

## *Anabaena* sp. Strain PCC 7120 *bifA* Gene Encoding a Sequence-Specific DNA-Binding Protein Cloned by In Vivo Transcriptional Interference Selection

TAI-FEN WEI, T. S. RAMASUBRAMANIAN, FRANCES PU, AND JAMES W. GOLDEN\*

Department of Biology, Texas A&M University, College Station, Texas 77843-3258

Received 27 January 1993/Accepted 2 May 1993

VF1 is a DNA-binding protein from the cyanobacterium *Anabaena* sp. strain PCC 7120. VF1 was originally identified on the basis of its binding affinity to the upstream region of *xisA*, which encodes a heterocyst-specific site-specific recombinase. VF1 also binds to the *glnA*, *rbcl*, and *nifH* promoters in vitro, suggesting that VF1 interacts with genes expressed in both vegetative cells and heterocysts. The role of VF1 in regulating gene expression in PCC 7120 is unknown. As a step towards the goal of understanding the role of VF1 in regulating gene expression, we have cloned the *bifA* gene by using a genetic selection strategy. *bifA* encodes a protein, BifA, that has chromatographic and DNA-binding properties indistinguishable from those of VF1. The cloning strategy was based on a transcriptional interference assay in which a strong synthetic promoter, *conII*, interferes with the expression of an *aadA* gene, which provides resistance to spectinomycin and streptomycin (S. J. Elledge, P. Sugiono, L. Guarente, and R. W. Davis, Proc. Natl. Acad. Sci. USA 86:3689–3693, 1989). A selection plasmid, pAM994, which has the *conII* promoter negatively regulated by a VF1-binding site, was used to enrich for VF1-producing clones from an expression library containing PCC 7120 DNA fragments. Mobility shift assays were used to identify a 672-bp open reading frame that encoded VF1-like binding activity. The deduced BifA amino acid sequence shows 77% identity to NtcA, which is a global regulator involved in nitrogen control in *Synechococcus* sp. strain PCC 7942. Both BifA and NtcA belong to the cyclic AMP receptor protein (CRP) family of prokaryotic regulatory proteins. Genes similar to *envM*, *hisB*, and ORF60-5 were found near the *bifA* gene.

*Anabaena* sp. strain PCC 7120 is a filamentous cyanobacterium that grows as chains of photosynthetic vegetative cells in the presence of reduced nitrogen such as ammonia or nitrate. When deprived of a reduced nitrogen source, cells at semiregular intervals along each filament terminally differentiate to form heterocysts, which are specialized for nitrogen fixation (22, 42). The ultimate product of nitrogen fixation in the heterocyst is glutamine, which is exported to neighboring vegetative cells in exchange for the carbohydrate products of CO<sub>2</sub> fixation that are necessary for the production of ATP and reductant in the heterocyst.

Heterocyst formation requires substantial changes in gene expression that alter the physiology and morphology of the differentiating cell. For example, genes involved in nitrogen fixation (*nif*) are turned on, while genes involved in carbon fixation (*rbclS*) are turned off (17, 23). In addition to transcriptional gene regulation, two site-specific chromosomal rearrangements are tightly coupled to heterocyst differentiation: the excision of an 11-kb element from the *nifD* gene in the *nifHDK* operon and the excision of a 55-kb element from the *fdxN* gene in the *nifB-fdxN-nifS-nifU* operon (17–19, 24, 29, 30). The rearrangements produce continuous coding sequences and functional operons essential for nitrogen fixation. The *nifHDK* operon encodes the structural proteins of nitrogenase, and the *nifB-fdxN-nifS-nifU* operon is thought to encode proteins required for the maturation of nitrogenase (23, 30, 31).

We are interested in how these two *nif* gene rearrangements are regulated during heterocyst differentiation. The *xisA* gene, which is present within the 11-kb *nifD* element,

encodes the site-specific recombinase required for the excision of that element (21, 25). The regulation of *xisA* during heterocyst development is not understood, and attempts to map its promoter were not successful. Expression of *xisA* in *Escherichia coli* is sufficient to cause rearrangement of an artificial-substrate plasmid (6). In PCC 7120, forced *xisA* expression from a shuttle vector in vegetative cells, as measured by the rearrangement of the vegetative cell chromosome, requires the deletion of sequences upstream of the gene (6). This suggested that a *trans*-acting factor might be present in vegetative cells that could block expression of the *xisA* gene by interaction with this putative regulatory region.

A DNA-binding protein, VF1, which binds to *xisA* upstream sequences, has been identified and partially purified from extracts of PCC 7120 vegetative cells (7). DNase I footprinting and exonuclease III deletion analyses of the *xisA* upstream region mapped three adjacent VF1-binding sites to a 66-bp region that overlaps the sequences involved in suppression of *xisA* expression in vegetative cells (6, 7). VF1 is likely to be involved in the regulation of *xisA* expression, but its function may not be limited to *xisA* because it also binds to the *glnA*, *rbcl*, and *nifH* promoters in vitro (7). Although the binding of VF1 protein to these different promoter sequences has been partially characterized in vitro, the role VF1 may play in their in vivo regulation has not been determined. VF1's possible effects on gene expression are particularly enigmatic because of the differences in the regulation of the genes it interacts with.

To further study the role of VF1 in the regulation of gene expression in PCC 7120, we cloned the *bifA* gene by a cloning strategy based on a transcriptional interference assay (11, 12). The strategy provides a genetic selection for expressed clones of genes encoding sequence-specific DNA-

\* Corresponding author.

TABLE 1. Bacterial strains and plasmids

Strain or plasmid	Relevant characteristics	Source or reference
<i>Anabaena</i> sp. strain PCC 7120	Wild type	R. Haselkorn
<i>E. coli</i>		
DH5 $\alpha$ MCR	F <sup>-</sup> <i>mcrA</i> $\Delta$ ( <i>mrr-hsdRMS-mcrBC</i> ) $\phi$ 80 <i>dlacZ</i> $\Delta$ M15 $\Delta$ ( <i>lacZYA-argF</i> ) <i>U169 deoR</i>	BRL <sup>a</sup>
JM107	$\Delta$ ( <i>lac-proAB</i> ) <i>thi-1 gyrA96 endA1 hsdR17 relA1 supE44 mcrA</i> {F' <i>traD36 proAB lacI<sup>q</sup>Z</i> $\Delta$ M15}	43
Plasmids		
pBluescript SK(+/-)	Ap <sup>r</sup>	Stratagene
pSE380	Ap <sup>r</sup> , <i>trc</i> promoter, <i>lacI<sup>q</sup></i> , ribosome binding site, superlinker	5
pNN388	Cm <sup>r</sup> Sp <sup>r</sup> +Sm <sup>r</sup> , low copy number	11
pNN396	Ap <sup>r</sup> , <i>conII</i> promoter and polylinker	11
pAM251	Ap <sup>r</sup> , <i>xisA</i> gene with upstream regulatory sequence	7
pAM709	Ap <sup>r</sup> , 230-bp <i>XbaI-ScaI</i> fragment from pAM251 containing the VF1-binding site in <i>XbaI-SmaI</i> site of pBluescript SK(-)	This study
pAM993	Cm <sup>r</sup> , <i>conII</i> promoter regulated by <i>xisA</i> VF1-binding site from pAM709 in pNN388	This study
pAM994	Cm <sup>r</sup> , pAM1156 modified by removal of 4 bp at <i>KpnI</i> site	This study
pAM995	Ap <sup>r</sup> , original <i>bifA</i> clone in pSE380	This study
pAM996	Ap <sup>r</sup> , subclone of pAM995 insert as <i>SmaI-PstI</i> fragment into pBluescript SK(+)	This study
pAM1156	Cm <sup>r</sup> , <i>conII</i> promoter regulated by a 35-bp <i>glnA</i> VF1-binding site in pNN388	This study

<sup>a</sup> BRL, Bethesda Research Laboratories.

binding proteins. BifA protein produced in *E. coli* showed properties indistinguishable from those of VF1. The deduced BifA protein sequence showed that it is very similar to the cyanobacterial regulatory protein NtcA (39) and that it belongs to the cyclic AMP receptor protein (CRP) family of prokaryotic regulatory proteins.

## MATERIALS AND METHODS

**Strains and growth conditions.** Strains and plasmids used in this study are listed in Table 1. For protein extractions, *Anabaena* sp. strain PCC 7120 was grown in 2 liters of Allen and Arnon liquid medium diluted eightfold (AA/8), with modifications as previously described (20). Cells were grown at 30°C with illumination at approximately 100 microeinsteins m<sup>-2</sup> s<sup>-1</sup> and bubbled with 1% CO<sub>2</sub> in air. *E. coli* strains were grown as previously described (20). Media were supplemented as needed with appropriate antibiotics and 1 mM isopropyl- $\beta$ -D-thiogalactopyranoside (IPTG). *E. coli* DH5 $\alpha$ MCR (Bethesda Research Laboratories, Life Technologies, Inc., Gaithersburg, Md.) was used for maintenance and amplification of the expression library and all plasmids. *E. coli* JM107 (43) was used as the host for the in vivo genetic selection.

