MOLECULAR AND CELLULAR BIOLOGY, Sept. 1989, p. 4074–4078
0270-7306/89/094074-05$02.00/0

Vol. 9, No. 9

DNA Determinants of Structural and Regulatory Variation within the Murine β-Glucuronidase Gene Complex

CYNTHIA J. WAWRZYNIAK,† PATRICIA M. GALLAGHER, MELANIE A. D'AMORE, JEAN E. CARTER, S. DIANE LUND, EUGENE M. RINCHIK,‡ AND ROGER E. GANSCHOW*

Division of Basic Science Research, The Children's Hospital Research Foundation, Cincinnati, Ohio 45229

Received 17 January 1989/Accepted 12 June 1989

The murine β-glucuronidase (GUS) gene complex, [*Gus*], encompasses the GUS structural element, *Gus-s*, and a set of regulatory elements which serve to modulate *Gus-s* expression. Three common GUS haplotypes representing virtually all inbred strains of laboratory mice have been compared with respect to GUS mRNA sequence. Results of such comparisons revealed sequence variations which target the location of one of the GUS regulatory elements to sequences within *Gus-s* and which account for known electrophoretic and heat stability differences among GUS allozymes of the three common GUS haplotypes.

The murine β-glucuronidase (GUS) gene complex, designated [*Gus*] and located on chromosome 5, provides a useful model system for examining the structure and function of mammalian regulatory elements. Elements of [*Gus*] have been identified through characterization of natural variants of GUS expression (for a review, see reference 23).

The elements of [*Gus*] include the GUS structural gene, *Gus-s*, and a set of regulatory elements which serve to modulate *Gus-s* expression. Three common alleles of *Gus-s* (*Gus-s*^{a}, -*s*^{b}, and -*s*^{h}) specify allozymes which differ in electrophoretic mobility, heat stability, or both (15, 18, 22, 27). Included among the GUS regulatory elements is a *cis*-active element, designated *Gus-u*, which controls the levels of GUS synthesis in all tissues at all times (17). *Gus-u*, while exerting control over the rate of GUS synthesis, does not influence the levels of GUS mRNA in tissue, suggesting that the effects of this regulatory element on GUS synthetic rates occur subsequent to the processing of GUS transcripts, possibly at the translational level (28). Combinations of specific alleles of *Gus-s* and its regulatory elements define three common GUS haplotypes: [*Gus*]^a, [*Gus*]^b, and [*Gus*]^h (Table 1 defines each haplotype insofar as is necessary for this report).

We are attempting to utilize the natural variants of GUS structure and regulation to identify and characterize the DNA determinants of the underlying elements within [*Gus*]. One aspect of this strategy is comparative sequencing of exons of *Gus-s* among the common GUS haplotypes. Results of this analysis account for electrophoretic and heat stability differences which distinguish GUS allozymes among the three common haplotypes of [*Gus*]. More important, these results permit us to infer the location of the *Gus-u* regulatory element within [*Gus*].

Cloning and structural characterization of *Gus-s*^{b} and *Gus-s*^{h}. Using a cDNA probe representing exons 4 through 12 of the murine GUS structural gene, *Gus-s*, we have identified and isolated single recombinant genomic clones which contain the entire 14-kilobase (kb) gene from each of the three common murine GUS haplotypes, A, B, and H (Fig. 1).

An H haplotype genomic library was prepared from spleen DNA of the C3H/HeJ inbred mouse strain and constructed in the λ bacteriophage vector EMBL3. A random primer-labeled (6) fragment (1.4-kb *Pst*I-*Hin*dIII) from the GUS cDNA clone pGUS-1 (7, 24) was used as a hybridization probe to screen approximately 10^6 recombinant phage in this library by in situ hybridization (1, 20). A cloned genomic DNA, designated λHGus-1, was isolated in this manner and contains all of *Gus-s*^{h} within a 20-kb insert (Fig. 1). λHGus-1 was purified (19) and characterized by restriction mapping with *Eco*RI, *Sal*I, and *Hin*dIII.

cosBGus-1 was isolated from a cosmid library prepared with DNA from the YBR (B haplotype) inbred mouse strain in the laboratory of Miriam Meisler at the University of Michigan. cosBGus-1 contains all of *Gus-s*^{b} within a 40-kb insert (Fig. 1).

