JOURNAL OF BACTERIOLOGY, Feb. 1989, p. 1017–1027
0021-9193/89/021017-11$02.00/0

Vol. 171, No. 2

Physical and Genetic Map of the Major *nif* Gene Cluster from *Azotobacter vinelandii*

MARTY R. JACOBSON,[1] KEVIN E. BRIGLE,[1] LISA T. BENNETT,[1] ROBERT A. SETTERQUIST,[1] MARK S. WILSON,[1] VALERIE L. CASH,[1] JIM BEYNON,[2] WILLIAM E. NEWTON[3] AND DENNIS R. DEAN[1]*

Department of Anaerobic Microbiology, Virginia Polytechnic Institute and State University, Blacksburg, Virginia, 24061[1]; BioTechnica International, Inc., Cambridge, Massachusetts, 02140[2]; and Western Regional Research Center, Agricultural Research Service, U.S. Department of Agriculture, Albany, California 94710[3]

Received 29 August 1988/Accepted 2 November 1988

Determination of a 28,793-base-pair DNA sequence of a region from the *Azotobacter vinelandii* genome that includes and flanks the nitrogenase structural gene region was completed. This information was used to revise the previously proposed organization of the major *nif* cluster. The major *nif* cluster from *A. vinelandii* encodes 15 *nif*-specific genes whose products bear significant structural identity to the corresponding *nif*-specific gene products from *Klebsiella pneumoniae*. These genes include *nifH*, *nifD*, *nifK*, *nifT*, *nifY*, *nifE*, *nifN*, *nifX*, *nifU*, *nifS*, *nifV*, *nifW*, *nifZ*, *nifM*, and *nifF*. Although there are significant spatial differences, the identified *A. vinelandii nif*-specific genes have the same sequential arrangement as the corresponding *nif*-specific genes from *K. pneumoniae*. Twelve other potential genes whose expression could be subject to *nif*-specific regulation were also found interspersed among the identified *nif*-specific genes. These potential genes do not encode products that are structurally related to the identified *nif*-specific gene products. Eleven potential *nif*-specific promoters were identified within the major *nif* cluster, and nine of these are preceded by an appropriate upstream activator sequence. A+T-rich regions were identified between 8 of the 11 proposed *nif* promoter sequences and their upstream activator sequences. Site-directed deletion-and-insertion mutagenesis was used to establish a genetic map of the major *nif* cluster.

Azotobacter vinelandii is capable of diazotrophic growth using any of three distinct nitrogen fixation systems (see references 6 and 32). Molybdenum-dependent nitrogenase is a complex metalloenzyme composed of two component proteins called the Fe protein and the MoFe protein. The Fe protein acts as a specific ATP-binding, one-electron reductant of the MoFe protein, which contains the site(s) for substrate binding and reduction. Native Fe protein is an approximately 60,000-dalton dimer of identical subunits. A single 4Fe-4S cluster is believed to be symmetrically bridged between the Fe protein subunits. The MoFe protein is an $\alpha_2\beta_2$ protein with an M_r of about 220,000 and it contains two Mo atoms and about 32 Fe and 32 S^- atoms per molecule. About 16 of these Fe atoms can be extruded from each MoFe protein molecule in the form of 4Fe-4S clusters by treatment of the native protein with thiols in a denaturing organic solvent. All or most of the remaining Fe and both Mo atoms constitute the two identical iron-molybdenum cofactors. In addition to magnesium ATP, a source of reducing equivalents, protons, and an anaerobic environment are required for nitrogenase turnover. For a recent review, see reference 45.

Two other nitrogen-fixing complexes that are structurally related to but genetically distinct from the Mo-dependent nitrogenase discussed above have recently been identified in the aerobic diazotroph *A. vinelandii* (see discussion in reference 15). These systems apparently have catalytic components that are analogous to the Fe protein and the MoFe protein discussed above. A major difference in these various nitrogen-fixing systems appears to reside within the metal composition of their cofactor species (15, 21). Interestingly, the various nitrogen-fixing systems appear to share some gene products that are required for the maturation of their respective catalytic components (28, 29).

The biochemical complexity of nitrogenase is reflected in the genetic organization and in the regulation of expression of the components required for the synthesis of a catalytically competent nitrogenase. For example, activation of the nitrogenase MoFe protein requires the biosynthetic activity of at least six gene products that participate in the assembly of the FeMo cofactor center (see discussion in reference 40). In the facultative anaerobe *Klebsiella pneumoniae*, all of the *nif*-specific genes are clustered within a single region of the genome and organized into eight transcriptional units (see Fig. 1). At least two of these transcriptional units overlap (5, 37). There are now 20 proposed *nif*-specific genes from *K. pneumoniae*, and the nucleotide sequence for all of them, *nifJ* (M. Cannon, F. Cannon, V. Buchanan-Wollaston, D. Ally, A. Ally, and J. Beynon, Nucleic Acids Res., in press), *nifH* (42, 47), *nifD* (26, 42), *nifK* (24), *nifT* (J. Beynon, M. Cannon, A. Ally, V. Buchanan-Wollaston, R. Setterquist, D. R. Dean, and F. Cannon, Nucleic Acids Res., in press), *nifY* (Beynon et al., in press), *nifE* (43), *nifN* (43), *nifX* (Beynon et al., in press), *nifU* (4), *nifS* (4), *nifV* (4), *nifW* (37; Beynon et al., in press), *nifZ* (37), *nifM* (37), *nifF* (18), *nifL* (19), *nifA* (14, 17), *nifB* (13), and *nifQ* (13), has been determined. In addition to the work cited above, the nucleotide sequence of the *K. pneumoniae nif* cluster was also determined independently (W. Arnold, A. Rump, W. Klipp, V. B. Priefer, and A. Puhler, J. Mol. Biol., in press). The *K. pneumoniae nif*-specific genes encode the following components: *nifH* (Fe protein subunit); *nifD* (MoFe protein α subunit); *nifK* (MoFe protein β subunit); *nifF* and *nifJ* (electron transport components); *nifE*, *nifN*, *nifV*, *nifB*, and *nifQ* (FeMo cofactor biosynthetic enzymes); *nifM* (Fe protein maturation component); *nifL* (negative regulatory ele-

* Corresponding author.

ment); *nifA* (positive regulatory element); *nifT*, *nifY*, *nifX*, *nifU*, *nifS*, *nifW*, and *nifZ* (functions of products not known).

Recent studies in our laboratories have focused on a comparative analysis of *nif*-specific genes from *K. pneumoniae* and the obligate aerobe *A. vinelandii*. Twelve *nif*-specific genes from *A. vinelandii* were previously isolated and sequenced, and the products of these genes share structural identity with the corresponding *K. pneumoniae* gene products (2–4, 8, 9, 11, 16, 28). These genes are clustered into two different linkage groups. The *nifA*, *nifB*, and *nifQ* genes are contained within one of these gene clusters (2, 28), and the *nifH*, *nifD*, *nifK*, *nifE*, *nifN*, *nifU*, *nifS*, *nifV*, and *nifF* genes are contained within the other (3, 4, 8, 9, 11, 16). Although they are split into two different linkage groups, all of the *nif*-specific genes identified from *A. vinelandii* have a sequential arrangement within their respective clusters identical to that found in *K. pneumoniae*.

In the present study, we completed the nucleotide sequence of a 28,793-base-pair region from the *A. vinelandii* genome that includes the major *nif* cluster from *A. vinelandii* and compared this region with the *K. pneumoniae nif* cluster. In addition, we genetically analyzed this region from *A. vinelandii* by introducing specific deletion, insertion, and insertion-plus-deletion mutations into the *A. vinelandii* chromosome.

MATERIALS AND METHODS

Materials. All of the restriction endonucleases and the large fragment of DNA polymerase I (Klenow) used in this study were purchased from either Bethesda Research Laboratories, Inc. (Gaithersburg, Md.), New England BioLabs, Inc. (Beverly, Mass.), or Boehringer Mannheim Biochemicals (Indianapolis, Ind.). Deoxynucleoside triphosphates (dATP, dCTP, dTTP, and 7-deazo-dGTP), dideoxynucleoside triphosphates (ddATP, ddCTP, ddGTP, and ddTTP), and plasmid pUC4-KAPA were purchased from Pharmacia (Piscataway, N.J.). [α-^{35}S]dATP was purchased from Dupont, NEN Research Products (Boston, Mass.). Ultrapure acrylamide and DNA grade formamide were obtained from Bethesda Research Laboratories. Ultrapure urea was purchased from Boehringer Mannheim Biochemicals. Cesium chloride, ethidium bromide, and all of the antibiotics used were purchased from Sigma Chemical Co. (St. Louis, Mo.). All other chemicals were reagent grade and available commercially.

Growth of *A. vinelandii* strains. The wild-type and mutant strains of *A. vinelandii* OP were grown at 30°C on modified Burk medium (46) containing 0.01 mM $NaMoO_4$. Thus, only diazotrophic growth using the molybdenum-containing nitrogenase system was examined in this study. When a fixed nitrogen source was required, ammonium acetate was added to a final concentration of 30 mM. Kanamycin was added to a final concentration of 0.5 μg/ml, and rifampin was added to a final concentration of 5.0 μg/ml as required.

***Escherichia coli* growth and DNA fragment isolation.** *E. coli* 71-18 [Δ(*lac-proAB*) *thi supE* (F′ *proAB lacI*^{q}*ZM15*)], which served as a host for all of the plasmids and the M13 bacteriophage used in this study, was grown and maintained as described by Messing (33). When necessary, the medium was supplemented with 100 μg of ampicillin per ml or 50 μg of kanamycin per ml. Recombinant M13 phage was grown and M13 single-stranded and replicative form DNAs were prepared as described in detail by Messing (33). DNA fragments used for cloning were purified as described by Maniatis et al. (30).

