STUDIES ON THE STRUCTURE AND FUNCTION OF VARIOUS NIF AND NIF-ASSOCIATED GENE PRODUCTS ENCODED WITHIN THE AZOTOBACTER VINELANDII NIF GENE CLUSTER

by

Kevin Eugene Brigle

Dissertation submitted to the Faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Anaerobic Microbiology

APPROVED:

D. R. Dean, Chairman

R. C. Bates

J. S. Chen

R. E. Ebel

J. L. Johnson

May, 1989

Blacksburg, Virginia

STUDIES ON THE STRUCTURE AND FUNCTION OF VARIOUS NIF AND NIF-ASSOCIATED GENE PRODUCTS ENCODED WITHIN THE AZOTOBACTER VINELANDII NIF GENE CLUSTER

by

Kevin Eugene Brigle

Committee Chairman: Dennis R. Dean
Anaerobic Microbiology

(ABSTRACT)

The present study investigates the structural and functional roles of the metalloclusters present within the MoFe protein of nitrogenase from Azotobacter vinelandii. A gene replacement strategy was developed for oligonucleotide-directed mutagenesis of these proteins and the resulting biological and biochemical effects of these changes were examined. Identification of structurally important regions in the MoFe protein subunits and assignment of specific amino acid residues as potential metal cluster ligands were based upon several criteria: i. metallocluster extrusion requirements; ii. spectroscopic properties of the MoFe protein; iii. interspecies and intersubunit comparisons; iv. comparison of the MoFe protein subunit sequences to iron molybdenum cofactor biosynthetic gene products. This mutagenesis strategy has permitted the construction of thirty three mutant strains having specific amino acid substitutions within the MoFe protein subunits. Based on the diazotrophic growth characteristics and substrate reduction capabilities of these mutant strains, a model is presented in which potential metallocluster binding sites within the MoFe protein subunits are defined. In addition to analysis of the MoFe protein subunits,

this site-directed mutagenesis and gene replacement strategy can be used to place specific mutations into any gene product encoded within the A. vinelandii nif gene cluster.

Finally, nucleotide sequence analysis of the regions flanking the nifEN genes revealed the presence of three nif genes (nifT, nifY, and nifX) and four open reading frames (ORF1, ORF2, ORF3, and ORF4). Two of these genes, nifX and ORF3, were shown to be under nif control and synthesis of their products elevated in response to a demand for fixed nitrogen. Mutant strains with deletions in ORF3 appeared to accumulate an excess amount of MoFe protein when compared to wild type. The ORF3 gene product has been overproduced in E. coli. This provides an important step toward characterizing the protein and elucidating the molecular basis for its control of nifDK gene expression.

ACKNOWLEDGMENTS

I thank the following people whose guidance and friendship have made the past several years a time of personal and intellectual growth: My committee members Dr. Bates, Dr. Chen, Dr. Ebel, and Dr. Johnson whose time and efforts were most appreciated; Dr. Dennis (The Great Camel) Dean for seven years of patience, time and effort toward the development of my intellectual and personal growth; My collaborators, Dr. Bill Newton, Dr. Hal May, and , for their intellectual input and technical expertise; My fellow graduate students and lab mates, especially , , , and Dr. Marty Jacobson for their friendship both in and out of the lab; , for his patience and brute force computer efforts; My parents for their constant love and support; My new Virginia family - , , and - for their remarkable understanding and love.

Finally and foremost, my deepest love and thanks to my wife , for her never ending support, love, patience, and understanding which all too often went unrewarded.

TABLE OF CONTENTS

ABSTRACT ii

ACKNOWLEDGEMENTS iv

TABLE OF CONTENTS v

LIST OF ILLUSTRATIONS vii

LIST OF TABLES x

LITERATURE REVIEW 1

Introduction
Enzymatic Reaction
The nif Genes: Products and Functions
nifH: The Fe Protein
nifD and nifK: The MoFe Protein Subunits
nifM: Maturation of the Fe Protein
nifW and nifZ: Maturation of the MoFe Protein?
nifH, nifE, nifN, nifV, nifB, nifQ: Biosynthesis of FeMoco
nifF and nifJ: Electron Transport to Nitrogenase
nifU, nifS, nifT, nifY,and nifX: Unknown Functions
nifA and nifL: Regulation
Alternative Nitrogen Fixing Systems

RESEARCH GOALS 38

MATERIAL AND METHODS 40

E. coli: Strains and Growth
A. vinelandii: Strains and Growth
Determination of Whole Cell Nitrogenase Activity
β-galactosidase Assay
Oligonucleotide-Directed Mutagenesis
Overproduction of Proteins in E. coli
Two-Dimensional Gel Electrophoresis
Computer Analysis

SITE-DIRECTED MUTAGENESIS OF THE MoFe PROTEIN 48

Nomenclature
Construction of Mutant Strains
Rational for Amino Acid Replacements
Specific Amino Acid Replacements
Characterization of Mutants
A Model for Sites of Metallocluster Binding within Nitrogenase

GENETIC ANALYSIS OF REGIONS FLANKING THE nifE AND nifN GENES . . 90

Sequence Comparisons
Gene Expression
Mutagenesis of nifY, nifX, ORF3
Biochemical Analysis
Overexpression of the ORF3 gene product in E. coli
Summary