λAGus-4, previously published by this laboratory, contains a 20-kb insert within which is found the A allele of *Gus-s* (4). We have previously reported the complete sequence organization of this allele, which includes 14,009 nucleotides and 12 exons (4). Restriction mapping of λHGus-1 and cosBGus-1 revealed patterns very similar to those of λAGus-4 (Fig. 1).

Comparison of exon sequences among the three common alleles of *Gus-s*. The sequencing strategy for the informative clones used in determining the exon sequences of the B and H alleles of *Gus-s* is shown in Fig. 1. Sequence comparisons among the exons of the A, B, and H alleles are presented as composite mRNAs in Fig. 2.

The mRNA sequence of the A allele (Fig. 2) is that previously reported by our laboratory (8). Comparison at the nucleotide level revealed that the A and B mRNAs differ at 17 positions, 14 of which are transitions, while 3 are transversions (Fig. 2). The mRNA sequence of the H allele differs from that of the B allele by a single transition at position 272 and from that of the A allele by a total of 18 residues. The single difference between the B and H alleles predicts a restriction-fragment-length polymorphism resulting from the change of a *Hph*I site in the A and B alleles to a *Fok*I site in the H allele. This difference provides the only restriction-fragment-length polymorphism which distinguishes the B and H alleles among inbred mouse strains (7).

Comparison of the polypeptide sequences of the A, B, and H

* Corresponding author.
† Present address: Miami Valley Laboratories, Cincinnati, OH 45239.
‡ Present address: Oak Ridge National Laboratory, Biology Division, Oak Ridge, TN 37831-8077.

TABLE 1. Definitions of the three common β-glucuronidase haplotypes[a]

[*Gus*] haplotype	Allele of:		Allozyme	Representative inbred strain
	Gus-s	*Gus-u*		
A ([*Gus*]^a)	*a*	*b*	GUS-A	BALB/cJ
B ([*Gus*]^b)	*b*	*b*	GUS-B	YBR
H ([*Gus*]^h)	*h*	*h*	GUS-H	C3H/HeJ

[a] Information is summarized from references 17 and 23.

allozymes. Of the 17 nucleotide differences between A and B mRNAs, 1 was in the 5′ untranslated region, while 6 clustered in the 3′ untranslated region. The remaining 10 differences appeared in the amino acid-coding portion of GUS mRNA. Seven of the ten differences were silent, producing no amino acid changes, while the remaining three specified differences in two amino acids (Fig. 2). The variation at positions 806 and 807 between A mRNA and that of either B or H predicted an aspartic acid at residue 265 in the A polypeptide versus a glycine in the B or H polypeptide. At position 970, a guanine in A mRNA was replaced by an adenylate residue in B and H mRNAs. This difference predicted a valine at residue 320 of the A polypeptide versus isoleucine in the B or H polypeptide. Of the two amino acid differences described, only the difference at residue 265 resulted in a change in the net charge of the GUS polypeptide, providing the structural basis for the observed difference in electrophoretic mobility between GUS-A and either GUS-B or GUS-H (15, 18).

As described above, the mRNA sequence of the H allele differed from that of the B allele by a single base at position 272. This difference predicted an isoleucine at residue 87 in the H polypeptide versus a threonine in the A or B polypeptide, resulting in a slight polarity difference between the polypeptides. Since this variation represented the only difference between the H and B polypeptides, it must provide the structural basis for the historically observed heat lability of the enzyme product of the H allele relative to that of either B or A (9).

According to the algorithm of Garnier et al. (10), the single amino acid difference between B and H polypeptides at residue 87 predicts a perturbation in protein secondary structure. This algorithm predicts an extended or beta-sheet conformation in the B polypeptide from amino acid 74 through 80 followed by five residues primarily in random coil-and-turn configurations and then by an α-helical stretch from residue 86 through 94. In contrast, the same algorithm predicts a beta-sheet conformation for the H polypeptide from residue 74 to 85, with an interruption by turns or random coils of only three residues, followed by a region of α-helix extending from residue 86 to 93 (Fig. 3).

The algorithm of Chou and Fasman (3) also predicts a difference in protein secondary structure as a result of the single amino acid difference between the B and H polypeptides (data not shown).

Furthermore, the algorithms of Kyte and Doolittle (14) and Hopp and Woods (13) predict that over a stretch of approximately 10 amino acids extending from residue 83 to 92, the B polypeptide is significantly more hydrophilic than the same stretch of amino acids in the H polypeptide (Fig. 3B).

