Proc. Natl. Acad. Sci. USA
Vol. 82, pp. 5720–5723, September 1985
Biochemistry

Azotobacter vinelandii nifD- and *nifE*-encoded polypeptides share structural homology

(FeMo cofactor/DNA sequence/nitrogenase/nitrogen fixation)

DENNIS R. DEAN AND KEVIN E. BRIGLE

Battelle–C. F. Kettering Research Laboratory, 150 East South College Street, Yellow Springs, OH 45387

Communicated by Masayasu Nomura, May 17, 1985

ABSTRACT **The *Azotobacter vinelandii nifE* gene was isolated and its complete nucleotide sequence was determined. The amino acid sequences deduced from the *A. vinelandii nifE* and *nifD* gene sequences were compared and found to share striking primary sequence homology. This homology implies a functional and possibly an evolutionary relationship between these two gene products. The structural homology is discussed with regard to the potential FeMo cofactor binding properties of these polypeptides and the possible role of a *nifEN* product complex as a surrogate MoFe protein.**

Nitrogen fixation is catalyzed by nitrogenase, a complex enzyme composed of two separately purifiable proteins (for an overview, see ref. 1). Three polypeptides, the *nifH* gene product (Fe protein), the *nifD* gene product (MoFe protein α subunit), and the *nifK* gene product (MoFe protein β subunit) compose the structural components of the active nitrogenase complex. Of the nitrogenase protein components, the smaller Fe protein (or component II) acts as an ATP-binding specific electron donor to the larger MoFe protein (or component I). The MoFe protein is an $\alpha_2\beta_2$ protein of ≈230,000 daltons with at least six metal-containing prosthetic groups, including a Mo-containing species. Very little is known about the structure, redox properties, and function in the catalytic process of the individual prosthetic groups or their spatial arrangement within the MoFe protein. With regard to the Mo-containing species, Shah and Brill (2) discovered that an FeMo cofactor could be isolated intact and free of the MoFe protein by acid/base treatment of purified MoFe protein followed by extraction into *N*-methylformamide. The extracted FeMo cofactor can activate nitrogenase in crude extracts of certain *Klebsiella pneumoniae* and *A. vinelandii* mutants that direct the synthesis of an inactive MoFe protein species. In the case of *K. pneumoniae*, the FeMo cofactor reconstitutable mutants have lesions that fall into three complementation groups—*nifB*, *nifE*, and *nifN* (3). Mutation in another *K. pneumoniae* allele; *nifV* leads to the synthesis of a MoFe protein that carries an altered FeMo cofactor species and exhibits dramatic changes in substrate recognition and reactivity (4). These results indicate that FeMo cofactor plays an important role in substrate binding and reduction, and consequently the structure and biosynthesis of FeMo cofactor is of considerable importance.

Several pathways for FeMo cofactor biosynthesis can be imagined. FeMo cofactor could be assembled stepwise into the apo-MoFe protein, FeMo cofactor could be synthesized and inserted into the apo-MoFe protein, or a combination of these two mechanisms could occur. In relation to this question, Ugalde *et al.* (3) have recently reported that FeMo cofactor or a closely related species accumulates in strains of *K. pneumoniae* that completely lack the MoFe protein. In these mutant strains, the FeMo cofactor moiety apparently accumulates on protein species, indicating that such protein species could represent intermediates in the FeMo cofactor biosynthetic pathway. Importantly, these species do not accumulate in *K. pneumoniae nifB*, *nifE*, or *nifN* mutants, implying that one or more of their respective products could carry a form of FeMo cofactor. Of these three *nif* alleles, the *nifEN* genes form a single transcription unit adjacent to the *nif* structural gene cluster (5). Furthermore, the *nifE* and *nifN* gene products have mutual stability requirements, indicating that these polypeptides probably form a complex (6). This same mutual stability relationship is also found for the MoFe protein subunits (6).

Considering the similar genetic organization of the *nifDK* and *nifEN* genes, the mutual stability requirements of their respective products, and their apparent FeMo cofactor binding properties, we wondered whether the *nifDK* and *nifEN* gene products might share some structural similarities. We have previously sequenced the entire *nif* structural gene cluster isolated from the aerobic nitrogen fixer *Azotobacter vinelandii* (unpublished data). Here we present the sequence determination for the entire *A. vinelandii nifE* gene and compare the *nifD*- and *nifE*-encoded polypeptide sequences deduced from their respective coding sequences.

