Isolation and characterization of the rat tyrosine aminotransferase gene

(exon–intron structure/glucocorticoid control/CAMP control/regulatory DNA sequence)

TAKEHISA SHINOMIYA*, GERD SCHERER†, WOLFGANG SCHMID*, HANSWALTER ZENTGRAF†, AND GÜNTER SCHÜTZ*

*Institute of Cell and Tumor Biology, and †Institute of Virus Research, German Cancer Research Center, Immuenheimer Feld 280, D-6900 Heidelberg, Federal Republic of Germany

Communicated by M. Lindauer, November 14, 1983

ABSTRACT Tyrosine aminotransferase (TAT; L-tyrosine:2-oxoglutarate aminotransferase, EC 2.6.1.5) from rat liver is subject to glucocorticoid, cAMP, and developmental control. To study the underlying regulatory mechanisms, the TAT structural gene was isolated from a λ bacteriophage rat DNA library. Heteroduplex analysis revealed that the 2.4-kilobase-long TAT mRNA is encoded by a gene that extends over 11 kilobases and is interrupted by 11 introns. To characterize the presumptive control region, the DNA sequence around the 5′ end of the gene was determined and the start site of transcription was identified by nuclease S1 protection experiments. A short sequence homology in an equivalent position relative to the cap site was detected between TAT and tryptophan oxygenase, another glucocorticoid-controlled gene from rat liver. This sequence is related to the sequence 5′-5'-G-T-T-T-C-T-3′ found in regions of the long terminal repeat of mouse mammary tumor virus, which has been shown to interact with the glucocorticoid receptor (Scheideret, C., Geisse, J., Westphal, H. M. & Beato, M. (1983) Nature (London) 304, 749–752).

A classical example for steroid hormone control is the induction of the enzymatic activities of both tyrosine aminotransferase (TAT; L-tyrosine:2-oxoglutarate aminotransferase, EC 2.6.1.5) and tryptophan oxygenase [TO; tryptophan 2,3-dioxygenase, L-tryptophan:oxygen 2,3-oxidoreductase (de-cyclizing), EC 1.13.11.11] in rat liver. The increase in activity of these two glucogenic enzymes after glucocorticoid stimulation results from an increased rate of enzyme synthesis (1–3), which derives from an increase in the amount of translatable mRNA (4–6). Apart from glucocorticoid induction, the activity of TAT is also inducible by cAMP (7, 8). Since both hormonal controls for TAT are also operative in a number of established rat hepatoma cell lines, the hormonal induction of TAT in hepatoma cells has become an attractive model system to study regulated gene expression in mammalian cells (9).

In addition to hormonal modulation, both TAT and TO activity are also under developmental control. While TAT enzyme activity appears around birth (10), TO enzyme activity can be detected around postnatal day 15 (11).

By genetic and biochemical analysis of several albino lethal mutants of the mouse, a control region required for expression and inducibility of mouse TAT has been assigned to the region of the albino locus on chromosome 7 (12). In these mutants, the activity of several liver enzymes is affected. The basal level of TAT is severely decreased and the enzyme activity is no longer inducible by glucocorticoids, although the structural gene for TAT is still present (13).

To study the complex control mechanisms operating on these two liver-specific genes at the molecular level, we have isolated the TAT and TO genes from the rat genome. We have previously described the isolation and characterization of cDNA as well as genomic clones for TO (14). With these molecular clones as probes, some of us have shown that glucocorticoid induction of TO occurs at the transcriptional level (15). Similar analysis has shown that glucocorticoid and cAMP induction of TAT also result from transcriptional activation of the gene (unpublished data).

In this report, we describe the isolation of the TAT gene from a rat phage library, using a recently isolated cDNA clone as probe (16). The structural features of the rat TAT gene are presented. To study the mechanism of transcriptional activation and to enable the design of hybrid genes, the DNA sequence at the 5′ end of the TAT gene was determined. Comparison of this sequence with the DNA sequence in the 5′ region of TO and with sequences in the long terminal repeat of mouse mammary tumor virus DNA known to interact with the glucocorticoid receptor (17) revealed short stretches of sequence homology.