TABLE 1. Plasmids used as a primary source of DNA fragments used for sequencing experiments

Plasmid	Fragment[a]	Region cloned[b]	*nif* gene(s) or ORF(s) cloned[c]	Vector
pDB1	*Xho*I	---10297	*HDKTY, 1, 2, EN′*	pKT230
pDB6	*Sma*I	716–6349	*HDKTY′*	pUC8
pDB7	*Sma*I	6350–9743	*′Y 1, 2 EN′*	pUC8
pDB17	*Sal*I	208–1984	*H′*	pBR325
pDB32	*Bgl*II	1447–6164	*′HDKTY′*	pUC8
pDB38	*Hin*dIII	16539–24448	*′5, 6, USV, 7, 8, WZM, 9′*	pKT230
pDB42	*Hin*dIII	16539–24448	*′5, 6, USV, 7, 8, WZM, 9′*	pUC8
pDB49	*Sal*I	18674–20390	*′USV′*	pUC8
pDB50	*Sal*I	16627–17673	*′5, 6′*	pUC8
pDB54	*Bgl*II	6165–12562	*′Y, 1, 2, ENX, 3, 4*	pUC8
pDB56	*Eco*RI	15857–19405	*′10, 5, 6, US′*	pBR322
pDB58	*Eco*RI	9029–14940	*′ENX 3, 4, 11*	pBR322
pDB59	*Eco*RI	19406–25243	*′SV, 7, 8, WZM, 9′*	pBR322
pDB89	*Xho*I	25114–27070	*′9, F*	pKT230
pDB90	*Xho*I	27071–28789		pKT230
pDB104	*Sal*I	18323–18673	*′U′*	pUC8
pDB106	*Sma*I	20017–21488	*′SV, 7′*	pUC7
pDB124	*Xho*I	22718–24225	*′WZM′*	pUC7
pDB129	*Xho*I	17071–20504	*′5, 6, USV′*	pKT230
pDB151	*Sma*I	24092–25095	*′M, 9′*	pUC9
pDB170	*Sal*I	24815–26294	*′9, F′*	pUC7
pDB176	*Xho*I	20505–22223	*′V, 7, 8′*	pUC7
pDB177	*Pst*I	22318–23148	*′8, WZ′*	pUC7

[a] Restriction enzyme sites on the left and right termini of the cloned *A. vinelandii* genomic fragment.

[b] The numbers refer to the region cloned and correspond to the region sequenced. The numbers can be correlated to the numbering on the physical map shown in Fig. 1. The broken line preceding the number in the pDB1 line indicates that the sequence of the left end of that clone was not determined.

[c] Proposed genes contained within the cloned fragment. A prime on the left side of a gene or ORF indicates that the amino-terminal portion of that gene was not included in the cloned region, and a prime on the right side of a gene or ORF indicates that the carboxy-terminal portion of that gene or ORF was not included in the cloned region.

Plasmids and DNA sequence analysis. Preparation, restriction enzyme digestion, and ligation of hybrid plasmid DNAs were performed by previously described techniques (9). About 200 different hybrid plasmids containing DNA from the region sequenced in this study were constructed. Some of the plasmids that served as the major source of DNA for sequencing experiments are shown in Table 1. All DNA sequence analyses were performed by the dideoxy-chain termination procedure (41) with hybrids of M13 filamentous phage vectors described by Messing (33). For sequencing experiments, the appropriate hybrid plasmid was digested with the individual restriction enzyme *Pst*I, *Sma*I, *Xho*I, *Sal*I, *Hin*dIII, *Eco*RI, *Kpn*I, *Sst*I, or *Bam*HI or some combination of these enzymes, and the resulting DNA fragments were purified. The individually purified DNA fragments were either ligated into the appropriately digested M13 vector DNA for sequence analysis or further digested with the individual restriction enzyme *Sau*3A, *Hin*PI, *Alu*I, *Rsa*I, or *Msp*I and shotgun cloned into the appropriately digested M13 phage vector DNA for random sequence analysis. All sequences were determined in an overlapping fashion in both directions, except for short segments adjacent to the 5′ and 3′ ends of the entire sequenced region. These regions were sequenced in only one direction but from at least two differently isolated templates. To reexamine previously published sequences, we prepared synthetic oligonucleotides (Applied Biosystems 381A automated oligonucleotide synthesizer) to use as sequencing primers for these experiments.

Mutagenesis of *A. vinelandii* genes. Transformations of *A. vinelandii* were performed in liquid cultures as described by Page and von Tigerstrom (36). Specific *nif* deletions contained within various hybrid plasmids (see Table 2) were recombined into the *A. vinelandii* chromosome by congression (coincident transfer of unlinked genetic markers), with rifampin resistance as the selected marker. This procedure was described in detail previously (39). Congression was accomplished by adding equal amounts (1 μg each) of genomic Rif^r DNA and a particular *nif* deletion plasmid DNA to competent cultures of wild-type Rif^s *A. vinelandii* cells. After allowing time for phenotypic lag, the transformed cultures were spread on Burk ammonium acetate-supplemented agar plates containing rifampin. Rif^r transformants were scored on Burk nitrogen-free and Burk ammonium acetate-supplemented agar plates to identify Nif$^-$ transformants. The frequency of congression was highly variable, ranging from about 2 to 0.01% of the total Rif^r transformants. This frequency correlated with the lengths of the *A. vinelandii* sequences contained within the recombinant plasmid and were therefore available for reciprocal recombination with the chromosome.

The kanamycin resistance-encoding cartridge (including the natural kanamycin resistance-encoding gene promoter), which had been cloned into the central portion of *A. vinelandii* DNA sequences contained within a particular hybrid plasmid, was recombined into the *A. vinelandii* genome by double-reciprocal recombination events during transformation. For a detailed description of this procedure, see reference 3. All mutant constructs involved double-crossover events during transformation. Such double-crossover events resulted in exchange of either an insertion or a deletion plus an insertion with the homologous region on the host chromosome and subsequent loss of the plasmid vector. The plasmid vectors used in these experiments are not capable of autonomous replication in *A. vinelandii*. Double-crossover event recombinants were easily distinguished from single-crossover event recombinants, since single-crossover event recombinants became endowed with the antibiotic resistance phenotype. (Amp^r) carried on the plasmid vector. Transformed cells were grown for about 18 h in liquid Burk medium supplemented with ammonium acetate. After growth, cells that had recombined the kanamycin resistance determinant into their chromosome via double-crossover events were recovered by being plated on Burk agar plates supplemented with ammonium acetate and kanamycin. The location of the Kan^r determinant within the genome of each of the resulting mutant strains was confirmed by Southern hybridization, marker rescue experiments, or two-factor crosses.

A third procedure, which is a combination of the above two methods, was used to delete regions from the *A. vinelandii* chromosome that did not result in a discernible Nif$^-$ phenotype. In step 1, a kanamycin resistance-encoding cartridge was introduced into a specific position within the chromosome as described above. If the resultant strain had a Nif$^+$ phenotype, as well as the Kan^r character, a deletion could be placed at the position of the Kan^r insertion by transformation with the appropriate deletion plasmid DNA, followed by a test for congression. In these experiments, the source of Rif^r DNA was a strain that was deleted for a large portion of the major *nif* cluster (DJ40). The resultant Rif^r Nif$^+$ transformants were then scored for the Kan^s phenotype.

Computer analysis. DNA sequences and deduced protein sequences were analyzed with the Pustell DNA sequence computer program available from International Biotechnologies Inc. (New Haven, Conn.). Computer-assisted analysis was used to identify potential open reading frames (ORFs). The codon usage bias for the identified *nif*-specific genes from within the major *nif* cluster was used to generate the preferred codon usage bias in this analysis.

RESULTS AND DISCUSSION

DNA sequence and physical organization. In the present study, we determined the nucleotide sequence of approximately 17,000 base pairs from the *A. vinelandii* chromosome within the major *nif* cluster. This sequence includes all of the regions located between the previously reported sequences within the major *nif* cluster, as well as about 1 kilobase pair preceding the *nifH* gene and 2 kilobase pairs following the *nifF* gene. The *nif* structural gene region was also resequenced. This analysis confirmed revisions within the original structural gene region sequence (9) proposed by Hiratsuka and Roy (23), except that we observed two T residues at positions 2179 and 2180 located within the *nifH-nifD* intergenic region (for numbering, see Fig. 1). There were also four neutral nucleotide substitutions within the reported *nifH* sequence and four errors within the region between the *nifH* promoter and the upstream activator sequence in the originally published sequence (9). This information completed the sequence for 28,793 continuous base pairs from within the major *A. vinelandii nif* cluster and permitted the revision of our previously proposed physical organization of this region (see reference 32 for the most recent comparative discussion of *A. vinelandii*, *A. chroococcum*, and *K. pneumoniae nif* gene organizations). The sequence has been submitted to the Genbank library and has the accession code M20568. The sequence can also be obtained from us in diskette or printout format.

