Complete nucleotide sequence of the *E. coli* glutathione synthetase gsh-II

Hiroshi Gushima, Shuhay Yasuda, Eiichi Soeda, Masami Yokota, Masatoshi Kondo and Akira Kimura

1Research Institute for Food Science, Kyoto University, Uji, Kyoto 611, 2Central Research Laboratory of Yamanouchi Pharmaceutical Company, 1-1-8 Azusawa, Itabashi-ku, Tokyo 174, and 3National Institute of Genetics, 1-111 Yata, Mishima 411, Japan

Received 24 October 1984; Accepted 7 November 1984

ABSTRACT

The nucleotide sequence of the cloned DNA, 1,478 bp in length coding for glutathione synthetase (GSH-II) of *E. coli* B has been determined. Amino acid and nucleotide sequence analyses have assigned the open reading frame for GSH-II, starting with the ATG near its 5' terminus. The molecular weight calculated from the predicted amino acid sequence is 35,559 daltons, being in good agreement with that of a GSH-II subunit estimated by the SDS-PAGE method. Several signal sequences conserved in the promoter regions of *E. coli* were found in the non-coding regions of the gsh-II gene. They include the Shine-Dalgarno sequence, the Pribnow box and the sequence conserved in the "-35 region" with a preferable spacing from each other for an efficient transcription. Downstream from the termination codon, the inverted repeat sequences were present, followed by 6 successive T's. These structural features found in the non-coding regions have suggested to be involved in regulatory functions for the gsh-II gene expression.

INTRODUCTION

Glutathione synthetase (EC 6.3.2.3) (GSH-II), the second enzyme in the glutathione biosynthetic pathway in *E. coli* B, catalyzes the condensation of γ-glutamylcysteine and glycine in the presence of ATP to give glutathione. As shown previously (1-4), we have been studying the production of glutathione by *E. coli* cells in a bioreactor system. In order to attain a higher production of glutathione, the gene for GSH-II has been cloned into pBR322 and designated pGS200 (2). Recently the GSH-II has been purified from the cell extract of *E. coli* B transformed with pGS200 and well characterized enzymologically (5). An apparent molecular weight of native GSH-II was estimated to be about 152 kd by gel filtration method. This enzyme is active in a tetrameric form of the identical subunit with a molecular weight of approximately 38 kd. In the present work, we have sequenced the cloned DNA fragment, which includes the entire gsh-II gene as well as the flanking regulatory regions.
MATERIALS AND METHODS

Bacterial Strains

Strain E. coli B (C1001) and E. coli K-12 [C600(F−hsdR hsdM recA thy leu thi lacY supE tonA)] were used. The strain C1001 was deficient in GSH-II activity.

Enzymes and Biochemicals

Restriction endonucleases were obtained from Takara Shuzo Co. Ltd, and Nippon Gene Co. Ltd. Nuclease P1 was from Yamasa Shoyu Co. Ltd. [α-32P]dCTP was purchased from RCC Amersham.

M13 cloning and sequencing kits were obtained from either RCC Amersham or Takara Shuzo Co. Ltd. Tetramethylthiuram disulfide and other reagents used for enzyme assay were purchased from Sigma Chemical Co. Ltd.

DNA sequencing procedure

pGS200 was digested with the combination of BamHI and HindIII restriction endonucleases and the resulting 1400 bp fragment was separated from the vector by electrophoresis in a 10% low-melting-temperature agarose gel. The fragment was redigested with either RsaI or HaeIII, inserted into the corresponding the cloning sites of M13 mp8 and M13 mp9 phage vectors and subcloned according to the supplier's specifications (RCC Amersham).

Alternatively, the fragment was sonicated into pieces and inserted at the SmaI site of M13 mp8 DNA after processed with nuclease P1 and T4 DNA polymerase to generate flush ends. A shotgun library was constructed by the modified method (Yasuda et al., submitted) of Deininger (8). The resulting recombinant phage DNA was sequenced by the "dideoxy sequencing method" of Sanger et al (9).

DNA sequence analyses were performed with DNASIS (Hitachi SK Co.)

Purification of GSH-II and amino acid sequence determination

GSH-II was purified from E. coli C600 cells transformed with pGS200. The purification method was essentially as reported previously (5) except that FPLC system with Mono Q column (Pharmacia) was used in the final step. The amino-terminal sequence of the purified GSH-II at one nmol scale was determined by the use of Model 470A protein sequencer (Applied Biosystems).

RESULTS

Subcloning of the gsh-II gene

pGS200 was 6900 bp in length and contained the gsh-II gene derived from E. coli B chromosomal DNA fragment (2600 bp) inserted at the HindIII site of pBR322 (2). To obtain a plasmid containing the shorter gsh-II gene, we
Figure 1. Physical map of the recombinant plasmid pGS200 (Murata et al., 1983) and the strategy for the DNA sequencing of gsh-II gene.