LITERATURE CITED 106

APPENDIX I 124

APPENDIX II 131

APPENDIX III 135

APPENDIX IV 146

APPENDIX V 153

APPENDIX VI 159

APPENDIX VII 162

APPENDIX VIII 165

APPENDIX IX 171

APPENDIX X 174

APPENDIX XI 176

VITA 207

LIST OF ILLUSTRATIONS

1. Transfer of electrons through nitrogenase. 5

2. Comparison of the organization of the major nif cluster from A. vinelandii with that of K. pneumoniae. 7

3. Proposed ligation of the [4Fe-4S] cluster in the Fe protein. 13

4. Model of the coupling of electron transport to nitrogenase with obligatory electron transport to oxygen. 28

5. Strategy for the construction of site-directed mutant strains. 50

6. Alignment of MoFe protein α-subunit sequences. 54

7. Alignment of MoFe protein β-subunit sequences. 56

8. Comparison of the predicted secondary structures of the MoFe protein α- and β-subunits. 58

9. Comparison of structurally similar domains within the α- and β-subunits of the MoFe protein. 61

10. Comparison of the MoFe protein α-subunit and the nifE-encoded polypeptide. 63

11. Comparison of the MoFe protein β-subunit and the nifK-encoded polypeptide. 64

12. Comparison of the predicted secondary structures of the MoFe protein α-subunit and the nifE-encoded polypeptide. 65

13. Comparison of the predicted secondary structures of the MoFe protein β-subunit and the nifN-encoded polypeptide. 66

14. Comparison of the hydrophobicity of the MoFe protein α-subunit and the nifE-encoded polypeptide. 67

15. Comparison of the hydrophobicity of the MoFe protein β-subunit and the nifN-encoded polypeptide. 68

16. Diazotrophic growth of wild type and mutant strains having substitutions for nonconserved cysteine residue **α-cys-45**. 73

17. Diazotrophic growth of wild type and mutant strains having substitutions for conserved cysteine residue **α-cys-183**. 74

18. Diazotrophic growth of wild type and mutant strains having substitutions for conserved cysteine residue **β-cys-153**. 76

19. Diazotrophic growth of wild type and mutant strains having substitutions for amino acid residues within α-subunit region III. 78

20. Diazotrophic growth of wild type and mutant strains having substitutions for highly conserved histidine residues present within region I of both subunits. 80

21. Diazotrophic growth of wild type and mutant strains having substitutions in conserved nitrogen-donor residues within α-subunit region V. 82

22. Diazotrophic growth of wild type and mutant strains having substitutions for amino acid residues within α-subunit region V near conserved cysteine residue **α-cys-275**. 84

23. Proposed models for the sites of metallocluster binding within nitrogenase. 86

24. Physical map of the A. vinelandii nif cluster surrounding nifX and construction of A. vinelandii strains used to analyze genes within this region. 95

25. Diazotrophic growth of wild type and mutant strains having deletions in nifY, nifX, and ORF3. 97

26. Normalized specific activity of MoFe protein and of Fe protein in wild type and mutant strain DJ44. 98

27. Physical map of the A. vinelandii nif cluster surrounding nifX and construction of plasmids used to overproduce the ORF3 gene product in E. coli. 101

28. Expression of the ORF3 gene product in E. coli JM105. 102

29. Potential secondary structure predicted to form in the nifK - nifT intercistronic region and potential secondary structure predicted to form after removal of the 40-bp EagI fragment. 155

30. Physical map of the A. vinelandii nif cluster surrounding nifT and construction of A. vinelandii strains used to analyze genes within this region. 156

31. Diazotrophic growth of wild type and a mutant strain which carries a deletion in the nifK - nifT intercistronic region. 157

32. Time course of β-galactosidase activity of strains DJ297 and DJ299 upon derepressing for nitrogenase activity. 158

33. Physical map of the A. vinelandii nif cluster surrounding the structural genes and construction of A. vinelandii nif:lacZ fusion strains used to analyze expression of these genes. 161

LIST OF TABLES

1. The nif gene products and their functions. 8

2. S_{AB} values calculated for nifH-deduced amino acid sequences. 10

3. S_{AB} values calculated for nifD- and nifK-deduced amino acid sequences. 17

4. Mutant strains used in this study. 41

5. Targeted amino acid residues and the rationale for their targeting. . . . 70

6. Growth characteristics and whole cell and component protein activities of wild type and site-directed mutant strains. 72

7. Features of the nifT, nifY, and nifX gene products of A. vinelandii and K. pneumoniae. 92

8. Growth characteristics and component protein activities of wild type and mutant strains having deletions in nifY, nifX, and ORF3. . . . 99

LITERATURE REVIEW

Introduction.