In Fig. 1, the physical organization of the *K. pneumoniae nif* cluster and the major *nif* cluster from *A. vinelandii* are compared. We identified 30 potential genes within this region in *A. vinelandii*. Individual *A. vinelandii nif*-specific genes were assigned on the basis of significant sequence identity of the deduced gene product with its counterpart among the *K. pneumoniae nif*-specific gene products. The products of *A. vinelandii nif*-specific genes whose sequence has not been reported previously (*nifT*, *nifY*, *nifX*, *nifW*, *nifZ*, and *nifM*) are compared with their *K. pneumoniae nif* gene counterparts in Fig. 2. Although no essential functions related to nitrogen fixation have been established for the proposed *nifT*-, *nifY*-, *nifX*-, *nifW*-, and *nifZ*-encoded gene products, the similar organizations of these genes within the respective clusters, as well as the conservation in the primary sequence of their products, strongly suggest that they are likely to have *nif*-related functions. Kennedy et al. (29) have previously shown that the *K. pneumoniae nifM* gene can be used to complement an *A. vinelandii nifM* insertion mutation. This result demonstrated that the *K. pneumoniae nifM* gene product is functional in *A. vinelandii*. It was, therefore, surprising to discover the low level of sequence identity when the respective *nifM* gene products were compared (Fig. 2). These two gene products showed an overall amino acid sequence identity of only 16%. This sequence identity is confined to the C-terminal third of the polypeptides, and remarkably, the sequences that make up the N-terminal two-thirds of the polypeptides share little sequence identity. These comparisons suggest that the active portion of the *nifM* product is likely to be located within the C-terminal region of the polypeptide.

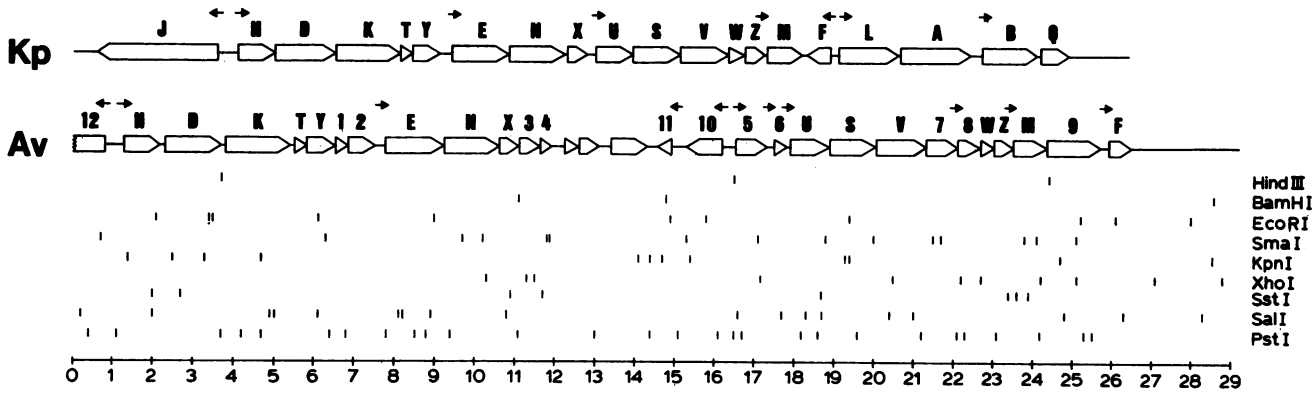

FIG. 1. Physical map of the major *nif* cluster from *A. vinelandii* and comparison with that of *K. pneumoniae*. The upper portion corresponds to the organization of *nif*-specific genes from *K. pneumoniae* (Kp), and the lower portion corresponds to the physical map for the major *nif* cluster from *A. vinelandii* (Av). Arrows indicate the approximate position and direction of identified or proposed *nif*-specific promoters in each organism. *A. vinelandii* genes whose deduced products have sequence identity when compared with *K. pneumoniae* *nif*-specific gene products have the appropriate *nif* genotypic designations. ORFs that are potentially cotranscribed with the identified *nif*-specific genes but do not share sequence identity with any of the identified *nif*-specific genes are numbered. ORFs preceded by a canonical *nif*-specific promoter sequence and an upstream activator sequence are also numbered. ORFs not apparently cotranscribed with any of the identified *nif*-specific genes and not preceded by a *nif*-specific promoter sequence are not labeled. Positions of restriction enzyme sites within the *A. vinelandii* major *nif* cluster are indicated by bars below the gene map. Numbering refers to the distance in kilobase pairs from the first nucleotide in the continuous sequence submitted to the Genbank library (accession code, M20568). This numbering can also be used to correlate cloned regions, positions of proposed genes, positions of proposed promoter sequences, positions of A+T-rich regions, and positions of specific mutations given in the other figures and tables.

ORFs which encode gene products that either appear to be cotranscribed with the identified *A. vinelandii nif*-specific genes or are preceded by potential *nif*-specific promoter sequences (see below) are numbered in Fig. 1. No significant sequence identity was observed when the products of the numbered genes were compared to the gene products encoded within the *K. pneumoniae nif* cluster. However, it is not known whether analogs to the numbered genes from *A. vinelandii* exist elsewhere on the *K. pneumoniae* chromosome. Each of these ORFs is preceded by a ribosome-binding site and has a biased codon usage similar to that for the identified *A. vinelandii nif*-specific genes. ORF12 is

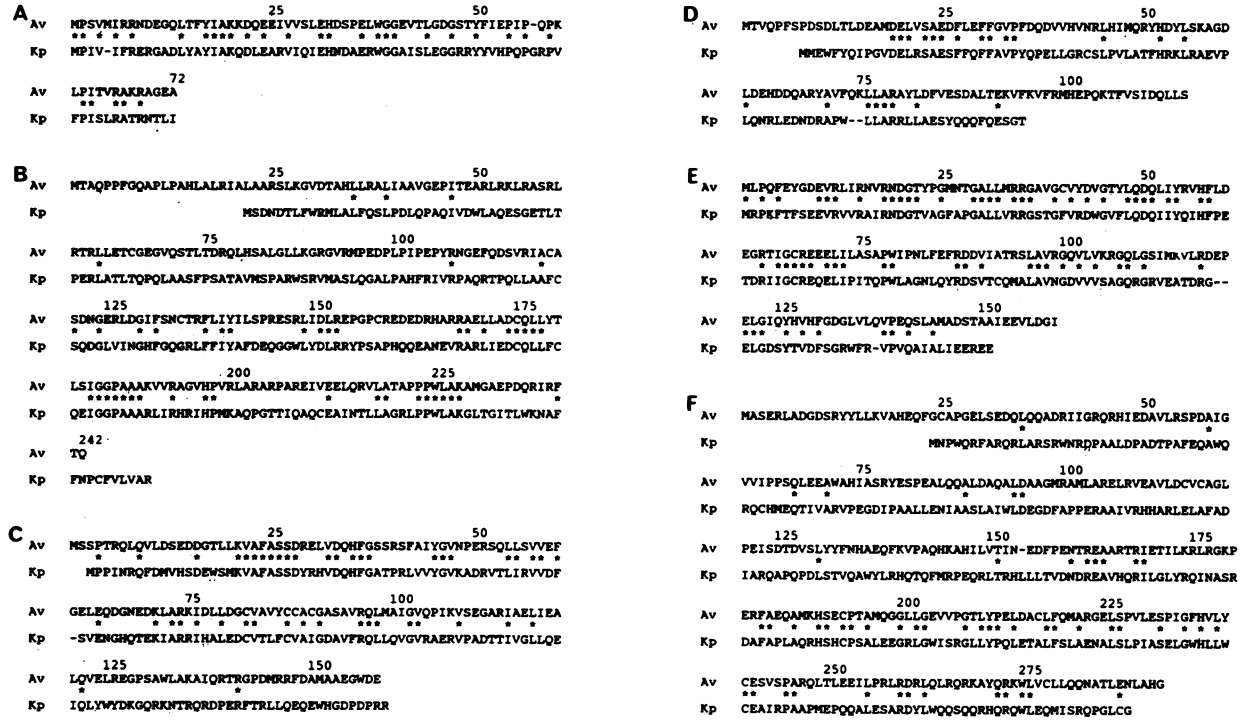

FIG. 2. Comparison of *A. vinelandii* (Av) and *K. pneumoniae* (Kp) *nifT* (A), *nifY* (B), *nifX* (C), *nifW* (D), *nifZ* (E), and *nifM* (F) gene products. Perfect identities are indicated by asterisks. Alignment adjustments made to provide the best fit are indicated by hyphens. Numbers refer to the *A. vinelandii* sequence. The entire sequence of each polypeptide is shown.

located in the same relative position as the *nifJ* gene from *K. pneumoniae* (Fig. 1). Comparison of the available ORF12-encoded product sequence with the *K. pneumoniae nifJ* gene product sequence (Cannon et al., in press) revealed no significant sequence identity. This feature, however, does not rule out the possibility that ORF12 encodes a product with a *nifJ*-like product function, because sequence identity could be restricted to only a small portion of the respective gene products. The comparison of the *A. vinelandii* and *K. pneumoniae nifM*-encoded products revealed one such example of analogous *nif* gene products having limited sequence identity (Fig. 2). Three other potential ORFs which are located between ORF4 and ORF11 (Fig. 1) and exhibit a codon usage bias characteristic of the other *A. vinelandii nif*-specific genes were also identified. Although preceded by reasonable translation initiation signals, these potential genes are not located immediately adjacent to any of the *nif*-specific genes or preceded by a *nif*-like promoter sequence. Consequently, we did not assign these ORFs numerical or genotypic designations. One of these proposed genes potentially encodes a rather small polypeptide (only 51 residues) that has two striking features. This potential gene product has six alanine residues located at the N terminus. Furthermore, four cysteines are located within this polypeptide spaced in the arrangement Cys-X-X-Cys-X-X-X-Cys-X-X-X-Cys. There were also numerous other potential ORFs within the *A. vinelandii* sequence. However, these either were not preceded by the appropriate translation initiation sequences or did not have the appropriate codon usage bias (see later discussion). In Table 2, the positions of the assigned coding regions within the major *nif* cluster from *A. vinelandii* and the deduced molecular masses and pIs of the corresponding gene products are given. A number of these genes have overlapping translation initiation and termination signals. These genes include *nifT-nifY*, *nifE-nifN-nifX*, *nifU-nifS*, *nifV*-ORF7, ORF8-*nifW*, and *nifZ-nifM*-ORF9. Such overlapping translational signals are frequently observed for cotranscribed genes, and this feature is indicative of translational coupling between adjacent genes (1, 35).

