JOURNAL OF BACTERIOLOGY, Sept. 1987, p. 4024–4029
0021-9193/87/094024-06$02.00/0
Copyright © 1987, American Society for Microbiology
Vol. 169, No. 9

Comparative Organization of Nitrogen Fixation-Specific Genes from *Azotobacter vinelandii* and *Klebsiella pneumoniae*: DNA Sequence of the *nifUSV* Genes

JIM BEYNON,[1] ABDUL ALLY,[1] MAURA CANNON,[1] FRANK CANNON,[1] MARTY JACOBSON,[2] VALERIE CASH,[2] AND DENNIS DEAN[2]*

BioTechnica International, Inc., Cambridge, Massachusetts 02140,[1] and Department of Anaerobic Microbiology, Virginia Polytechnic Institute and State University, Blacksburg, Virginia 24060[2]

Received 4 March 1987/Accepted 15 June 1987

In the facultative anaerobe *Klebsiella pneumoniae* 17 nitrogen fixation-specific genes (*nif* genes) have been identified. Homologs to 12 of these genes have now been isolated from the aerobic diazotroph *Azotobacter vinelandii*. Comparative studies have indicated that these diverse microorganisms share striking similarities in the genetic organization of their *nif* genes and in the primary structure of their individual *nif* gene products. In this study the complete nucleotide sequences of the *nifUSV* gene clusters from both *K. pneumoniae* and *A. vinelandii* were determined. These genes are identically organized on their respective genomes, and the individual genes and their products exhibit a high degree of interspecies sequence homology.

Nitrogen fixation is catalyzed by the enzyme nitrogenase, a complex two-component metalloenzyme. In the facultative anaerobe *Klebsiella pneumoniae* there are at least 17 genes whose products are likely to be required for the synthesis and assembly of a fully active nitrogen-fixing system. These genes include *nifHDK* (nitrogenase structural components); *nifF* and *nifJ* (electron transport components); *nifQ*, *nifB*, *nifE*, *nifN*, and *nifV* (FeMo-cofactor biosynthetic components); *nifA* and *nifL* (positive and negative regulatory elements); *nifM* (Fe protein maturation component); and *nifY*, *nifX*, *nifU*, and *nifS* (functions not known). For a recent review, see reference 7. The *nif* genes are clustered on the *K. pneumoniae* chromosome and are organized into eight transcriptional units (1). A gene-product relationship has been established for many of the *K. pneumoniae nif*-specific components (9, 18), yet only those products encoded by the nitrogenase structural genes and those encoded by the electron transport-specific genes (16, 24) have been purified in their native form. Since the function of a number of the uncharacterized *nif* gene products is likely to involve a catalytic action upon a nitrogenase structural component, for example, metallocenter assembly, it is probable that some of these gene products are present in nitrogen-fixing cells in only low amounts. This possibility and the prospect that many of the ancillary *nif* gene products will share with the structural components an extreme sensitivity to oxygen potentially present a formidable challenge for *nif* gene product isolation and biochemical characterization.

One approach for initiating the characterization of *nif* gene products is the determination of their polypeptide sequences deduced from DNA sequence analysis. Such an approach can provide (i) insights into the structural features of the encoded polypeptides, (ii) DNA sequence information necessary for the construction of hybrid plasmids that can be used to enhance the production of the individual *nif* polypeptides, and (iii) the information and materials needed for the directed mutagenesis of the individual *nif* genes. In this regard, comparative studies of the primary sequences of individual *nif*-specific components from diverse diazotrophic species are of particular relevance because conserved sequences among homologous proteins are likely to indicate their important structural features.