Predictions of protein secondary structure and hydropathicity were made by using the DNASTAR Sequence Analysis Program (DNASTAR, Inc., Madison, Wis.).

Targeting of the GUS regulatory element, *Gus-u*. The solitary nucleotide difference between the B and H mRNAs suggested a structural basis for the *cis*-active regulatory

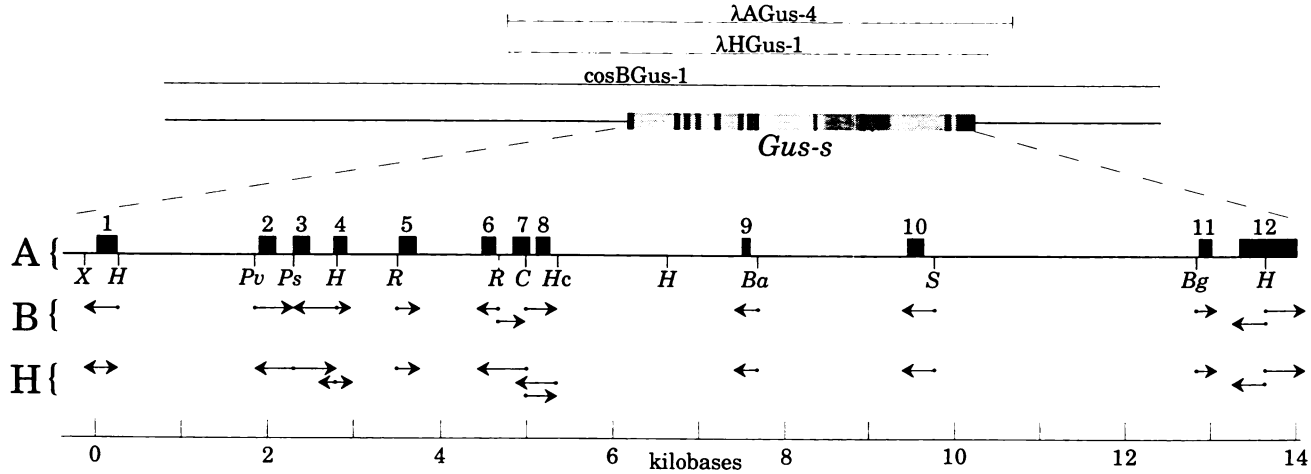
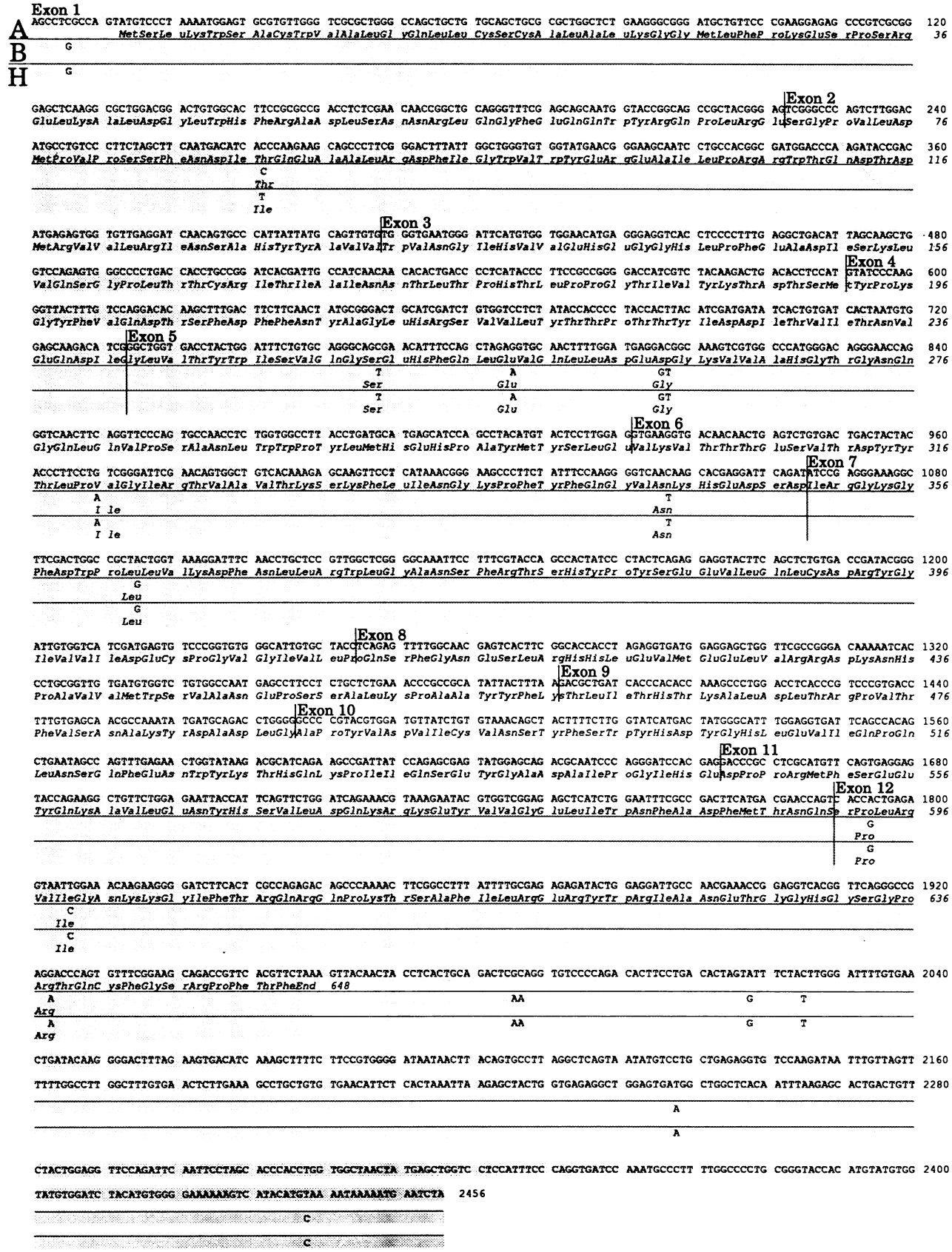


