JOURNAL OF BACTERIOLOGY, Sept. 1991, p. 5771–5777
0021-9193/91/185771-07$02.00/0

Vol. 173, No. 18

Characterization of Insertion Sequence IS*892* and Related Elements from the Cyanobacterium *Anabaena* sp. Strain PCC 7120

YUPING CAI

MSU-DOE Plant Research Laboratory, Michigan State University, East Lansing, Michigan 48824-1312

Received 6 May 1991/Accepted 2 July 1991

IS*892*, one of the several insertion sequence (IS) elements discovered in *Anabaena* sp. strain PCC 7120 (Y. Cai and C. P. Wolk, J. Bacteriol. 172:3138–3145, 1990), is 1,675 bp with 24-bp near-perfect inverted terminal repeats and has two open reading frames (ORFs) that could code for proteins of 233 and 137 amino acids. Upon insertion into target sites, this IS generates an 8-bp directly repeated target duplication. A 32-bp sequence in the region between ORF1 and ORF2 is similar to the sequence of the inverted termini. Similar inverted repeats are found within each of those three segments, and the sequences of these repeats bear some similarity to the 11-bp direct repeats flanking the 11-kb insertion interrupting the *nifD* gene of this strain (J. W. Golden, S. J. Robinson, and R. Haselkorn, Nature [London] 314:419–423, 1985). A sequence similar to that of a binding site for the *Escherichia coli* integration host factor is found about 120 bp from the left end of IS*892*. Partial nucleotide sequences of active IS elements IS*892*N and IS*892*T, members of the IS*892* family from the same *Anabaena* strain, were shown to be very similar to the sequence of IS*892*.

Insertion sequences (IS) are transposable elements that are generally smaller than transposons and normally bear only genes related to transposition (for a review, see reference 12). Because IS elements lack selectable markers, they are usually identified by their ability to introduce insertional mutations in marker genes and operons (9, 14). Most IS elements from a variety of prokaryotes are 0.8 to 2.5 kb in size and have a pair of near-perfect inverted terminal repeats ranging from 8 to 41 bp. Almost all bacterial IS elements characterized to date generate directly repeated duplications of their target DNA sequences upon insertion. Many elements make a duplication of a fixed number of base pairs, ranging from 2 to 13 bp, as a characteristic of the element (12).

Cyanobacteria (blue-green algae) are prokaryotes that differ physiologically and phylogenetically (43) from other eubacteria. Two active IS elements have been isolated and sequenced from cyanobacteria: IS*701* from *Calothrix* sp. strain PCC 7601 (12, 27) and IS*891* from *Anabaena* sp. strain M-131 (4). IS*701* appears to be a typical IS element with inverted terminal repeats and the generation of target duplications (27). By contrast, IS*891* lacks inverted terminal repeats and fails to generate a target duplication upon insertion (4). In addition, genetic elements, collectively denoted the *mys* family, from the cyanobacterium *Anabaena* sp. strain PCC 7120 were suggested to be IS on the basis of the structural similarity of their nucleotide sequence to that of typical IS elements (1).

In a study of spontaneous mutation of the conditionally lethal gene *sacB* in *Anabaena* sp. strain PCC 7120, 10 IS elements of five different sizes were detected and one of them, IS*892*, was restriction mapped (7). I report here a more detailed study of IS*892* and of related IS elements in this *Anabaena* strain.

MATERIALS AND METHODS

Bacterial strains, growth media, cultural conditions, and most molecular biological techniques used in this study are as previously described (7). A nitrocellulose filter that was to be rehybridized for further Southern analysis was first stripped of radioactivity by immersion for 10 min in 1 liter of boiling 5 mM EDTA (pH 8.0). [α-^{32}P]dATP used in Southern analysis was purchased from Amersham Corp., Arlington Heights, Ill.

Plasmid DNA was extracted from *Anabaena* sp. strains PCC 7120 and M-131 by a modified boiling procedure (22): cells from a 50-ml liquid culture in early stationary phase were harvested, washed with 1.0 ml of H_2O, and mixed with 700 μl of STET solution (2% sucrose, 5% Triton X-100, 50 mM EDTA [pH 8.0], 10 mM Tris-HCl [pH 8.0]) plus 50 μl of 10 mg of lysozyme per ml of H_2O. After 5 min at room temperature, the suspension was heated in boiling water for 40 s. The mixture was then centrifuged at 21,000 × *g* for 20 min at 4°C, and the pellet was removed. The supernatant solution was extracted with phenol and then with chloroform. DNA was precipitated by isopropanol at −70°C and resuspended in 50 μl of $T_{1/10}E$ solution (10 mM Tris-HCl, 0.1 mM EDTA [pH 8.0]).

DNA sequence was determined by using synthetic DNA primers and ordered deletions of fragments subcloned into vectors pUC118 and pUC119 (41). Deletions were made by a combination of timed digestion by exonuclease III and treatment with mung bean nuclease according to a protocol provided by Stratagene, La Jolla, Calif. DNA oligonucleotide primers were synthesized and purified with equipment and reagents supplied by Applied Biosystems, Inc., Foster City, Calif. Double-stranded DNA and the chain termination technique using dideoxynucleotides (34) were used in DNA sequencing, utilizing Sequenase version 2.0 from U.S. Biochemical Corp., Cleveland, Ohio, and adenosine 5′-α-[^{35}S]thiotriphosphate from Amersham Corp. Electrolyte gradient polyacrylamide sequencing gels were prepared and run as described previously (35). The gels were fixed in a solution of 5% methanol plus 5% acetic acid (6) and dried at 80°C in vacuo before autoradiography. Both strands of the nucleotide sequence presented in Fig. 2 were sequenced. Nucleotide and amino acid sequences were analyzed with the software Editbase (Purdue Research Foundation and USDA/ARS), HIBIO DNASIS, and HIBIO PROSIS (Hitachi America Ltd., San Bruno, Calif.).

