SI-1

Localization of F plasmid SopB protein and Gene silencing via protein-mediated subcellular localization of DNA

Sook-Kyung Kim and James C. Wang

Department of Molecular and Cellular Biology, Harvard University, Cambridge, MA02138

Abstract

The subcellular localization of the SopB protein, which is encoded by the *Escherichia coli* F plasmid and is involved in the partition of the single-copy plasmid, was directly visualized through the expression of the protein fused to the jellyfish green fluorescent protein (GFP). The fusion protein was found to localize to positions close but not at the poles of exponentially growing cells. Examination of derivatives of the fusion protein lacking various regions of SopB suggests that the signal for the cellular localization of SopB resides in a region close to its N terminus. Overexpression of SopB led to silencing of genes linked to, but well-separated from, a cluster of SopB-binding sites termed *sopC*. In this SopB-mediated repression of *sopC*-linked genes, all but the N-terminal 82 amino acids of SopB can be replaced by the DNA-binding domain of a sequence-specific DNA-binding protein, provided that the *sopC* locus is also replaced by the recognition sequence of the DNA-binding domain. These results suggest a mechanism of gene silencing: patches of closely packed DNA-binding protein is localized to specific cellular sites; such a patch can capture a DNA carrying the recognition site of the DNA-binding domain and sequestrate genes adjacent to the recognition site through nonspecific binding of DNA.

Introduction

The *Escherichia coli* fertility or F plasmid encodes two proteins, SopA and SopB, that are known to be required for the stable inheritance of the plasmid (1-3). The 323-residue SopB protein is a DNA-binding protein that specifically binds to the F plasmid *sopC* locus, which is composed of 12 tandem imperfect 43-bp repeats. SopA, a 388-residue protein with a DNA-dependent ATPase

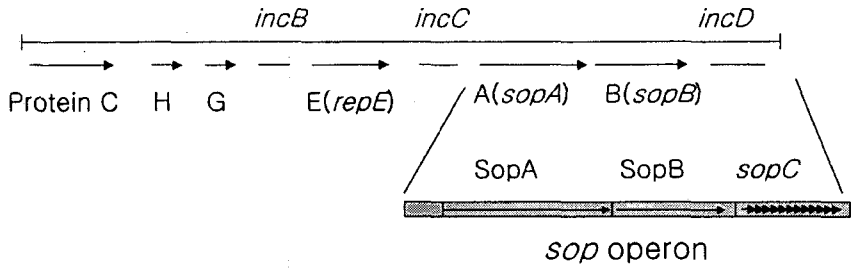

Fig. 1. The *sop* operon region of *Escherichia coli* F plasmid.

The genetic map of the *sopA* and *sopB* genes and the cis-acting *sopC/incD* region. The arrows indicate 12 tandom repeats of the 43 bp in the *sopC* region. *inc*, incompatibility; Protein H and G, *ccdA* and *ccdB* (coupled cell division); Protien C, replication protein (*oriV*); Protein E, replication protein (*oriS*).

activity, is involved in the autoregulation of the *sopAB*-operon and appears to bind, by itself or perhaps together with SopB, to four repetitive sequence elements in the regulatory region of the *sopAB* operon (Fig. 1). Proteins that are structurally and functionally similar to SopA and SopB are known to be encoded by other low-copy-number plasmids such as the P1 prophage (ParA and ParB, respectively, see refs 4-5).

The F plasmid SopB protein has also been implicated in a phenomenon termed the *IncG* type incompatibility. Inheritance of a single-copy F or *E. coli oriC* replicon is compromised by the expression of a high level of SopB if the replicon carries the *sopABC* locus (6). Two finding that are likely to be related to the *IncG* type incompatibility. It was found that in *E. coli* cells expressing a high level of SopB, plasmids bearing either the complete *sopC* element or as few as one single 43-bp *sopC* motif have a much higher linking number than the same plasmids without the *sopC* motif. Furthermore, in such cells either chromosomally located or plasmid-borne genes that are linked to *sopC* were found to be silenced; the distance over which *sopC* can exert this silencing effect is at least 10-kb. Genes within this silenced region appear to be inaccessible to cellular proteins, as evidenced by monitoring the sites of modification by *dam* methylase and the sites of cleavage by DNA gyrase (7).

