Vol. 170, No. 1, 1990 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
July 16, 1990 Pages 176-181

DEDUCED PRIMARY STRUCTURE OF RAT TRYPTOPHAN-2,3-DIOXYGENASE

Katsumi Maezono, Kosuke Tashiro and Toshikazu Nakamura*

Department of Biology, Faculty of Science
Kyushu University, Fukuoka 812, Japan

Received May 23, 1990

SUMMARY: The complete amino acid sequence of the tryptophan 2, 3-dioxygenase (TO) of rat liver was determined from the nucleotide sequence of a full length TO cDNA isolated from a rat liver cDNA library and determined its primary structure. TO was encoded in a mRNA of about 1.7 kb containing an open reading frame of 1218 bp. According to the deduced amino acid sequence, the monomeric polypeptide of TO consisted of 406 amino acid residues with a calculated molecular weight of 47,796 daltons. It has twelve histidine residues around its hydrophobic region, which has homology with some heme proteins and oxygenase, suggesting that this hydrophobic region might to be the core of TO for the activity.

Tryptophan-2,3-dioxygenase (EC 1.13.1.12, TO) is a liver specific enzyme that plays a key role in the L-tryptophan metabolic pathway, in which TO catalyzes the reaction of L-tryptophan with molecular oxygen to yield N-formyl-L-kynurenine[1]. The ferrous enzyme changes into an oxygenated reaction intermediate that binds oxygen and L-tryptophan[2]. TO is a tetramer composed of four identical 40 kD polypeptides[3]. Because TO contains two protoheme IV per molecule, the rat liver enzyme is represented as an $\alpha_2\beta_2$ structure. However, there is little information about the structure of TO because it is unstable and difficult to crystalize. Studies by electron paramagnetic resonance suggested that the ligand of the heme was a nitrogen atom from a histidine residue[4].

TO is induced by substrate L-tryptophan or glucocorticoids[5][6]. Using a primary culture of rat hepatocytes, we showed that glucagon,

*To whom reprint requests should be addressed.

Abbreviations: TO, tryptophan 2,3-dioxygenase; pcTO, partial TO cDNA clone.

0006-291X/90 $1.50

dibutyryl cAMP, insulin and epinephrine also act as regulators of TO level in hepatocytes[7)8)] and regulation of TO level by various hormones operates through a change in its translatable mRNA activity[9)]. Recently, we isolated a partial cDNA clone of TO mRNA by the hybrid-translation method from enriched polysomal mRNA. Using this cDNA, we demonstrated that the changes in TO mRNA activity were due to changes in the amount of TO mRNA through its transcription; and moreover, this transcriptional regulation was apparantly mediated by a short-lived protein[10)]. To study the molecular mechanism TO regulation, the gene must be isolated and the *cis*- and *trans*-acting elements must be identified.

In this report, we cloned the full length TO cDNA and determined the primary structure of TO. From the amino acid sequence and its hydropathy plot, we predicted the sequence of the core region of TO.

MATERIALS AND METHODS

Materials —— Enzymes for DNA manipulations were obtained from Takara Shuzo Co. and Nippon Gene Co. The multiprime DNA labelling system was from Amersham. Sequenase™ was from Toyobo Co. Radioisotopes were obtained from Amersham.

Screening of rat liver cDNA library —— The partial TO cDNA clone, pcTO, was isolated as described previously[10)]. Essentially, TO mRNA was extracted from enriched polysomes by immunoadsorption and used for construction of the cDNA library. pcTO was obtained from this cDNA library by the hybrid-translation method. A 570-bp insert DNA of pcTO was labelled with [α-^{32}P]dCTP by the multiprime labelling system and used as a probe for screening. Its specific activity was about 5.0 X 10^8 dpm/μg. The λgt10 rat liver cDNA library was constructed with oligo(dT) as a primer. *In situ* plaque hybridization was performed according to standard methods[11)]. Positive plaques were isolated, and phage DNAs were prepared. Inserted DNAs were recloned into the EcoRI site of plasmid pUC119 or pUC118.

Sequence analysis —— To determine the nucleotide sequence, a positive clone was deleted to various sizes using exonuclease III and mung bean nuclease. These plasmid subclones were sequenced by the dideoxy nucleotide method[12)] using Sequenase™.