The physical organizations of *nif* genes from *K. pneumoniae* and the *nif* genes contained within the major *nif* cluster from *A. vinelandii* are remarkably similar. The *nif* genes identified from both organisms are all arranged in the same sequential order. However, a striking difference is the presence of potential genes within the *A. vinelandii* cluster that are not present within the *K. pneumoniae* gene cluster. In addition to those described herein, Joerger and Bishop (28) found that the *A. vinelandii nifB* and *nifQ* genes are separated by two potential genes. In *K. pneumoniae*, *nifB* and *nifQ* are immediately adjacent to one another and are contained within a single transcription unit (13; Fig. 1). On the basis of restriction enzyme mapping data, we previously proposed that *nifF* from *A. vinelandii* is transcribed in the opposite direction relative to the structural gene cluster (3), as in *K. pneumoniae*. However, the DNA sequence revealed that this conclusion was incorrect and that *nifF* is transcribed in the same direction as *nifHDK*. This feature, therefore, represents a significant difference between the organizations of the respective *A. vinelandii* and *K. pneumoniae nif* clusters (Fig. 1).

Promoter regions. Beynon et al. (5) proposed a canonical *nif*-specific promoter sequence that was based on a comparison of sequences preceding the individual transcription initiation sites for the *nifH*, *nifE*, *nifU*, *nifM*, *nifF*, *nifL*, and *nifB* genes from *K. pneumoniae*. This sequence (CTGG-N_8-TTGCA), which is characteristic of NtrA (RpoN)-dependent *nif* promoters (see reference 32), is generally conserved among the known and proposed *nif*-specific promoters from all gram-negative diazotrophs (see, for example, *nif*-specific promoters from *Bradyrhizobium japonicum* [22]), and the subsequence GG-N_{10}-GC is invariant. In addition, Buck et al. (12) have identified a *nif*-specific upstream activator sequence (TGT-N_{10}-ACA) characteristic of NifA-dependent promoters, located about 100 to 150 base pairs from the various *nif*-specific transcription initiation sites. In the case of *A. vinelandii*, it was shown that expression of the nitrogenase structural gene products is dependent on the presence of both the *nifA* gene product (2) and the *ntrA* (*rpoN*) gene product (48). We therefore examined the *A. vinelandii* sequence for potential consensus *nif* promoter and upstream activator sequences.

Eleven potential *nif* promoters (Fig. 1, arrows) were identified, and nine of these are preceded by potential upstream activator sequences (Fig. 3). In the case of the proposed *nifU* and *nifM* promoters, there are no apparent activator sequences located immediately upstream. However, the activator sequence preceding the proposed ORF6 promoter and that preceding the proposed ORF8 promoter could potentially serve to activate the *nifU* promoter and the *nifM* promoter, respectively (Fig. 3). The *K. pneumoniae nifU* promoter and the proposed *A. vinelandii nifU* promoter are both located at about the same distance from the proposed translation initiation sites of their respective *nifU* genes (4). A similar situation exists for the *K. pneumoniae*

TABLE 2. Features of proposed products encoded by genes within the major *nif* cluster

nif gene or ORF	Coding region[a]	Total amino acids	Product molecular mass (daltons)[b]	Calculated product pI
H	1240–2112	290	31,496	5.46
D	2238–3717	492	55,267	6.34
K	3817–5388	523	59,438	6.34
T	5520–5738	72	8,038	5.51
Y	5743–6471	242	26,683	11.41
1	6495–6757	89	9,552	4.93
2	6772–7506	244	27,963	5.99
E	7765–9189	474	52,141	6.34
N	9199–10575	458	49,187	6.06
X	10580–11056	158	17,270	5.39
3	11087–11569	160	18,087	5.79
4	11589–11798	69	7,855	5.92
5	16527–17288	253	27,491	11.85
6	17511–17834	107	11,029	5.13
U	17899–18837	312	33,251	5.71
S	18839–20047	402	43,578	6.04
V	20120–21277	385	41,633	6.04
7	21274–22071	265	28,327	9.28
8	22100–22639	179	19,899	5.46
W	22636–22983	115	13,399	5.25
Z	23003–23482	159	17,774	5.56
M	23472–24350	292	32,782	5.99
9	24343–25665	440	47,961	6.04
F	25908–26450	180	19,643	5.18
10	16155–15079	358	39,627	10.88
11	14899–14681	72	7,985	7.00
12	786–			

[a] The numbers refer to the position of the proposed coding region within the sequenced portion of the *A. vinelandii* genome and can be correlated with the numbering on the physical map shown in Fig. 1. The A residue in the initiation codon and position 3 in the termination codon are included in these numbers.
[b] The calculated molecular mass includes the initiating amino acid.

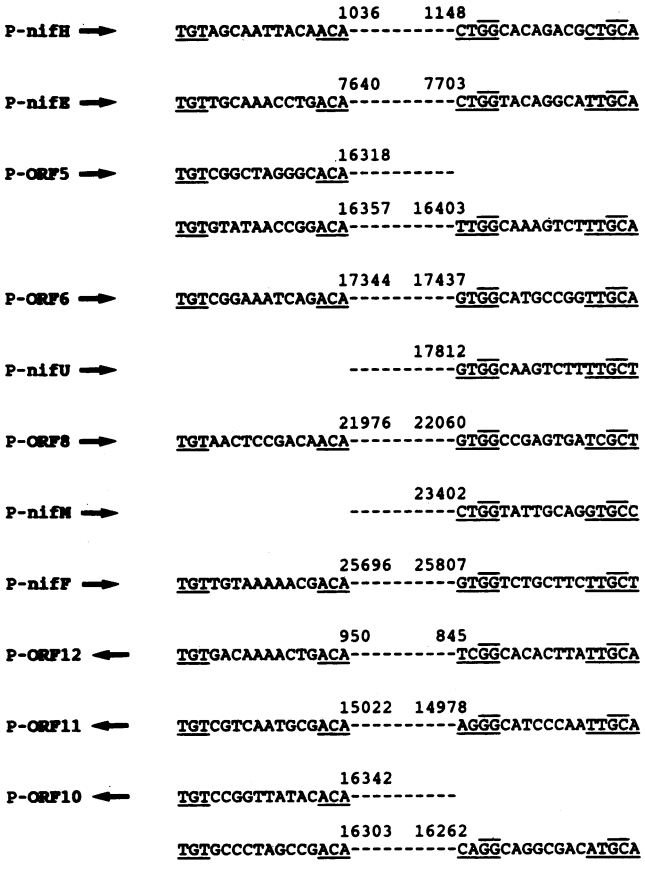

FIG. 3. Comparison of potential promoter sequences and upstream activator sequences identified within the major *nif* cluster from *A. vinelandii*. Arrows indicate the directions of transcription of the proposed promoter sequences as shown in Fig. 1. Numbers refer to the positions of the proposed upstream activator sequence and the promoter sequence within the sequenced portion of the *A. vinelandii* genome and can be correlated with the numbering on the physical map shown in Fig. 1. An extra C residue was placed between residues 1033 and 1034 in the proposed upstream activator region for *nifH* in the original sequence (9). Regions corresponding to the proposed canonical upstream activator sequence (12) or the canonical *nif*-specific promoter sequence (5) are underlined. Sequences that are invariant in proposed or identified *nif*-specific promoters are overlined. Potential promoters are named by the designations of the genes that immediately follow.

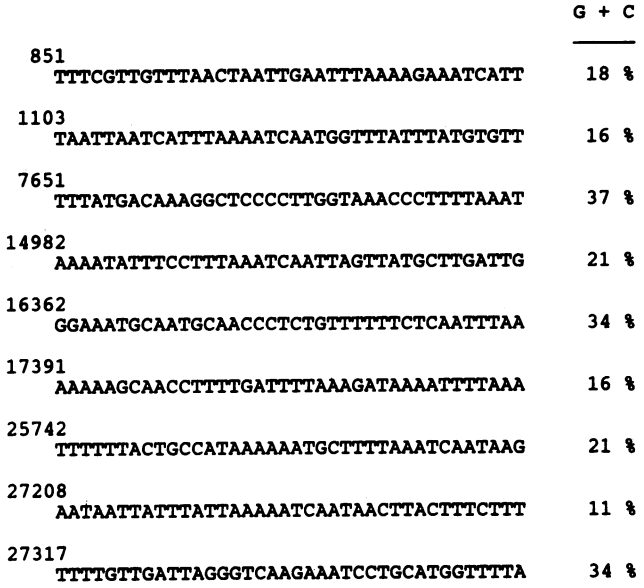

FIG. 4. Positions of A+T-rich regions within and following the major *nif* cluster from *A. vinelandii*. Numbers refer to the location of each A+T-rich region located within the sequenced portion of the *A. vinelandii* genome and can be correlated with the numbering shown on the physical map in Fig. 1. All sequences in this figure read left to right relative to the map shown in Fig. 1.

and *A. vinelandii nifM* promoters as well, which are located about 60 base pairs upstream from their respective *nifM* coding sequences, within the *nifZ* coding region (Fig. 1). The proposed ORF5 and ORF10 promoters, which are oriented in opposite and divergent directions, have tandem and potentially overlapping upstream activator sequences (Fig. 3). The proposed ORF8 promoter is close to the ribosome-binding site for ORF8, and consequently, it seems unlikely that this promoter can direct expression of ORF8.

