JOURNAL OF BACTERIOLOGY, Apr. 1987, p. 1547–1553
0021-9193/87/041547-07$02.00/0

Vol. 169, No. 4

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*

KEVIN E. BRIGLE,[1] MARY C. WEISS,[2] WILLIAM E. NEWTON,[2] AND DENNIS R. DEAN[1]*

Department of Anaerobic Microbiology, Virginia Polytechnic Institute and State University, Blacksburg, Virginia 24060,[1] and Western Regional Research Center, Agricultural Research Service, U.S. Department of Agriculture, Albany, California 94710[2]

Received 8 September 1986/Accepted 9 January 1987

The genes from *Azotobacter vinelandii*, which are homologous to the iron-molybdenum cofactor biosynthetic genes, *nifE* and *nifN*, from *Klebsiella pneumoniae*, have been cloned and sequenced. These genes comprise a single transcription unit and are located immediately downstream from the nitrogenase structural gene cluster (*nifHDK*). DNA sequence analysis has revealed that the products of the *nifE* and *nifN* genes contain considerable homology when compared with the *nifD* (MoFe protein α subunit) and the *nifK* (MoFe protein β subunit) gene products, respectively. These striking sequence homologies indicate a structural and functional relationship between a proposed *nifEN* product complex and the nitrogenase MoFe protein as well as imply an ancestral relationship between these gene clusters. The isolation and characterization of strains which contain deletions within the *nifEN* gene cluster demonstrate a role for these products in iron-molybdenum cofactor biosynthesis in *A. vinelandii*.

The catalytic components of nitrogenase comprise two separable protein species: the Fe protein, a dimer of identical subunits (native M_r, ca. 60,000; product of the *nifH* gene), and the MoFe protein, a 2α-2β tetramer (native M_r, ca. 220,000; products of the *nifD* and *nifK* genes, respectively). During enzymatic turnover, these two component proteins associate and dissociate (9) with the reduced Fe protein serving as an ATP-binding, specific electron donor to the MoFe protein, which contains the sites for substrate binding and reduction (10) (for a recent review, see reference 14). Because of its direct role in dinitrogen binding and reduction, the structure and reactivity of the MoFe protein species are of particular interest. The native MoFe protein probably contains four [4Fe-4S] clusters (14), two iron-molybdenum cofactor centers (FeMo cofactors) (14), and probably two other Fe-containing centers, called S centers (14). The FeMo cofactor centers are relatively small inorganic clusters, composed of Mo, Fe, and S, and are likely to occupy a pivotal role in substrate binding and reduction because (i) *Klebsiella pneumoniae* strains lacking FeMo cofactor (i.e., *nifB*, *nifE*, and *nifN* mutants) lack catalytic activity (17, 20); (ii) cofactorless MoFe protein from such mutant strains can be reactivated by addition of purified FeMo cofactor (20); and (iii) mutation in another *K. pneumoniae* allele, *nifV*, gives rise to strains with a MoFe protein which contains an altered form of FeMo cofactor, and this MoFe protein species exhibits dramatic alterations in substrate recognition and reactivity (11).

Other than the identification of genes whose products are necessary for FeMo cofactor biosynthesis and maturation (*nifBNEV*), little is known about the biochemical events leading to FeMo cofactor biosynthesis, and information is not available regarding the spatial arrangement of the FeMo cofactor between or within the individual MoFe protein subunits. In the case of *Azotobacter vinelandii* (18) and *K. pneumoniae* (23), it has been shown that the MoFe protein subunits (*nifD* and *nifK* gene products) are not required for FeMo cofactor biosynthesis. These results imply that the FeMo cofactor is synthesized and inserted into an immature form of the MoFe protein rather than being synthesized stepwise into the MoFe protein. Thus, one or more of the *nifB*, *nifE*, and *nifN* gene products appeared likely to share with the MoFe protein the ability to bind a molybdenum species. If so, we reasoned that a comparison of the primary structure of FeMo cofactor biosynthetic gene products with those of the MoFe protein subunits could provide information on the function of the individual FeMo cofactor biosynthetic gene products as well as reveal potential domains for FeMo cofactor binding within the MoFe protein (7). Towards this end, we have previously isolated the *nifEN* gene cluster from *A. vinelandii*, sequenced the *nifE* gene, and compared the *nifE*-encoded gene product with the *nifD*-encoded gene product (MoFe protein α subunit) (7). That study showed that the *nifD* gene and the *nifE* gene products do indeed share striking primary-sequence homology. In the present study, we extended our analysis of the *nifD-nifE* gene product homologies, completed the nucleotide sequence of the *nifN* gene, compared the *nifK* gene and *nifN* gene products, and isolated and characterized strains which contain deletions within the *nifEN* gene cluster.

MATERIALS AND METHODS

Cell growth and nitrogenase derepression. The wild-type and mutant strains of *A. vinelandii* were cultured in a modified Burk medium (21). This medium was supplemented to a final concentration of 10 mM with filter-sterilized urea when a fixed source of nitrogen was included in the medium. For derepression of nitrogenase synthesis, all cultures were initially grown in Burk urea-supplemented medium to the mid-logarithmic phase and were harvested by centrifugation. Harvested cells were washed once with Burk medium, suspended in the original volume of Burk nitrogen-free medium, and incubated for an additional 3 h. All cultures were grown by shaking vigorously at 30°C. Culture volumes

* Corresponding author.