In the case of the nitrogenase structural components, a high degree of interspecies sequence homology has permitted the isolation of nitrogenase structural genes from numerous diazotrophic species (20). Similarly, the demonstration of interspecies homologies among other individual *nif* genes should also permit the identification and isolation of such genes from a variety of diazotrophs. Summarized here are results obtained by using this approach for the isolation and mapping of *nif* genes from the aerobic diazotroph *Azotobacter vinelandii*. A comparison of the physical organization of *nif* genes established for *K. pneumoniae* with that currently developed for *A. vinelandii* is presented. In addition, we have determined and compared the nucleotide sequences of the *nifUSV* gene clusters from both *K. pneumoniae* and *A. vinelandii* and compared their respective encoded polypeptides.

MATERIALS AND METHODS

The results of physical and genetic mapping procedures used to define the *K. pneumoniae nifUSV* gene cluster are summarized in reference 1. The identification of the *A. vinelandii nifUSV* gene cluster was based on its homology to the *K. pneumoniae nifUSV* gene cluster (see Results). The DNA sequences of the *A. vinelandii* and *K. pneumoniae nifUSV* gene clusters were determined by the dideoxy chain termination procedure (21) by using hybrids of filamentous phage vectors described by Messing (15). For *A. vinelandii* sequencing experiments, the hybrid plasmids used as DNA sources are shown in Fig. 1A. For these experiments purified DNA fragments were digested with the individual restriction enzymes *Sau*3A, *Eco*RI, *Sma*I, *Xho*I, *Pst*I, *Sal*I, *Hin*cII, *Hin*PI, and *Msp*I or some combination of these restriction enzymes and ligated into the appropriately digested replicative form of the filamentous phage vector DNA. Approximately 200 base pairs were determined in each sequencing experiment. All sequences were determined in an overlapping fashion and in both directions. For *K. pneumoniae* sequencing experiments, the hybrid plasmid pWK25 (17)

* Corresponding author.

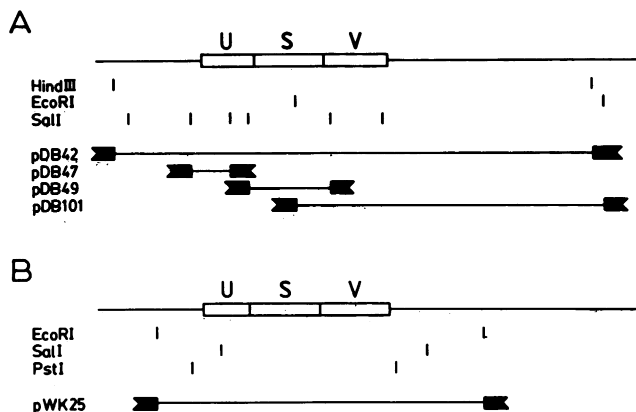

FIG. 1. Partial restriction map and hybrid plasmids used as a DNA source for sequencing experiments. (A) *A. vinelandii*; (B) *K. pneumoniae*. Hybrid plasmids containing *A. vinelandii* DNA were derived from pUC8 (15). Hybrid plasmid pWK25 was obtained from W. Klipp and A. Puhler (17).

served as the ultimate source of DNA (Fig. 1B). The individual *Sal*I and *Pst*I restriction enzyme fragments (Fig. 1B) were sonicated by the procedure of Messing (15), cloned into M13 sequencing vector DNA, and sequenced. Sufficient clones were sequenced to provide overlapping sequences from both strands.

RESULTS AND DISCUSSION

Comparative organization of *nif* genes from *K. pneumoniae* and *A. vinelandii*. The current status of *A. vinelandii nif* gene mapping experiments is shown in Fig. 2, and the results are compared with the physical map established for *K. pneumoniae* (1). The isolation and physical mapping of the *nifHDK* (2, 8, 19), *nifEN* (3, 5, 13), and *nifM* (13) genes from *A. vinelandii* have been previously described. The assignment of the *A. vinelandii nifF* gene was based on the location of a flavodoxin-specific gene sequence within the *nif* cluster and a high degree of sequence homology between this gene and the *K. pneumoniae nifF* gene (6; unpublished results). The *nifY* and *nifX* gene assignments were based on homologies between the respective genes from *K. pneumoniae* and *A. vinelandii* indicated by DNA sequence analysis (unpublished results). However, the assignment of *nifY* and *nifX* as *nif* genes must be considered tentative even for *K. pneumoniae* because neither a gene-product relationship nor a physiological function has been established for these genes and their products. The identification of the *A. vinelandii nifUSV* genes was determined by comparison of the complete nucleotide sequence of this region with that of the corresponding region of *K. pneumoniae* (this study).

