Nucleotide sequence of the thioredoxin gene from Escherichia coli

Jan-Olov Höög, Hedvig von Bahr-Lindström, Staffan Josephson, Betty J. Wallace, Sidney R. Kushner, Hans Jörnvall, and Arne Holmgren

1Department of Chemistry I, Karolinska Institutet, S-104 01 Stockholm; 2Department of Chemistry, KabiGen AB, S-112 87 Stockholm, Sweden; and 3Department of Molecular and Population Genetics, University of Georgia, Athens, GA 30602, U.S.A.

(Received 27 September 1984)

The nucleotide sequence of the thioredoxin gene from Escherichia coli was determined. The structural gene was identified on a cloned 3-kb PvuII fragment by hybridization with a synthetic oligodeoxyribonucleotide corresponding to a part of the amino acid sequence of thioredoxin. Restriction-enzyme fragments were used as templates in the dideoxy sequence method, directly and after subcloning into M13mp8. A segment of 450 nucleotides was determined using both strands, alternatively, without extensive overlaps. The sequence contains the thioredoxin coding region, a potential ribosome-binding site, and a putative promotor region. The predicted amino acid sequence differs by two inversions from the previously given thioredoxin sequence. The revised sequence is presented and the results further show that thioredoxins from E. coli B and K12 are identical.

Thioredoxin from Escherichia coli has suggested functions as a general protein disulphide reductase (1) and as a hydrogen donor for ribonucleotide reductase (1) and is also an essential subunit of phage T7 DNA polymerase (2).

Thioredoxin from E. coli consists of 108 amino acids, and a primary structure for this protein has been deduced (3). The tertiary structure has also been determined (4) as well as the relationship to glutaredoxin (5-7) and intracellular localizations (8). Mapping experiments have located the gene for thioredoxin, trxA, at 84 min on the E. coli chromosome (9). Recently, a segment of the E. coli chromosome containing the thioredoxin gene has been ligated into the plasmid pBR325 (10). The resulting plasmid pBHK8 transformed into E. coli K12 (strain: SK3981) overproduces thioredoxin 150- to 200-fold (10).
Knowledge of the nucleotide sequence of the thioredoxin gene will enable studies using site-directed mutagenesis to define structure-function relationships of thioredoxin.

In this study, we report determination of the nucleotide sequence of the thioredoxin gene from \textit{E. coli}, independent of similar studies elsewhere (11). We analyze the protein and compare the amino acid prediction from DNA with the primary structure of thioredoxin.

**Materials and Methods**

**Preparation of DNA**

The plasmid pBHK8 was isolated from \textit{E. coli} SK3981 with the alkaline lysis method (10,12). Digestions of plasmid DNA by restriction endonucleases were performed using the conditions suggested by the manufacturer (Boehringer). Restriction endonuclease fragments were separated by agarose-gel electrophoresis and isolated by electro-elution in dialysis bags (12).

**Synthesis of oligodeoxyribonucleotides**

Oligodeoxyribonucleotides were synthesized by the solid-phase phosphoamidite method using an automatic synthesizer (13,14). The concentrations of the four 5'-O-dimethoxytrityl-2'-deoxy-nucleoside-3'-O-methyl phosphoramidite reagents were measured (14) prior to making the mixtures so as to get equal ratios between reagents. Initially, a mixed 14-mer was synthesized corresponding to the known amino acid sequence (3). When part of the DNA structure was known, a 15-mer was synthesized corresponding to the nucleotide sequence.

**Hybridization with synthetic oligodeoxyribonucleotides**

Oligodeoxyribonucleotides (0.01 \(\mu\)g/\(\mu\)l) were labelled at the 5' end by transfer from \([^{32}\text{P}]\text{ATP} \) (Amersham; 3000 Ci/mmol; 1:2 molar ratio \([^{32}\text{P}]\text{ATP}:\text{oligodeoxyribonucleotides} \) by using T4 polynucleotide kinase (New England Biolabs; 2 U/\(\mu\)l) in 50-ram Tris/HCl pH 7.5, 10-mM MgCl\(_2\), 10-mM dithiothreitol for 45 min at 37°C. Excess reagents were removed by separation on Sephadex G-50 (0.25 x 15 cm) in 10-mM Tris/HCl, 1-mM EDTA, pH 7.5. Radioactively labelled oligodeoxyribonucleotides (3 x 10\(^8\) c.p.m./\(\mu\)g) were used as probes for hybridization analysis by the Southern blot (15,16) and dot-blot (17) techniques.