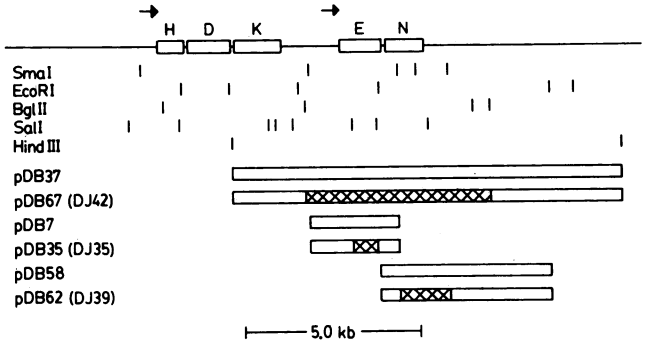

FIG. 1. Physical map of the *A. vinelandii* genome surrounding the *nif* structural gene cluster and plasmids used for deletion strain constructions. Plasmid pDB37 is a *nif*-pKT230 (1) hybrid. The horizontal arrows above the *nifHDK* and *nifEN* gene clusters indicate the direction of transcription. The indicated *nif-Hin*dIII fragment was cloned into the *Hin*dIII site of pKT230. pDB67 is a deletion derivative of pDB37. The internal *Bgl*II fragments contained within pDB37 are deleted in pDB67. pDB6 contains a *Sma*I fragment cloned into the *Sma*I site of pUC8 (12). For the construction of pDB35, an internal *Sal*I fragment was deleted from pDB7. pDB58 contains an *Eco*RI fragment cloned into the *Eco*RI site of pBR322. For the construction of pDB62, the internal *Sma*I fragments were deleted from pDB58. The open regions indicate cloned *A. vinelandii* genomic DNA, and the cross-hatched areas indicate deletions. The names of the deletion strains that were constructed with the deletion plasmid DNAs are indicated in parentheses adjacent to the individual plasmid designations. kb, Kilobases.

were 500 ml dispensed into 2-liter flasks. Derepressed cells were harvested by centrifugation, washed once with 0.05 M Tris (pH 8.0), repelleted by centrifugation at 10,000 × *g* for 10 min, and stored at −80°C until used for extract preparation.

***Escherichia coli* growth and plasmid preparation.** Growth of *E. coli* strains carrying hybrid *nif*-containing plasmids and the preparation, restriction enzyme digestion, and ligation of hybrid plasmid DNAs were all performed as described previously (4). Details of plasmid constructions are described in the legend to Fig. 1.

***A. vinelandii* transformations and construction of *nif*-deletion strains.** Transformations of *A. vinelandii* wild-type cells with purified *A. vinelandii* chromosomal DNA which carries an uncharacterized Rif^r marker and hybrid *nif*-containing plasmid DNA were performed in liquid cultures, as described by Page and von Tigerstrom (15). Specific *nif* deletions contained within various hybrid plasmids (Fig. 1) were transferred to the *A. vinelandii* chromosome by using congression (coincident transfer of unlinked genetic markers). In this case, rifampin resistance was the selected character. This procedure was followed exactly as described in detail previously (18).

DNA sequence analysis. DNA sequence analysis was performed by the dideoxy chain termination procedure (19) with hybrids of the filamentous M13 bacteriophage vectors described by Messing and Vieira (12). DNA fragments were prepared from the *nifEN* gene cluster by digestion with individual restriction enzymes, *Rsa*I, *Hae*III, *Sau*3A, *Eco*RI, *Sma*I, *Xho*I, *Pst*I, *Sal*I, and *Hin*cII, or some combination of these restriction enzymes and ligated into the appropriately digested replicative form of the filamentous M13 phage vector DNA. Approximately 200 base pairs was determined from each sequencing experiment. The sequences were determined in an overlapping fashion in both directions and in all cases by more than one sequencing run (Fig. 2).

Enzyme assays and reconstitutions. Crude extracts (approximately 25 mg of protein per ml) were prepared from derepressed cells in 0.05 M Tris (pH 8.0) by passage through a chilled, argon-flushed French pressure cell at 12,000 lb/in^2, followed by centrifugation at 14,000 × *g* for 20 min at 4°C. They were then degassed, and their Fe protein and MoFe protein activities were measured by the acetylene reduction assay in the presence of saturating amounts of the complementary protein. These assay procedures are essentially as described previously (18). Fe protein and MoFe protein were purified from *A. vinelandii* (strain OP), as described previously (5). The purified Fe protein and MoFe protein used in this study had specific activities of 850 and 2,020 nmol of C_2H_2 reduced per min per mg of protein, respectively. Reconstitution of MoFe protein activity of mutant strains involved the titration of 0.25-ml portions of each extract with purified FeMo cofactor to constant C_2H_2 reduction activity, which was measured after allowing reconstitution to proceed at 30°C for 60 min. The isolated FeMo cofactor used was 2.9 mM in Mo and had a specific activity of 230 nmol of C_2H_2 reduced per min per ng-atom of Mo. Protein concentrations were determined by the biuret method (8).