Nitrogen is an essential component of proteins, nucleic acids and other molecules which comprise the living cell. However, most of the earth's nitrogen exists in a chemically stable, molecular form (N_2) which cannot be utilized by most forms of life. Therefore, most living organisms must obtain nitrogen in some usable, fixed form such as nitrate, nitrite, ammonia, or other more complex compounds such as amino acids. These usable fixed forms of nitrogen are very scarce and are continually being recycled to N_2 by the combined microbial processes of nitrification and denitrification. Certain microorganisms are able to convert chemically stable N_2 to a usable fixed form. These organisms provide virtually all of the fixed nitrogen necessary for survival of the earth's ecosystem. In this bioconversion of nitrogen, which is termed biological nitrogen fixation, N_2 is reduced to ammonia (NH_3) via a very complex and poorly understood catalytic process.

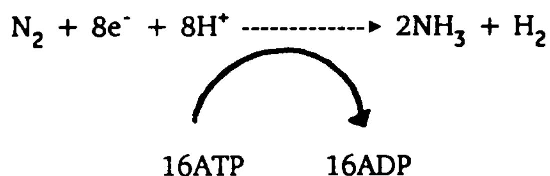
The first report of biological nitrogen fixation was presented on 20 September 1886, when Dr. Hermann Hellriegel and Dr. Hermann Wilfarth described their findings on the nitrogen nutrition of legumes before the 59th assembly of German scientists and physicians in Berlin. Their results, which contemporaries described as 'epoch-making', marked the beginning of research into the process of biological

nitrogen fixation (85). In the 102 years that have followed this initial discovery, a wide variety of prokaryotic organisms have been shown to be capable of fixing nitrogen: (A) the free-living aerobe Azotobacter vinelandii; (B) the free-living, facultative anaerobe Klebsiella pneumoniae; (C) the free-living obligate anaerobe Clostridium pasteurianum; (D) the photoautotrophic blue-green alga Rhodobacter capsulatus; (E) the fast- and the slow-growing legume symbionts Rhizobium meliloti and Bradyrhizobium japonicum, respectively; (F) the non-leguminous symbiont Azospirillum brasilense; (G) the mesophilic methanogenic archaebacterium Methanobacterium ivanovii; and (H) the thermophilic methanogenic archaebacterium Methanococcus thermolithotrophicus. This represents only a partial list of nitrogen fixing bacteria but it demonstrates the diverse physiologies and habitats of those organisms capable of this process. Research on the nitrogen fixing properties of azotobacters began in 1901 when Beijerinck reported that aerobic Azotobacter chroococcum and A. agilis could fix nitrogen (see 3 in reference 18). A. vinelandii was first isolated in 1903 from soil samples in Vineland, New Jersey. These are Gram-negative to Gram-variable, ovoid cells which are motile by peritrichous flagella. In a century of research on biological nitrogen fixation, it is only within the last decade that alternative nitrogen fixing systems have been demonstrated. These alternative vanadium-containing and the iron-dependent enzyme systems will be addressed, but unless otherwise specified 'nitrogenase' will refer to the molybdenum-dependent nitrogen fixing complex.

Enzymatic Reaction.

Minimally, biological nitrogen fixation requires the enzyme nitrogenase, MgATP, a strongly reducing electron donor, protons, and an anaerobic environment. Nitrogenase is composed of two separately purified component proteins called the Fe protein (Component II) and the MoFe protein (Component I). These proteins, especially the Fe protein, are rapidly and irreversibly inactivated by oxygen (14). The native Fe protein is a 63 kD dimer of two identical subunits believed to be symmetrically bridged by a single redox active [4Fe-4S] cluster (54). The Fe protein contains two sites for ATP binding and is the obligate one electron donor to the MoFe protein during substrate reduction. The MoFe protein (native $M_r \approx 230{,}000$) is a $2\alpha2\beta$ tetramer of two non-identical subunits and contains the site(s) of substrate binding and reduction. The MoFe protein contains two atoms of Mo and ≈ 32 Fe and ≈ 32 acid-labile sulfur atoms. While these atoms can be extracted from the protein in the form of six metal containing prosthetic groups of two distinct types (four 'P-clusters' and two 'FeMo cofactors'), the actual structures and stoichiometry of these metal clusters within the native MoFe protein are unknown.

The enzymatic reaction catalyzed by nitrogenase is as follows:



$$N_2 + 8e^- + 8H^+ \longrightarrow 2NH_3 + H_2$$
$$16ATP \curvearrowright 16ADP$$

The evolution of one molecule of H_2 appears to be an obligate wasteful side

reaction that is coupled to N_2 reduction (15, 73, 156). Many nitrogen fixing bacteria can recapture some of the energy wasted in hydrogen production by recycling the evolved hydrogen via an uptake hydrogenase (127, 161). The flow of the electrons through nitrogenase during turnover is shown in Fig. 1. The physiological source of reducing equivalents necessary for nitrogenase reduction has been identified only for Klebsiella pneumoniae. In this organism, pyruvate:flavodoxin oxidoreductase couples the oxidation of pyruvate to the reduction of flavodoxin yielding acetyl-SCoA and CO_2 (153). The flavodoxin, functioning between the semi-quinone and hydroquinone states, transfers one electron to the Fe protein. The reduced Fe protein then associates with the MoFe protein and, in a reaction coupled to the hydrolysis of two MgATP, transfers a single electron to the MoFe protein (47). In what is considered to be the rate limiting step in the overall reaction (170), the oxidized Fe protein must dissociate from the MoFe protein before the Fe protein can be reduced again (47). Substrate reduction occurs at a site(s) on the MoFe protein and, as seen from the above equation, reduction of N_2 to NH_3 requires that this sequence of reactions occurs eight times. The most comprehensive kinetic model which describes the mechanism for nitrogenase function has been presented in a series of papers by Thornley and Lowe (90, 91, 171, 172). Their model, which is based on experimental observation and computer modeling, is extremely complex and defines seventeen rate constants and forty-six intermediates. In addition to N_2 and protons, nitrogenase is also capable of reducing other triple-bonded substrates such as acetylene, alkyl cyanides,

