Cloning, Sequence Analysis, and Expression in Escherichia coli of the Gene Encoding the Candida utilis Urate Oxidase (Uricase)

Yasuji Koyama,1 Toshio Ichikawa, and Eiichi Nakano
Research and Development Division, Kikkoman Corporation, Noda 389, Noda, Chiba 278

Received for publication, June 17, 1996

A urate oxidase (uricase) gene was cloned from Candida utilis with an oligonucleotide probe based on the amino acid sequence of cyanogen bromide-cleaved uricase. The uricase gene contains 909 base pairs and encodes a protein with a predicted mass of 34,193 Da. Candida uricase was similar (49% match in amino acid sequence) to the uricase from Aspergillus flavus. The uricase from Candida utilis has four cysteines and one of them, Cys168, participates in the enzyme activity. This enzyme was expressed to a level of about 20% of total cellular protein in an Escherichia coli cell as a soluble and functional form.

Key words: Candida utilis, nucleotide sequence analysis, overexpression in E. coli, oxidation of cysteine residue, uricase.

Urate oxidase [EC 1.7.3.3] (uricase), is an enzyme in the purine degradation pathway and catalyzes the oxidation of uric acid to allantoin. Humans lack uricase and the accumulation of uric acid in blood causes gout symptoms (1, 2). Uricase, therefore, is used for the enzymatic determination of uric acid to diagnose gout. Uricase has been found in various sources and the corresponding genes cloned from some mammals (3-5), Drosophila (6), soybean (7), Aspergillus (8, 9), and Bacillus (10), but the gene encoding uricase of Candida utilis, which is widely available commercially, has not been cloned. The Candida uricase does not require any cofactor for the enzymatic oxidation. So it is important to improve the productivity and to investigate the structure-function relationship of this enzyme. In this report we describe the cloning of the uricase gene from C. utilis and a role of cysteine residues.

MATERIALS AND METHODS

Strains—C. utilis IF00988 (11) was used as enzyme source and DNA donor. Escherichia coli LE392 and P2392 (12), obtained from Funakosi (Tokyo), were used as host strains for a genetic library constructed in the EMBL4 bacteriophage vector (13). E. coli JM109 (14) obtained from Takara Shuzo (Kyoto) was used as a host for DNA sequencing and gene expression.

Materials—Restriction enzymes and DNA-modifying enzymes were obtained from Takara Shuzo and Boehringer Mannheim (Tokyo). Gigapack II Gold for in vitro packaging was purchased from Funakosi. [γ-32P]ATP was from Amersham (Tokyo). All other reagents were of analytical grade.

Purification and Amino Acid Sequencing of Uricase from C. utilis IF00988—C. utilis was grown at 30°C for 48 h in a medium containing 5% glucose, 1% polypeptone, 0.5% yeast extract, 0.85% urea, 0.3% (NH4)2SO4, 0.2% K2HPO4, and 0.05% MgSO4, and was harvested by centrifugation at 5,000 × g for 10 min. For induction of uricase, the cells were suspended in a medium containing 5% glucose, 0.06% uric acid, 0.2% K2HPO4, and 0.05% MgSO4, and incubated at 30°C for 4 h. After induction, the cells were harvested by centrifugation, resuspended in 0.2 M sodium carbonate buffer (pH 10) and extracted with chloroform at 30°C for 20 h. The enzyme was absorbed to DEAE-cellulose in a batch system and eluted with 0.15 M NaCl. The eluate was purified to homogeneity by HPLC on a hydroxyapatite column and a reverse-phase column CAPCELL PAK C18 (Shiseido, Tokyo). The enzyme was digested with cyanogen bromide (15) and the resulting peptides were separated by reverse-phase HPLC and sequenced with a model 470A protein sequencer (Perkin Elmer Japan, Urayasu).