MATERIALS AND METHODS

Construction and Screening of the Rat Genomic Library. High molecular weight DNA (100 µg) from livers of two adult male Wistar rats was partially digested with Sau3A and fractionated by electrophoresis on a 0.5% low-melting-temperature agarose gel. DNA fragments 15–19 kilobases (kb) long were recovered as described (18) and ligated to BamHI-cut DNA of the λ phage vector EMBL3. This cloning vector, a derivative of the vector λ1059 (18), allows recovery of the cloned DNA as Sal I fragments (19). The ligated DNA was packaged in vitro yielding a total of 11 × 10⁶ recombinant phages. To establish a library, 3 × 10⁶ recombinants were amplified, and 1 × 10⁶ original recombinants were screened directly (20) by using nick-translation of pUC-TAT-3 DNA (16) as probe. Five independent clones were isolated, one of which, λTAT1, is described in this report.

Subcloning and DNA Sequence Analysis. The 2.7-kb Sal I fragment of λTAT1 was subcloned into the polylinker region of plasmid pUC8 (21) yielding plasmid pUTATS2.7. Deletions removing segments of the 5′ flanking region were generated essentially as described by Frischau et al. (22). In brief, the plasmid was linearized at random by partial digestion with DNase I and ligated to BamHI linkers. After digestion with BamHI, which cuts once in pUC8, the DNA preparation was fractionated on an agarose gel and DNA fragments of the desired lengths (5.4–3.5 kb) were circularized with T4 DNA ligase and cloned. Appropriate clones, identi-
fied by restriction analysis, were sequenced (23). After cleavage at the single BamHI site, the DNA was labeled at the 5' or 3' end and recut with EcoRI. Both strands were always sequenced.

Heteroduplex Analysis. Heteroduplex analysis was carried out as described (14), except that enriched TAT mRNA was obtained by hybridizing total rat liver RNA to pcTAT-3 DNA immobilized to macroporous Sephacryl S-500 (24).

Nuclease S1 Mapping. Nuclease S1 analysis was carried out essentially as described (14), except that RNA-DNA hybrids were formed at 48°C. The nuclease S1 probe was obtained as follows: the subclone pUTATS2.7 was digested with Mst II, treated with calf intestine alkaline phosphatase, and 5'-end-labeled with [γ-32P]ATP using T4 polynucleotide kinase. After recutting with EcoRI, the 1.7-kb EcoRI/Mst II fragment, uniquely labeled at the Mst II 5' end at position +84, was isolated.

RESULTS

Isolation and Characterization of a λ Recombinant Phage Containing the Complete TAT Structural Gene. Radiolabeled pcTAT-3 DNA, a cDNA clone containing approximately one-third of the TAT mRNA sequences (16) (Fig. 1C) was used to screen a phage library of rat liver DNA generated by partial digestion with Sau3A. From 1 × 10^6 recombinant phages screened, five independent phages were isolated and shown by restriction endonuclease analysis and Southern blotting (not shown) to span a contiguous region of 27 kb. The restriction map of one genomic clone, λrTAT1, which contains the entire TAT structural gene, is shown in Fig. 1A. This clone carries a 17.5-kb insert with 1.5 kb of 5'-flanking DNA, 11 kb of gene sequences, and 5 kb of 3'-flanking DNA (Fig. 1A).

The structure of the TAT gene was determined by analysis of heteroduplexes formed between λrTAT1 DNA and rat liver poly(A)^+ RNA enriched for TAT mRNA. A representative electron micrograph with corresponding interpretation is shown in Fig. 2.

Fig. 2. Electron microscopy of a TAT heteroduplex. (A) Hybrid formed between TAT mRNA and the DNA of λrTAT1. (Bar = 0.1 μm.) (B) Interpretive drawing with RNA sequences represented by a dashed line and DNA sequences by a solid line. The lengths of exons A–L and introns 1–11 are given in Fig. 1B.