RESULTS AND DISCUSSION

To clone the cDNA containing a whole sequence of TO mRNA, we screened the λgt10 cDNA library of rat liver with the insert of pcTO, which was isolated by the hybrid-translation method[10)]. By the screening of 2 X 10^5 clones of the rat liver cDNA library, we obtained eight positive ones. One of the positive clones, pTO14B,

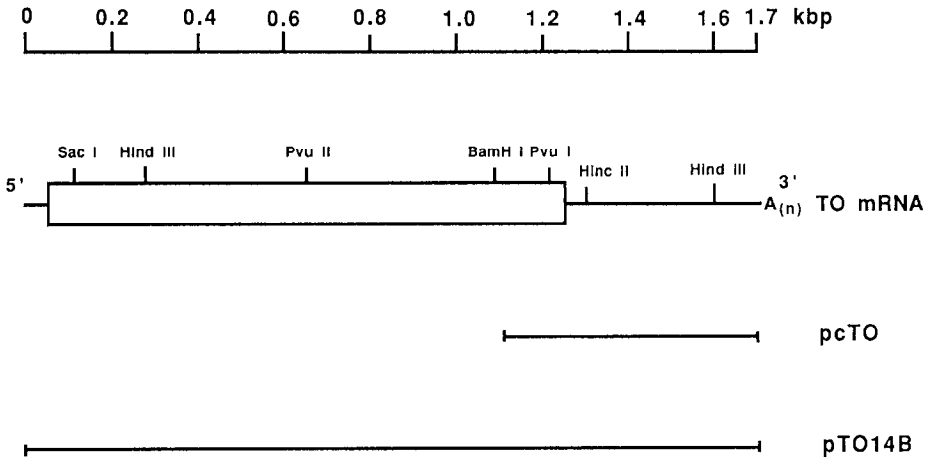

Fig. 1. Schematic representation of rat TO mRNA and cDNA.
The rat TO mRNA is diagrammed below the scale. Restriction sites are shown on the structure of TO mRNA. Non-coding regions are represented by a line and the coding region is boxed. cDNA clones are represented below the mRNA.

has a 1690 bp insert that is almost identical to the length of TO mRNA detected by Northern blot analysis[10]. The insert DNA of pTO14B was subcloned into plasmid pUC119 and sequenced by the dideoxy method using a double strand plasmid as a template. Figure 1 shows the structure of TO mRNA and the restriction cleavage map of TO cDNA. pTO14B contains the open reading frame of 1221 bp and most, if not all, the 5'-untranslated region and the whole 3'-untranslated region. Polyadenylation signals exist at just upstream of the poly(A) tail. Figure 2 shows the nucleotide sequence of TO cDNA and the deduced amino acid sequence. TO is encoded in the open reading frame of 406 amino acid residues. Alternatively, the open reading frame may be initiated at either nucleotide 58, 322 or 343. However, because the CTACC sequence from nucleotide -5 to -1 matches well with Kozak's consensus sequence[13] for translation of the initiation site, we predict that the ATG at nucleotides 1 to 3 is the initiation codon of TO. The calculated molecular weight of the TO monomeric polypeptide is 47,796 daltons.