FIG. 1. Sequencing strategy for *Gus-s* exons of the three common murine GUS haplotypes. λAGus-4, λHGus-1, and cosBGUS-1 designate genomic clones containing *Gus-s* alleles carried by the A, H, and B haplotypes, respectively. Restriction fragments from λHGus-1 and cosBGus-1 were isolated by gel electrophoresis through low-melting-temperature agarose and were ligated into M13mp18 or M13mp19. Single-stranded M13 templates were sequenced by the quasi-end-labeling adaptation of the dideoxy chain termination method (5). In addition, some restriction fragments were 5′ end labeled with [γ-^{32}P]ATP (3,000 Ci/mmol) and sequenced by the method of Maxam and Gilbert (21). Resulting DNA sequences were entered, stored, and analyzed by using the Microgenie DNA Sequence Analysis Program (Beckman Instruments, Inc., Fullerton, Calif.) (26). Below the genomic clones is a graphic representation of the published exon-intron organization for the A haplotype allele of *Gus-s* (4), in which darkened blocks represent exons of *Gus-s*. Indicated by vertical lines extending down from the expanded exon map of *Gus-s* are informative restriction sites which are designated as follows: Ba, *Bam*HI; Bg, *Bgl*II; C, *Cla*I; H, *Hin*dIII; Ps, *Pst*I; Pv, *Pvu*II; R, *Eco*RV; S, *Sph*I; and X, *Xba*I. The dashed line at approximately +4.7 kb designates an *Eco*RV site which is absent in the A haplotype but present in the B and H haplotypes. Sequencing strategies for exons of the B and H haplotypes are depicted by horizontal arrows indicating the direction and extent of each sequence determination, with haplotype designations shown to the left of each set of arrows.


