Biochemical and Biophysical Research Communications **291,** 979–986 (2002)
doi:10.1006/bbrc.2002.6559, available online at http://www.idealibrary.com on IDEAL®

Evidence for "Pre-recruitment" as a New Mechanism of Transcription Activation in *Escherichia coli:* The Large Excess of SoxS Binding Sites per Cell Relative to the Number of SoxS Molecules per Cell

Kevin L. Griffith,[1] Ishita M. Shah,[1] Todd E. Myers, Michael C. O'Neill, and Richard E. Wolf, Jr.[2]
Department of Biological Sciences, University of Maryland Baltimore County, Baltimore, Maryland 21250

Received January 30, 2002

In response to the oxidative stress imposed by redox-cycling compounds like paraquat, *Escherichia coli* induces the synthesis of SoxS, which then activates the transcription of ~100 genes. The DNA binding site for SoxS-dependent transcription activation, the "soxbox," is highly degenerate, suggesting that the genome contains a large number of SoxS binding sites. To estimate the number of soxboxes in the cell, we searched the *E. coli* genome for SoxS binding sites using as query sequence the previously determined optimal SoxS binding sequence. We found ~12,500 sequences that match the optimal binding sequence under the conditions of our search; this agrees with our previous estimate, based on information theory, that a random sequence the size of the *E. coli* genome contains ~13,000 soxboxes. Thus, fast-growing cells with 4–6 genomes per cell have ~65,000 soxboxes. This large number of potential SoxS binding sites per cell raises the interesting question of how SoxS distinguishes between the functional soxboxes located within the promoters of target genes and the plethora of equivalent but nonfunctional binding sites scattered throughout the chromosome. To address this question, we treated cells with paraquat and used Western blot analysis to determine the kinetics of SoxS accumulation per cell; we also determined the kinetics of SoxS-activated gene expression. The abundance of SoxS reached a maximum of 2,500 molecules per cell 20 min after induction and gradually declined to ~500 molecules per cell over the next 1.5 h. Given that activation of target gene expression began almost immediately and given the large disparity between the number of SoxS molecules per cell, 2,500, and the number of SoxS binding sites per cell, 65,000, we infer that SoxS is not likely to activate transcription by the usual "recruitment" pathway, as this mechanism would require a number of SoxS molecules similar to the number of soxboxes. Instead, we propose that SoxS first interacts in solution with RNA polymerase and then the binary complex scans the chromosome for promoters that contain a soxbox properly positioned and oriented for transcription activation. We name this new pathway "pre-recruitment."

Key Words: **oxidative stress; "pre-recruitment;" transcription activation; Western blot.**

SoxS is the direct transcription activator of the *Escherichia coli* superoxide regulon, also known as the SoxRS regulon (1–4). Upon sensing a change in the cellular redox potential, induced, for example, by treatment of cells with the redox-cycling compound paraquat, SoxR undergoes a conformational change and activates transcription of the *soxS* gene (5, 6). Newly synthesized SoxS then binds to the "soxbox" DNA sequence in the promoter regions of its target genes and activates their transcription (2, 7, 8). Recent DNA macroarray analyses have indicated that SoxS may activate transcription of as many as 100 genes, though only about 20 are known to play a direct role in the defense against oxidative stress (9).

SoxS is a member of the AraC/XylS family of bacterial transcription activators (10, 11). SoxS is particularly closely related (~45% amino acid sequence identity) to two other members of the protein family, MarA and Rob. Induction of MarA synthesis and artificial hyper-expression of Rob are known to regulate transcription of gene sets that are highly coincident with that regulated by SoxS, although each protein regulates many of the respective target genes to different degrees (12–15). MarA resides in the *marRAB* operon, whose transcription is repressed by MarR and induced by aromatic weak acids like salicylate; *de novo* synthesis of MarA, just like *de novo* synthesis of SoxS, acti-

[1] These authors contributed equally to this work.

[2] To whom correspondence and reprint requests should be addressed at Department of Biological Sciences, University of Maryland Baltimore County, 1000 Hilltop Circle, Baltimore, Maryland 21250. Fax: (410) 455-3875. E-mail: wolf@umbc.edu.

0006-291X/02 $35.00

vates the transcription of the genes of the *mar* regulon, whose expression confers resistance to a variety of structurally unrelated antibiotics (16, 17). Unlike SoxS and MarA, Rob is expressed constitutively at about 5,000–10,000 molecules per cell (18, 19). Though a *rob*::*kan* null mutation does not confer an obvious growth phenotype, it does reduce by fivefold the basal expression of *micF*, a gene whose transcription is known to be activated by SoxS and MarA (13, 19). Moreover, artificial over-expression of Rob from a plasmid activates transcription of many of the same genes as SoxS and MarA, although with a different expression profile (14). Thus, in *E. coli*, three proteins activate transcription of a large and common set of genes now known as the *sox*/*mar*/*rob* regulon; induction of the regulon by any of the three proteins confers resistance to multiple antibiotics, heavy metals, and redox-cycling compounds as well as tolerance to organic solvents (6, 16, 20).

