Complete Nucleotide Sequence of the Metapyrocatechase Gene on the TOL Plasmid of Pseudomonas putida mt-2

(Received for publication, August 3, 1982)

Chieko Nakai, Hiroyuki Kagamiyama, and Mitsuhiro Nozaki
From the Department of Biochemistry, Shiga University of Medical Science, Seta, Ohtsu, Shiga 520-21, Japan

Teruko Nakazawa, Sachiyu Inouye, Yousuke Ebina, and Atsushi Nakazawa
From the Department of Biochemistry, Yamaguchi University School of Medicine, Ube, Yamaguchi 755, Japan

Metapyrocatechase which catalyzes the oxygenative ring cleavage of catechol to form α-hydroxyymuconic c semialdehyde is encoded by the xylE gene on the TOL plasmid of Pseudomonas putida mt-2. We have cloned the xylE region in Escherichia coli and determined the nucleotide sequence of the DNA fragment of 985 base pairs around the gene. The fragment included only one open translational frame of sufficient length to accommodate the enzyme. The predicted amino acid sequence consisted of 307 residues, and its NH2- and COOH-terminal sequences were in perfect agreement with those of the enzyme recently determined (Nakai, C., Hori, K., Kagamiyama, H., Nakazawa, T., and Nozaki, M. (1983) J. Biol. Chem. 258, 2916-2922). A mutant plasmid was isolated which did not direct the synthesis of the active enzyme. This plasmid had a DNA region corresponding to the NH2-terminal two-thirds of the polypeptide. From the deduced amino acid sequence, the secondary structure was predicted.

Around 10 base pairs upstream from the initiator codon for metapyrocatechase, there was a base sequence which was complementary to the 3'-end of 16 S rRNAs from both E. coli and Pseudomonas aeruginosa. A preferential usage of C- and G-terminated codons was found in the coding region xylE, which contributed to the relatively high G + C content (57%) of this gene.

Metapyrocatechase (catechol: oxygen 2,3-oxidoreductase, (decatlyzing), catechol 2,3-dioxygenase, EC 1.13.11.2) from Pseudomonas putida (arvilla) mt-2 is a ferrous iron containing dioxygenase obtained in a crystalline form (1, 2). It has been established recently that the enzyme is encoded by the xylE gene on the TOL plasmid (3). The location and orientation of the gene have also been determined (4).

Education of the reaction mechanism requires detailed knowledge of the active site structure, which can be obtained only after determining the primary structure. However, no information on this aspect with the enzyme is available. In the preceding paper (5), we reported that the enzyme consists of four identical subunits; each has M₉ = 35,000. Further, the partial amino acid sequences of the enzyme, the NH2-terminal sequence (53 residues), and the COOH-terminal sequence (6 residues) were determined (5). Since these results of the sequence analyses have permitted assignment of the coding region of the enzyme in the plasmid, we have performed the determination of the nucleotide sequence of the xylE gene to deduce the amino acid sequence of the enzyme.

In this paper, we describe the complete nucleotide sequence of the structural gene for metapyrocatechase and the predicted primary structure of the enzyme. The secondary structure predicted from the primary structure is also discussed.

EXPERIMENTAL PROCEDURES

Materials

Bacterial Strains and Plasmids—The bacterial strains used were Escherichia coli 2080 (thi lac mal mtl ara xyl rpsL) and E. coli GM31 (thr leu dcm his ara thi lac gel xyl mtl rpsL tonA supE) (6). The original strain carrying the TOL plasmid was P. putida (arvilla) mt-2 (ATCC 23873) which was used in the preceding paper (5). A recombinant plasmid of TOL and RP4, pTN2, constructed in vitro (7) was used as a DNA donor to make a pACYC184-TOL hybrid, pTS89, as described (8). The plasmid vectors used were pBR322 (9), pACYC177 (10), and pACYC184 (11). Plasmid DNAs were purified as described previously (7).