Oligonucleotides—Oligonucleotides were synthesized on a DNA model 382 synthesizer (Perkin Elmer Japan). Oligonucleotide U1 (CCNCARAAYCCNAARAA, where N is A, C, G, or T, R is A or G, and Y is C or T), which was used to screen the DNA library, was based on the peptide sequence PQNPKK. Oligonucleotides UN (ATGTCAACACGGTCTCATCATCCACCTAC) and UC (TTACAACTTGGTCTTCTCCTTACGGACAAC), based on the sequences of amino and carboxyl terminals, respectively, were used as primers for amplification of the uricase-coding region by PCR.

DNA Manipulation—Isolation of genomic DNA and mRNA from C. utilis and hybridization were carried out as previously described (16). Genomic DNA was partially digested with Sau3AI. Fragments of 9 to 20 kb were purified by agarose gel electrophoresis and ligated to BamHI-digested EMBL4 vector. About 20,000 plaques were screened by plaque hybridization using the U1 probe labeled with [γ-32P]ATP.

A 4.5-kb EcoRI fragment of a positive plaque was subcloned into pUC118 (17) to construct pUO501 and sequenced by the dideoxy chain termination method using Kilo-Sequence Deletion Kit (Takara Shuzo) and Dye Primer Cycle Sequencing Kit (Perkin Elmer Japan).
Computer analysis was done using DNASTAR-Mac (Hitachi Software Engineering, Yokohama).

Expression of the Uricase Gene in E. coli—For expression of PCR-amplified fragments in E. coli, the expression vector pUTE200 was used (18). The 912-bp DNA fragment encoding uricase was amplified by PCR using pU0501 as a template. This fragment was inserted into the HpaI site of pUTE200 to obtain pUX201.

Site-Directed Mutagenesis—Substitution of Cys to Ser was carried out by the method of Kunkel et al. (19) with a Muta-Gene™ in vitro mutagenesis kit (Bio-Rad Japan, Tokyo) using a single-strand DNA prepared from pU0501-harboring JM109. The synthetic oligonucleotides used are listed in Table I. After the mutant structures had been confirmed by DNA sequence determination, DNA fragments containing mutated sites were cut off with suitable restriction enzymes and replaced into pUX201.

Uricase Assay and Protein Analysis—E. coli carrying pUX201 was cultured in L broth containing 1 mM IPTG and 100 µg/ml ampicillin at 37°C for 16 h and disrupted by ultrasonication to obtain a crude extract. A reaction mixture containing 2 ml of 0.001% uric acid in 0.1 M borate buffer (pH 8.5), 0.7 ml of water, and 0.1 ml of crude extract was incubated at 25°C for 5 min and the reaction was stopped by addition of 0.2 ml of 20% KOH. The activity of uricase was calculated from the absorbance decrease of uric acid at 293 nm. One unit was defined as the amount of enzyme necessary to transform 1 µmol of uric acid into allantoin in 1 min at 25°C and pH 8.5. To examine thermal stability, crude extract was incubated at 60°C for 15 min. After centrifugation, residual activity of supernatant was measured.

Nucleotide Sequence Accession Number—The nucleotide sequence data reported in this paper will appear in the GSDB, DDBJ, EMBL, and NCBI nucleotide sequence databases with the accession number D32043.

RESULTS

Cloning of the Uricase Gene—We concluded that the amino terminus of this protein was blocked since the amino-terminal sequence of the uricase purified from C. utilis IFO 0988 could not be determined. After treatment with cyanogen bromide, four peptide fragments could be collected and sequenced. Probe U1 designed from one of the peptide sequences, PQNPKK, was synthesized and used for plaque hybridization studies. About 20,000 plaques were screened with probe U1 and three positive clones were obtained. An EcoRI fragment of about 4.5 kb which hybridized to the probe was subcloned into pUC118 and the resultant plasmid was designated pUX0501.

