electrophoresis through low-melting agarose or by electrophoresis onto NA45 paper according to the manufacturer's instructions. Purified fragments were ligated into M13mp18 or M13mp19 and transformed into *E. coli* JM109 (Messing et al., 1981; Yanisch-Perron et al., 1985). M13 clones too large to be sequenced completely were modified by rapid deletion subcloning according to Dale et al. (1985), creating a nested set of subclones. Single-stranded M13 templates were sequenced by the quasi-end-labeling adaptation of the dideoxy chain-termination method (Duncan, 1985). The products of these reactions were separated by electrophoresis on denaturing 5% and 6% polyacrylamide gels. Gels were dried under vacuum with heat and then autoradiographed.

In addition, several genomic DNA fragments were 5′ end labeled with [γ-^{32}P]ATP (3000 Ci/mmol) by T4 polynucleotide kinase, following treatment with bacterial alkaline phosphatase, and sequenced by the method of Maxam and Gilbert (1980). Sequencing reactions were run on denaturing 6% and 20% polyacrylamide gels. Gels were dried under vacuum with heat and then autoradiographed.

Resulting DNA sequences were entered, stored, and analyzed with the Microgenie DNA Sequence Analysis Program from Beckman Instruments (Queen & Korn, 1980).

Exon Mapping. A 2.6-kb *Eco*RI–*Hin*dIII genomic fragment containing the putative first exon and flanking sequence of *Gus-s* (see Figure 3) was cloned into a Bluescribe vector between the *Hin*dIII and *Eco*RI sites. The insert was oriented such that transcription from the T3 RNA polymerase promoter (Melton et al., 1984) yielded a [^{32}P]UTP-labeled transcript containing sequence complementary to that of the sense strand in the corresponding exon sequence within GUS mRNA. The resulting antisense RNA was isolated in low-melting agarose, visualized with ethidium bromide staining, and excised. This RNA was removed from the gel by phenol and phenol–chloroform extractions, precipitated with ethanol, and resuspended in 15 μL of hybridization buffer (Melton et al., 1984). The radiolabeled RNA was added to 100 μg of total kidney RNA from androgen-treated female mice, heated to 85 °C for 15 min, and hybridized at 45 °C overnight. The products of the reaction were treated with RNase A (40 μg/mL) and RNase T1 (500 μg/mL) for 30 min at 30 °C, after which SDS was added to a final concentration of 0.02%. Proteinase K was added to a final concentration of 0.15 μg/μL, and the reaction was incubated for 15 min at 37 °C, extracted with phenol and phenol–chloroform, and ethanol precipitated. To assess the extent of RNase protection of the probe, samples were analyzed on a denaturing 6% polyacrylamide gel.

Primer Extension Analysis. Primers for primer extension analyses were generated from a 2.5-kb *Kpn*I–*Eco*RI subclone of M13 containing the putative first exon of *Gus-s* and 5′ flanking sequence (Figure 4) extending to the 5′ most *Eco*RI site in Figure 1. The M13 universal 17-mer primer was hybridized to this M13 clone, quasi end labeled (Duncan, 1985) with [α-^{32}P]dATP (3000 Ci/mmol), and extended with avian reverse transcriptase in the presence of 2.5 mM dNTPs. The resulting double-stranded M13 DNA was divided into two aliquots, one of which was digested with *Pvu*II and the other with *Sac*I. The digestion products were run on a denaturing 5% polyacrylamide gel to isolate the single-stranded primers which were located by autoradiography and excised. The DNA was eluted from the gel overnight at 37 °C in 0.5 M NH_4OAc–1 mM EDTA and collected by precipitation with ethanol and resuspended in 12.5 μL of 0.1× SET (1 mM Tris-HCl, pH 7.5, 0.5 mM EDTA, 0.1% SDS).

Total RNA was isolated according to Chirgwin et al. (1979), and poly(A+) RNA was prepared by oligo(dT)–cellulose chromatography (Aviv & Leder, 1972). The hybridization solution contained 6 μg of poly(A+) RNA, [α-^{32}P]dATP-labeled primer [(~3–5) × 10^5 cpm], 33.2 μL of hybridization buffer (Degen et al., 1987), and 50% formamide with a final volume of 98.7 μL. After overnight incubation at 39 °C, the nucleic acids were precipitated twice with ethanol and dissolved in 20 μL of reverse transcriptase mix (Degen et al., 1987). The extension reactions were incubated at 37 °C for 1 h and stopped by ethanol precipitation. The samples were analyzed by electrophoresis on a denaturing 6% polyacrylamide gel, which was then dried and autoradiographed.

Results and Discussion

Isolation and Mapping of Murine GUS Genomic Clones. A BALB/cJ mouse DNA library was screened for GUS structural gene sequences with the murine GUS cDNA probe, pGUS-1 [for description of this probe, see Palmer et al. (1983) and Gallagher et al. (1987)]. From 10^6 phage screened, 24 positive clones were isolated. Characterization by restriction enzyme mapping, as described under Materials and Methods, revealed five unique, overlapping clones, designated λAGus-1 through λAGus-5, which span approximately 40 kb of DNA (Figure 1).

Nucleotide Sequence of Gus-s with Reference to Available GUS cDNAs. Restriction fragments from the genomic clones λAGus-1, λAGus-4, and λAGus-5 were used for DNA sequencing (see Figure 1). Generation of DNA fragments and clones for sequencing, as well as sequencing techniques, is described under Materials and Methods. The sequencing strategy showing the informative clones is depicted in Figure 1. For 98% of the gene and flanking regions, both DNA strands were sequenced or determined more than once in independently derived subclones. The sequence of the remaining genomic DNA was determined with single subclones. Sequence comparisons between overlapping murine GUS cDNAs (Gallagher et al., 1988) and GUS genomic DNAs permit identification of the 3′-terminal 10 exons of *Gus-s* and most of the exon which immediately precedes this group. The 5′ end of the incomplete exon was localized 18 base pairs upstream of the 5′ GUS cDNA termination point (Gallagher et al., 1988) by comparisons of our genomic sequence with

