JOURNAL OF BACTERIOLOGY, May 1991, p. 2864–2871
0021-9193/91/092864-08$02.00/0

Vol. 173, No. 9

Two Divergently Transcribed Genes, *soxR* and *soxS*, Control a Superoxide Response Regulon of *Escherichia coli*

JIE WU AND BERNARD WEISS*

Department of Pathology, University of Michigan Medical School, Ann Arbor, Michigan 48109-0602

Received 26 December 1990/Accepted 2 March 1991

***soxR* governs a superoxide response regulon that contains the genes for endonuclease IV, Mn^{2+}-superoxide dismutase, and glucose 6-phosphate dehydrogenase. The *soxR* gene encodes a 17-kDa protein; some mutations of this gene cause constitutive overexpression of the regulon. Induction by paraquat (methyl viologen) requires both *soxR* and a new gene, *soxS*. *soxS* is adjacent to *soxR*, it encodes a 13-kDa protein, and it is required for paraquat resistance. These functions were revealed by studies in which the sequence of the 1.1-kb *soxR-soxS* region was determined, the 5′ ends of the mRNAs were mapped, and complementation tests were performed with *soxRS* plasmids containing deletions of known sequence. The two genes are divergently transcribed, and the transcripts overlap. The *soxS* promoter is within the 85-nucleotide intergenic region, whereas the *soxR* promoter is within *soxS*. *soxS* mRNA increases after induction. Both protein products have possible DNA-binding (helix-turn-helix) domains. SoxR contains four cysteines (CX_2CXCX_5C) that might be part of a sensor region. SoxS shows 17 to 31% homology to the C-terminal portions of members of the AraC family of positive regulators.**

Bacteria undergo specific multigene global responses upon exposure to noxious stimuli or to new nutrient environments (30). The responses are largely mediated by regulons, groups of genes and operons under common control. The means of control are diverse and sometimes elaborate. They may involve new transcriptional activators or repressors, proteolytic cleavage or other posttranslational modifications, positive or negative feedback loops, and regulatory cascades. These pathways provide new models for understanding multigene regulation in all organisms. This study deals with a regulon governing part of the response of *Escherichia coli* K-12 to superoxide.

Superoxide anion radicals are generated during normal aerobic metabolism by the incidental autooxidation of respiratory chain coenzymes. Some one-electron redox reagents like paraquat (methyl viologen) may also generate superoxide via redox cycling in vivo. Upon exposure to such compounds, *E. coli* increases the synthesis of about 40 proteins (21, 47). Nine of these proteins are produced by a regulon controlled by *soxR*, a gene located at 92 min on the linkage map (22, 46). The gene products known to be governed by *soxR* include the DNA repair enzyme endonuclease IV, glucose 6-phosphate dehydrogenase, Mn^{2+}-superoxide dismutase, paraquat diaphorase, a modified ribosomal protein, and an antisense inhibitor of *ompF*, the gene coding for a major porin. *soxR* was identified through mutants obtained in two laboratories by independent means, namely, by looking for mutations affecting *nfo* (endonuclease IV) gene expression (46) or for mutations increasing resistance to superoxide generators (22).

In previous work, we found that *soxR* plasmids expressed a 17-kDa protein that was altered by mutations that caused a constitutive overexpression of the regulon (46). *soxR* was defined as the structural gene for this protein. However, the plasmids also contained a divergently transcribed gene encoding a 13-kDa polypeptide, the production of which was eliminated by some mutations affecting *soxR* complementation. Because the transcripts of the divergent genes might overlap, we did not know whether the 13-kDa product was also a regulatory protein or whether a mutation in its gene was primarily affecting transcription of *soxR* (46, 51). In this study, we resolve this question. Through DNA sequencing, mRNA mapping, and complementation analysis with sequenced deletions, we show that the 13-kDa SoxS protein is required together with SoxR for induction of the regulon and that SoxS is structurally similar to some known transcriptional activators.

MATERIALS AND METHODS

Media and reagents. TY medium (52) was used for routine bacterial growth and 2× YT medium (35) was used for propagation of M13 phages. Reagents used in generating M13 deletions (15) were obtained from International Biotechnologies, and reagents used in DNA sequencing (Sequenase 2.0) were from U.S. Biochemicals. Avian myeloblastosis virus reverse transcriptase was obtained from Boehringer-Mannheim. Oligonucleotides were synthesized at the DNA Synthesis Facility at the University of Michigan.

Bacterial strains. Strains BW841 (Δ*sox-8*::*cat*) and BW840 (*soxS3*::Tn*10*) contained prophage λIT1 (46), which bears an *nfo′-lacZ* operon fusion. They were F^+ derivatives of strains BW829 and BW803 (46) prepared by conjugation (27) with the F^+ strain CR63 (4) at a donor/recipient ratio of 10:1; counterselection was for streptomycin resistance. The F^+ phenotype was confirmed by cross-streaking a loopful of cells with a needle dipped in a suspension (10^9/ml) of the F-inhibited phage T7 and comparing the zones of killing after incubation. Strain BW856 had the same genotype as BW841, except that it bore a noninducible analog of prophage λIT1 isolated as a temperature-resistant derivative of the *c*I857 (Ts) *ind* mutant prophage λIT2 (46). Noninducibility of the prophage was verified by exposure to UV light (41); strain BW856 showed an 18% increase in β-galactosidase activity 1 h after exposure and no eventual lysis, whereas strain BW841 displayed a 3.7-fold increase and almost complete lysis. Strain BW824(pIT24), which carries *soxR4*::*cat* muta-

* Corresponding author.

tions both on its chromosome and on a pBR322-derived plasmid, was described previously (Table 6 of reference 46).