Abundant genetic and biochemical evidence has demonstrated unequivocally that SoxS, MarA, and Rob bind as monomers to the same 20 bp DNA sequences in the promoters of the genes they regulate, though they do so with different affinities (12, 13, 21–25). This is not surprising, since they activate the same set of genes, but to different degrees, and have a high degree of amino acid sequence conservation, particularly within their dual helix-turn-helix DNA binding motifs (11, 26, 27). However, an interesting and unusual property of these three proteins is that there is a rather large amount of sequence divergence among the various DNA sites they bind. From a set of 14 biologically active sequences known to be used for DNA binding and transcription activation *in vivo*, Martin *et al.* deduced the consensus sequence AYnGCACnnWnn-RYYAAAYn (Y = C, T; n = A, C, G, T; W = A, T; R = A, G) (12). In agreement, a systematic mutagenesis of the *zwf* and *fpr* soxboxes by Griffith and Wolf (28, 29) showed that the optimal binding sequence for SoxS contains four DNA binding determinants, the "invariant A" at the first position, Recognition Element 1 with sequence GCAC, Recognition Element 2 with sequence CAAA and an A/T-rich spacer; moreover, this work produced the optimal sequence for SoxS binding, An-VGCACWWWnKRHCAAAHn (V = A, C, G; K = G, T; H = A, C, T), a sequence consistent with the less defined consensus of Martin *et al.* Lastly, information theory applied by Wood *et al.* (30) to a collection of 16 sequences known to bind SoxS *in vitro* provided a sequence logo depicting the relative frequency of each base at each position in the set of soxboxes; the conserved residues within the logo are consistent with the degeneracy of the consensus sequence of Martin *et al.* (12) and with the DNA binding determinants of Griffith and Wolf (28, 29). More importantly, the information theory approach produced a measure of the average sequence conservation among the SoxS binding sites: the average binding site contains but 9.5 bits of information (30). Thus, in random sequence, one soxbox would be expected every 724 bases ($1/\{2^{9.5}\}$) on a given DNA strand. This means that a random sequence the size of the *E. coli* genome (4.6×10^6 bp $\times$ 2) would be expected to have ~12,700 SoxS binding sites! In turn, this analysis raises the intriguing question of how SoxS is able to bind to and activate transcription of its ~100 target genes when these binding sites represent only about 1% of the total SoxS binding sites in the genome. The same question pertains to transcription activation by MarA and Rob.

"Recruitment" is commonly accepted as the main mechanism of transcription activation in bacteria (31). In the typical pathway of recruitment, the activator first binds its DNA target. Then, through protein-protein interactions with the DNA-bound activator, RNA polymerase (RNAP) is recruited to the promoter and a stable, open transcription initiation complex is formed As exemplified by catabolite gene activator protein (CAP, also known as cyclic AMP receptor protein, CRP), and lambda repressor, the activator can recruit RNAP by enhancing the binding of RNAP to the promoter, i.e., closed complex formation, and/or by stimulating the rate of conversion of the unstable RNAP-DNA closed complex to the stable open complex (31). Accordingly, for recruitment to be the mechanism by which SoxS activates transcription of its target genes, it would appear that the cell would need to produce a very large number of SoxS molecules (or the sum of SoxS, MarA, and Rob), enough to ensure the binding of SoxS to the functional sites in the promoters of its target genes in the presence of the vast excess of equivalent but nonfunctional sites scattered throughout the chromosome. Given that fast-growing *E. coli* cells have 4–6 genomes (32), then such fast-growing cells would contain ~65,000 SoxS binding sites (13,000 sites/genome $\times$ 5 genomes) and transcription activation by recruitment would require a similarly large number of SoxS molecules.

Here, we determine by computer analysis that the *E. coli* genome contains a number of SoxS binding sites similar to that predicted by information theory for a random sequence and we use immunoblot analysis to determine the number of SoxS molecules in *E. coli* cells growing in broth cultures. Finding that SoxS-dependent gene expression begins almost immediately after induction and finding a maximum of only 2,500 SoxS molecules per cell, we infer that recruitment is not likely to be the mechanism of transcription activation employed by SoxS. We propose instead that activation may proceed by a new mechanism, which we refer to as "pre-recruitment."

MATERIALS AND METHODS

Computer analysis. We used computer program Alignall to search both strands of the *E. coli* genomic sequence (33) using the

optimal SoxS binding sequence AnVGCACWWWnKRHCAAAHn. (n = A, C, G, T; V = A, C, G; W = A, T; K = G, T; R = A, G; H = A, C, T) as the query sequence (28, 29) and allowing mismatches at 3 positions and 4 positions. The results were tallied with an editor. The program can be found at the following URL: http://www.research.umbc.edu/~moneill/software.

Western blots and determination of the number of SoxS molecules per cell. Rabbits (Pocono Rabbit Farm) were chosen for immunization with purified his_6-SoxS (28) and preparation of anti-SoxS serum after first screening pre-immune sera for low titers of antibodies directed against low MW *E. coli* proteins. The antiserum was collected after six immunizations spaced 1 month apart. For the determination of SoxS abundance, overnight cultures of strain GC4468 (Δ*lacU169 rpsL*) (34) grown in LB medium at 37°C were diluted 1:100 into 110 ml of fresh LB medium and incubated at 37°C until the culture density reached A_{600} ~0.1 at which point SoxS synthesis was induced by addition of 0.5 mM paraquat (methyl viologen; Sigma). At each time point, 11 ml samples were taken and placed on ice. One ml of each sample was removed and used for cell count determination as described below. The cells from the remaining 10 ml were immediately collected by centrifugation. The cell pellets were resuspended in 0.5 ml of sonication buffer (50 mM Tris–HCl, pH 7.9, 3 mM DTT, 1 mM EDTA, and 6 M urea) and subjected to sonication with a Branson sonifier for two pulses of 1 min. The insoluble cell material was removed by centrifugation at 13,000*g* for 30 min. The entire 500 μl sonic extract from each sample was added to a microfuge tube containing 250 μl of Laemmli gel loading buffer. Then, 10 μl of each sample, containing the proteins from $1–5 \times 10^7$ cells, was loaded into a well of an 18% Tris-glycine gel (Invitrogen) and subjected to SDS–PAGE. Proteins were transferred by electroblotting (Invitrogen) to polyvinylidene difluoride membrane and Western blot analysis was carried out with the ECF chemifluorescent substrate under the conditions described by the manufacturer (Amersham). The signals were detected and quantified with a Storm PhosphorImager (Molecular Dynamics) using ImageQuant software. Each gel contained 0.3, 1.0, and 3.0 ng of purified his_6-SoxS. The signals of purified protein were directly proportional to the input concentration ($r^2 = 0.99$) and were used to generate a standard curve from which the amount of SoxS in each lane of the gel could be accurately determined. From these values, the number of SoxS molecules per ml of culture in each sample was calculated. Control experiments confirming the identity of the band containing SoxS were conducted with strains DJ901 (Δ*soxRS*) (34), BW709 ($soxR1^{con}$) (3), and GC4468/pBAD18-His_6-SoxS. We also note that the sonic extracts contained a band of insoluble cross-reacting material of MW slightly less than that of SoxS. Its presence would interfere with SoxS quantitation, but most of it was removed by centrifugation. The insoluble cross-reacting material does not contain SoxS, because its relative abundance is the same in extracts of strains DJ901 and GC4468; moreover, the relative amount of the cross-reacting material is unaffected by paraquat treatment.