Eight of the eleven proposed *nif* promoter sequences have A+T-rich regions between them and their respective upstream activator sequences (Fig. 4). The overall G+C content of the entire *A. vinelandii* region sequenced is 64%, while the identified A+T-rich regions range from about 11 to 36% G+C (Fig. 4). This feature is particularly striking because such A+T-rich regions are not found elsewhere within the proposed *nif* cluster, although there are two other A+T-rich regions located beyond the *nifF* gene. Similar A+T-rich sequences were also recognized in regions located immediately preceding the *K. pneumoniae nifH*, *nifE*, *nifU*, and *nifB* promoters (5). Beynon et al. (5) suggested that such A+T-rich regions which precede certain of the *nif*-specific promoters are more likely to provide generalized RNA polymerase recognition sites rather than serving as *nif*-specific regulatory signals. However, the possibility remains that the extent and position of the A+T-rich regions which precede certain *nif*-specific promoters could affect the frequency of transcription initiation.

Codon usage. A summary of the codon usage for all of the proposed genes that are labeled in Fig. 1 is shown in Table 3. This codon usage shows a strong bias for G and C residues. This observation is consistent with the high overall G+C content of the *A. vinelandii* genome. Comparison of the codon usage for the individual genes showed that the *nifH* gene exhibits the strongest bias and that *nifD* and *nifK* gene codon usages are also more strongly biased than those of the other genes. Several codons that are never used within the nitrogenase structural genes are used moderately within most of the other proposed genes. For example, the leucine codon UUG is never used within the *nifHDK* genes but occurs a total of 45 times within the other proposed genes. Similarly, the arginine codon CGG is never used within the *nifHDK* genes, yet it occurs a total of 58 times within the other proposed genes. This distribution of codon usage could represent a mechanism for dampening the translation of nitrogenase-processing gene products while ensuring a high level of translation of the structural components. This hypothesis presumes that the biased codon usage within the *nif* structural gene cluster reflects the relative abundance of the corresponding iso-accepting tRNA species within *A. vinelandii*. However, no such relationship has yet been established.

Identification of essential *nif*-specific genes. In Materials and Methods, procedures for introducing specific deletion, insertion, and insertion-plus-deletion mutations into the *A.*

TABLE 3. Summary of codon usage for all of the proposed genes designated in Fig. 1

Codon	Amino acid	No. of times used	Codon	Amino acid	No. of times used	Codon	Amino acid	No. of times used	Codon	Amino acid	No. of times used
UUU	F	10	UCU	S	7	UAU	Y	50	UGU	C	16
UUC	F	198	UCC	S	106	UAC	Y	111	UGC	C	96
UUA	L	0	UCA	S	0	UAA		4	UGA		17
UUG	L	45	UCG	S	74	UAG		2	UGG	W	59
CUU	L	13	CCU	P	10	CAU	H	37	CGU	R	94
CUC	L	140	CCC	P	107	CAC	H	105	CGC	R	259
CUA	L	8	CCA	P	7	CAA	Q	35	CGA	R	6
CUG	L	463	CCG	P	171	CAG	Q	213	CGG	R	58
AUU	I	29	ACU	T	18	AAU	N	27	AGU	S	21
AUC	I	274	ACC	T	244	AAC	N	146	AGC	S	123
AUA	I	4	ACA	T	4	AAA	K	34	AGA	R	2
AUG	M	164	ACG	T	25	AAG	K	227	AGG	R	14
GUU	V	39	GCU	A	51	GAU	D	112	GGU	G	110
GUC	V	189	GCC	A	428	GAC	D	287	GGC	G	330
GUA	V	27	GCA	A	28	GAA	E	193	GGA	G	24
GUG	V	224	GCG	A	146	GAG	E	277	GGG	G	39

vinelandii genome were described. These procedures were used to place such mutations throughout the major *nif* cluster (Table 4). All of the mutant strains described here were examined for the ability to grow on nitrogen-free minimal medium (46) containing 0.01 mM MoO_4^{2-}. It has been demonstrated that at this concentration of molybdenum the molybdenum-containing nitrogenase system is expressed while the alternative nitrogen fixation systems are repressed (6, 27). Therefore, under the physiological conditions used in this study, only the effect of each of the individual mutations on the molybdenum-containing nitrogenase system is examined.

Mutant strains that were deleted from within the *nifH* coding sequence to beyond the region sequenced in this study (DJ40 and DJ41) lost their diazotrophic growth capabilities. Although these two strains also had a reduced growth rate on Burk medium supplemented with a fixed nitrogen source, they remained viable. This demonstrated that no genes whose products have an essential cellular function are interspersed among the *nif*-specific genes. All mutant strains with deletion or insertion mutations within the structural genes (*nifHDK*), either individually (DJ13, DJ54, DJ77, and DJ100) or in combination (DJ11, DJ33, DJ34, DJ40, DJ41, and DJ46), lost their diazotrophic growth capabilities (Table 4). Strains with specific deletions within *nifE* (DJ35), *nifN* (DJ72), *nifU* (DJ99), and *nifM* (DJ136) are also phenotypically Nif^-. To rule out the possibility that the Nif^- character of strains deleted for *nifN* or *nifM* was the consequence of polarity caused by the resultant frame shift within these mutant strains (Table 4), mutations were placed in regions immediately following the *nifN* or *nifM* gene. These strains (DJ44, DJ57, and DJ135) all remained Nif^+, and consequently the *nifN* and *nifM* gene products are essential for diazotrophic growth.

On the basis of complementation experiments, as well as Mu reversion studies, Roberts and Brill (38) suggested that *nifU* might not be essential for nitrogen fixation in *K. pneumoniae*. In contrast, Merrick et al. (31) reported evidence (based on complementation studies using a single mutant strain, UN2461 *nifU*) that *nifU* does represent a discrete complementation group within *K. pneumoniae*. The deletion that we placed within the *A. vinelandii nifU* gene did not result in a coding frame shift, and consequently the Nif^- phenotype of this strain is unlikely to be the consequence of polarity on the expression of distal genes. Since the *nifU* gene products from both *A. vinelandii* and *K. pneumoniae* share significant sequence identity (4), it is likely that the *nifU* gene product performs the same function in both organisms. If the *nifU* gene product is not essential for diazotrophic growth in *K. pneumoniae*, this difference might be a reflection of the very different physiologies of the respective organisms, i.e., a facultative anaerobe versus an obligate aerobe. Roberts and Brill (38) also suggested the possibility that the *nifU* and *nifS* gene products could form a complex because the apparent *nifU* gene product was less stable in the absence of the *nifS* gene product. Even if the *nifU* and *nifS* gene products do form a complex in *A. vinelandii*, the *nifU* gene product must have some function in the absence of the *nifS* product because a *nifS* deletion strain (DJ116) is still capable of weak diazotrophic growth (see later discussion).

Mutations that do not affect diazotrophic growth. Deletion of the entire *nifT* gene (DJ86) did not affect diazotrophic growth, indicating that its product does not have a function essential for nitrogen fixation under the physiological conditions used in this study. It is also unlikely that *nifY* is an essential *nif*-specific gene, since the deletion within mutant strain DJ86 removed the ribosome-binding site for *nifY*. A strain that has an insertion mutation within *nifT* (DJ87) also exhibited normal diazotrophic growth. Thus, *nifY*, ORF1, and ORF2 are either not essential *nif*-specific genes or the kanamycin resistance cartridge that is inserted into the *nifT* gene in strain DJ87 is not sufficiently polar upon expression of the distal genes such that there is an observable phenotype. There is no evidence, based on the DNA sequence, that transcription initiation occurs from the region within *nifK* through ORF2. Deletion-and-insertion mutagenesis of the region between ORF4 and ORF5 (strains DJ133 and DJ134) had no effect on diazotrophic growth. This indicates that the region between ORF4 and ORF11 probably does not encode *nif*-related gene products (see previous discussion). Furthermore, the potential ORF11 and ORF10 gene products (Table 4, mutant strains DJ133 and DJ134) are apparently not required for normal diazotrophic growth. A mutant