Kp nifH	C A C G G	C T G G	T A T G T T C C	C T G C A	C T T C T
Av nifH	G G G T G	C T G G	C A C A G A C G	C T G C A	T T A C C
Kp nifE	C G C T T	C T G G	A G C G C G A A	T T G C A	T C T T C
Av nifE	G C C T T	C T G G	T A C A G G C A	T T G C A	A T G A T
Kp nifU	A T T C T	C T G G	T A T C G C A A	T T G C T	A G T T C
Av nifU	T G G T T	G T G G	C A A G T C T T	T T G C T	T G T T G
Kp nifF	G C A A C	C T G G	C A C A G C C T	T C G C A	A T A C C
Av nifF	G G G G A	G T G G	T C T G C T T C	T T G C T	G T T A C

FIG. 3. Comparison of identified *nif* promoters from *K. pneumoniae* (Kp) and regions containing proposed *nif* promoters from *A. vinelandii* (Av). Consensus *nif* promoter sequences are boxed.

Mutant *A. vinelandii* strains that have a NifB-like phenotype (23) and a NifA-like phenotype (22) have been described; however, no linkage has yet been established between these genes and the major *nif* cluster in *A. vinelandii*. Two factor crosses and marker rescue experiments performed in our laboratory indicated that the *nifB*-like and *nifA*-like alleles were not located within 5 kilobases of the *nifF* gene in *A. vinelandii*. At present there is no information available concerning the presence or absence of a *nifJ*-like gene in *A. vinelandii*.

The overall organization of *nif* genes in *K. pneumoniae* and *A. vinelandii* is remarkably similar. The only striking differences are the large gaps between *nif* gene clusters in *A. vinelandii* (Fig. 2). Whether there are other genes that are specific for nitrogen fixation in *A. vinelandii* encoded within these gaps is not yet known. The regions between the identified *nif* gene clusters in *A. vinelandii* do not, however, include any genes that encode an essential cellular function, since mutant strains with large deletions that spanned the entire region between the *nifH* gene and the *nifF* gene were viable, provided a fixed nitrogen source was added to the culture medium.

Promoters. Promoter sequences have been identified in *K. pneumoniae* in regions preceding the *nifH*, *nifE*, *nifU*, *nifB*, *nifM*, *nifF*, and *nifL* genes (1). These *nif* promoters have an unusual and characteristic structure consisting of two regions of conserved sequence homology. DNA sequences preceding the *nifH*, *nifE*, *nifU*, and *nifF* genes in *A. vinelandii* are compared with the homologous regions in *K. pneumoniae* in Fig. 3. This comparison indicated that *A. vinelandii nif* promoters share a striking structural homology with their *K. pneumoniae* counterparts. In addition, a proposed activator sequence located upstream from *nif* promoters has been identified in the available *K. pneumoniae* and *A. vinelandii nif* promoter-proximal sequences (4). These features and the demonstration that the *K. pneumoniae nifA* gene product can stimulate *A. vinelandii nif* gene expression (14) suggest that the regulation of *nif* gene expression in both organisms is likely to be similar. However, there must also be some differences, because the *K. pneumoniae nifH* pro-