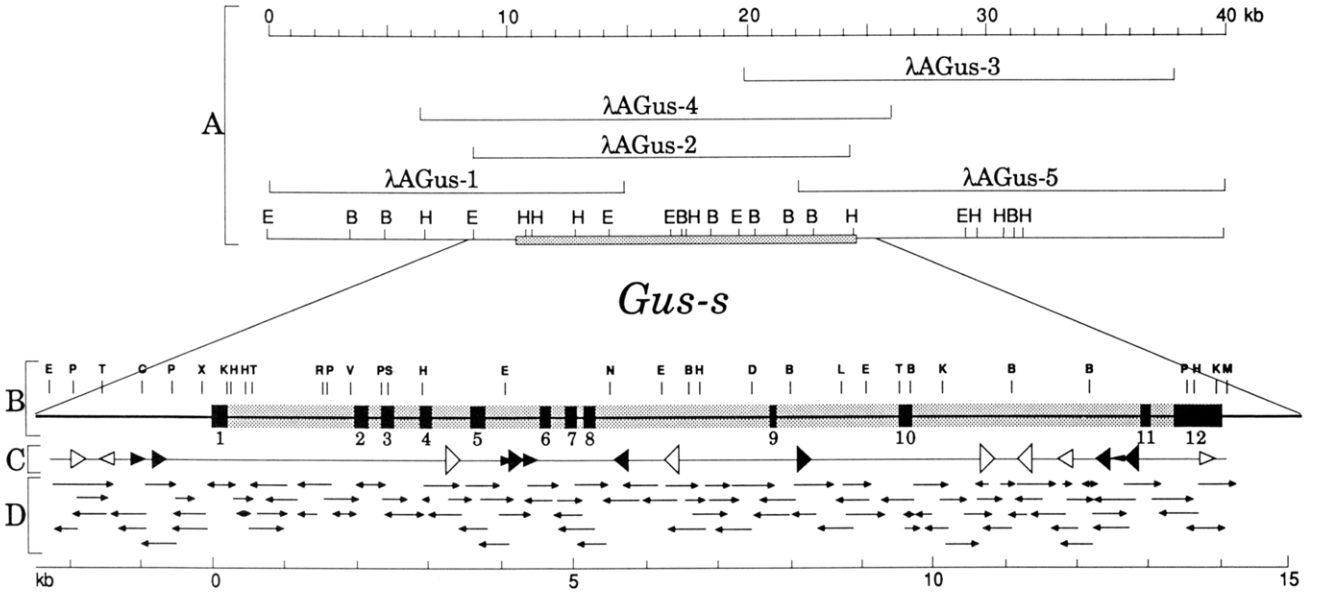


FIGURE 1: Organization of the murine β-glucuronidase gene. (A) Overlapping genomic clones of the A haplotype are shown with a kilobase scale. The stippled area within the partial restriction map marks the boundaries of *Gus-s*. Restriction sites indicated in (A) and (B) include the following: B, *Bam*HI; C, *Sac*I; D, *Dra*I; E, *Eco*RI; H, *Hin*dIII; K, *Kpn*I; L, *Sal*I; M, *Mnl*I; N, *Hin*cII; P, *Pst*I; R, *Rsa*I; S, *Sph*I; T, *Taq*I; V, *Pvu*II; X, *Xba*I. (B) The enlarged region from (A) shows a partial restriction map of *Gus-s* with darkened blocks indicating exons and stippled areas representing introns. (C) Locations of B1 and B2 repetitive elements are designated; the solid arrowheads represent B1 elements, and the open arrowheads represent B2 elements. The orientation of the arrow represents the direction of transcription from the putative RNA polymerase III promoters of these repeats. The vertical height designates the relative size of the element. (D) Sequenced fragments of the gene; arrows represent the area and direction sequenced; double-ended arrows designate complementary clones sequenced.

published rat GUS cDNA sequence (Nishimura et al., 1986) and with consensus splice junction sequences (Mount, 1982). However, this exon does not represent the first exon of *Gus-s* since DNA sequence encoding the amino-terminal 10 amino acids of mature mouse GUS (A. J. Lusis and K. Wilson, personal communication) is localized within an open reading frame approximately 1.5 kb upstream at nucleotide positions 79–108 in Figure 2.

Identification and Characterization of the First Exon and the Major Site of Transcription Initiation. We have utilized several approaches to identify and characterize the first exon of *Gus-s*. Initially, we searched in the upstream direction of the GUS genomic sequence for a continuation of the rat GUS cDNA sequence similarity not represented in the previously identified 11 exons of *Gus-s*. The continuation of sequence similarity to rat GUS cDNA sequence begins at residue 222 in Figure 2. This continuation begins on the appropriate side of a donor splice junction consensus sequence, is separated by 1660 base pairs from the 5′ end of the first of the previously identified group of 11 exons, and continues uninterrupted through the 5′ end of the published rat GUS cDNA to residue 8 in Figure 2. This similarity includes sequences encoding the amino-terminal 10 amino acids of the mature GUS polypeptide, a 22 amino acid signal sequence, and all or a portion of the 5′ untranslated sequence of *Gus-s*. That this sequence represents almost all of the first exon of the *Gus-s* transcriptional unit is inferred from results of nuclease protection and primer extension experiments described below.