METHODS

Wild-type *A. vinelandii* was used as the primary source of DNA. A genomic library of *A. vinelandii* DNA was prepared by ligating partial *Sau*3A restriction enzyme-digested, size-fractionated [10–20 kilobases (kb)] DNA with purified λ Charon 30 *Bam*HI arms. The ligated DNA mixture was packaged *in vitro* into λ and used to infect *Escherichia coli* KH802 cells. The infected cells were spread directly onto LB soft agar plates, and after incubation the resultant plaques were lifted onto GeneScreen filters (New England Nuclear) and the filters were hybridized to radioactively labeled purified *K. pneumoniae nifHDK* probes (5). DNA was prepared from several positive hybridizing phage, and a 12-kb *Xho* I restriction enzyme fragment containing the *A. vinelandii nifHDK* genes and their flanking sequences was subcloned into the plasmid pKT230 (7). This hybrid plasma was ultimately the source for all DNA preparations and sequence analyses. The *A. vinelandii nifE* gene was located on the cloned 12-kb fragment by hybridization with purified *K. pneumoniae nifE* sequences (5). All DNA sequence determinations were performed by using the dideoxy chain-termination procedure (8) and hybrids of filamentous phage vectors described by Messing (9).

RESULTS AND DISCUSSION

The assignment of the *A. vinelandii nifE* gene product to the polypeptide sequence shown in Fig. 2 is based on the following: (*i*) the homologous location of the *nifE* coding sequences relative to the *nif* structural genes in both *K.*

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

pneumoniae (5) and *A. vinelandii* (Fig. 1), (*ii*) hybridization of *nifE*-specific *K. pneumoniae* DNA with the appropriate *A. vinelandii* sequences, (*iii*) striking sequence homology between the *K. pneumoniae nifE* promoter region (5) and the *A. vinelandii* sequences preceding the *nifE* coding region (Fig. 2), and (*iv*) the *A. vinelandii nifE* polypeptide deduced from the nucleotide sequence is very similar in size and charge to the *K. pneumoniae nifE* polypeptide (6).

The *nifE* initiation codon is preceded by an excellent ribosome binding site 9 base pairs (bp) upstream and a perfect consensus *nif* promoter (ref. 5; Fig. 2), CTGG-8 bp-TTGC. Our preliminary DNA sequence experiments indicate that another polypeptide is encoded immediately downstream from *nifE*, and we have tentatively assigned this sequence as the *nifN* coding region. The *nifE* gene product is an acidic 456 amino acid polypeptide of 50,236 daltons.

We have previously isolated the entire *A. vinelandii nif* structural gene region (Fig. 1) and determined its complete nucleotide sequence (unpublished data). After completion of the DNA sequences, the *nifE* polypeptide sequence was compared to the *nifD* and *nifK* polypeptide sequences. Striking amino acid sequence homology was found only between the *nifE* product and the *nifD* product, and these homologies are shown in Fig. 3. There is a region of 182 amino acids where there is 37% homology. Only one adjustment in alignment was necessary for this comparison, and the adjustment is indicated by an asterisk at amino acid 325 in the *nifD* encoded polypeptide sequence. There are four regions where the sequence homologies are most striking: 222–248 (15/27 amino acids conserved), 288–305 (10/18 amino acids conserved), 344–358 (12/15 amino acids conserved), and 370–380 (8/10 amino acids conserved). If neutral substitutions are considered (for example, glutamic acid for aspartic acid), the *nifE–nifD* product homologies between amino acids 222 and 380 become even more significant.

It was also of interest to compare the *A. vinelandii nifE* polypeptide to *nifD* gene products from other diazotrophic species. In a number of cases, amino acids not conserved between the *A. vinelandii nifE* and *nifD* gene products are conserved between the *A. vinelandii nifE* and other *nifD* products. For example, amino acid 249 is cysteine in the *A. vinelandii nifD* sequence, valine in the *A. vinelandii nifE* sequence, and also valine in the corresponding *Anabaena* (10), *C. pasteurianum* (11), and *R. parasponium* (unpublished data cited in ref. 11) *nifD* sequences. Generally, the sequences conserved between the *A. vinelandii nifD* product and the *A. vinelandii nifE* product are also highly conserved among all *nifD* product sequences. Most substitutions in these regions can be considered neutral.