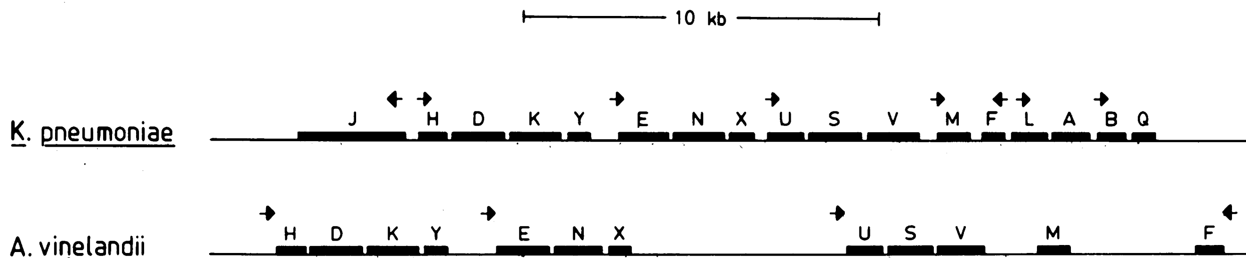

FIG. 2. Comparative organization of *nif* genes from *K. pneumoniae* and *A. vinelandii*. Arrows indicate the location of promoters and their direction of transcription. A promoter and direction of transcription have not been determined for *A. vinelandii nifM*. kb, Kilobases.

A

B

FIG. 4. Complete nucleotide sequences of the *A. vinelandii nifUSV* genes (A) and the *K. pneumoniae nifUSV* genes (B). For both *A. vinelandii* and *K. pneumoniae* the gene order is *nifU*, *nifS*, and *nifV*. The regions corresponding to the proposed *A. vinelandii nifU* promoter and the identified *K. pneumoniae nifU* promoter are boxed. A region of dyad symmetry located within the *A. vinelandii nifS-nifV* intercistronic region is underlined.

TABLE 1. Comparison of the *A. vinelandii* and *K. pneumoniae nifUSV* genes and their products[a]

Gene	Bacterium	Product mol wt	Product pI	% Homology at: DNA level	% Homology at: Protein level
nifU	*A. vinelandii*	33,274	4.93	64.5	54.5
nifU	*K. pneumoniae*	29,498	4.80		
nifS	*A. vinelandii*	43,587	5.79	65.0	59.0
nifS	*K. pneumoniae*	43,259	7.76		
nifV	*A. vinelandii*	41,658	5.71	56.5	44.4
nifV	*K. pneumoniae*	41,194	6.11		

[a] Calculation of the gene product molecular weights and isoelectric points and the percentage of DNA and protein sequence homologies was done by using the sequencing program from DNASTAR, Inc., Madison, Wis.

moter is not expressed when introduced into *A. vinelandii* (12) and because *A. vinelandii nif* structural gene mRNA does not accumulate under conditions of Mo deprivation (11), while *K. pneumoniae nif* structural gene mRNA accumulation is apparently insensitive to cellular Mo levels (10).

DNA sequences of the *nifUSV* gene clusters. The identification of the *A. vinelandii nifUSV* gene cluster was accomplished by performing random sequence analysis of DNA fragments isolated from the genomic region between the *nifN* gene and the *nifF* gene and comparing the sequences with preliminary DNA sequence data obtained for the *nifUSV* gene cluster from *K. pneumoniae*. A complete sequence determination and comparison of approximately 3 kilobases of DNA from the *nifUSV* regions of both organisms revealed three open reading frames which corresponded to the encoded *nifU*, *nifS*, and *nifV* gene products (Fig. 4A and B). A comparison of the region preceding the *nifU* gene from both organisms revealed consensus *nif* promoter sequences at approximately the same distance from the *nifU* translation initiation site (Fig. 4A and B). In *A. vinelandii* only a single nucleotide separated the termination signal for the *nifU* gene product and the initiation codon for the *nifS* gene product. There was also a relatively short space (30 base pairs) separating *nifU* and *nifS* in *K. pneumoniae*. Only 14 base pairs separated *nifS* and *nifV* in *K. pneumoniae*, while a much larger gap (72 base pairs) separated *nifS* and *nifV* in *A. vinelandii*. A region of dyad symmetry was located in the *A. vinelandii nifS-nifV* intercistronic space. Similar structures have been recognized in *nifH-nifD* and *nifD-nifK* intercistronic regions in *A. vinelandii* (2), but their physiological function is not yet known. No obvious *rho*-independent transcription termination signal structures were apparent in the regions immediately following the respective ends of the *nifV* genes, although there was a pyrimidine-rich region located just past the *K. pneumoniae nifV* gene. A strong transcription termination signal was not expected past *nifV* in *K. pneumoniae*, since transcription readthrough between *nifV* and *nifM* has been demonstrated (1).