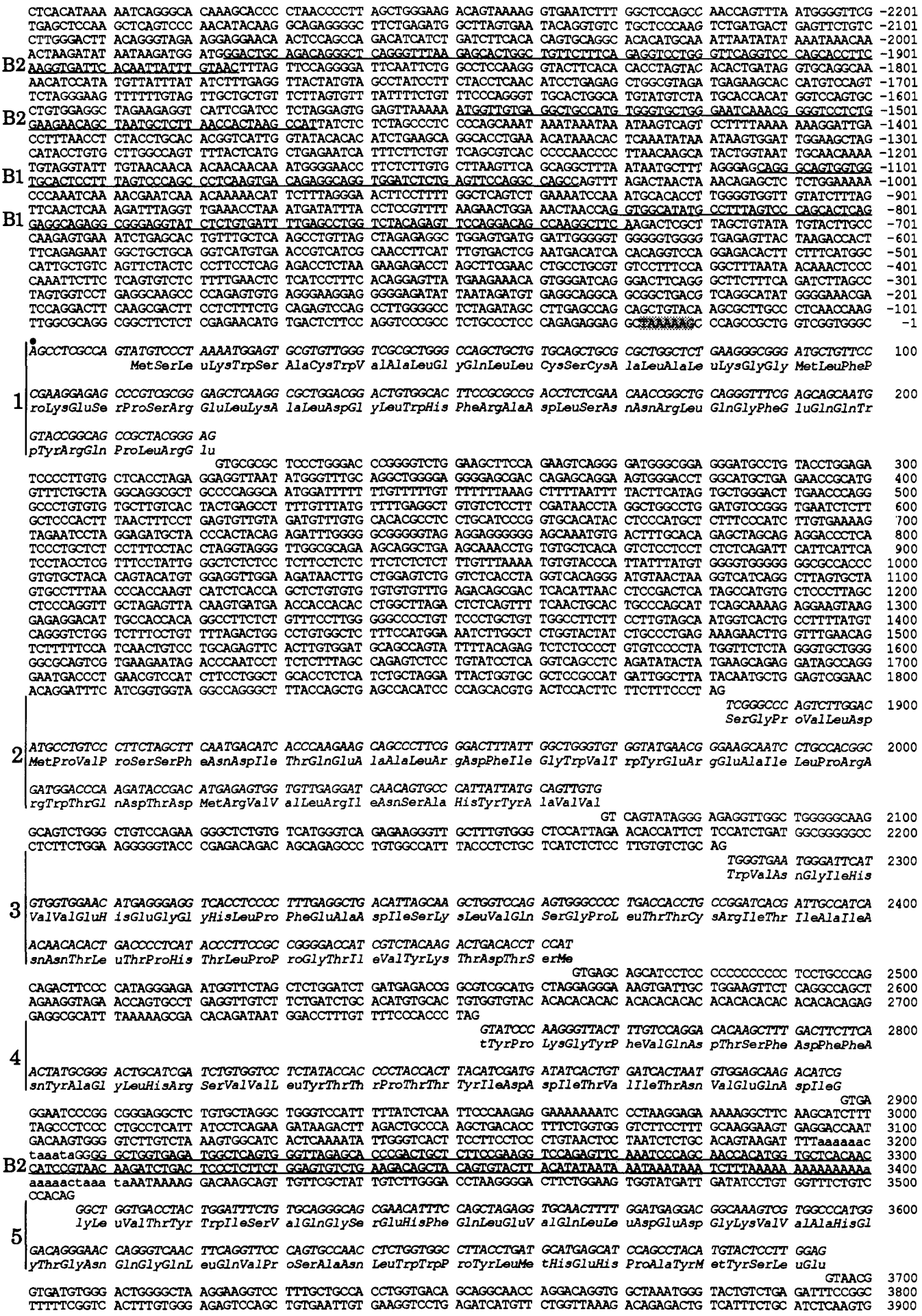
In order to size the putative exon 1, a restriction fragment containing the putative first exon of *Gus-s* plus 3′ and 5′ flanking sequences was ligated into the Bluescribe vector as described under Materials and Methods and in the legend to Figure 3. The antisense strand of this insert was transcribed in the presence of [^{32}P]UTP, hybridized to total RNA from androgen-treated female mouse kidney, and digested with RNases A and T1, and the products were separated on a 6% polyacrylamide sequencing gel. The results, shown in Figure 3, reveal a RNA:RNA protected fragment comigrating with a 220 base pair marker (Figure 3). This result indicates that the size of the putative exon 1 is approximately 220 nucleotides.

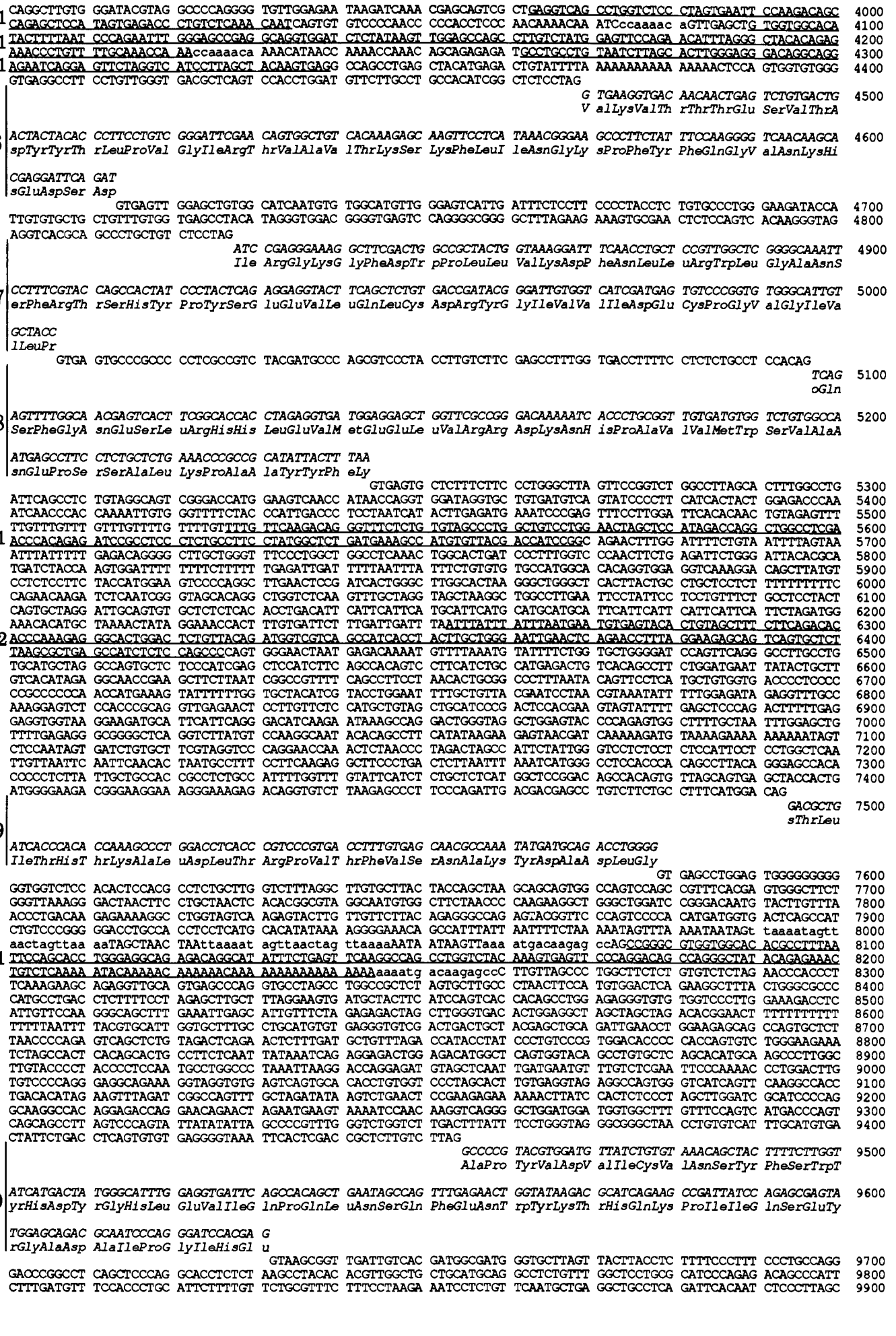
To identify the site of transcription initiation, two nested antisense primers containing a portion of the putative first exon were created from a single M13 subclone (*Kpn*I/*Eco*RI). Both primers begin at the same restriction site (*Kpn*I, residue 200) yet end at staggered positions (*Pvu*II, residue 52; *Sac*I, residue 121) within the exon (See Figures 2 and 4). Each radiolabeled antisense primer was hybridized to mouse kidney poly(A+) RNA and extended with reverse transcriptase, and the products of extension were separated electrophoretically in a 6% polyacrylamide sequencing gel in parallel with a chain termination sequencing reaction of the *Kpn*I/*Eco*RI M13 subclone from which the primers were derived. As shown in Figure 4, the length of each extended primer corresponds to an extension which terminates in a doublet at a T residue of the antisense strand of *Gus-s* corresponding to the A nucleotide designated as residue 1 of Figure 2 in the genomic sequence. This A residue lies 12 base pairs upstream from the beginning of the sequence which encodes the signal domain of the GUS polypeptide. That this site represents the beginning of the first exon, and thus the *Gus-s* transcription initiation site rather than an acceptor splice junction, is strongly supported by the size of the protected fragment in the exon sizing studies and the unambiguous absence of an acceptor splice junction consensus sequence at or near this residue.