M13 phage clones. A 3.6-kb *Pst*I-*Bam*HI segment of the *soxRS* plasmid pIT15 (46) was subcloned in phages M13 mp18 and mp19 (53), and the recombinant was kindly provided by Irina R. Tsaneva. The cloned DNA consisted of 2.6 kb of chromosomal DNA bracketed by *Pst*I-*Pvu*I and *Pst*I-*Bam*HI segments of plasmid pBR322. Because the large plasmids were unstable, partial deletion of the mp19 derivative was done by the method of Dale et al. (15). The resulting plasmid retained the 0.1-kb pBR322 *Pst*I-*Pvu*I piece and 1.1 kb of chromosomal DNA clockwise of the chromosomal *Pvu*I site. The 1.2-kb insert was excised by cutting at vector *Hin*dIII and *Eco*RI sites and then cloned in the opposite orientation in M13 mp18. Further unidirectional deletions were obtained. They extended into the insert from the *Eco*RI site in phage mp19 and from the *Hin*dIII site in mp18. The resulting plasmids were used for both DNA sequencing and complementation analysis.

DNA sequencing. DNA sequencing was performed with Sequenase version 2.0 (U.S. Biochemicals Corp.); ambiguities were resolved by the use of Ssb protein, dITP, and Mn^{2+} according to the manufacturer's directions. Both strands were completely sequenced.

mRNA mapping. Bacterial strains were grown with and without paraquat induction as described below under "Complementation analysis." Cellular RNA was extracted as described elsewhere (36) except that the cells were frozen and thawed for two cycles in lysis buffer (7). Different protocols were used for primer extension mapping of *soxS* mRNA (36) and *soxR* mRNA (3). The primers were synthetic oligonucleotides complementary to nucleotides 2 to 21 of *soxR* and to nucleotides 3 to 23 of *soxS* (Fig. 1). The extended primers were measured against DNA sequencing ladders prepared from the same radiolabeled primers on phage DNA templates. The RNA sources were BW841 (M13 mp18::*soxRS*) for *soxR* and BW824(pIT24[*soxR4*::*cat*]) for *soxS*. The DNA templates were M13 mp19::*soxRS* for *soxR* and M13 mp18::*soxRS* for *soxS*.

Complementation analysis. To 10 ml of 2× YT medium were added 10^{11} to 10^{12} recombinant M13 phages and 0.1 ml of a saturated culture of the host cells. The mixture was shaken at 37°C in a 250-ml flask with a 14-mm-diameter side cuvette (Bellco), and growth was monitored with a Klett colorimeter and a 560-nm filter. When the reading for the mixture was 40 Klett units (1×10^8 to 2×10^8 cells per ml), 4-ml portions were transferred to two 50-ml flasks and paraquat was added to one flask to a final concentration of 0.2 mg/ml. Incubation was continued with vigorous aeration (400 rpm on a gyratory shaker) for 1 h. Assays for β-galactosidase were done as previously described (46). At the time of harvesting, the cultures were streaked on TY agar plates, and 24 subclones of each strain were tested for phage production. In each case, 75 to 100% of the cells were infected. Tests for paraquat sensitivity were performed on fresh saturated cultures of phage-infected cells on gradient plates (14) containing 0 to 90 μg/ml of paraquat. Strains having ≥90% of the resistance of the parental strain (i.e., of BW841 containing intact *soxRS* plasmids) were scored as paraquat resistant; paraquat-sensitive strains were two to three times more sensitive.

Other methods. General cloning techniques and propagation of M13 phages were done as described elsewhere (35). Computer software for sequence analysis (16, 32) was obtained from the Genetic Computer Group, University of Wisconsin. Promoter sequences and integration host factor binding sites were analyzed with the MacTargSearch program (19).

Nucleotide sequence accession number. The DNA sequence of the *soxRS* region (Fig. 1) has been submitted to GenBank under accession number M60111.

RESULTS

DNA sequence. The Δ*sox-8*::*cat* mutation has a 1.7-kb deletion that was originally generated on a plasmid and then transferred to the chromosome. The mutation causes noninducibility of the *sox* regulon and paraquat sensitivity (46). A 1.1-kb segment of DNA from this region was subcloned in M13 phages to provide templates for DNA sequencing by the dideoxynucleotide method. The recombinant phage plasmids fully complemented the phenotypic defects of the Δ*sox-8*::*cat* mutant.

Proteins encoded by *sox* plasmids were previously identified by radiolabeling (46), and the genes were located by deletion and insertion mutations. Mutations affecting the superoxide response spanned a region producing two polypeptides from genes that were divergently oriented. The SoxR protein (about 17 kDa) was identified as the product of a gene containing a *Sma*I site. The other protein, which we hereby designate SoxS, was about 13 kDa. The DNA sequence (Fig. 1) confirms these findings; the only intact large open reading frames specify proteins of 17.1 and 12.9 kDa and are preceded by potential ribosome binding sites. *soxR* and *soxS* have codon preference statistics ($P = 0.705$ and 0.798, respectively), consistent with bona fide genes that are weakly expressed (23). In further confirmation of these results, the *soxR4*::*cat* mutation, which consists of a Tn*9* segment inserted into *soxR* (46), is predicted, from our sequence and that of Tn*9* (1, 31), to produce a 20.5-kDa fusion protein; a 21-kDa protein was previously observed (46).

The restriction sites indicated by the sequence are consistent with previous work (46) that placed it at 92.2 min on the current linkage map (5) and at 4357 kb on the physical map (25) of *E. coli*. The sequence has one *Eco*RV site, which is in *soxS* (nucleotide 292), and one *Kpn*I site, which is in *soxR* (nucleotide 999). They correspond to adjacent sites on the physical map (25). As predicted, there was an absence of recognition sites for *Bam*HI, *Hin*dIII, *Eco*RI, *Pst*I, and *Pvu*II. In addition to an expected *Bgl*I site (nucleotide 1041), the sequence contains an unexpected one (nucleotide 812) that may have been missed during physical mapping because of its proximity to the former.

