MOLECULAR AND CELLULAR BIOLOGY, Dec. 1985, p. 3634–3639
0270-7306/85/123634-06$02.00/0

Vol. 5, No. 12

Structure of the Rat α_1-Acid Glycoprotein Gene

YU-CHENG J. LIAO,[1] JOHN M. TAYLOR,[2] JAMES L. VANNICE,[3] GARY A. CLAWSON,[1] AND EDWARD A. SMUCKLER[1]*

Department of Pathology, School of Medicine, University of California, San Francisco, California 94143[1]; Gladstone Foundation, San Francisco General Hospital, San Francisco, California 94140[2]; and Department of Molecular Biology, Genentech Inc., South San Francisco, California 94080[3]

Received 25 June 1985/Accepted 16 September 1985

The complete nucleotide sequence of the rat α_1-acid glycoprotein gene has been determined from an isolated lambda recombinant bacteriophage. Southern blot analysis and DNA sequencing indicate that there is only one gene per genome; it contains six exons and is located within a 3,200-base-pair fragment starting from a TATA box and extending to the polyadenylation signal AATAAA. Transcription starts 37 base pairs upstream from the beginning of the translation codon ATG. The TATA box (TATAAA) lies 26 base pairs upstream from this site. The gene contains several potential glucocorticoid receptor-binding sites, both inside and outside the structural gene.

α_1-Acid glycoprotein (α_1-AGP) is a single polypeptide with a molecular mass of 23 kilodaltons plus three to five carbohydrate side chains; the latter account for about 45% of its total mass of 45 kilodaltons (21). Although its precise function is not known, the protein is highly conserved in evolution and maintained during life, even during severe starvation. Synthesis occurs in the liver and probably only in that organ. Acute inflammation caused by physical injury, by bacterial or protozoan infections, or associated with lupus erythematosus or neoplasia results in prompt elevation of the serum levels of this protein. Experimental subcutaneous injection of turpentine to produce sterile abscesses in the rat results in a 90- to 100-fold increase of α_1-AGP mRNA in hepatic cytoplasm, and its translation product becomes a major serum glycoprotein (15). Exposure of hepatic cells in culture to glucocorticoid hormones results in a similar induction of α_1-AGP (1, 26). In vitro studies indicate that α_1-AGP inhibits platelet aggregation (24), lymphocyte transformation (4), cell-collagen interaction (5), and the invasion of malarial protozoa into erythrocytes (6). The resemblance of α_1-AGP to erythrocyte membrane sialoglycoproteins, the homology of its amino acid sequence in the carboxyl terminal region with that in the immunoglobulin G H-chain constant region (22), and its induction by glucocorticoid hormones suggest that it might be involved in the immune system and might act both as a protective decoy during parasitism and as an immunosuppressant during acute inflammation in response to tissue damage.

The remarkable induction of α_1-AGP by glucocorticoids and the requirement of ongoing protein synthesis during gene expression, which is in contrast to mouse mammary tumor virus, make α_1-AGP a more complex and complete model for studying steroid hormone regulation in mammalian genes; in addition, the physiological role of this protein in inflammatory homeostasis provides a model for multiorgan investigation.

Ricca and Taylor (16) have previously cloned and sequenced rat AGP cDNA, and Yamamoto and colleagues (10) have cloned and characterized rat glucocorticoid receptor cDNA. In the present study, we used rat AGP cDNA to isolate the corresponding gene sequence from a rat genomic library. Analysis of this gene indicates it exists as a single copy consisting of six exons and has four potential glucocorticoid-binding sites. Additionally, there are several sequences that may be involved in the regulation of its expression.

Synthetic 17-base primer and four dideoxynucleotide triphosphates for dideoxy sequencing were purchased from P-L Biochemicals, Inc., Milwaukee, Wis. Nitrocellulose (BA 85) was purchased from Schleicher & Schuell, Inc., Keene, N.H. Restriction enzymes and DNA modification enzymes were obtained from New England BioLabs, Inc., Beverly, Mass., and were used according to the specifications of the manufacturer.