Another experiment utilizing a primer derived from within exon 2 results in an extension product whose size concurs with that found in the extension studies described above. Moreover, these results substantiate that intron 1 contains no coding sequence for GUS mRNA (data not shown).

Taken together, the results of cDNA sequence comparisons, primer extension analyses, and nuclease protection studies enabled us to define the structure of *Gus-s*.

Features of Gus-s. Gus-s spans 14 009 bp from the point of transcription initiation (Figure 2, residue 1) through a single polyadenylation signal (AATAAA) at position 13 995 (Proudfoot & Brownlee, 1976) to the first in a stretch of 14


B2
B2
B1
B1
B2
1
2
3
4
5



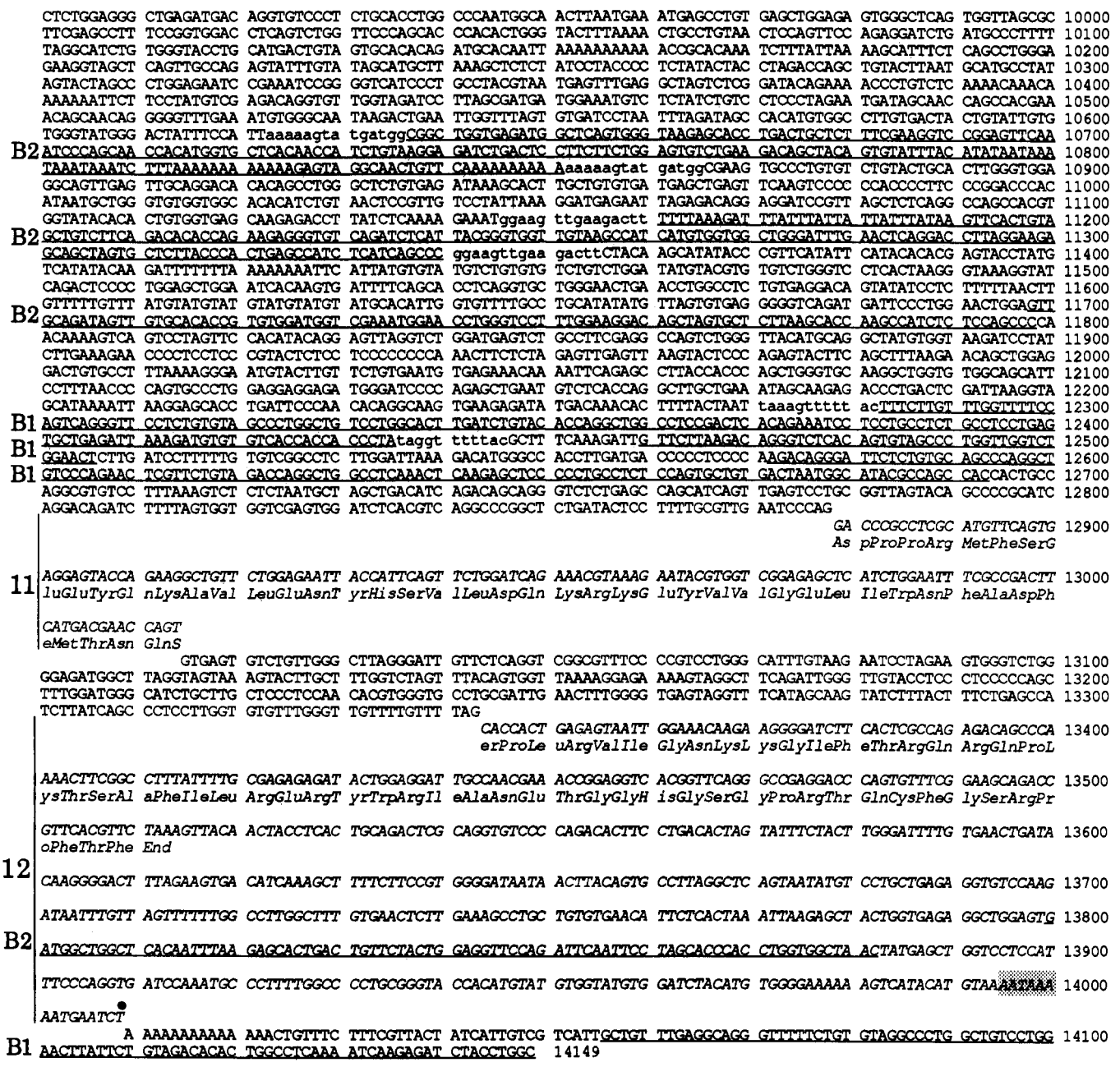


FIGURE 2: Nucleotide sequence of the murine β-glucuronidase gene and its 5′ and 3′ flanking regions. The transcription initiation site (nucleotide 1) and the putative (see text) poly(A+) addition site (nucleotide 14 009) are identified by solid dots. The numbers to the right of each line correspond to the terminal nucleotide of that line. The 2300 nucleotides of 5′ flanking sequence presented are numbered −1 to −2300. The numbers to the left of the sequence designate exons. The deduced amino acid sequence is indicated by the three-letter code below the coding sections of the exons. Stippled boxes beginning at nucleotides −28 and 13 995 indicate the positions of the TATA box and polyadenylation signal, respectively. Repetitive elements of the B1 or B2 families are underlined and are distinguished by labels to the left of the sequence. Lower-case letters indicate direct repeats associated with certain of the repetitive elements (see text and Table II). A 23-nucleotide direct repeat located at 7990–8012 and 8024–8046 and indicated by lower-case letters is not associated with a repetitive element.

A residues located 10 bp downstream from the polyadenylation signal. We are unable to define the base at which polyadenylation begins because of the presence of this homopolymeric sequence. However, on the basis of GUS cDNA sequence (Gallagher et at, 1988), it is clear that the site at which polyadenylation of GUS mRNA occurs is either within or at one end of this homopolymeric A interval.