RESULTS AND DISCUSSION

***nifN* sequence analysis.** The isolation of the *nifEN* gene cluster from *A. vinelandii* and determination of the complete nucleotide sequence of the *nifE* gene was previously described (7). The genetic and physical organization of the *nifHDK* and *nifEN* gene clusters is shown in Fig. 1. Both the *nifH* (4) and the *nifE* (7) genes from *A. vinelandii* are preceded by excellent consensus *nif* promoter sequences, suggesting that the *nifHDK* and the *nifEN* gene clusters each represent independently regulated transcription units as found for the analogous gene clusters from *K. pneumoniae* (3). In Fig. 2, the complete nucleotide sequence for *nifN* is shown. The *nifN* gene product is an acidic 458-residue polypeptide and has a molecular weight of 49,186 if the amino-terminal methionine residue is considered in the calculation. Because there is no DNA or amino acid sequence information available regarding the *nifEN* genes or their products from any other organism, their identification in *A. vinelandii* is based on their proximity to the *nifHDK* cluster, their hybridization with *K. pneumoniae nifEN* genes, and their function in FeMo cofactor biosynthesis (discussed later). Only 12 nucleotides separate the *nifE* and *nifN* genes, and the apparent ribosome binding site for *nifN* is partially overlapped by the termination signal for the *nifE* gene product (Fig. 2). Such overlapping translational stop and start signals often indicate translational coupling (13), and such coupling is frequently found for genes whose products form macromolecular complexes in equimolar amounts, for example, ribosomal proteins (2). In the region just distal to the end of *nifN*, there is another potential open reading frame. Whether this sequence actually represents a *nif*-specific gene which is transcribed and translated is not known. In the case of *K. pneumoniae*, there is a proposed gene, called *nifX*, which is reported to be cotranscribed with and located just distal to *nifN* (16). The function of the product of this putative *K. pneumoniae* gene remains obscure.