Aliquots of the samples taken at each time were serially diluted and the number of cells per ml was determined with a Petroff–Hausser chamber. Then, the number of SoxS molecules per cell at each time point was determined by dividing the number of SoxS molecules per ml by the total number of cells per ml. The number of SoxS molecules per cell was determined in five independent experiments and the values presented are the averages from those experiments. The standard deviation values were about 25% of the mean values.

Assay of β-galactosidase activity. Fusions of SoxS-dependent promoters to *lac* were carried on λ prophages integrated into the chromosome of strain RA4468 (GC4468 *rob::kan*) in single copy (13). The lysogens were grown under the same conditions as described above for determination of the number of SoxS molecules per cell. The β-galactosidase specific activity in the samples taken at different times following paraquat treatment was determined as described previously (35), except that the cultures were grown in flasks, not in 96-well polypropylene blocks. The values at each time after induction of SoxS synthesis with paraquat were divided by the values in the corresponding uninduced cultures to yield the induction ratios.

RESULTS

Estimating the Number of SoxS Binding Sites in the E. coli Genome

Since the sequence of the *E. coli* genome is nonrandom (33), we wanted to determine whether it contains as many SoxS binding sites as predicted by information theory. We used the optimal SoxS binding sequence (28, 29) AnVGCACWWWnKRHCAAAHn (n = A, C, G, T; V = A, C, G; W = A, T; K = G, T; R = A, G; H = A, C, T) as a query sequence. We chose the optimal binding sequence rather than the consensus sequence of Martin *et al.* (12) because the optimal sequence contains information at 17 of 20 positions and thus using it as query sequence might be a more rigorous screen for binding sites. Since the members of the set of 16 soxbox sequences known to bind SoxS *in vitro* (30) match the optimal sequence at an average of 13.3 positions of the 17 informative positions, we searched the *E. coli* genome (33) under conditions allowing 3 and 4 mismatches and we screened both strands of the genome. At 3 mismatches, we identified 2,490 sites and at 4 mismatches, we identified 16,809 sites. By interpolation, this means that the *E. coli* genome contains ~12,500 sites that match the optimal binding sequence at 13.3 positions. Thus, the number of potential SoxS binding sites in the *E. coli* genome estimated by this analysis is similar to the number predicted by information theory. More importantly, since genome expression analysis has shown that SoxS regulates expression of ~100 genes (9), it is clear that the soxboxes functional in gene expression are but a small fraction of the total number of SoxS binding sites in the cell. As discussed above, this raises the question of whether SoxS activates transcription by the recruitment mechanism, which would require enough SoxS molecules to bind to most of the ~65,000 soxboxes in fast-growing cells, or whether by some means SoxS is able to distinguish the functional sites from the vast excess of equivalent but non-functional sites.

The Number of SoxS Molecules in Fast-Growing Cells Following Induction with Paraquat

We grew strain GC4468 in broth and induced SoxS synthesis with 500 μM paraquat, a concentration known to produce the maximum amount of SoxS-dependent gene activation (36). We took samples at various times after induction, determined the number of cells per ml, and disrupted the cells by sonication. After removing the cellular debris by centrifugation, a step necessary because of the presence in the cell extracts of insoluble, cross-reacting material that would

