Proc. Natl. Acad. Sci. USA
Vol. 92, pp. 791–795, January 1995
Genetics

Programed DNA rearrangement of a cyanobacterial *hupL* gene in heterocysts

(*Anabaena*/nitrogen fixation/uptake hydrogenase/site-specific recombinase)

CLAUDIO D. CARRASCO, JILL A. BUETTNER, AND JAMES W. GOLDEN*

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

Communicated by Robert Haselkorn, University of Chicago, Chicago, IL, October 20, 1994 (received for review August 5, 1994)

ABSTRACT **Programed DNA rearrangements that occur during cellular differentiation are uncommon and have been described in only two prokaryotic organisms. Here, we identify the developmentally regulated rearrangement of a hydrogenase gene in heterocysts of the cyanobacterium *Anabaena* sp. strain PCC 7120. Heterocysts are terminally differentiated cells specialized for nitrogen fixation. Late during heterocyst differentiation, a 10.5-kb DNA element is excised from within the *hupL* gene by site-specific recombination between 16-bp direct repeats that flank the element. The predicted HupL polypeptide is homologous to the large subunit of [NiFe] uptake hydrogenases. *hupL* is expressed similarly to the nitrogen-fixation genes; *hupL* message was detected only during the late stages of heterocyst development. An open reading frame, named *xisC*, identified near one end of the *hupL* DNA element is presumed to encode the element's site-specific recombinase. The predicted XisC polypeptide is homologous with the *Anabaena* sp. strain PCC 7120 site-specific recombinase XisA. Neither XisC nor XisA shows sequence similarity to other proteins, suggesting that they represent a different class of site-specific recombinase.**

Programed DNA rearrangements have been described in eukaryotic and prokaryotic organisms, but relatively few genome rearrangements are specifically associated with a differentiated cell type. In eukaryotes, examples include rearrangement of the immunoglobulin and T-cell receptor genes in the vertebrate immune system (1, 2), chromatin diminution in nematodes (3), and yeast mating-type switching (4). In prokaryotes, two organisms undergo developmentally regulated DNA rearrangements (5). During sporulation in *Bacillus subtilis*, a 42-kb element is deleted from the *sigK* gene in the terminally differentiating mother cell restoring the intact gene (6). The *sigK* rearrangement requires a site-specific recombinase encoded by the *spoIVCA* gene, which is present on the 42-kb element (6, 7). In *Anabaena* sp. strain PCC 7120 (*Anabaena* 7120), two DNA elements are excised from within nitrogen-fixation operons during heterocyst development (8).

Anabaena 7120 is a filamentous heterocyst-forming cyanobacterium that is capable of oxygen-evolving photosynthesis and nitrogen fixation. Under nitrogen-limiting conditions, *Anabaena* 7120 grows as a simple multicellular organism composed of two interdependent cell types: vegetative cells and heterocysts. Heterocysts are highly specialized, terminally differentiated cells that supply vegetative cells with fixed nitrogen (9). Heterocysts are formed about every 10th cell along a filament in a one-dimensional semiregular pattern (10, 11).

Two genome rearrangements occur during the late stages of heterocyst differentiation at about the same time that the nitrogen-fixation genes begin to be transcribed (8, 12). The *nifD* rearrangement is the excision of an 11-kb element from within the *nifD* open reading frame (8). Excision of the element occurs by site-specific recombination between two 11-bp direct repeats and requires the product of the *xisA* gene, which is located on the *nifD* element (13–15). An *xisA* null mutant forms heterocysts but is unable to excise the *nifD* element or grow on medium lacking a source of combined nitrogen (14).

The *fdxN* rearrangement is the excision of a 55-kb element from within the *fdxN* gene, which is part of the *nifB–fdxN–nifS–nifU* operon (16, 17). Rearrangement occurs by site-specific recombination between two 5-bp direct repeats that border the element (18). The *fdxN* rearrangement requires the *xisF* gene, which is located on the element and encodes a protein belonging to the resolvase class of site-specific recombinases (19). An *xisF* knockout mutant forms heterocysts but is unable to excise the *fdxN* element or grow on medium lacking a source of combined nitrogen (19).

We report the identification of a third programed genome rearrangement that occurs during *Anabaena* 7120 heterocyst differentiation and involves the excision of a DNA element from within a hydrogenase gene by site-specific recombination.† A putative recombinase gene, *xisC*, was identified on the element. *xisC*, along with the *Anabaena* 7120 *xisA* gene, does not belong to the integrase or resolvase class of site-specific recombinases.

MATERIALS AND METHODS

***Anabaena* Growth Conditions.** *Anabaena* 7120 was grown and induced to form heterocysts as described (12). Heterocysts were isolated as described (19).

DNA Manipulations and Plasmid Constructions. DNA manipulations and recombinant DNA techniques were performed by standard procedures (20). *Anabaena* 7120 DNA was prepared from vegetative cells, purified heterocysts, and the heterocyst-induction samples as described (12). Restriction endonucleases and other DNA-modifying enzymes were used according to the manufacturer's recommendations or standard protocols. Restriction enzyme sites were mapped on the overlapping cosmid clones 8D9, 3C7, and 13B3. The map was confirmed by Southern hybridizations of vegetative cell and heterocyst genomic DNA. For Southern analysis, DNA was transferred from agarose gels to MagnaCharge membrane (MSI) with 50 mM NaOH/1 M NaCl. DNA fragments were labeled with a random primer kit (Boehringer Mannheim), and Southern hybridization was performed as described (12).