Sequence homologies. Several factors have led us to propose recently that the *nifEN* gene products form a FeMo cofactor binding complex structurally analogous to that

```
GAA TTC GGC TAT GGC GGC TAC GAC CGC ATG CTG GAA CTG GTG CGT
GLU PHE GLY TYR GLY GLY TYR ASP ARG MET LEU GLU LEU VAL ARG

CAC GTC TGC ATC ACC CTG GAA TGC CCG GTG TGG GAG GCG GTG CGC
HIS VAL CYS ILE THR LEU GLU CYS PRO VAL TRP GLU ALA VAL ARG

CGC CCC GCG CCC TGG GAC ATC CCC GCC AGC CAG GAC GCC CGC CCG
ARG PRO ALA PRO TRP ASP ILE PRO ALA SER GLN ASP ALA ARG PRO

AGC GGC GGC CCG TTC GGC GAA TGA  GGAGAACGC  ATG GCC GAG ATC
SER GLY GLY PRO PHE GLY GLU END             MET ALA GLU ILE
                          10
ATC AAT CGC AAC AAG GCC CTG GCC GTC AGC CCG CTG AAG GCC AGT
ILE ASN ARG ASN LYS ALA LEU ALA VAL SER PRO LEU LYS ALA SER
 20                                      30
CAG ACC ATG GGC GCC GCG CTG GCC ATC CTC GGC CTG GCT CTC AGC
GLN THR MET GLY ALA ALA LEU ALA ILE LEU GLY LEU ALA LEU SER
                          40
ATG CCG CTG TTT CAC GGC TCG CAG GGT TGC ACC GCG TTC GCC AAG
MET PRO LEU PHE HIS GLY SER GLN GLY CYS THR ALA PHE ALA LYS
 50                                      60
GTG TTC TTC GTT CGC CAT TTC CGC GAG CCG GTC CCG CTG CAG ACC
VAL PHE PHE VAL ARG HIS PHE ARG GLU PRO VAL PRO LEU GLN THR
                          70
ACC GCC ATG GAC CAG GTC AGT TCG GTG ATG GGC GCC GAC GAG AAC
THR ALA MET ASP GLN VAL SER SER VAL MET GLY ALA ASP GLU ASN
 80                                      90
GTC GTC GAG GCG CTG AAG ACC ATC TGC GAA CGG CAG AAT CCC TCG
VAL VAL GLU ALA LEU LYS THR ILE CYS GLU ARG GLN ASN PRO SER
                         100
GTC ATC GGC CTG CTC ACC ACC GGC CTG TCG GAA ACC CAG GGC TGC
VAL ILE GLY LEU LEU THR THR GLY LEU SER GLU THR GLN GLY CYS
110                                     120
GAC CTG CAC ACC GCC CTG CAC GAG TTC CGT ACC CAG TAC GAG GAG
ASP LEU HIS THR ALA LEU HIS GLU PHE ARG THR GLN TYR GLU GLU
                         130
TAC AAG GAC GTG CCG ATC GTT CCG GTG AAC ACG CCG GAC TTC AGC
TYR LYS ASP VAL PRO ILE VAL PRO VAL ASN THR PRO ASP PHE SER
140                                     150
GGC TGC TTC GAG AGC GGC TTC GCC GCG GCG GTC AAG GCG ATC GTC
GLY CYS PHE GLU SER GLY PHE ALA ALA ALA VAL LYS ALA ILE VAL
                         160
GAG ACG CTG GTG CCC GAA CGG CGC GAT CAG GTG GGC AAG CGT CCG
GLU THR LEU VAL PRO GLU ARG ARG ASP GLN VAL GLY LYS ARG PRO
170                                     180
CGC CAG GTC AAC GTG CTG TGC TCG GCC AAT CTC ACC CCG GGC GAT
ARG GLN VAL ASN VAL LEU CYS SER ALA ASN LEU THR PRO GLY ASP
                         190
CTG GAG TAC ATC GCC GAA AGC ATC GAG AGC TTC GGT CTG CGC CCC
LEU GLU TYR ILE ALA GLU SER ILE GLU SER PHE GLY LEU ARG PRO
200                                     210
TTG CTG ATC CCC GAC CTG TCC GGC TCG CTC GAC GGC CAC CTG GAC
LEU LEU ILE PRO ASP LEU SER GLY SER LEU ASP GLY HIS LEU ASP
                         220
GAG AAC CGT TTC AAT GCC CTG ACC ACC GGC GGA CTG AGC GTG GCC
GLU ASN ARG PHE ASN ALA LEU THR THR GLY GLY LEU SER VAL ALA
230                                     240
GAA CTG GCC ACC GCC GGA CAG AGC GTC GCC ACT CTG GTG GTC GGG
GLU LEU ALA THR ALA GLY GLN SER VAL ALA THR LEU VAL VAL GLY
                         250
CAG AGC CTG GCG GGT GCG GCC GAC GCC CTG GCC GAG CGC ACC GGC
GLN SER LEU ALA GLY ALA ALA ASP ALA LEU ALA GLU ARG THR GLY
260                                     270
GTG CCC GAC CGG CGC TTC GGC ATG CTC TAC GGT CTG GAT GCG GTC
VAL PRO ASP ARG ARG PHE GLY MET LEU TYR GLY LEU ASP ALA VAL
                         280
GAT GCC TGG CTG ATG GCG CTG GCC GAG ATC AGC GGC AAT CCG GTG
ASP ALA TRP LEU MET ALA LEU ALA GLU ILE SER GLY ASN PRO VAL
290                                     300
CCC GAC CGC TAC AAG CGC CAG CGT GCC CAA TTG CAG GAC GCC ATG
PRO ASP ARG TYR LYS ARG GLN ARG ALA GLN LEU GLN ASP ALA MET
                         310
CTC GAC ACC CAC TTC ATG CTC AGT TCC GCA CGC ACG GCC ATC GCC
LEU ASP THR HIS PHE MET LEU SER SER ALA ARG THR ALA ILE ALA
320                                     330
GCC GAT CCC GAT CTG CTG CTC GGT TTC GAT GCC CTG CTG CGC AGC
ALA ASP PRO ASP LEU LEU LEU GLY PHE ASP ALA LEU LEU ARG SER
                         340
ATG GGC GCG CAC ACG GTA GCC GCC GTG GTG CCG GCC CGG GCC GCC
MET GLY ALA HIS THR VAL ALA ALA VAL VAL PRO ALA ARG ALA ALA
350                                     360
GCG CTG GTC GAT TCG CCT CTG CCC TCC GTG CGG GTC GGC GAC CTG
ALA LEU VAL ASP SER PRO LEU PRO SER VAL ARG VAL GLY ASP LEU
                         370
GAG GAC CTC GAG CAT GCC GCC CGC GCC GGC CAG GCC CAA CTG GTG
GLU ASP LEU GLU HIS ALA ALA ARG ALA GLY GLN ALA GLN LEU VAL
380                                     390
ATC GGC AAC AGC CAC GCC CTG GCC AGC GCC CGT CGC CTC GGT GTG
ILE GLY ASN SER HIS ALA LEU ALA SER ALA ARG ARG LEU GLY VAL
                         400
CCA CTG TTG CGT GCC GGC TTC CCG CAG TAC GAT CTG CTG GGC GGT
PRO LEU LEU ARG ALA GLY PHE PRO GLN TYR ASP LEU LEU GLY GLY
410                                     420
TTC CAA CGC TGC TGG TCC GGC TAC CGC GGC AGC AGT CAG GTG CTG
PHE GLN ARG CYS TRP SER GLY TYR ARG GLY SER SER GLN VAL LEU
                         430
TTC GAT CTG GCC AAC CTG CTG GTC GAA CAC CAC CAG GGT ATC CAG
PHE ASP LEU ALA ASN LEU LEU VAL GLU HIS HIS GLN GLY ILE GLN
440                                     450
CCC TAT CAT TCG ATC TAT GCG CAA AAA CCG GCA ACC GAA CAG CCG
PRO TYR HIS SER ILE TYR ALA GLN LYS PRO ALA THR GLU GLN PRO

CAA TGG AGG CAC TGAGCGATGTCCAGCCCGACCCGACAATTGCAGGTATTGGATA
GLN TRP ARG HIS END

GCGACCACGACGGCACCCTGCTGAAGGTCGCCTTCGCCTCGTCCGACCGCGAACTGGTC

GAC
```

FIG. 2. Complete nucleotide sequence of the *A. vinelandii nifN* gene. Only one strand, that with the polarity of the mRNA, is shown. The deduced amino acids are shown below the coding sequence. The numbered sequences represent the *nifN* gene. The translated upstream sequences represent *nifE* gene coding sequences. The complete nucleotide sequence of *nifE* can be found in reference 7. The probable ribosome binding site, GAGGAG, for the *nifN* gene overlaps with the termination signal, TGA, for the *nifE* gene. Another potential ribosome binding site, GAGG, is found in the end of the *nifN* coding sequence, and a potential initiation codon and open reading frame are located three nucleotides past the *nifN* termination codon.

found in the MoFe protein (*nifDK* gene products) (7). These factors include (i) a similar relationship in molecular size and charge when *K. pneumoniae nifD*- and *nifK*-encoded polypeptides are compared with the *nifE*- and *nifN*-encoded polypeptides, respectively (17); (ii) a mutual stability relationship between *nifE*- and *nifN*-encoded polypeptides in *K. pneumoniae*, indicating that these gene products could form a macromolecular complex (17); and (iii) the likely FeMo cofactor binding properties of the *nifEN* gene products indicated by their role in FeMo cofactor biosynthesis. Striking sequence homology between the *nifD* gene product and the *nifE* gene product strongly supported that notion (7).