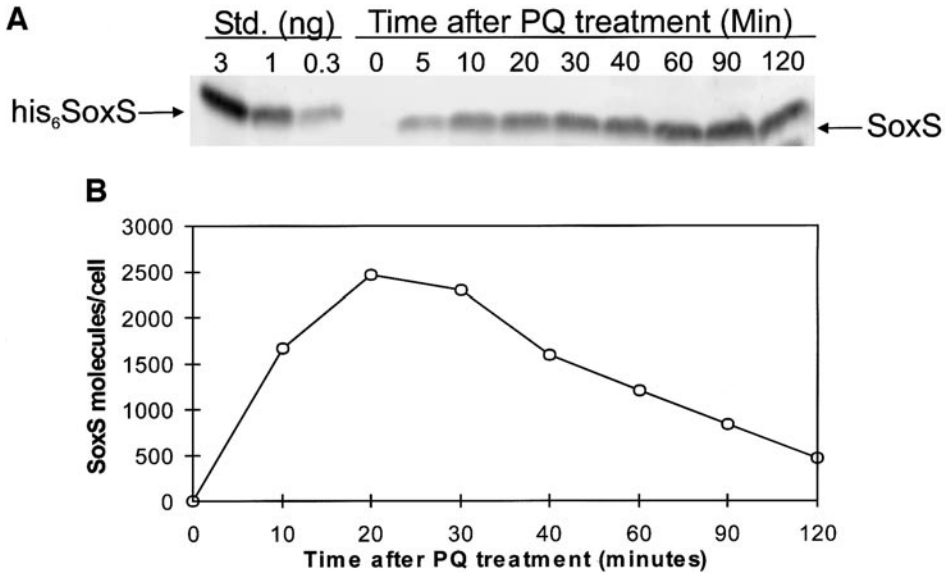

FIG. 1. Determination of the number of SoxS molecules per cell by Western blot analysis. (A) Representative Western blot with anti-SoxS serum and detection with the ECF chemifluorescent substrate. Strain GC4468 was grown in LB medium at 37°C and SoxS synthesis was induced by treatment with 0.5 mM paraquat (PQ). Samples were taken at the indicated times, subjected to SDS–PAGE in an 18% Tris-glycine gel, and processed for Western blot analysis and determination of the number of cells per milliliter as described under Materials and Methods. Samples containing 0.3, 1.0, and 3.0 ng of purified his_6-SoxS were included in the gel and used to generate a standard curve relating chemifluorescence signal to the concentration of SoxS. (B) Increase in the number of SoxS molecules per cell as a function of time after treatment with paraquat. After determining the number of cells per milliliter in each sample by counting in a Petroff–Hauser chamber and determining the number of SoxS molecules per cell by Western blot analysis as shown in A, the number of SoxS molecules per cell at each time point was calculated and the kinetics of accumulation plotted. The data are the average of five independent experiments.

otherwise interfere with quantitation of SoxS, we subjected the total soluble proteins to SDS–PAGE and immunoblotting with antiserum directed against highly purified his_6-SoxS, as described under Materials and Methods. Using a standard curve prepared for each gel from three known amounts of SoxS over a 10-fold range, we determined the number of SoxS molecules per cell as a function of time after paraquat treatment.

Figure 1 shows a typical immunoblot (Fig. 1A) and the time course of the change in the number of SoxS molecules per cell after induction (Fig. 1B). The experiments were repeated five times and the average values are presented. Prior to induction, the amount of SoxS was below the detection limit of the assay, which we estimate to be 50–100 molecules per cell. After induction, a protein band appeared whose intensity varied as a function of time after paraquat treatment. Control experiments (not shown), using a strain deleted of SoxS (DJ901), a strain constitutively expressing SoxS (BW799), and a strain expressing plasmid-borne his_6-SoxS, showed that the protein band appearing after paraquat treatment is SoxS and not a cross-reacting protein induced by SoxS synthesis.

Following induction, the abundance of SoxS rapidly increased to a maximum of 2,500 molecules per cell and then gradually declined to a level of about 500 molecules/cell. Thus, the maximum number of SoxS molecules per cell, 2,500, is far below the estimates of the number of soxboxes per cell, 50,000–65,000. Accordingly, if the soxboxes, both functional and non-functional, are more or less equivalent with respect to their ability to bind SoxS, then on average only about 4% of the sites will be bound at any given instant because there is a 25-fold excess of total SoxS binding sites relative to total SoxS molecules. Moreover, since the number of functional SoxS binding sites, ~500 per cell (~100 genes/genome × 5 genomes/cell), is small compared to the total number of soxboxes, then it would seem highly unlikely that SoxS would be able to activate transcription of its target genes without employing a means to discriminate between the functional and non-functional sites. In turn, this suggests that SoxS probably does not employ the recruitment mechanism described by Ptashne and Gann (31) to activate transcription of its target genes.

Kinetics of SoxS-Dependent Gene Expression Following Paraquat Induction of SoxS and the Effects of Hyperexpression of SoxS and the Presence of Rob on Target Gene Expression

A caveat to the inference that the number of SoxS molecules per cell is too small to allow transcription activation by recruitment is the possibility that activated gene expression is gradual and sub-optimal. To

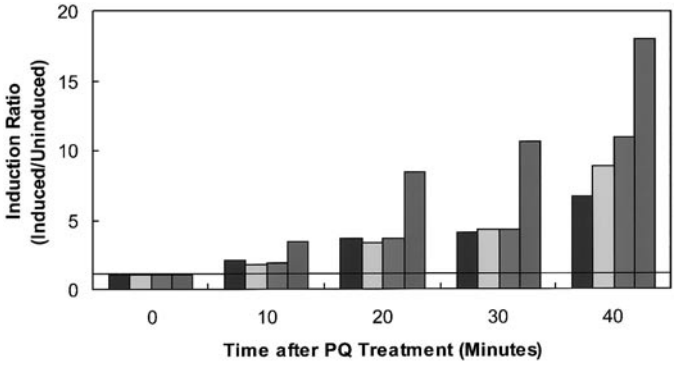

FIG. 2. Kinetics of SoxS-dependent gene expression following induction of SoxS synthesis with paraquat. As indicated by bar groups reading left to right, strains carrying *lac* fusions to the *micF* (left) *zwf* (second to left), *fpr* (second to right), and *fumC* (right) promoters were grown in LB broth at 37°C and treated with paraquat (PQ) to induce SoxS synthesis. Samples taken at different times after induction were assayed for β-galactosidase activity as described under Materials and Methods. The specific activity values at each time were divided by the specific activity values in corresponding uninduced cultures to yield the induction ratios. The specific activity values in the uninduced cultures were 77, 261, 186, and 208 Miller Units for *micF*, *zwf*, *fpr*, and *fumC*, respectively. The horizontal line represents an induction ratio of 1.0.

address this possibility, we determined the kinetics of SoxS-dependent gene expression in four fusion strains following induction of SoxS with paraquat. Figure 2 shows that the level of β-galactosidase produced from *lac* fusions to the *fpr*, *fumC*, *micF*, and *zwf* promoters begins to increase within a few mins after treating the culture with paraquat. This shows that significant SoxS-dependent gene expression occurs almost immediately, even before the maximum number of SoxS molecules per cell is produced.