Two plausible mechanisms were proposed to account for these observations. In the first, binding of SopB to *sopC* is postulated to nucleate the formation of a nucleoprotein filament, within which the DNA is inaccessible to transcription. In the second, binding of SopB to *sopC* is postulated to sequestrate the DNA to a subcellular region inaccessible to cellular enzymes including RNA polymerase. Whereas these recent findings with ParB (8) provide new support of the

nucleoprotein filament model proposed earlier (7), in the meaning time we have been entertaining alternative interpretations. There were two subtle hints from our earlier studies that favored the DNA sequestration model. First, SopB-mediated gene silencing is known to be abolished by the deletion of the N-terminal 71 amino acids of the protein (9). This particular N-terminal region of SopB can apparently mediate the polar localization of the protein in *E. coli* cells. Visualization of SopB and its fragments fused to the green fluorescence protein (GFP) in *E. coli* cells indicate that intact SopB is localized to the "quarter-cell" positions near the cell poles, and deletion of the N-terminal 71 amino acids abolishes this localization. N-terminal SopB fragments as short as 82 amino acids also show a polar localization, though not confined to the quarter-cell positions (10).

Fusion proteins in which fragments of SopB are linked to the DNA binding domain of the yeast *GAL4* gene product, or a multiple zinc-finger protein engineered for binding to a specific DNA sequence. We found that such fusion proteins containing an N-terminal segment of SopB as short as 82 amino acids could silence genes linked to their respective sequence-specific DNA binding sites. These results are not readily explained by the nucleoprotein-filament model. We propose instead that gene silencing can be effected by the formation of a cellular patch of DNA binding domains through their localization to specific cellular sites: A DNA carrying the recognition site of a DNA binding protein can first bind to the patch through interactions between the recognition site and a DNA binding domain located in the patch; genes adjacent to the recognition site are then silenced through nonspecific binding to the closely packed DNA binding domains within the same patch.

Methods and Results

1. Subcellular localization of F plasmid SopB protein

1) Expression of a functional SopB protein with a GFP fused to its C terminus

To visualize the subcellular localization of the F plasmid SopB protein, we used gene fusion to link the widely used jellyfish GFP to the C terminus of SopB. Functionality of the fusion protein was tested by using the ability of SopB in silencing genes that linked to *sopC*. HB101 cells bearing a pair of

plasmids ptacSopB, which expresses SopB from the tac promoter, and pASLS4, which carries the tetracyclin-resistance and chloramphenicol-resistance markers and the *sopC* sequence, induction of SopB expression by the addition of IPTG led to repression of both *sopC*-linked drug-resistance markers, and colony formation on chloramphenicol or tetracyclin plates was abolished. Identical results were obtained when ptacSopB was replaced by pSopB-GFP fusion protein is functional *in vivo*. The functionality of the fusion protein was confirmed by the use of a strain ASL1270, which was derived from HB101 by the insertion of a DNA segment containing *sopC* and a chloramphenicol-resistance marker into the *λattB* locus of the *E. coli* chromosome. In the presence of IPTG, colony formation of ASL1270 cells transformed with either ptacSopB or pSopB-GFP was observed on plates containing ampicillin but not on plates containing chloramphenicol.