Transcription. The 5′ ends of the transcripts were mapped by primer extension (Fig. 2), and the results are recorded in Fig. 1. *soxS* and *soxR* mRNA each produced one major product when used as templates for reverse transcriptase. A minor band of material 6 nucleotides shorter than most of the *soxS* mRNA (Fig. 2, lane a) is of uncertain significance. It is unlikely to be a secondary start site because it is not properly distanced from any recognizable promoter sequence. In parallel lanes omitted from Fig. 2, no *soxS* or *soxR* mRNA was detected in a strain lacking these genes (i.e., a Δ*sox-8* mutant infected with M13 mp19). *soxR* mRNA has two unusual features: it starts 218 nucleotides upstream of *soxR*, and it extensively overlaps *soxS*, which is transcribed in the opposite direction (Fig. 1). The overlapping transcripts raise the possibility of coregulation.

Likely promoters were found near the transcriptional start sites (Fig. 1). The putative *soxR* and *soxS* promoters have similarity scores (28, 29) of 56 and 39%, respectively, placing

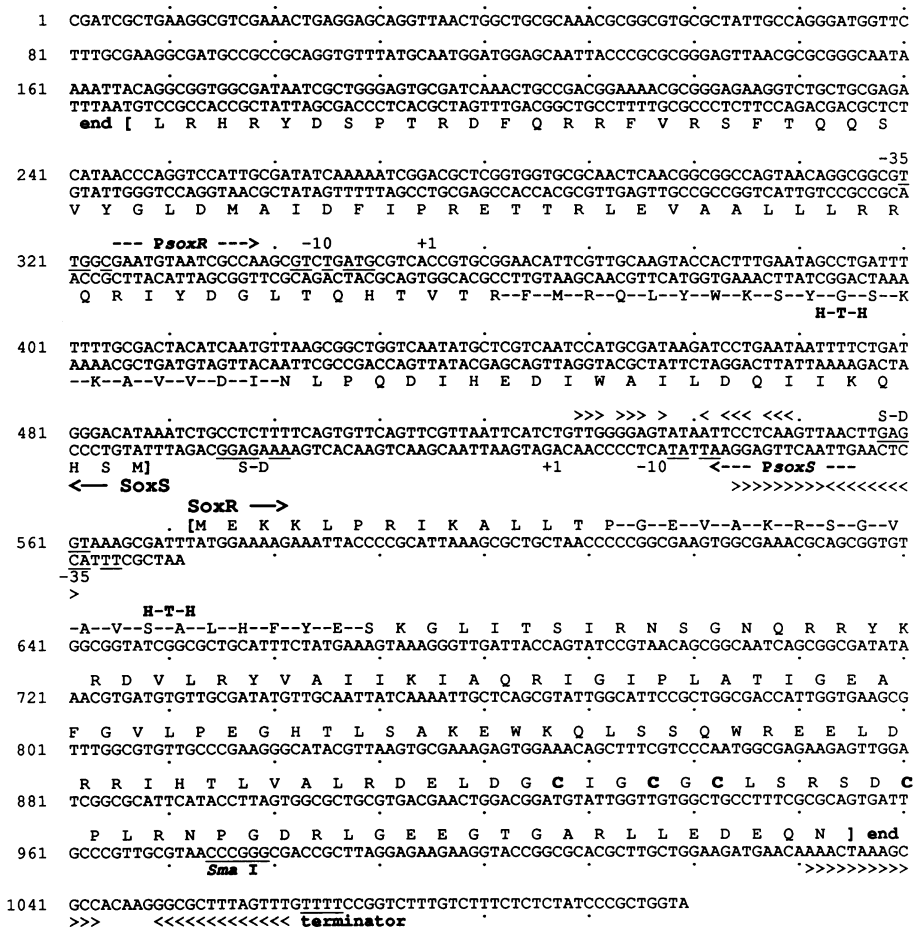

FIG. 1. DNA sequence and features of the *soxRS* region. Numbered lines are the 5′→3′ DNA sequence starting at a chromosomal *Pvu*I site and oriented clockwise with respect to the linkage map and physical map of *E. coli*. Also shown are the deduced protein sequences and the complement of nucleotides 161 to 571, the sense strand for *soxS*. Abbreviations: H-T-H, helix-turn-helix; S-D, Shine-Dalgarno sequence; +1, beginning of mRNA; P*soxR* and P*soxS*, promoter regions. Convergent arrowheads indicate regions of dyad symmetry. Underlined bases are those matching a consensus sequence.

them only at the 99% confidence level for σ^{70} promoters among random segments of a sequence with this base composition. Low scores are typical of weakly expressed or positively activated genes (see Discussion).

The *soxRS* region contains no sequences strongly homologous to known binding sites for integration host factor (19) or for catabolite activator protein (54). We could detect no similarities among the 5′-flanking regions of the sequenced paraquat-inducible genes, namely, *soxS*, *nfo* (36), or *sodA* (45). Therefore, if there is a *sox* box, i.e., a consensus regulatory sequence for genes of the *soxRS* regulon, it is unlikely to contain a nondegenerate contiguous sequence of sufficient length to be detected by the computer algorithm.

Paraquat produced an increase in the *soxS* message in a strain carrying a pBR322-derived *soxS* plasmid (Fig. 2, lanes a and f). In separate experiments, this effect was also seen in cells carrying M13::*soxRS* plasmids. Therefore, *soxS* is inducible. Between the genes are two overlapping segments displaying dyad symmetry. One is an 18-bp palindrome, and the other has a stem-and-loop motif that would generate a 9-bp mRNA stem containing two G · U base pairs. The dyads are potential binding sites for regulatory proteins, e.g., a possible transcriptional activator of *soxS* or a possible transcriptional or translational repressor of *soxR*.

At the end of *soxR* is a region of dyad symmetry with poly(T) sequences at one end and poly(A) sequences at the other. Computer analysis (10) indicated that this region should serve as a bidirectional, rho-independent, transcriptional terminator with scores better than those of 10^4 random sites. The long untranslated leader sequence of *soxR* mRNA might preclude rho-dependent transcription. No similar terminator was found for *soxS*, but there is no evidence that it is in an operon; the sequenced region downstream of that gene does not contain any open reading frame preceded by a possible ribosome binding site. It is likely, therefore, that the termination of *soxS* transcription is rho dependent.