***nifUSV* gene product comparisons.** The complete sequences of the *nifUSV* gene clusters from *K. pneumoniae* and *A. vinelandii* permitted the calculation of the pI and

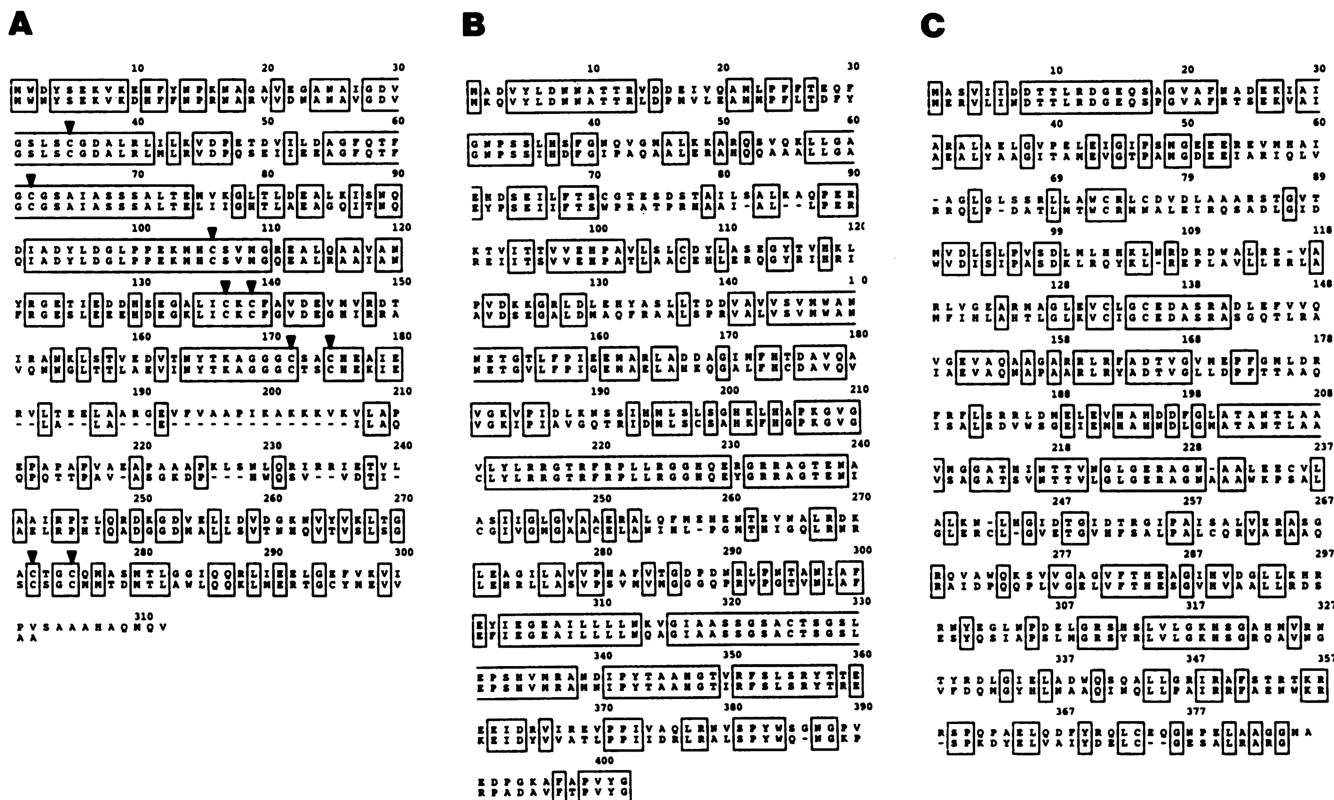

FIG. 5. Comparison of the *A. vinelandii* and *K. pneumoniae nifU* (A), *nifS* (B), and *nifV* (C) gene products. For each comparison the upper sequences represent the *A. vinelandii* gene product and the lower sequences represent the *K. pneumoniae* gene product. Perfect homologies are boxed. Alignment adjustments were made to provide the best fit and are indicated by hyphens. Numbers above the sequences correspond to the *A. vinelandii* sequences in Fig. 4A. The entire sequence of each polypeptide is shown. In panel A conserved cysteinyl residues are indicated by arrowheads.

molecular weight of the encoded proteins by using DNA sequence data (Table 1). The values for *K. pneumoniae* were in good agreement with estimations for the *nifUSV*-encoded polypeptides based on two-dimensional gel analyses (9, 18).

Primary sequence comparisons of nitrogenase components from a wide variety of diazotrophs have been used in attempts to identify structurally important regions of these polypeptides. As in the case of interspecies nitrogenase comparisons (20), the respective *nifU*, *nifS*, and *nifV* gene products from *A. vinelandii* and *K. pneumoniae* showed a high degree of homology at both the DNA and protein sequence levels (Fig. 4 and 5 and Table 1). Interspecies nitrogenase structural component homologies were greatest in the amino-terminal halves of the polypeptides and were largely centered around cysteinyl residues (see the Discussion in reference 2). A comparison of *nifU*-encoded polypeptides from *A. vinelandii* and *K. pneumoniae* also revealed this pattern. There were nine conserved *nifU* gene product cysteinyl residues, and seven of these were found in highly conserved regions. The nitrogenase structural components are known to be Fe- and S-containing proteins, and as such they are expected to share functional domains in regions surrounding cysteinyl ligands. The pattern of cysteinyl residue conservation in the *nifU* gene products raises the possibility that the *nifU*-encoded polypeptides could be metalloproteins. However, the function of *nifU* is not yet known, nor is there any biochemical information available regarding its potential metallocenter composition.