The murine GUS gene is composed of 12 exons separated by 11 introns (Figures 1 and 2 and Table I). The first exon contains the 5′ untranslated region (12 bp), an appropriate translation initiation motif (CCAGTATGT) at position 8 (Kozak, 1984), and the coding sequence of the 22 amino acid GUS signal sequence as well as that for the first 48 amino acids of the mature GUS polypeptide. The coding sequence of the first 10 amino acids of mature GUS as determined by Lusis and Wilson (personal communication) begins at nucleotide residue 79 (Figure 2). The second amino acid within this sequence of 10 differs from that determined by protein sequencing (Lys instead of Ile). Exon 1 is extremely GC rich (65%). Exon 12 encodes the carboxy-terminal 56 amino acids of GUS as well as the entire 499-nucleotide 3′ untranslated region of GUS mRNA including a single polyadenylation signal, AATAAA (Proudfoot & Brownlee, 1976). The 12 exons vary in size from 85 bases in exon 9 to 666 bases in exon 12.

Comparisons of sequences of *Gus-s* exons with published human GUS cDNA sequence (Oshima et al., 1987) reveal that all of exon 6 and several nucleotides of exon 5 of the murine

Table I: Locations and Sizes of Exons and Introns within the Murine β-Glucuronidase Gene

exon	nucleotide positions	length (bp)	amino acids	intron	nucleotide positions	length (bp)
1	1–222	222	1–70[a]	1	223–1882	1660
2	1883–2068	186	71–132	2	2069–2282	214
3	2283–2464	182	133–193	3	2465–2753	289
4	2754–2896	143	193–241	4	2897–3506	610
5	3507–3694	188	241–303	5	3695–4469	775
6	4470–4613	144	304–351	6	4614–4827	214
7	4828–5006	179	352–411	7	5007–5096	90
8	5097–5243	147	411–460	8	5244–7493	2250
9	7494–7578	85	460–488	9	7579–9454	1876
10	9455–9631	177	489–547	10	9632–12878	3247
11	12879–13014	136	548–593	11	13015–13343	329
12	13344–14009	666	593–648[b]			

[a] Exon 1 also contains 12 nucleotides of 5′ noncoding sequence preceding the ATG initiation codon. [b] Exon 12 also contains 496 nucleotides of 3′ noncoding sequence following the TAA termination codon.

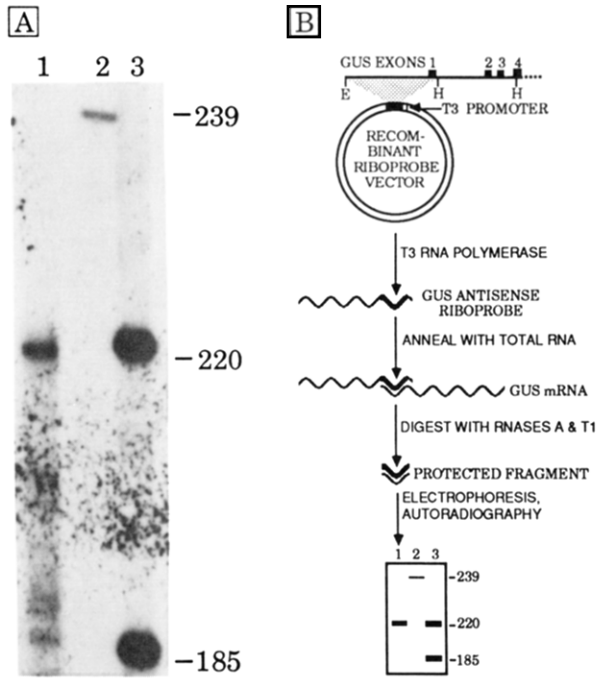


FIGURE 3: Exon mapping of *Gus-s* exon 1. (A) The 220 base pair RNA:RNA (GUS-specific riboprobe:GUS mRNA) hybrid molecule resulting from exon mapping for the first exon of *Gus-s* is shown in lane 1. This fragment was generated from hybridization of an antisense GUS-specific riboprobe (containing the putative exon 1 plus flanking sequence) with total kidney RNA from androgen-treated female mice, followed by digestion with RNases A and T1. Lane 2 contains a 239 base pair radiolabeled marker derived from a *Hin*fI digestion of a GUS/pUC construct. Lane 3 shows a 220 base pair marker from digestion of M13mp10 with *Hph*I. (B) Schematic of GUS exon mapping. The 5′ most *Hin*dIII/*Eco*RI genomic fragment containing the putative exon 1 was cloned into corresponding restriction sites in a Bluescribe riboprobe vector. Labeled, antisense RNA was generated from the T3 promoter. The GUS antisense riboprobe was annealed with total kidney RNA from androgen-treated female mice and subsequently digested with RNases A and T1. The protected fragment (220 base pairs) was sized on a 6% sequencing gel and subjected to autoradiography.

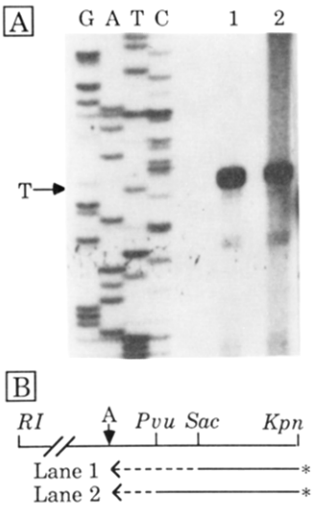


FIGURE 4: Determination of the site of transcription initiation by primer extension analysis. (A) The gel displays the sequence (labeled GATC) of the *Kpn*I/*Eco*RI M13 subclone from which two nested M13 primers used in primer extension were derived. The products of two primer extension reactions using a *Kpn*I/*Sac*I M13 primer and a *Kpn*I/*Pvu*II M13 primer (lanes 1 and 2) comigrate with the T residue (antisense with respect to *Gus-s*) designated to the left of the sequence. Upon lower exposure, the band in lanes 1 and 2 resolves into two bands. (B) Schematic of primer extension strategy showing origin of M13 primers and result. The A residue is the transcription initiation site which corresponds to the antisense T residue identified in (A).

GUS gene are missing from a human GUS cDNA obtained from both placenta and fibroblasts. Whether this deletion represents differential or incorrect splicing of a primary transcript is not known at this time, but of note is the observation that the deleted human sequence begins with a GTG motif commonly found at splice junctions.