The rat genomic library, provided by T. Sargent (National Institutes of Health), was composed of a partial *Eco*RI digest of rat hepatic DNA cloned into bacteriophage Charon 4A. The 792-base-pair (bp) *Pst*I-*Pst*I DNA fragment derived from rat α_1-AGP cDNA plasmid pAGP663 (16) was used as a hybridization probe after labeling with [α-^{32}P]dCTP by the nick translation method (8). Screening and hybridization were carried out as described by Maniatis et al. (8). Positive plaques were purified by three further cycles of plating at densities of 200 plaques per plate. Phage DNA was prepared as described by Maniatis et al. (8).

Recombinant phage DNA was analyzed by Southern blot hybridization (25) after restrictive digestion and electrophoresis in 0.75% agarose gels. Hybridizations were performed at 68°C with 5×10^6 cpm/ml of the PstI-PstI fragment of rat α_1-AGP plasmid pAGP663 as a probe. Filters were autoradiographed at −70°C with intensifier screens.

DNA sequencing was carried out by the dideoxy chain-termination method (17, 23) by cloning restricted DNA fragments into various polylinker sites of bacteriophages M13, mp18, and mp19 and then by priming for chain extension and for base-specific termination. DNA sequence data were analyzed with the computer programs of Martinez et al. (9).

Isolation of a recombinant bacteriophage containing the rat α_1-AGP gene and strategy of sequencing. The genomic library was prepared by Sargent et al. (18) from a partial *Eco*RI digest of Sprague-Dawley rat DNA cloned into lambda bacteriophage Charon 4A. Plaques (4×10^6) equivalent to 10 genomes were screened with nick-translated cDNA probe of AGP. Among nine positive clones, the one with the strongest hybridization signal was selected and purified by low-density

* Corresponding author.