**Construction of selection plasmids.** Three selection plasmids were constructed based on in vitro studies of VF1-binding sites in the *glnA* promoter (32) and the *xisA* upstream region (7). A 35-bp oligonucleotide containing the 24-bp VF1-binding site from the *glnA* promoter was synthesized and ligated into the *SmaI* (destroyed)-*BamHI* site of pNN396 (Fig. 1). pNN396 is a multicopy plasmid containing the *conII* promoter and a polylinker (11). The VF1-binding site together with the upstream *conII* promoter was subsequently released as a *NotI-HindIII* restriction fragment and subcloned into the same sites of pNN388 (11) to produce pAM1156. To improve its selection characteristics, pAM1156 was digested with *KpnI*, blunted with T4 polymerase, and religated to form pAM994. This modification positioned an adenine at the predicted transcription start site

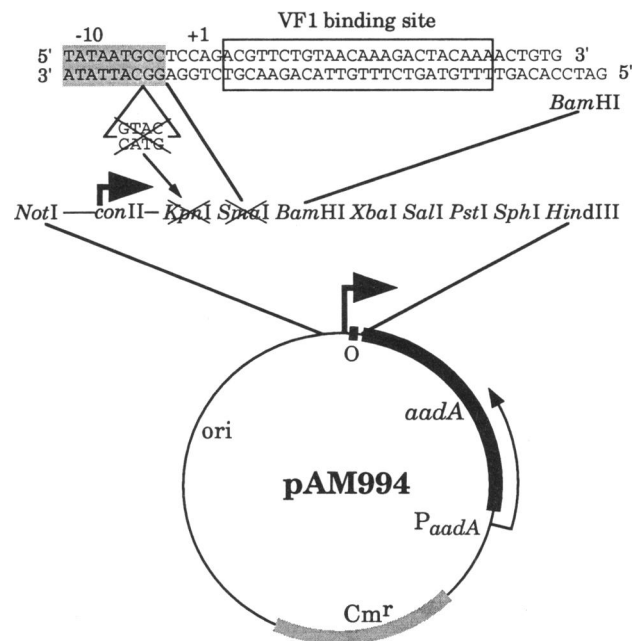


FIG. 1. Physical map of *bifA* selection plasmid pAM994. The nucleotide sequence at the top is a synthetic double-stranded oligonucleotide containing the VF1-binding site in the *glnA* promoter. The VF1-binding site acts as an artificial operator (O) when cloned immediately downstream of the *conII* promoter. The shaded sequence is from pNN396 and shows the removal of 4 bp at the *KpnI* site in the final pAM994 plasmid. The predicted transcription start site is labeled +1, and the -10 consensus region of the *conII* promoter is marked. The pNN396 *NotI-HindIII* fragment containing the *conII* promoter regulated by the VF1-binding site was cloned into the low-copy-number vector pNN388 downstream of, and antisense to, the *aadA* gene, which is driven by its weak native promoter (*P<sub>aadA</sub>*). The *aadA* gene provides resistance to SP and SM.

and placed the VF1-binding site 4 bp closer to the *conII* promoter (Fig. 1).

A selection plasmid based on the VF1-binding site in the *xisA* upstream region was also constructed. The 251-bp *PstI*-*BamHI* fragment from pAM709 containing the three adjacent VF1-binding sites in the *xisA* upstream region was cloned into the same sites of pNN396. The *NotI*-*HindIII* fragment containing the *conII* promoter regulated by the three *xisA* binding sites was subsequently cloned into the same sites of pNN388 to form pAM993.

**Selection conditions.** To establish the antibiotic sensitivities of the selection plasmids, JM107 strains containing pAM993, pAM994, pAM1156, or pNN388 were transformed with pSE380 (5), the vector used to construct our expression library. Their resistances to spectinomycin (SP) and streptomycin (SM) were determined empirically. Luria-Bertani (LB) agar plates were supplemented with four antibiotics: chloramphenicol (CM, 34  $\mu\text{g/ml}$ ) for the selection plasmid, ampicillin (AP, 50  $\mu\text{g/ml}$ ) for the library vector, and a 1:1 ratio of SP and SM (abbreviated as SP+SM) at combined concentrations of 0, 20, 30, 40, 60, 80, 100, and 150  $\mu\text{g/ml}$  to determine the activity of the *aadA* gene for each plasmid. A 1:1 ratio of SP and SM was used to reduce the number of spontaneous mutants that were resistant to both antibiotics. In addition, IPTG (1 mM) was added for induction of  $P_{trc}$  on pSE380.

**Construction of the pSE380 expression library.** DNA was isolated from PCC 7120 filaments as described previously (19) and was partially digested with *Sau3AI*. Gel-purified *Sau3AI* fragments with an average size of 4.4 kb (3.5 to 6.5 kb) were ligated into the *BamHI* site of pSE380. pSE380 is an expression vector, selectable by AP resistance, that has a superlinker directly downstream of a  $P_{trc}$  promoter regulated by *lacI<sup>r</sup>* carried on the plasmid (5). *E. coli* DH5 $\alpha$ MCX was transformed by electroporation with the ligation mixture and spread on LB plates supplemented with AP (100  $\mu\text{g/ml}$ ). The primary expression library before amplification contained approximately 34,300 colonies. Analysis of plasmids from a random sample of 24 colonies showed that 67% contained apparently random inserts of the expected size range. The size of the PCC 7120 genome is 6.4 Mb (3); therefore, the library represents approximately 14 genome equivalents. The primary expression library was amplified and stored as purified plasmid DNA by collecting the cells from the original LB-AP plates and preparing plasmid DNA from the pooled colonies by an alkaline lysis procedure (2).

**Isolation of VF1-like proteins from *E. coli* clones.** After initial selection, a pool of 106 clones (see Results) was tested for the production of VF1 DNA-binding activity. A 5-ml culture of each clone was grown to mid-log phase and induced with IPTG (1 mM). The 106 cultures were pooled, and the cells were collected by centrifugation at  $10,000 \times g$  for 15 min. The cell pellet was suspended in 25 ml of extraction buffer (50 mM Tris-HCl [pH 7.5], 1 mM EDTA, 10% glycerol, 0.5% Triton X-100, 2 mM dithiothreitol [DTT], 1 mM phenylmethylsulfonyl fluoride). The cell suspension was split into two equal portions, and each sample was homogenized with an equal volume of glass beads (0.1-mm diameter) four times for 1 min each at 0 to 4°C in a homogenizing mill (Braun). The cell lysate was cleared by centrifugation at  $31,000 \times g$  for 25 min and then  $142,000 \times g$  for 90 min at 4°C. The supernatant was fractionated by loading it onto a 3-ml heparin-Sepharose CL6B (Pharmacia LKB Biotechnology Inc., Piscataway, N.J.) column equilibrated with buffer A (50 mM Tris-HCl [pH 7.5], 0.1 mM EDTA, 0.1% Triton X-100, 10% glycerol, 0.1 M ammonium

sulfate, 2 mM DTT, 0.5 mM phenylmethylsulfonyl fluoride) as described previously (7). The column was washed with 45 ml of buffer A and eluted with a 44-ml linear gradient of 0.1 to 1.0 M ammonium sulfate in buffer A. The column fractions (2 ml) were collected, and 1  $\mu\text{l}$  of each fraction was assayed for VF1 DNA-binding activity by mobility shift assays with both *glnA* and *xisA* upstream probes.

For the identification of individual clones, the original 106 colonies were grouped into smaller pools and a simplified extraction procedure was used to obtain samples for mobility shift assays. Individual 5-ml cultures were grown and induced as before and then pooled in small groups or assayed individually. The cell pellet of each sample was suspended in 5 or 10 ml of extraction buffer, and the cells were broken by ultrasonic treatment at 4°C for 3 min at a 50% duty cycle, output control setting 5 (Branson Sonifier Cell Disruptor 350 equipped with a tapered microtip). The cell lysate was cleared as before. Ammonium sulfate fractionation was used to partially purify VF1-like proteins to avoid running heparin-Sepharose columns for a large number of samples. Solid ammonium sulfate was added to 20% saturation, and the protein precipitate was removed by centrifugation for 35 min at  $35,000 \times g$  at 4°C. Solid ammonium sulfate was then added to the supernatant to 60% saturation. The protein precipitate was collected by centrifugation as before. The precipitate from the 20 to 60% cut was suspended in 0.5 to 2 ml of extraction buffer, and 1 to 2  $\mu\text{l}$  was assayed for VF1-like DNA-binding activity.