Exon 1
Exon 2
Exon 3
Exon 4
Exon 5
Exon 6
Exon 7
Exon 8
Exon 9
Exon 10
Exon 11
Exon 12
A
B
H

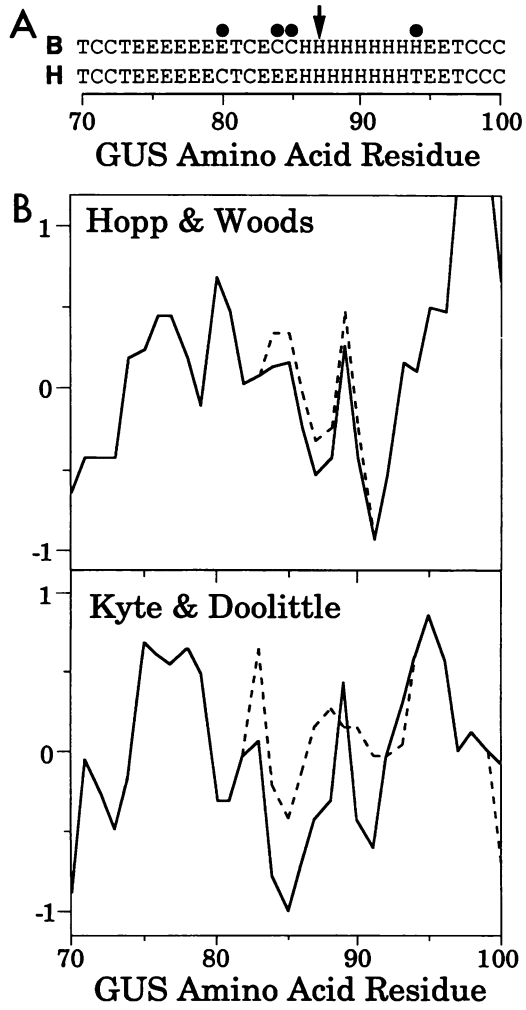


FIG. 3. (A) Comparison of predicted secondary structure characteristics of the region of the GUS polypeptide which contains the solitary difference in amino acid sequence between the B and H haplotypes. Conformational states predicted by the algorithm of Garnier et al. (10) for each amino acid between residues 70 and 100 are summarized. Designations of conformational state are as follows: C, coil; E, extended chain; H, α helix; T, turn. Solid circles represent residues at which differences in conformational states are predicted between the B and H polypeptides. Location of the single amino acid difference which distinguishes the B from the H polypeptides is indicated by the arrow. (B) Comparative hydropathicity profiles for the GUS polypeptides of the B and H haplotypes. Hydropathicity data derived from the Hopp and Woods algorithm (13) for the GUS polypeptides between amino acid residues 70 and 100 are presented in the top panel, while the bottom panel represents similar data derived from the algorithm of Kyte and Doolittle (14). For each panel, regions greater than zero are hydrophilic, while regions less than zero are hydrophobic. Solid and dashed lines represent hydropathicities for polypeptides of the B and H haplotypes, respectively.

element, *Gus-u*. Mice of the A and B haplotypes are homozygous for the wild-type allele, *Gus-u*^{b}, while the three- to fivefold-lower levels of GUS activity in tissue which are characteristic of the H haplotype are controlled by the mutant allele, *Gus-u*^{h} (17).

We have previously demonstrated that the mutant allele of *Gus-u* in H haplotype mice alters the rates of GUS synthesis without altering the levels of GUS transcripts, thus precluding an effect on transcriptional or transcript-processing mechanisms (28). From this we inferred that *Gus-u* encoded a recognition site within either the processed GUS transcript or the GUS polypeptide. Since only a single nucleotide difference was observed for GUS mRNAs in mice which manifest the variant alleles of *Gus-u*, the location of this difference targeted *Gus-u* to a region within the second exon of *Gus-s*, thus placing it within the coding sequence of the gene over which it exerts control.

The effect of *Gus-u* on the expression of GUS must occur subsequent to the processing of GUS transcripts but prior to the acquisition by GUS of the ability to react with polyclonal antibodies used in establishing the rate of GUS synthesis. In their evaluation of the parameters responsible for the differential expression of GUS between murine tissues, Bracey and Paigen (2) suggest that such differences arise primarily from alterations in translational yield, which they define as the number of mature enzyme molecules formed per minute per mRNA molecule. Variations in translational yield can be due either to alterations in translational efficiency or to an event which promotes the maturation of GUS polypeptides to a point at which they can be recognized by GUS antibodies.

If *Gus-u* regulates translational yield of GUS polypeptides, then the expression of this element could vary between the GUS haplotypes so that transcripts derived from the wild-type *Gus-u*^{b} allele are translated more efficiently than those from the *Gus-u*^{h} allele. Differences between transcripts involving primary sequences or secondary structure or both are known to alter the rate of attachment to ribosomes or other stages of translation initiation or both. (11).