In the cases of interspecies *nifS* and *nifV* comparisons, there were regions of relatively high sequence homology throughout the polypeptides (Fig. 5B and C). The *nifS* gene products showed very high levels of sequence homology in their carboxy-terminal halves, while the *nifV* gene products exhibited much less compact homology than did the *nifU* and *nifS* gene products. The *nifUSV* gene products did not share any striking regions of sequence homology among themselves, nor did they share sequence homology with other available *nif* gene products.

Summary. Extensive genetic and biochemical studies on the facultative anaerobe *K. pneumoniae* have indicated that there are at least 17 genes whose products are involved in the process of nitrogen fixation. Homologs to 12 of these genes have now been identified in the aerobic diazotroph *A. vinelandii*. The high degree of conservation in *nif* gene organization and in the primary sequences of the individual *nif* genes between these diverse microorganisms indicates that the *nif* regulons from both species arose from a common ancestral origin.

Interspecies *nif* structural gene sequence homologies (20) have been used for the isolation of nitrogenase structural components from a wide variety of diazotrophs by heterologous DNA-DNA hybridizations. The high degree of sequence homology demonstrated here for the *nifUSV* gene clusters from *A. vinelandii* and *K. pneumoniae* indicates that a similar approach is likely to be successful for the isolation of *nif* genes from other organisms. Such strong conservation of sequence homology among the various *nif*-specific components also indicates to us that *nif* gene product structure-function studies on the more biochemically amenable free-living diazotrophs *A. vinelandii* and *K. pneumoniae* will have application towards efforts to understand and manipulate the nitrogen-fixing capacity of agronomically important symbiotic diazotrophs such as the rhizobia. Finally, the DNA sequence information provided here on the *nifUSV* genes should substantially aid efforts to overproduce and specifically alter the gene products for the purpose of elucidating their respective functions in relation to nitrogen fixation.