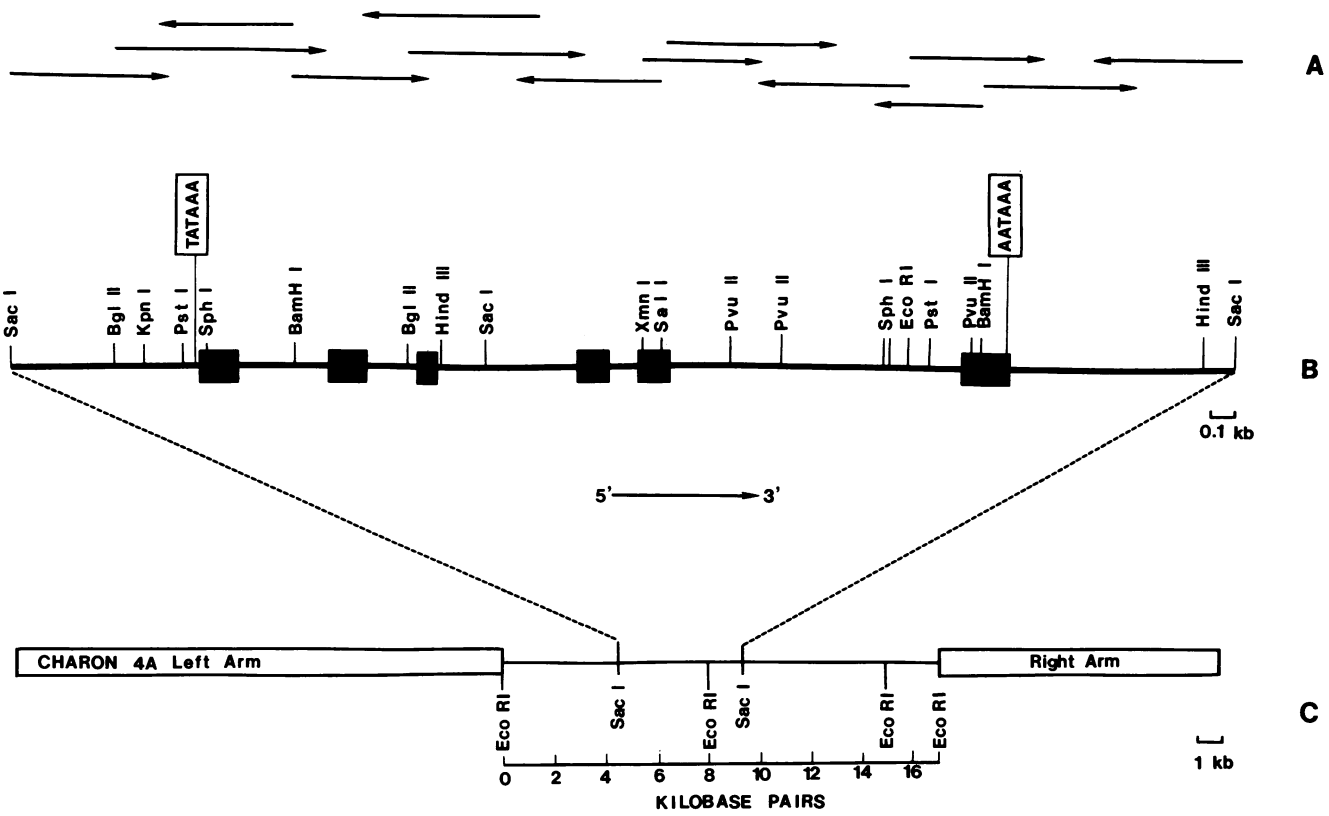

FIG. 1. Map of the rat α_1-AGP gene. (A) Strategy of sequencing. Arrows indicate the directions and lengths of the M13 subclones used in dideoxynucleotide sequencing. (B) Restriction sites of sequenced rat α_1-AGP gene. The six exons of the gene are indicated by solid boxes. (C) Organization of rat DNA insertion and the recombinant clone. Only *Eco*RI and *Sac*I sites are used as landmarks.

plating. Restrictive digestion with *Eco*RI and Southern blot analysis indicated that this clone contains a 17-kilobase-pair (kbp) insert with two internal and two flanking *Eco*RI sites. Based on the intensity of the hybridization signals, 80% of the exons were located in the 8.5-kbp *Eco*RI fragment and 20% were in the 7-kbp *Eco*RI fragment; none were found in the 2-kbp fragment. The order and arrangement of *Eco*RI sites (Fig. 1) were determined by single- and double-restriction mapping and patterns of Southern hybridization.

To confirm the copy number and absence of sequence rearrangements of the α_1-AGP gene during recombinant DNA manipulations, we performed Southern hybridizations with various restriction enzymes on both cellular and recombinant phage DNA. The results indicated that there is only one copy of the α_1-AGP gene in rat genome and that no rearrangement occurred during cloning and bacteriophage propagations. All fragments detected by ^{32}P-labeled cDNA probe were identical in size in cellular and recombinant phage DNA with different restriction digests (Fig. 2).

A sequencing strategy was formulated by taking advantage of the existing *Sal*I and *Bam*HI sites in the α_1-AGP cDNA. Recombinant bacteriophage DNA was digested with *Sal*I, *Bam*HI, and *Eco*RI and then directly subcloned into M13 bacteriophage polylinker sites. Because of the rarity of the *Sal*I site in the mammalian genome and lambda bacteriophage, we were successful in cloning the reading frame on the first trial. The sequencings were then expanded from the *Sal*I site toward both the 5′ and 3′ ends of the gene. Further subclonings were undertaken with identified restriction sites of the new sequence and convenient *Sac*I sites found on positively hybridized fragments on Southern blot analysis. The total sequencing was completed by overlapping sequences of M13 subclones and extended about 1 kbp beyond the TATA box and the polyadenylation signal AATAAA at the 5′ and 3′ ends, respectively. The 5′- to 3′-end orientation of the α_1-AGP gene in recombinant lambda bacteriophage was determined from the nucleotide sequences of exons and from the locations of restriction sites, including *Eco*RI, *Sac*I, *Bam*HI, and *Sal*I, on Southern hybridizations.

Organization of AGP gene. The rat AGP gene consists of six exons and five introns and strictly observes Chambon's rule that all introns begin with dinucleotide G-T and end with A-G (Fig. 3). The more detailed consensus sequences of donor G-T-$\frac{A}{G}$-T-G-T and acceptor $(\frac{T}{C})_{11}$-N-$\frac{C}{T}$-A-G proposed by Breathnach and Chambon (2) are reasonably matched by the splicing junctions of the introns and exons in the rat α_1-AGP gene.