Sequence of the Uricase Gene—We sequenced 1,641 bp

| TABLE I. Oligonucleotides for site-directed mutagenesis. Asterisks indicate Ser anticodon. Each primer was designed to generate a new restriction site to facilitate screening of mutant. |
|---|---|
| Position | Mutagenic oligonucleotide |
| 39 | 5'-AACCCACCTCTCAAGCACTATTGAGCTCGCTCATCAA TTTTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTATTAT

Fig. 1. DNA sequence of the uricase gene and its flanking region. The deduced amino acid sequence is shown under the DNA sequence. The sequences of four peptides obtained by cyanogen bromide cleavage and confirmed by peptide sequencing are underlined. Probe U1 was synthesized on the basis of the peptide sequence PQNPKK, which is double-underlined. The putative TATA box is boxed. The putative copper-binding site (H-X-H) is indicated by asterisks.
We have cloned and sequenced the uricase gene from a genomic DNA of *C. utilis*, which encodes a 34-kDa protein with 303 aa residues. A putative TATA box (TATAAAA) was found at position –99 (Fig. 3) but no other consensus regulatory sequences were identified.

The genomic gene amplified by PCR could be expressed in *E. coli* under the control of a lac promoter, suggesting that this gene contains no introns. The recombinant uricase accumulated to a level of approximately 20% of soluble intracellular protein and had the same specific activity as uricase purified directly from *C. utilis*.

Compared with uricases from other sources, the amino acid sequence of *C. utilis* uricase possesses 49% identity to fungi, 33% to mammals, 37% to fly, 33% to soybean, and 21% to *Bacillus*. Bairoch (21) and Legoux et al. (8) identified consensus patterns for eukaryotic uricase (motif 1 and motif 2 in Fig. 3). The sequences Val-Leu-Lys-Ser-Thr-Gly-Ser (154-160 in Fig. 1) and Ser-Pro-Ser-Val-Gln-Ala-Thr-Met-Phe (231-239 in Fig. 1) are slightly different from motif 1 (Val-Leu-Lys-Thr-Thr-Gln-Ser) and motif 2 (Ser-Pro-Ser-Val-Gln-Lys/His/Asn-Thr-Leu-Tyr), respectively. Recently, Yamamoto et al. (10) reported that the corresponding sequence of *Bacillus* sp. TB-90 was not highly homologous and that another highly conserved region was identified (region A in Fig. 3). This sequence, Asn-Ser-X-Val/Ile-Val/Ile-Ala/Pro-Thr-Asp-Ser/Thr-X-Lys-Asn, is found in the sequence of uricase from *C. utilis*.

**DISCUSSION**

We have cloned and sequenced the uricase gene from a genomic DNA of *C. utilis*, which encodes a 34-kDa protein with 303 aa residues. A putative TATA box (TATAAAA) was found at position –99 (Fig. 3) but no other consensus regulatory sequences were identified.

The genomic gene amplified by PCR could be expressed in *E. coli* under the control of a lac promoter, suggesting that this gene contains no introns. The recombinant uricase accumulated to a level of approximately 20% of soluble intracellular protein and had the same specific activity as uricase purified directly from *C. utilis*.

Compared with uricases from other sources, the amino acid sequence of *C. utilis* uricase possesses 49% identity to fungi, 33% to mammals, 37% to fly, 33% to soybean, and 21% to *Bacillus*. Bairoch (21) and Legoux et al. (8) identified consensus patterns for eukaryotic uricase (motif 1 and motif 2 in Fig. 3). The sequences Val-Leu-Lys-Ser-Thr-Gly-Ser (154-160 in Fig. 1) and Ser-Pro-Ser-Val-Gln-Ala-Thr-Met-Phe (231-239 in Fig. 1) are slightly different from motif 1 (Val-Leu-Lys-Thr-Thr-Gln-Ser) and motif 2 (Ser-Pro-Ser-Val-Gln-Lys/His/Asn-Thr-Leu-Tyr), respectively. Recently, Yamamoto et al. (10) reported that the corresponding sequence of *Bacillus* sp. TB-90 was not highly homologous and that another highly conserved region was identified (region A in Fig. 3). This sequence, Asn-Ser-X-Val/Ile-Val/Ile-Ala/Pro-Thr-Asp-Ser/Thr-X-Lys-Asn, is found in the sequence of uricase from *C. utilis*.