TABLE 4. Mutant strains used in this study

A. vinelandii strain	Mutation[a]	Location[b]	*nif* gene(s) or ORF(s) disrupted[c]	Parental plasmid[d]	Plasmid used[e]	Nif phenotype[f]	Reference
DJ78	Δ::	----720	12	pDB44	pDB109	+	
DJ40	Δ	716------	*H-F*	pAV80	pDB41	−	
DJ41	Δ	1447-----	*H-F*	pAV80	pDB40	−	
DJ34	Δ	210–8929	*HDKTY, 1, 2, E*	pDB1	pDB34	−	
DJ11	Δfs	1447–6164	*HDKTY*	pDB3	pDB11	−	
DJ46	Δ	2034–2758	*HD*	pDB6	pDB70	−	
DJ54	Δfs	1450–1988	*H*	pDB6	pDB81	−	40
DJ77	Δ	1712–1840	*H*	pMJH5	pDB115	−	
DJ33	Δ	2546–4739	*DK*	pDB32	pDB33	−	39
DJ100	Δ	2546–3367	*D*	pDB14	pDB31	−	39
DJ13[g]	Δ	4223–4697	*K*			−	7[g]
DJ86	Δ	5451–5733	*T*	pDB6	pDB119	+	
DJ87	::	5734	*T*	pDB6	pDB116	+	
DJ42	Δ	6165–13130	*Y, 1, 2, ENX, 3, 4*	pDB37	pDB67	−	11
DJ35	Δ	8162–8929	*E*	pDB7	pDB35	−	11
DJ39	Δ	9744–11845	*NX, 3, 4*	pDB58	pDB62	−	11
DJ72	Δfs	9744–10237	*N*	pDB92	pDB95	−	
DJ44	Δ	10930–11565	*X, 3*	pDB58	pDB61	+	
DJ57	Δ::	10930–11565	*X, 3*	pDB54	pDB82	+	
DJ133	Δ::	11846–15311	*11, 10*	pDB137	pDB140	+	
DJ134	Δ::	14156–15387	*11, 10*	pDB137	pDB141	+	
DJ43	Δ	17101–18875	*5, 6, US*	pDB56	pDB63	−	
DJ37	Δ	18777–23791	*USV, 7, 8, WZM*	pDB42	pDB57	−	
DJ99	Δ::	18243–18590	*U*	pDB111	pDB127	−	
DJ105	Δ	18243–18590	*U*	pDB111	pDB114	−	
DJ74	Δfs	20025–20127	*SV*	pDB101	pDB108	−	
DJ47	Δfs	19303–19391	*S*	pDB42	pDB79	Slow	
DJ48	Δ::	19303–19391	*S*	pDB42	pDB71	−	
DJ116	Δ	18876–20024	*S*	pDB133	pDB136	Slow	
DJ38	Δ	20735–21723	*V, 7*	pDB42	pDB52	Slow	
DJ71	Δ	20391–20984	*V*	pDB106	pDB107	Slow	
DJ142	Δ::	21489–21706	*7*	pDB172	pDB175	Reduced	
DJ155	Δ	21489–21706	*7*	pDB172	pDB174	+	
DJ137	::	22318	*8*	pDB162	pDB165	Slow	
DJ175	Δ	22111–22317	*8*	pDB190	pDB191	+	
DJ132	Δfs	22972–23148	*WZ*	pDB124	pDB148	Slow	
DJ144	::	22718	*W*	pDB177	pDB180	Slow	
DJ67	Δ	23368–23791	*ZM*	pDB97	pDB98	−	
DJ149	::>	23149	*Z*	pDB124	pDB132	Slow	
DJ150	::<	23149	*Z*	pDB124	pDB134	−	
DJ98	::	23683	*M*	pDB97	pDB134	−	
DJ136	Δfs	23785–24091	*M*	pDB152	pDB156	−	
DJ135	::	24816	*9*	pDB151	pDB153	+	
DJ58	::>	26295	*F*	pDB93	pDB99	+	3
DJ60	Δ::>	26017–26294	*F*	pDB94	pDB100	+	3

[a] Δ indicates a deletion, fs indicates a frame shift, :: indicates insertion of a kanamycin resistance-encoding gene cartridge, and < or > indicates the direction of transcription of the kanamycin resistance gene cartridge relative to transcription of the *nifHDK* genes.

[b] The numbers refer to the left and right termini of deletion mutations located within or bordering the sequenced region. A broken line to the left or right of a number indicates that the location of the left or right terminus of the deletion is not known. A single number (with no dashed lines) indicates the position of an insertion mutation that does not result in the deletion of any genomic DNA. The numbers can be correlated with the numbering shown in Fig. 1.

[c] Proposed gene(s) physically disrupted by the mutation. Potential polar effects are not considered as a physical disruption.

[d] Parental plasmid from which the plasmid used to construct the mutant strain was derived.

[e] Plasmid actually used to construct the mutant strain.

[f] A minus indicates that the mutant strain was incapable of diazotrophic growth, slow indicates that the mutant strain was capable of only very slow diazotrophic growth, reduced indicates that the mutant strain was capable of diazotrophic growth but at a rate noticeably lower than that of the wild-type strain, and a plus indicates that the mutant strain was capable of normal diazotrophic growth. All growth experiments were performed on Burk nitrogen-free agar plates. Slight reductions in diazotrophic growth rates would not be recognized by this procedure.

[g] This strain was constructed by crossing *A. vinelandii* CA13 (7) with wild-type *A. vinelandii*.

strain with an insertion mutation within ORF9 (DJ135) also remained Nif$^+$.

Interpretation of results from the mutagenesis of the ORF7-ORF8 region is complicated. A specific deletion within ORF7 (DJ155) or ORF8 (DJ175) had no affect on diazotrophic growth, indicating that these gene products are dispensible. In contrast, a strain with a deletion-and-insertion mutation within ORF7 (DJ142) exhibited a reduced rate of diazotrophic growth, while a strain with an insertion mutation within ORF8 (DJ137) showed an even lower rate of diazotrophic growth. This result indicated that transcription must occur through both ORF7 and ORF8 and that the insertion mutations within the respective genes caused at least a partial polar effect. Whether the difference between the diazotrophic growth rates of DJ142 (Δ::ORF7) and DJ137 (::ORF8) is the consequence of a difference in the orientations of the kanamycin resistance cartridges placed within the respective genes or results from the potential ability for

transcription initiation to take place at the proposed ORF8 promoter is not known. The orientations of the kanamycin resistance cartridges within the respective genes of mutant strains DJ137 and DJ142 are also not known.

In *K. pneumoniae*, a pyruvate-flavodoxin oxidoreductase (*nifJ* gene product) couples the oxidation of pyruvate to the reduction of flavodoxin (*nifF* gene product) yielding acetyl coenzyme A and CO_2. Reduced flavodoxin subsequently acts as a low-potential electron donor to oxidized Fe protein (34, 44). The pathway for electron transport to nitrogenase in *A. vinelandii* is much less clear (see discussion in reference 3). A strain with a deletion-plus-insertion mutation within ORF12 (DJ78) was still capable of normal diazotrophic growth. This does not, however, rule out the possibility that ORF12 encodes a *nifJ*-like function (see previous discussion), because *A. vinelandii* strains that have similar deletion-plus-insertion mutations within *nifF* (DJ58 and DJ60) also retained normal diazotrophic growth capability (3).

Mutations that markedly reduce diazotrophic growth. *A. vinelandii* mutant strains with an in-frame deletion within *nifS* (DJ116) or *nifV* (DJ71) were capable of very slow diazotrophic growth on Burk nitrogen-free agar plates. Similarly, *nifS*-specific mutants from *K. pneumoniae* (38) and *B. japonicum* (20) also have been shown to exhibit a leaky Nif$^-$ phenotype. Nonpolar *nifV* point mutants from *K. pneumoniae* are also leaky (38). Thus, the *nifS* and *nifV* gene products are not essential for diazotrophic growth, yet they are required for the formation of a fully active nitrogen-fixing capability. Hoover et al. (25) proposed that *nifV* encodes a homocitrate synthase. Homocitrate is believed to be required for proper assembly of the FeMo cofactor, and it is possible that homocitrate (or a molecule derived from homocitrate) is an integral part of FeMo cofactor entity (25).

To test whether the effects of deleting the *nifS* and *nifV* genes were cumulative, we constructed a strain deleted for both genes. This strain (DJ74) exhibited a strictly Nif$^-$ phenotype, indicating that deletion of both *nifS* and *nifV* does cause a cumulative effect on nitrogen-fixing capability within the mutant strain. This indicates that the *nifS* and *nifV* gene products are unlikely to catalyze sequential steps in the same pathway. A deletion-and-frameshift mutation within the central portion of the *nifS* gene is apparently not strongly polar on *nifV* expression, because this mutant strain (DJ47) was still capable of slow diazotrophic growth. Thus, the translation of *nifS* and *nifV* does not appear to be strongly coupled. In contrast, a deletion-plus-insertion mutation located within *nifS* is polar upon *nifV* expression, since the strain with this mutation (DJ48) exhibited a strict Nif$^-$ phenotype. This result supports the notion that transcription initiation does not occur in the region immediately preceding *nifV*.

Interpretation of the phenotypes of strains altered within the *nifWZM* region is also difficult. A strain with an insertion mutation within *nifW* (DJ144) was capable of only very slow diazotrophic growth. This could be due to specific inactivation of *nifW* or it could be the result of polarity on *nifZM* gene expression. If the phenotype is the consequence of polarity, then the polar effect must be incomplete, since an intact *nifM* gene is absolutely required for diazotrophic growth (Table 4, DJ136 and DJ98). A strain with a deletion-and-frameshift mutation (DJ132) which spans the *nifW* and *nifZ* coding sequences also exhibited only slow diazotrophic growth. This effect could be due to inactivation of *nifW* or *nifZ* or both. Another possible explanation for this result is that the frameshift mutation within DJ132 causes an incomplete polar effect on *nifM* gene expression. The *nifZ* and *nifM* coding sequences overlap by eight base pairs, and thus, translational coupling could be required for efficient expression of *nifM*. The phenotypes of insertion mutants altered within *nifZ* depended on the orientation of the kanamycin resistance cartridge used to interrupt the *nifZ* gene. A mutant strain (DJ149) with the kanamycin resistance cartridge inserted within *nifZ* such that the natural kanamycin resistance gene promoter is pointed in the same direction as *nifM* was capable of slow diazotrophic growth. In contrast, a mutant strain (DJ150) with the same kanamycin resistance cartridge pointed in the reverse direction was strictly Nif$^-$. This indicates that the degree of polarity caused by insertion of the kanamycin cartridge into the *A. vinelandii* genome depends on its orientation. These data do not clearly indicate whether or not *nifZ* gene expression is required for normal diazotrophic growth.

