Sequence of a 1.4 kb Eco RI fragment of Azotobacter vinelandii nif DNA

Koji Hiratsuka and Kenneth L.Rov*

Department of Microbiology, University of Alberta, Edmonton, Alberta T6G 2E9, Canada Submitted December 17, 1987

A 28 kb Hind III fragment of Azotobacter vinelandii UW DNA, identified by its homology to the Klebsiella pneumoniae nifHDKY genes in the plasmid pSA30 (1), was cloned into the broad host-range cosmid vector pSa747 (2,3). A 1.41 kb EcoRI segment, subcloned into pUC8, was completely sequenced using predominantly the chain termination method of Sanger et al. (4). The sequence obtained corresponded to the distal 20 amino acids of the nifH gene product (Fe protein), the first 409 amino acids of the nifD gene product (MoFe protein \alpha-subunit), and the included intergenic region. When compared with the sequence reported by Brigle et al. (5) a total of 13 base changes were found including the C to G change at position 196 which results in the introduction of an MspI site (which is confirmed by digestion). These changes in base sequence result in five amino acid differences including a glycine at position 3 of the nifD gene product. This is consistent with the amino acid sequence found by Lundell and Howard (6).

```
160 170 180 190 200 210 220 230 240 250 ACCCOCCATTATGAACOCTAAGGCAAGAGCAACTCATGCACATTATGAACOCTAAGGCAAGAGCAAGTCATACCCACAAGAGCAAGAGCAAGTCATACCCACAAAGTCATCACGAAAGTCATCACGAAAGTTCATCCCACAAAGTCATCACGAAAGTCATCACGAAAGTTCATCCCACAAA
               520
                                                                                                                                                                                                                                                                                550
                                                                                                                                 ThrThrGlyValAsnAla
                                                   620
                                                                                                                                  650 660
GAGTGCCCGATCGGCC
GluCysProIleGlyI
                                                                                                                                                                                            770
                                                                  780 790
ACCACATOGCCAACGACGC
                                                                                                                                     800
CCCACT
                                                                                                                                                                  810
                                                                                                                                                                  960 970 980 990 1000 1010 1020 1030
COCCECTOATTCCGATCAACCCCGAAGGTCAACCTGGATCACCTGCATCACCTCGATCAACTGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGGAACTGAACTGGAACTGGAACTGGAACTGGAACTGAACTGAACTGAACTGAACTGAACTGAACTGAACTGAACTGAACTGAACTGAACTGAACTGAACTGAACTGAACTGAACTGAACTGAACTGAACTGAACTGAACTGAACTGAACTGAACTGAACTGAACTGAACTGAACTGAACTGAACTGAACTGAACTGAACTGAACTGAACTGAACTGAACTGAACTGAACTGAACTGAACTGAACTGAACTGAACTGAACTGAACTGAACTGAACTGAACTGAACTGAACTGAACTGAACTGAACTGAACTGAACTGAACTGAACTGAACTGAACTGAACTGAACTGAACTGAACTGAACTGAACTGAACTGAACTGAACTGAACTGAACTGAACTGAACTGAACTGAACTGAACTGAACTGAACTGAACTGAACTGAACTGAACTGAACTGAACTGAACTGAACTGAACTGAACTGAACTGAACTGAACTGAACTGAACTGAACTGAACTGAACTGAACTGAACTGAACTGAACTGAACTGAACTGAACTGAACTGAACTGAACTGAACTGAACTGAACTGAACTGAACTGAACTGAACAACTGAACTGAACAACTGAACTGAACAACTGAACAACAACAACAACAA
 1360 1370 1380 1390 1400
ACCATGAAAGAATGGGTGACTCCACCCTGCTGTGAGGTGACGTGA
TheMetLysGluMetGlyAspSerTheLeuLeuTyzAspAspVelT
```

Differences between this sequence and that of Brigle et al. (5) are underlined. Arrows indicate changed amino acids.

ACKNOWLEDGEMENTS: This work was supported in part by an operating grant from the Natural Sciences and Engineering Research Council of Canada. pSA30 was provided by Dr. F. Ausubel and pSa747 was a gift of Dr. C. Kado.

*To whom correspondence should be addressed.

REFERENCES:

- 1. Cannon, F.C. et al. (1979) Molec. Gen. Genet. 174: 59-66.
- 2. Tait, R.C. et al. (1983) Bio/technol. 1: 269-275.
- 3. Doran, J.L. et al. (1987) J. Gen. Micro. 133: 2059-2072.
- Sanger, F. et al. (1977) Proc. Nat. Acad. Sci. U.S.A. 74: 5463-5467.
 Brigle, K.E. et al. (1985) Gene 37: 37-44.
- 6. Lundell, D.J. and Howard, J.B. (1981) J. Biol. Chem. 256: 6385-6391.