In Fig. 3, we present a more complete *nifD-nifE* product sequence homology comparison than that previously published and also compare *nifK* and *nifN* gene product sequences. These comparisons reveal not only that the *nifD* and *nifE* gene products are structurally homologous but that the *nifK* and *nifN* gene products are also structurally homologous. The *nifE* and *nifN* gene products show only very weak homology when compared with each other. Although there is no biochemical information on the potential subunit composition of a proposed *nifEN* product complex, all of the above data provide evidence that such a complex does exist.

The considerable sequence homologies apparent in Fig. 3 potentially reflect a number of important structural features, which could have functions shared by both the MoFe protein and the proposed *nifEN* product complex. For example, these two species could share primary protein sequence homology in regions surrounding FeMo cofactor binding domains, subunit-subunit interface sites, and possibly Fe-S

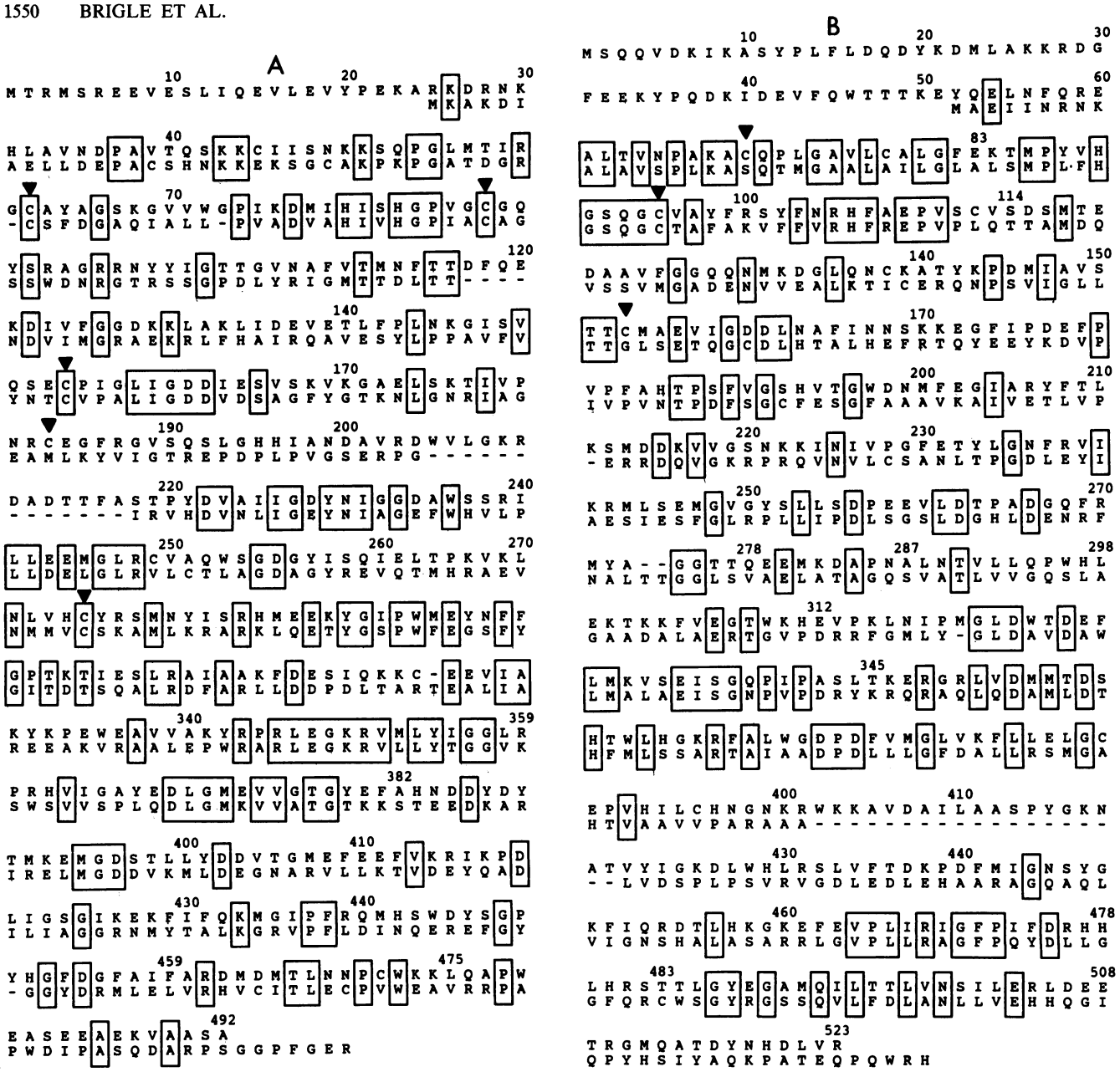

FIG. 3. Comparison of the *nifD*- and *nifE*-encoded (A) and the *nifK*- and *nifN*-encoded (B) polypeptides. The upper sequences represent the *nifD*- or the *nifK*-encoded polypeptides, and lower sequences represent the *nifE*- or *nifN*-encoded polypeptides. Perfect homologies are boxed. Alignment adjustments were made to provide the best fit, and these adjustments are indicated by hyphens. The entire sequence of each polypeptide is shown. Cysteine residues which are conserved among those species whose *nifD* product and *nifK* product sequences have been determined are indicated by arrowheads.

cluster binding sites. The existence of FeMo cofactor binding or assembly site(s) within the proposed *nifEN* product complex must, however, be the paramount structural feature to be considered in any analysis because of the essential role that the *nifEN*-encoded proteins occupy in the biogenesis of native FeMo cofactor.