Nucleotide sequence accession number. The nucleotide

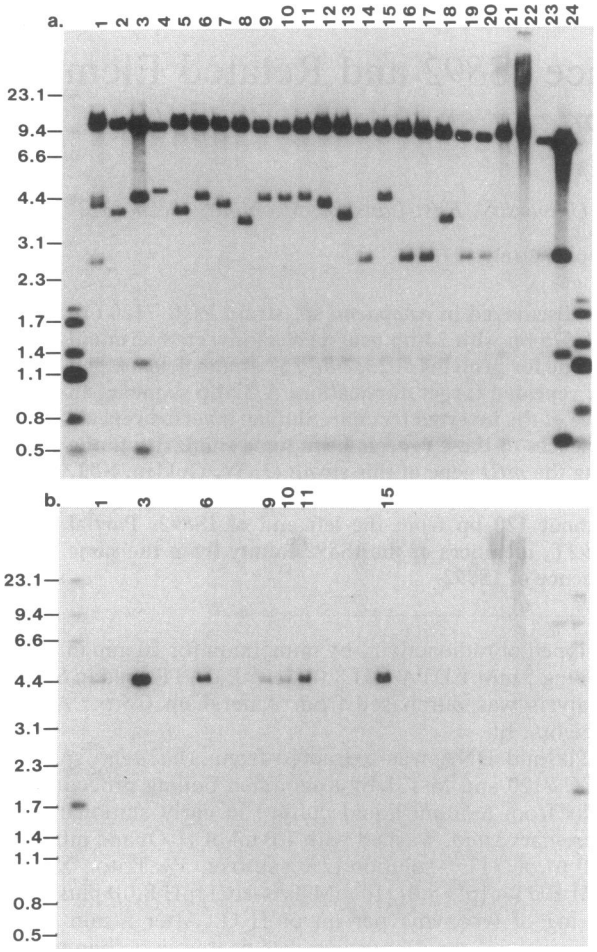

FIG. 1. Southern analysis of DNA from colonies of *Anabaena* sp. strain PS250-N (N = 1, 2, ..., 22). Total DNA from PS250-1 to PS250-22 (lanes 1 to 22, respectively) and from PCC 7120(pRL250) (lane 23) and DNA of plasmid pRL250 (lane 24) were digested with *Pst*I. (a) Filter probed with labelled plasmid pRL250. The band of 2.6 kb corresponds to the *Pst*I fragment that bears the entire *sacB* gene, while the bands of 9.5, 1.2, and 0.5 kb are from the rest of plasmid pRL250 (7). (b) The same filter, stripped of radioactivity from the above hybridization, was reprobed with the internal *Dra*I-*Eco*RV fragment from IS*892* (Fig. 2). The unnumbered lanes on the left in both panels are size markers of DNA (in kilobases). (Material used for panel a, lanes 1 through 10, was presented as lanes 1 to 10 in Fig. 4 of reference 7. Fig. 1a presented here represents a larger sample pool and more-comprehensive data.)

sequence of IS*892* shown in Fig. 2 has been deposited in GenBank under accession number M64297.

RESULTS AND DISCUSSION

Spontaneous mutations in *Anabaena* sp. strain PCC 7120. Plasmid pRL250 (see Fig. 1a of reference 7) carries the gene *sacB* and a fragment of the cyanobacterial plasmid pDU1 that confers autonomous replication in *Anabaena* species. *Anabaena* sp. strain PCC 7120 bearing this plasmid cannot grow on sucrose-containing solid medium because of the presence of a functional *sacB* gene (7). One such *Anabaena* colony was subcultured continuously for 2 months in liquid medium, and about 10^7 cells were then plated on solid medium plus 5% sucrose. Close to 300 colonies were recovered after 10 days. Plasmids isolated from 22 of these colonies (denoted PS250-N, where N = 1, 2, ..., 22) were analyzed by Southern hybridization (Fig. 1a). In 15 of these plasmids, the 2.6-kb *sacB*-containing *Pst*I fragment of pRL250 was replaced by a larger *Pst*I fragment, while other *Pst*I fragments of the plasmid were unchanged. In the remaining seven cases, the 2.6-kb fragment showed no visible change in size or appeared to have been deleted entirely. In the strains that showed no visible change in the 2.6-kb fragment, it is likely that their *sacB* gene had been inactivated by a point mutation or a small deletion, thus accounting for the viability of the strains on sucrose-containing medium.

The 15 variants of pRL250 that showed an increase in size of the 2.6-kb fragment were transferred to *Escherichia coli* by transformation and analyzed. Although colony PS250-1 appeared to contain two variants of pRL250, only the one with a 1.5-kb insertion was recovered. The other, which has a 1.7-kb insertion that hybridizes to IS*892* (Fig. 1b), was not. On the basis of limited data from restriction mapping and Southern analysis, the 15 presumptive IS elements have been tentatively grouped and named (Table 1). A detailed study of the IS element (denoted IS*895*) from colony PS250-5 is presented separately (2). IS*891*, isolated from the closely related *Anabaena* sp. strain M-131, hybridized to the genome of *Anabaena* sp. strain PCC 7120 (4) but failed to hybridize to the 15 active IS discovered in this experiment.