Our rationale for comparing the *nifE* product to the *nifD* and *nifK* products was based on the observation that the *nifDK* complex (MoFe protein) and probably a *nifEN* product complex bind a form of FeMo cofactor and thus these species might share primary sequence homology in their FeMo cofactor binding regions. The striking homology between the *nifD* product and the *nifE* product clearly supports this idea. Based on these results, we predict that the *nifK* product and the *nifN* product will also share sequence homology. It is likely that a detailed comparison of the *nifDK* products to the *nifEN* products will ultimately contribute to the elucidation of the spatial arrangement of FeMo cofactor within the MoFe protein. In particular, these sequence comparisons should be considered in site-directed mutagenesis strategies aimed at identifying amino acid ligands that participate in binding FeMo cofactor to the MoFe protein. These striking sequence homologies should also stimulate discussion on the possible evolutionary relationship between the *nifD* and *nifE* genes as well as the evolutionary development of the *nif* regulon.

The result of this study is relevant to two other, as yetunexplained, phenomena regarding the expression and activity of nitrogenase from *A. vinelandii*. Shah *et al.* (12) have characterized many *A. vinelandii* mutant strains that lack MoFe protein cross-reacting material. However, all of these strains have detectable but very low levels of MoFe protein activity. Considering the homologous structure of the *nifE* and *nifD* polypeptides, we wonder if the residual MoFe protein activity detected in extracts of these mutant strains is supplied by other *nif*-specific gene products—in particular, the postulated *nifEN* complex.

Using the same mutant strains, Bishop and coworkers (13) have demonstrated the phenotypic reversal of the Nif^- character using either special physiological conditions or by isolation of pseudorevertant strains. Based on this result, Bishop proposed the existence of an alternative nitrogenase system in *A. vinelandii* and has identified four polypeptides, by using two-dimensional analysis of crude extracts, that accumulate under the appropriate physiological conditions and in the pseudorevertant strains. Two of these polypeptides have approximately the same size and charge as the *K. pneumoniae nifEN* gene products (6, 13). Thus, it should be considered a possibility that the MoFe protein-like activity

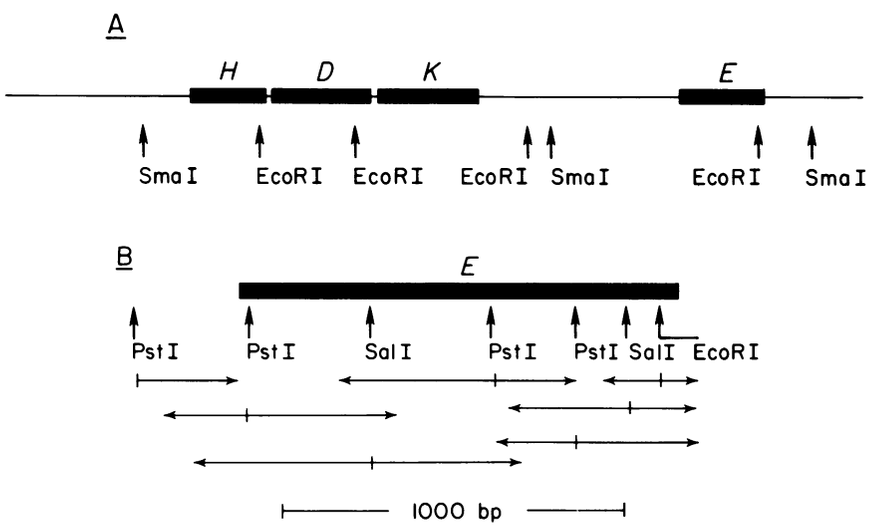

FIG. 1. (*A*) Physical map of the *A. vinelandii* genome in the *nif* structural gene region. (*B*) Sequencing strategy for the determination of the complete nucleotide sequence of the *nifE* gene.