Comparison of exons of the AGP gene with previously published cDNA sequences shows that there is only one base difference, G instead of A at the fifth nucleotide after the translation stop codon TAG on the 3′ end. The difference could be due to intraspecies polymorphism or replication error of lambda bacteriophage. It has no effect on amino acid coding sequences.

5′-End region. The transcriptional initiation site was determined previously by reverse transcriptase-catalyzed cDNA synthesis in the presence of a 67-bp *Dde*I cDNA fragment as primer and chain-terminating dideoxynucleotide

triphosphates (16). The nucleotide residue A, 27 bases downstream from the TATA box, is the transcriptional start.

The Goldberg-Hogness box (TATAAAA) is located between 21 and 27 bases upstream from the transcriptional start. There is no apparent consensus sequence of a CAAT box in its anticipated (50 bp) upstream position; however, there may be a mixed TATA and CAAT sequence between 262 and 272 bases upstream from the transcriptional start. Its distant location indicates that it may not be a CAAT box; however, shortly repeated GA and CT, which resembles Z-DNA structure, on its flanking areas suggests a potential involvement in expressional regulation. A hexanucleotide sequence, T-G-T-T-C-T, which has been postulated to be a binding site for the glucocorticoid receptor complex in mouse mammary tumor virus (20), is found 431 bp upstream from the transcriptional start. The hexanucleotide sequence was also found previously 451 bp upstream from the transcriptional start of the human renin gene; hence, the renin gene was speculated to be under steroid hormone regulation (11).

Intron regions. The potential glucocorticoid receptor-binding sequence of T-G-T-T-C-T is also present three times, within introns located at positions 1,511, 3,281, and 3,553, respectively (Fig. 3). A survey of intron regions shows unusual patterns of nucleotide sequences, including the existence of an alternating purine-pyrimidine or Z-DNA-like structure, followed by 14 Gs at location 3,383 (Fig. 3). The neighborhoods of the glucocorticoid-binding hexanucleotide consist of partially symmetrical $G(T)_{2-5}$ repeats, GTTTTT (TGTTCT)GTTTTGTTTGTTGTTTTGTTTTT, located at 3,547 bp (Fig. 3). This is followed two times by a sequence of ACCCTGCA 7 and 24 bp downstream. Intron 4 has only 130 bp and contains sequences of $T(C)_5A(C)_3ACA$ and CCCATT

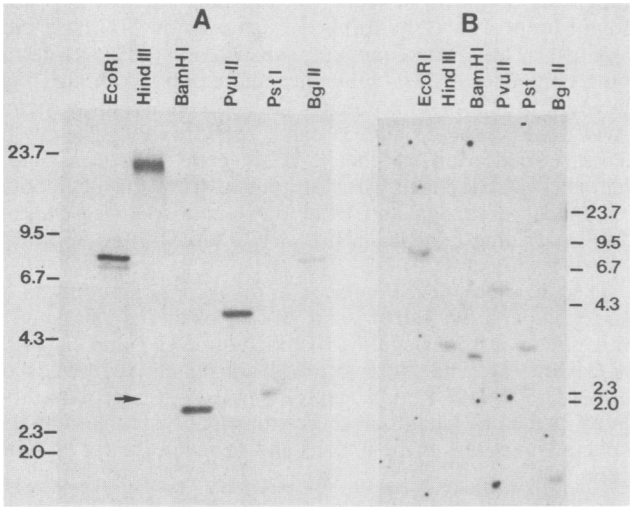

FIG. 2. Southern hybridization of total genomic DNA and recombinant phage DNA. (A) Southern blot of the recombinant phage DNA. DNA (1 μg) was digested with the appropriate enzyme, electrophoresed on a 0.75% agarose gel, transferred to nitrocellulose, and hybridized to a nick-translated probe derived from cDNA of the rat α_1-AGP gene. The arrow indicates the location of a positively hybridized *Hin*dIII fragment after extended exposure. (B) Southern blot of genomic DNA. Rat hepatoma cell DNA (5 μg) was digested with appropriate enzymes, electrophoresed on a 1% agarose gel, and hybridized under the same conditions as described for panel A.

in the middle, with their symmetrical counterparts at the acceptor side of the splicing junction (Fig. 3).