General features of IS*892*. A restriction map of IS*892*, isolated from colony PS250-3, was reported previously (7). IS*892* (Fig. 2) is 1,675 bp in length and has 24-bp near-perfect (21 out of 24 bp) inverted terminal repeats that show no significant sequence similarity to termini of other known bacterial IS elements (12). Two open reading frames (ORFs) are present in tandem on the same DNA strand. When the first methionine residue is taken as the translational initiation codon, the two ORFs, ORF1 and ORF2, are predicted to code for proteins of 233 and 137 amino acids, respectively. Possible alternative start codons TTG (19, 44) and GTG (32) could extend ORF1 to 262 amino acids and ORF2 to 188 or 173 amino acids (Fig. 2). As has been observed in a number of cyanobacterial genes (39) and the other three sequenced cyanobacterial IS elements (2, 4, 38), neither ORF of IS*892*, in defined or extended version, is preceded by a typical ribosome binding sequence. The complementary strand does not contain complete ORFs of more than 82 codons with a reasonably positioned initiation codon. An incomplete reading frame, initiated at the first methionine codon at bp 357 to 355, extends for 118 codons and out of the left (L) of IS*892* end without a stop codon.

An *E. coli*-type promoter (29) is present 5′ to ORF1 (Fig. 2). The −35 region of that presumed promoter, 5′ TTACTA 3′, lies within the L-end terminal repeat sequence. A −35re-

TABLE 1. IS found in *Anabaena* sp. strain PCC 7120

IS element	Size (kb)	Source of insertion	Comment
IS*892*	1.7	PS250-3	Also PS250-6, -9, -10, -11, and -15; see text
IS*893*	1.2	PS250-2	Possibly also PS250-13
IS*894*	1.9	PS250-4	
IS*895*	1.2	PS250-5	A *mys* element (2)
IS*897*	1.5	PS250-7	Possibly also PS250-1 and PS250-12
IS*898*	1.0	PS250-8	Possibly also PS250-18