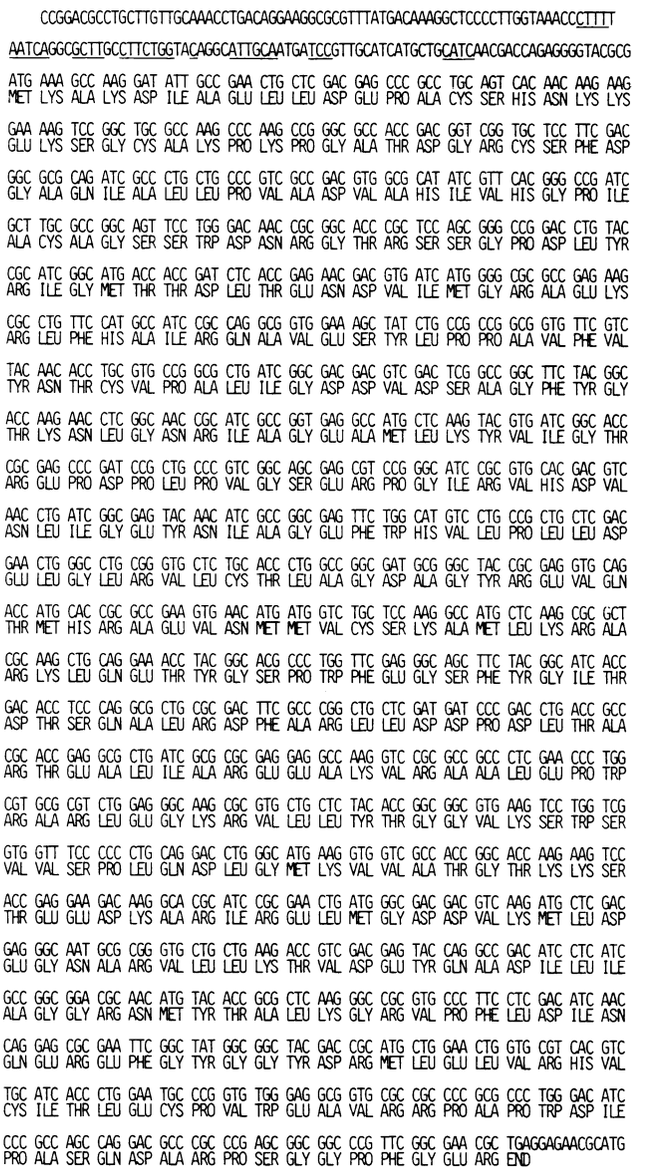

FIG. 2. The complete nucleotide sequence of the *A. vinelandii nifE* gene. Only the coding strand is shown and the correct direction of transcription is left to right, using the orientation in Fig. 1. Underlined nucleotides are homologous to the *K. pneumoniae nifE* promoter (5). A perfect consensus *nif* promoter (5) CTGG-8 bp-TTGCA is found in this region.