Of all the conserved sequences observed in Fig. 3, it is interesting that the region from amino acids 220 to 379 in the *nifD-nifE* product homologies is clearly the most striking. We have previously suggested that the region surrounding Cys residue 275 in the α subunit of the MoFe protein is a likely candidate for FeMo cofactor binding (4). This suggestion was based on the observation that there is possibly only a single thiol ligand in the FeMo cofactor (6) (potentially contributed by Cys residue 275; also see below), and the suggestion was made because of the concentration of interspecies *nifD* gene product conservation of amino acids containing amide functional groups in the region immediately surrounding Cys residue 275. Such amino acids can be considered to have functional groups which most closely duplicate *N*-methylformamide, the agent most frequently used to extrude FeMo cofactor from native MoFe protein (24). It is, therefore, probably significant that the homologous region within the *nifE* product contains a Cys residue homologous to the *nifD*-encoded Cys residue 275 as well as strong *nifD* product sequence homology upstream and downstream from this residue (Fig. 3A). However, such *nifD* product-*nifE* product sequence conservation provides only

an indication of one potential FeMo cofactor binding region. The possibility of the contribution of distantly located amino acids within or among the individual subunits as ligands to a single FeMo cofactor molecule must also be considered. It should be kept in mind that, whatever the form of FeMo cofactor bound to the proposed *nifEN* gene products complex, it must ultimately escape from that species during the maturation of the MoFe protein. Consequently, the specific amino acid ligands, which actually bind a form of FeMo cofactor to the proposed *nifEN* gene products complex, could be quite different from those ligands which bind the mature FeMo cofactor to the MoFe protein. The potential involvement of the peptide backbone itself in FeMo cofactor binding remains yet another possibility (24).

Sequence comparisons of MoFe protein subunits from a wide variety of diazotrophic species have been used in attempts to identify structurally important regions of these polypeptides. Such comparisons have largely centered on sequence homologies surrounding cysteinyl residues because of the well-known role of Cys-thiol ligation to iron-containing prosthetic groups within a wide variety of metalloproteins (see references 4 and 22 for further discussion). These sequence comparisons have revealed five conserved Cys residues in all available α-subunit sequences (residues 62, 88, 154, 183, and 275, by the numbering system in Fig. 3) and three conserved Cys residues in all available β-subunit sequences (residues 70, 95, and 153, by the numbering system in Fig. 3). All of the conserved Cys residues are located in regions of generally high sequence homology when interspecies MoFe protein sequence comparisons are made. It is unlikely that these conserved Cys residues compose all of the metallocluster ligands to be found in the MoFe protein (see discussions in references 4 and 22), yet it is reasonable to expect that all or most of these conserved Cys residues participate in the binding of metallocenters within the MoFe protein complex.

Comparison of the *nifD* product with the *nifE* product reveals that four of the five conserved interspecies MoFe protein α-subunit Cys residues can also be considered as conserved within the *nifE* product. Among these is Cys residue 275, discussed above. Comparison of the *nifK* product with the *nifN* product reveals that only one of three conserved interspecies Cys residues within the MoFe protein β subunit can be considered conserved within the *nifN* gene product. However, one of those unconserved residues (Cys residue 70) is replaced by Ser in a highly homologous portion of the *nifN* product. In the absence of a source of purified *nifEN* gene product complex, which could be subjected to spectroscopic as well as other biophysical techniques already used to study the MoFe protein, it is not possible to assign a functional significance to these Cys residues conserved within both the MoFe protein subunits and the *nifEN*-encoded products. However, the results do suggest that it is highly possible that there are other metal-containing centers within the proposed *nifEN* products complex in addition to a form of the FeMo cofactor.

Evolutionary relationship between the MoFe protein and the EN protein. The strong sequence homology for the *nifD* and *nifE* gene products as well as for the *nifK* and *nifN* gene products suggests that the *nifEN* genes and the *nifDK* genes could have evolved from a common ancestral origin. This possibility finds additional support from comparison of the molecular size and charge of the individual gene products (4, 7, 17) as well as the genetic organization of *nifDK* and *nifEN* within their respective gene clusters (3, 7). The physical location of the *nifEN* gene cluster immediately downstream from the *nifHDK* cluster also indicates that one gene cluster could have easily originated from the tandem duplication of the other. Comparison of the DNA sequences immediately preceding and distal to the *nifDK* and *nifEN* gene sequences does not, however, reveal an obvious site for such a recombination event. The strong primary sequence homologies between the MoFe protein and the proposed *nifEN* gene products complex also raise the possibility that one protein species might functionally replace the other (7). However, we have found that *A. vinelandii* mutant strains, which are deleted for the *nifDK* genes (18), as well as mutant strains deleted for the *nifEN* genes (this study) are Nif$^-$ when cultured in a normal Burk medium lacking a source of fixed nitrogen.