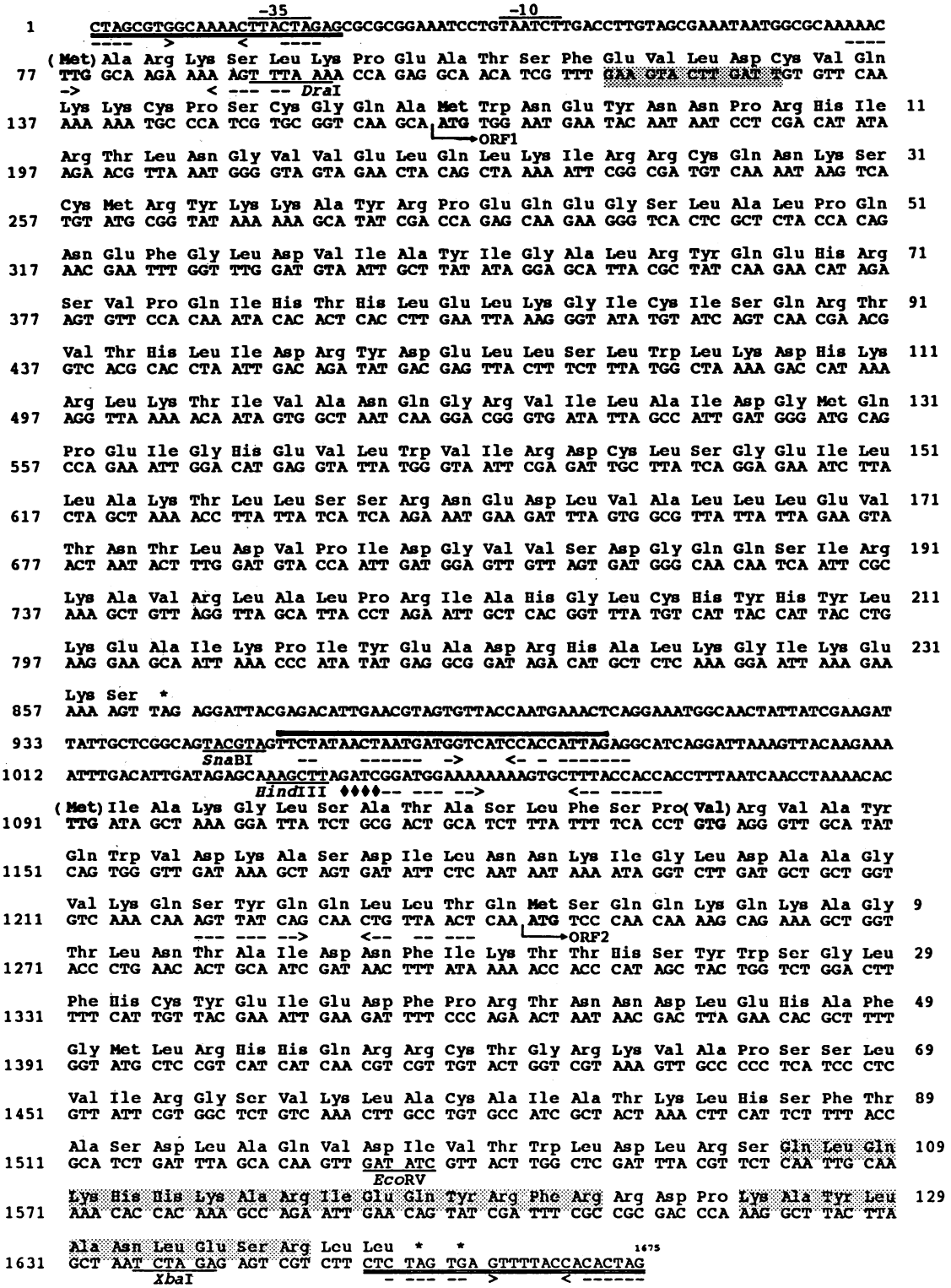

FIG. 2. Complete nucleotide sequence of transposable element IS*892* from the cyanobacterium *Anabaena* sp. strain PCC 7120. The noncoding strand of the sequence is presented in 5′ (left [L] end) to 3′ (right [R] end) direction. Numbering of the nucleotide sequence and of the amino acid sequences of ORF1 and ORF2 are presented, respectively, to the left and to the right of the sequence. The deduced amino acid sequences of ORF1 and ORF2 and their possible 5′ extensions are displayed above the DNA sequence, with presumed translational initiation sites printed in boldface. Possible alternative start codons are indicated in parentheses. Also shown are the −35 and −10 regions of a presumed promoter for ORF1. The inverted termini are double underlined. Potential stem-loop structures are indicated by pairs of counterpointing arrows under the nucleotide sequence. Restriction sites mentioned in the text are underlined, and the site for *dam* methylation is marked by filled diamonds. The M sequence, from bp 953 to 984, which is similar to the terminal repeats (see text), is highlighted by a bar above the sequence. The shaded DNA sequence around bp 125 is a possible IHF binding site (see text), and the shaded amino acid sequences at the carboxyl end of ORF2 may be a potential helix-turn-helix DNA-binding structure.

gion of an *E. coli*-like promoter (29), pointing outward, can be found in the inverted termini: 5′ TTGCCA 3′ at the L end and 5′ TTACCA 3′ at the right (R) end. Such outward-pointing half-promoters have been found at the ends of many IS elements, and their implication has been discussed previously (12).

The sequence 5′ GAAGTACTTGATT 3′, from bp 116 to 128 (Fig. 2), matches well with the consensus sequence 5′ TAAnTnnTTGATT 3′ (17) of binding sites for the *E. coli* integration host factor (IHF). In *E. coli*, the histonelike protein IHF (8, 11) has been shown to participate in the transposition of IS*1* and IS*10*, which have IHF binding sites at or near their end sequences (13, 28). Possible IHF binding sites have been found in IS elements of various origins (12). The presence of a putative IHF binding site near one end of IS*892*, in IS*701* (12), and in IS*895* (2) makes it tempting to speculate that an IHF-like protein, although not yet observed in cyanobacteria, could be involved in the transposition of their IS elements.

Several strains of cyanobacteria, including *Anabaena* sp. strain PCC 7120, exhibit *dam* methylation of their DNA (31). Sites for *dam* methylation found at the ends of IS*10*, IS*50*, and IS*903* have been reported to influence transposition (5, 33). The sole site for *dam* methylation found in IS*892* is located at the end of a stem-loop structure 5′ to ORF2 (Fig. 2). Whether that site influences the activity of IS*892* is, however, unknown.

The two proteins predicted by ORF1 and ORF2 in IS*892*, 26.8 and 15.8 kDa in molecular mass, respectively, are likely cytosolic proteins because each has an overall hydropathy index of −0.1 with no peaks over ±1.0 (window size, 19 amino acid residues [42]). Although calculated isoelectric points are close to 7.0, both proteins have a moderately high content of basic amino acids (arginine and lysine residues account for ca. 15% of the amino acids in each protein), consistent with possible interactions of the proteins with DNA (12). Prediction of secondary structure by the Garnier-Robson method (42) suggested that a helix-turn-helix conformation, a structural motif found repeatedly in prokaryotic DNA-binding proteins (30), could form at the carboxyl end of ORF2 (Fig. 2).

The G+C contents and codon usages of the genome of *Anabaena* sp. strain PCC 7120 and of IS*892* were compared (Tables 2 and 3). IS*892* has a markedly lower G+C content than that of the genome, and its codon usage differs extensively from that of abundantly expressed chromosomal genes: of the 18 amino acids that have multiple synonymous codons, only five amino acids (Pro, Gln, Ala, Val, and Glu) are represented by similar codon preferences. The altered codon usage in IS*892* seems consistent with its lower G+C content: almost all changed codon preferences favor codons ending with A or U. Alternatively, codon usage, often not correlated with the G+C content of an organism (23), may be attributable to the presumed low expressivity of the genes of IS*892* (18).

TABLE 2. Comparison of G+C contents of IS*892* and the chromosome and *nifD* element of *Anabaena* sp. strain PCC 7120[a]

G+C content	*Anabaena* sp. strain PCC 7120	*nifD* element	IS*892*
Overall	42.5	38.7	38.7
Of ORFs	47.9	40.5	39.7

[a] Data on overall G+C content of the strain are as previously published (21), and that of the *nifD* element are calculated from published sequence data (24, 25). The G+C content in the ORFs of this *Anabaena* strain is calculated from eight abundantly expressed chromosomal genes (39), that of the *nifD* element is calculated from five proposed ORFs in the sequenced region (24, 25), and that of IS*892* is calculated from its two ORFs.

Comparisons (26, 42) between IS*892* and other cyanobacterial IS elements IS*701* (38), IS*891*, and IS*895* failed to identify regions of significant similarity of nucleic acid or protein sequence. A search covering both the GenBank and the EMBL data bases also failed to recognize a known transposable element that shares significant sequence similarity with IS*892*.