General properties of SoxR and SoxS. The structure and properties of both proteins were deduced from the DNA sequence. SoxR (pI = 9.4) and SoxS (pI = 10.7) are both basic proteins. SoxR and SoxS contain α-helix–β-turn–α-helix motifs with scores of 1,466 and 1,497, respectively, on the scale of Dodd and Egan (17), corresponding to a 56% probability of being DNA-binding proteins. Neither appear to contain ATP-binding sites (20), flavin-binding sites (12), or transmembrane binding motifs (18).

SoxR contains four cysteines. The spacing of the four cysteines in *soxR*, CX_2CXCX_5C, is not similar to the spacing of those in iron-sulfur centers (12). It is shared only by

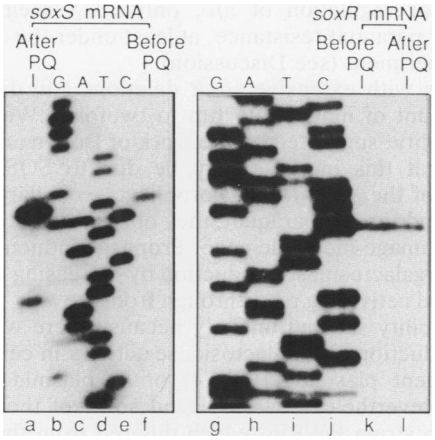

FIG. 2. Mapping of the 5′ ends of mRNAs. 5′-^{32}P-labeled primers that were complementary to *soxS* or *soxR*, as indicated, were hybridized to similar amounts of cellular RNA, extended with avian myeloblastosis virus reverse transcriptase, and separated by electrophoresis (lanes a, f, k, and l). DNA sequencing ladders (lanes b to e and g to j) were generated from the same primers on recombinant M13 DNA templates. RNA was isolated after incubation of cells with paraquat (PQ) (lanes a and k). Apart from unannealed probe DNA (not shown), the only bands seen in the mRNA lanes (lanes a, f, k, and l) were those shown. Results are recorded in Fig. 1.

proteins in the National Biomedical Research Foundation data base like laminin or metallothionines that have so many cysteines they are apt to have regions of random similarity. The cysteines might be involved in polypeptide folding, multimer formation, or even redox sensing. It may be significant that the cysteines are located near the *Sma*I site. Insertion or deletion mutations near this region produce a dominant constitutive overexpression of the *sox* regulon, which led to the suggestion that normal induction may be mediated through changes in secondary structure of the protein in this region (46).

SoxS is related to the AraC family of proteins. SoxS was compared with translated DNA sequences in the GenBank and EMBL data bases. The strongest match (49% homology) was with an open reading frame in Tn*10*, the position of which corresponded to that of *tetD* (8, 39), a gene of unknown function. The National Biomedical Research Foundation protein data base was scanned for sequences similar to SoxS and to a SoxS-TetD consensus sequence. Proteins that were found to be similar to SoxS throughout its length were all of bacterial origin (Fig. 3). Additional homologs (not listed) were proteins, from other bacteria, with functions identical to that of *E. coli* AraC (11). SoxR and SoxS also showed homology to the predicted products of adjacent open reading frames (ORFs) in Tn*917* of *Streptococcus faecalis* (40); SoxR had a 21% identity in a 116-amino-acid overlap with ORF5, and there was a 32% identity between a 50-amino-acid C-terminal segment of SoxS and a portion of ORF6.

From 17 to 49% of the amino acids in the aligned sequences (Fig. 3) were identical to ones at the same position in SoxS, and many others represented conservative substitutions by related amino acids. Although the sequences were selected by a computer algorithm on the basis of evolutionary relatedness of their amino acids, similarities persisted in a comparison based on chemical relatedness (Fig. 3). Uncertainty about the significance of homologies of less than 30% are dispelled when one notes the following: (i) only a few small gaps are needed for optimum alignment; (ii) helix-turn-helix motifs, when present, are also aligned, and (iii) there are seven positions at which an amino acid in SoxS is also found in at least eight of its nine homologs (Fig. 3, boldface type). Except for TetD, the other proteins in Fig. 3 are known to be regulatory (see Discussion).

Both genes are needed for induction. The Δ*sox-8*::*cat* mutation (46) consists of the substitution of a segment of Tn*9*

FIG. 3. Homology between SoxS and other bacterial proteins. The entire protein sequence of SoxS is aligned with segments of other bacterial proteins shown in Table 1. Symbols: |, an amino acid identical to one in SoxS at that position; :, a functionally similar amino acid (groups are MILV, YWF, QNED, MILV, PAGST, and HKR); -, a gap of one amino acid introduced for alignment. Underlined segments are helix-turn-helix motifs with a >55% probability of being DNA-binding domains (17). The first and last amino acids in each aligned segment are numbered, and length is shown as the number of amino acids in the whole polypeptide. Percent homology is that between the aligned segments only.

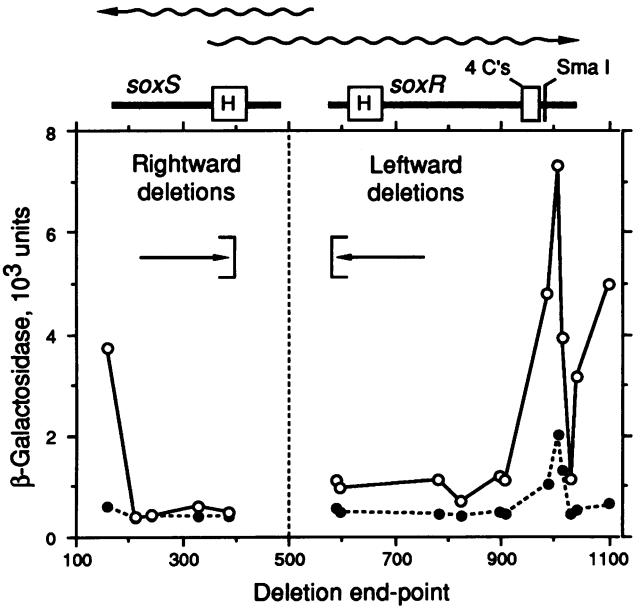

FIG. 4. Effect of partial deletions of *soxS* and *soxR* on the constitutive expression and inducibility of an *nfo'-lac* fusion. The 1,099-bp *soxRS* region (Fig. 1) cloned in M13 phages was partially deleted rightward from nucleotide 1 or leftward from nucleotide 1099. Strain BW841 {Δ*soxRS* [λ Φ(*nfo'-lac*)]} was infected with each phage clone and assayed without paraquat (●) or after paraquat treatment (○). Each data point represents the average of two to five experiments. At the top of the figure is a map of the *soxRS* region in which the transcripts (wavy arrows), *soxS* and *soxR* genes (solid bars), regions encoding helix-turn-helix motifs (H), and region encoding 4 cysteines (4 C's) are shown.