TO is known to be a heme protein and contains two molecules of protoheme IV. Moreover, its heme binding ligand has been shown to be histidine residues[4]. There are twelve histidine residues in the deduced amino acid sequence. Marnett *et al.* determined the heme binding residue of cyclooxygenase and showed that the His-Tyr-Pro-Arg sequence of this enzyme was the reaction center[14]. However, we could not find this sequence in TO. Figure 3 shows a hydropathy plot of the deduced amino acid sequence of TO. We found that the

```
                                                                TCCTAGCAAACCTGTGTGCTCCTGGGACGCATCACTACC    -1

Met Ser Gly Cys Pro Phe Ser Gly Asn Ser Val Gly Tyr Thr Leu Lys Asn Leu Ser Met    20
ATG AGT GGG TGC CCA TTT TCA GGA AAC AGT GTA GGA TAT ACT TTG AAA AAC TTA TCT ATG    60

Glu Asp Asn Glu Glu Asp Gly Ala Gln Thr Gly Val Asn Arg Ala Ser Lys Gly Gly Leu    40
GAA GAC AAT GAA GAA GAC GGA GCT CAA ACT GGT GTA AAC AGA GCC AGC AAA GGA GGA CTT   120

Ile Tyr Gly Asp Tyr Leu Gln Leu Glu Lys Ile Leu Asn Ala Gln Glu Leu Gln Ser Glu    60
ATC TAT GGG GAC TAC TTG CAG TTG GAG AAG ATT TTG AAT GCA CAA GAA CTT CAA AGT GAA   180

Ile Lys Gly Asn Lys Ile His Asp Glu His Leu Phe Ile Ile Thr His Gln Ala Tyr Glu    80
ATC AAA GGG AAT AAA ATC CAC GAC GAG CAC CTC TTT ATT ATA ACT CAC CAA GCT TAT GAA   240

Leu Trp Phe Lys Gln Ile Leu Trp Glu Leu Asp Ser Val Arg Glu Ile Phe Gln Asn Gly   100
CTT TGG TTT AAA CAA ATT CTC TGG GAA CTT GAT TCT GTT CGT GAG ATT TTT CAA AAT GGC   300

His Val Arg Asp Glu Arg Asn Met Leu Lys Val Met Thr Arg Met His Arg Val Val Val   120
CAT GTC AGG GAT GAG AGG AAC ATG CTC AAG GTG ATG ACT CGG ATG CAC CGT GTG GTG GTC   360

Ile Phe Lys Leu Leu Val Gln Gln Phe Ser Val Leu Glu Thr Met Thr Ala Leu Asp Phe   140
ATC TTC AAG CTC CTG GTA CAG CAG TTC TCG GTT CTG GAA ACA ATG ACT GCC TTG GAC TTC   420

Asn Asp Phe Arg Glu Tyr Leu Ser Pro Ala Ser Gly Phe Gln Ser Leu Gln Phe Arg Leu   160
AAT GAC TTC AGA GAG TAC CTG TCT CCA GCA TCA GGC TTC CAG AGT CTT CAG TTC CGG CTG   480

Leu Glu Asn Lys Ile Gly Val Leu Gln Ser Leu Arg Val Pro Tyr Asn Arg Lys His Tyr   180
CTA GAA AAT AAG ATA GGT GTT CTT CAG AGC TTG AGA GTC CCT TAC AAC AGG AAA CAC TAT   540

Arg Asp Asn Phe Glu Gly Asp Tyr Asn Glu Leu Leu Leu Lys Ser Glu Gln Glu Gln Thr   200
CGT GAT AAC TTT GAA GGA GAC TAC AAT GAG CTG CTG CTG AAA TCG GAG CAG GAG CAG ACG   600

Leu Leu Gln Leu Val Glu Ala Trp Leu Glu Arg Thr Pro Glu Leu Glu Pro His Gly Phe   220
CTA TTG CAG CTG GTG GAG GCA TGG CTG GAA CGC ACA CCT GGC TTA GAG CCA CAT GGA TTC   660

Asn Phe Trp Gly Lys Phe Glu Lys Asn Ile Leu Lys Gly Leu Glu Glu Glu Phe Leu Lys   240
AAT TTC TGG GGA AAG TTT GAA AAA AAT ATC TTG AAG GGT CTG GAA GAG GAG TTC CTA AAG   720

Ile Gln Ala Lys Lys Asp Ser Glu Glu Lys Glu Glu Gln Met Ala Glu Phe Arg Lys Gln   260
ATT CAG GCG AAA AAG GAC TCT GAA GAA AAA GAG GAA CAG ATG GCA GAG TTC CGG AAG CAG   780

Lys Glu Val Leu Leu Cys Leu Phe Asp Glu Lys Arg His Asp Tyr Leu Leu Ser Lys Gly   280
AAA GAG GTG CTG CTC TGC TTG TTC GAT GAG AAG CGT CAT GAC TAC CTT CTG AGT AAA GGT   840

Glu Arg Arg Leu Ser Tyr Arg Ala Leu Gln Gly Ala Leu Met Ile Tyr Phe Tyr Arg Glu   300
GAA CGA CGA CTG TCA TAC CGT GCA CTC CAG GGA GCA CTG ATG ATA TAT TTT TAC AGG GAG   900

Glu Pro Arg Phe Gln Val Pro Phe Gln Leu Leu Thr Ser Leu Met Asp Ile Asp Thr Leu   320
GAG CCT CGA TTC CAG GTC CCT TTC CAG TTG CTG ACC TCA CTT ATG GAC ATT GAC ACA CTC   960

Met Thr Lys Trp Arg Tyr Asn His Val Cys Met Val His Arg Met Leu Gly Ser Lys Ala   340
ATG ACC AAA TGG AGA TAT AAT CAT GTG TGC ATG GTG CAC AGG ATG CTA GGC AGC AAG GCT  1020

Gly Thr Gly Gly Ser Ser Gly Tyr Tyr Tyr Leu Arg Ser Thr Val Ser Asp Arg Tyr Lys   360
GGC ACT GGG GGA TCC TCA GGC TAT TAT TAT CTG CGC TCA ACT GTG AGC GAC AGG TAC AAG  1080

Val Phe Val Asp Leu Phe Asn Leu Ser Ser Tyr Leu Val Pro Arg His Trp Ile Pro Lys   380
GTG TTC GTG GAT TTA TTT AAC CTC TCA TCG TAC CTG GTT CCC CGA CAC TGG ATA CCA AAG  1140

Met Asn Pro Ile Ile His Lys Phe Leu Tyr Thr Ala Glu Tyr Ser Asp Ser Ser Tyr Phe   400
ATG AAT CCG ATC ATT CAC AAG TTC CTT TAC ACA GCT GAG TAC AGC GAC AGC TCC TAC TTC  1200

Ser Ser Asp Glu Ser Asp ***                                                       406
AGC AGC GAT GAA TCA GAT TGA GTTCTTCTGAACATCAGTCCAGGCTACAGGATTCCCAGTCAACTTTTATTT

TATAAATTTTTACAAATATGTGATTGGTGTAACATATTTATATTTGTAGTTCAGAGACGTGATGTTGTGGTCCAATCCT
GGAAAAAATTATGATTTCGCATATCATGATGATGTATGATTAAGCAGATTAAGCATTATGATAAAAATAACTTGGTAAA
ATGTTAGCATCATCATACATATGATGTATTCTGGTTATAACTCAATTTACCCTGACACTTACCTCCATAGAAACACTTT
AAGTAATTAGTTCCTTATTGCTTCATACTTTATAAAGCTTGGTCACCAGTTCTTTTATACTATAGATGCAATAAATACT
ATTCTTCTGTACAAAATTTATTCAAATGAATCTTTAATTAATAAATTTAGTTTTTGTCTGCGAAAAAAAAAAAAAAAA  1705
```