3′-End region. The polyadenylation signal (AATAAA) is found 18 nucleotides upstream from the beginning of the poly(A) tail-formation site. A decanucleotide sequence of $TCCT^{C}_{T}CT^{CCC}_{TTT}$ is repeated four times, overlapping one another within a 30-bp region located 15 bp downstream from the poly(A) addition signal and immediately followed by a complementary GGAGAGGAGG. These particular sequences create five *Mnl*I restriction sites within 40 bp of length and produce a highly symmetrical region centering at the second *Mnl*I. Its significance is unknown, although it may play a role in posttranscriptional processing or have a regulatory function in gene expression.

Our findings concerning the nucleotide sequence of this gene and its structure point to several interesting features, including potential steroid receptor-binding sites (Table 1). We have found a sequence consistent with a glucocorticoid receptor-binding site upstream of the α_1-AGP mRNA initiation site and three others within the structural gene. An identical hexanucleotide sequence has also been found in the human renin gene (19). The hexanucleotide T-G-T-T-C-T in the human renin gene is located almost identically with that in the rat α_1-AGP gene at 450 and 430 bp upstream from the mRNA initiation site, respectively. Because this area of DNA has been postulated to be the site of glucocorticoid receptor complex binding, the two genes may be under similar glucocorticoid control.

The second T-G-T-T-C-T sequence of the rat α_1-AGP gene is located within intron 2. Sequences 3 and 4 are in intron 5. Although the functional significance of additional binding sites is unknown, such sites have been found in the human growth hormone genes (12). Furthermore, Charnay et al. (3) and Wright et al. (28) have demonstrated that gene-specific controlling elements of human α- and β-globin genes can be located within the structural gene instead of in the promoter region.

It should be noted that hexanucleotide 4 is surrounded by a $G(T)_{2-5}$ element of $G(T)_5[TGTTCT]G(T)_4G(T)_3G(T)_2G(T)_4G(T)_5$. A Z-DNA-like structure of seven continuous dinucleotide G-Ts followed by $(G)_{10}A(G)_3$ at position 3,382 (Fig. 3) is also located in this intron 139 bp upstream from the fourth glucocorticoid-binding sequence. $(GT)_n$ oligomers have the potential for forming a Z-DNA structure under particular conditions, including ionic strength and superhelical density, and are widely distributed throughout eucaryotic genomes (7, 13). It has been suggested that long $(GT)_n$ oligomers ($n > 20$), which do not form a typical nucleosomal structure, may act as chromatin-structural elements only and have no regulatory function. However, smaller Z-DNA sequences ($n < 10$), such as those found in the rat α_1-AGP gene, have been found in a number of enhancer and transcriptional control sequences from DNA and RNA viruses (14).

Since the rat α_1-AGP gene does contain potential glucocorticoid receptor-binding sites, gene expression may not be entirely because of secondary action of the hormone as previously suggested (1, 26). We propose that the rat α_1-AGP gene may be directly regulated by the glucocorticoid receptor complex, but that a permissive labile factor supplied by continuous protein synthesis in particular cell types must be present to sustain high levels of stable mature α_1-AGP mRNA. However, the sequences we mentioned may not represent bona fide glucocorticoid receptor-binding sites. This possibility is supported by two facts. (i) The homology of these putative binding sites does not extend