for a region containing the *soxR* and *soxS* genes. The deletion extends from an *Hpa*I site at nucleotide 35 of Fig. 1 to one about 250 nucleotides beyond the end of *soxR*. A Δ*sox-8*::*cat* mutant has an increased sensitivity to killing by paraquat, and its *soxRS* regulon is not induced by paraquat (46). These defects were reversed in *trans* by M13 plasmids containing the 1.1-kb *soxRS* region shown in Fig. 1. We then tested a series of plasmid deletion mutants in which the *soxS* or *soxR* gene was truncated (Fig. 4). The only plasmids that fully complemented the Δ*sox-8*::*cat* chromosome were those whose deletions did not encroach on *soxR* or *soxS*. One such deletion terminated 9 nucleotides from the end of *soxS*, and another deletion terminated 6 nucleotides from the end of *soxR*. Truncation of *soxS* destroyed paraquat inducibility, even when the deletions fell far short of the *soxR* promoter region.

Removal of 21, 28, or 47 nucleotides from the end of *soxR* led to a two- to fourfold constitutive overexpression of *nfo'-lacZ* and to its hyperinducibility (Fig. 4). These results were consistent with previous findings (46) that insertions in or around the *Sma*I site or deletions to the right of it resulted in the constitutive phenotype. The constitutive deletion mutations (Fig. 4) did not extend into the region encoding the cysteines in SoxR. More-extensive deletion of *soxR* led to relative noninducibility, thus providing the first evidence that *soxR* is actually required for induction.

The phage-infected strains depicted in Fig. 4 were also tested for sensitivity to killing by paraquat. All of the *soxS* mutants, but none of the *soxR* mutants, had an increased sensitivity. Therefore, although both *soxR* and *soxS* were required for regulation of *nfo*, only *soxS* seemed to be needed for paraquat resistance, at least under the conditions of this experiment (see Discussion).

Plasmids with extensive *soxR* deletions still displayed a small amount of inducibility (up to twofold). We explored the possibility, suggested by the work of Brawn and Fridovich (9), that this increase may be due to SOS-mediated induction of the λ Φ(*nfo'-lac*) prophage secondary to DNA damage produced by paraquat; they observed delayed induction of a damage-inducible gene. Prophage induction should increase β-galactosidase production by increasing gene copy number and activating read-through from the λ p_L promoter. This possibility seemed unlikely because there was no significant induction of β-galactosidase activity in cells bearing *soxS*-deficient plasmids (Fig. 4) or in plasmid-free cells (≤20%). Nevertheless, we repeated some of these experiments with strain BW856, which differed from that used in Fig. 4 only in that it contained a λ Φ(*nfo'-lac*) prophage whose replication was noninducible. The results were similar. We still found a small (30 to 90%) increase in β-galactosidase activity after paraquat treatment of cells whose plasmids had extensive *soxR* deletions and a sixfold increase in those with an intact plasmid. Therefore, under the conditions of our experiments, paraquat did not appear to evoke a significant SOS response.

***sox*-3::Tn*10* is a *soxS* mutation.** The *sox*-3::Tn*10* mutation was produced by random chromosomal insertion of a mini-Tn*10* transposon. It was found to be cotransducible with *soxR* and Δ*sox-8*::*cat* mutations, to specify noninducibility of the regulon, and to result in an increased sensitivity to superoxide generators (46). These properties are consistent with the properties we now associate with a *soxS* mutant. In experiments similar to those in Fig. 4, the noninducibility and paraquat sensitivity of strain BW840 (*sox-3*::Tn*10*) were fully complemented by $soxS^+$ plasmids containing extensive *soxR* deletions, but not by ones containing *soxS* deletions (data not shown). Therefore, *sox-3*::Tn*10* (*soxS3*::Tn*10*) is an allele of *soxS*, and its previously described traits (46) must be interpreted in this new light (see Discussion).

DISCUSSION

The DNA sequence (Fig. 1) confirms earlier work on the proteins specified by intact and mutant *sox* plasmids (46) that indicated that the *sox* region contains divergently transcribed genes encoding a 17-kDa SoxR protein and a 13-kDa protein that is now designated as SoxS. Although some mutations affecting *soxR* complementation also affected the SoxS protein (46), because of the orientation of the genes it was possible that the *soxR* promoter lay within *soxS*, an arrangement confirmed by the DNA sequence (Fig. 1). Therefore, there was no previous direct evidence that *soxS* was involved in the superoxide response. This study provides such evidence in the form of the following observations. (i) The *soxS3*::Tn*10* mutation, which specified paraquat sensitivity and paraquat noninducibility, was complemented by $soxS^+$ plasmids containing extensive *soxR* deletions. (ii) *soxS* mutants displayed a phenotype (paraquat sensitivity) not found in *soxR* mutants. (iii) Partial deletions of *soxS* that were 78 or 108 bp distant from the *soxR* promoter resulted in noninducibility. (iv) *soxS* mRNA increased after paraquat induction. (v) SoxS is partly homologous to known regulatory proteins.