The restriction sites of pGS200 DNA used for sequencing indicated at the map coordinated on the top column. At the bottom column in the expanding scale, the extent and direction of DNA sequencing are indicated by horizontal arrows. The open and closed circles represent the 5' termini of the cloned fragments generated by restriction enzyme digestion and sonication, respectively.

digested pGS200 with BamHI restriction endonuclease and cloned into the BamHI site of pBR325. These plasmids were used for transformation of E.coli B C1001 (gsh-II'). The resulting transformants (gsh-II') were selected according to the method of Murata et al. (2). The restriction mapping has indicated that the gsh-II gene is located within the HindIII-BamHI fragment (1400 bp) of pGS200 (Fig. 1).

Nucleotide sequence of the gsh-II gene

To determine the nucleotide sequence of the gsh-II gene, we first digested pGS200 with the combination of BamHI and HindIII to isolate the 1400 bp DNA fragment. The fragment was dissected further by either restriction enzymes or sonication and inserted into the corresponding cloning sites of the M13 vector. After transfection in E.coli JM101 strain, the recombinant phages were propagated and the resulting phage DNA was submitted to sequencing (Fig. 1).

The sequence of the 1,478 bp chromosomal DNA containing the gsh-II gene is shown in Figure 2. Examination of the nucleotide sequence shows only one
Nucleic Acids Research