A cosmid library containing random fragments of the *Anabaena* 7120 vegetative cell genome was made in the conjugal shuttle vector pDUCA7M, which was modified from

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviation: RT-PCR, reverse transcription polymerase chain reaction.

*To whom reprint requests should be addressed.

†The sequences reported in this paper have been deposited in the GenBank data base [accession nos. U08013 (*hupL*) and U08014 (*xisC*)].

pDUCA7 (21) by eliminating the *Sal* I site between *oriT* and *oriV*. *Anabaena* 7120 genomic DNA was partially digested with *Cpf* I to yield fragments that averaged 20–40 kb in size and then dephosphorylated. Vector arms, each containing a *cos* site, were isolated after digestion with *Bam*HI and *Sal* I or *Bam*HI and *Kpn* I. The insert DNA was ligated to the pDUCA7M vector arms, packaged *in vitro*, and the resulting phage particles were used to infect *Escherichia coli* strain DH5α-MCR.

Plasmids were maintained in *E. coli* strain DH10B as described (12). The plasmid pAM1268 contains the 6.5-kb *Hin*dIII fragment from cosmid 3C7 in the *Hin*dIII site of pBluescript SK (+) (Stratagene). pAM1283 contains a 1.7-kb *Hpa* I fragment from pAM1268 ligated into the *Sma* I site of pBluescript SK (+). pAM1311 contains a 3.2-kb *Hin*dIII fragment from cosmid 3C7 in the *Hin*dIII site of pBluescript SK (+). pAM1314 contains an 8.4-kb *Eco*RI fragment from cosmid 3C7 in the *Eco*RI site of pBluescript SK (+).

RNA Manipulations. RNA for Northern blot analysis and reverse transcription polymerase chain reaction (RT-PCR) was prepared from frozen samples of vegetative cells, heterocysts, and heterocyst-induction samples as described (12).

DNA Sequencing. Double-stranded DNA sequencing was performed on plasmid clones or PCR products with a SequiTherm cycle sequencing kit (Epicentre Technologies, Madison, WI) and [α-^{32}P]dATP (DuPont/NEN Research Products). DNA sequence was determined for both strands. The heterocyst recombination sites were sequenced directly from fragments amplified by PCR from heterocyst genomic DNA. Sequencing data were analyzed by using the sequence analysis software package of the Genetics Computer Group (22) and the NCBI GenBank BLAST (23) e-mail server.

PCR and RT-PCR. RT-PCR was performed essentially as described (20). Purified total RNA was treated with DNase I before RT-PCR analysis. Reverse transcription with avian myeloblastosis virus reverse transcriptase (Boehringer Mannheim) was performed according to the manufacturer's instructions. The reverse transcription reactions (20 μl) contained 0.5 μg of total RNA and 34 ng of oligonucleotide primers. The antisense primers were 5′-TTCTACTGCACCGCGTT-3′ for *nifH* and 5′-CGTTCCAAAGAACAACCC-3′ for *hupL*.

PCR reactions were carried out in a MiniCycler (MJ Research, Cambridge, MA) with *Taq* DNA polymerase (Promega) essentially according to the manufacturer's instructions. All PCRs were 50 μl and contained 34 ng of each oligonucleotide primer and a final concentration of 1.5 mM $MgCl_2$. Reactions with genomic DNA contained 0.1 μg of purified DNA. For PCR amplification of the rearranged heterocyst chromosome, the primers were 5′-GTTCGCCATTGACC-3′ and 5′-GTGGTACGGTCAAAGG-3′ and the reaction was cycled 30 times at 94°C for 1 min, 43°C for 1 min, and 72°C for 2 min. For PCR amplification of the excised *hupL* element, the primers were 5′-GCTTCGACTAACTTCTG-3′ and 5′-GAGTTTAGACGATTTTGGGG-3′ and the reaction was cycled 30 times at 94°C for 1 min, 48°C for 1 min, and 72°C for 2 min. The RT-PCR amplification reactions included a hot start at 94°C for 10 min before addition of polymerase, contained cDNA from 5 μl of the reverse transcription reaction, and used the same antisense primer as that used for reverse transcription. PCR amplification of *nifH* cDNA used the sense primer 5′-CGGCGACTGTTCTACAT-3′ and was cycled 40 times at 94°C for 1 min, 53°C for 2 min, and 72°C for 1 min. PCR amplification of *hupL* cDNA was the same except the sense primer was 5′-GTTCGCCATTGACC-3′ and the annealing temperature was 48°C.