The resulting structural map of the TAT gene relative to the restriction map is shown in Fig. 1. The direction of transcription was initially established by determining which strand of the cDNA insert in clone pcTAT-3 is protected from nuclease S1 digestion by TAT mRNA (data not shown).
and later confirmed by the nuclease S1 mapping experiments described below. As shown in Fig. 1, the TAT gene spans 11 kb and has a total of 11 introns. The heteroduplex measurements give a total length for the 12 exons of 2300 base pairs, in good agreement with our earlier size estimate for the TAT mRNA of about 2400 bases (16).

Restriction fragments carrying repetitive sequences, identified by hybridization to 32P-labeled rat liver DNA, can be found both within and flanking the TAT gene (Fig. 1A).

**Determination of the Sequence of the 5'-End Region and of the Start Site of Transcription of the TAT Gene.** To measure the distance of the first exon from the left arm of the EMBL3 vector, heteroduplex experiments were carried out as described but including single strands of an unrelated EMBL3 clone. This clone can form hybrid molecules with clone pUTAT1 via the common vector arms, while TAT mRNA can

---

**Fig. 3.** Construction of deletion mutants for sequence analysis. Plasmid pUTATS2.7 contains the 2.7-kb Sal I fragment of λrTAT cloned in pUC8. Deletions were generated (22) by inserting BamHI linkers at random and deleting the DNA segments between the BamHI site in the polylinker region of pUC8 and the inserted linker, as indicated by arrows. The map is not drawn to scale. The filled box symbolizes exon A as determined by heteroduplex measurements, and the arrow shows the start site of transcription. Five deletion clones were selected to establish the sequence shown in Fig. 4. The following regions were analyzed: Δ2/8 (−600 to −418), Δ5/3 (−436 to −173), Δ5/4 (−220 to −7), Δ7/1 (−11 to +151), and Δ4/6 (+97 to +300).

**Fig. 4.** Nucleotide sequence of the 5'-end region of the rat TAT gene. DNA sequences around exon A of the TAT gene as determined by deletion sequence analysis (see Fig. 3). Position +1 indicates the presumed cap site (see Fig. 5). The 5' end of exon A has not been mapped, but is definitely beyond the Mst II site. "TATA"- and CCAAT-like sequences are boxed. The sequence homologous to a sequence found at position −101 to −110 in the rat T0 gene (14) is indicated by asterisks. Sequences similar to the hexanucleotide sequence T-G-T-T-C-T found in regions of the long terminal repeat of MMTV interacting with the glucocorticoid receptor (17) are overlined.

**Fig. 5.** Localization of the TAT mRNA start site by nuclease S1 mapping. Lanes G, A, T, and C: sequencing ladder starting from the Mst II site at position +4. Note that the labeling of the sequencing lanes has been transposed so the mRNA sequence can be read directly. Lane 1, nuclease S1-resistant products obtained by digesting a hybrid between total poly(A)+ RNA isolated from livers of dexamethasone-induced animals and the Mst II probe fragment with 200 units of nuclease S1 per ml. Lane 2, same as lane 1, but using 1600 units of nuclease S1 per ml. Lane 3, same as lane 1, but with total poly(A)+ RNA isolated from dibutyryl cAMP-injected animals. Lane 4, same as lane 1, but with RNA isolated from adrenalectomized rats. Lane 5, same as lane 1, except that yeast tRNA was incubated with the probe. Two micrograms of RNA was used in all cases. The putative transcription start site is indicated by an arrow at the sequence of interest.
still hybridize to the insert region in λTAT1 DNA. This analysis allowed us (data not shown) to map the beginning of the first exon at about 1.5 kb from the left vector arm, or from the Sal I site at the arm-insert junction (Fig. 1A).