**Nucleotide sequence and protein analysis**

\textit{Sau3A} fragments were cloned into the \textit{BamHI} site of M13mp8 (18). Transfection of \textit{E. coli} JM 103 was performed as described (19).

Sequence analysis by the dideoxy method (20,21) utilizing single-stranded M13-templates (22) and restriction-enzyme fragments after heat-induced strand separation (23) was carried out with either an M13-specific universal primer (Amersham) or with the thioredoxin-specific oligodeoxyribonucleotides as primers. The labelled nucleotide was \([^{35}\text{S}]\text{dATP} \) (Amersham; 600 Ci/mmol) and the reaction mixture was separated by electrophoresis in a 0.2-mm-thick 6% urea-polyacrylamide gel.
Fig. 1. Identification of the fragment from pBHK8 containing the thioredoxin gene. Lane A: Ethidium bromide stain after agarose-gel electrophoresis of a PvulI digest of pBHK8, showing the 3-kb and 2.6-kb fragments (10). Lane B: Autoradiograph of lane A after transfer to nitrocellulose by the method of Southern (15) and hybridization with the 32p-labelled mixed 14-mer oligodeoxyribonucleotide (Table I).

Thioredoxin from *E. coli* SK3981 was prepared as described (10) and manual sequence analysis was carried out with the DABITC method (24,25).

**Results**

The plasmid pBHK8 was cleaved with *Pvu*II and the digest was separated by agarose-gel electrophoresis (Fig. 1). A 3-kb fragment containing the thioredoxin gene was identified (Fig. 1) after transfer to nitrocellulose by the method of Southern and hybridization with the mixed 14-mer synthetic probe (Table I). The 3-kb fragment was isolated from agarose gel and used for sequence analysis directly and after subcloning into M13mp8.

**Table 1. Synthetic oligodeoxyribonucleotides used as hybridization probes and primers in dideoxy sequence analysis**

<table>
<thead>
<tr>
<th>Amino acid sequence</th>
<th>14-mer</th>
<th>15-mer</th>
</tr>
</thead>
<tbody>
<tr>
<td>28 Trp - Ala - Glu - Trp - Cys -</td>
<td>3' ACCCGTCTTACAC 5'</td>
<td>5' ACTGACCGTGGCAAA 3'</td>
</tr>
<tr>
<td>32 Lys - Leu - Thr - Val - Ala - Lys</td>
<td>5'</td>
<td></td>
</tr>
</tbody>
</table>

NUCLEOTIDE SEQUENCE OF *E. coli* THIOREDOXIN GENE
Fig. 2. Summary of nucleotide sequence data. Line A represents the complete structure determined, with amino acid sequence numbering above the line and nucleotide sequence numbering below. Line B represents the two Sau3A fragments sequenced in the M13 system. Lines C and D show the regions sequenced directly from the 3-kb fragment using as primer, respectively, the mixed 14-mer and the 15-mer oligodeoxyribonucleotides (Table I).

The 3-kb fragment was digested with Sau3A, and the mixture obtained was cloned into M13mp8. A 48-nucleotide fragment, identified as complementary to the mixed 14-mer probe by the dot-blot hybridization technique, was sequenced (Fig. 2). Based on the sequence of another Sau3A fragment (Fig. 2) from the thioredoxin gene, a 15-mer oligodeoxyribonucleotide was synthesized (Table I) for use as a primer in the sequence analysis.