A family of IS*892*-related IS. The internal *Dra*I-*Eco*RV fragment of IS*892* (Fig. 2) was used to reprobe the filter used in Fig. 1a. Six bands of the same size as the band from PS250-3 showed strong hybridization (Fig. 1b), suggesting that the IS elements from these seven colonies are homologous. Plasmids were recovered from these colonies, except for colony PS250-1, and the L-end portions of their insertions were partially sequenced. The partial sequences (ca. 400 bp) of the insertions from PS250-6, -11, and -15 were identical to the corresponding sequence of IS*892*. However, sequences of the IS elements from PS250-9 and PS250-10 (denoted IS*892*N and IS*892*T, respectively) differed slightly from that of IS*892* as well as from each other. Compared with IS*892*, IS*892*N and IS*892*T have a 1-bp insertion following the L-end terminal and have 20 and 16 single-base-pair changes (transition/transversion ≈ 2:1), respectively. There are eight single-base-pair differences between IS*892*N and IS*892*T. Most of the changes within ORF1 affect the second or the third base of a codon and do not result in any amino acid replacement. A few changes affect the first base of a codon and generate conservative amino acid replacements (data not shown). None of the base pair changes leads to the disruption of ORF1 (even in the extended version), supporting the idea that this ORF encodes a functional protein.

Target sequence and specificity. The junctions produced by insertion of members of the IS*892* family into the *sacB* gene were sequenced. All such insertions were shown to lie within the ORF of *sacB* (36). The data showed that IS*892* makes directly repeated 8-bp target duplications (Table 4). All three members of the IS*892* family inserted into A- or T-rich target sites. The sequence AAAT(a/t) appeared in all the target sites for IS*892* and IS*892*T. The site 5′ AAAATATC 3′ appears to be particularly favorable because at least two independent insertions by IS*892*, in colonies PS250-3 and PS250-11, targeted this site. These two insertions recognized the same site but inserted in opposite orientations, indicating that the orientation of a target site does not necessarily dictate the orientation of insertion.

The changing genome of *Anabaena* sp. strain PCC 7120 and the IS*892* family. The presence of active IS may strongly influence the structure and stability of the genome by transposition and by acting as substrates for homologous recombination (12, 20). When three batches of total DNA, extracted in 1-year intervals from serially subcultured wild-type *Anabaena* sp. strain PCC 7120 (always in liquid medium, with or without fixed nitrogen source; inoculation interval, ca. 2.5 weeks), were digested with *Xba*I or *Eco*RV and probed with the *Dra*I-*Eco*RV fragment of IS*892*, three similar but nonidentical patterns of hybridization were observed; a more dramatically different banding pattern was observed in the genome of a culture recovered from an 8-year-old frozen sample (data not shown). The changes may reflect the activity of the IS*892* family, although DNA rearrangement not related to IS*892* remains a possibility. By

TABLE 3. Comparison of codon usages of IS*892* and the chromosome and *nifD* element of *Anabaena* sp. strain PCC 7120[a]

Amino acid	Gene codon	Codon usage frequency of:			Amino acid	Gene codon	Codon usage frequency of:		
		Anabaena sp. strain PCC 7120	*nilD* element	IS*892*			*Anabaena* sp. strain PCC 7120	*nilD* element	IS*892*
Arg	CGA	1.5	13.7	20.0	Ile	AUA	1.4	31.8	29.6
	CGC	29.4	28.8	13.3		AUC	69.5	23.5	18.5
	CGG	6.6	11.0	10.0		AUU	29.1	44.7	51.9
	CGU	51.5	20.5	23.3					
	AGA	9.6	20.5	26.7	Lys	AAA	54.7	80.5	79.2
	AGG	1.5	5.5	6.7		AAG	45.3	19.5	20.8
Leu	CUA	11.1	17.9	14.9	Asn	AAC	91.7	32.3	28.6
	CUC	14.6	10.6	12.8		AAU	8.3	67.7	71.4
	CUG	19.4	13.0	4.2					
	CUU	4.4	9.7	12.8	Gln	CAA	80.4	74.0	73.7
	UUA	16.6	30.9	48.9		CAG	19.6	26.0	26.3
	UUG	40.0	17.9	6.4					
					Ilis	CAC	91.0	33.3	36.8
Ser	UCA	6.8	21.0	33.3		CAU	9.0	66.7	63.2
	UCC	28.4	13.6	9.5					
	UCG	0.0	6.2	0.0	Glu	GAA	80.4	68.0	76.2
	UCU	39.8	23.4	28.6		GAG	19.6	32.0	23.8
	AGC	21.6	14.8	4.8					
	AGU	3.4	21.0	23.8	Asp	GAC	55.4	34.2	27.8
						GAU	44.6	65.8	72.2
Thr	ACA	25.6	37.8	5.9					
	ACC	59.5	28.4	29.4	Tyr	UAC	80.2	38.1	46.2
	ACG	2.4	9.5	17.6		UAU	19.8	61.9	53.8
	ACU	12.5	24.3	47.1					
					Cys	UGC	68.2	45.4	12.5
Pro	CCA	26.4	27.6	54.5		UGU	31.8	54.6	87.5
	CCC	23.6	25.9	27.3					
	CCG	0.0	3.4	0.0	Phe	UUC	73.6	25.0	0.0
	CCU	50.0	43.1	18.2		UUU	26.4	75.0	100.0
Ala	GCA	26.8	33.3	26.9	Met	AUG			
	GCC	9.0	21.2	19.2					
	GCG	7.1	7.4	7.7	Trp	UGG			
	GCU	57.1	37.1	46.1					
					Stop	UAA	66.7	40.0	0.0
Gly	GGA	7.0	35.6	38.9		UAG	22.2	20.0	66.7
	GGC	17.1	27.1	5.5		UGA	11.1	40.0	33.3
	GGG	2.3	8.5	22.2					
	GGU	73.6	28.8	33.3					
Val	GUA	46.8	25.4	35.0					
	GUC	6.4	20.0	10.0					
	GUG	6.4	20.0	15.0					
	GUU	40.4	34.6	40.0					