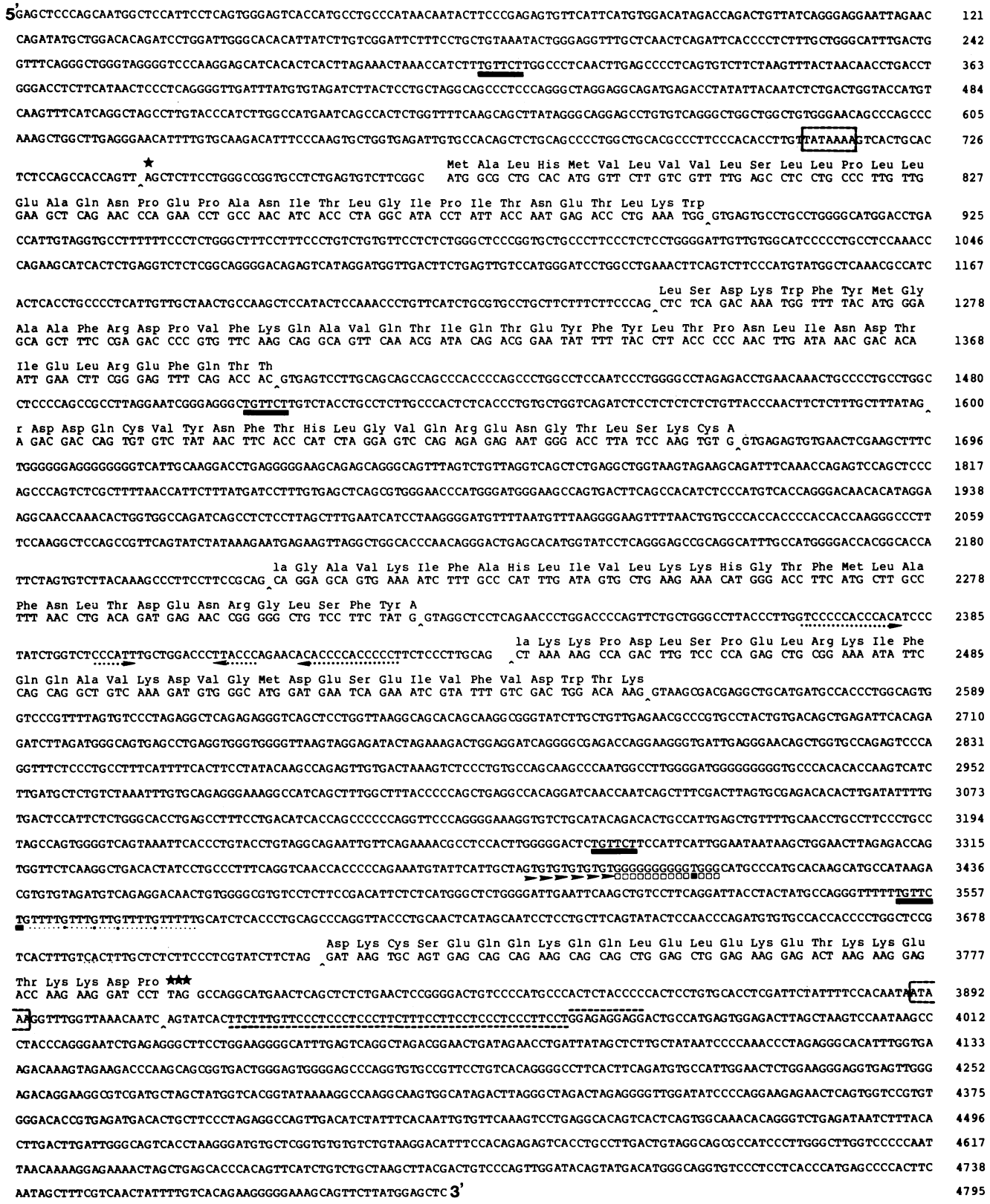

FIG. 3. DNA sequence of the rat α_1-AGP gene. Exons are flanked with lower signs, and translational codons are printed as triplets and lined with their designated amino acids. The TATA box and polyadenylation signal (AATAAA) are boxed. A single asterisk indicates the beginning of transcription, and triple asterisks indicate the codon (TAG) of translational stop. The hexanucleotides T-G-T-T-C-T of potential glucocorticoid receptor-binding sequences are underlined with a solid line. Palindromic sequences in intron 4 are marked by arrows and dots; dinucleotide G-T repeats followed by poly(G) are marked by arrowheads and open squares. $G(T)_{2-5}$ repeats after the fourth hexanucleotide T-G-T-T-C-T are underlined with open circles and dots. The *Mnl*I-rich area is underlined with a broken line and its following complementary sequence is marked with a broken line at the top.