In this study, we used paraquat resistance and the paraquat-mediated induction of an *nfo'-lac* fusion as indicators of regulon function and showed that *soxS* was required for

TABLE 1. Proteins similar to SoxS

Protein	Source	Regulatory target	Reference(s)
TetD	Tn*10*	Unknown	8, 39
RhaS	*E. coli*	Rhamnose operon	44
RhaR	*E. coli*	Rhamnose operon	44
VirF	*Yersinia* plasmid	Outer membrane protein regulon	13
MelR	*E. coli*	Melibiose operon	50
XylS	*Pseudomonas* plasmid	Plasmid xylene regulon	24, 42
AraC	*E. coli*	Arabinose regulon	43
EnvY	*E. coli*	Porin genes	26
AppY	*E. coli*	Growth-phase-dependent genes	2

both. However, *soxS* must also be required for resistance to other superoxide generators and for the induction of other genes of the *soxR* regulon because all of these traits are affected by the *sox-3*::Tn*10* mutation (46), which we have just found to be an allele of *soxS*.

Although *soxR* was needed for regulon induction, it did not appear to be needed for paraquat resistance, suggesting that *soxS* may have a function not dependent on *soxR*. However, this curious result may have merely been an artifact of our experimental procedure. Suppose that in wild-type *E. coli*, *soxR* expression is essential for *soxS* induction, which in turn is needed for paraquat resistance. This requirement for *soxS* induction may have been eliminated in our complementation tests because the gene was on a multicopy plasmid. To verify this hypothesis, we must have a mutation that does not currently exist, namely, a chromosomal *soxR* mutation specifying the noninducible phenotype. It should be noted that the basal (uninduced) level of *nfo* expression is independent of the number of copies of *soxS* in the cell (46; also Fig. 4). Therefore, either an increase in SoxS alone is not sufficient to induce *nfo* or constitutive production of SoxS is independent of copy number (via negative regulation). Thus, at present, we cannot dismiss the possibility that the only direct function of *soxR* might be to control *soxS*.

It was previously concluded that the *soxRS* region must exert a net positive control over the regulon because deletions rendered the regulon noninducible (22, 46). Moreover, its control is probably exerted at the level of transcription because the *nfo'-lac* fusion used in some of these studies (46) was an operon fusion in which the *lacZ* gene was preceded by its own ribosome binding site. Consistent with these assumptions are the properties of proteins to which SoxS is similar (Table 1). Except for TetD, the proteins are known to serve as positive regulators; they activate other genes in *trans*. AraC (37), RhaR (44), and MelR (49) have been shown to bind near promoter regions. Most of the other proteins had been previously found to have significant homology with AraC, one of the best-studied transcriptional activators.

What are the individual roles of SoxR and SoxS? Transcriptional regulators may fulfill one or both of two tasks. They must interact with the inducer (sensor function) and affect transcription by binding to DNA and possibly to RNA polymerase (regulator function). In two-component regulatory systems, one protein may be the sensor and the other may be the regulator (34). As reviewed by Ramos et al. (33), both functions may be combined in individual proteins of the AraC family. The similar C-terminal regions of the AraC-like proteins, which contain their putative DNA-binding (helix-turn-helix) domains, probably mediate their similar regulatory functions. It is this portion to which SoxS is similar over almost its entire length; therefore, SoxS may be a regulator. Their more variable N-terminal regions may be primarily responsible for recognizing their respective inducers; it is this portion of AraC and XylS in which mutations occur that alter the affinity or specificity for effectors (33). This region has no counterpart in the much shorter SoxS protein. Its function may therefore be fulfilled by SoxR, a protein having four closely spaced cysteines that might be involved in conformational changes accompanying induction and that might even be a redox center that detects superoxide. Therefore, we speculate that SoxS is a regulator and SoxR is a sensor, but their roles may not be so simple. SoxR contains a helix-turn-helix motif that is as good a putative DNA-binding domain as that in SoxS. Thus, the two proteins may interact while bound to DNA.

In two-component regulatory systems, the sensor protein is usually synthesized at much lower levels than that of the regulator (34). In maxicell preparations, less SoxR was produced than SoxS (46), a probable result of a relatively weaker match between the *soxR* promoter and the σ^{70} promoter consensus sequence and the higher frequency of rare codons in *soxR*. Unfortunately, we cannot estimate the relative amounts of *soxS* versus *soxR* mRNA from the data in Fig. 2 because different protocols were used for each. (Different procedures were required to optimize each primer extension past the secondary structures formed by the symmetrical sequences in the intergenic region.) The *soxS* promoter, though predicted to be better than that of *soxR*, is still outmatched by about 0.5% of random sequences. However, *soxS* appears to be inducible, and low similarity to a promoter site consensus is characteristic of positively activated promoters (28).

The face-to-face arrangement of the *soxR* and *soxS* promoters provides opportunity for coordinate and even reciprocal regulation (for a review, see reference 6). Convergent transcription may inhibit the expression of the gene with the weaker promoter (38, 48), perhaps via steric interference between RNA polymerase molecules moving in opposite directions. Therefore, as *soxS* transcription is induced, we may expect the transcription of *soxR* to decrease. However, in repetitions of the experiment in Fig. 2, we have not seen a consistent significant change. Translational control is an alternative possibility. Annealing of the complementary portions of the two mRNAs could reduce gene expression, and the long leader sequence in *soxR* mRNA may have yet another function in translation. In unpublished preliminary studies, we find that paraquat treatment decreases the expression of a SoxR-LacZ fusion protein while increasing that of a SoxS-LacZ fusion. Reduced synthesis of nascent SoxR protein would keep it from competing with previously activated molecules. That such competition can indeed occur is suggested by an earlier finding that a plasmid *soxR*$^+$ gene can partially suppress constitutive overexpression of *nfo* in mutants producing altered forms of SoxR (46).

ACKNOWLEDGMENTS

We gratefully acknowledge the capable technical assistance of Fred Kung.

This study was supported by a research grant (DMB-8922562) from the National Science Foundation and in part by an NIH grant (MO1 RR 00042).