[a] Codon usage frequency is presented as a percentage of the total usage of corresponding sets of synonymous codons. Methionine and tryptophan are not compared because a single codon corresponds to each of these amino acids. See footnote *a* of Table 2 for references on ORFs used in the calculations.

contrast, IS*895* did not show any changed pattern of hybridization to total DNA of several cultures of wild-type *Anabaena* sp. strain PCC 7120, including one with which the hybridization pattern of IS*892* had changed (2; unpublished observation). Given that comparison and the fact that they were most frequently observed among insertions into the *sacB* gene, members of the IS*892* family appear to transpose actively. *Anabaena* species have multiple copies of the chromosome per cell (7) and are filamentous. It remains uncertain whether random chance or some unidentified selective pressure led to conversion of all copies of the chromosome to the new configuration in a relatively short period of time.

Southern analysis of *Eco*RV- or *Xba*I-digested total DNA from four cultures (one shown in Fig. 3) of wild-type *Anabaena* sp. strain PCC 7120 showed that there were at least nine copies of the members of IS*892* family in the genome (two cultures showed 10 distinct bands). It was not determined which hybridizing band corresponded to a particular member of the IS*892* family or whether all hybridizing copies were capable of transposition. No data are available to indicate whether members of the IS*892* family transpose conservatively or replicatively.

Unique nucleotide structure of IS*892*. When properly aligned, a 32-bp sequence, herein denoted the M sequence (bp 953 to 984 [Fig. 2]), in the region between ORF1 and ORF2 could be viewed as an imperfect direct repeat of the R-end terminal sequence (therefore an imperfect inverted

TABLE 4. Target duplications produced by insertions of members of the IS*892* family in *Anabaena* sp. strain PCC 7120[a]

Source of insertion	IS element	Target duplication
PS250-3	IS*892*	5′ ^{684}AAAATATC^{691} 3′
PS250-15	IS*892*	5′ ^{684}AAAATATC^{691} 3′
PS250-11	IS*892*	5′ ^{691}*GATATTTT^{684}* 3′
PS250-6	IS*892*	5′ ^{1508}*TTTTAAAG^{1501}*3′
PS250-9	IS*892*N	5′ ^{1468}*GTTAGATG^{1461}*3′
PS250-10	IS*892*T	5′ ^{1231}CAAATACT^{1238} 3′

[a] Target sequences in the *sacB* gene are presented from 5′ to 3′ where insertion immediately follows in the same orientation as shown in Fig. 2. Base-pair numbering (superscripts) is after the published *sacB* sequence (36), and numbers with an asterisk indicate sequence of the complementary strand. The R-end junctions of IS*892*N and IS*892*T were not satisfactorily sequenced, so 8-bp duplications are partially assumed.

repeat of the L end). Discounting the six bases that introduce gaps in the alignment, 20 of 26 bases in the M sequence are identical, in order, to the 26 bases at the R end. A similar inverted repeat was also found within each of those three sequences (the inverted repeats in the M and R-end sequences are capable of forming high-energy stem-loop structures). The L-end and M sequences thus make possible a mini-IS*892* element bearing only ORF1. The M and R-end sequences, on the other hand, are direct repeats containing intrinsic inverted repeats. Such a structure surrounding ORF2 is reminiscent of the *aadA* and *sat* genes in Tn*7* and related Tn*1825*, which were suggested to have inserted into the transposons via integrase-mediated site-specific recombinations (10, 37). Downstream from the M sequence is another stem-loop structure which is immediately preceded by the *dam* methylation site (Fig. 2). The sequence of one arm of this stem-loop, 5′ TTTACCAC 3′ (bp 1060 to 1067), is exactly repeated in the R-end sequence, which suggests the possible formation of an alternative stem-loop structure with ORF2 in the loop region.

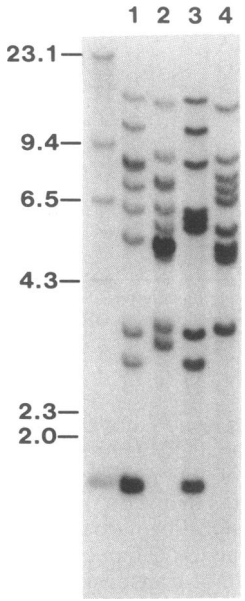

FIG. 3. Southern analysis to determine copy number of IS*892*-like elements in the genomes of *Anabaena* sp. strains PCC 7120 and M-131. Total DNAs from strain PCC 7120 (lanes 1 and 2) and strain M-131 (lanes 3 and 4) were digested with *Eco*RV (lanes 1 and 3) or *Xba*I (lanes 2 and 4) and probed with the radioactively labelled internal *Dra*I-*Eco*RV fragment from IS*892*. The unnumbered lane on the left indicates sizes of DNA (in kilobases).