TABLE 1. Comparison of nucleotide sequences involved in glucocorticoid receptor-binding activities

Gene	Nucleotide sequence[a]
Mouse mammary tumor virus	−187 5′GGTTACAAACTGTTCTTAAAACAAGG3′ −129 5′TGGTATCAAATGTTCTGATCTGAGCT3′ −108 5′GAGCTCTTAGTGTTCTATTTTCCTAT3′ −92 5′TTGATTCCTCTGTTCTTTTGGAATCT3′
Chicken lysozyme[b]	−50 3′TTGATTCCTCTGTTCTACAGTTGTCA5′
Rat α-1 AGP	−449 5′AAACCATCTTTGTTCTTGGCCCTCAA3′ +844 5′TCGGGAGGGCTGTTCTTGTCTACCTG3′ +4004 5′TGGGGGACTCTGTTCTTCCATTCATT3′ +4286 5′CAGGGTTTTTTGTTCTGTTTTGTTTG3′
Human growth hormone	+97 5′GGGCACATGTGTCCTGAGGGGAGAG3′
Human metallothionein	−246 5′CGGTACACTGTGTCCTCCCGCTGCAC3′
Human renin	−202 5′TTCCCACAGCTGTCCTTCCTCCAGCG3′ −314 5′CTGACAGTGCTGTTTCTCATCAGCCT3′

[a] Underlining indicates the consensus hexanucleotide sequences, and asterisks indicate the unmatched nucleotides. Negative numbers indicate the nucleotide position in relationship before translational start (position 0), and positive numbers indicate nucleotide position after transcriptional start.

[b] Chicken lysozyme genes are arranged from the 3′ toward the 5′ end.

further into regions of homology among human growth hormone, human metallothionein II, and mouse mammary tumor virus genes (12; Table 1); (ii) The transcription rate of the rat α_1-AGP gene is only marginally increased by the addition of glucocorticoid as measured by in vitro transcription in isolated nuclei (27). Glucocorticoid receptor-binding assays should resolve this issue.

We are particularly grateful to David Geller and Barbara Poetter for help in preparing and editing this manuscript. We also thank Thomas Sargent for sharing his rat genomic library.

This work was supported by Public Health Service grant CA-21141 from the National Institutes of Health and a grant from the Council for Tobacco Research.

ADDENDUM

During the preparation of this manuscript, the structure of this gene was described by R. Reinke and P. Feigelson (J. Biol. Chem. **260**:4397–4403, 1985). Our work confirms a majority of their sequences and extends this structure.