**Partial purification of VF1 from PCC 7120 vegetative cells.** PCC 7120 vegetative cell filaments were harvested from late-exponential-growth-phase 2-liter cultures, and VF1 was extracted and partially purified by heparin-Sepharose CL6B column chromatography as previously described (7). One of the peak VF1-containing column fractions was used as the VF1 standard in mobility shift assays.

**Analysis of DNA-protein complexes with mobility shift assays.** Mobility shift assays (14, 15) were performed as described by Ausubel et al. (2) and Chastain et al. (7). The DNA probes were a 252-bp *XbaI*-*HindIII* fragment of the *xisA* upstream region from pAM709 and a 35-bp synthetic oligonucleotide containing the 24-bp VF1-binding site in the *glnA* promoter region (Fig. 1). The DNA fragments were 3' end labeled by using [ $\alpha$ - $^{32}\text{P}$ ]dCTP and the Klenow fragment of DNA polymerase I by a standard protocol (2). Labeled DNA probes (0.2 to 0.5 ng) with an average specific activity of  $10^7$  cpm/ $\mu\text{g}$  were incubated with 1 or 2  $\mu\text{l}$  of partially purified protein extract in binding buffer (4 mM Tris-HCl [pH 8.0], 12 mM *N*-2-hydroxyethylpiperazine-*N'*-2-ethanesulfonic acid [pH 7.9], 12% glycerol, 60 mM KCl, 0.5 mM EDTA, 1 mM DTT) containing 0.5  $\mu\text{g}$  of poly(dI-dC) in a final volume of 20  $\mu\text{l}$ . The mixture was incubated at room temperature for 15 min and loaded on a 5% high-ionic-strength polyacrylamide gel (30:1 acrylamide-bisacrylamide, 50 mM Tris-HCl [pH 8.5], 380 mM glycine, 2 mM EDTA, 2.5% glycerol). Electrophoresis was performed at 25°C in a Tris-glycine running buffer (50 mM Tris-HCl [pH 8.5], 380 mM glycine, 2 mM EDTA) for 2 to 2.5 h at 30 mA. The gels were dried and autoradiographed.

**Sequencing of the *bifA* clone and computer analysis.** The insert from the pAM995 *bifA* clone was subcloned from the pSE380 vector as a *SmaI*-*PstI* fragment into pBluescript SK(+) to form pAM996. Double-strand sequencing of the entire insert was done by the Advanced DNA Technologies Laboratory in the Department of Biology, Texas A&M University. Sequential deletions were made with exonuclease III along one strand by using an Erase-a-Base kit

(Promega), and the nested deletions were sequenced by the dideoxy-chain termination method (34). The second strand and gaps in the first strand were sequenced by using a series of complementary synthetic oligonucleotide primers. Sequencing data were analyzed by using the sequence analysis software package of the Genetics Computer Group (GCG) (9) and the NCBI GenBank BLAST e-mail server (1).

**Nucleic acid methods.** Standard methods were used for *E. coli* plasmid DNA isolation and transformation by calcium chloride treatment or electroporation (2). Restriction endonucleases and other DNA-modifying enzymes were used according to the manufacturer's recommendations or standard methods (2). Vegetative cell DNA and heterocyst DNA (isolated from purified heterocysts) were prepared as previously described (19). The Southern transfer was from a 0.5% agarose gel run in TBE buffer (2) to a GeneScreen nylon membrane (DuPont, NEN Research Products) with 10× SSPE (1× SSPE is 0.18 M NaCl, 10 mM NaH<sub>2</sub>PO<sub>4</sub>, and 1 mM EDTA [pH 7.4]) (27). The *bifA* hybridization probe was labeled with a random primer kit from Boehringer Mannheim. Southern hybridization was performed in 5× SSPE–2% sodium dodecyl sulfate at 65°C, and filters were washed in 0.5× SSPE–0.2% sodium dodecyl sulfate at 65°C.

**Nucleotide sequence accession number.** The nucleotide sequence reported here has been deposited in GenBank under accession number L10036.

## RESULTS

**Strategy for genetic selection.** We used a genetic selection method designed to facilitate the cloning of genes encoding DNA-binding proteins from expression libraries (12). A single-copy-number plasmid, pNN388, carrying an *aadA* gene under the control of its natural weak promoter, P<sub>*aadA*</sub>, was used to construct the selection plasmids. The *aadA* gene encodes aminoglycoside adenylyltransferase, which adenylylates antibiotics such as SP or SM and renders them ineffective. The selection plasmids were constructed by subcloning the strong *conII* promoter (12) together with a VF1-binding site downstream and antisense to the *aadA* gene such that the promoters are transcribed convergently. The *conII* promoter is a constitutively expressed derivative of the *tac* promoter in which the *lac* operator has been replaced with a polylinker. This allows the positioning of a specific DNA-binding sequence near the transcriptional start site so that it can function as an artificial operator.

Transcription from the *conII* promoter, if not regulated by DNA-protein interactions at the artificial operator, would interfere with the transcription from the *aadA* promoter and result in relative sensitivity of the host cells to SP+SM. Protein binding at the artificial operator would repress transcription from the *conII* promoter and alleviate interference of *aadA* transcription, resulting in a relatively high level of SP+SM resistance. The different level of SP+SM tolerance permitted by the repressed and the derepressed *conII* promoters allows for enrichment of clones encoding sequence-specific DNA-binding proteins from large expression libraries. Elledge et al. (12) have demonstrated that the selection scheme will function for a variety of known prokaryotic and eukaryotic DNA-binding proteins.

**Enrichment for clones encoding VF1 DNA-binding activity.** We constructed three selection plasmids, pAM993, pAM994, and pAM1156 (Table 1), based on VF1-binding sites in the *glnA* promoter (32) and the *xisA* upstream region (7). These sites have been characterized by DNase I footprint analysis. In vitro studies showed that VF1 binds with

higher affinity to the *glnA* promoter than to the *xisA*, *rbCL*, and *nifH* upstream regions and that VF1 forms multiple complexes with the *xisA* upstream region by binding to three tandem sites. We chose the *glnA* and *xisA* binding sites as good candidates for use as artificial operators in the in vivo genetic selection.

The SP+SM resistance of each selection plasmid was compared with that of the unmodified pNN388. The SP+SM resistance was determined under the same conditions that were to be used for enrichment of the expression library; JM107 containing each selection plasmid along with pSE380 was plated on LB agar containing AP (50 µg/ml), CM (34 µg/ml), IPTG (1 mM), and a range of SP+SM concentrations. pNN388, which does not contain the antisense *conII* promoter, provided the maximum possible resistance to SP+SM from its *aadA* gene. pNN388 allowed normal growth at SP+SM concentrations up to 100 µg/ml. The range of SP+SM concentrations at which each selection plasmid allowed no, or very poor, growth compared with that of the control strain containing pNN388 was identified. The SP+SM selection conditions were determined to be 30 to 100 µg/ml for pAM993, 75 to 100 µg/ml for pAM1156, and 40 to 100 µg/ml for pAM994.

Our initial attempt to enrich for clones encoding VF1 DNA-binding activity from the pSE380 expression library was with pAM993. JM107 harboring pAM993 was transformed with the PCC 7120 expression library by electroporation and plated on LB agar containing AP, CM, IPTG, and SP+SM at 100 µg/ml. Colonies of various sizes appeared on these plates after 1 day of growth. Attempts to identify clones by characterizing their resistance to SP+SM in the presence and absence of IPTG, which would induce the promoter on pSE380, were unsuccessful. Several larger colonies were selected to determine if VF1 DNA-binding activity could be detected in cell extracts by using mobility shift assays. VF1-binding activity was not detected in any of the clones.

Our first selection plasmid based on the *glnA* binding site, pAM1156, provided a relatively narrow SP+SM selection range, presumably because of poor transcriptional interference by the regulated *conII* promoter. Attempts to use this plasmid to enrich for clones from the expression library were unsuccessful. pAM1156 was subsequently modified by deleting 4 bp between the *conII* promoter and the VF1-binding site to produce pAM994 (Fig. 1). The SP+SM selection range for pAM994 was significantly wider than that for pAM1156, as stated above.