To observe whether the postulated mini-IS*892* transposes or whether ORF2 is removable from IS*892*, the 0.85-kb *Dra*I-*Sna*BI fragment (L end to M) and the 0.5-kb *Hin*dIII-*Eco*RV fragment (M to R end) from IS*892* (see Fig. 2) were individually used to probe *Eco*RV- or *Xba*I-digested total DNA. Patterns of hybridization by both probes were identical to that by the *Dra*I-*Eco*RV fragment containing both ORF1 and ORF2 (Fig. 3), suggesting that transposition of the complete IS*892* is the predominant event. A computer search of both ORFs of IS*892* failed to identify a structural motif similar to the one that is conserved in the integrase family of site-specific recombinases (3).

Resemblance of IS*892* and the *nifD* element. The *nifD* element is an 11-kb sequence interrupting the *nifD* gene in the chromosome of vegetative cells of the same *Anabaena* strain, and it must be excised from the chromosome by site-specific recombination between its 11-bp directly repeated border sequences in a late stage of heterocyst differentiation in order to create a functional *nifD* gene (15, 16, 24). Six or 7 bases of the sequence of that 11-bp recombination site were found to be identical to corresponding bases (no gaps introduced in the alignment) in the sequences comprising the intrinsic repeats in the termini of IS*892* and in the M sequence. The *nifD* element also has essentially the same G+C content and codon usage as IS*892* (Tables 2 and 3). No significant sequence similarity was found between IS*892* and the *nifD* element, except for their termini.

Distribution of IS*892*. IS*892* from *Anabaena* sp. strain PCC 7120 hybridized to genomic DNA from several other *Anabaena* and *Nostoc* strains which are filamentous, including *Anabaena* sp. strain M-131 (40) (Fig. 3, lanes 3 and 4) but did not hybridize to two unicellular cyanobacterial strains tested (40). However, attempts to entrap an active IS*892*-like transposable element from *Anabaena* sp. strain M-131, which is also sensitive to sucrose when bearing pRL250 (7), failed (unpublished results). It is possible that the IS*892*-hybridizing elements in *Anabaena* sp. strain M-131 transpose relatively infrequently or no longer at all.

ACKNOWLEDGMENTS

I thank S. E. Curtis, N. Tandeau de Marsac, and T. Thiel for providing unpublished data during the course of this study and D. Holland for helpful advice on sequencing. I am very grateful to C. P. Wolk and J. Elhai for critical reading and revision of the manuscript.

This work was supported by the U.S. Department of Energy under contract DE-AC02-76ERO-1338 and grant DE-FG02-90ER20021.

REFERENCES

1. **Alam, J., and S. E. Curtis.** 1985. Characterization of a family of putative insertional elements from the cyanobacterium *Anabaena*, abstr. OR-22-07. 1st Int. Congr. Plant Molecular Biology.
2. **Alam, J., J. M. Vrba, Y. Cai, J. A. Martin, L. J. Weislo, and S. E. Curtis.** 1991. Characterization of the IS*895* family of insertion sequences from the cyanobacterium *Anabaena* sp. strain PCC 7120. J. Bacteriol. **173**:5778–5783.
3. **Argos, P., A. Landy, K. Abremski, J. B. Egan, E. Haggard-Ljungquist, R. H. Hoess, M. L. Kahn, H. Kalionis, S. V. L. Narayana, L. S. Pierson III, N. Sternberg, and J. M. Leong.** 1986. The integrase family of site-specific recombinases: regional similarities and global diversity. EMBO J. **5**:433–440.
4. **Bancroft, I., and C. P. Wolk.** 1989. Characterization of an insertion sequence (IS*891*) of novel structure from the cyano-