Genetic analysis. Formal proof that the *A. vinelandii* genes we have isolated and sequenced are indeed homologs to the previously identified *K. pneumoniae nifEN* genes requires that a functional role in FeMo cofactor biosynthesis be demonstrated for the *A. vinelandii* gene products. Such proof requires the demonstration that mutant strains altered in *nifE* and *nifN* are deficient in FeMo cofactor biosynthesis. In Fig. 1, the physical maps of the *nifEN* coding region from three *A. vinelandii* mutant strains, which carry deletions in portions or all of the *nifEN* gene cluster, are shown. These strains were constructed by using hybrid plasmid DNAs, which contain various defined deletions within the cloned regions of *nif*-specific *A. vinelandii* DNA (Fig. 1). The exact procedures for such constructions are described elsewhere (18). Strain DJ42 (Δ*nifEN*) is deleted for the entire *nifEN* cluster as well as for sequences extending shortly upstream and downstream from *nifE* and *nifN* (Fig. 1). Strain DJ35 (Δ*nifE*) contains an in-frame deletion located entirely within the *nifE* coding sequences, and strain DJ39 (Δ*nifN*) contains a deletion beginning from the first *Sma*I restriction enzyme site located within the *nifN* gene and ending shortly downstream from *nifN* coding sequences (Fig. 1). Each of these mutant strains exhibits a Nif$^-$ phenotype when cultured in a Burk minimal medium lacking a fixed nitrogen source, and each mutant strain grows at rates comparable to the wild-type strain when a fixed source of nitrogen is added to that medium. It is highly unlikely that the deleted regions located upstream or downstream from the *nifEN* coding sequences are responsible for the Nif$^-$ character in strains DJ42 and DJ39. This conclusion is based on the observed Nif$^+$ phenotype of other strains which we have constructed that carry a deletion or an insertion (or both a deletion and an insertion) in these regions but which does not extend into the *nifEN* coding sequences. However, strains with such mutations located immediately downstream from the *nifN* coding sequences do have decreased Fe protein activity in crude extracts (D. R. Dean and W. E. Newton, unpublished data). Various purified DNA fragments which include all (or portions) of the *nifEN* cluster were used for transformations in marker rescue experiments and permitted the confirmation of the size and the location of the deletions shown in Fig. 1 (data not shown).

Assays and reconstitutions. In Table 1, the specific activities for the Fe protein and MoFe protein in crude extracts prepared from the parental wild-type strain and the individual mutant strains are shown. Each of the mutant strains almost completely lacked MoFe protein activity, yet each strain retained significant amounts of Fe protein activity. For strain DJ35 (Δ*nifE*), about 75% of wild-type Fe protein was observable, while for strains DJ39 and DJ42, where the deleted regions extend beyond the *nifEN* coding sequences, a much lower (about 25%) Fe protein activity was observed.

TABLE 1. Component protein activities in mutant and wild-type strains

Strain	Sp act of:	
	Fe protein[a]	MoFe protein[b]
Wild type	48.0	57.4
DJ35 (Δ*nifE*)	30.4	0.7
DJ39 (Δ*nifN*)[c]	9.7	0.7
DJ42 (Δ*nifEN*)[c]	10.3	0.9

[a] Nanomoles of ethylene formed per minute per milligram of crude extract of protein in the presence of saturation levels of purified *A. vinelandii* MoFe protein. All values represent the average of three determinations.

[b] Nanomoles of ethylene formed per minute per milligram of crude extract of protein in the presence of saturation levels of purified *A. vinelandii* Fe protein.

[c] Genotypic designations are added only for clarity and do not strictly indicate the complete *nif* genotype of the organism.

In these latter strains, this effect is likely a result of sequences deleted downstream from *nifN* and not the consequence of alterations in *nifEN* gene products (see above). However, the mutant strains needed to resolve this issue completely are not yet available.

The results of MoFe protein reconstitution experiments performed by the incubation of crude extracts prepared from the individual mutant strains in the presence of extracted FeMo cofactor are shown in Table 2. A significant amount of MoFe protein activity was restored by this treatment; for DJ35, reconstituted activity was comparable to the activities of UW45 (a previously described FeMo cofactorless *A. vinelandii* strain) extracts (17). Extracts of DJ39 and DJ42 consistently had lower activities, a phenomenon that appears to parallel their lowered Fe protein activity. These results confirm that the physiological defect of these mutant strains resides in their inability to accumulate FeMo cofactor containing MoFe protein.

Summary. The data presented in this study show that the functional *nifE* and *nifN* genes, which have now been isolated and sequenced from the diazotroph *A. vinelandii*, encode products which bear striking sequence homology when compared with the *nifD*- and *nifK*-encoded polypeptides, respectively. Such sequence homology is an indication that the *nifEN*-encoded products could form a complex structurally analogous to the MoFe protein and that the *nifEN* and *nifDK* genes are themselves ancestrally related. It is likely that FeMo cofactor or an intermediate in FeMo cofactor biosynthesis resides upon the proposed *nifEN* product complex sometime before cofactor donation to the MoFe protein. Whether the assembly of FeMo cofactor takes place wholly or partially on the proposed *nifEN* product complex remains to be determined. Biochemical efforts aimed at purifying a native form of the proposed *nifEN* product complex should be substantially aided by information provided here and elsewhere (7) on the primary structure and likely composition of this species. Moreover, the availability of *A. vinelandii* mutant strains containing defined deletions within the *nifEN* cluster should prove to have a practical application towards the elucidation of the biochemical sequence of events during FeMo cofactor biogenesis. Finally, the *nifDK-nifEN* product homologies described here are being used in our laboratory to provide insights for strategies in site-directed mutagenesis experiments aimed at identifying specific amino acid ligands that participate in the binding of the FeMo cofactor to the MoFe protein.