REFERENCES

1. **Alton, N. K., and D. Vapnek.** 1979. Nucleotide sequence analysis of the chloramphenicol resistance transposon Tn*9*. Nature (London) **282**:864–869.
2. **Atlung, T., A. Nielsen, and F. G. Hansen.** 1989. Isolation, characterization, and nucleotide sequence of *appY*, a regulatory gene for growth-phase-dependent gene expression in *Escherichia coli*. J. Bacteriol. **171**:1683–1691.
3. **Ausubel, F. M., R. Brent, R. E. Kingston, D. D. Moore, J. G. Seidman, J. A. Smith, and K. Struhl (ed.).** 1989. Current protocols in molecular biology. John Wiley & Sons, Inc., New York.
4. **Bachmann, B. J.** 1972. Pedigrees of some mutant strains of *Escherichia coli* K-12. Bacteriol. Rev. **36**:525–557.
5. **Bachmann, B. J.** 1990. Linkage map of *Escherichia coli* K-12, edition 8. Microbiol. Rev. **54**:130–197.
6. **Beck, C. F., and R. A. Warren.** 1988. Divergent promoters, a common form of gene organization. Microbiol. Rev. **52**:318–326.
7. **Belfort, M., J. Pedersen-Lane, D. West, K. Ehrenman, G. Maley, F. Chu, and F. Maley.** 1985. Processing of the intron-containing thymidylate synthase (*td*) gene of phage T4 is at the RNA level. Cell **41**:375–382.
8. **Braus, G., M. Argast, and C. F. Beck.** 1984. Identification of additional genes on transposon Tn*10*: *tetC* and *tetD*. J. Bacteriol. **160**:504–509.
9. **Brawn, M. K., and I. Fridovich.** 1985. Increased superoxide radical production evokes inducible DNA repair in *Escherichia coli*. J. Biol. Chem. **260**:922–925.
10. **Brendel, V., G. H. Hamm, and E. N. Trifonov.** 1986. Terminators of transcription with RNA polymerase from *Escherichia coli*: what they look like and how to find them. J. Biomol. Struct. Dyn. **3**:705–723.
11. **Burke, K. A., and G. Wilcox.** 1987. The *araC* gene of *Citrobacter freundii*. Gene **61**:243–252.
12. **Cole, S. T., K. Eiglmeier, S. Ahmed, N. Honore, L. Elmes, W. F. Anderson, and J. H. Weiner.** 1988. Nucleotide sequence and gene-polypeptide relationships of the *glpABC* operon encoding the anaerobic *sn*-glycerol-3-phosphate dehydrogenase of *Escherichia coli* K-12. J. Bacteriol. **170**:2448–2456.
13. **Cornelis, G., C. Sluiters, C. L. de Rouvroit, and T. Michiels.** 1989. Homology between VirF, the transcriptional activator of the *Yersinia* virulence regulon, and AraC, the *Escherichia coli* arabinose operon regulator. J. Bacteriol. **171**:254–262.
14. **Cunningham, R. P., S. M. Saporito, S. G. Spitzer, and B. Weiss.** 1986. Endonuclease IV (*nfo*) mutant of *Escherichia coli*. J. Bacteriol. **168**:1120–1127.
15. **Dale, R. M. K., B. A. McClure, and J. P. Houchins.** 1985. A rapid single-stranded cloning strategy for producing a sequential series of overlapping clones for use in DNA sequencing: application to sequencing the corn mitochondrial 18 S rDNA. Plasmid **13**:31–40.
16. **Devereux, J., P. Haeberli, and O. Smithies.** 1984. A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. **12**:387–395.
17. **Dodd, I. B., and J. B. Egan.** Systematic method for the detection of potential λ Cro-like DNA-binding regions in proteins. J. Mol. Biol. **194**:557–564.
18. **Engelman, D. M., T. A. Steitz, and A. Goldman.** 1986. Identifying nonpolar transbilayer helices in amino acid sequences of membrane proteins. Annu. Rev. Biophys. Biophys. Chem. **15**:321–353.
19. **Goodrich, J. A., M. L. Schwartz, and W. R. McClure.** 1990. Searching for and predicting the activity of sites for DNA binding proteins: compilation and analysis of the binding sites for *Escherichia coli* integration host factor (IHF). Nucleic Acids Res. **18**:4493–5000.
20. **Gottesman, S., C. Squires, E. Pichersky, M. Carrington, M. Hobbs, J. S. Mattick, B. Dalrymple, H. Kuramitsu, T. Shiroza, T. Foster, W. P. Clark, B. Ross, C. L. Squires, and M. R. Maurizi.** 1990. Conservation of the regulatory subunit for the Clp ATP-dependent protease in prokaryotes and eukaryotes. Proc. Natl. Acad. Sci. USA **87**:3513–3517.
21. **Greenberg, J. T., and B. Demple.** 1989. A global response induced in *Escherichia coli* by redox-cycling agents overlaps with that induced by peroxide stress. J. Bacteriol. **171**:3933–3939.
22. **Greenberg, J. T., P. Monach, J. H. Chou, P. D. Josephy, and B. Demple.** 1990. Positive control of a global antioxidant defense regulon activated by superoxide-generating agents in *Escherichia coli*. Proc. Natl. Acad. Sci. USA **87**:6181–6185.
23. **Gribskov, M., J. Devereux, and R. R. Burgess.** 1984. The codon preference plot: graphic analysis of protein coding sequences and prediction of gene expression. Nucleic Acids Res. **12**:539–549.
24. **Inouye, S., A. Nakazawa, and T. Nakazawa.** 1986. Nucleotide sequence of the regulatory gene *xylS* on the *Pseudomonas putida* TOL plasmid and identification of the protein product. Gene **44**:235–242.
25. **Kohara, Y., K. Akiyama, and K. Isono.** 1987. The physical map of the whole *E. coli* chromosome: application of a new strategy for rapid analysis and sorting of a large genomic library. Cell **50**:495–508.
26. **Lundrigan, M. D., M. J. Friedrich, and R. J. Kadner.** 1989. Nucleotide sequence of the Escherichia coli porin thermoregulatory gene envY. Nucleic Acids. Res. **17**:800.
27. **Miller, J. H.** 1972. Experiments in molecular genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
28. **Mulligan, M. E., D. K. Hawley, R. Entriken, and W. R. McClure.** 1984. *Escherichia coli* promoter sequences predict in vitro RNA polymerase selectivity. Nucleic Acids Res. **12**:789–800.
29. **Mulligan, M. E., and W. R. McClure.** 1986. Analysis of the occurrence of promoter-sites in DNA. Nucleic Acids Res. **14**:109–126.
30. **Neidhardt, F. C.** 1987. Multigene systems and regulons, p. 1313–1317. *In* F. C. Neidhardt, J. L. Ingraham, K. B. Low, B. Magasanik, M. Schaechter, and H. E. Umbarger (ed.), *Escherichia coli* and *Salmonella typhimurium*: cellular and molecular biology. American Society for Microbiology, Washington, D.C.
31. **Ohtsubo, H., and E. Ohtsubo.** 1978. Nucleotide sequence of an insertion element, IS*1*. Proc. Natl. Acad. Sci. USA **75**:615–619.
32. **Pearson, W. R., and D. J. Lipman.** 1988. Improved tools for biological sequence comparison. Proc. Natl. Acad. Sci. USA **85**:2444–2448.
33. **Ramos, J. L., F. Rojo, L. Zhou, and K. N. Timmis.** 1990. A family of positive regulators related to the *Pseudomonas putida* TOL plasmid XylS and the *Escherichia coli* AraC activators. Nucleic Acids Res. **18**:2149–2152.
34. **Ronson, C. W., B. T. Nixon, and F. M. Ausubel.** 1987. Conserved domains in bacterial regulatory proteins that respond to environmental stimuli. Cell **49**:579–581.
35. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
36. **Saporito, S. M., and R. P. Cunningham.** 1988. Nucleotide sequence of the *nfo* gene of *Escherichia coli* K-12. J. Bacteriol. **170**:5141–5145.
37. **Schleif, R.** 1987. The L-arabinose operon, p. 1473–1481. *In* F. C. Neidhardt, J. L. Ingraham, K. B. Low, B. Magasanik, M. Schaechter, and H. E. Umbarger (ed.), *Escherichia coli* and *Salmonella typhimurium*: cellular and molecular biology. American Society for Microbiology, Washington, D.C.
38. **Schmeissner, U., D. Court, H. Shimatake, and M. Rosenberg.** 1980. Promoter for the establishment of repressor synthesis in bacteriophage λ. Proc. Natl. Acad. Sci. USA **77**:3191–3195.
39. **Schollmeier, K., and W. Hillen.** 1984. Transposon Tn*10* contains two structural genes with opposite polarity between *tetA* and IS*10*$_R$. J. Bacteriol. **160**:499–503.
40. **Shaw, J. H., and D. B. Clewell.** 1985. Complete nucleotide sequence of macrolide-lincosamide-streptogramin B-resistance transposon Tn*917* in *Streptococcus faecalis*. J. Bacteriol. **164**: 782–796.
41. **Silhavy, T. J., M. L. Berman, and L. W. Enquist.** 1984. Experiments with gene fusions. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