The 11 introns of *Gus-s*, which vary in size from 90 base pairs in intron 7 to 3247 base pairs in intron 10, account for 82.5% of the gene. The splice junction sequences (Figure 5) flanking these introns conform to consensus splice junction sequences and the GT-AG rule (Mount, 1982; Breathnach & Chambon, 1981). Availability of intron sequences is useful, although the functional significance of the GUS introns and their role in GUS regulation are not known. Intron sequence may provide important regulatory information as shown by the identification of a tissue-specific enhancer in an intron of a rearranged γ_{2b} immunoglobulin gene (Gillies et al., 1983). In addition, Brinster et al. (1988) recently demonstrated that introns play a role in facilitating transcription of microinjected genes. However, this facilitation of transcription is not observed by transfection of the identical genes into cultured cells, demonstrating that this effect may be only on genes exposed to developmental influences.

Repetitive Elements within Gus-s. The *Gus-s* gene and its flanking sequences have been searched for repetitive elements, direct repeats, and inverted repeats. Several representatives of the B1 and B2 families of murine repetitive elements are found within *Gus-s*, the locations of which are shown in Figures 1 and 2, and a summary of their characteristics is presented in Table II. Comparisons of the sequences of B1 and B2 elements with reported consensus sequences (Kalb et al., 1983; Krayev et al., 1982) are shown in Figure 6. Three of 11 B1 elements (B1d, B1g, and B1f) and 3 of 8 B2 elements (B2c,

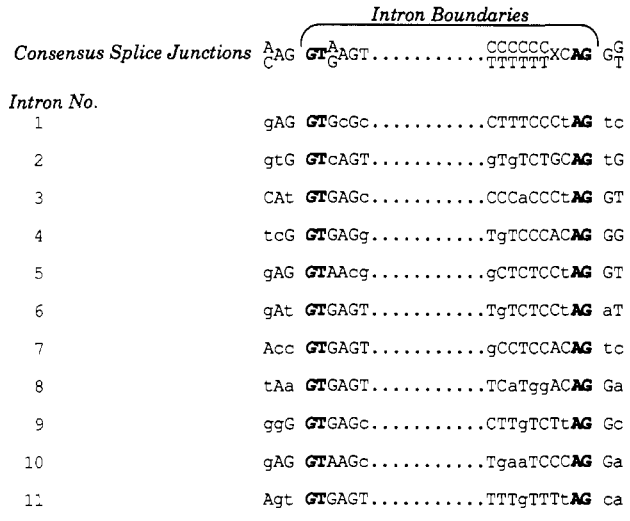


FIGURE 5: Exon–intron splice-junction sequences of *Gus-s*. These sequences are compared to the consensus sequence at the top. Matches are indicated by capital letters; mismatches, by lower-case letters. The intron boundaries are indicated. The invariant GT and AG residues at the respective 5′ and 3′ ends of the introns are in bold-face type.

B2e, and B2f) are complete and are flanked by direct repeats. There is a single complete B1 element (B1f) which is not flanked by direct repeats. One of the complete B2 elements, found within intron 4 (see Figures 1 and 2), creates a restriction fragment length polymorphism that distinguishes DNA of A haplotype mice from that of B haplotype mice (Gallagher et al., 1987), providing a useful GUS haplotype marker. Comparisons of the B1 elements with consensus sequences in Figure 6 indicate that many of the partial elements are lacking portions of the putative RNA polymerase III promoter. A partial B2 element within exon 12 of *Gus-s* is found within the 3′ untranslated region of GUS mRNA (Gallagher et al., 1988).

Another repetitive sequence motif is found within intron 3 of *Gus-s*, consisting of a 38 base pair alternating AC homocopolymer tract (positions 2659–2696). Such alternating purine–pyrimidine homocopolymers have the ability to form left-handed DNA conformational helices known as Z-DNA (Hamada et al., 1984a). Transfection studies with pSV2-cat constructs containing various-length alternating homocopolymers linked to the SV40 promoter and the coding sequence for CAT show as much as a 10-fold increase in CAT gene expression when compared to controls (Hamada et al., 1984b). The affect of this homocopolymer stretch on *Gus-s* gene expression remains to be elucidated.

A novel complex repeat not associated with B1 or B2 elements is located within intron 9 of *Gus-s* and is described in