RESULTS

Identification of the Heterocyst-Specific DNA Rearrangement. Our studies of a set of cosmid clones that complemented an *Anabaena* 7120 heterocyst pattern-formation mutant, which formed double heterocysts instead of single heterocysts, led to the serendipitous discovery of a third heterocyst-specific genome rearrangement. Southern analysis with the cosmid clone 8D9 as a probe showed differences in the pattern of hybridizing fragments between vegetative cell and heterocyst DNA samples. The cosmid 8D9 was used to identify two additional overlapping cosmid clones, 3C7 and 13B3, from the same library. Southern experiments with isolated restriction fragments were used to identify a 6.5-kb *Hin*dIII fragment from cosmid 3C7 that contained a DNA breakpoint. Fig. 1 shows a Southern blot of *Hae* II-digested DNA hybridized with the 6.5-kb *Hin*dIII fragment. A 16.5-kb fragment present in vegetative cells was replaced in heterocysts by a 6.0-kb fragment and a slowly migrating band presumed to be an open (nicked) circular molecule, which we named the *hupL* element. The sheared linear form of the excised *hupL* element migrated at 10.5 kb. An adjacent 2.2-kb *Hae* II fragment also hybridized with the probe.

The new rearrangement occurs late during heterocyst differentiation at approximately the same time as the *nifD* and *fdxN* rearrangements (8). Southern analysis of genomic DNA isolated at various times during heterocyst differentiation and from purified heterocysts detected the new rearrangement 24 hr after induction and showed that the rearrangement is under tight developmental control (Fig. 1). Our Southern analysis did not show evidence for specific amplification or degradation of the excised *hupL* element in heterocysts.

Southern analysis of genomic DNA fragments separated by pulsed-field gel electrophoresis showed that the *hupL* element is located on the *Bln* I-E and *Pst* I-A fragments (data not shown), which places the element between 0.4 and 0.9 Mb on the *Anabaena* 7120 chromosome, which is >700 kb from the *nifD* and *fdxN* elements (24).

The Rearrangement Is a Conservative Site-Specific Recombination. Restriction analysis of vegetative cell and heterocyst DNA showed that a 10.5-kb element was deleted from the heterocyst chromosome as a circular molecule (Fig. 2*A*). The DNA breakpoints on the vegetative cell chromosome were localized to small restriction fragments that were then sequenced. Our maps of vegetative cell and heterocyst DNA, along with the analysis of partial open reading frames on the sequenced vegetative cell fragments, were used to select PCR primers that we predicted would allow the amplification of the two DNA breakpoints from heterocyst DNA. PCR products of the expected sizes were obtained and their sequences were

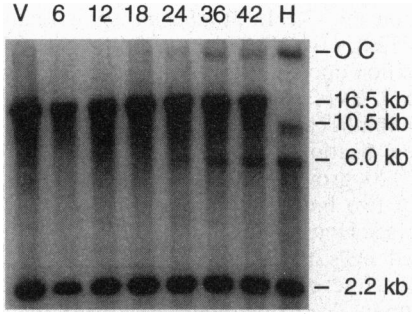

FIG. 1. Southern analysis of DNA isolated during heterocyst induction in *Anabaena* 7120. *Hae* II-digested DNA from vegetative cells (V), purified heterocysts (H), and filaments isolated at 6, 12, 18, 24, 36, and 42 hr after nitrogen step-down was fractionated by electrophoresis on a 0.7% agarose gel and transferred to a nylon membrane. The membrane was hybridized with a radiolabeled probe generated from the 6.5-kb *Hin*dIII fragment in pAM1268. In this induction, proheterocysts were formed by 18 hr and mature heterocysts (with polar granules) were present at 24 hr. OC, open (nicked) circular form.