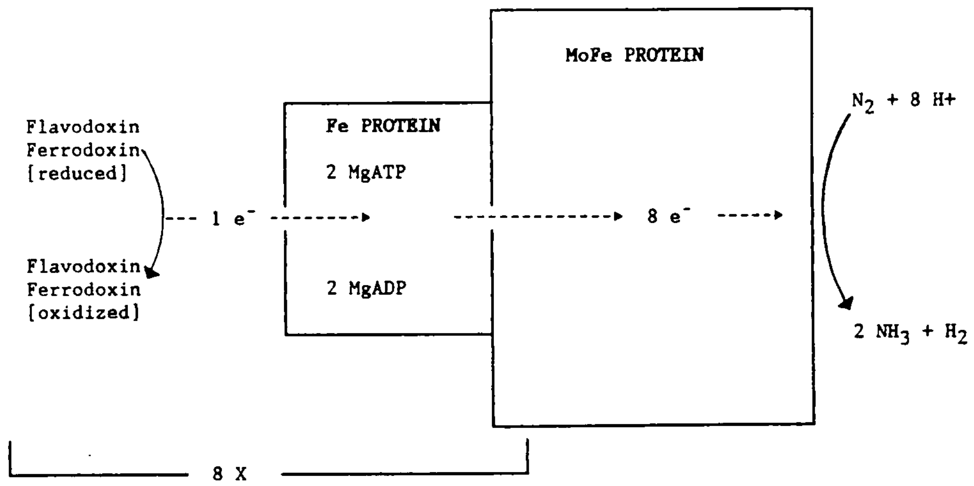


FIGURE 1. Transfer of electrons through nitrogenase. (Adapted from Haaker and Klugkist87).

allene, azide, cyclopropene, diazirine, and hydrogen cyanide (106). Due to the extreme sensitivity of ethylene detection, the reduction of acetylene to ethylene is the most widely used assay for in vivo and in vitro nitrogenase activity. However, most alternative substrates are used as probes for the number and nature of the sites of substrate interaction on nitrogenase, and as probes for the types of intermediates that might be formed and electron pathways that might be used during N_2 reduction.

The nif Genes: Products and Functions.

By convention, the nitrogen fixation (nif) genes of all organisms are designated on the basis of their structural or functional homology to the twenty nif genes identified in the K. pneumoniae nif cluster (1). In this model organism, the twenty nif genes are clustered within a single region of the chromosome and are organized into eight transcriptional units (Fig. 2). In A. vinelandii, homologs to each of these nif genes except nifJ have been identified (Appendix III). As in K. pneumoniae, these genes are clustered within a single region of the chromosome and they exhibit the same relative sequential arrangement (Fig. 2). In contrast to K. pneumoniae, there are a number of additional open reading frames which may be under nif regulation interspersed throughout the A. vinelandii nif gene cluster. The three catalytic protein components of nitrogenase are encoded by the structural genes nifHDK. As shown in Table 1, the remaining nif gene products are involved either in regulation, metallocluster processing, electron transport, or maturation of an

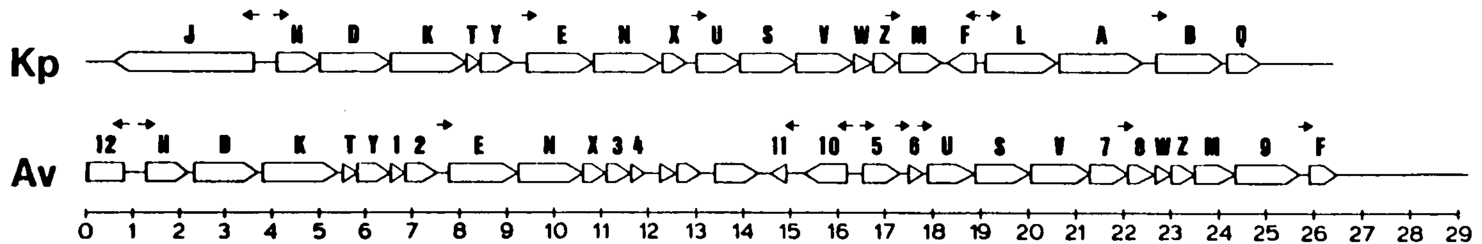


FIGURE 2. Comparison of the organization of the major nif cluster from A. vinelandii (lower portion) with that of K. pneumoniae (upper portion). Arrows indicate the approximate position and direction of identified or proposed nif-specific promoters in each organism. A. vinelandii genes whose deduced products have sequence identity when compared with K. pneumoniae nif-specific gene products have the appropriate nif genotypic designations. Open reading frames (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 upstream activator sequence are also numbered. ORFs apparently not cotranscribed with any of the identified nif-specific genes and not preceded by a nif-specific promoter sequence are not labeled. Numbering refers to the distance in kilobase pairs from the first nucleotide in the sequence given in Appendix XI. (Adapted from Appendix III - Fig. 1).