We also introduced plasmid pBAD18-his$_6$-SoxS into the fusion strains and determined the time course of β-galactosidase expression following treatment of the cultures with 0.2% arabinose. When expressed from a plasmid under control of the arabinose-inducible *araBAD* promoter, about 100-fold more SoxS is produced from the plasmid under these conditions than is produced by induction of the chromosomal *soxS* gene with paraquat. Nonetheless, the amount of reporter gene expression was the same. Thus, the amount of SoxS in cells treated with paraquat is not limiting. If recruitment were the mechanism for SoxS activation of transcription of the small number of target genes relative to the excess of potential binding sites, then one might expect that the amount of SoxS would be subsaturating and that this would produce sub-maximal amounts of activated target gene expression. Accordingly, this experiment provides additional evidence that recruitment is not the mechanism of transcription activation by SoxS.

Another possible way in which SoxS might employ recruitment as the mechanism for transcription activation even when the number of SoxS molecules is small compared to the total number of binding sites would be if activation is dependent on the sum of the newly synthesized SoxS molecules and the 5,000-10,000 molecules of constitutively expressed Rob (18, 19). In this case, the total number of molecules potentially able to activate target gene expression, ~7,500–12,500, would be three- to fivefold closer to the number of soxboxes than if Rob does not contribute to SoxS-dependent gene expression. However, a *rob*::*kan* null mutation has no effect on the amount of target gene expression after induction of SoxS synthesis with paraquat (13). We conclude that Rob does not make a significant contribution to SoxS-dependent gene activation.

DISCUSSION

Recruitment is commonly accepted as the primary mechanism of transcription activation in *E. coli* (31). Typically, the activator binds with relatively high affinity, e.g., $K_D \leq 10^{-9}$ M, to a small number of DNA sites with relatively high sequence conservation. This binding then promotes open complex formation through protein-protein interactions between the activator and RNAP. This binding order is not obligatory as RNAP could form an unstable, closed complex before the activator binds its recognition sequence. While it is presumed that the number of activator molecules is approximately equal to or exceeds the number of specific binding sites, few studies have directly addressed this point.

Our work has provided evidence that SoxS, and by inference, MarA, and Rob, probably do not function by recruitment and we describe below arguments supporting its action by a new mechanism, which we term "pre-recruitment." Inconsistent with activation by recruitment, the data presented here show that the number of SoxS binding sites in fast-growing *E. coli* cells far exceeds the number of SoxS molecules. Using information theory, we had already calculated that a 4.6×10^6 bp genome of random sequence would have 13,000 SoxS binding sites such that fast-growing cells with 4–6 genomes per cell would contain 65,000 binding sites (28–30). Given the macroarray data showing that SoxS regulates ~100 genes (9), the number of functional SoxS binding sites would represent only about 1% of the total SoxS binding sites in fast-growing cells. The computer analysis presented above supports this estimate. Using the optimal SoxS binding sequence to interrogate the *E. coli* genome, we found ~12,500 sites when we allowed 3.7 mismatches among the 17 informative positions in the query sequence. Thus, this estimate of the number of sites in the *E. coli* genome is consistent with the estimate based on information theory and random sequence. Recently, Martin *et al.* (37) used their consensus sequence to estimate the number

of SoxS/MarA/Rob binding sites in the *E. coli* genome. In good agreement with our estimates, they found that 9,158 sites in the genome match the consensus sequence in at least 11 of the 14 specified positions (37). Moreover, and in perfect agreement, Martin *et al.* determined directly that the *E. coli* genome does indeed contain ~13,000 binding sites: they conducted gel mobility shift experiments with total *E. coli* DNA partially digested with *Sau*IIIa and *Hpa*I and found that the amounts of DNA shifted by SoxS and MarA were equivalent to the presence of 13,000 binding sites per genome (37). Thus, these four measures consistently agree that the *E. coli* genome contains ≥10,000 binding sites.

We used Western blot analysis employing purified his_6-SoxS as a standard to determine the number of SoxS molecules in uninduced cells and as a function of time after induction of SoxS synthesis with paraquat (Fig. 1). Prior to induction, the number of SoxS molecules was below the limit of detection by our sensitive chemifluorescence assay system, i.e., <50–100 molecules per cell. Immediately after induction there was a rapid increase in the number of SoxS molecules per cell so that in 20 min, the cells contained about 2,500 molecules. After reaching this maximum, the number of SoxS molecules per cell gradually declined. By rearranging the data into a differential plot of the number of molecules per ml as a function of the number of cells per ml (not shown), we found that SoxS initially accumulates at a rate producing ~2,500–3,000 molecules per cell and later accumulates at a rate producing ~850 molecules per cell. We do not know the basis for this biphasic accumulation rate of SoxS. It is not likely to be due to a rapid burst of synthesis followed by total shutoff of SoxS synthesis and then dilution through growth and cell division of the newly synthesized SoxS molecules because, in work to be reported elsewhere, we have found that SoxS is very unstable, with a half-life of ~3 min. One possible explanation for the biphasic kinetics of SoxS accumulation is that by ~20 min after induction, the activation of the regulon's genes has begun to restore to normal the physiological state of the cell, e.g., its redox potential, such that partially activated SoxR produces a sub-maximal accumulation rate of SoxS. Alternatively, the reduction in the rate of SoxS accumulation may be due to the negative autoregulation of SoxS synthesis described by Nunoshiba *et al.* (38). Regardless of the biphasic kinetics of SoxS accumulation, it is clear that the maximum number of SoxS molecules produced after paraquat treatment is small compared to the total number of SoxS binding sites in the cell and thus that recruitment is not likely to be the mechanism of transcription activation.