JM107 harboring pAM994 was transformed with the PCC 7120 expression library by electroporation and plated on LB agar containing AP, CM, IPTG, and SP+SM at 70, 85, and 100 µg/ml. We used three different SP+SM concentrations for enrichment because we were unable to predict the degree of repression of the *conII* promoter by VF1 binding to its binding site. Only full repression of the *conII* promoter would provide resistance to SP+SM at 100 µg/ml. A total of 20,000 transformants were plated. Colonies appeared on all plates, but with decreased frequency at the higher SP+SM concentrations. Analysis of several individual clones from the highest SP+SM concentration for VF1-binding activity did not identify any positive clones and showed that the genetic selection was not definitive in discriminating positive clones from background clones. We selected 106 clones for further study: 84 from 100-µg/ml, 10 from 85-µg/ml, and 12 from 70-µg/ml SP+SM selection plates. We picked more colonies from the highest SP+SM concentration under the assumption that VF1 binding would efficiently suppress the

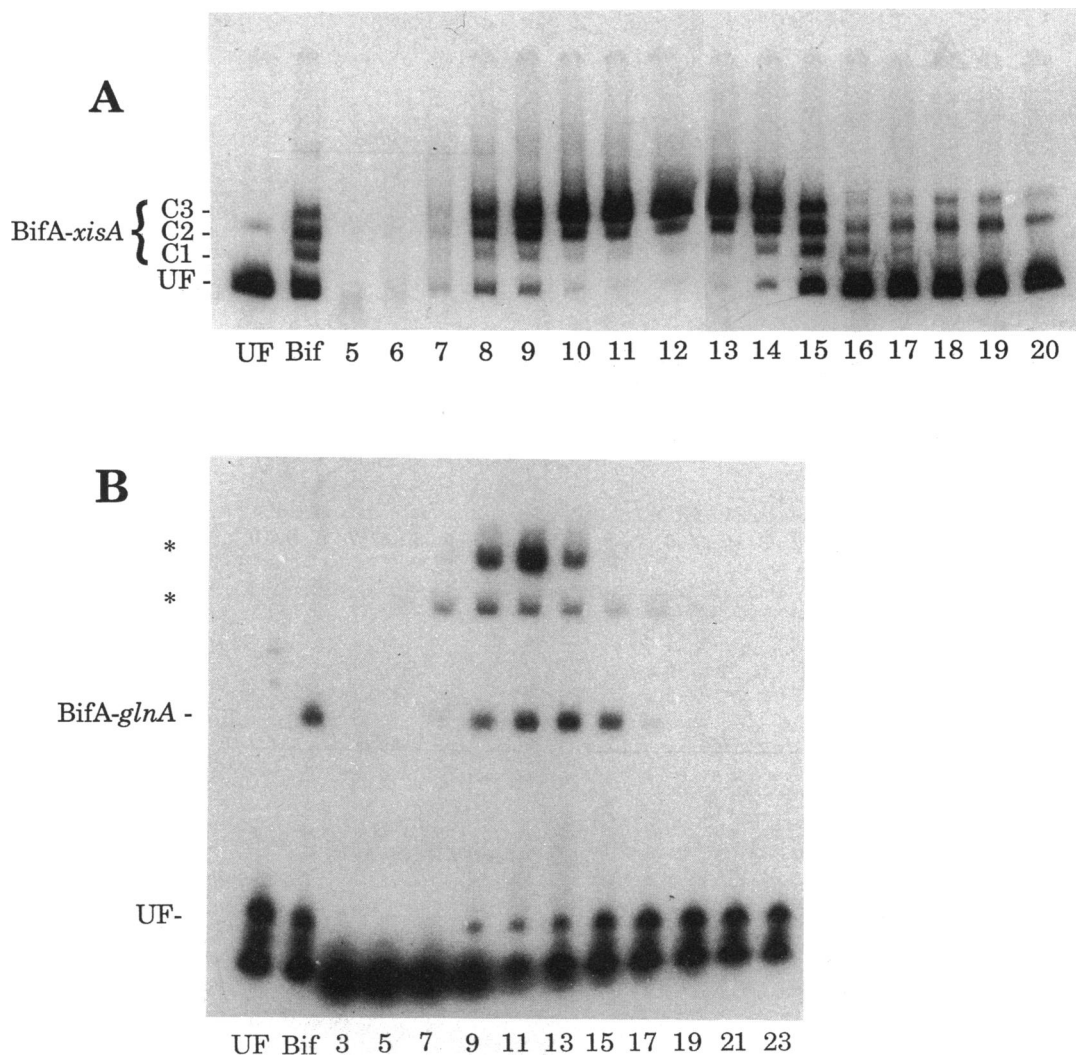


FIG. 2. Mobility shift assays showing BifA binding to *xisA* and *glnA* probes in extracts from a pool of 106 *E. coli* clones. Samples (1  $\mu$ l) of selected heparin-Sepharose column fractions from a linear 0.1 to 1.0 M ammonium sulfate gradient were equilibrated with 10,000 cpm of labeled DNA fragment and analyzed for mobility shift on a 5% high-ionic-strength polyacrylamide gel. A VF1-containing heparin-Sepharose fraction from PCC 7120 vegetative cell extracts was used as the standard. (A) Labeled *xisA* upstream fragment from pAM709. Three complexes (C1, C2, and C3) are formed between BifA and the *xisA* fragment. (B) Labeled synthetic 35-bp fragment containing the VF1-binding site on the *glnA* promoter. \*, complexes of unknown composition; UF, unbound DNA fragment; Bif, PCC 7120 vegetative cell VF1 standard. Lane numbers refer to heparin-Sepharose column fractions.

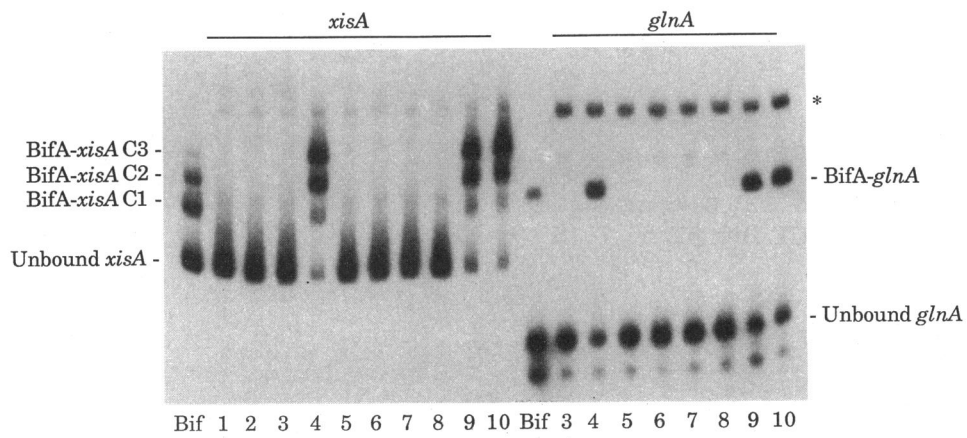
*conII* promoter and allow greater enrichment of positive clones at the higher drug concentration; this turned out not to be the case. Clones encoding VF1 DNA-binding activity were apparently present at higher frequencies at the lower SP+SM concentrations.

**Identification of clones encoding VF1 DNA-binding activity.** VF1 DNA-binding activity was identified in partially purified protein fractions obtained from a pooled protein sample extracted from the 106 clones. Heparin-Sepharose column fractions from the combined extracts were analyzed by mobility shift assays using *xisA* (Fig. 2A) and *glnA* (Fig. 2B) probes. A VF1-containing fraction from PCC 7120 vegetative cell extracts served as a standard for comparison. The elution profile and the position of the shifted bands for both probes were similar to those obtained with PCC 7120 extracts and showed the expected three DNA-protein complexes for the VF1-*xisA* interaction (7). As was originally

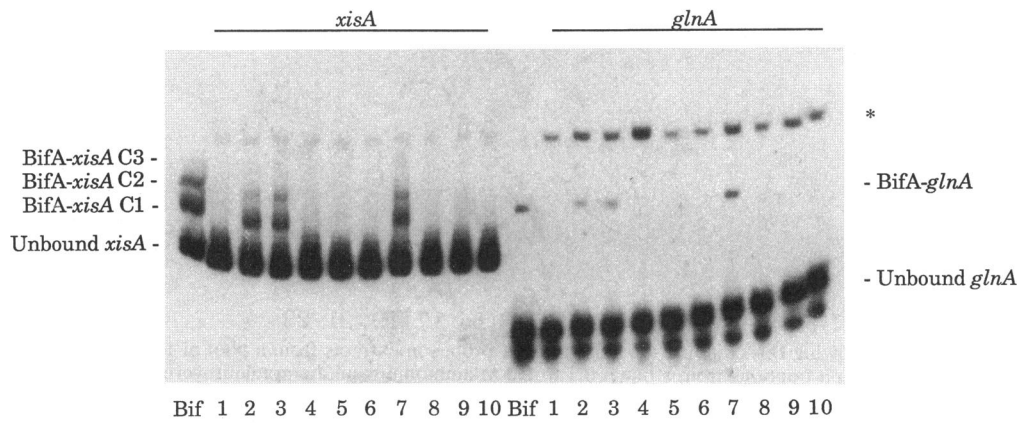
observed for VF1-*xisA* interactions, a shift to the slower-migrating complexes was seen in peak fractions (7). This result showed that at least one clone expressing VF1 DNA-binding activity was present among the 106 isolates.