TABLE 1. Proposed functions of the *nif* gene products.

nif GENE	FUNCTION OF GENE PRODUCT
H	Fe protein subunit
	FeMoco biosynthesis
D	α-subunit of MoFe protein
K	β-subunit of MoFe protein
M	Maturation of Fe protein
W	Maturation/stability of MoFe protein?
Z	Maturation/stability of MoFe protein?
E	FeMoco biosynthesis
N	FeMoco biosynthesis
B	FeMoco biosynthesis
Q	FeMoco biosynthesis
V	FeMoco biosynthesis (homocitrate synthase)
F	Electron transport to nitrogenase (flavodoxin)
J	Electron transport to nitrogenase (pyruvate:flavodoxin oxidoreductase)
A	Positive activator of *nif* transcription
L	Negative regulator of NifA function
U	Unknown
S	Unknown
T	Unknown
Y	Unknown
X	Unknown

active nitrogenase enzyme. (The 28,793 bp nucleotide sequence A. vinelandii nif cluster, and the deduced amino acid sequences of all nif genes and ORF's encoded in this sequence are given Appendix XI. Unless otherwise specified, all numbering of amino acid residues discussed below will refer to the A. vinelandii nif gene products.)

The Nitrogenase Structural Genes: nifH, nifD, and nifK.

nifH: The Fe Protein.

The complete amino acid sequence of the Fe protein (nifH) has been deduced from the nucleotide sequence from thirteen eubacteria and three methanogenic archaebacteria: Anabaena 7120 (104), A. vinelandii (11), A. chroococcum (140), Bradyrhizobium (Br.) japonicum (37), Clostridium pasteurianum (19, 181), K. pneumoniae (147, 165), Rhizobium meliloti (174), Rhizobium phaseoli (129), Rhizobium sp. Parasponia (149), Rhizobium ORS571 (118), Rhizobium trifolii (148), Rhodobacter capsulatus (78, 146), Thiobacillus ferrooxidans (126), Methanobacterium (Mb.) ivanovii (159), Methanococcus (Mc.) thermolithotrophicus (159), and Mc. voltae (160). With the exception of R. capsulatus and T. ferrooxidans, Souillard et al. (159) calculated the similarity coefficients (S_{AB} values) between each of these sequenced nifH gene products. The comparisons reveal strong sequence conservation among all of the Fe proteins, and S_{AB} values between specific pairs of Fe proteins which are consistent with the phylogenetic positions of the different organisms (Table 2). The Fe proteins of A. chroococcum and of Mc.

TABLE 2. S_{AB} values calculated for nifH-deduced amimo acid sequences of different bacteria: Mb. ivanovii (Mbi), Mc. thermolithotrophicus (Mct), Mc. voltae (Mcv), C. pasteurianum (Cp), A. vinelandii (Av), A. chroococcum (Ac), K. pneumoniae (Kp), Anabaena 7120 (An), Br. japonicum (Rj), Rhizobium sp. Parasponia (RP), Rhizobium ORS571 (Rs), R. meliloti (Rm), R. trifolii (Rt). (See text for references.) (From Souillard et al. 1988.)

	Mcv	Mbi	Cp1	Cp2	Av	Ac	Kp	An	Rj	RP	Rs	Rm	Rt	Rp
Mct	55	50	50	51	48	47	48	46	46	46	45	48	46	48
Mcv		50	52	54	50	50	51	50	49	49	49	49	49	49
Mbi			53	54	50	51	48	46	48	48	48	47	50	48
Cp1				92	66	65	64	61	61	61	61	62	61	62
Cp2					66	65	65	61	61	63	61	63	63	63
Av						89	88	71	75	74	73	68	69	70
Ac							85	72	74	71	71	71	73	73
Kp								70	74	74	75	67	68	70
An									70	70	72	69	70	71
Rj										97	92	77	78	79
RP											93	76	77	79
Rs 1 and 2												79	77	80
Rm													91	93
Rt														90

thermolithotrophicus have the highest (89%) and lowest (48%) sequence identities respectively with the A. vinelandii Fe protein. Examination of the aligned nifH gene products reveals features common to either all or most of the polypeptides (19, 52, 53, 137, 148, 160, 159): (1) presence of five conserved cysteine residues, with the exception of the last cysteine residue in both Mc. thermolithotrophicus and Mc. voltae; (2) strong sequence conservation in the regions surrounding the five conserved cysteine residues; (3) conservation of an arginine at residue 100; (4) absence of a tryptophan residue, with the exception of R. trifolii; (5) presence of a consensus ATP-binding site. Conservation of these features among the nifH gene products is likely due to the many interactions required of the Fe protein (MoFe protein, electron donor proteins, ATP) and the different reactions in which the Fe protein participates (reduction of the MoFe protein, FeMo-cofactor biosynthesis).