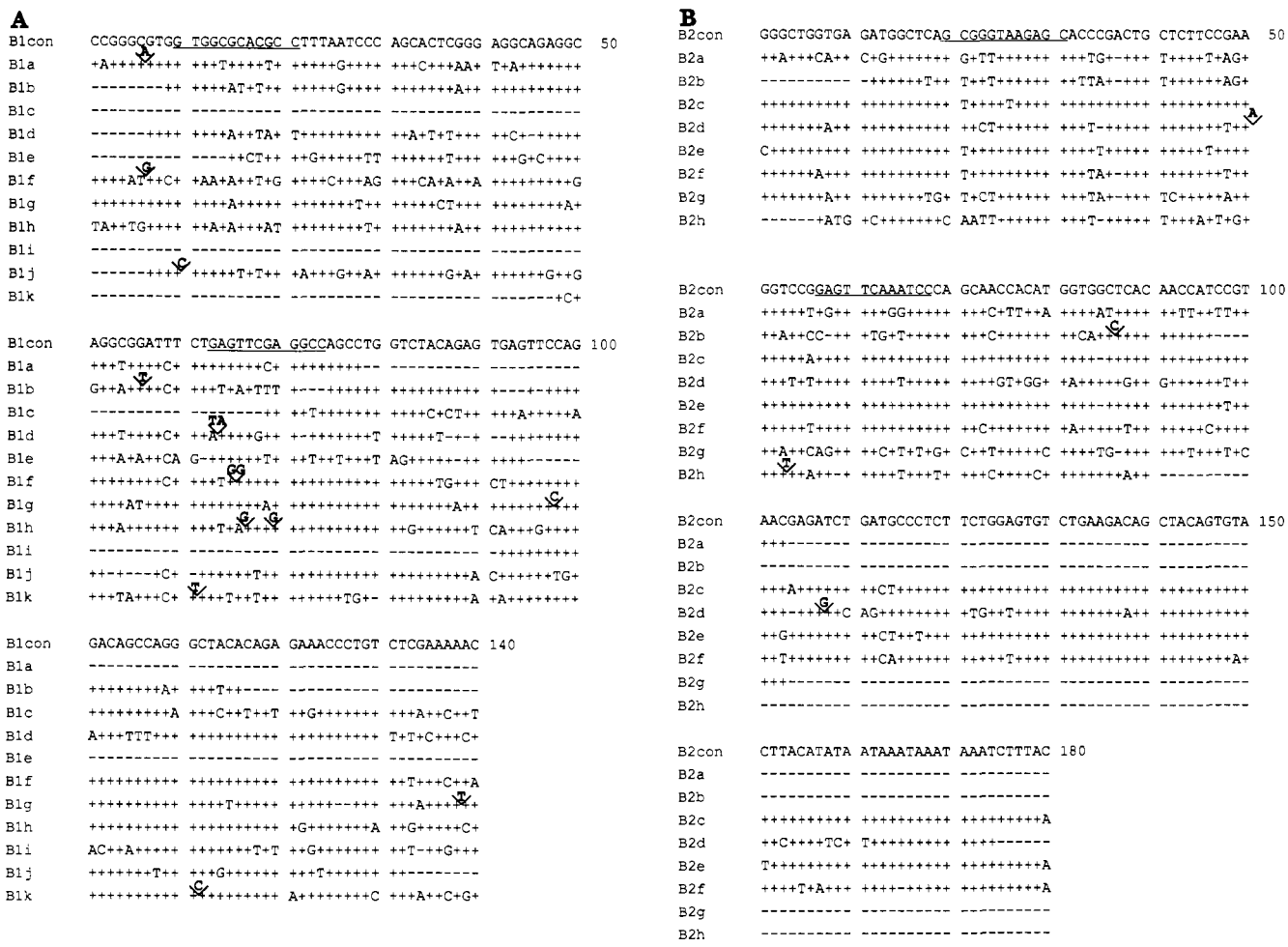


FIGURE 6: Comparison of the B1 and B2 repetitive sequences of the murine β-glucuronidase gene (*Gus-s*). (A) A B1 consensus sequence reported by Kalb et al. (1983) is shown on the top line with the underlined sequences indicating the putative split-promoter for RNA polymerase III. The B1 sequences within and surrounding *Gus-s* are listed below, and their locations are given in Table II; (+) indicates nucleotide identity with the consensus sequence, (−) indicates the nucleotide is missing relative to the consensus sequence, and additional nucleotides that are not found in the consensus sequence are listed above each line. (B) A B2 consensus sequence reported by Krayev et al. (1982) is shown on the top line with the underlined sequences indicating the putative RNA polymerase III split-promoter. The B2 sequences within and surrounding *Gus-s* are listed below, and their locations are given in Table II. The designations within the sequences are as in (A).

Table II: Characteristics of B1 and B2 Repetitive Sequences in the Murine β-Glucuronidase Gene

	nucleotide position	strand[a]	% sequence similarity[b]	size of direct repeat	direct repeat
B1a*[c]	(1114)–(1036)[d]	+	85		
B1b*	(831)–(730)	+	81		
B1c*	3963–4034	+	81		
B1d	4090–4222	+	79	8	CCAAAACA
B1e*	4262–4339	+	74		
B1f	5527–5669	–	82		
B1g	8075–8244	+	90	15	AAAATGACAAGAGCC
B1h	12283–12435	–	84	12	TA$^{G}_{AA}$GTTTTTAC[e]
B1i*	12460–12506	–	82		
B1j*	12572–12693	–	83		
B1k*	14056–14149	–	83		
B2a*	(1977)–(1876)	+	72		
B2b*	(1550)–(1467)	–	76		
B2c	3209–3400	+	96	13	AAAAAACTAAATA
B2d*	6253–6426	–	83		
B2e	10637–10851	+	93	14	AAAAAGTATGATGG
B2f	11170–11340	–	89	15	GGAAGTTGAAGACTT
B2g*	11698–11798	–	72		
B2h*	13800–13882	+	75		

[a] With respect to direction of transcription from the RNA polymerase III promoter: (+) represents transcription from 5′ to 3′ from the same strand as GUS mRNA, and (–) represents transcription from the complementary strand. [b] B1a–B1k sequences are compared to the B1 consensus sequence reported by Kalb et al. (1983). B2a–B2h sequences are compared to the B2 consensus sequence reported by Krayev et al. (1982). [c] Asterisks represent partial B1 or B2 repetitive sequences. [d] Numbers in parentheses represent nucleotide positions upstream from transcription initiation. [e] Mismatches between direct repeats are indicated by the placement of the alternate bases at a single position within the sequence.

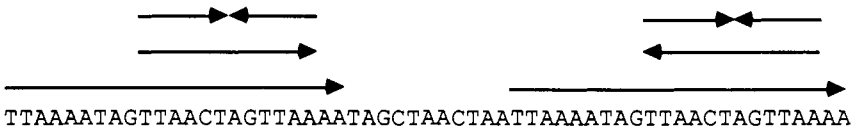


FIGURE 7: Diagram and sequence of a complex repeat within intron 9 of *Gus-s*. Sequence for the complex repeat is given with nucleotide numbers corresponding to those in Figure 2. The 23-base direct repeat is indicated by the two long arrows (7990–8012; 8024–8046). The 12 base pair inverted repeat, contained within the direct repeat, is designated by the arrows spanning positions 7999–8010 and 8033–8044. A palindrome within each inverted repeat is indicated by the adjacent arrows at positions 7999–8004 and 8005–8010 and positions 8033–8038 and 8039–8044.

Figure 7. This repeat consists of a 23 base pair direct repeat (residues 7990–8012 and 8024–8046 in the nucleotide sequence, Figures 2 and 7) separated by 11 bases. Within the direct repeat is a 12-base inverted repeat (7999–8010 and 8033–8044, figure 7) which itself is palindromic (TTAACTAGTTAA). The significance of this repeat, if any, is unknown. However, such inverted and palindromic repeats are capable of forming secondary structures which may serve to control and/or modify gene expression.

Promoter Elements. Two putative RNA polymerase II promoter regions are discernible within and surrounding the 5′ end of *Gus-s*. The location of one of these regions suggests that it controls transcription of *Gus-s* while a second motif toward the 3′ end of the first intron of *Gus-s* apparently controls another transcriptional unit (Wang et al., unpublished experiments).

Appropriately located and positioned at –28 relative to the transcription initiation site of *Gus-s* is a TATA box sequence (Breathnach & Chambon, 1981) which is imbedded within a GC-rich region (62% over 500 bases) containing a number of CpG dinucleotides. Lack of methylation of CpG dinucleotides has been implicated as a possible control mechanism of transcriptional activation in housekeeping genes as well as certain other genes [Wolf & Migeon, 1985; for review, see Bird (1986)]. Whether the TATA box and CpG dinucleotides affect expression of *Gus-s* remains to be established. Sequence flanking the 5′ end of *Gus-s* does not contain a CAAT box motif commonly found in RNA polymerase II transcription units (Breathnach & Chambon, 1981), nor does it have any binding sites for the SpI transcription factor (Briggs, 1986).