DNA variations responsible for differences in features of GUS structure. The amino acid difference at residue 265, which predicted an aspartic acid in the A polypeptide as opposed to a glycine at this position in the B and H polypeptides, resulted in a change in the net charge of the protein. This amino acid difference provided the structural basis for the observed difference in electrophoretic mobility between GUS-A and either GUS-B or GUS-H, since GUS-A migrates more rapidly toward the anode at alkaline pH than either GUS-B or GUS-H does (15, 18). This conclusion is consistent with a previous prediction that the difference in electrophoretic mobility among the GUS allozymes is due to a single charge difference in the GUS polypeptide (23).

A single amino acid difference between the B and H polypeptides at residue 87 provides a structural explanation

FIG. 2. Derived sequences of murine GUS mRNAs and their encoded polypeptides for each of the common GUS haplotypes (A, B, and H). The GUS mRNA sequence for the A haplotype was published previously (8), while those of the B and H haplotypes represent a compilation of novel sequence from *Gus-s* exons contained within the genomic clones (Fig. 1). Nucleotides are numbered in the 5′-to-3′ direction from 1 to 2456. Nucleotide 1 represents the base at which transcription is initiated (4). Amino acids are numbered in italics (*1* to *648*), beginning with the first residue of the signal sequence (4). Exon locations are demarcated by vertical lines, with exon numbers located above the nucleotide sequence at the beginning of each exon. A single line of sequence shown for a region of mRNA or polypeptide or both indicates identity among the haplotypes (e.g., nucleotide sequence 121 to 240). However, if sequence differences exist within a line, then only the sequences representing such differences for the B and H haplotypes are shown beneath the complete A haplotype sequence (e.g., nucleotide sequence 241 to 360).

for observed differences in heat stability of GUS activity among the A, B, and H allozymes (9). The heat lability of GUS-H activity relative to that of GUS-B is observed regardless of pH, buffering anion, buffer molarity, or ionic strength (12). The sequence variation between the B and H polypeptides also underlies significant differences in the predicted secondary structure and hydropathy of the protein, thus providing a sound basis for the established difference in heat stability between GUS-H and GUS-B activities.

Evolution of the three common GUS haplotypes. The nucleotide sequence data for the A, B, and H mRNAs can be used to calculate the point at which chromosomes carrying these GUS structural alleles diverged. Assuming a mutation rate of approximately 4.2×10^{-9} to 7.0×10^{-9} substitutions per site per year for neutral mutations (16, 25), the A and B alleles diverged approximately 8.7×10^5 to 14.5×10^5 years ago. This estimate is based on the observation that of 17 nucleotide differences detected between the A and B mRNAs, 15 were silent.

We acknowledge Kenneth Blumenthal and Steven Small for helpful discussions during the preparation of the manuscript.

This work was supported by Public Health Service grants DK 14770 to R.E.G. and GM 12089 to C.J.W. from the National Institutes of Health. C.J.W. was also supported by a Trustee Grant from the Children's Hospital Research Foundation, Cincinnati, Ohio. M.A.D. and S.D.L. are predoctoral trainees of the Graduate Program in Developmental Biology, University of Cincinnati, Cincinnati, Ohio. The manuscript was coauthored by a contractor of the U.S. government under contract DE-AC05-84OR21400.