The individual clones responsible for the VF1 DNA-binding activity detected in the combined extract were identified by progressively subdividing the 106 clones into smaller groups. The 106 clones were first divided into 10 groups of 10 or 11 clones each. For each group, a combined protein extract was partially purified by 20 to 60% ammonium sulfate fractionation and analyzed for VF1 activity by mobility shift assays. Three of the 10 groups showed VF1 activity with both *xisA* and *glnA* probes, indicating the presence of at least 3 positive clones among the 31 colonies that constituted the three positive groups (Fig. 3A). The 31 clones were divided into 10 groups of three or four each. Among these 10 groups, 3 showed VF1 DNA-binding activ-

**A**



**B**



**C**

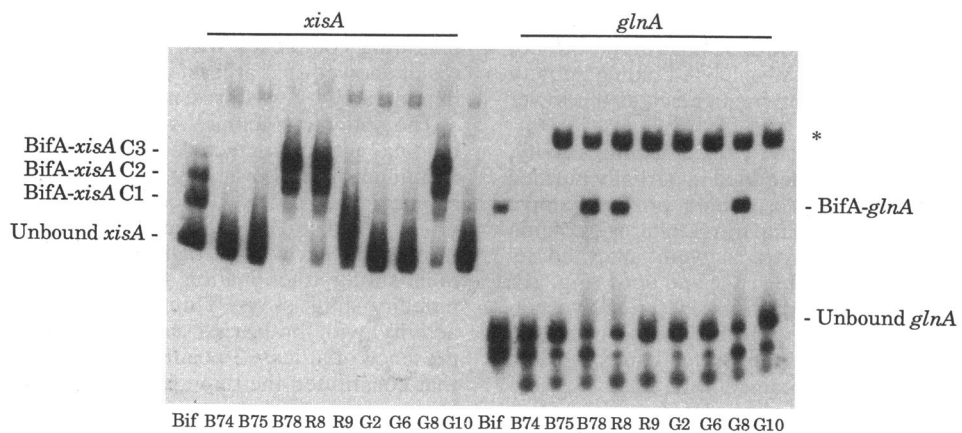


FIG. 3. Identification of *bifA* clones. *xisA* and *glnA* probes were used in mobility shift assays of ammonium sulfate-precipitated protein extracts from subgroups of the original 106 individual clones. The *xisA* probe is to the left and the *glnA* probe is to the right in all three panels. Mobility shift assays used 10,000 cpm of each probe and 2  $\mu$ l of ammonium sulfate-fractionated protein extract. (A) First set of 10 subgroups contained 10 or 11 clones per group. (B) Positive groups 4, 9, and 10 from panel A, which together contributed 31 clones, were subgrouped into 10 groups of three or four clones each. (C) Nine clones from the positive groups 2, 3, and 7 in panel B were individually analyzed. Lane numbers in panels A and B refer to the group numbers; those in panel C refer to the individual clones. Labels are as in Fig. 2.

ity (Fig. 3B). The nine clones from the three positive groups were then assayed individually, and three positive clones were identified (Fig. 3C), one from each of the SP+SM selection concentrations. The three clones were analyzed by digestion with several restriction enzymes. All three clones showed the same size insert fragment (4.3 kb) and identical restriction digest patterns with *TaqI*, *MspI*, and *SmaI-DraI*, which produced a significant number of small fragments. We concluded that the three clones were identical, which is consistent with our use of an amplified expression library, and we selected one clone, pAM995, for further studies.

**Sequence analysis.** The nucleotide sequence of the entire 4.3-kb insert in pAM995 [subcloned in pBluescript SK(+) to make pAM996] revealed the presence of three complete and two partial open reading frames (ORFs) (Fig. 4A). ORF1, ORF2, and ORF3 are transcribed in the opposite direction relative to ORF4 and ORF5. The predicted protein sequences of the five ORFs were compared with the SwissProt data base by using the WORDSEARCH program of the GCG package. The 279-bp partial ORF1 encodes a 93-amino-acid

polypeptide that showed 63% identity and 78% similarity to the N-terminal region of the ORF60-5-encoded protein in *Synechocystis* sp. strain PCC 6803 (8). The deduced ORF60-5-encoded protein is thought to be homologous to a family of peripheral ATP-binding subunits of membrane-protein complexes involved in metabolite transport. The 630-bp ORF2 encodes a 23-kDa protein with a pI of 6.59 that showed 50.5% identity and 67.3% similarity to the HisB protein of *Azospirillum brasilense* (13). The 795-bp ORF3 encodes a 28-kDa protein with a pI of 6.13 that showed 48.6% identity and 70.2% similarity to the EnvM protein of *Salmonella typhimurium* (38). The 672-bp ORF4 encodes a 25-kDa basic protein with a pI of 9.70 that showed 77% identity and 87% similarity to the *Synechococcus* NtcA protein, which is thought to encode a transcriptional activator of genes subject to nitrogen control (39). The peptide deduced from the 498-bp incomplete ORF5 did not show significant similarity to other known proteins.

We suspected that either ORF4 or the incomplete ORF5 or both might encode VF1 DNA-binding activity. ORF4 and ORF5 were subcloned from pAM996 into pSE380 together and separately (Fig. 4B). These clones and several exonuclease III-generated deletions of pAM996 (Fig. 4C) were assayed for VF1 DNA-binding activity by using mobility shift gels. These experiments showed that VF1 DNA-binding activity was encoded by ORF4, which we have named the *bifA* (binding factor A) gene (Fig. 4). A direct correspondence between the *bifA* gene product, BifA, and the PCC 7120 VF1 protein cannot be established since VF1 has not been purified and its amino acid sequence is not known.

A Southern blot of *EcoRI*- and *XbaI*-digested PCC 7120 vegetative cell and heterocyst DNA was hybridized with a *bifA* probe. Single fragments of 6.3 and 10 kb were identified in the *EcoRI* and *XbaI* digests, respectively (Fig. 5). No differences were detected between vegetative cell and heterocyst DNA. The *bifA* gene was mapped to the region between approximately 4.9 and 5.3 Mb on the PCC 7120 chromosome (3) by hybridization to Southern blots of pulse-field gels; the *bifA* probe hybridized to *BlnI* (*AvrII*) fragment A and *PstI* fragment D (16). These results show that *bifA* is a single-copy gene and that it is not closely linked to other mapped genes known to be involved in heterocyst differentiation or nitrogen fixation.

The nucleotide sequence and deduced amino acid sequence of *bifA* (ORF4) are shown in Fig. 6. A putative ribosome binding site (RBS) and the restriction sites used for making the ORF subclones are marked.

Computer analysis using the NCBI GenBank BLAST e-mail server (1) showed significant similarity between BifA and a set of prokaryotic regulatory proteins related to *E. coli* CRP. BifA showed 37% identity and 69% similarity to *E. coli* CRP (33), 37% identity and 74% similarity to *Rhizobium meliloti* FixK (4), and 27% identity and 54% similarity to *E. coli* Fnr (36). A cyanobacterial member of the CRP family, the *Synechococcus* CysR protein, which is involved in the regulation of sulfur assimilation (26), was identified in these searches and showed 32% identity and 53% similarity to

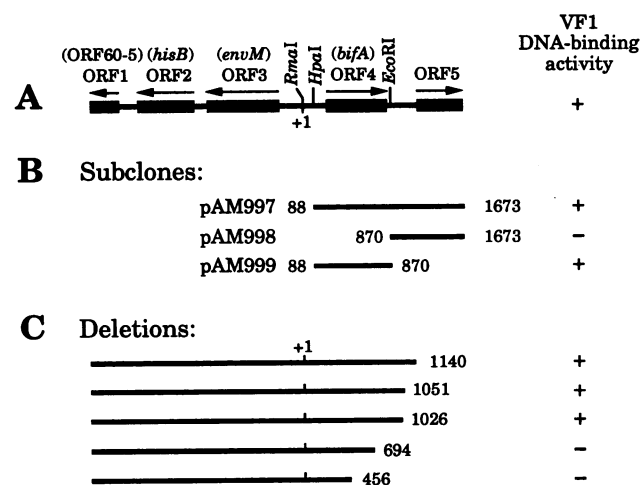


FIG. 4. Subcloning and deletion analyses of the *bifA* clone pAM995. (A) Organization of the three complete and two partial ORFs in the 4.3-kb pAM995 insert. Numbers are nucleotide positions relative to the first residue of the *RmaI* site, which was arbitrarily chosen as nucleotide +1. Arrows indicate the orientations of the ORFs. ORF1 and ORF5 are incomplete. (B) ORF subclones in pSE380. ORF4 and the incomplete ORF5 were obtained by cloning a 1,680-bp *HpaI* fragment from pAM996 into the *SmaI* site of pSE380 to make pAM997. Plasmid pAM998, which lacks ORF4, was obtained by deleting an *EcoRI* fragment from pAM997. ORF4 (*bifA*) was obtained by cloning an *EcoRI* fragment from pAM997 into the *EcoRI* site of pSE380 to make pAM999. (C) Nested deletions of pAM996. Selected deletions generated for DNA sequence analysis by exonuclease III digestion were used for localization of *bifA*. The bars show DNA remaining in the deletion plasmids. VF1 DNA-binding activities in protein extracts of strains containing the subclone and deletion plasmids were analyzed by mobility shift assays using a labeled *xisA* upstream fragment.