Chemicals—A restriction endonuclease, Alul, was purified according to the method of Greene et al. (11). Other restriction endonucleases were obtained from Bethesda Research Laboratories, Boehringer Mannheim, and Takara Shuzo (Kyoto, Japan). [γ-32P]ATP was enzymatically synthesized with carrier-free 32P (New England Nuclear) according to the procedure of Walseth and Johnson (12). E. coli alkaline phosphatase, T4 polynucleotide kinase, and T4 DNA polymerase were purchased from Worthington Biochemicals, P-L Biochemicals, and Bethesda Research Laboratories, respectively.

Methods

Restriction Enzyme Analysis and Molecular Cloning—Analysis of restriction enzyme fragments by gel electrophoresis and purification of digested DNA fragments by an electrophoretic elution procedure were performed as described previously (7). Ligation of DNA fragments and transformation experiments were carried out as described previously (4). Metapyrocatechase-producing transformants were detected as yellow colonies after spraying a solution of 0.1 M catechol onto the colonies.

Preparation of Cell Extracts and Enzyme Assays—Cells were grown at 37 °C with shaking overnight in LB broth (1% tryptone, 0.5% yeast extract, 0.5% NaCl) containing ampicillin (100 μg/ml). Preparation of crude extracts of and assay conditions for metapyrocatechase were as described previously (13).

DNA Sequencing—pTS815 DNA was prepared from a DNA cytosine methylation-negative (dcm) host, GM31, harboring the plasmid, and digested with several restriction endonucleases. Preparation of DNA fragments for sequencing was carried out as described by Maxam and Gilbert (14) with slight modifications. Fragments were separated in 4, 5, 7.5, or 10% polyacrylamide gels followed by extraction overnight at 37 °C with 10 mM Tris-HCl buffer (pH 8.0) containing 0.5 mM EDTA, 0.1% sodium dodecyl sulfate, and 500 mM ammonium.
Complete DNA Sequence of Metapyrocatechase Gene

2924

acetic acid (15). Chemical sequencing procedures were those of Maxam and Gilbert (14). 5`-Termini of each fragment were dephosphorylated with alkaline phosphatase and endo-labeled with [y-32P]ATP with T4 polynucleotide kinase. The fragments were subjected to secondary restriction cleavage or to strand separation with dimethyl sulfoxide. For efficient labeling of restriction fragments containing blunt ends or 3'-protruding ends, the termini were first converted to 5'-protruding ends by 3'-exonuclease with T4 polymerase (16).

Preparation of Cyanogen Bromide Peptides and Amino Acid Analyses—Cyanogen bromide cleavage of metapyrocatechase, fractionation of the resulting peptides, and amino acid analyses of these peptides were performed as described in the preceding paper (5). For the amino acid analyses, samples were hydrolyzed with 0.5 ml of constant boiling HCl in evacuated sealed tubes at 105°C.

RESULTS

Molecular Cloning of the Metapyrocatechase Gene (xylE)—Metapyrocatechase is encoded by the xylE gene on the TOL plasmid. From restriction mapping of insertion and deletion mutants of the TOL plasmid and cloning of TOL fragments in E. coli, the xylE gene was located in a 1.8-kilobase BamHI-XhoI fragment. The direction of transcription was deduced from a read-through transcription from the known promoter on the vector (4, 7). To isolate a smaller DNA fragment containing xylE, pTS89, a pACYC184-TOL hybrid plasmid (8) was used as a starting material. Since the 2.2-kilobase XhoI fragment in the TOL segment was considered to contain xylE, the XhoI fragment from pTS89 was inserted into the XhoI site within the kanamycin gene of the vector, pACYC177, and the resulting plasmid was introduced into E. coli 20SO. Plasmid DNAs of transformants with ampicillin-resistant, kanamycin-sensitive, and metapyrocatechase-positive phenotypes were purified and subjected to restriction cleavage analysis. As expected, one of such plasmids, designated pTS115, yielded the 2.2-kilobase XhoI fragment. Fig. 1 shows the restriction map of the XhoI fragment of pTS115 and other plasmids constructed in this study. The direction of transcription of xylE and drug resistance genes are also presented. pTS115 contained the xylE gene downstream from the kanamycin promoter (10) of the vector. The crude extract of cells carrying this plasmid contained a significant amount of metapyrocatechase. To the contrary, cells carrying pTS88, a plasmid containing the same XhoI fragment as in pTS115 but with the opposite orientation with respect to the vector, did not produce the enzyme. Therefore, the expression of xylE in pTS115 is regarded to be due to a readthrough transcription from the vector promoter.