LITERATURE CITED

1. **Benton, W. D., and R. W. Davis.** 1977. Screening λgt recombinant clones by hybridization to single plaques in situ. Science **196**:180–182.
2. **Bracey, L. T., and K. Paigen.** 1987. Changes in translational yield regulate tissue-specific expression of β-glucuronidase. Proc. Natl. Acad. Sci. USA **84**:9020–9024.
3. **Chou, P. Y., and G. D. Fasman.** 1978. Empirical predictions of protein conformations. Annu. Rev. Biochem. **47**:251–276.
4. **D'Amore, M. A., P. M. Gallagher, T. R. Korfhagen, and R. E. Ganschow.** 1988. Complete sequence and organization of the murine β-glucuronidase gene. Biochemistry **27**:7131–7140.
5. **Duncan, C. H.** 1985. Quasi-end labeling in M13 dideoxy sequence analysis. NEN Product News **4**:6–7.
6. **Feinberg, A. P., and B. Voglestein.** 1983. A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Anal. Biochem. **132**:6–13.
7. **Gallagher, P. M., M. A. D'Amore, S. D. Lund, R. W. Elliott, J. Pazik, C. Hohman, T. R. Korfhagen, and R. E. Ganschow.** 1987. DNA sequence variation within the β-glucuronidase gene complex among inbred strains of mice. Genomics **1**:145–152.
8. **Gallagher, P. M., M. A. D'Amore, S. D. Lund, and R. E. Ganschow.** 1988. The complete nucleotide sequence of murine β-glucuronidase mRNA and its deduced polypeptide. Genomics **2**:215–219.
9. **Ganschow, R. E., and K. Paigen.** 1968. Glucuronidase phenotypes of inbred mouse strains. Genetics **59**:335–349.
10. **Garnier, J., D. J. Osguthorpe, and B. Robson.** 1978. Analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular proteins. J. Mol. Biol. **120**:97–120.
11. **Gehrke, L.** 1987. Differential translation of eukaryotic messenger RNAs: the role of messenger RNA secondary structure, p. 367–378. *In* J. Ilan (ed.), Translational regulation of gene expression. Plenum Publishing Corp., New York.
12. **Herrup, K., and R. J. Mullen.** 1977. Biochemical and genetic factors in the heat inactivation of murine β-glucuronidase. Biochem. Genet. **15**:641–653.
13. **Hopp, T. P., and K. R. Woods.** 1981. Prediction of protein antigenic determinants from amino acid sequences. Proc. Natl. Acad. Sci. USA **78**:3824–3828.
14. **Kyte, J., and R. F. Doolittle.** 1982. A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. **157**:105–132.
15. **Lalley, P. A., and T. B. Shows.** 1974. Lysosomal and microsomal glucuronidase: genetic variant alters electrophoretic mobility of both hydrolases. Science **185**:442–444.
16. **Li, W.-H., C.-I. Wu, and C.-C. Luo.** 1985. A new method for estimating synonymous and nonsynonymous rates of nucleotide substitution considering the relative likelihood of nucleotide and codon changes. Mol. Biol. Evol. **2**:150–174.
17. **Lusis, A. J., V. M. Chapman, R. W. Wangenstein, and K. Paigen.** 1983. *Trans*-acting temporal locus within the β-glucuronidase gene complex. Proc. Natl. Acad. Sci. USA **80**: 4398–4402.
18. **Lusis, A. J., and K. Paigen.** 1978. The large scale isolation of mouse β-glucuronidase and comparison of allozymes. J. Biol. Chem. **253**:7336–7345.
19. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor. N.Y.
20. **Maniatis, T. R., C. Hardison, E. Lacy, J. Lauer, C. O'Connell, D. Quon, G. K. Sim, and A. Efstratiadis.** 1978. The isolation of structural genes from libraries of eukaryotic DNA. Cell **15**: 687–700.
21. **Maxam, A. M., and W. Gilbert.** 1980. Sequencing end-labeled DNA with base-specific cleavages. Methods Enzymol. **65**:499–560.
22. **Paigen, K.** 1961. The effect of mutation on the intracellular location of β-glucuronidase. Exp. Cell Res. **25**:286–301.
23. **Paigen, K.** 1979. Acid hydrolases as models of genetic control. Annu. Rev. Genet. **13**:417–466.
24. **Palmer, R., P. M. Gallagher, W. L. Boyko, and R. E. Ganschow.** 1983. Genetic control of levels of murine kidney glucuronidase mRNA in response to androgen. Proc. Natl. Acad. Sci. USA **80**:7596–7600.
25. **Perler, F., A. Efstratiadis, P. Lomedico, W. Gilbert, R. Kolodner, and J. Dodgson.** 1980. The evolution of genes: the chicken preproinsulin gene. Cell **20**:555–566.
26. **Queen, C., and L. J. Korn.** 1984. A comprehensive sequence analysis program for the IBM personal computer. Nucleic Acids Res. **12**:581–599.
27. **Swank, R. T., K. Paigen, and R. E. Ganschow.** 1973. Genetic control of glucuronidase in mice. J. Mol. Biol. **81**:225–243.
28. **Wawrzyniak, C. J., S. A. Meredith, and R. E. Ganschow.** 1989. Two genetic elements regulate murine β-glucuronidase synthesis following transcript accumulation. Genetics **121**:119–124.