2) Visualization of the subcellular localization of the SopB-GFP by fluorescence and phase-contrast microscopy

In cells overexpressing SopB-GFP, the fusion protein was specifically located to regions close to but not at the cell poles (Fig. 2B, 2C, 2D). Cells expressing GFP itself showed a uniform distribution of the fluorescent protein (Fig. 2A) indicating that the observed localization of SopB-GFP reflects a

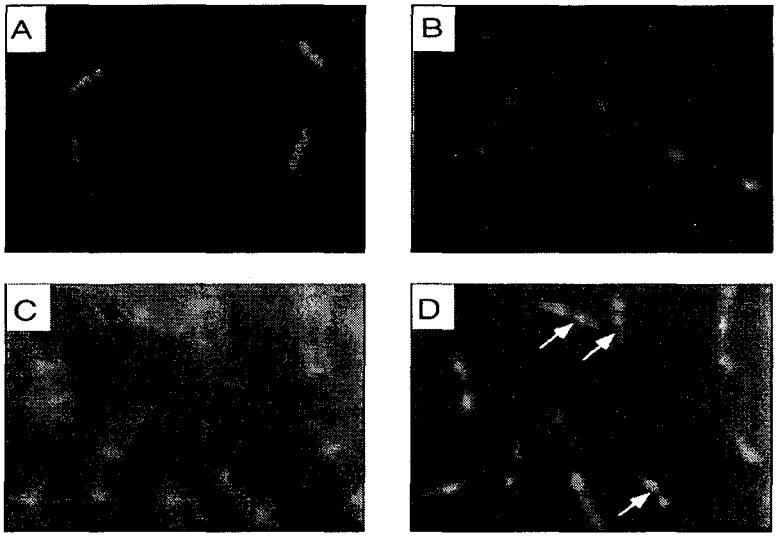

Fig. 2. Color image obtained by fluorescent microscopy of *Escherichia coli*.
(a) *E. coli* cells expressing GFP only. (b) *E. coli* expressing GFP-tagged full-length SopB. (c) Fluorescent-phase micrographs of *E. coli* (pSopB-GFP) (d) DAPI-stained cells (pSopB-GFP).

property of the SopB protein and not that of the GFP fused to it. For a population of cells growing exponentially, a single fluorescent dot was usually observed in the smaller cells and a pair of widely separated dots observed in the longer ones.

3) Involvenent of the N-terminal region of a SopB in cellular localization

To deduce which regions of SopB are involved in its specific subcellular localization, plasmids were constructed for the expression of various N- and C-terminal truncation of SopB fused to the GFP. Visualization revealed that deletion of the C-terminal 49 or 79 residues, which is known to abolish binding of SopB to *sopC*, had little effect on the cellular localization of the protein. In contrast, localization of SopB-GFP to positions close to the poles was no longer observed upon missing the N-terminal 71 residues of SopB.

2. Gene silencing via protein-mediated subcellular localization of DNA

1) Design of fusion protein and lacZ reporter expression plasmids

Two sets of expression plasmids were constructed. In one, codons for full length SopB or its fragment were joined to those encoding the DNA-binding domain of a sequence-specific DNA-binding protein. In the other, a *lacZ* reporter gene was placed under the control of a constitutive promoter, and a cluster of sites for the binding of the sequence-specific DNA-binding protein was placed about 1 kb upstream of this promoter. In the SopB(1-323) Gal4 plasmid, the 323 codons of intact SopB are placed under the control of an inducible promoter Ptac, ands 147 codons for the DNA-binding domain of yeast Gal4 proteins are added to the end of the SopB ORF. In the *lacZ* reporter plasmid carrying (UAS)4, a cluster of four yeast Gal4 protein-binding sites, the β-galactosidase ORF is placed under the control of a constitutive promoter Ptet.

2) Effect of fusion protein expression on β-galactosidase

In the above experiments, β-galactosidase activity was measured 2 hr after induction of fusion protein expression. Results obtained with the pairs of plasmids described above as well as additional pairs are summarized in Table 1. In all cases, expression of a chimeric protein in which a sequence-specific DNA

Table 1. Summary of protein-mediated gene silencing in *E. coli*.