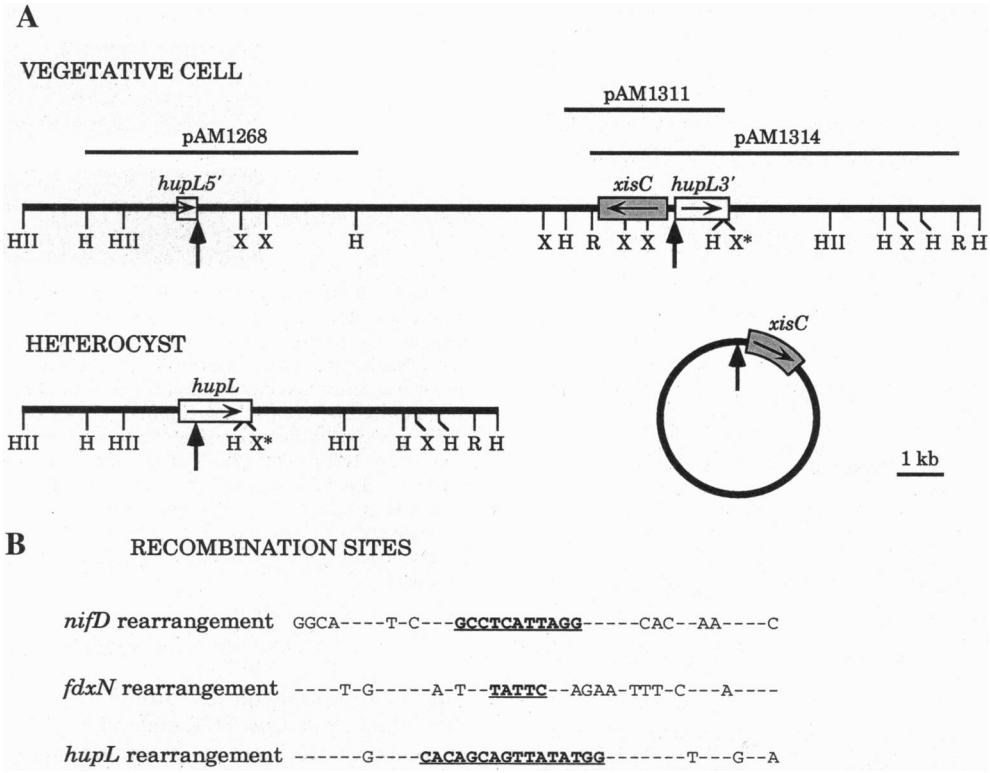

FIG. 2. (*A*) Map of the *hupL* rearrangement. The vegetative cell and heterocyst chromosomes and the excised 10.5-kb element are shown. The region of the vegetative cell chromosome shown is contained in the cosmid clone 3C7. Restriction sites: *Eco*RI (R), *Hin*dIII (H), *Hae* II (HII), and *Xba* I (X). *Xba* I sites marked with an asterisk (X*) do not cut in genomic DNA isolated from *Anabaena* 7120. Restriction sites are not marked on the excised *hupL* element. Boxes mark the *hupL* and *xisC* open reading frames and arrows show their orientation. The 5′ and 3′ ends of the *hupL* gene are shown on the vegetative cell chromosome. Vertical arrows indicate recombination sites. The *Hin*dIII fragments in pAM1268 and pAM1311 and the *Eco*RI fragment in pAM1314 were used as hybridization probes and for DNA sequence analysis. (*B*) Comparison of the recombination sites involved in the *nifD, fdxN*, and *hupL* rearrangements. In each case, site-specific recombination between the directly repeated sequences that flank the element results in excision of the element from the heterocyst chromosome. Nucleotides that are repeated at the borders of the elements are shown. Recombination occurs within the nucleotide sequences shown in boldface type and underlined.

determined. The DNA sequence showed that excision of the element occurs by reciprocal site-specific recombination between two 16-bp direct repeats that border the element. The positions of the recombination sites are shown in Fig. 2*A* and the DNA sequence of the 16-bp repeat is shown in Fig. 2*B*.

The recombination sites for the *nifD* (8), *fdxN* (18), and *hupL* rearrangements show no sequence similarities (Fig. 2*B*), which suggests that the *hupL* rearrangement requires its own site-specific recombinase.

A *hupL* Gene Is Rearranged. The 10.5-kb element interrupts an open reading frame in *Anabaena* 7120 vegetative cells that shows homology to hydrogenase genes. *hupL* genes encode the large subunit of [NiFe] hydrogenases (25, 26). Fig. 3 shows the nucleotide and predicted amino acid sequences of the rearranged *Anabaena* 7120 *hupL* gene. The predicted *Anabaena* 7120 HupL polypeptide is homologous to a large number of H_2-uptake [NiFe] hydrogenases; for example, it is 55% similar and 31% identical to the membrane-bound HupL of *Rhodobacter capsulatus* (27) and 53% similar and 31% identical to the periplasmic HydB from *Desulfovibrio fructosovorans* (28). The putative nickel-binding motifs found in the N- and C-terminal domains of [NiFe] hydrogenase large subunits are present in the *Anabaena* 7120 protein (Fig. 3); however, the *Anabaena* 7120 C-terminal motif contains a serine in a position occupied by a proline in other hydrogenases (26). The 3′ end of a partial open reading frame that shows similarity to *hupS*, which encodes the small subunit of [NiFe] hydrogenases, is present upstream of the *hupL* gene (Fig. 3 and data not shown).

The 10.5-kb DNA element is inserted within the first one-third of the *hupL* gene (Fig. 3). On the vegetative cell chromosome, the HupL polypeptide would be truncated after 170 amino acids (out of 531), and a stop codon is encountered within the element only four codons downstream of the 16-bp recombination site. It is unlikely that the unrearranged *hupL* gene would produce a functional product.

The *hupL* gene shows transcriptional regulation during heterocyst induction similar to that of the *nifH* gene. The *nifHDK* operon is transcribed late during heterocyst induction (12) in differentiating heterocysts (29). Northern blot analysis of *hupL* expression with a random primer-labeled DNA probe or a high-specific-activity RNA probe showed very weak hybridization signals to degraded transcripts only in total RNA samples isolated from late stages of heterocyst induction and from purified heterocysts (data not shown). These results suggested that the *hupL* gene is developmentally regulated and showed that *hupL* is only weakly expressed and that its message has a relatively short half-life. RT-PCR was used to detect *hupL* message during heterocyst development to confirm the Northern analysis results (Fig. 4). PCR amplified products from reverse-transcribed cDNA for *nifH* and *hupL* showed similar patterns of expression during heterocyst induction. For both messages, RT-PCR product was produced from total RNA isolated from heterocysts and from filaments 30 hr after nitrogen step-down. A small amount of *hupL* product was produced in the 24-hr sample and *nifH* product could be detected as early as the 12-hr sample. These data suggest that *hupL* is normally expressed only after excision of the 10.5-kb element in heterocysts.