1 AAGCT

6 TCAGCAGTGCCAGAAGATTGCATTCGCTCTGTGACCGATGTTCTGTAACCGGGTGCC

66 GAAATCCCTCCAGGCGATGGATCTGGAAGCCTGGTGTGCAGAGCAGGATGAAGGACTGAAA

126 CGCTATCAATTACTGATATCCTGTTGGACCTCGCGTTTTGCGTACAGAGACAACTGCG

186 CGCTATCAATTACTGATATCCTGTTGGACCTCGCGTTTTGCGTACAGAGACAACTGCG

-35 SQ

Pribnow start of mRNA

306 CTCAACCGCATTACCGCGCTTACAATACGATTTGGCGATTTGGGCTACAGAGACAACTGCG

366 ATG ATC AAG CTC GGC ATC GTG ATG GAC CCC ATC GCA AAC ATC AAC

1 Met Ile Lys Leu Gly Thr Val Met Asp Pro Ile Ala Asn Ile Asn

416 ATC AAG AAA GAT TCC AGT TTT GCT ATG TTG CTG GAA GCA CAG CGT

16 Ile Lys Lys Asp Ser Ser Phe Ala Met Leu Leu Glu Ala Glu Arg

476 CGT GGT TAC GAA CTT CAC TAT ATG GAG ATG GGC GAT CTG TAT CTG

31 Arg Gly Tyr Glu Leu Thr Met Glu Met Asp Leu Tyr Leu

536 CAG AAG AAC GAT TCC AGT TTT GCT ATG TTG CTG GAA GCA CAG CGT

46 Ile Asn Gly Glu Thr Val Met Asp Pro Ile Ala Asn Val Lys

601 CGT GGT TAC GAA CTT CAC TAT ATG GAG ATG GGC GAT CTG TAT CTG

31 Arg Gly Tyr Glu Leu Thr Met Glu Met Asp Leu Tyr Leu

661 GAA GAG AAA GGG ACC CTC GGC ATC GTT AAG AAC CCG CAG ACC CTC

106 Glu Glu Lys Gly Thr Leu Ile Val Asn Lys Pro Glu Ser Leu Arg

721 GAC TGT AAG CAG AAA CTG GTT ACC GCC TGG TTC TCT GAC TTA ACG

121 Asp Cys Gly Thr Leu Thr Arg Ser Glu Thr Phe Ser Thr Phe Leu Thr

771 CCA GAA ACG CTG GTT ACG GCC AAT AAA CCG CAG CTA AAA CCG TTC

136 Pro Glu Thr Leu Val Thr Arg Asn Lys Ala Glu Leu Lys Ala Phe

826 TGG GAG AAA CAC ACC GAC ATC ATT CTT AAG CCG CTG GAC GGT ATG

151 Trp Glu Thr His Ser Asp Ile Leu Lys Pro Leu Asp Gly Met

881 GCC GGC GCC TCG ATT GCC GTC GAC ATT AAG CCG CTG GCC ATC AGC

166 Gly Gly Ala Ser Ile Phe Arg Val Gly Gly Asp Glu Arg

931 GGC GTG ATT GCC GAA ACC CTG ACT GAG CAT GCC ACT GCC TAC TGC

181 Gly Val Ile Ala Glu Thr Leu Thr Glu His Gly Thr Arg Tyr Cys

981 ATG GCC CAA AAT TAC CTG CCA GCC ATT AAT CAC GTC ATG GGC GAC

196 Met Ala Glu Asn Tyr Leu Pro Ala Ile Lys Asp Gly Asp Lys Arg

9302
The nucleotide sequence of the gene is indicated and nucleotides are numbered from the HindIII site. The amino acid sequence of GSH-II predicted from the sequences is given below the sequence. Several regulatory sequences flanking the gsh-II gene are indicated with underlines. They include the Shine-Dalgarno (SD) sequences, the Pribnow box and the conserved sequence (-35 SQ) located at about -10 and -35 nucleotides, respectively, upstream from the starting point of mRNA synthesis which was tentatively assigned to position 268 or 267 as judged from the topology of the regulatory signal sequence described above. Downstream from the presumed termination codon of the gsh-II gene at positions 1314-1316, the inverted repeat sequences (IR) are presented at positions 1356-1364 and 1382-1390, followed by 6 successive T's which is a preferable site for the termination of mRNA synthesis.

possible open reading frame, starting at position 364 and terminating at position 1314, which was sufficiently long enough to code for a polypeptide of about 38 kd in molecular weight. In order to ascertain the initiation site of the gsh-II, we determined the amino terminal sequence of purified GSH-II. They were Met-Ileu-Lys-Leu-Gly-Ileu-Val-Met-Asp-Pro-Ileu-Ala-Asn-Ileu-Asn-Ileu-Lys-Lys-Asn-Ser-X-Phe-Ala-Met-Leu-Leu-Glu (Table 1). This sequence completely agrees with that of the first 27 amino acid predicted from the DNA sequence. This open reading region can code for a polypeptide of 316 amino acids. The amino acid sequence predicted for the GSH-II is
Table 1. Coincidence of the amino acid sequence at the NH₂-terminus of GSH-II with that predicted from nucleotide sequence.

<table>
<thead>
<tr>
<th>NH₂-terminal</th>
<th>Codon</th>
<th>Amino Acid</th>
<th>NH₂-terminal</th>
<th>Codon</th>
<th>Amino Acid</th>
</tr>
</thead>
<tbody>
<tr>
<td>Residue Number</td>
<td></td>
<td></td>
<td>Residue Number</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>ATG</td>
<td>Met</td>
<td>15</td>
<td>AAC</td>
<td>Asn</td>
</tr>
<tr>
<td>2</td>
<td>ATC</td>
<td>Ile</td>
<td>16</td>
<td>ATC</td>
<td>Ile</td>
</tr>
<tr>
<td>3</td>
<td>AAG</td>
<td>Lys</td>
<td>17</td>
<td>AAG</td>
<td>Lys</td>
</tr>
<tr>
<td>4</td>
<td>CTC</td>
<td>Leu</td>
<td>18</td>
<td>AAA</td>
<td>Lys</td>
</tr>
<tr>
<td>5</td>
<td>GGC</td>
<td>Gly</td>
<td>19</td>
<td>GAT</td>
<td>Asp</td>
</tr>
<tr>
<td>6</td>
<td>ATC</td>
<td>Ile</td>
<td>20</td>
<td>TCC</td>
<td>Ser</td>
</tr>
<tr>
<td>7</td>
<td>GTG</td>
<td>Val</td>
<td>21</td>
<td>AGT</td>
<td>X</td>
</tr>
<tr>
<td>8</td>
<td>ATG</td>
<td>Met</td>
<td>22</td>
<td>TTT</td>
<td>Phe</td>
</tr>
<tr>
<td>9</td>
<td>GAC</td>
<td>Asp</td>
<td>23</td>
<td>GCT</td>
<td>Ala</td>
</tr>
<tr>
<td>10</td>
<td>CCC</td>
<td>Pro</td>
<td>24</td>
<td>ATG</td>
<td>Met</td>
</tr>
<tr>
<td>11</td>
<td>ATC</td>
<td>Ile</td>
<td>25</td>
<td>TTG</td>
<td>Leu</td>
</tr>
<tr>
<td>12</td>
<td>GCA</td>
<td>Ala</td>
<td>26</td>
<td>CTG</td>
<td>Leu</td>
</tr>
<tr>
<td>13</td>
<td>AAC</td>
<td>Asn</td>
<td>27</td>
<td>GAA</td>
<td>Glu</td>
</tr>
<tr>
<td>14</td>
<td>ATC</td>
<td>Ile</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

X indicates unidentified amino acid shown in Fig.2. The amino acid composition to be as follows: Ala₂₂, Arg₁₉, Asn₁₂, Asp₂₀, Cys₄, Gln₁₁, Glu₂₇, Gly₂₅, His₄, Ileu₂₆, Leu₃₃, Lys₁₈, Met₁₀, Phe₁₁, Pro₁₇, Ser₁₀, Thr₁₉, Trp₄, Tyr₉, and Val₁₅. The molecular weight calculated from the predicted amino acid sequence is 35,559 daltons, which agrees well with the molecular weight of the GSH subunit (38 kd) estimated by SDS-PAGE (5). Since an apparent molecular weight of the GSH-II in an active state estimated by molecular sieve chromatography is about 152 kd, active GSH-II seems to be a tetrameric enzyme composed of four identical subunits. Codon usage for E.coli GSH-II derived from DNA sequence data is shown in Table 2.