In addition, run-off experiments using an in vitro transcription system revealed a start site for an RNA polymerase II transcript 580 ± 10 base pairs upstream of the leftmost EcoRI site, again 1.5 kb from the Sal I site at the left arm-insert junction (data not shown).

Therefore, the 2.7-kb Sal I fragment from the left end of the cloned TAT mRNA is needed only for transcription initiation (Fig. 1A) was subcloned into pUC8, and the region of interest was analyzed (23) using the deletion sequencing strategy of Frischau et al. (22). The construction of the deletion mutants is schematically shown in Fig. 3, with the resulting nucleotide sequence presented in Fig. 4. In-spection of the DNA sequence reveals the sequence T-A-T-T-T-T-C-T-A-A-A-A-AAA consensus sequence characteristic for RNA polymerase II promoter sites (23), located between nucleotides −31 and −25 relative to the +1 start site defined below. In addition, between nucleotides −77 and −60 is the "CCAAAT"-box sequence found in a number of eukaryotic genes at a corresponding distance from the cap site (26).

The start site of transcription of the TAT gene was determined by nuclease S1 mapping (27). A 1.7-kb EcoRI/Mst II fragment, 5'-end labeled in the coding strand of the TAT gene at the Mst II site at position +84 (Fig. 4), was used as a probe. Hybridization to total poly(A)+ RNA from livers of dexamethasone-stimulated rats yields two protected DNA fragments two nucleotides apart (Fig. 5, lane 1). The same two protected fragments are observed with RNA isolated from livers of dibutyryl cAMP- injected or adrenorecepto- mized animals, respectively (Fig. 5, lanes 3 and 4). In this experiment, the RNA samples isolated from the induced animals show a 5- to 10-fold increase in TAT mRNA, in confirmation of our earlier RNA blotting data (16). Note that uninduced and induced RNA samples originate from the same start site.

Because the upper of the two protected fragments becomes weaker with the addition of increasing higher amounts of nuclease S1 (compare lanes 1 and 2 of Fig. 5), and because most eukaryotic mRNAs initiate with an adenine residue preceded by a cytosine residue (25), we define the adenine residue labeled +1 in Figs. 4 and 5 as the cap site of the TAT mRNA, taking into account that nuclease S1-generated fragments are retarded by 1.5 nucleotides relative to chemically cleaved DNA on a sequencing gel (28).

**DISCUSSION**

We have isolated and characterized the structural gene encoding rat TAT. This gene extends over 11 kb and is interrupted by 11 introns. It is thus larger by a factor of 4.6 compared to the 2400-base-long mRNA that it encodes. The largest exon of the TAT gene is a 1.0-kb segment located at the 3' end. It very likely encodes mainly untranslated sequences, because only about 1400 nucleotides of the 2400-base-long TAT mRNA are needed to encode the 52,000 Dalton TAT protein.

Because the restriction maps of the five independently isolated TAT clones from our rat liver genomic library overlap to form a unique map, we conclude that the rat TAT gene is a single copy gene. This conclusion is corroborated by restriction analysis of TAT clones isolated from other phage libraries during the initial phase of this work. Two different clones were isolated from the rat liver DNA library of Sargent et al. (29), and five different clones were isolated from a rat HTC cell DNA library obtained by K. Yamamoto. While none of these clones contained the 5'-end region of the TAT gene, their restriction maps and the heteroduplex measurements made on two HTC clones agree with the data for the clones from our own library. The only difference was a length heterogeneity, observed in intron 3, in which clone λTAT1 has 200 base pairs less sequence information than the corresponding HTC clone. This difference is probably due to the gain or loss of a repetitive sequence element (unpublished data). Furthermore, genomic blots of rat liver DNA probed with various nonrepetitive restriction fragments derived from λTAT1 did not reveal evidence for a second TAT gene (data not shown).