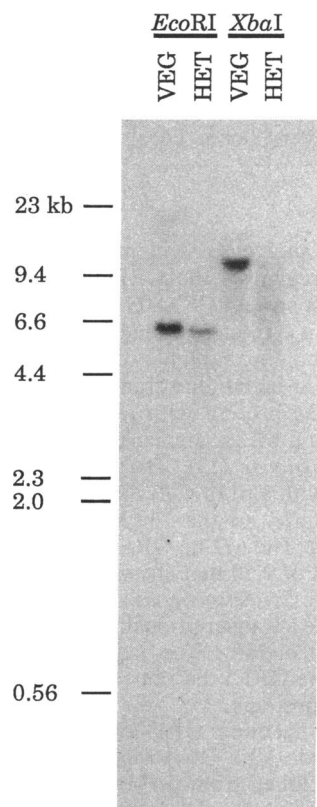


FIG. 5. *bifA* Southern analysis. A Southern blot of PCC 7120 vegetative cell (VEG) and heterocyst (HET) DNA digested with *EcoRI* and *XbaI* was hybridized with a random primer-labeled *EcoRI* fragment (834 bp) from pAM997 that contains the *bifA* gene. The heterocyst DNA is partially degraded and fails to clearly show the 10-kb *XbaI* fragment. The numbers on the left are the sizes of molecular weight markers.

**BifA.** Recently, the sequence of a cyanobacterial regulatory gene, *ntcA*, related to the CRP family was published (39). Alignment of the PCC 7120 BifA and the *Synechococcus* NtcA proteins by using the GCG BESTFIT program showed 77% identity and 87% similarity between the amino acid sequences.

A comparison of the protein sequences for some of these regulatory proteins, using the GCG PILEUP program, is shown in Fig. 7. A helix-turn-helix motif in the carboxy-terminal domain, which has been shown to be involved in specific DNA binding for CRP (10, 35), and the four glycine residues in the amino-terminal domain, which are associated with a  $\beta$ -roll structure essential for the regulatory properties of CRP (41), are potentially present in all five proteins. The cysteine residues present in the amino terminus of Fnr (37), presumably the binding site for metal ions that constitute the oxygen sensor, are not found in BifA. Residues involved in cyclic AMP binding in CRP (28, 41) are not well conserved in BifA.

## DISCUSSION

Relative to other organisms such as *Klebsiella*, *Rhizobium*, and *Azotobacter* spp., little is known about the regulation of nitrogen fixation in cyanobacteria. We have been studying the heterocyst-specific rearrangement of *nif*

operons in *Anabaena* sp. strain PCC 7120. As part of these studies, we have isolated the gene encoding a potential regulatory protein, BifA, which interacts with several genes that are differentially expressed in PCC 7120 vegetative cells and heterocysts. The *bifA* gene was cloned by using an in vivo selection scheme (12) with the anticipation that a reverse genetics approach could then be used to develop a better understanding of the role of *bifA* in PCC 7120 gene regulation.

BifA may be identical to VF1, a protein identified in vegetative cell extracts of PCC 7120 by its interaction with the *xisA* gene (7). *xisA* encodes a developmentally regulated site-specific recombinase that is required for the formation of the *nifHDK* operon in heterocysts (6, 21, 25). BifA and VF1 are indistinguishable in their chromatographic and DNA-binding properties. They show similar elution profiles on heparin-Sepharose columns and identical mobility shifts when assayed with the *glnA* 35-bp oligonucleotide and the *xisA* upstream fragment that forms three characteristic complexes. The VF1-binding site in the *xisA* upstream region overlaps with a regulatory region involved in *xisA* expression (6, 7). Therefore, it was originally thought that VF1 was a vegetative cell-specific protein that might be involved in suppression of the *xisA* recombinase and excision of the *nifD* element. However, VF1's putative regulatory role is not so simple because it binds not only to the upstream region of the *xisA* gene but also to the *glnA*, *rbcL*, and *nifH* promoters (7).

Little is known about the regulation of VF1. Preliminary experiments showed that the DNA-binding activity of VF1 was nearly absent in cell extracts from early-exponential-growth-phase cultures and increased in mid-exponential-phase to stationary-phase cultures and that it also persists in heterocysts (32). These observations could indicate that VF1 binding activity may be influenced by the metabolic state of the cells, the growth rate of the culture, or environmental factors such as the availability of combined nitrogen or light intensity. Clearly, analysis of strains containing mutations in the *bifA* gene will be required to establish its role in PCC 7120.

Initial efforts to clone *bifA* from an expression library by probing protein bound to replica filters with labeled specific DNA ligands failed. Reconstruction experiments showed that the VF1 protein loses its DNA-binding activity when bound to a nitrocellulose filter. Preliminary attempts to obtain purified VF1 for protein microsequencing were also unsuccessful. A genetic selection method which depends on the protein's in vivo DNA-binding properties was therefore used to identify the *bifA* gene in a plasmid expression library.

The success of this cloning strategy depends on the positioning of a well-characterized DNA-binding site that has high affinity for the protein of interest in the selection plasmid. Placing the binding site close to the *conII* promoter enhances its repression in the presence of the specific DNA-binding protein, allowing for increased expression of *aadA* and greater SP+SM resistance (11, 12). In addition, the engineered *conII* promoter must effectively interfere with *aadA* expression in the absence of binding protein. pAM994, a derivative of pAM1156 in which the VF1-binding site was closer to the *conII* promoter and the position of the predicted transcription start site was at a pyrimidine nucleotide, gave an enhanced SP+SM selection range.

After our initial attempts at applying the genetic selection scheme, it became clear that the method could not definitely identify *bifA* clones. In the first screens of the expression



```

CTAGCTAAAAATAGCAAAGTTGGGTATCATTATGAACAAATAACCATAAAGTTGTACGTTTGTAGTATAGGAAAGTACAGAAAGGTTAACGGTGCTTTAT 100
RmaI
TGATTTTTCAGGTGATTCTTAGGTAATTAGTTTTAGTTTTTTGTTATTCGGCATTGGGTAGGGGAAGGGAGATGATCGTGACACAAAGATAAGGCCT 200
HpaI
RBS M I V T Q D K A L
AGCAAATGTTTTTCGTCAGATGGCAACCGGAGCTTTTCCTCCTGTTGTGCGAAACGTTTGAACGCAATAAAACGATCTTTTTCTGGCGATCCTGCCGAA 300
A N V F R Q M A T G A F P P V V E T F E R N K T I F F P G D P A E
CGAGTCTACTTTCTTTTGAAGGGGCTGTGAAACTTCCAGGGTGTACGAGGAGGAGAAGAGATTACAGTAGCACTACTACGGGAAAATAGCGTTTTTG 400
R V Y F L L K G A V K L S R V Y E A G E E I T V A L L R E N S V F G
GTGTCCTGTCTTTGTTGACAGGAAACAAGTCGGATAGGTTTTACCATGCGGTGGCATTACTCCAGTAGAATTGCTTTCTGCACCAATGAACAAGTGA 500
V L S L L T G N K S D R F Y H A V A F T P V E L L S A P I E Q V E
GCAAGCACTGAAGGAAAATCCTGAATTATCGATGTTGATGCTGCGGGTCTGTCTTCGCGGATTCTACAACAGAGATGATGATTGAAACCTTAGCGCAC 600
Q A L K E N P E L S M L M L R G L S S R I L Q T E M M I E T L A H
CGAGATATGGGTTTCGAGATTGGTGAGTTTTCTGTTAATCTCTGTCGTGATTTTGGTGTTCCTTGTGCGATGGAATCACAATTGATTTAAAGTTATCTC 700
R D M G S R L V S F L L I L C R D F G V P C A D G I T I D L K L S H
ATCAGCGCATCGCCGAAGCAATTGGCTCTACTCGCGTTACTGTTACTAGGCTACTAGGGATTTCGCGGAGAAAAGATGATTTCCATCCACAAAAGAA 800
Q A I A E A I G S T R V T V T R L L G D L R E K K M I S I H K K K
GATTACTGTGCATAAACCTGTGACTCTCAGCAGACAGTTCACTTAAAGAAGAATAGCAGGAGTCAGAATTCAGAATGTGGCTAATATTAGCACAGGCTA 900
I T V H K P V T L S R Q F T *
EcoRI
AATCCTACTCATTATAGATTTAGTTGGTTTCTTCAATCCGTTTATTTAACTCAAAGTGACGGTGAACCCATACTGAAATCTGGCTCTGCCTTCTGCAC 1000
TCTTTAAACAATTGGAACCATCAGAGGTGGTACGTGTTATATTTTTTCTCAATGACACCCATAGTTAGAAGTAAATCGAAAATCTTGAGTAACTTCAGCTA 1100

```

FIG. 6. Nucleotide sequence of a 1,100-bp region containing *bifA* and its deduced amino acid sequence. A potential ribosome binding site (RBS) and the restriction sites shown in Fig. 4 are underlined and marked below the sequences. Nucleotide numbering starts at the *RmaI* site as in Fig. 4.

library, large fast-growing colonies were selected from plates containing high drug concentrations. However, none of these putative *bifA* clones produced VF1 DNA-binding activity. We concluded that the selection scheme could be used to enrich for *bifA* clones but that we would need to develop a screening procedure to identify the true clones from the background colonies that grew on the selection plates.