There were three SalI sites on the XhoI fragment as shown in Fig. 1. To see whether these sites were involved in the xylE gene, pTS115 DNA was partially digested with SalI, and the resulting two fragments were purified. One fragment that consisted of 1,700 bp 1 from the left SalI site to the right SalI site was ligated to SalI-cleaved pACYC184, and introduced into E. coli 20SO. Chloramphenicol-resistant, tetracycline-sensitive, and metapyrocatechase-positive transformants were selected. Such transformants carried a plasmid, pTS119, that contained xylE downstream from the tetracycline promoter. The other fragment consisting of 1180 bp from the left SalI site to the middle SalI site was ligated to SalI-cleaved pBR322 and introduced into E. coli 20SO. Ampicillin-resistant and tetracycline-sensitive transformants were obtained, but none was metapyrocatechase-positive. These transformants were found to have a plasmid, pTS114, that contained the ligated fragment of 1180 bp with the same orientation as in pTS119 with respect to the direction of the tetracycline gene of the vectors. The crude extract of cells carrying pTS114 did not contain metapyrocatechase. These results suggested that the xylE gene started in the left SalI fragment and ended in the right SalI fragment on the map in Fig. 1.

DNA Sequencing and Primary Structure of Metapyrocatechase—Metapyrocatechase consists of four identical subunits with Mr 35,000 (5), which corresponds to a gene of approximately 1000 bp. Since the above results showed the COOH-terminal region to be located in the right SalI fragment, the NH2 terminus should be located between 600 and 1100 bp from the 5`-end of the XhoI fragment (Fig. 1). In addition, the NH2-terminal amino acid sequence of 53 residues determined previously (5) provided information on the putative DNA sequence and its restriction sites in the NH2-terminal coding region. Comparing the putative restriction sites with the fine restriction map of the region, the PvuII site at the 1030 bp was suggested to be involved in the NH2-terminal region. Thus, a detailed restriction map was made for the assumed xylE region (Fig. 2).

Using the chemical method of Maxam and Gilbert (14), we determined the complete nucleotide sequence of the structural gene of metapyrocatechase according to the sequence strategy presented in Fig. 2. Sixty-two and 33% of the total nucleotide sequence were determined at least once on each strand, and twice on one strand, respectively.

The nucleotide sequence of xylE and the deduced amino

FIG. 1. Restriction maps of pTS115 and its derivatives. Stripped boxes indicate the TOL segment, and open boxes indicate the region of vectors. Gene symbols are as follows: xylE, metapyrocatechase gene; amp, ampicillin resistance gene; kan, kanamycin resistance gene; tetr, tetracycline resistance gene; cam, chloramphenicol resistance gene. Genes of vectors inactivated by fragment insertion are shown in parentheses. Closed circles indicate the promoter sites, and arrows indicate the direction of transcription. Symbols for restriction sites are as follows: ○, XhoI; ▼, BamHI; ●, SalI; ▼, EcoRI.

FIG. 2. Restriction map and sequence strategy of the DNA region for the metapyrocatechase gene. Restriction sites used for sequence determination are indicated by vertical lines. Horizontal arrows indicate the extent and the direction of sequencing from each site. Hatched region represents the metapyrocatechase gene.

1 The abbreviation used is: bp, base pairs.
Fig. 3. DNA sequence of the metapyrocatechase gene and the predicted polypeptide sequence. Nucleotide counting starts with the first letter of the initial methionine and is shown above the nucleotides. Amino acid numbering starts with the methionine and is shown in italics above the amino acid residues.