The presence of additional nifH or nifH-like sequences has been reported in several diazotrophs (118, 129, 140, 181). In A. vinelandii, there are two additional copies of the nifH gene (72, 77, 130). Transcripts of these genes are observed only under molybdenum starved, nitrogen fixing conditions and therefore, these additional nifH-coding sequences are believed to be involved in the alternative nitrogen fixing systems (9). This is also true for the additional copy of nifH found in A. chroococcum (138, 140). In C. pasteurianum, four additional nifH-like sequences are transcribed under nitrogen fixing conditions, but it is not known if any of the mRNAs are translated or if any of the reiterated sequences are actually essential for diazotrophic growth (181). In both Rhizobium ORS571 and R.

phaseoli, all of the reiterated nifH genes which are present are simultaneously transcribed under nitrogen fixing conditions. In both organisms, mutagenesis studies indicate that each of the reiterated genes contributes to total nitrogenase activity (118, 111).

Based upon studies in which iodoacetic acid was used to label exposed cysteinyl residues in the A. vinelandii Fe protein, Hausinger and Howard (54) proposed a model in which conserved cysteine residues 97 and 132 act as ligands to bind a single [4Fe-4S] cluster symmetrically between the two identical subunits (Fig.3). The validity of this model has been supported by studies in which serines have been substituted for the individual conserved cysteine residues in the Fe protein (65). In order to function in substrate reduction, the Fe protein must be activated in some unknown manner by the product of the nifM gene (133, 67, 123). This modification is not required for the Fe protein to carry out its unknown functional role in FeMo-cofactor biosynthesis (133). For the Fe protein to carry out both of these roles however, Howard et al. (65) has suggested that it must contain a properly ligated (cysteine residues 97 and 132) Fe-S cluster.

In response to certain environmental stimuli, a variety of diazotrophic bacteria (40, 45, 51, 81, 113) exhibit a rapid and generally reversible short term regulation of nitrogenase activity called 'switch off' regulation (189). The most well characterized 'switch off' system is that found in Rhodospirillum (Rs.) rubrum where nitrogenase activity is regulated by a reversible ADP-ribosylation of residue arginine-100 on a single subunit of the dimeric Fe protein (124). In Rs. rubrum, addition

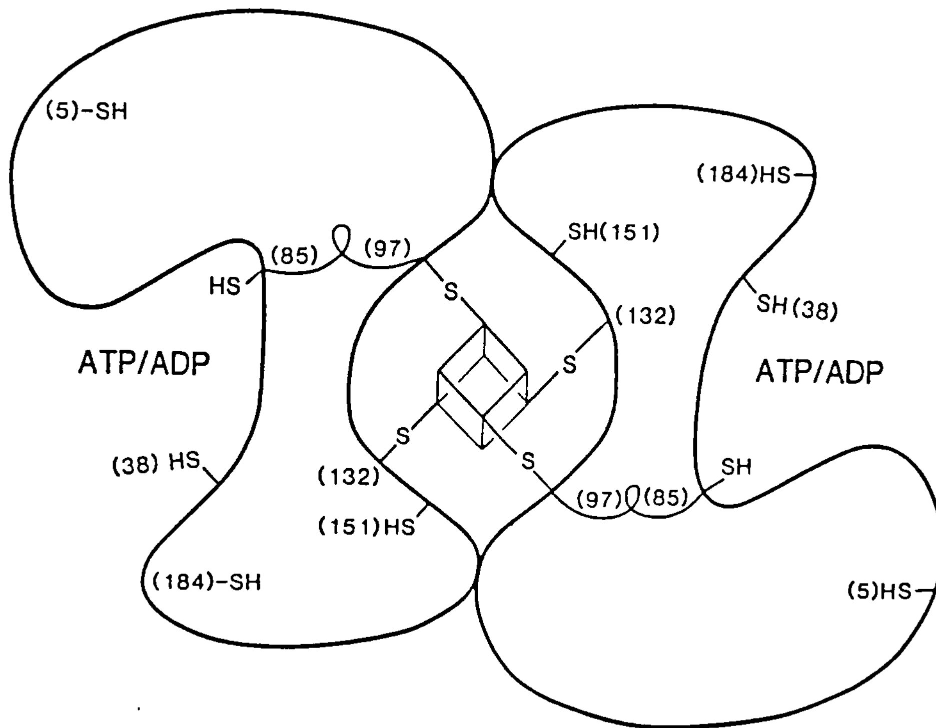


FIGURE 3. Proposed ligation of the [4Fe-4S] cluster in the Fe protein. (From Hausinger and Howard, 1983).