We were able to identify *bifA* clones within a pool of a large number of selected colonies by analyzing the combined protein extracts for VF1 DNA-binding activity. However, this approach required that the specific DNA-protein interaction could be detected in *E. coli* protein extracts even when only a few percent of the tested colonies were actually producing the desired protein.

Our experience in cloning the *bifA* gene showed that the degree of repression, because of protein binding to the artificial operator, of the activity of the *conII* antisense promoter is somewhat unpredictable. One positive *bifA* clone was found at each of the three SP+SM concentrations used for selection, even though fewer colonies from the lower antibiotic concentrations were screened for VF1 binding activity. We conclude that the percentage of positive clones among the background colonies may not be proportional to the selection stringency. In our experiments, the genuine *bifA* clones provided only moderate levels of SP+SM tolerance and represented a very small fraction of the colonies that grew at the highest SP+SM concentrations. Although we did not quantitatively establish the frequency of *bifA* clones in the original expression library and for each of the selection conditions, if we assume that the library was relatively random and that our frequency of positive clones was representative of the true frequency, then the in vivo

genetic selection provided a significant enrichment for *bifA* clones. Several factors may influence the degree of enrichment for a particular DNA-binding protein, assuming that the protein can be efficiently expressed in *E. coli* from a single DNA fragment present in an expression library. These factors include the affinity of the DNA-binding protein for its specific binding sequence in vivo and the degree of transcriptional interference provided by the regulated *conII* promoter in the presence and absence of the binding protein. Uncertainties associated with the properties of the regulated *conII* promoter suggest that several variations of the selection plasmid should be constructed and screened at a variety of SP+SM concentrations. An additional potential problem with this strategy is that *E. coli* proteins may interfere with the binding of the target protein or, more likely from our observations, cause a high background of SP+SM-resistant colonies because of interaction with the artificial operator binding site in the selection plasmid. We have observed a variety of DNA-protein interactions with *E. coli* protein extracts in mobility shift assays using several different PCC 7120 DNA probes, including the 35-bp *glnA* probe used for Fig. 2 and 3.

What role BifA (VF1) may play in PCC 7120 remains unclear. BifA shows strong similarity to the recently reported cyanobacterial regulatory protein NtcA, isolated from *Synechococcus* sp. strain PCC 7942 (39, 40). NtcA is proposed to be a transcriptional activator required for full expression of genes subject to ammonium repression in *Synechococcus* sp. strain PCC 7942. Wild-type *ntcA* is required for the full expression of a number of genes involved in nitrogen assimilation, including those for nitrite reductase, nitrate reductase, and glutamine synthetase. BifA's interaction with the *glnA* and *nifH* genes are consis-

```

Fnr (25) ISQLCIPFTL NEHELDQLDN IIERKKPIQK GQTLFKAQDE LKSLYAIRSG
FixK (14) VEHLG.PAPM SGRPLVA... ..TYKP GREIYAQODL NDKCYQVSTG
Crp (1) MVLGKPKQTPD TLEWFLSHC... ..HIHKYPS KSTLIHQQEK AETLYIVKQ
NtcA (1) .MLANENSLT TMFRELGSRK LPLQIEQFER GRTIFFPQDP AERVYLLVKQ
BifA (1) MIVTQDKALA NVFRQMATGA PFPVVFETFER NKTIFFPQDP AERVYLLVKQ
pileup m.ll..pall ..frlla.g. ....iet.er gktiff.gdp aer.Y.v.vkg

Fnr (75) TIKSYTITEQ *GDEQITGFHL AGDLVG...F DAIGSGHPS FAQALETSMV
FixK (54) AVRIYRLLSD GRRQVVSFHL PGEMFG...F EA..GSNHSF FAEAITETTL
Crp (47) SVAVLIKDEE QKEMILSYLN QGDFIGELGL FEEOQ.ERSA WVRKTACEV
NtcA (50) AVKLSRVYES GEEITVALLR ENSVFGVLSL LTGQRSDRFY HAVAFTPVQL
BifA (51) AVKLSRVYES GEEITVALLR ENSVFGVLSL LTGQRSDRFY HAVAFTPVEL
pileup avkl.rvye. Gee..v..l. eg.vfg.l.l 1.gg.sdr.y avAftpvvel

Fnr (122) CEIPFETLDD LSGKMPNLRQ QMRLMSGEI KGDQDMILL SKKNAEERLA
FixK (99) AIFGRRNMQE RS... ..ELLALALTM ARAQHLHVI GRQCAVERIA
Crp (96) AEISYKFKRQ LIQVNPDIEM RLSAQMARRL QVTSEKVGNL AFLDVTGRIA
NtcA (100) FSVPIEFMOK ALIERPELAN VMLQGLSSRI LQTEMMIETL AHRDMGSRV
BifA (101) LSAPIEQVEQ LKMPENPELAN LMLRGLSSRI LQTEMMIETL AHRDMGSRV
pileup a.pie.mqg ...enpel.m .ml...ssri lqt.mmi.tl ahrd.g.Rla

Fnr (172) AFYINLSRRF AQ.RGFSPRE FRLTMTRGDI GNYLGLTIVET ISRLLGRFQK
FixK (143) AFLVDLC... ..E.RQGGGRQ LRLPMSRQDI ADYLDLTIET VSRVYTKLKE
Crp (146) QTLNLAKQP DAMTHPDG... ..MQIKITRQEI GQIVGCSRET VGRILKMLED
NtcA (150) SFLLILCRDF .GIPSPDGIT IDLKLHQAI AEATGSTRVT VTRLLGLDRE
BifA (151) SFLLILCRDF GVPCADGIT IDLKLHQAI AEATGSTRVT VTRLLGLDRE
pileup .fll.lcrdf .g...pdg.t i.lk.erg.i ae..p.tref v.rllgdlre

Fnr (221) SGMLAVK.GK YITIEENDAL AQLAGHTRNV A
FixK (188) RSLIALRDAR TIDIMKPEAL RSLCN... ..
Crp (194) QNLISAH.GK TIVVYGR... ..
NtcA (199) SKLIAIH.KK RITVFNEVAL SQQFS... ..
BifA (200) KKMISIH.KK KITVHKVETL SRQFT... ..
pileup skliaih..k titv..pval sq.f.....

```

FIG. 7. Amino acid sequence alignment of BifA and NtcA, as well as other prokaryotic regulatory proteins belonging to the CRP family. The alignment was determined by using the PILEUP program of the GCG software package. The sequence at the bottom is the consensus determined with PILEUP. Numbers in parentheses indicate positions of the first amino acid in each row. Residues conserved in all five proteins are shown in boldface and in capital letters in the pileup row. The conserved glycine residues in the N terminus that are associated with the CRP  $\beta$ -roll structure are marked with a star. The putative  $\alpha$ -helices of the DNA recognition helix-turn-helix motif in the C terminus are boxed.

tent with it being homologous to NtcA, but BifA's interaction with the *xisA* and *rbcl* genes does not fit the proposed functions for NtcA. We suspect that BifA acts to inhibit expression of *xisA*. The regulatory consequences of BifA's interaction with *rbcl* are unknown but are presumably unrelated to nitrogen metabolism. We propose that BifA can act as both a positive and a negative regulator, as do other members of the CRP family (4, 33) and that its regulatory role in PCC 7120 may involve nitrogen metabolism, as well as other, currently uncharacterized, functions.

#### ACKNOWLEDGMENTS

We thank Steve Elledge and Ronald Davis for providing us with the plasmids necessary for the in vivo selection and for helpful suggestions and Conrad Halling for pointing out the ORF1 homology. We thank Rita Ghosh and Seema Endley for providing the pulsed-field gel *bifA* mapping results.

This work was supported by Public Health Service grant GM36890 from the National Institute of Health, National Science Foundation grant DMB-8553185, and a grant from the Searle Scholars Program of the Chicago Community Trust.

#### REFERENCES

- Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. Basic local alignment search tool. *J. Mol. Biol.* **215**:403-410.
- Ausubel, F. M., R. Brent, R. E. Kingston, D. D. Moore, J. G.