LITERATURE CITED

1. **Baumann, H., G. P. Jahreis, and K. C. Gaines.** 1983. Synthesis and regulation of acute plasma proteins in primary culture of mouse hepatocytes. J. Cell Biol. **97**:866–876.
2. **Breathnach, R., and P. Chambon.** 1981. Organization and expression of eukaryotic split genes coding for proteins. Annu. Rev. Biochem. **50**:349–384.
3. **Charnay, P., R. Treisman, P. Mellon, M. Chao, R. Axel, and T. Maniatis.** 1984. Differences in human α- and β-globin gene expression in mouse erythroleukemia cell: the role of intragenic sequences. Cell **38**:251–263.
4. **Chiu, K. M., R. F. Mortenson, A. P. Osmand, and H. Gewurz.** 1977. Interactions of α_1-acid glycroprotein with the immune system. Immunology **32**:997–1005.
5. **Franzblau, C., K. Schmid, B. Farris, J. Beldekas, P. Garvin, H. M. Kagan, and B. J. Baum.** 1976. The interaction of collagen with α_1-acid glycoprotein. Biochim. Biophys. Acta **427**:302–314.
6. **Friedman, J. M.** 1983. Control of malaria virulence by α_1-acid glycoprotein (orosomucoid), an acute-phase (inflammatory) reactant. Proc. Natl. Acad. Sci. USA **80**:5421–5424.
7. **Hamada, H., M. G. Petrino, and T. Kakunga.** 1982. A novel repeated element with Z-DNA-forming potential is widely found in evolutionary divers eukaryotic genomes. Proc. Natl. Acad. Sci. USA **79**:6465–6469.
8. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual, p. 109–112. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
9. **Martinez, H. M., B. Katzung, and T. Farrah.** 1984. Sequence analysis program. Biomathematics Computation Laboratory, Department of Biochemistry and Biophysics, University of California, San Francisco.
10. **Miesfeld, R., S. Okret, A.-C. Wikstrom, P. Weange, J.-A. Gustafsson, and K. R. Yamamoto.** 1984. Characterization of a steroid hormone receptor gene and mRNA in wild-type and mutant cells. Nature (London) **312**:779–781.
11. **Miyazaki, H., A. Fukamizu, S. Hirose, T. Hayashi, H. Hori, H. Ohkubo, S. Nakanishi, and K. Murakami.** 1984. Structure of the human renin gene. Proc. Natl. Acad. Sci. USA **81**:5999–6003.
12. **Moore, D. D., A. R. Marks, D. I. Buckley, G. Kapler, F. Payvar, and H. Goodman.** 1985. The first intron of the human growth hormone gene contains a binding site for glucorticoid receptor. Proc. Natl. Acad. Sci. USA **82**:699–702.
13. **Nordheim, A., and A. Rich.** 1983. The sequence (dC-dA)n · (dG-dT)n forms left-handed Z-DNA in negatively supercoiled plasmids. Proc. Natl. Acad. Sci. USA **80**:1821–1825.
14. **Nordheim, A., and A. Rich.** 1983. Negatively supercoiled simian virus 40 DNA contains Z-DNA segments within transcriptional enhancer sequences. Nature (London) **303**:674–679.
15. **Ricca, G. A., R. W. Hamilton, J. W. McLean, A. Conn, J. E. Kalinyak, and J. M. Taylor.** 1981. Rat α_1-acid glycoprotein messenger RNA. J. Biol. Chem. **256**:10362–10368.
16. **Ricca, G. A., and J. M. Taylor.** 1981. Nucleotide sequence of rat α_1-acid glycoprotein messenger RNA. J. Biol. Chem. **256**:11199–11202.
17. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
18. **Sargent, T. D., J. Wu, J. M. Sela-Trepat, R. B. Wallace, A. A. Reyes, and J. Bonner.** 1979. The rat serum albumin gene: analysis of cloned sequences. Proc. Natl. Acad. Sci. USA **76**:3256–3260.
19. **Scheidereit, C., and M. Baeto.** 1984. Contact between hormone receptor and DNA double helix within a glucocorticoid regulatory element of mouse mammary tumor virus. Proc. Natl. Acad. Sci. USA **81**:3029–3033.
20. **Scheidereit, C., S. Geisse, H. M. Westphal, and M. Beato.** 1983. The glucocorticoid receptor binds to defined nucleotide sequences near the promotor of mouse mammary tumor virus. Nature (London) **304**:749–752.
21. **Schmid, K.** 1975. α_1-Acid glycoprotein, p. 184–228. *In* F. Putnam (ed.). The plasma proteins: structure, function, and genetic control, vol. 1, Academic Press, Inc., New York.
22. **Schmid, K., H. Kaufman, S. Isemura, F. Bauer, J. Emura, T. Motoyama, M. Ishiguro, and S. Nanno.** 1973. Structure of α_1-acid glycoprotein. The complete amino acid sequence, multiple amino acid substitutions, and homology with the immuno-

globulins. Biochemistry **12**:2711–2724.

23. **Schreier, P. H., and R. Cortese.** 1979. A fast and simple method for sequencing DNA cloned in the single-stranded bacteriophage M13. J. Mol. Biol. **129**:169–172.
24. **Snyder, S., and E. L. Coodley.** 1976. Inhibition of platelet aggregation by α_1-acid glycoprotein. Arch. Intern. Med. **136**: 778–781.
25. **Southern, E. M.** 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. **98**:503–517.
26. **Vannice, J. L., G. M. Ringold, J. W. McLean, and J. M. Taylor.** 1983. Induction of the acute-phase reactant, α_1-acid glycoprotein, by glucocorticoids in rat hepatoma cells. DNA **2**:205–212.
27. **Vannice, J. L., J. M. Taylor, and G. M. Ringold.** 1984. Glucocorticoid-mediated induction of α_1-acid glycoprotein: evidence for hormone-regulated RNA processing. Proc. Natl. Acad. Sci. USA **81**:4241–4245.
28. **Wright, S., A. Rosenthal, R. Flavel, and F. Grosveld.** 1984. DNA sequences required for regulated expression of β-globin genes in murine erythroleukemia cells. Cell **38**:265–273.