ACKNOWLEDGMENT

The portion of this work completed at Virginia Polytechnic Institute and State University was supported by U.S. Department of Agriculture competitive grant 84-CRCR-1-1767.

LITERATURE CITED

1. **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.
2. **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.
3. **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.
4. **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.
5. **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.
6. **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.
7. **Gussin, G. N., C. W. Ronson, and F. M. Ausubel.** 1986. Regulation of nitrogen fixation genes. Annu. Rev. Genet. **20**:567–592.
8. **Helfrich, R. J., J. M. Ligon, and R. G. Upchurch.** 1986. Identification and organization of *nif* genes of *Azotobacter vinelandii*, p. 507. *In* H. J. Evans, J. Bottomley, and W. E. Newton (ed.), Nitrogen fixation research progress. Martinus Nijhoff, Publishers BV, Dordrecht, The Netherlands.
9. **Houmard, J., D. Bogusz, R. Bigault, and C. Elmerich.** 1980. Characterization and kinetics of the biosynthesis of some nitrogen fixation (*nif*) gene products in *Klebsiella pneumoniae*. Biochimie **62**:267–275.
10. **Imperial, J., R. A. Ugalde, V. K. Shah, and W. J. Brill.** 1984. Role of the *nifQ* gene product in incorporation of molybdenum into nitrogenase in *Klebsiella pneumoniae*. J. Bacteriol. **158**:187–194.
11. **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.
12. **Kennedy, C., and M. H. Drummond.** 1985. The use of cloned *nif* regulatory elements from *Klebsiella pneumoniae* to examine *nif* regulation in *Azotobacter vinelandii*. J. Gen. Microbiol. **131**:1787–1795.
13. **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 Tn5 mutagenesis and isolation from pLAFR1 gene banks. Mol. Gen. Genet. **205**:318–325.
14. **Kennedy, C., and R. Robson.** 1983. Activation of *nif* gene expression in *Azotobacter* by the *nifA* gene product of *Klebsiella pneumoniae*. Nature (London) **301**:626–628.
15. **Messing, J.** 1983. New M13 vectors for cloning. Methods Enzymol. **101**:20–78.
16. **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.
17. **Puhler, A., and W. Klipp.** 1981. Fine-structure analysis of the gene region for N_2-fixation (*nif*) of *Klebsiella pneumoniae*, p. 276–286. *In* H. Boethe and A. Trebst (ed.), Biology of inorganic nitrogen and sulfur. Springer-Verlag, New York.
18. **Roberts, G. P., and W. J. Brill.** 1980. Gene-product relationships of the *nif* regulon of *Klebsiella pneumoniae*. J. Bacteriol.

144:210–216.

19. **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.
20. **Ruvkun, G. B., and F. M. Ausubel.** 1980. Interspecies homology of nitrogenase genes. Proc. Natl. Acad. Sci. USA **74:**191–195.
21. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **77:**5463–5467.
22. **Shah, V. K., L. C. Davis, J. K. Gordon, W. H. Orme-Johnson, and W. J. Brill.** 1973. Nitrogenase. III. Nitrogenaseless mutants of *Azotobacter vinelandii*: activities, cross-reactions and EPR spectra. Biochim. Biophys. Acta **292:**246–255.
23. **Shah, V. K., J. Imperial, R. A. Ugalde, P. W. Ludden, and W. J. Brill.** 1986. *In vitro* synthesis of the iron-molybdenum cofactor of nitrogenase. Proc. Natl. Acad. Sci. USA **83:**1636–1640.
24. **Shah, V. K., G. Stacey, and W. J. Brill.** 1983. Electron transport to nitrogenase. J. Biol. Chem. **258:**12064–12068.