(inactivation of the Fe protein) and removal (activation of the Fe protein) of the ADP-ribosyl moiety is catalyzed by a transferase and a glycohydrolase, respectively (94, 142). Fe protein isolated from A. vinelandii and C. pasteurianum are both in vitro substrates for the ADP-ribosyl transferase isolated from Rs. rubrum (93, 95). Fe protein from C. pasteurianum forms a tight, catalytically incompetent complex with A. vinelandii MoFe protein (31, 32, 33). Once formed, the C. pasteurianum Fe protein in the complex no longer functions as a substrate for the ADP-ribosyl transferase. Additionally, formation of the tight complex is prevented if the C. pasteurianum Fe protein is first ribosylated. These results have led Ludden and his coworkers (124, 114) to suggest that in Rs. rubrum, the arginine-100 residue is present at a component protein interaction site and that ribosylation of this residue acts to prevent a functional Fe protein/MoFe protein association. They further suggest that while the presence of an arginine at residue 100 is not essential for Fe protein/MoFe protein interaction, it plays a critical role in allowing the MgATP-dependent transition of the Fe protein to the fully open conformation that is required for electron transfer to the MoFe protein (92). It is possible that the role(s) of the arginine-100 residue is not the same in all diazotrophs. For example, while Fe protein from A. vinelandii, K. pneumoniae, and C. pasteurianum are all in vitro substrates for ADP-ribosyl transferase from Rs. rubrum, only with A. vinelandii does the treatment have no effect on nitrogenase activity (92, 95, 114). Additionally, although there is in vivo evidence for 'switch off' regulation in A.

vinelandii (25, 81, 86), it likely involves noncovalent mechanisms quite different from ADP-ribosylation (44, 86).

NifD and nifK: The MoFe Protein Subunits.

The complete amino acid sequence of the α-subunit (nifD) has been deduced from the nucleotide sequence from eight organisms: Anabaena 7120 (39, 89), A. vinelandii (11), Br. japonicum (79), C. pasteurianum (179), K. pneumoniae (1, 69, 147), Rhizobium sp. Parasponia (183), cowpea Rhizobium strain IRc78 (186), and Rhodobacter capsulatus (146). The complete amino acid sequence of the β-subunit (nifK) has been deduced from the nucleotide sequence from six organisms: Anabaena 7120 (99), A. vinelandii (11), Br. japonicum (169), C. pasteurianum (180), K. pneumoniae (1, 59, 162), Rhizobium sp. Parasponia (183). A comparison of all nifD-encoded sequences, with the exception of R. capsulatus, has been presented by Ioannidis and Buck (69). The R. capsulatus nifD-encoded sequence has been omitted from all published α-subunit sequence comparisons probably because of the questionable validity of the sequence data. The R. capsulatus nifH nucleotide sequence, determined in the same study as the nifD sequence, was later found to have a significant number of errors (78). Wang et al. (180) have presented a detailed comparison of all β-subunit sequences with the exception of the K. pneumoniae sequence. Modified and updated versions of the α-subunit (with exception of R. capsulatus) and β-subunit sequence alignments will be given in the body of this work (Figs. 6 and 7, respectively). An additional alignment of the α-

subunit sequences which includes the questionable R. capsulatus sequence is given in Appendix X. Based upon the subunit alignments in Figs. 6 and 7, S_{AB} values (36) have been calculated for each of the MoFe protein subunits and are presented in Table 3. With the exception of C. pasteurianum, the overall sequence identity is quite high among the respective MoFe protein subunits. The S_{AB} values calculated for pairings with A. vinelandii correlate with the level of activity observed for heterologous reconstitutions of the A. vinelandii MoFe protein with the Fe proteins from these other species (32). In addition to sequence comparisons, a number of groups have done secondary structure analyses for several of the sequenced α- and β-subunits (69, 79, 89, 180, 183). From these different types of comparisons, several features common to all of the MoFe protein subunits have emerged. (1) Within each subunit the primary sequence, predicted secondary structure, and charge distribution (hydrophobic index) is highly conserved between the different species. These features are least conserved within the C. pasteurianum MoFe protein subunits, in accordance with the unusual biochemical properties of the nitrogenase enzyme isolated from this organism. (2) There are five conserved cysteine residues among the α-subunits (62, 88, 154, 183, & 275), three among the β-subunits (70, 95, & 153), and each is present within a highly conserved region of the respective protein. (3) All of the conserved cysteine residues are located at or within a predicted β-turn suggesting a near surface location in the native MoFe protein. This further suggests that the cysteine residues would be readily accessible as potential ligands for the various metal clusters. (4) Only α-subunit cysteine-154 is associated

TABLE 3. SAB values calculated for nifD-deduced (upper panel) and nifK-deduced (lower panel) amino acid sequences of different bacteria: Anabaena 7120 (An), A. vinelandii (Av), Br. japonicum (Bj), C. pasteurianum (Cp), cowpea Rhizobium strain IRc78 (cR), K. pneumoniae (Kp), Rhizobium sp. Parasponia (pR). (See text for references.) Calculations for the α- and β-subunit comparisons based upon alignments in Figs. 6 and 7, respectively.