The isolated 3-kb fragment was used as template in direct dideoxy sequence analysis after heat-induced strand separation. The mixed 14-mer oligodeoxyribonucleotide was utilized as primer to determine the nucleotide sequence of the coding strand on the 5' side of the region complementary to the primer (Fig. 2) (26). The 15-mer oligodeoxyribonucleotide was used in the same way to determine the sequence of the non-coding strand on the 3' side of the oligodeoxyribonucleotide (Fig. 2). The resulting nucleotide sequence of the thioredoxin gene is shown in Fig. 3.

The N-terminal sequence of thioredoxin from E. coli strain SK3981 was determined with the DABITC method and found to be Ser-Asp-Lys-Ile-, which is identical to that of E. coli strain B (3).

Discussion

The nucleotide sequence of the thioredoxin gene from E. coli K12 is given in Fig. 3 together with the predicted amino acid sequence. The predicted protein structure was found to differ from the previously reported amino acid sequence (3) at two places. The differences concern positions 16-17 and 71-72, which have been quoted as Leu-Val and Ile-Gly, respectively (Fig. 5 in ref. 3). However, these structures were mis-quotations of original protein data, giving Val-Leu (27) and Gly-Ile (Fig. 4 in ref. 3), and the orders are now verified and established by the present nucleotide sequence. The entire gene was not analyzed on both strands. This was considered unnecessary since the nucleotide sequence agrees with the amino acid sequence.
Fig. 3. Nucleotide sequence of the thioredoxin gene and its corresponding amino acid sequence. The nucleotide sequence of 450 nucleotides comprising the non-coding strand of the thioredoxin gene is shown in the 5'-3' direction. The proposed ribosome-binding site is boxed and the -35 and -10 sites of the proposed promoter are indicated by lines below the nucleotide sequence.

The *E. coli* K12 strain SK3981 containing the plasmid pBHK8 produces thioredoxin in amounts greater than 300 mg per litre of culture medium (10), which is 150-200 times more than wild-type *E. coli* B. A normal processing of the thioredoxin N-terminus by the overproducing strain is indicated by the N-terminal serine residue, which is identical to that of the wild-type protein. This contrasts with N-termini determined for glutaredoxin, a small redox-protein belonging to the same superfamily as thioredoxin (5). The glutaredoxin from the overproducing *E. coli* K12 strain C10-17 has an N-terminal methionine residue, followed by a glutamine residue (5), compared to the protein from *E. coli* B which starts with glutamic acid/glutamine (28). Altogether, the nucleotide sequence determination and the presently determined N-terminal amino acid sequence compared with the previously known amino acid sequence shows that there is no strain variation for thioredoxin from *E. coli* B and K12 strain SK3981.
The thioredoxin coding region is initiated by an AUG triplet. It is preceded by a potential ribosome-binding site (G-G-A-G-U-U) showing five out of six residues identical to the Shine-Dalgarno sequence (G-G-A-G-G-U) (29). The coding region is terminated by an ochre (U-A-A) codon. The best putative promotor region is 70 base pairs upstream of the translation-initiation codon. It comprises a Pribnow box and a -35 region (Fig. 3); both showing four out of six residues identical to the respective consensus sequences (30) in E. coli. Therefore, it seems as if the transcription of E. coli thioredoxin is governed by its own promotor region, which is consistent with the fact that thioredoxin is an abundant protein (10^4 copies/cell (31)). The results also show that thioredoxin is not synthesized with a signal sequence, although it is localized in the periphery of E. coli cells.

During the identification and nucleotide sequence determination of a plasmid containing the rho gene from E. coli K12, the presence of an unidentified, expressed polypeptide of 12 000-14 000 Da was observed (32). The sequence of the first 60 nucleotides of the 390 preceding the coding part of the rho gene (32) is identical to the last residues of the coding part of the thioredoxin gene (Fig. 3). This identifies the unknown polypeptide as thioredoxin and furthermore locates the gene for thioredoxin upstream of the rho gene in E. coli K12.

Acknowledgements
We thank Drs. Lars Hellman, Erik Holmgren, and Lars Wieslander for valuable discussions and Eva Lagerholm for synthesis of the oligodeoxyribonucleotides.
This work was supported by grants from the Swedish Cancer Society, the Swedish Medical Research Council, Magn. Bergvall's Foundation, and the National Institute of Health (6m27997).

References