Fig. 2. Nucleotide sequence of rat TO cDNA and the deduced amino acid sequence.

Nucleotide and amino acid numbers are at the right. Polyadenylation signals are double underlined. Histidine residues are boxed.

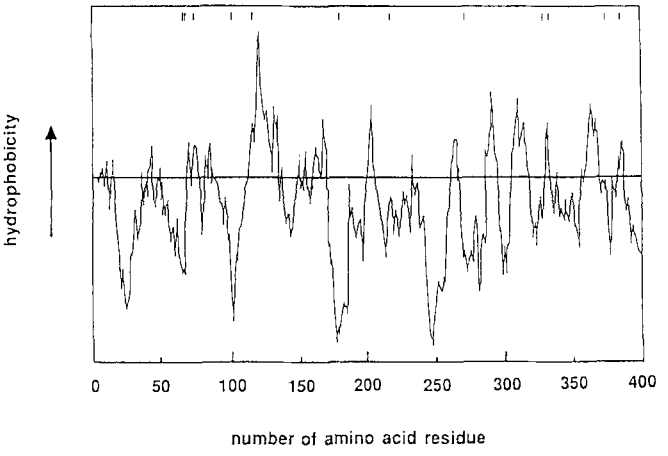

Fig. 3. Hydropathy plot of deduced amino acid sequence of TO.
The hydropathy plot was constructed using the algorithm of Kyte and Doolittle[24] and a window size of five residues. Values above the baseline are hydrophobic, while those below are hydrophilic. The bar above the hydropathy profile indicates the position of the histidine residues.

amino acid sequence from residues 110-170 was markedly hydrophobic. Although TO has no significant sequence homology to any other protein, the hydrophobic region contents a sequence that is partially homologous to the hydrophobic region of other oxygenase[15] and heme proteins such as a family of cytochrome P450[16] (data not shown). Moreover, histidine residues exist around this region. Therefore, this region might correspond to the core center of this enzyme and might be essential for its activity.