- Seidman, J. A. Smith, and K. Struhl. 1987. Current protocols in molecular biology. Greene Publishing Associates and Wiley Interscience, New York.
- Bancroft, I., C. P. Wolk, and E. V. Oren. 1989. Physical and genetic maps of the genome of the heterocyst-forming cyanobacterium *Anabaena* sp. strain PCC 7120. *J. Bacteriol.* **171**:5940-5948.
- Batut, J., M.-L. Daveran-Mingot, M. David, J. Jacobs, A. M. Garnerone, and D. Kahn. 1989. *fixK*, a gene homologous with *fnr* and *crp* from *Escherichia coli*, regulates nitrogen fixation genes both positively and negatively in *Rhizobium meliloti*. *EMBO J.* **8**:1279-1286.
- Brosius, J. 1989. Superpolylinkers in cloning and expression vectors. *DNA* **8**:759-777.
- Brusca, J. S., C. J. Chastain, and J. W. Golden. 1990. Expression of the *Anabaena* sp. strain PCC 7120 *xisA* gene from a heterologous promoter results in excision of the *nifD* element. *J. Bacteriol.* **172**:3925-3931.
- Chastain, C. J., J. S. Brusca, T. S. Ramasubramanian, T.-F. Wei, and J. W. Golden. 1990. A sequence-specific DNA-binding factor (VF1) from *Anabaena* sp. strain PCC 7120 vegetative cells binds to three adjacent sites in the *xisA* upstream region. *J. Bacteriol.* **172**:5044-5051.
- Chitnis, P. R., and N. Nelson. 1991. Molecular cloning of the genes encoding two chaperone proteins of the cyanobacterium *Synechocystis* sp. PCC 6803. *J. Biol. Chem.* **266**:58-65.
- Devereux, J., P. Haeberli, and O. Smithies. 1984. A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Res.* **12**:387-395.
- Ebright, R. H., P. Cossart, B. Gicquel-Sanzey, and J. Beckwith. 1984. Mutations that alter the DNA sequence specificity of the catabolite gene activator protein of *E. coli*. *Nature (London)* **311**:232-235.
- Elledge, S. J., and R. W. Davis. 1989. Position and density effects on repression by stationary and mobile DNA-binding proteins. *Genes Dev.* **3**:185-197.
- Elledge, S. J., P. Sugiono, L. Guarente, and R. W. Davis. 1989. Genetic selection for genes encoding sequence-specific DNA-binding proteins. *Proc. Natl. Acad. Sci. USA* **86**:3689-3693.
- Fani, R., M. Bazzicalupo, G. Damiani, A. Bianchi, C. Schipani, V. Sgaramella, and M. Polsinelli. 1989. Cloning of histidine genes of *Azospirillum brasilense*: organization of the *ABFH* gene cluster and nucleotide sequence of the *hisB* gene. *Mol. Gen. Genet.* **216**:224-229.
- Fried, M., and D. M. Crothers. 1981. Equilibria and kinetics of *lac* repressor-operator interactions by polyacrylamide gel electrophoresis. *Nucleic Acids Res.* **9**:6505-6525.
- Garner, M. M., and A. Revzin. 1981. A gel electrophoretic method for quantifying the binding of proteins to specific DNA regions: applications to components of the *Escherichia coli* lactose operon regulatory system. *Nucleic Acids Res.* **9**:3047-3060.
- Ghosh, R., S. Endley, and J. W. Golden. Unpublished data.
- Golden, J. W., C. D. Carrasco, M. E. Mulligan, G. J. Schneider, and R. Haselkorn. 1988. Deletion of a 55-kilobase-pair DNA element from the chromosome during heterocyst differentiation of *Anabaena* sp. strain PCC 7120. *J. Bacteriol.* **170**:5034-5041.
- Golden, J. W., M. E. Mulligan, and R. Haselkorn. 1987. Different recombination site specificity of two developmentally regulated genome rearrangements. *Nature (London)* **327**:526-529.
- Golden, J. W., S. J. Robinson, and R. Haselkorn. 1985. Rearrangement of nitrogen fixation genes during heterocyst differentiation in the cyanobacterium *Anabaena*. *Nature (London)* **314**:419-423.
- Golden, J. W., L. L. Whorff, and D. R. Wiest. 1991. Independent regulation of *nifHDK* operon transcription and DNA rearrangement during heterocyst differentiation in the cyanobacterium *Anabaena* sp. strain PCC 7120. *J. Bacteriol.* **173**:7098-7105.
- Golden, J. W., and D. R. Wiest. 1988. Genome rearrangement and nitrogen fixation in *Anabaena* blocked by inactivation of *xisA* gene. *Science* **242**:1421-1423.

22. Haselkorn, R. 1978. Heterocysts. *Annu. Rev. Plant Physiol.* **29**:319–344.
23. Haselkorn, R., S. E. Curtis, J. W. Golden, P. J. Lammers, S. A. Nierzwicki-Bauer, S. J. Robinson, and N. E. Tumer. 1985. Organization and transcription of *Anabaena* genes regulated during heterocyst differentiation, p. 261–266. *In* J. A. Hoch and P. Setlow (ed.), *Molecular biology of microbial differentiation*. American Society for Microbiology, Washington, D.C.
24. Haselkorn, R., J. W. Golden, P. J. Lammers, and M. E. Mulligan. 1986. Developmental rearrangement of cyanobacterial nitrogen-fixation genes. *Trends Genet.* **2**:255–259.
25. Lammers, P. J., J. W. Golden, and R. Haselkorn. 1986. Identification and sequence of a gene required for a developmentally regulated DNA excision in *Anabaena*. *Cell* **44**:905–911.
26. Laudenbach, D. E., and A. R. Grossman. 1991. Characterization and mutagenesis of sulfur-regulated genes in a cyanobacterium: evidence for function in sulfate transport. *J. Bacteriol.* **173**:2739–2750.
27. Maniatis, T., E. F. Fritsch, and J. Sambrook. 1982. *Molecular cloning: a laboratory manual*. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
28. Moore, J., M. Kantorow, D. Vanderzwaag, and K. McKenney. 1992. *Escherichia coli* cyclic AMP receptor protein mutants provide evidence for ligand contacts important in activation. *J. Bacteriol.* **174**:8030–8035.
29. Mulligan, M. E., W. J. Buikema, and R. Haselkorn. 1988. Bacterial-type ferredoxin genes in the nitrogen fixation regions of the cyanobacterium *Anabaena* sp. strain PCC 7120 and *Rhizobium meliloti*. *J. Bacteriol.* **170**:4406–4410.
30. Mulligan, M. E., and R. Haselkorn. 1989. Nitrogen-fixation (*nif*) genes of the cyanobacterium *Anabaena* sp. strain PCC 7120: the *nifB-fdxN-nifS-nifU* operon. *J. Biol. Chem.* **264**:19200–19207.
31. Orme-Johnson, W. H. 1985. Molecular basis of biological nitrogen fixation. *Annu. Rev. Biophys. Biophys. Chem.* **14**:419–459.
32. Ramasubramanian, T. S., T.-F. Wei, and J. W. Golden. Unpublished data.
33. Reznikoff, W. S. 1992. Catabolite gene activator protein activation of *lac* transcription. *J. Bacteriol.* **174**:655–658.
34. Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74**:5463–5467.
35. Schultz, S. C., G. C. Shields, and T. A. Steitz. 1991. Crystal structure of a CAP-DNA complex: the DNA is bent by 90°. *Science* **253**:1001–1007.
36. Shaw, D. J., D. W. Rice, and J. R. Guest. 1983. Homology between CAP and Fnr, a regulator of anaerobic respiration in *Escherichia coli*. *J. Mol. Biol.* **166**:241–247.
37. Trageser, M., and G. Unden. 1989. Role of cysteine residues and of metal ions in the regulatory functioning of Fnr, the transcriptional regulator of anaerobic respiration in *Escherichia coli*. *Mol. Microbiol.* **3**:593–599.
38. Turnowsky, F., K. Fuchs, C. Jeschek, and G. Hogenauer. 1989. *envM* genes of *Salmonella typhimurium* and *Escherichia coli*. *J. Bacteriol.* **171**:6555–6565.
39. Vega-Palas, M. A., E. Flores, and A. Herrero. 1992. NtcA, a global nitrogen regulator from the cyanobacterium *Synechococcus* that belongs to the Crp family of bacterial regulators. *Mol. Microbiol.* **6**:1853–1859.
40. Vega-Palas, M. A., F. Madueno, A. Herrero, and E. Flores. 1990. Identification and cloning of a regulatory gene for nitrogen assimilation in the cyanobacterium *Synechococcus* sp. strain PCC 7942. *J. Bacteriol.* **172**:643–647.
41. Weber, I. T., and T. A. Steitz. 1987. Structure of a complex of catabolite gene activator protein and cyclic AMP refined at 2.5 Å resolution. *J. Mol. Biol.* **198**:311–326.
42. Wolk, C. P. 1982. Heterocysts, p. 359–386. *In* N. G. Carr and B. A. Whitton (ed.), *The biology of cyanobacteria*. University of California Press, Berkeley.
43. 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.