As a solution to the conundrum, we propose that SoxS uses a pre-recruitment mechanism for transcription activation. In this mechanism, SoxS would first make protein-protein contacts with RNAP, either in solution, or as the two molecules collide during their independent diffusion along the chromosomal DNA. The SoxS-RNAP binary complexes would then scan the chromosome for promoters containing SoxS binding sites in the proper position and orientation for transcription activation. Once found, the complexes might rearrange to allow activation at the two types of SoxS-dependent promoters (2, 12, 30). Thus, by this mechanism, SoxS "pre-recruits" RNAP by interacting with it before either has bound specifically to the promoter regions of activatable genes. In so doing, the deficit in sequence conservation of the SoxS binding sites is overcome by adding to the total information, the sequence conservation of the promoter hexamers, as recognized by the σ subunit of RNAP. In other words, the binary complex scans the chromosome for sites containing both the sequences recognizes by RNAP as well as the sequence recognized by SoxS. As such, by using the sequence information residing in the promoter, the binary complexes would be able to distinguish the functional SoxS binding sites from the vast excess of binding sites that are not functional because an activatable promoter does not reside nearby. To summarize the basic mechanistic difference between recruitment and pre-recruitment, we note that the hallmark of recruitment is that the protein-protein interactions between the activator and RNAP occur after one or both proteins have bound specifically to their DNA targets, whereas in pre-recruitment the protein-protein interactions between the activator and RNAP occur in the absence of specific DNA binding; thus, the protein-protein interactions in recruitment produce a DNA-activator-RNAP ternary complex whereas in pre-recruitment, the protein-protein interactions result in formation of an activator-RNAP binary complex.

In work to be described in detail elsewhere, we provide genetic evidence supporting pre-recruitment as the mechanism of transcription activation by SoxS. Briefly, we found that SoxS mutants severely defective in DNA binding are *trans*-dominant negative to the wild type SoxS allele, while positive control mutants of SoxS are recessive. When recruitment is the mechanism of transcription, positive control mutants of the activator are *trans*-dominant negative, while mutants defective in DNA binding are recessive.

Martin *et al.* have proposed a similar model for transcription activation by MarA and SoxS, a model they refer to as "DNA scanning" (37). Moreover, they have provided strong, direct evidence for the model by demonstrating that MarA and SoxS are able to form binary complexes with RNAP in solution and that the equilibrium binding constant for the interaction is ~0.3 μM (37). Rob may also activate transcription by this mechanism as it too is able to interact with RNAP in solution, though with lower affinity than MarA and SoxS (37). They also showed that ternary complexes containing RNAP, promoter DNA, and either MarA or SoxS

are more stable than the corresponding ternary complex containing Rob (37). Thus, in addition to the theoretical arguments presented here that lead to the hypothesis that pre-recruitment/DNA scanning, not recruitment, is the mechanism of transcription activation by SoxS, MarA, and Rob, both genetic and biochemical experiments are available to provide direct support for the model.

In further support of pre-recruitment/DNA scanning as the pathway of transcription activation by SoxS, we point out that a fundamental feature of the SoxRS and MarRAB systems is at odds with recruitment as the mechanism. In most genetic regulatory systems, both the sensor and the regulator are present at all times (and are often functions of a single protein) so that when a signal is received, the response can be initiated immediately and at a maximum rate. This is not the case with the two gene, two stage SoxRS and MarRAB systems, since SoxS and MarA must be synthesized *de novo* before they can begin to mount the respective defense responses (7, 8, 17, 39). Having to synthesize the regulator before initiating a response against toxic agents like redox-cycling compounds and antibiotics would appear to be a bad regulatory design. However, the SoxRS and MarRAB systems can theoretically begin to respond rapidly, because with SoxS and MarA being only107 and 129 amino acids in length, respectively, the polypeptides can be synthesized in only ~7–8 sec, since the polypeptide chain elongation rate is 15 amino acids/sec in fast-growing cells (32). Moreover, since SoxS and MarA bind DNA and activate transcription as monomers (23, 27), the cell does not have to accumulate a high enough concentration to allow oligomerization and thus the first SoxS and MarA molecules synthesized have the potential to initiate the defense response. However, the fact that the soxboxes/marboxes residing in SoxS-/MarA-dependent promoters represent only a small fraction of the total number of SoxS/MarA binding sites means that activation of the target gene transcription is not likely to occur efficiently by recruitment, because most newly synthesized molecules would bind to non-functional sites and not sites where they could activate transcription. Because of this limitation, it is even more difficult to see how recruitment would allow the cell to induce gene expression within mins after receiving the stress signal. Pre-recruitment, on the other hand, provides a solution to the problem, since the first SoxS/MarA molecules synthesized would be able to interact immediately with molecules of RNAP to form SoxS-RNAP or MarA-RNAP binary complexes, which would then be able to scan the chromosome in search of SoxS-/MarA-dependent promoters. Accordingly, pre-recruitment would provide a means for the two gene, two stage SoxRS and MarRAB systems to respond quickly to potentially lethal stresses and to do so even though the direct regulator of the defense response binds to a highly degenerate binding site.