***nifX* and *nifY* gene products have sequence identity.** The *nifD* and *nifK* gene products from *A. vinelandii* were shown to have considerable primary sequence identity when compared with the respective *nifE* and *nifN* gene products (11, 16). On the basis of this observation, we suggested that the *nifHDK* cluster and the *nifEN* cluster bear an evolutionary relationship to one another and that one cluster might have originated by tandem duplication of the other (11). In the present study, potential gene products that are apparently cotranscribed with the *nifHDK* cluster (*nifT*, *nifY*, ORF1, and ORF2; Fig. 1) were respectively compared with the potential gene products apparently cotranscribed with the *nifEN* cluster (*nifX*, ORF3, and ORF4; Fig. 1). Some significant sequence identity was observed when the *nifY* gene product was compared with the *nifX* gene product (Fig. 5A), but no other significant sequence identities were recognized among the other gene products. The sequence identity found between the *A. vinelandii nifY* and *nifX* genes is located toward the C-terminal end of the respective gene products (Fig. 5A). We also compared the *K. pneumoniae nifY* and *nifX* gene products, and these gene products also share significant sequence identity (Fig. 5B). However, the regions of sequence identity found between the *nifY* and *nifX* gene products from *K. pneumoniae* are different from those observed for the *A. vinelandii* gene products. Although it seems that *nifY* and *nifX* could have common evolutionary origins, it appears that the sequences conserved between the *nifY* and *nifX* gene products do not necessarily indicate conservation in their present functions. No function has been established for the *nifY* or *nifX* gene product, and both genes are apparently dispensible in *A. vinelandii*.

The nucleotide sequence of the 20 proposed *nif*-specific genes from *K. pneumoniae* is now known, and homologs to 18 of these genes from *A. vinelandii* have been isolated, sequenced, and mutagenized. The possibility of the presence of a *nifJ*-like gene in *A. vinelandii* has not been eliminated, as previously discussed. Similarly, the possibility of the presence of a *nifL*-like gene has not been eliminated because of the conserved identities observed in a comparison between the *K. pneumoniae nifL* gene product and an ORF identified in *A. vinelandii* which precedes *nifA* (2). Most of the *nif* genes from *A. vinelandii* are grouped within the same region of the genome and have the same sequential arrangement as the corresponding *K. pneumoniae nif*-specific genes but a different spatial arrangement. The placement of specific deletion, insertion, and insertion-plus-deletion mutations within the major *nif* cluster of *A. vinelandii* permitted the development of a genetic map. The genes from *A. vinelandii* identified as absolutely required for diazotrophic growth include *nifH*, *nifD*, *nifK*, *nifE*, *nifN*, *nifU*, and *nifM*. The

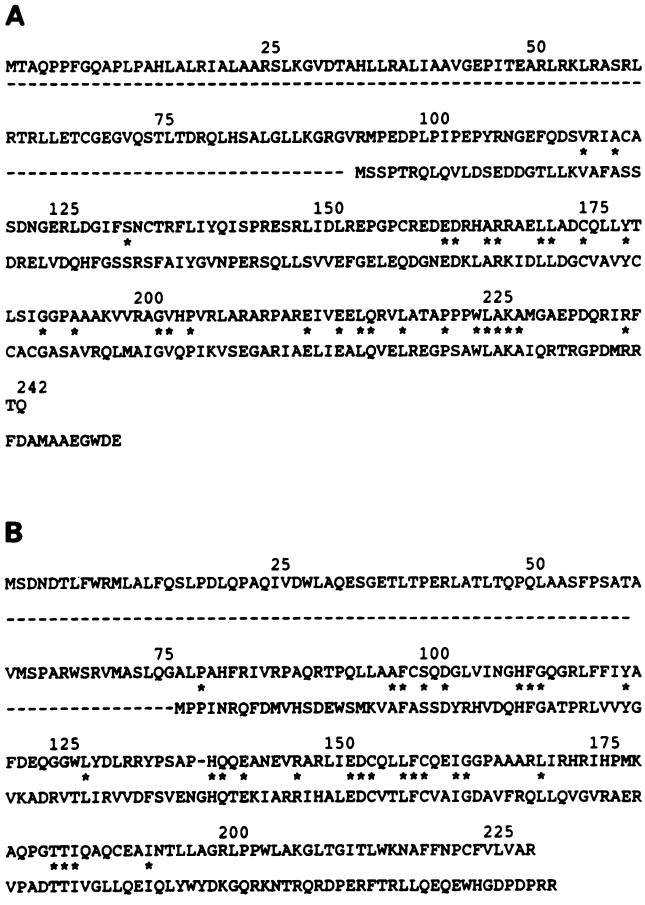

FIG. 5. Comparison of the *A. vinelandii nifY* and *nifX* gene products (A) and the *K. pneumoniae nifY* and *nifX* gene products (B). For each comparison, the upper sequence represents the *nifY* gene product and the lower sequence represents the *nifX* gene product. Perfect identities are indicated by asterisks. Dashed lines preceding the respective *nifX* gene products are added so that the respective *nifY* and *nifX* amino-coding regions can be readily distinguished. The entire sequence for each of the gene products is shown.

nifS and *nifV* genes were not absolutely required for diazotrophic growth, but deletion of these genes resulted in severe reduction of the diazotrophic growth capabilities of the resultant mutant strains. ORF12, *nifT*, *nifY*, ORF1, ORF2, *nifX*, ORF3, ORF4, ORF11, ORF10, ORF7, ORF8, ORF9, and *nifF* all appeared not to be required for normal diazotrophic growth in *A. vinelandii* under the physiological conditions used in this study. However, the genetic map of the major *nif* cluster is not complete, because either a requirement for ORF5, ORF6, *nifW*, or *nifZ* gene expression for diazotrophic growth was not tested or the results were inconclusive. Moreover, it is important to establish which of the potential promoters identified in this study are actually subject to control by the global *nif*-regulatory elements.

Mutant strains of *A. vinelandii* described here that have defined deletions or insertion mutations or both within individual genes should serve several useful purposes. The potential functions of the individual gene products in relation to the maturation of the nitrogenase structural components can now be analyzed by examining the catalytic consequences of each mutation. Furthermore, extracts from individual mutants can be mixed in attempts to reconstitute full nitrogenase activity as an approach to defining in vitro assays for the individual products. Such an approach was required to determine that *nifV* is necessary for the synthesis of homocitrate (25). The specific mutagenesis of multiple sites can also be accomplished by using the mutagenesis procedures described here sequentially. For example, a mutant strain deleted for both *nifN* and *nifB* was recently constructed (D. Dean, unpublished data). Such mutants will be useful for determining the sequence of events during nitrogenase component maturation. The collection of mutant strains, plasmid constructs, and DNA sequence information described here, as well as the site-directed mutagenesis procedure described previously (10), should also be valuable for identification and modification of the catalytic sites of the individual gene products encoded within the major *nif* cluster. These mutant strains can be used to determine which of the potential gene products encoded within the major *nif* cluster are required for maturation of either or both of the alternative nitrogen-fixing systems present in *A. vinelandii*.

ACKNOWLEDGMENTS

We thank the many people who contributed to this effort by supplying unpublished sequence information, technical information, and suggestions. We are especially grateful to Frank Cannon, Maura Cannon, Christina Kennedy, Mike Merrick, Bill Page, Rob Robson, and Greg Upchurch.

This work was supported by a grant from the U.S. Department of Agriculture (87-CRCR-1-2459), and M. R. Jacobson was supported by a postdoctoral fellowship (1 F 32 AIO7873-01) from the National Institutes of Health.

LITERATURE CITED

1. **Baughman, G., and M. Nomura.** 1983. Localization of the target site for translational regulation of the L11 operon and direct evidence for translational coupling in *Escherichia coli*. Cell **34**:979–988.
2. **Bennett, L. T., F. Cannon, and D. R. Dean.** 1988. Nucleotide sequence and mutagenesis of the *nifA* gene from *Azotobacter vinelandii*. Mol. Microbiol. **2**:315–321.
3. **Bennett, L. T., M. R. Jacobson, and D. R. Dean.** 1988. Isolation, sequencing, and mutagenesis of the *nifF* gene encoding flavodoxin from *Azotobacter vinelandii*. J. Biol. Chem. **263**:1364–1369.
4. **Beynon, J., A. Ally, M. Cannon, F. Cannon, M. Jacobson, V. Cash, and D. Dean.** 1987. Comparative organization of nitrogen fixation-specific genes from *Azotobacter vinelandii* and *Klebsiella pneumoniae*: DNA sequence of the *nifUSV* genes. J. Bacteriol. **169**:4024–4029.
5. **Beynon, J., M. Cannon, V. Buchanan-Wollaston, and F. Cannon.** 1983. The *nif* promoters of *Klebsiella pneumoniae* have a characteristic primary structure. Cell **34**:665–671.
6. **Bishop, P. E., R. Premakumar, R. D. Joerger, M. R. Jacobson, D. A. Dalton, J. R. Chisnell, and E. D. Wolfinger.** 1988. Alternative nitrogen fixation systems in *Azotobacter vinelandii*, p. 71–79. *In* H Bothe, F. J. de Bruijn, and W. E. Newton (ed.), Nitrogen fixation: hundred years after. Gustav Fischer Verlag, Stuttgart, Federal Republic of Germany.
7. **Bishop, P. E., T. M. Rizzo, and K. F. Bott.** 1985. Molecular cloning of *nif* DNA from *Azotobacter vinelandii*. J. Bacteriol. **162**:21–28.
8. **Brigle, K. E., and D. R. Dean.** 1988. Revised nucleotide sequence of the *Azotobacter vinelandii nifE* gene. Nucleic Acids Res. **16**:5214.
9. **Brigle, K. E., W. E. Newton, and D. R. Dean.** 1985. Complete nucleotide sequence of the *Azotobacter vinelandii* nitrogenase structural gene cluster. Gene **37**:37–44.
10. **Brigle, K. E., R. A. Setterquist, D. R. Dean, J. S. Cantwell, M. C. Weiss, and W. E. Newton.** 1987. Site-directed mutagenesis of the nitrogenase MoFe protein of *Azotobacter vinelandii*.