Protein	DNA recognition site	Repression of lacZ
SopB(1-323)	*sopC*	yes
SopB(1-323)	none	no
SopB(1-323)	$(UAS)_4$	no
SopB(1-323)	$(N/\!/Z)_8$	no
SopB(1-323)Gal4	none	no
SopB(1-323)Gal4	$(UAS)_4$	yes
SopB(1-323)Gal4	$(N/\!/Z)_8$	no
SopB(1-82)Gal4	none	no
SopB(1-82)Gal4	$(UAS)_4$	yes
SopB(72-323)Gal4	$(N/\!/Z)_8$	no
Gal4	$(UAS)_4$	no
SopB(1-323)Zif	none	no
SopB(1-323)Zif	$(UAS)_4$	no
SopB(1-323)Zif	$(N/\!/Z)_8$	yes
SopB(1-82)Zif	none	no
SopB(1-82)Zif	$(UAS)_4$	no
SopB(1-82)Zif	$(N/\!/Z)_8$	yes
SopB(72-323)Zif	none	no
SopB(72-323)Zif	$(UAS)_4$	no
SopB(72-323)Zif	$(N/\!/Z)_8$	no
Zif	$(N/\!/Z)_8$	no

binding domain is fused to the C-terminus of SopB, or the SopB N-terminal fragment spanning amino acids 1 to 82, leads to the repression of the lacZ reporter gene if and only if a cluster of recognition sites of the DNA binding domain is present. SopB lacking the first 71 amino acids does not show this gene-silencing effect when fused to the same DNA binding domains, and similarly the expression of Gal4 or Zif itself has no effect on the expression of lacZ whether the reporter gene is linked to a cluster of its recognition sites.

DISCUSSION

Our results indicate that abutting the N-terminal 82 amino acid residues of SopB to any sequence-specific DNA binding domain yields a protein capable of repressing genes linked to the recognition site of the particular DNA binding domain. The three DNA binding proteins examined, SopB, Gal4, and Zif268//NRE, are not known to share common structural features in their

interactions with their respective DNA binding sites. As a consequence, the SopB N-terminal 82 amino acid fragment fused to each of these proteins is expected to assume a different spatial position relative to the DNA binding surface of the fusion protein. It is therefore difficult to envision, structurally, how these fusion proteins could all form a nucleoprotein filament with the differently positioned SopB fragment serving as the common glue. Our earlier observation that SopB readily saturates its binding sites within *sopC* but does not form a nucleoprotein filament even at a very high concentration also argues against the presence of a sticky SopB segment capable of acting as an amorphous cement.

According to this gene sequestration model, fusion of a DNA binding domain to the N-terminal 82 amino acid fragment of SopB localizes the DNA binding domains to the cell poles, presumably through specific interactions between the N-terminal region of SopB and a membrane-bound protein. Patches of closely packed fusion protein molecules, with their DNA binding domains exposed to the cytoplasmic side, are thus formed at these cellular locations. For a DNA segment carrying a cluster of recognition sites of the DNA binding domain of the localized fusion protein, these recognition sites would first bind to such a patch through sequence-specific DNA-protein interactions. DNA segments adjacent to the bound cluster of recognition sites would then bind, through nonspecific interactions, to nearby DNA binding domains in the same patch. In such a sequestration model, the transcription machinery is either excluded from the particular subcellular location, perhaps owing to its preferential localization to other cellular regions, or is incapable of accessing the DNA because of the fixation of the DNA to multiple points within the patch of DNA binding domains.

A number of proteins that interact with DNA, including the chromosomally encoded SopB homologs of *Bacillus subtilis*, *Pseudomonas putida*, and *Caulobacter crescentus*, and bacterial DNA topoisomerase IV A-subunit and MinD protein, have recently been shown to have a polar localization (11-15). In *B. subtilis*, localization of a number of replication proteins to a mid-cell location has also been reported (16). Thus silencing of genes through protein-mediated localization of a DNA to cell poles, and perhaps to other particular subcellular locations as well, may play significant roles in chromosome partition as well as other chromosomal transactions in various bacteria. It should be pointed out, however, that the gene sequestration model of gene silencing proposed here requires the formation of a patch of closely packed DNA binding domains but

does not require specific localization of such a patch to cell poles.