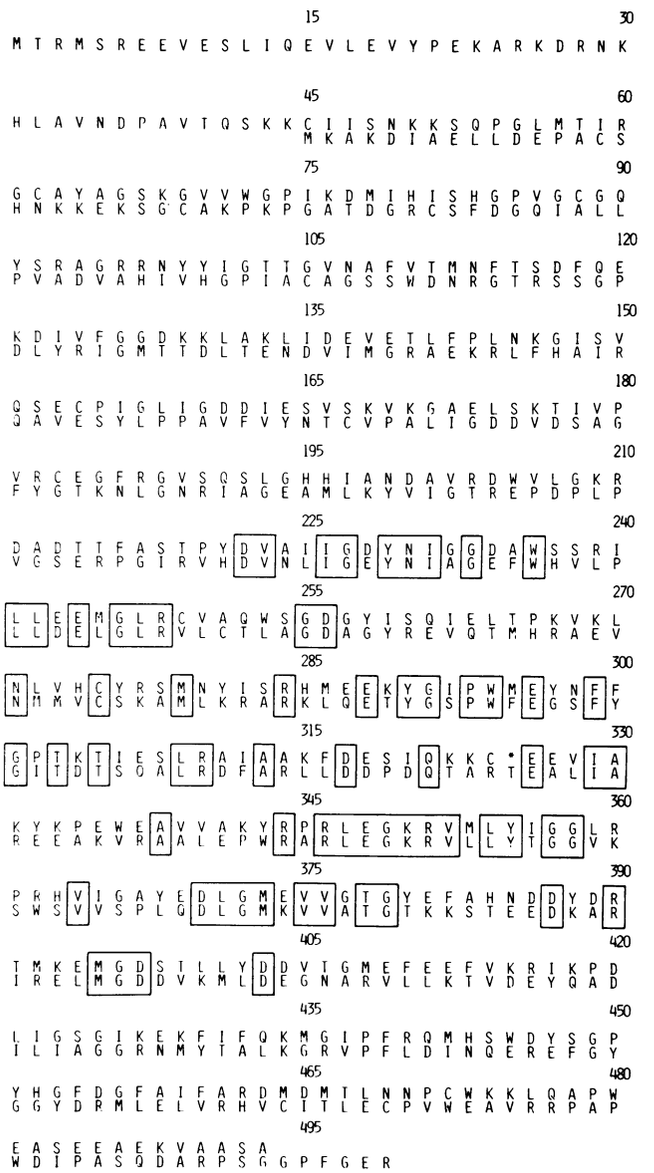

FIG. 3. Comparison of the *nifD*- and *nifE*-encoded polypeptides. Upper sequences represent the *nifD* polypeptide, and lower sequences represent the *nifE* polypeptide. Perfect homologies considered significant are boxed. An alignment adjustment of one amino acid is made in the *nifD* sequence and is indicated by an asterisk. Other sequence homologies, not boxed in the figure, are also apparent further upstream if realignments are made. For example, there is sequence homology surrounding cysteine-88 in the *nifD* product sequence when compared to cysteine-105 in the *nifE* product sequence; and there is homology surrounding cysteine-154 in the *nifD* product sequence when compared to cysteine-167 in the *nifE* product sequence. These homologies are probably significant but are not as striking as those boxed in the figure. Amino acids are identified by the single-letter code.

ascribed by Bishop to an alternative nitrogenase is due to the accumulation of the postulated *nifEN* complex under certain physiological conditions and in the pseudorevertant strains. The isolation of the *A. vinelandii nifEN* gene products and their further genetic manipulation will certainly allow an answer to these questions. These experiments and the completion of the *A. vinelandii nifN* sequence determination will provide further insight on the role of the *nifEN* products in nitrogen fixation.

We thank Dr. Greg Upchurch for providing his hybridization data prior to its publication. This work was supported by Grant 84-CRCR-1-1519 from the United States Department of Agriculture.

1. Burgess, B. K. (1984) in *Advances in Nitrogen Fixation Research*, eds. Newton, W. E. & Veeger, C. (Kluwer Boston, Hingham, MA), pp. 103–114.
2. Shah, V. K. & Brill, W. J. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 3249–3253.
3. Ugalde, R. A., Imperial, J., Shah, V. K. & Brill, W. J. (1984) *J. Bacteriol.* **159,** 888–893.
4. Hawkes, T. R., MacClean, P. A. & Smith, B. E. (1984) *Biochem. J.* **217,** 317–321.
5. Beynon, J., Cannon, M., Buchanan-Wollaston, V. & Cannon, F. (1983) *Cell* **34,** 665–671.
6. Roberts, G. P., MacNeil, T., MacNeil, D. & Brill, W. J. (1978) *J. Bacteriol.* **136,** 267–279.
7. Bagdasarian, M., Lurz, R., Ruckert, B., Franklin, F. C. H., Bagdasarian, M. M., Frey, J. & Timmis, K. N. (1981) *Gene* **16,** 237–247.
8. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 5463–5467.

9. Messing, J. (1983) in *Methods Enzymol.* **101,** 20–78.
10. Lammers, P. J. & Haselkorn, R. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 4723–4727.
11. Hase, T., Wakabayashi, S., Nakano, T., Zumft, W. G. & Matsubura, H. (1984) *FEBS Lett.* **166,** 39–43.
12. Shah, V. K., Davis, L. C., Gordon, J. K., Orme-Johnson, W. H. & Brill, W. J. (1973) *Biochim. Biophys. Acta* **292,** 246–255.
13. Bishop, P. E., Jarlenski, D. M. L. & Heatherington, D. R. (1980) *Proc. Natl. Acad. Sci. USA* **77,** 7342–7346.