**DISCUSSION**

We have cloned and sequenced the uricase gene from a genomic DNA of *C. utilis*, which encodes a 34-kDa protein with 303 aa residues. A putative TATA box (TATAAAA) was found at position –99 (Fig. 3) but no other consensus regulatory sequences were identified.

The genomic gene amplified by PCR could be expressed in *E. coli* under the control of a lac promoter, suggesting that this gene contains no introns. The recombinant uricase accumulated to a level of approximately 20% of soluble intracellular protein and had the same specific activity as uricase purified directly from *C. utilis*.

Compared with uricases from other sources, the amino acid sequence of *C. utilis* uricase possesses 49% identity to fungi, 33% to mammals, 37% to fly, 33% to soybean, and 21% to *Bacillus*. Bairoch (21) and Legoux et al. (8) identified consensus patterns for eukaryotic uricase (motif 1 and motif 2 in Fig. 3). The sequences Val-Leu-Lys-Ser-Thr-Gly-Ser (154-160 in Fig. 1) and Ser-Pro-Ser-Val-Gln-Ala-Thr-Met-Phe (231-239 in Fig. 1) are slightly different from motif 1 (Val-Leu-Lys-Thr-Thr-Gln-Ser) and motif 2 (Ser-Pro-Ser-Val-Gln-Lys/His/Asn-Thr-Leu-Tyr), respectively. Recently, Yamamoto et al. (10) reported that the corresponding sequence of *Bacillus* sp. TB-90 was not highly homologous and that another highly conserved region was identified (region A in Fig. 3). This sequence, Asn-Ser-X-Val/Ile-Val/Ile-Ala/Pro-Thr-Asp-Ser/Thr-X-Lys-Asn, is found in the sequence of uricase from *C. utilis*.
Fig. 3. A multiple alignment of uricases. The uricases from *C. utilis*, *A. flavus* (A38097), soybean, rat, and *Bacillus* (D49974) are aligned. The conserved residues among the five proteins are boxed. Motif 1 and motif 2 were proposed as consensus patterns for uricase by Bairoch (21) and Legoux (8), respectively. Regions A and B were newly identified conserved sequences.

(54-65 in Fig. 1). In addition, there is another conserved sequence, Tyr/His-Gly-Lys in the amino terminal region of all uricases (region B in Fig. 3). A mutant with this region deleted had no uricase activity (data not shown).

Thus, this sequence is important for the enzyme activity. Eukaryotic uricases are known to be copper-binding enzymes (22) and the copper-binding site has been proposed (2). But uricases of *C. utilis* and *A. flavus* contain no copper (20, 23), despite having a putative copper-binding site (His-Asp-His, 119-121 in Fig. I and Ref. 8). Yamamoto et al. reported that the copper-binding site was not found in the uricase of *Bacillus* sp. TB-90, which does not contain copper (10). It is interesting to investigate how microbial uricases compensate functionally for the absence of copper in the active center.

Uricase of *C. utilis* is well characterized (20, 24, 25). Nishimura et al. (20) reported that it contains three cysteines, of which one is located near the active site of the enzyme but is not necessary for uricase activity. It is known that subunits of *Candida* uricase can associate to form an enzyme with lower specific activity under aerobic conditions (25). Based on our sequence data, the uricase of *C. utilis* IF00988 contains four cysteines. We showed that the oxidation or substitution of Cys168 resulted in reduction of the uricase activity. One hypothesis is that a chemically reactive sulfhydryl group located on the surface of the molecule is spontaneously oxidized during purification and forms a disulfide bond linking the two molecules.

We thank Ms. K. Saito for technical assistance.

REFERENCES


Cloning of the Uricase Gene from Candida utilis