TABLE 2. Complementation of extracts of *A. vinelandii* mutant strains with isolated FeMo cofactor

Strain	Addition	Sp act[a]
DJ35 (Δ*nifE*)	None	0.1
	FeMo cofactor	15.1
DJ39 (Δ*nifN*)[b]	None	0.2
	FeMo cofactor	3.7
DJ42 (Δ*nifEN*)[b]	None	0.2
	FeMo cofactor	7.1

[a] Nanomoles of ethylene formed per minute per milligram of crude extract protein in the presence or the absence of added FeMo cofactor. All values represent the average of three determinations. All activity measurements were performed in the presence of saturation levels of purified *A. vinelandii* Fe protein.

[b] Genotypic designations are added only for clarity and do not strictly indicate the complete *nif* genotype of the organism.

ACKNOWLEDGMENTS

We thank Stephen F. Gheller for the preparation of FeMo cofactor and Marty Jacobson for helpful discussions.

This work was supported by grant 85-CRCR-1-1906 from the U.S. Department of Agriculture and by grant DMB-8510029 from the National Science Foundation.

LITERATURE CITED

1. **Bagdasarian, M., R. Lurz, B. Ruckert, F. C. H. Franklin, M. M. Bagdasarian, J. Frey, and K. N. Timmis.** 1981. Specific-purpose plasmid cloning vectors. Gene **16**:237–247.
2. **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.
3. **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.
4. **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.
5. **Burgess, B. K., D. B. Jacobs, and E. I. Stiefel.** 1980. Large-scale purification of high activity *Azotobacter vinelandii* nitrogenase. Biochim. Biophys. Acta **614**:196–209.
6. **Burgess, B. K., E. I. Stiefel, and W. E. Newton.** 1980. Oxidation-reduction properties and complexation reactions of the iron-molybdenum cofactor of nitrogenase. J. Biol. Chem. **255**:353–356.
7. **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.
8. **Gornall, A. G., C. J. Bardawall, and M. M. David.** 1948. Determination of serum proteins by means of the biuret reaction. J. Biol. Chem. **177**:751–766.
9. **Hageman, R. V., and R. H. Burris.** 1978. Nitrogenase and nitrogenase reductase associate and dissociate with each catalytic cycle. Proc. Natl. Acad. Sci. USA **75**:2699–2702.
10. **Hageman, R. V., and R. H. Burris.** 1979. Changes in the EPR signal of dinitrogenase from *Azotobacter vinelandii* during the lag period before hydrogen evolution begins. J. Biol. Chem. **254**:11189–11192.
11. **Hawkes, T. R., P. A. McLean, and B. E. Smith.** 1984. Nitrogenase from *nifV* mutants of *Klebsiella pneumoniae* contains an altered form of the iron-molybdenum cofactor. Biochem. J. **217**:317–321.
12. **Messing, J., and J. Vieira.** 1982. A new pair of M13 vectors for selecting either DNA strand of double-digest restriction fragments. Gene **19**:269–276.
13. **Oppenheim, D. S., and C. Yanofsky.** 1980. Translational coupling during expression of the tryptophan operon in *Escherichia coli*. Genetics **95**:785–795.

14. **Orme-Johnson, W. H.** 1985. Molecular basis of biological nitrogen fixation. Annu. Rev. Biophys. Biophys. Chem. **14**:419–459.
15. **Page, W. J., and M. von Tigerstrom.** 1979. Optimal conditions for transformation of *Azotobacter vinelandii*. J. Bacteriol. **139**: 1058–1061.
16. **Puhler, A., M. O. Aguilar, M. Hynes, P. Muller, W. Klipp, U. Priefer, R. Simon, and G. Weber.** 1984. Advances in the genetics of free-living and symbiotic nitrogen fixing bacteria, p. 609–619. *In* C. Veeger and W. E. Newton (ed.), Advances in nitrogen fixation research. Martinus Nijhoff Publishing, The Hague, The Netherlands.
17. **Roberts, G. P., T. MacNeil, D. MacNeil, and W. J. Brill.** 1978. Regulation and characterization of protein products coded by the *nif* (nitrogen fixation) genes of *Klebsiella pneumoniae*. J. Bacteriol. **136**:267–279.
18. **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.
19. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain termination inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
20. **Shah, V. K., and W. J. Brill.** 1977. Isolation of an iron-molybdenum cofactor from nitrogenase. Proc. Natl. Acad. Sci. USA **74**:3249–3253.
21. **Strandberg, G. W., and P. W. Wilson.** 1968. Formation of the nitrogen-fixing enzyme systems in *Azotobacter vinelandii*. Can. J. Microbiol. **14**:25–31.
22. **Thony, B., K. Kaluza, and H. Hennecke.** 1985. Structural and functional homology between the α and β subunits of the nitrogenase MoFe protein as revealed by sequencing the *Rhizobium japonicum nifK* gene. Mol. Gen. Genet. **198**:441–448.
23. **Ugalde, R. A., J. Imperial, V. K. Shah, and W. J. Brill.** 1984. Biosynthesis of iron-molybdenum cofactor in the absence of nitrogenase. J. Bacteriol. **159**:888–893.
24. **Walters, M. A., S. K. Chapman, and W. H. Orme-Johnson.** 1986. The nature of amide ligation to the metal sites of FeMoco. Polyhedron **5**:561–565.