The pre-recruitment/DNA scanning pathway of transcription activation may be used widely throughout biology. For example, eukaryotic transcription activators often recognize a short DNA sequence that is present at thousands of sites in the genome. We argue that to overcome this low information content, activators frequently form multi-protein transcription initiation complexes, e.g., "enhanceosomes" (40–42), where the ability to bind specifically to target promoters is achieved by combining the DNA binding specificities of several proteins in the complex. This process is equivalent to the pre-recruitment/DNA scanning model that we propose here and that Martin *et al.* (37) have proposed elsewhere.

ACKNOWLEDGMENTS

We are grateful to R. G. Martin and J. L. Rosner for fruitful discussions, the sharing of data prior to publication, and comments on the manuscript. We also thank S. O. Rosenberg for advice on the Western blot experiments. This work was supported by Public Health Service Grant GM27113 from the National Institutes of Health.

REFERENCES

1. Amábile Cuevas, C. F., and Demple, B. (1991) Molecular characterization of the *soxRS* genes of *Escherichia coli:* Two genes control a superoxide stress regulon. *Nucleic Acids Res.* **19,** 4479–4484.
2. Jair, K.-W., Fawcett, W. P., Fujita, N., Ishihama, A., and Wolf, R. E., Jr. (1996) Ambidextrous transcriptional activation by SoxS: Requirement for the C-terminal domain of the RNA polymerase alpha subunit in a subset of *Escherichia coli* superoxide-inducible genes. *Mol. Microbiol.* **19,** 307–317.
3. Tsaneva, I. R., and Weiss, B. (1990) *soxR,* a locus governing a superoxide response regulon in *Escherichia coli* K-12. *J. Bacteriol.* **172,** 4197–4205.
4. Wu, J., and Weiss, B. (1991) Two divergently transcribed genes, *soxR* and *soxS,* control a superoxide response regulon of *Escherichia coli. J. Bacteriol.* **173,** 2864–2871.
5. Pomposiello, P. J., and Demple, B. (2001) Redox-operated genetic switches: The SoxR and OxyR transcription factors. *Trends Biotechnol.* **3,** 109–114.
6. Demple, B. (1996) Redox signaling and gene control in the *Escherichia coli soxRS* oxidative stress regulon–A review. *Gene* **179,** 53–57.
7. Nunoshiba, T., Hidalgo, E., Amábile Cuevas, C. F., and Demple, B. (1992) Two-stage control of an oxidative stress regulon: The *Escherichia coli* SoxR protein triggers redox-inducible expression of the *soxS* regulatory gene. *J. Bacteriol.* **174,** 6054–6060.
8. Wu, J., and Weiss, B. (1992) Two-stage induction of the *soxRS* (superoxide response) regulon of *Escherichia coli. J. Bacteriol.* **174,** 3915–3920.
9. Pomposiello, P. J., Bennik, M. H., and Demple, B. (2001) Genome-wide transcriptional profiling of the *Escherichia coli* responses to superoxide stress and sodium salicylate. *J. Bacteriol.* **183,** 3890–3902.
10. Gallegos, M.-T., Schleif, R., Bairoch, A., Hofmann, K., and

Ramos, J. L. (1997) AraC/XylS family of transcriptional regulators *Microbiol. Mol. Biol. Rev.* **61,** 393–410.