42. **Spooner, R. A., K. Lindsay, and F. C. Franklin.** 1986. Genetic, functional and sequence analysis of the *xylR* and *xylS* regulatory genes of the TOL plasmid pWWO. J. Gen. Microbiol. **132**:1347–1358.
43. **Stoner, C. M., and R. Schleif.** 1982. Is the amino acid but not the nucleotide sequence of the *Escherichia coli araC* gene conserved? J. Mol. Biol. **154**:649–652.
44. **Tobin, J. F., and R. F. Schleif.** 1987. Positive regulation of the *Escherichia coli* L-rhamnose operon is mediated by the products of tandemly repeated regulatory genes. J. Mol. Biol. **196**:789–799.
45. **Touati, D.** 1988. Transcriptional and posttranscriptional regulation of manganese superoxide dismutase biosynthesis in *Escherichia coli*, studied with operon and protein fusions. J. Bacteriol. **170**:2511–2520.
46. **Tsaneva, I. R., and B. Weiss.** 1990. *soxR*, a locus governing a superoxide response regulon in *Escherichia coli* K-12. J. Bacteriol. **172**:4197–4205.
47. **Walkup, L. K. B., and T. Kogoma.** 1989. *Escherichia coli* proteins inducible by oxidative stress mediated by the superoxide radical. J. Bacteriol. **171**:1476–1484.
48. **Ward, D. F., and N. E. Murray.** 1979. Convergent transcription in bacteriophage λ: interference with gene expression. J. Mol. Biol. **133**:249–266.
49. **Webster, C., L. Gardner, and S. Busby.** 1989. The *Escherichia coli melR* gene encodes a DNA-binding protein with affinity for specific sequences located in the melibiose-operon regulatory region. Gene **83**:207–213.
50. **Webster, C., K. Kempsell, I. Booth, and S. Busby.** 1987. Organisation of the regulatory region of the *Escherichia coli* melibiose operon. Gene **59**:253–263.
51. **Weiss, B., I. R. Tsaneva, and J. Wu.** 1990. A superoxide response regulon in *Escherichia coli*. UCLA Symp. Mol. Cell. Biol. New Ser. **136**:287–295.
52. **White, B. J., S. J. Hochhauser, N. M. Cintrón, and B. Weiss.** 1976. Genetic mapping of *xthA*, the structural gene for exonuclease III in *Escherichia coli* K-12. J. Bacteriol. **126**:1082–1088.
53. **Yanisch-Perron, C., J. Vieira, and J. Messing.** 1985. Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. Gene **33**:103–119.
54. **Zhang, X., and R. H. Ebright.** 1990. Identification of a contact between arginine-180 of the catabolite gene activator protein (CAP) and base pair 5 of the DNA site in the CAP-DNA complex. Proc. Natl. Acad. Sci. USA **87**:4717–4721.