Identification of the Recombinase Gene, *xisC*. A 1.5-kb open reading frame, the *xisC* gene, was identified 115 bp inside the right border of the *hupL* element (Fig. 2*A*). Comparison of the predicted XisC amino acid sequence with GenBank sequences identified apparent homology with a single protein: XisA, the

```
ACCGTATTTAAAACCCAAACCATCATGGGAGTACCCAAAGAACTACCGCCAGGAGTAAGC    60
T  V  F  K  T  Q  T  I  M  G  V  P  K  E  L  P  P  G  V  S
AATAAAAACTACGCTGTCCTCACAATGGTGGCTAAAGACACAGCCCCCAAATGGGCAGAA   120
N  K  N  Y  A  V  L  T  M  V  A  K  D  T  A  P  K  W  A  E
GAAGACTTTTTTCACAGTTTAGTCAGTGGTCAGTAGTCAATGGTCAGTTGCTACTGACAA   180
E  D  F  F  H  S  L  V  S  G  Q  *
CCGACAACCGACAACCAACAACTAACAAATATGACAATCAAAACATTAGATATATCACCC   240
                        M  T  I  K  T  L  D  I  S  P
GTCGGTAGAGTCGAGGGTGATTTAGATGTCCGTGTGGAAATAGAAGACGGAAGGGTAGTT   300
V  G  R  V  E  G  D  L  D  V  R  V  E  I  E  D  G  R  V  V
AACGCCTGGACACACGCCGAACTATTTCGCGGCTTTGAAATCATTCTTCGTGGTAAAGAC   360
N  A  W  T  H  A  E  L  F  R  G  F  E  I  I  L  R  G  K  D
CCCCAAGCTGGATTAATTGTTACGCCTCGTATCTGCGGTATCTGCGGTGCTTCTCACTTA   420
P  Q  A  G  L  I  V  T  P  R  I  C  G  I  C  G  A  S  H  L
ACCTCTGCATCATGGGCATTAGATACAGCATGGAATACAACCGTTCCCCGCAACGCCATC   480
T  S  A  S  W  A  L  D  T  A  W  N  T  T  V  P  R  N  A  I
TTAGCCAGAAACCTCGGTCAAATTGTCGAAACCATCCAAAGCATCCCCCGCTACTTTTAT   540
L  A  R  N  L  G  Q  I  V  E  T  I  Q  S  I  P  R  Y  F  Y
GGATTGTTCGCCATTGACCTAACAAATAAAAAATACCGTAGTAGCCGCTTCTACGATGAA   600
G  L  F  A  I  D  L  T  N  K  K  Y  R  S  S  R  F  Y  D  E
GCCGTCCGCCGCTTTTCTGCCTACACAGGTAAATCTTATGAACTCGGCGTGACCATTTCC   660
A  V  R  R  F  S  A  Y  T  G  K  S  Y  E  L  G  V  T  I  S
AGCAAACCCGTAGAAATTTACGCCCTGTTTGGCGGACAATGGCCGCACAGCAGTTATATG   720
S  K  P  V  E  I  Y  A  L  F  G  G  Q  W  P  H  S  S  Y  M
GTGCCTGGTGGTGTGATGTGCGCCCCCACCTTAACCGACATTACCCGCGCTTGGGCAATT   780
V  P  G  G  V  M  C  A  P  T  L  T  D  I  T  R  A  W  A  I
CTCGAATACTTCCGCACCAACTGGTTAGAACCTGTGTGGTTGGGTTGTTCTTTGGAACGC   840
L  E  Y  F  R  T  N  W  L  E  P  V  W  L  G  C  S  L  E  R
TACGAAGAAATTCAAACTTACGATGACTTCATGGACTGGTTAGAAGCAGACATCAAGCAT   900
Y  E  E  I  Q  T  Y  D  D  F  M  D  W  L  E  A  D  I  K  H
CGTGAGTCAGACTTGGGTTTCTATTGGCGTATGGGTTTAGATATCGGTTTGGATAGATAC   960
R  E  S  D  L  G  F  Y  W  R  M  G  L  D  I  G  L  D  R  Y
GGGGCTGGTGTGGGTAAATACGTCTCCTGGGGATACCTACCCCATGAAGATAAATACCAA  1020
G  A  G  V  G  K  Y  V  S  W  G  Y  L  P  H  E  D  K  Y  Q
AAGCCCACCATCGAAGGACGCAACGCCGCCATGATTATGAAAAGTGGTGTGTATGACAGC  1080
K  P  T  I  E  G  R  N  A  A  M  I  M  K  S  G  V  Y  D  S
TTCGAGAATACACACACTTTGATGGATCACACCTTCGCCCGTGAGAATACCACCCACGCC  1140
F  E  N  T  H  T  L  M  D  H  T  F  A  R  E  N  T  T  H  A
TGGTATGATGAAGGTAACGCAGATGTTCACCCCTTTGACCGTACCACCAAACCCACCCAC  1200
W  Y  D  E  G  N  A  D  V  H  P  F  D  R  T  T  K  P  T  H
AAAAATACTAAAGACTTCAAAAATGCCTACTCTTGGTCAACAGCCGTACTACACCAAGAC  1260
K  N  T  K  D  F  K  N  A  Y  S  W  S  T  A  V  L  H  Q  D
TTCGGACGCTTGGAAGTCGGCCCCCTAGCCCGCCAGCTAGTTGCAGGTGGTCAGCATGGC  1320
F  G  R  L  E  V  G  P  L  A  R  Q  L  V  A  G  G  Q  H  G
GAATCTTGGCAACATTACGACGGTTTTATCTTGGATGCTTTCCAAAAAATGGGTGGTGCT  1380
E  S  W  Q  H  Y  D  G  F  I  L  D  A  F  Q  K  M  G  G  A
AGTATTCATTTGCGTCAACTTGCACGAGTTCACGAAATTGTCAAGTTATATCGCCAAGCA  1440
S  I  H  L  R  Q  L  A  R  V  H  E  I  V  K  L  Y  R  Q  A
GAAAGATGTCTGCGTGAGTTTGTCTTAAATGATCCTTGGTATATCAAACCCAAAGAAAAA  1500
E  R  C  L  R  E  F  V  L  N  D  P  W  Y  I  K  P  K  E  K
GACGGACGTGGTTGGGGTGCAACGGAAGCCTCACGGGGTTCTTTGTGTCACTGGATAGAT  1560
D  G  R  G  W  G  A  T  E  A  S  R  G  S  L  C  H  W  I  D
ATAGAAGGTGGCAAGATTAAGAATTATCAGGTGATTGCTGCGACTACTTGGAACGTCGGC  1620
I  E  G  G  K  I  K  N  Y  Q  V  I  A  A  T  T  W  N  V  G
CCCCGTGATAGTGAAGGGGTACGCGGCCCTATTGAGGAAGCTTTAATCGGGACTCCTATT  1680
P  R  D  S  E  G  V  R  G  P  I  E  E  A  L  I  G  T  P  I
GAAGATTCTAGAGATCCTGTGGAAGTGGGACACGTTGCGCGATCGTTTGACTCTTGTCTG  1740
E  D  S  R  D  P  V  E  V  G  H  V  A  R  S  F  D  S  C  L
GTGTGTACTGTCCACGCCCATGATGCGAAGACAGGCGAGGAGTTGGCGCGTTTTCGGACT  1800
V  C  T  V  H  A  H  D  A  K  T  G  E  E  L  A  R  F  R  T
GCTTAAGTTTAAACGCAGAGGGGCGCGGAGGTCAGCGCGGAGGTACGCGGAGAGTGTGCT  1860
A  *
GTTAATTAATGAGTGGCTAATGGCTAGTTATGATTTAGTTAGTCATTAGCTCATTAATAT  1920
TTCATGGTTGAGAATTGCTGAGTGCTGTTGCTACGATGTG                      1960
```

