Nucleotide Sequence and Protein Overproduction of Bacteriophage T4 Thioredoxin

DAVID M. LeMASTER

Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, Connecticut 06511

Received 18 February 1986/Accepted 10 June 1986

The bacteriophage T4 nrdC gene was cloned and physically mapped to 47.6 kilobases from the reference BamHI site. The DNA sequence is consistent with that reported from earlier protein sequence studies. The gene was subcloned into a lambda pL overexpression vector which allowed for the isolation of approximately 5 mg/liter.

The bacteriophage T4 nrdC gene encodes a small dithiol-disulfide couple redox protein, thioredoxin. It has been mapped between genes 49 and H (15), which is the most poorly characterized region of the T4 genome. Although numerous in vitro transcripts have been found in this region, recent genetic-physical correlations have not located any genes in the 15-kilobase (kb) stretch between genes 49 and tk (3). This note describes the mapping, cloning, and sequence of the nrdC gene as well as a closely adjacent open reading frame.

Additional interest in T4 thioredoxin arises from the considerable similarity between its tertiary structure and that of Escherichia coli thioredoxin (2). This is in marked contrast to the virtual absence of sequence homology (12). As a result, the two proteins are believed to offer a useful comparative system for the study of protein dynamics. E. coli thioredoxin has already proven suitable for such studies, as evidenced by nuclear magnetic resonance (4, 7), circular dichroism (9), fluorescence (6, 14), and cleavage and reconstitution (5, 13) experiments. Unfortunately, the solution properties of the T4 thioredoxin are far less well known. This is primarily due to the difficulties involved in purifying adequate amounts of the material from its natural source. To provide larger amounts of the protein as well as render labeling and mutagenesis experiments more tractable, we have subcloned the T4 thioredoxin gene into a lambda pL induction vector for overproduction.

The nrdC gene was mapped to the Psrl-Xbal fragment extending from 45.7 to 49.1 kb by the Southern blot technique with a 32P-labeled synthetic oligonucleotide (A-APUR-T-G-PYR-G-T-PUR-T-A-PYR-T-G-PYR-G-A), which corresponds to the active-site protein sequence, with degenerate mixes added for the third codon positions. The 5.3-kb Xbal fragment covering this region was preparatorily isolated by gel electrophoresis and cut with Psrl and Kpnl. The gene was found to reside on the 1.9-kb Kpnl-Xbal fragment which was cloned into the plasmid pUC19. Restriction analysis of this clone indicated that the gene was located near the Kpnl terminus. A 490-base-pair (bp) HaeIII fragment extending from the 5’ flanking region of the gene to the HaeIII site adjacent to the polylinker sequence was cloned in both orientations into the M13 derivative mp9 and sequenced by the dyeoxynucleotide method. The 620-bp Rsal fragment covering the 5’ flanking region was partially sequenced as well. The DNA sequence of the thioredoxin gene and the flanking regions is given in Fig. 1. The predicted protein sequence identically matches the sequence previously reported (12). The T-A-A sequence which is only one base pair away from the initiator codon serves to terminate an open reading frame, at minimum, 300 bp long. Unfortunately, the paucity of genetic data for this region of the T4 map makes it impossible to suggest an identity for the gene product which may be encoded. Furthermore, a survey of the National Biomedical Research Foundation protein data bank failed to reveal any strongly homologous sequences. There is also a short 105-bp open reading frame starting at the 3’ terminus of the thioredoxin coding region.

The 490-bp HaeIII fragment was cloned into the Hpal site of pKC30 (10) to yield pDL51. The plasmid pKC30 carries the lambda pL promoter upstream from the Hpal site. When pDL51 is transformed into M5219 (8), a strain carrying only part of the lambda functions, including the temperature-sensitive repressor C557, T4 thioredoxin was produced as roughly 5% of the total protein. This higher level of synthesis facilitated simplification of the previously published purification procedure (1, 11).

A culture was grown in rich medium to mid-log phase at 35°C, shifted to 43°C, and grown for 3 h. The cells were ruptured by sonication in 50 mM sodium phosphate (pH 7.0) containing 5 mM 2-mercaptoethanol. Streptomycin sulfate and ammonium sulfate precipitations were performed as described earlier (1) by using narrower (35 to 70%) ammonium sulfate fractionation. The (NH4)2SO4 pellet from 20 to 25 g (wet weight) of cells was suspended in 50 mM ammonium acetate (pH 6.1) and fractionated on an Ultragel AcA54

FIG. 1. Nucleotide sequence of bacteriophage T4 nrdC gene. The predicted protein sequence of T4 thioredoxin is given below the nucleotide sequence.
column (2.5 by 90 cm) equilibrated in the same buffer. The pooled fractions were then loaded onto a Biorex 70 column equilibrated in the acetate buffer, and the T4 protein was eluted with a gradient to 0.8 M ammonium acetate. After dialysis and lyophilization, approximately 25 mg of T4 thioredoxin was recovered, which appeared homogeneous by sodium dodecyl sulfate-polyacrylamide gel electrophoresis.

This work was supported by project grant PCM8305203 from the National Science Foundation and Public Health Service program grant GM22778 from the National Institutes of Health.

I thank F. M. Richards and T. Platt for the use of their laboratory facilities and for their advice during this work. I also thank T. Lin and E. Spicer for providing the phage DNA and N. Grindley for offering the use of his Biosearch SAM-1 oligonucleotide synthesizer. The assistance of M. Bannon and J. Mouning in preparing the manuscript is also appreciated.

LITERATURE CITED