Unambiguous proof that clone λTAT1 contains the authentic rat TAT gene comes from the observation that stable transformants of mouse 3T3 cells, obtained by cotransfer of λTAT1 DNA together with the chicken thymidine kinase gene, do express TAT enzyme activity (unpublished data).

To allow sequence comparisons with other glucocorticoid-controlled genes and the design of chimeric genes for functional studies, a region of 800 bases around the 5' end of the TAT gene has been analyzed, and the start site of transcription has been determined by nuclease S1 mapping. It is not clear whether the two closely spaced nuclease-resistant fragments observed in this experiment (Fig. 5) result from incorrect nuclease S1 digestion or if they reflect a true start-site microheterogeneity as observed in a number of other well-characterized systems (30, 31).

The 5' end of the first exon has not yet been determined, but must lie beyond the Mst II site used for nuclease S1 mapping. There are four potential G-T splice-donor sequences within the next 50 bases following the Mst II site, with the DNA sequence around the G-T at position +104 having the best homology to the splice-donor consensus sequence (25, 32). If this site is actually used, the first exon will just terminate with an ATG codon. Note also that there is a dinucleotide repeat within the first intron, with 25 consecutive T-C residues starting at position +210.

We have used computer analysis to search for sequence homologies in the 5'-flanking regions of the TAT and the TO gene (14), because both genes are under glucocorticoid control in rat liver. We have previously noted a sequence at position −101 to −110 in the TO gene, which is similar to a sequence present at position −102 to −109 in the long terminal repeat of mouse mammary tumor virus (MMTV) (14). In the TAT sequence shown in Fig. 4, we find a sequence identity with a fit of 8/10 nucleotides to this TO sequence, located at position −83 to −92, again at a similar position relative to the RNA start site. By nuclease protection experiments, two regions in the long terminal repeat of MMTV have recently been identified to interact with the glucocorticoid receptor protein, and the hexanucleotide sequence T-G-T-T-C-T has been found to be present four times within these regions (17; K. Yamamoto, personal communication). Interestingly, the short TAT-TO homology described above is closely related to this T-G-T-T-C-T sequence, with the first thymine replaced by an adenine residue. We note, however, that sequences deviating by one nucleotide from the T-G-T-T-C-T sequence motif of MMTV do occur at a total of six positions in the 5'-flanking region of the TAT gene, indicated by overlining in Fig. 3. There is no perfectly matching hexanucleotide T-G-T-T-C-T in our TAT sequence. We do not find any region in the TAT gene having more than 60% homology to the 24-nucleotide-long sequence reported by Coch et al. (33), which is shared by several glucocorticoid regulated genes.

Besides being controlled by glucocorticoids, the expression of the TAT gene is also regulated by cAMP (7, 8). We have recently shown that the increase in TAT mRNA activity in rat liver after cAMP stimulation results from a rapid increase in the transcription rate of the gene (unpublished data). We therefore compared the 5'-flanking sequences of the TAT gene with cAMP digestion data if they reflect a true start-site microheterogeneity (34), which is expressed in the pituitary and also rapidly responds to cAMP by increasing its nuclear tran-
scription rate (35, 36). While we observe a number of short sequence homologies, it is at the moment impossible to assess their biological significance. Gene transfer studies with chimeric genes containing the TAT promoter region may allow us to probe the importance of these and of the other sequences discussed above in the various controls operating on the TAT gene.

We thank T. Sargent and K. Yamamoto for their phage libraries, V. Pirrotta for gifts of packaging extracts, C. M. Strange for help in the initial phase of this work, S. Hashimoto for doing the run-off experiments, and S. Suhai for help with computer assisted sequence analysis. We are grateful to M. Jantzen, R. Miksicek, and R. Renkawitz for helpful comments on the manuscript and J. Purkert and M. Cole for typing. This work was supported by a grant from the Humboldt-Stiftung to G. Schütz and grants from the Deutsche Forschungsgemeinschaft (Schu 51/4-1 and Schu 51/5-3) and the Fonds der Chemischen Industrie to G. Schütz.