Genomic cloning experiments indicated that TO is encoded in a single gene (data not shown). The analysis of *cis*-acting elements on the TO gene promoter for induction of TO expression by hormones has shown that two glucocorticoid responsive elements are present upstream of the TO gene[17][18]. In addition to the regulation by hormones, TO expression is also regulated in the liver during the development of rats[19]-[21]. It first appears in the liver at 2 weeks after birth so that TO expression is a suitable marker for terminal differentiation of the liver. Previously, using TO cDNA, we found that the increased expression of TO during the postnatal development of rats results from an increase in the amount of TO mRNA through activation of TO gene transcription[22][23]. However, the molecular mechanism for the activation of the dormant TO gene is still unknown. Identification of *cis*- and *trans*-acting elements controlling TO gene activation during differentiation of immature hepatocytes should be helpful for solving this problem.

ACKNOWLEDGMENT

This work was supported by a Research Grant from the Ministry of Education, Science and Culture of Japan.

REFERENCES

1. Hayaishi, O., Rothberg, S., Mehler, A. H. and Saito, Y. (1957) J. Biol. Chem. 229, 889-896
2. Ishimura, Y., Nozaki, M., Hayaishi, O., Nakamura, T., Tamura, M. and Yamazaki, I. (1970) J. Biol. Chem. 245, 3595-3602
3. Schutz. G. and Feigelson, P (1972) J. Biol. Chem. 242, 5327-5337
4. Henry. Y., Ishimura, Y., and Peisach, J. (1976) J. Biol. Chem. 251, 1578-1581
5. Knox, W. E. and Mehler, A. H. (1950) J. Biol. Chem. 187, 419-430
6. Feigelson, P. and Greengard, O. (1962) J. Biol. Chem. 237, 3714-3717
7. Nakamura, T., Shinno, H., and Ichihara, A. (1980) J. Biol. Chem. 255, 7533-7535
8. Noda, C., Nakamura, T. and Ichihara, A. (1983) J. Biol. Chem. 258, 1520-1525
9. Niimi, S., Nakamura, T., Nawa, K., and Ichihara, A. (1983) J. Biochem. 94, 1697-1706
10. Nakamura, T., Niimi, S., Nawa, K., Noda, C., Ichihara, A., Takagi, Y., Anai, M. and Sakaki, Y. (1987) J. Biol. Chem. 262, 727-733
11. Maniatis, T., Fritsch, E. F. and Sambrook, J (1982) Molecular Cloning: A Laboratory Mannual, Cold Spring Harbor, N.Y.
12. Sanger, F., Coulson, A. R., Barrell, B. G., Smith, A. J. Hand Roe, B. A. (1980) J. Mol. Biol. 143, 161-178
13. Kozak, M. (1984) Nucleic Acids Res. 12, 8857-8872
14. Marnett, L. J. Chen, Y. N. P. Maddipati, K. R. Ple, P. and Labeque, R. (1988) J. Biol. Chem., 263, 16532-16535
15. Lederer, F., Ghrir, R., Guiard, B., Cortial, S. and Ito, A. (1983) Eur. J. Biochem., 132, 95-102
16. Phillips, I. R., Shephard, E. A., Ashworth, A. and Rabin, B. R. (1985) Proc. Natl. Acad. Sci. USA 82, 983-987
17. Danesch, U., Hashimoto, S., Renkawitz, R and Shutz, G. (1983) J. Biol. Chem. Chem. 258, 4750-4753
18. Schule, R., Muller, M., Otuka-Murakami, H., and Renkawi, R (1988) Nature 332, 87-90
19. Greengard, O. and Dewey, H. K. (1971) Proc. Natl. Acad. Sci. USA. 68, 1698-1701
20. Roper, M. D. and Franz, J. M. (1977) J. Biol. Chem. 252, 4354-4360
21. Killewich, L. A. and Feigelson, P. (1977) Proc. Natl. Acad. Sci. USA. 74, 5392-5396
22. Nagao, M., Nakamura, T., and Ichihara, A., (1986) Biochim. Biophys. Acta 867, 179-186
23. Nakamura, T., Nagao, O. and Ichihara, A., (1987) Exp. Cell. Res. 169, 1-14
24. Kyte, J. and Doolittle, R. F. (1982) J. Mol. Biol. 157, 105-132