Table I

<table>
<thead>
<tr>
<th>Amino acid</th>
<th>Predicted</th>
<th>Direct analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lys</td>
<td>14</td>
<td>15.0</td>
</tr>
<tr>
<td>His</td>
<td>16</td>
<td>14.8</td>
</tr>
<tr>
<td>Arg</td>
<td>19</td>
<td>18.8</td>
</tr>
<tr>
<td>Asp</td>
<td>27</td>
<td>35.5</td>
</tr>
<tr>
<td>Asn</td>
<td>27</td>
<td>24.8</td>
</tr>
<tr>
<td>Thr</td>
<td>17</td>
<td>16.5</td>
</tr>
<tr>
<td>Ser</td>
<td>10</td>
<td>9.8</td>
</tr>
<tr>
<td>Gln</td>
<td>21</td>
<td>27.8</td>
</tr>
<tr>
<td>Gin</td>
<td>8</td>
<td>29</td>
</tr>
<tr>
<td>Pro</td>
<td>12</td>
<td>12.3</td>
</tr>
<tr>
<td>Gly</td>
<td>24</td>
<td>24.8</td>
</tr>
<tr>
<td>Ala</td>
<td>32</td>
<td>24.3</td>
</tr>
<tr>
<td>Cys</td>
<td>3</td>
<td>3.0</td>
</tr>
<tr>
<td>Val</td>
<td>21</td>
<td>23.0</td>
</tr>
<tr>
<td>Met</td>
<td>11</td>
<td>10.0</td>
</tr>
<tr>
<td>Ile</td>
<td>8</td>
<td>10.0</td>
</tr>
<tr>
<td>Leu</td>
<td>34</td>
<td>35.0</td>
</tr>
<tr>
<td>Tyr</td>
<td>12</td>
<td>6.5</td>
</tr>
<tr>
<td>Phe</td>
<td>16</td>
<td>15.0</td>
</tr>
<tr>
<td>Trp</td>
<td>5</td>
<td>3.8</td>
</tr>
<tr>
<td>Total</td>
<td>307</td>
<td>305.9</td>
</tr>
</tbody>
</table>

These values were calculated from the data of Nozaki et al. (2).

Table II

<table>
<thead>
<tr>
<th>Peptide position</th>
<th>2-6</th>
<th>66-68</th>
<th>147-156</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amino acid</td>
<td>Found*</td>
<td>Predicted</td>
<td>Found</td>
</tr>
<tr>
<td>Lys</td>
<td>1.00</td>
<td>1</td>
<td>0.23</td>
</tr>
<tr>
<td>His</td>
<td>1.00</td>
<td>1</td>
<td>1.05</td>
</tr>
<tr>
<td>Arg</td>
<td>1.18</td>
<td>1</td>
<td>1.90</td>
</tr>
<tr>
<td>Asp</td>
<td>1.01</td>
<td>1</td>
<td>1.44</td>
</tr>
<tr>
<td>Thr</td>
<td>0.36</td>
<td>1</td>
<td>2.98</td>
</tr>
<tr>
<td>Ser</td>
<td>0.31</td>
<td>1</td>
<td>3.31</td>
</tr>
<tr>
<td>Met</td>
<td>0.73</td>
<td>1</td>
<td>0.77</td>
</tr>
<tr>
<td>Phe</td>
<td>0.19</td>
<td>1</td>
<td>1.18</td>
</tr>
<tr>
<td>Gly</td>
<td>0.14</td>
<td>1</td>
<td>1.02</td>
</tr>
</tbody>
</table>

* Amino acid composition of fraction VII in Fig. 7 in the preceding paper (5).
  
Amino acid composition of fraction VIII in Fig. 7 in the preceding paper (6).
  
Amino acid composition of fraction VI in Fig. 7 in the preceding paper (5).
  