A second promoter motif is located within intron 1 and on the same strand as *Gus-s*. The RNA polymerase II recognition signals, CCAAT (position 1623) and TATA (position 1674), are appropriately oriented and spaced relative to each other (Breathnach & Chambon, 1981). Downstream from the TATA box is a consensus translation initiation sequence, at position 1764 (CCGCCATGA) (Kozak, 1984), followed by an open reading frame which extends to nucleotide 1858 (Figure 2). We have identified by nuclease protection analysis a transcription start site at one of the A residues at position 1702 or 1703 (Wang et al., unpublished experiments).

To initiate study of the [*Gus*] complex at the molecular level, we have used a GUS cDNA probe to isolate a series of overlapping GUS genomic clones. Reported here is the complete sequence of the A haplotype of the GUS structural gene, *Gus-s*[a], which provides essential information for comparisons of the various GUS structural alleles. Functional comparisons by genetic manipulation of the DNA within the GUS gene complex among the available GUS haplotypes may reveal differences in DNA structure and function which are causally related to the well-characterized variation in GUS expression among inbred mouse strains.

ACKNOWLEDGMENTS

We acknowledge Drs. C. Duncan, D. Wiginton, J. Degen, S. Degen, and J. Robbins for helpful discussions and advice. We thank Michael Frand, Angela Schmulbach Field, and Lorie Schaefer for expert technical assistance. We also thank Dr. J. C. States for advice with data processing.

REFERENCES

Aviv, H., & Leder, P. (1972) *Proc. Natl. Acad. Sci. U.S.A. 69*, 1408–1412.

Benton, W. D., & Davis, R. W. (1977) *Science (Washington, D.C.) 196*, 180–182.

Bird, A. P. (1986) *Nature (London) 321*, 209–213.
Breathnach, R., & Chambon, P. (1981) *Annu. Rev. Biochem. 50*, 349–383.
Briggs, M. R., Kadonaga, J. T., Bell, S. P., & Tijan, R. (1986) *Science (Washington, D.C.) 234*, 47–52.
Brinster, R. L., Allen, J. M., Behringer, R. R., Gelinas, R. E., & Palmiter, R. D. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 836–840.
Chirgwin, J., Przybyla, A., McDonald, R., & Rutter, W. (1979) *Biochemistry 18*, 5294–5299.
Dale, R. M. K., McClure, B. A., & Houchins, J. P. (1985) *Plasmid 13*, 31–40.
Degen, S. J. F., Heckel, J. L., Reich, E., & Degen, J. L (1987) *Biochemistry 26*, 8270–8279.
Duncan, C. H. (1985) *NEN Product News 4*, 6–7.
Gallagher, P. M., D'Amore, M. A., Lund, S. D., Elliott, R. W., Pazik, J., Hohman, C., Korfhagen, T. R., & Ganschow, R. E. (1987) *Genomics 1*, 145–152.
Gallagher, P. M., D'Amore, M. A., Lund, S. D., & Ganschow, R. E. (1988) *Genomics 2*, 215–219.
Gillies, S. D., Morrison, S. L., Oi, V. T., & Tonegawa, S. (1983) *Cell (Cambridge, Mass.) 33*, 717–728.
Hamada, H., Petrino, M. G., Kakunaga, T., Seidman, M., & Stollar, B. D. (1984a) *Mol. Cell. Biol. 4*, 2610–2621.
Hamada, H., Seidman, M., Howard, B. H., & Gorman, C. A. (1984b) *Mol. Cell. Biol. 4*, 2622–2630.
Kalb, V. F., Glasser, S., King, D., & Lingrel, J. B. (1983) *Nucleic Acids Res. 11*, 2177–2184.
Kozak, M. (1984) *Nucleic Acids Res. 12*, 3873–3893.
Krayev, A. S., Markusheva, T. V., Kramerov, D. A., Ryskov, A. P., Skryabin, K. G., Bayev, A. A., & Georgiev, G. P. (1982) *Nucleic Acids Res. 10*, 7461–7475.
Lalley, P. A., & Shows, T. B. (1974) *Science (Washington, D.C.) 185*, 442–444.
Lund, S. D., Miller, D., Chapman, V., & Ganschow, R. E. (1988) *Genetics 119*, 151–156.
Lusis, A. J., Chapman, V. M., Wangenstein, R. W., & Paigen, K. (1983) *Proc. Natl. Acad. Sci. U.S.A. 80*, 4398–4402.
Maniatis, T., Hardison, R. C., Lacy, E., Lauer, J., O'Connell, C., Quon, D., Sim, G. K., & Efstratiadis, A. (1978) *Cell (Cambridge, Mass.) 15*, 687–701.
Maniatis, T., Fritsch, E. F., & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
Maxam, A. M., & Gilbert, W. (1980) *Methods Enzymol. 65*, 499–560.
Melton, D. A., Kreig, P. A., Rebagliati, M. R., Maniatis, T., Zinn, K., & Green, M. R. (1984) *Nucleic Acids Res. 12*, 7035–7056.
Meredith, S. A., & Ganschow, R. E. (1978) *Genetics 90*, 725–734.
Messing, J., Crea, R., & Seeburg, P. H. (1981) *Nucleic Acids Res. 9*, 309–321.
Mount, S. M. (1982) *Nucleic Acids Res. 10*, 459–472.
Nishimura, Y., Rosenfeld, M. G., Kreibich, G., Gubler, U., Sabatini, D. D., Adesnik, M., & Andy, R. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 7292–7296.
Oshima, A., Kyle, J. W., Miller, R. D., Hoffmann, J. W., Powell, P. P., Grubb, J. H., Sly, W. S., Tropak, M., Guise, K. S., & Gravel, R. A. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 685–689.
Paigen, K. (1961) *Exp. Cell Res. 25*, 286–301.
Paigen, K. (1979) *Annu. Rev. Genet. 13*, 417–466.
Palmer, R., Gallagher, P. M., Boyko, W. L., & Ganschow, R. E. (1983) *Proc. Natl. Acad. Sci. U.S.A. 80*, 7596–7600.
Proudfoot, N. J., & Brownlee, G. G. (1976) *Nature (London) 263*, 211–214.
Queen, C., & Korn, L. J. (1984) *Nucleic Acids Res. 12*, 581–599.
Rigby, P. W. J., Dieckmann, M., Rhodes, C., & Berg, P. (1977) *J. Mol. Biol. 113*, 237–251.
Swank, R. T., Paigen, K., & Ganschow, R. E. (1973) *J. Mol. Biol. 81*, 225–243.
Wolf, S. F., & Migeon, B. R. (1985) *Nature (London) 314*, 467–469.
Yanisch-Perron, C., Vieira, J., & Messing, J. (1985) *Gene 33*, 103–119.