FIG. 3. Nucleotide and predicted amino acid sequences of the *Anabaena* 7120 *hupL* gene. The 1593-bp *hupL* open reading frame encodes a 531-amino acid polypeptide with a predicted molecular weight of 60,202. The C-terminal 51 amino acids of *hupS* are shown at the beginning of the DNA sequence. The 16-bp recombination site on the heterocyst chromosome is shown in boldface type and doubly underlined. Amino acids highly conserved among membrane-bound [NiFe] hydrogenase large subunits are shown in boldface type. The four underlined cysteines, two in an N-terminal domain and two in a C-terminal domain, are thought to be involved in binding nickel. The nonconserved serine in the C-terminal domain is shown in italics. Asterisk, stop codon.

Anabaena 7120 *nifD*-element recombinase. Although XisC is 25% longer than XisA, the XisA amino acid sequence can be aligned along its entire length with XisC (Fig. 5); the sequences are 61% similar and 43% identical. The XisA sequence shown in Fig. 5 starts at the first of two in-frame ATG codons (15). Although recombinase activity is retained in *xisA* deletions missing the first ATG (13), it is not known which translation start site is used *in vivo*.

The position of the *xisC* gene on the *hupL* element and its homology with *xisA* strongly suggest that it encodes the site-specific recombinase responsible for the programed rearrangement of the *hupL* gene.

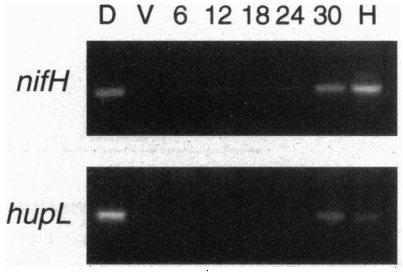

FIG. 4. *hupL* transcripts are detected during only the late stages of heterocyst development. RT-PCR with primers for *nifH* or *hupL* message was performed on a genomic DNA control sample (lane D) and RNA isolated from vegetative cells (lane V), heterocysts (lane H), and whole filaments collected at the indicated number of hours after induction of heterocyst development (lanes 6, 12, 18, 24, and 30). Proheterocysts were present at 18 hr and mature heterocysts were present at 24 hr. The RT-PCR products were resolved by electrophoresis on a 1.5% agarose gel stained with ethidium bromide. The RT-PCR products for *nifH* and *hupL* were 320 bp and 293 bp, respectively. Control experiments with the RNA samples, in which reverse transcriptase was omitted, failed to produce any amplified products.