Proc. Natl. Acad. Sci. USA 84:7066–7069.
11. **Brigle, K. E., M. C. Weiss, W. E. Newton, and D. R. Dean.** 1987. Products of the iron-molybdenum cofactor-specific biosynthetic genes, *nifE* and *nifN*, are structurally homologous to the products of the nitrogenase molybdenum-iron protein genes, *nifD* and *nifK*. J. Bacteriol. **169**:1547–1553.
12. **Buck, M., S. Miller, M. Drummond, and R. Dixon.** 1986. Upstream activator sequences are present in the promoters of nitrogen fixation genes. Nature (London) **320**:370–378.
13. **Buikema, W. J., J. A. Klingensmith, S. L. Gibbons, and F. M. Ausubel.** 1987. Conservation of structure and location of *Rhizobium meliloti* and *Klebsiella pneumoniae nifB* genes. J. Bacteriol. **169**:1120–1126.
14. **Buikema, W. J., W. W. Szeto, P. V. Lemley, W. H. Orme-Johnson, and F. M. Ausubel.** 1985. Nitrogen fixation specific regulatory genes of *Klebsiella pneumoniae* and *Rhizobium meliloti* share homology with general nitrogen regulatory gene *ntrC* of *K. pneumoniae*. Nucleic Acids Res. **13**:4539–4555.
15. **Chisnell, J. R., R. Premakumar, and P. E. Bishop.** 1988. Purification of a second alternative nitrogenase from a *nifHDK* deletion strain of *Azotobacter vinelandii*. J. Bacteriol. **170**: 27–33.
16. **Dean, D. R., and K. E. Brigle.** 1985. *Azotobacter vinelandii nifD*- and *nifE*-encoded polypeptides share structural homology. Proc. Natl. Acad. Sci. USA **56**:5720–5723.
17. **Drummond, M., P. Whitty, and J. Wooton.** 1986. Sequence and domain relationships of *ntrC* and *nifA* from *Klebsiella pneumoniae*: homologies to other regulatory proteins. EMBO J. **5**: 441–447.
18. **Drummond, M. H.** 1985. The base sequence of the *nifF* gene of *Klebsiella pneumoniae* and homology of the predicted amino acid sequence of its protein product to other flavodoxins. Biochem. J. **232**:891–896.
19. **Drummond, M. H., and J. C. Wooton.** 1987. Sequence of *nifL* from *Klebsiella pneumoniae*: mode of action and relationship to two families of regulatory proteins. Mol. Microbiol. **1**:37–44.
20. **Ebeling, S., M. Hahn, H.-M. Fischer, and H. Hennecke.** 1987. Identification of *nifE*-, *nifN*- and *nifS*-like genes in *Bradyrhizobium japonicum*. Mol. Gen. Genet. **207**:503–508.
21. **Hales, B. J., E. E. Case, J. E. Morningstar, M. F. Dzeda, and L. A. Mauterer.** 1986. Isolation of a new vanadium-containing nitrogenase from *Azotobacter vinelandii*. Biochemistry **25**: 7251–7255.
22. **Hennecke, H., H.-M. Fischer, M. Gubler, B. Thony, D. Anthamatten, I. Kullik, S. Ebeling, S. Fritsche, and T. Zurcher.** 1988. Regulation of *nif* and *fix* genes in *Bradyrhizobium japonicum* occurs by a cascade of two consecutive gene activation steps of which the second one is oxygen sensitive, p. 339–344. *In* H. Bothe, F. J. de Bruijn, and W. E. Newton (ed.), Nitrogen fixation: hundred years after. Gustav Fischer Verlag, Stuttgart, Federal Republic of Germany.
23. **Hiratsuka, K., and K. L. Roy.** 1988. Sequence of a 1.4 kb *Eco*RI fragment from *Azotobacter vinelandii nif* DNA. Nucleic Acids Res. **16**:1207.
24. **Holland, D., A. Zilberstein, A. Zamir, and J. L. Sussman.** 1987. A quantitative approach to sequence comparisons of nitrogenase MoFe protein α- and β-subunits including the newly sequenced *nifK* gene from *Klebsiella pneumoniae*. Biochem. J. **247**:277–285.
25. **Hoover, T. R., A. D. Robertson, R. C. Cerny, R. N. Hayes, J. Imperial, V. K. Shah, and P. W. Ludden.** 1987. Identification of the V factor needed for synthesis of the iron-molybdenum cofactor of nitrogenase as homocitrate. Nature (London) **329**: 855–857.
26. **Ioannidis, I., and M. Buck.** 1987. Nucleotide sequence of the *Klebsiella pneumoniae nifD* gene and predicted amino acid sequence of the α-subunit of nitrogenase MoFe protein. Biochem. J. **247**:287–291.
27. **Jacobson, M. R., R. Premakumar, and P. E. Bishop.** 1986. Transcriptional regulation of nitrogen fixation by molybdenum in *Azotobacter vinelandii*. J. Bacteriol. **167**:480–486.
28. **Joerger, R. D., and P. E. Bishop.** 1988. Nucleotide sequence and genetic analysis of the *nifB-nifQ* region from *Azotobacter vinelandii*. J. Bacteriol. **170**:1475–1487.
29. **Kennedy, C., R. Gamal, R. Humphrey, J. Ramos, K. Brigle, and D. Dean.** 1986. The *nifH*, *nifM* and *nifN* genes of *Azotobacter vinelandii*: characterization by Tn*5* mutagenesis and isolation from pLARF1 gene banks. Mol. Gen. Genet. **205**:318–325.
30. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
31. **Merrick, M., M. Filser, R. Dixon, C. Elmerich, L. Sibold, and J. Houmard.** 1980. The use of translocatable genetic elements to construct a fine-structure map of the *Klebsiella pneumoniae* nitrogen fixation (*nif*) gene cluster. J. Gen. Microbiol. **117**: 509–520.
32. **Merrick, M. J.** 1988. Organization and regulation of nitrogen fixation genes in *Klebsiella* and *Azotobacter*, p. 293–302. *In* H. Bothe, F. J. de Bruijn, and W. E. Newton (ed.), Nitrogen fixation: hundred years after. Gustav Fischer Verlag, Stuttgart, Federal Republic of Germany.
33. **Messing, J.** 1983. New M13 vectors for cloning. Methods Enzymol. **101**:20–78.
34. **Nieva-Gomez, D., G. P. Roberts, S. Klevickis, and W. J. Brill.** 1980. Electron transport to nitrogenase in *Klebsiella pneumoniae*. Proc. Natl. Acad. Sci. USA **77**:2555–2558.
35. **Oppenheim, D. S., and C. Yanofsky.** 1980. Translational coupling during expression of the tryptophan operon in *Escherichia coli*. Genetics **95**:785–795.
36. **Page, W. J., and M. von Tigerstrom.** 1979. Optimal conditions for transformation of *Azotobacter vinelandii*. J. Bacteriol. **139**: 1058–1061.
37. **Paul, W., and M. Merrick.** 1987. The nucleotide sequence of the *nifM* gene of *Klebsiella pneumoniae* and identification of a new gene: *nifZ*. Eur. J. Biochem. **170**:259–265.
38. **Roberts, G. P., and W. J. Brill.** 1980. Gene-product relationships of the *nif* regulon of *Klebsiella pneumoniae*. J. Bacteriol. **144**:210–216.
39. **Robinson, A. C., B. K. Burgess, and D. R. Dean.** 1986. Activity, reconstitution, and accumulation of nitrogenase components in *Azotobacter vinelandii* mutant strains containing defined deletions within the nitrogenase structural gene cluster. J. Bacteriol. **166**:180–186.
40. **Robinson, A. C., D. R. Dean, and B. K. Burgess.** 1987. Iron-molybdenum cofactor biosynthesis in *Azotobacter vinelandii* requires the iron protein of nitrogenase. J. Biol. Chem. **262**: 14327–14332.
41. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
42. **Scott, K. F., B. G. Rolfe, and J. Shine.** 1981. Biological nitrogen fixation: primary structure of the *Klebsiella pneumoniae nifH* and *nifD* genes. J. Mol. Appl. Genet. **1**:71–81.
43. **Setterquist, R., K. E. Brigle, J. Beynon, M. Cannon, A. Ally, F. Cannon, and D. R. Dean.** 1988. Nucleotide sequence of the *nifE* and *nifN* genes from *Klebsiella pneumoniae*. Nucleic Acids Res. **11**:5215.
44. **Shah, V. K., G. Stacey, and W. J. Brill.** 1983. Electron transport to nitrogenase. J. Biol. Chem. **258**:12064–12068.
45. **Smith, B. E., M. Buck, R. R. Eady, D. J. Lowe, R. N. F. Thorneley, G. Ashby, J. Deistung, M. Eldridge, K. Fisher, C. Gormal, I. Ioannidis, H. Kent, J. Arber, A. Flood, C. D. Garner, S. Hasnain, and R. Miller.** 1988. Recent studies on the structure and function of molybdenum nitrogenase, p. 91–100. *In* H. Bothe, F. J. de Bruijn, and W. E. Newton (ed.), Nitrogen fixation: hundred years after. Gustav Fischer Verlag, Stuttgart, Federal Republic of Germany.
46. **Strandberg, G. W., and P. W. Wilson.** 1968. Formation of the nitrogen-fixing enzyme system in *Azotobacter vinelandii*. Can. J. Microbiol. **14**:25–31.
47. **Sundaresan, V., and F. M. Ausubel.** 1981. Nucleotide sequence of the gene coding for the nitrogenase iron protein from *Klebsiella pneumoniae*. J. Biol. Chem. **246**:2808–2812.
48. **Toukdarian, A., and C. Kennedy.** 1986. Regulation of nitrogen metabolism in *Azotobacter vinelandii*: isolation of *ntr* and *glnA* genes and construction of *ntr* mutants. EMBO J. **4**:399–407.