bacterium *Anabaena* sp. strain M-131. J. Bacteriol. **171**:5949–5954.
5. **Berg, D. E.** 1989. Transposon Tn*5*, p. 185–210. *In* D. E. Berg and M. M. Howe (ed.), Mobile DNA. American Society for Microbiology, Washington, D.C.
6. **Biggin, M. D., T. J. Gibson, and G. F. Hong.** 1983. Buffer gradient gels and ^{35}S label as an aid to rapid DNA sequence determination. Proc. Natl. Acad. Sci. USA **80**:3963–3965.
7. **Cai, Y., and C. P. Wolk.** 1990. Use of a conditionally lethal gene in *Anabaena* sp. strain PCC 7120 to select for double recombinants and to entrap insertion sequences. J. Bacteriol. **172**:3138–3145.
8. **Drlica, K., and J. Rouviere-Yaniv.** 1987. Histonelike proteins of bacteria. Microbiol. Rev. **51**:301–319.
9. **Fiandt, M., W. Szybalski, and M. H. Malamy.** 1972. Polar mutations in *lac*, *gal* and phage λ consist of a few IS-DNA sequences inserted with either orientation. Mol. Gen. Genet. **119**:223–231.
10. **Fling, M. E., J. Kopf, and C. Richards.** 1985. Nucleotide sequence of the transposon Tn*7* gene encoding an aminoglycoside-modifying enzyme, 3″(9)-*O*-nucleotidyltransferase. Nucleic Acids Res. **13**:7095–7106.
11. **Friedman, D. I.** 1988. Integration host factor: a protein for all reasons. Cell **55**:545–554.
12. **Galas, D. J., and M. Chandler.** 1989. Bacterial insertion sequences, p. 109–162. *In* D. E. Berg and M. M. Howe (ed.), Mobile DNA. American Society for Microbiology, Washington, D.C.
13. **Gamas, P., M. Chandler, P. Prentki, and D. J. Galas.** 1987. *Escherichia coli* integration host factor binds specifically to the ends of the insertion sequence IS*1* and to its major insertion hotspot in pBR322. J. Mol. Biol. **195**:261–272.
14. **Gay, P., D. Le Coq, M. Steinmetz, T. Berkelman, and C. I. Kado.** 1985. Positive selection procedure for entrapment of insertion sequence elements in gram-negative bacteria. J. Bacteriol. **164**:918–921.
15. **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.
16. **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.
17. **Goodrich, J. A., M. L. Schwartz, and W. R. McClure.** 1990. Searching for and predicting the activity of sites for DNA binding proteins: compilation and analysis of the binding sites for *Escherichia coli* integration host factor (IHF). Nucleic Acids Res. **18**:4993–5000.
18. **Gouy, M., and C. Gautier.** 1982. Codon usage in bacteria: correlation with gene expressivity. Nucleic Acids Res. **10**:7055–7074.
19. **Gren, E. J.** 1984. Recognition of messenger RNA during translational initiation in *Escherichia coli*. Biochimie **66**:1–29.
20. **Grindley, N. D. F., and R. R. Reed.** 1985. Transpositional recombination in prokaryotes. Annu. Rev. Biochem. **54**:863–896.
21. **Herdman, M., M. Janvier, J. B. Waterbury, R. Rippka, and R. Y. Stanier.** 1979. Deoxyribonucleic acid base composition of cyanobacteria. J. Gen. Microbiol. **111**:63–71.
22. **Holmes, D. S., and M. Quigley.** 1981. A rapid boiling method for the preparation of bacterial plasmids. Anal. Biochem. **114**:193–197.
23. **Ikemura, T.** 1985. Codon usage and tRNA content in unicellular and multicellular organisms. Mol. Biol. Evol. **2**:13–34.
24. **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.
25. **Lammers, P. J., S. McLaughlin, S. Papin, C. Trujillo-Provencio, and A. J. Ryncarz II.** 1990. Developmental rearrangement of cyanobacterial *nif* genes: nucleotide sequence, open reading frames, and cytochrome P-450 homology of the *Anabaena* sp. strain PCC 7120 *nifD* element. J. Bacteriol. **172**:6981–6990.
26. **Lipman, D. J., and W. R. Pearson.** 1985. Rapid and sensitive protein similarity searches. Science **227**:1435–1441.
27. **Mazel, D., A.-M. Castets, J. Houmard, and N. Tandeau de Marsac.** 1988. Cyanobacterial insertion elements: characterization and potential, p. 227. Abstr. VI Intl. Symp. Photosynthetic Prokaryotes.
28. **Morisato, D., and N. Kleckner.** 1987. Tn*10* transposition and circle formation *in vitro*. Cell **51**:101–111.
29. **Mulligan, M. E., D. K. Hawley, R. Entriken, and W. R. McClure.** 1984. *Escherichia coli* promoter sequences predict *in vitro* RNA polymerase selectivity. Nucleic Acids Res. **12**:789–800.
30. **Pabo, C. O., and R. T. Sauer.** 1984. Protein-DNA recognition. Annu. Rev. Biochem. **53**:293–321.
31. **Padhy, R. N., F. G. Hottat, M. M. Coene, and P. P. Hoet.** 1988. Restriction analysis and quantitative estimation of methylated bases of filamentous and unicellular cyanobacterial DNAs. J. Bacteriol. **170**:1934–1939.
32. **Reddy, K. J., G. S. Bullerjahn, D. M. Sherman, and L. A. Sherman.** 1988. Cloning, nucleotide sequence and mutagenesis of a gene (*irpA*) involved in iron-deficient growth of the cyanobacterium *Synechococcus* sp. strain PCC 7942. J. Bacteriol. **170**:4466–4476.
33. **Roberts, D., D. C. Hoopes, W. R. McClure, and N. Kleckner.** 1985. IS*10* transposition is regulated by DNA adenine methylation. Cell **43**:117–130.
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. **Sheen, J., and B. Seed.** 1988. Electrolyte gradient gels for DNA sequencing. BioTechniques **6**:942–944.
36. **Steinmetz, M., D. Le Coq, S. Aymerich, G. Gonzy-Tréboul, and P. Gay.** 1985. The DNA sequence of the gene for the secreted *Bacillus subtilis* enzyme levansucrase and its genetic control sites. Mol. Gen. Genet. **200**:220–228.
37. **Sundström, L., P. H. Roy, and O. Sköld.** 1991. Site-specific insertion of three structural gene cassettes in transposon Tn*7*. J. Bacteriol. **173**:3025–3028.
38. **Tandeau de Marsac, N.** 1988. Personal communication.
39. **Tandeau de Marsac, N., and J. Houmard.** 1987. Advances in cyanobacterial molecular genetics, p. 251–302. *In* P. Fay and C. Van Baalen (ed.), The cyanobacteria. Elsevier Science Publishing, Inc., New York.
40. **Thiel, T.** 1989. Personal communication.
41. **Vieira, J., and J. Messing.** 1987. Production of single-stranded plasmid DNA. Methods Enzymol. **153**:3–11.
42. **Von Heijne, G.** 1987. Sequence analysis in molecular biology: treasure trove or trivial pursuit? p. 81–121. Academic Press, Inc., New York.
43. **Woese, C. R.** 1987. Bacterial evolution. Microbiol. Rev. **51**:221–271.
44. **Zhang, C., M. Durand, R. Jeanjean, and F. Joset.** 1989. Molecular and genetical analysis of the fructose-glucose transport system in the cyanobacterium *Synechocystis* PCC6803. Mol. Microbiol. **3**:1221–1229.