11. Martin, R. G., and Rosner, J. L. (2001) The AraC transcriptional activators. *Curr. Microbiol.* **4,** 132–137.
12. Martin, R. G., Gillette, W. K., Rhee, S., and Rosner, J. L. (1999) Structural requirements for marbox function in transcriptional activation of *mar/sox/rob* regulon promoters in *Escherichia coli:* Sequence, orientation and spatial relationship to the core promoter, *Mol. Microbiol.* **34,** 431–441.
13. Martin, R. G., Gillette W. K. Rosner J. L. (2000) Promoter discrimination by the related transcriptional activators MarA and SoxS: Differential regulation by differential binding. *Mol. Microbiol.* **35,** 623–34.
14. Ariza, R. R., Li, Z., Ringstad, N., and Demple, B. (1995) Activation of multiple antibiotic resistance and binding of stress-inducible promoters by *Escherichia coli* Rob protein. *J. Bacteriol.* **177,** 1655–1661.
15. Ariza, R. R., Cohen, S. P., Bachhawat, N., Levy, S. B., and Demple, B. (1994) Repressor mutations in the marRAB operon that activate oxidative stress genes and multiple antibiotic resistance in *Escherichia coli. J. Bacteriol.* **176,** 143–148.
16. Alekshun, M. N., and Levy S. B. (1999) The mar regulon: Multiple resistance to antibiotics and other toxic chemicals. *Trends Microbiol.* **7,** 410–413.
17. Alekshun, M. N., and Levy S. B. (1997) Regulation of chromosomally mediated multiple antibiotic resistance: The mar regulon. *Antimicrob. Agents Chemother.* **41,** 2067–2075.
18. Azam, T. A., and Ishihama A. (1999) Twelve species of the nucleoid-associated protein from *Escherichia coli.* Sequence recognition specificity and DNA binding affinity. *J. Biol. Chem.* **274,** 33105–33113.
19. Skarstad, K., Thöny, B., Hwang, D., and Kornberg, A. (1993) A novel binding protein of the origin of the *Escherichia coli* chromosome. *J. Biol. Chem.* **268,** 5365–5370.
20. Nakajima, H., Kobayashi, K., Kobayashi, M., Asako, H., and Aono, R. (1995) Overexpression of the *robA* gene increases organic solvent tolerance and multiple antibiotic and heavy metal ion resistance in *Escherichia coli. Appl. Environ. Microbiol.* **61,** 2302–2307.
21. Fawcett, W. P., and Wolf, R. E., Jr. (1994) Purification of a MalE-SoxS fusion protein and identification of the control sites of *Escherichia coli* superoxide-inducible genes. *Mol. Microbiol.* **14,** 669–679.
22. Jair, K.-W., Martin, R. G., Rosner, J. L., Fujita, N., Ishihama, A., and Wolf, R. E., Jr. (1995) Purification and regulatory properties of MarA protein, a transcriptional activator of *Escherichia coli* multiple antibiotic and superoxide resistance promoters. *J. Bacteriol.* **177,** 7100–7104.
23. Jair, K.-W., Yu, X., Skarstad, K., Thöny, B., Fujita, N., Ishihama, A., and Wolf, R. E., Jr. (1996) Ambidextrous transcriptional activation of promoters of the superoxide and multiple antibiotic resistance regulons by Rob, a binding protein of the *Escherichia coli* origin of chromosomal replication. *J. Bacteriol.* **178,** 2507–2513.
24. Li, Z., and Demple, B. (1994) SoxS, an activator of superoxide stress genes in *Escherichia coli.* Purification and interaction with DNA. *J. Biol. Chem.* **269,** 18371–18377.
25. Li, Z., and Demple, B. (1996) Sequence specificity for DNA binding by *Escherichia coli* SoxS and Rob proteins. *Mol. Microbiol.* **20,** 937–945.
26. Kwon, H. J., Bennik, M. H. J., Demple, B., and Ellenberger, T. (2000) Crystal structure of the *Escherichia coli* Rob transcription factor in complex with DNA. *Nat. Struct. Biol.* **7,** 424–430.
27. Rhee, S., Martin, R. G., Rosner, J. L., and Davies, D. R. (1998) A novel DNA-binding motif in MarA: The first structure for an AraC family transcriptional activator. *Proc. Natl. Acad. Sci. USA* **95,** 10413–10418.
28. Griffith, K. L., and Wolf, R. E., Jr. (2001) Systematic mutagenesis of the DNA binding sites for SoxS in the *Escherichia coli zwf* and *fpr* promoters: Identifying nucleotides required for DNA binding and transcription activation. *Mol. Microbiol.* **40,** 1141–1154.
29. Griffith, K. L., and Wolf, R. E., Jr. (2001) Systematic mutagenesis of the DNA binding sites for SoxS in the *Escherichia coli zwf* and *fpr* promoters: Identifying nucleotides required for DNA binding and transcription activation. Erratum. *Mol. Microbiol.* **42,** 571.
30. Wood, T. I., Griffith, K. L., Fawcett, W. P., Jair, K.-W., Schneider, T. D., and Wolf, R. E., Jr. (1999) Interdependence of the position and orientation of SoxS binding sites in the transcriptional activation of the class I subset of *Escherichia coli* superoxide-inducible promoters. *Mol. Microbiol.* **34,** 414–430.
31. Ptashne, M., and Gann, A. (1997) Transcriptional activation by recruitment *Nature* **386,** 569–577.
32. Bremer, H., and Dennis, P. P. (1996) Modulation of Chemical Composition and Other Parameters of the Cell by Growth Rate, American Society for Microbiology, Washington, DC.
33. Blattner, F. R., Plunkett, G., III, Bloch, C. A., Perna, N. T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J. D., Rode, C. K., Mayhew, G. F., Gregor, J., Davis, N. W., Kirkpatrick, H. A., Goeden, M. A., Rose, D. J., Mau, B., and Shao, Y. (1997) The complete genome of *Escherichia coli* K-12. *Science* **277,** 1453–1474.
34. Greenberg, J. T., Monach, P., Chou, J. T., and Demple, B. (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.
35. Griffith, K. L., and Wolf, R. E., Jr. (2002) Measuring β-galactosidase activity in bacteria: Cell growth, permeabilization, and enzyme assays in 96-well arrays. *Biochem. Biophys. Res. Commun.* **290,** 397–402.
36. Fawcett, W. P., and Wolf, R. E., Jr.Jr. (1995) Genetic definition of the *Escherichia coli zwf* "soxbox," the DNA binding site for SoxS-mediated induction of glucose 6-phosphate dehydrogenase in response to superoxide. *J. Bacteriol.* **177,** 1742–1750.
37. Martin, R. G., Gillette, W. K., Martin, N. I., and Rosner, J. L. (2002) Complex formation between activator and RNA polymerase as the basis for transcription activation by MarA and SoxS in *Escherichia coli. Mol. Microbiol.* **43,** in press.
38. Nunoshiba, T., *et al.* (1993) Negative autoregulation by the *Escherichia coli* SoxS protein: A dampening mechanism for the soxRS redox stress response. *J. Bacteriol.* **175,** 7492–7494.
39. Gambino, L., Gracheck, S. J., and Miller, P. F. (1993) Overexpression of the MarA positive regulator is sufficient to confer multiple antibiotic resistance in *Escherichia coli. J. Bacteriol.* **175,** 2888–2894.
40. Carey, M. (1998) The enhanceosome and transcriptional synergy. *Cell* **92,** 5–8.
41. Echols, H. (1986) Multiple DNA-protein interactions governing high-precision DNA transactions. *Science* **233,** 1050–1056.
42. Grosschedl, R. (1995) Higher-order nucleoprotein complexes in transcription: Analogies with site-specific recombination. *Curr. Opin. Cell Biol.* **7,** 362–370.