Methionine was determined as homoserine plus homoserine lactone.

molecular weight was calculated to be 35,155 from the predicted amino acid composition (Fig. 3 and Table I). This was in excellent agreement with the subunit $M_t = 35,000$ determined by sodium dodecyl sulfate-polycrylamide gel electrophoresis (5). $NH_2$- and COOH-terminal analyses of the enzyme protein established that this is indeed the structural gene for the enzyme. As demonstrated in the preceding paper (5), the...
NH₂ terminus of the enzyme is characterized by a sequence of 53 amino acids starting with Met-Asn-Lys-Gly-Val-, while the COOH-terminal sequence is -Phe-Met-Thr-Val-Leu-Thr. These sequences are in perfect agreement with the predicted amino acid sequences of the NH₂ and COOH termini shown in Fig. 3.

There is a close agreement between the amino acid composition previously determined (2) and that predicted from the DNA sequence (Table I). The difference in tyrosine contents between the determined and predicted compositions was probably due to the oxidation of tyrosine during the acid hydrolysis. In the course of this work, we have redetermined the amino acid compositions of small peptides obtained by cyanogen bromide cleavage and the corresponding peptides predicted from DNA sequence are shown in Table II. The close agreement in amino acid compositions between the found and predicted values confirms the postulate that the peptides predicted from the DNA sequence do exist in the protein.

Secondary Structure Prediction—The translated sequence of metapyrocatechase was analyzed for the secondary structure using the method of Chou and Fasman (Fig. 4) (17). The contents of α-helix and β-sheet structures were 29 and 26%, respectively. β-Turns were predicted in six positions along the polypeptide.

Base Composition and Codon Usage—The structural gene of metapyrocatechase consists of 921 base pairs. No significant direct repeat sequence in the coding region was detected. The DNA sequence shown in Fig. 3 has a G + C content of 57 mol %, which is closer to that of the P. putida chromosome (50-51%) than to that of E. coli chromosome (60-63%). The codon usage of the metapyrocatechase gene was apparently different from that of many E. coli genes as shown.
in Table III; there was a preference for using the C- and G-terminated codons over the U- and A-terminated codons of Arg, Ser, Thr, Pro, Ala, Gly, Val, and Lys. Such preferential codon usage was not found in a number of E. coli genes, especially in the ribosomal protein genes (19, 20). On the other hand, there were some similarities in the codon usage for xylE to E. coli genes; high usage of CUG(Leu) and GGPyr(Gly), and minimum use of UAU(Ile) and AGP(Arg).

**Nucleotide Sequence of Initiation Region for Protein Synthesis**—The RNA sequence of 31 nucleotides upstream from the initiation codon, AUG, is presented in Fig. 5. There were four termination codons, UGA, in two reading frames. In addition, there was a purine-rich region complementary to sequences near the 3'-end of the 16 S rRNA of both E. coli and P. aeruginosa, which is regarded as the ribosome-binding site for metapyrocatechase synthesis both in E. coli and P. putida (21).

**DISCUSSION**

Metapyrocatechase is an extradiol-type dioxygenase; it contains the ferrous form of iron and is composed of four identical subunits (1, 5). On the other hand, protocatechuate 3,4-dioxygenase and pyrocatechase are intradiol-type dioxygenases; both contain the ferric form of iron and consist of nonidentical subunits, $\alpha$ and $\beta$ (22-24). The amino acid sequences of these two types of dioxygenases may provide basic information for understanding of the enzymatic mechanism of these dioxygenases.

By the molecular cloning technique, we have constructed a series of hybrid plasmids in E. coli which contain the structural gene for metapyrocatechase (4, 8). The enzyme synthesized in E. coli carrying such a plasmid is indistinguishable from that produced by P. putida with regard to catalytic activity, the molecular weight of the subunit, and the antigenicity (4). Therefore, we have determined the nucleotide sequence of the structural gene for metapyrocatechase to deduce the amino acid sequence. The NH$_2$- and COOH-terminal amino acid sequences predicted from the DNA sequence were in perfect agreement with those of the enzyme (5). Further, the molecular weight and amino acid composition of the polypeptide predicted from the DNA sequence were in agreement with those experimentally determined. A mutant plasmid (pT5114) was isolated which did not direct the synthesis of the active enzyme. This plasmid had a DNA region corresponding to the NH$_2$-terminal two-thirds of the polypeptide.