Transcriptional signals

Prokaryotic consensus sequences for transcriptional initiation have well been documented (10-12). In the precise DNA sequence shown in Figure 2, we find two hexanucleotides TTGCCA and TTTACT at positions 232 and 255, respectively, preceding to the initiation codon ATG at position 366. The former sequence matches in five out of six positions to the consensus
sequence TTGACA conserved in the "-35 region" upstream from the initiation site (10) and is identical to that reported for purf gene (13). The latter sequence also agrees in four out of six nucleotides with the Pribnow box TATAAT (10). Furthermore, these consensus sequences are separated from each other by 17 bp, presumably being the most preferred spacing for an efficient transcription (11). Regarding a transcriptional termination signal, it has not been studied so extensively. The inverted repeat sequences can be located at positions 1356 and 1382, which can form a stable hairpin loop structure. This inverted repeats are immediately followed by T-rich sequence. This sequence arrangements is often found in the prokaryotic terminal region of mRNA (10).
We have established the complete nucleotide sequence of *E. coli* B gsh-II gene, including the flanking regulatory regions. The DNA sequence upstream from the 5'-terminus of the gsh-II gene (position 220-270) is slightly A-T rich, where the promoter is expected to be located. The sequence of putative "-35 region" and Pribnow boxes of the gsh-II gene are both very similar to the known consensus nucleotide sequences (10). In this region, another Pribnow box like sequence TATCCT is found at position 264. However the sequence is separated from the "-35 region" by 25 bp, the spacing being far from the typical spacing of 17 nucleotides (10). Importance of the critical spacing between the two conserved sequence is further strengthened by the DNA of considering the direct contact sites of RNA polymerase with the promoter region on three-dimensional model (12). From these points, we assigned the sequence TTTACT at position 255 to be the Pribnow box. However, further biochemical and genetic experiments will be required to confirm our tentative promoter sequence as well as the transcriptional start point. Furthermore, about 10 nucleotides upstream from the initiation codon ATG a potential Shine-Dalgarno sequence GGAG (15) is located (Fig. 2).

About 40 bp downstream from the translational termination codon TAA, there exist GC-rich inverted repeated sequences followed by a stretch of successive Ts. This can form a very stable structure with a G of -23 kcal/mol, calculated according to the method of Tinoco et al (14). Presumably they constitute the transcriptional termination signal.

A molecular weight calculated from the 316 amino acid residues encoded from the gsh-II gene is 35,549 daltons, and is in good agreement with that estimated from SDS-PAGE (38 kd). Recently, by the high performance liquid chromatography using a SW 3000 column (Toyo Soda), a more precise molecular weight of 140 kd for the GSH-II holoenzyme was determined. As the GSH-II is comprised of four identical subunits (5), the subunit molecular weight (35.6 kd) calculated from the DNA sequence data fits even better to this estimate than does the value (38 kd) by SDS-PAGE. Codon usage in the gsh-II gene shows no strong bias from the common tendency (Table 2). We can estimate the frequency of optimal codon usage (Fop) to be 0.72, according to the method of Ikemura and Ozeki (16). Thus, gsh-II gene seems to be only moderately expressed in *E. coli* cells.

We have been establishing a bioreactor system for the production of glutathione by the use of the immobilized *E. coli* cells modified by genetic engineering techniques (1-4). When both gsh-I and gsh-II genes inserted
simultaneously into pBR325 were introduced into the E.coli K-12 C600 cells, the productivity of glutathione was very much enhanced. While the gsh-I gene is now being sequenced, the knowledge of nucleotide sequences of both genes will enable us to improve the expression rate of these genes by gene manipulation.

ACKNOWLEDGMENTS

We thank Prof. M. Takanami, Institute for Chemical Research of Kyoto University, Assoc. Prof. M. Sakaguchi, Dr. K. Murata, Dr. Y. Fukuda, Mr. K. Watanabe, Research Institute for Food Science, Kyoto University, and Mr. T. Miya, Research Laboratory of Kojin Company, Ltd., for many helpful discussions.

REFERENCES