DISCUSSION

The identification of the programed *hupL* rearrangement shows that *Anabaena* 7120 undergoes at least three developmentally regulated genome rearrangements during heterocyst differentiation. All three rearrangements involve the excision of DNA elements from the chromosome by site-specific recombination between short directly repeated sequences, and the rearrangements all occur late during heterocyst differentiation. In each case, the element is removed from the coding region of a developmentally regulated gene. Despite these similarities, each element encodes its own site-specific recombinase and is bordered by distinct recombination sites.

```
XisC  MSGQNQGNCEETYSTSTHEYANADSTDWFADMFADFEPQNTTDGSYRGTM 50
                                ..:: |.:.: |||:|| ..|||: |.
XisA  .....................MQNQGQDKYQQAFADLEPLSSTDGSFLGSS 30

XisC  AKIKATELQYQAVFEQKLVEANTNLKRERIRVSIKQTGNSLQLRATLPLK 100
      . ..   : .. . |.| ..| . |..:.:||:::...||||||||:|
XisA  LQAQQQREHMRTKVLQDLDKVNLRSKSAKTKVSVRESNGSLQLRATLPIK 80

XisC  PGD.GSLGKTKKQYDLSLGIPANLEGLKTAIEESYELGKLIARHTFEWNE 149
      | | :. |..:|||:|||.|||||:||||| ||.|||||||||.||||||
XisA  PRDKDTNGTGRKQYNLSLNIPANLDGLKTAEEEAYELGKLIARKTFEWND 130

XisC  KYLGIKSREKQEIKTIGELLDKFEEKYYQTRQKTITSQNTFPNYISVIKR 199
      |||| |. .|.: .|||:||:||.|.|:.|:.:|....:|| |:| ..|
XisA  KYLG.KEATKKDSQTIGDLLEKFAEEYFKTHKRTTKNKHTFFYYFSRTQR 179

XisC  NFPLTHLATKENFEEIINSVQGNK.KNELIAVTSVFIKTFNLGFQLDVKR 248
        ..||| ||: : |: :: :. : :   ..|.|. |||:::|.  .
XisA  YTNSKDLATAENLINSIEQIDKEWARYNAARAISAFCITFNIEIDLSQYS 229

XisC  DNVTPAHREIPEDDKIIYSFDLFEKFALNRKNTNISDEIDTWEMWRWVYG 298
      . .... |:||.|..|: ::. ||.: :.| |   .|  |.|::|||.||
XisA  KMPDRNSRNIPTDAEILSGITKFEDYLVTRGNQVNEDVKDSWQLWRWTYG 279

XisC  MLATFGLRPRELFVQPDINWWMSPQNLDHTWKVNKNTKTGYREVIPFVPE 348
      |||.|||||||:|:.|:|:|||:|.:|:| ||||:|:.||| |:.:|: .|
XisA  MLAVFGLRPREIFINPNIDWWLSKENIDLTWKVDKECKTGERQALPLHKE 329

XisC  WIELFDLKNPKPLKILEKKVTKIASVQ..NINWMRRDISRWFRKVGIEFQ 396
      ||: |||:||| |.:|.. :.| .... :|. : . ||:||||||:::|.
XisA  WIDEFDLRNPKYLEMLATAISKKDKTNHAEITALTQRISWWFRKVELDFK 379

XisC  PYDLRHACAIRAHLQGIPIKAAADNLGHTVDEHTKTYQRWFGIENRKKAF 446
      |||||||..:   :
XisA  PYDLRHAGQLERIF*..................................... 394
```

FIG. 5. Comparison of the predicted XisC and XisA protein sequences. The amino acid sequences of XisC (498 amino acids) and XisA (393 amino acids) were aligned by using the GAP program (default settings) from the Genetics Computer Group software package. The C-terminal 52 amino acids of XisC are not shown. Vertical lines indicate identical amino acids; double and single dots indicate evolutionarily conserved amino acid substitutions. Asterisk, stop codon.

It is not known how the rearrangements are coordinately regulated during heterocyst development. The recombinases must be active only during the late stages of heterocyst differentiation, but it is not known if the regulation occurs at the transcriptional or posttranscriptional level. The transcription of the *xisA* and *xisF* genes has been difficult to study because of their low levels of expression, and the transcription start site has not been identified for either gene. However, two *Anabaena* 7120 DNA-binding proteins, NtcA (BifA) and factor 2, have been shown to bind to sequences upstream of the *xisA* gene, and a consensus binding sequence for NtcA was determined (30, 31). An obvious NtcA binding site is not present in the upstream region of *xisC*.

All recombinases that have been studied *in vitro*, and many others identified genetically, can be assigned to one of two families: the resolvase family or the integrase family of recombinases (32). The *Anabaena* 7120 XisF recombinase, which is required for excision of the *fdxN* element, belongs to the resolvase family and is homologous to the *Bacillus subtilis* SpoIVCA site-specific recombinase (19). XisC and XisA represent another class of site-specific recombinases since they do not show similarity to either of these two families of recombinases or to any other protein in the data bases.

To our knowledge, the *Anabaena* 7120 *hupL* gene represents the only gene encoding a cyanobacterial membrane-bound [NiFe] uptake hydrogenase identified thus far. In diazotrophs, uptake hydrogenases function to utilize molecular hydrogen, which is a by-product of nitrogen fixation, for the energy-conserving reduction of electron acceptors (25, 26). Our results indicate that the *hupL* gene is only rearranged in heterocysts and only transcribed after heterocyst formation, about the same time that the 10.5-kb element is excised from the chromosome. We suspect that this *Anabaena* 7120 uptake hydrogenase is required to improve the efficiency of nitrogen fixation in heterocysts and that the putative *hupSL* operon is expressed only in heterocysts. Therefore, the *hupL* element could be carried in the vegetative cell genome without detriment to the organism.