The amino acid sequence of metapyrocatechase showed no apparent homology to those of either subunit of protocatechuate 3,4-dioxygenase. In addition, no sequence homology was observed between the enzyme and putidaredoxin, an iron-sulfur protein of P. putida which is a component of the camphor-5-exoxygenase system (28).

From the secondary structure prediction, metapyrocatechase contained the $\alpha$-helix region of 29% of the whole polypeptide. The low content of $\alpha$-helix of the enzyme was previously reported by Hirata et al. (29) based on studies of CD and ORD.

The DNA sequence of the metapyrocatechase gene is the first sequence that has been described for the structural genes derived from a pseudomonad source. The relatively high G + C content of the gene may reflect its origin from the genus *Pseudomonas*. The preferential usage of C- and G-terminated codons which was apparent in synonymous codons, contributed to the high G + C content of the gene; the frequencies of each nucleotide appearing in the third position of codons in xylE were calculated to be 18, 41, 10, and 31% for U, C, A, and G, respectively, whereas those appearing in the whole codons were 21, 29, 21, and 29% for the same nucleotides. It would be interesting to know whether the preferential codon usage is common to other *Pseudomonas* genes and accounts for the high G + C content of this genus.

The metapyrocatechase gene *xylE* is a component of the *xylDEFG* operon of the TOL plasmid, and the promoter region of the operon is separated from *xylE* by the gene for toluate dihydrodiol dehydrogenase (*xylD*) and probably also by the gene for toluate dihydrodiol dehydrogenase (4, 8). Several hybrid plasmids containing *xylE* were constructed *in vitro* in E. coli, pT5115 and a similar type of plasmid contained *xylE* downstream from the vector promoter without its own promoter. Such plasmids gave host bacteria a high ability to produce the enzyme (4). On the contrary, when *xylE* was under the control of its own operator-promoter, the expression was very low in E. coli even when activated by an inducer and an activator protein (4, 8). These results suggest that the low expression of *xylE* in *E. coli* is due to inefficient transcription of the foreign gene.

The 3'-terminal sequence of the 16 S rRNA of *E. coli* and *P. aeruginosa* has a high homology (21). It was reported that the ribosome-binding site of the coat cistron of an RNA phage, PP$_3$, of *P. aeruginosa* has a sequence complementary to that near 3'-ends of the 16 S rRNAs of these bacteria (30). Based on the initiation complex formation in *in vitro* between PP$_3$ RNA and ribosomes, it was shown that the initiation site for the coat protein synthesis is recognized by both *E. coli* and *P. aeruginosa* ribosomes. It has been shown here that the sequence of the 5'-region of the initiation codon of *xylE* is also complementary to the 3'-terminal sequence of the 16 S rRNAs. Although the 3'-terminal sequence of 16 S rRNA of *P. putida* has not been determined, it is likely to have a high homology with that of *P. aeruginosa*. It is plausible to assume that the sequence upstream from the initiation codon of *xylE* is recognized by these bacteria as a common ribosome-binding site for metapyrocatechase synthesis.

**Acknowledgments**—We are grateful to Dr. Takashi Miyata for secondary structure prediction by the method of Chou and Fasman. Thanks are also due to Dr. M. Iwaki and G. Yamazaki for their cooperation in analysis of amino acid sequence homology.

**REFERENCES**

Complete DNA Sequence of Metapyrocatechase Gene

Complete nucleotide sequence of the metapyrocatechase gene on the TOI plasmid of Pseudomonas putida mt-2.
C Nakai, H Kagamiyama, M Nozaki, T Nakazawa, S Inouye, Y Ebina and A Nakazawa


Access the most updated version of this article at http://www.jbc.org/content/258/5/2923

Alerts:
• When this article is cited
• When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

This article cites 0 references, 0 of which can be accessed free at http://www.jbc.org/content/258/5/2923.full.html#ref-list-1