It is likely that sequence-specific and nonspecific interactions between a long DNA segment and a patch of DNA binding proteins may represent a mode of gene regulation much more general than the special cases discussed here. Perinuclear localization of inactive regions of eukaryotic chromosomes, for examples, is well-known (17-20). In the repression of a reporter gene flanked by the silencer elements of yeast mating loci, repression of the reported gene occurs if such a cassette is placed close to, but not if it is placed far from, telomeres that appear to localize by immunofluorescence and *in situ* hybridization to a limited number of discrete sites close to the nuclear pores (reviewed in Ref. 21). Interestingly, overexpression of Gal4 DNA binding domain fused to integral membrane proteins was recently reported to facilitate transcriptional silencing of a modified yeast mating locus in which the silencer elements had been replaced by the Gal4 binding sites (22).

References

1. Nordstrom, K. & Austin, S. J. (1989) Annu. Rev. Genet. 23, 37-69.
2. Hiraga, S. (1992) Annu. Rev. Biochem. 61, 283-306.
3. Wake, R. G. & Errington, J. (1995) Annu. Rev. Genet. 29, 41-67.
4. Davis, M. A. Radnedge, L., Martin, K. A., Hayes, F., Youngren, B. & Austin, S. J. (1996) Mol. Microbiol. 21, 1029-1036.
5. Montellabi, V. M., Rouch, D. A. & Thomas, C. M. (1990) Mol. Microbiol. 4, 1455-1463.
6. Kusugawa, N., Mori, H., Kondo, A. & Hiraga, S. (1987) Mol. Gen. Genet. 208, 365-372.
7. Lynch, A. S. & Wang, J. C. (1995) Proc. Natl. Acad. Sci. USA 92, 1896-1900.
8. Rodionov, O., Lobocka, M. & Yarmolinsky, M. (1999) Science 283, 546-549.
9. Hanai, R., Liu, Z., Benedetti, P., Caron, P. R., Lynch, A. S. & Wang, J. C. (1996) J. Biol. Chem. 271, 17469-17475.
10. Kim, S.-K. & Wang, J. C. (1998) Proc. Natl. Acad. Sci. USA 95, 1523-1527.
11. Lin, D. C.-H., Levin, P. A. & Grossman, A. D. (1997) *Proc. Natl. Acad. Sci* **94**, 4721-472.
12. Ogasawa, N. & Yoshikawa, H. (1992) *Mol. Microbiol.* **6**, 629-634.
13. Mohl, D. A. & Gober, J. W. (1997) *Cell* **88**, 675-684.
14. Huang, W. M., Libbey, J. L., Hoeven, P. V. D. & Yu S. H. (1998) *Proc. Natl Sci. USA* **95**, 4652-4657.

15. Marston, A. L., Thomaides, H. B., Edwards, D. H., Sharpe,,M. E. & Erring (1998) *Genes Dev.* **12**,3419-3430.
16. Lemon, K. P. & Grossman, A. D. (1998) *Science* **282**, 1516-1519.
17. Rae, P. M. M., and Francke, W. W. (1972) *Chromosoma* **39**, 443-456.
18. Walker, C. L., Cargile, C. B., Floy, K. M., Delannoy, M., & Migeon, B. *Proc. Natl. Acad. Sci. USA* **88**, 6191-6195.
19. Klein, F., Laroche, T., Cardenas, M. E., Hofmann, J. F., Schweizer, D. & G S. M. (1992) *J. Cell Biol.* **117**, 935-948.
20. Gotta, M., Laroche, T., Formenton, A., Maillet, L., Scherthan, H. & Gasser (1996) *J. Cell Biol.* **134**, 1349-1363.
21. Gotta, M. & Gasser S. M. (1996) *Experientia* **52**, 1136-1147.
22. Andruilis, E. D., Neiman, A. M., Zappulla, D. C., and Sternglanz, R. (1998) *N* 394, 592-595.