Excision of the *nifD* and *fdxN* elements is required for the correct expression of the genes and operons in which they reside; failure to excise these elements results in heterocysts that are unable to fix nitrogen (14, 19). The *hupL* rearrangement may differ from the *nifD* and *fdxN* rearrangements in that failure to produce hydrogenase would not be expected to block nitrogen fixation in heterocysts or growth on nitrogen-free medium but would only degrade the efficiency of nitrogen fixation.

The third *Anabaena* 7120 genome rearrangement described here was independently identified by pulsed-field gel electrophoresis of vegetative cell and heterocyst DNA (33). Matveyev *et al.* (33) found that the vegetative cell 505-kb *Sal* I D fragment was shifted to a size 18 kb smaller in heterocyst DNA.

We thank T. S. Ramasubramanian and K. S. Ramaswamy for helpful discussions and comments on the manuscript. This work was supported by a grant (GM36890) from the National Institutes of Health.

1. Gellert, M. (1992) *Trends Genet.* **8,** 408–412.
2. Oettinger, M. A. (1992) *Trends Genet.* **8,** 413–416.
3. Tobler, H., Etter, A. & Muller, F. (1992) *Trends Genet.* **8,** 427–432.
4. Haber, J. E. (1992) *Trends Genet.* **8,** 446–452.
5. Haselkorn, R. (1992) *Annu. Rev. Genet.* **26,** 113–130.
6. Kunkel, B., Losick, R. & Stragier, P. (1990) *Genes Dev.* **4,** 525–535.
7. Sato, T., Samori, Y. & Kobayashi, Y. (1990) *J. Bacteriol.* **172,** 1092–1098.
8. Golden, J. W., Robinson, S. J. & Haselkorn, R. (1985) *Nature (London)* **314,** 419–423.
9. Wolk, C. P. (1982) in *The Biology of Cyanobacteria*, eds. Carr, N. G. & Whitton, B. A. (Univ. Calif., Berkeley), pp. 359–386.
10. Wolk, C. P. (1989) *Plant Syst. Evol.* **164,** 27–31.
11. Buikema, W. J. & Haselkorn, R. (1993) *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **44,** 33–52.
12. Golden, J. W., Whorff, L. L. & Wiest, D. R. (1991) *J. Bacteriol.* **173,** 7098–7105.
13. Brusca, J. S., Chastain, C. J. & Golden, J. W. (1990) *J. Bacteriol.* **172,** 3925–3931.
14. Golden, J. W. & Wiest, D. R. (1988) *Science* **242,** 1421–1423.
15. Lammers, P. J., Golden, J. W. & Haselkorn, R. (1986) *Cell* **44,** 905–911.
16. Golden, J. W., Carrasco, C. D., Mulligan, M. E., Schneider, G. J. & Haselkorn, R. (1988) *J. Bacteriol.* **170,** 5034–5041.
17. Mulligan, M. E. & Haselkorn, R. (1989) *J. Biol. Chem.* **264,** 19200–19207.
18. Golden, J. W., Mulligan, M. E. & Haselkorn, R. (1987) *Nature (London)* **327,** 526–529.
19. Carrasco, C. D., Ramaswamy, K. S., Ramasubramanian, T. S. & Golden, J. W. (1994) *Genes Dev.* **8,** 74–83.
20. Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A. & Struhl, K. (1994) *Current Protocols in Molecular Biology* (Greene & Wiley, New York).
21. Buikema, W. J. & Haselkorn, R. (1991) *J. Bacteriol.* **173,** 1879–1885.
22. Genetics Computer Group (1991) *Program Manual for the GCG Package* (Genetics Computer Group, Madison, WI).
23. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990) *J. Mol. Biol.* **215,** 403–410.
24. Kuritz, T., Ernst, A., Black, T. A. & Wolk, C. P. (1993) *Mol. Microbiol.* **8,** 101–110.
25. Przybyla, A. E., Robbins, J., Menon, N. & Peck, H. D., Jr. (1992) *FEMS Microbiol. Rev.* **8,** 109–135.
26. Wu, L. F. & Mandrand, M. A. (1993) *FEMS Microbiol. Rev.* **10,** 243–269.
27. Leclerc, M., Colbeau, A., Cauvin, B. & Vignais, P. M. (1988) *Mol. Gen. Genet.* **214,** 97–107.
28. Rousset, M., Dermoun, Z., Hatchikian, C. E. & Belaich, J.-P. P. (1990) *Gene* **94,** 95–101.
29. Elhai, J. & Wolk, C. P. (1990) *EMBO J.* **9,** 3379–3388.
30. Chastain, C. J., Brusca, J. S., Ramasubramanian, T. S., Wei, T.-F. & Golden, J. W. (1990) *J. Bacteriol.* **172,** 5044–5051.
31. Ramasubramanian, T. S., Wei, T.-F. & Golden, J. W. (1994) *J. Bacteriol.* **176,** 1214–1223.
32. Stark, W. M., Boocock, M. R. & Sherratt, D. J. (1992) *Trends Genet.* **8,** 432–439.
33. Matveyev, A. V., Rutgers, E., Söderbäck, E. & Bergman, B. (1994) *FEMS Microbiol. Lett.* **116,** 201–208.