α-Subunit SAB

	Av	Kp	pR	cR	Bj	An
Kp	72					
pR	72	69				
cR	71	68	93			
Bj	70	68	92	91		
An	68	67	70	70	69	
Cp	40	39	39	40	39	39

β-Subunit SAB

	Av	Kp	An	Bj	pR
Kp	67				
An	54	51			
Bj	53	51	60		
pR	51	49	57	87	
Cp	37	35	41	38	36

with an amino acid sequence similar to conserved sequences found among certain bacterial ferredoxins (12, 96). (5) There is significant spatial conservation and limited sequence identity between the α- and β-subunits within the regions surrounding the first three conserved cysteine residues of each subunit. However, these regions within the two subunits are predicted to have very different secondary structures and it is unclear if these spatial and sequential similarities actually translate into a functional similarity. Based on X-ray crystallographic analysis of the C. pasteurianum MoFe protein, Yamane et al. (185) have predicted low resolution identity between the subunits at the three-dimensional level. On the basis of sequence comparisons, Robson (137) has identified a possible low affinity ATP-binding sequence on the β-subunit. However, results of ATP-binding studies have been contradictory: Miller and co-workers have been able to detect ATP-binding to the MoFe protein of both K. pneumoniae and A. chroococcum (107, 108). However, Cordewener and co-workers (22, 23) have never observed binding of ATP to the MoFe protein of A. vinelandii and argue that it does not occur. While it seems unlikely that the MoFe proteins from these organisms would differ with respect to the binding of ATP, it is unclear as to which group is correct.

During the past decade, the results of cluster extrusion studies and the techniques of EPR, EXAFS, MCD, and Mossbauer spectroscopy have provided the following picture of the metallocluster structure MoFe protein: The protein is $\alpha_2\beta_2$ tetramer that contains two atoms of Mo and $\approx$ 32 atoms of Fe which can be extracted from the protein in the form of two cluster types referred to as M-centers

(FeMoco) and the P-clusters. Initial categorization of these clusters was provided by Mossbauer and EPR studies in which the ≈ 32 Fe atoms were found to be present in four spectroscopically distinct classes that were labeled 'D', 'Fe^{2+}', 'M', and 'S' (112). The labels 'D' and 'S' were given arbitrarily while the 'Fe^{2+}' label was chosen because this Fe species has a Mossbauer feature characteristic of a high-spin ferrous ion. The label 'M' was given to reflect that this Fe species is associated with a magnetic center detectable by EPR spectroscopy in the dithionite-reduced protein.

The 'S' iron species was believed to represent a single redox-invariant Fe atom and, because of the extreme complexity of the MoFe protein Mossbauer spectrum, it resisted further characterization until recently (discussed below). The 'D' and 'Fe^{2+}' iron species were found to be in an approximate stoichiometry of 3:1 (112). Based upon the changes in the Mossbauer and EPR spectra of the MoFe protein during thionine oxidation, Zimmermann et al. (187) concluded that these two iron species were arranged as four [4Fe-4S] cubane structures of composition [D_3Fe^{2+}-4S]. These were labeled 'P-clusters' because they were believed to be tightly bound to the protein. Compared to the Mossbauer spectra of other proteins which contain conventional [4Fe4S] clusters, the P-clusters exhibit very unusual spectroscopic features. The possibility that the usual thiol ligands where replaced by other nucleophiles or the possibility of a five-coordinate ligation sphere were suggested to explain these unique spectral properties. The 4Fe-cubane structure predicted for the P-clusters was corroborated in the following year by the extrusion of four [4Fe-4S] clusters from the MoFe protein (87). In this procedure, which is conducted

anaerobically, the protein is denatured in the presence of thiol compounds that trap the Fe-S centers in the form of easily identified and quantitated synthetic analogues. This extrusion method is only suggestive of [4Fe-4S] clusters and does not preclude the possibility that they result from the breakdown of larger clusters upon their release from the protein matrix. Additionally, although suggestive of thiol coordination, this extrusion method does not directly identify the terminal ligands to the P-clusters within the protein. Recently, two independent groups have presented new spectral data on the MoFe protein and both have proposed new models for the P-clusters which are different than that just described. McLean and co-workers (101) have been able to resolve the complex Mossbauer spectrum by utilizing MoFe protein with ^{57}Fe selectively enriched P-clusters and by performing the spectroscopic analysis at a lower temperature. Their data suggests that the D:Fe^{2+}:S ratios within the MoFe protein are 10:4:2 and that the 'S' iron species is actually a subset of the P-clusters. They propose that the MoFe protein contains two pairs of slightly inequivalent P-clusters: two of the [D_3Fe^{2+}-4S] type and two of the [D_2SFe^{2+}-4S] type. Alternatively, a detailed analysis of EPR data from thionine oxidized MoFe protein have led Hagen and co-workers (48) to suggest two other possibilities: (1) the MoFe protein contains only two P-clusters which are actually larger, eight Fe-containing entities uniquely able to transfer more than one electron; (2) the MoFe protein contains two classes of [4Fe-4S] P-clusters that behave very differently during enzyme turnover. Unlike the model put forth by McLean and co-workers, neither of the models suggested by Hagen's group offer any further characterization

or explanation of the 'S' iron species.

The magnetic, EPR-active center present in the MoFe protein was subsequently shown to be a novel iron- and molybdenum-containing cofactor called FeMoco, M-centers, or M-clusters (131). In 1977, Shah and Brill (151) successfully isolated FeMoco from the MoFe protein of several diazotrophs. In this procedure, the protein is denatured with citric acid, neutralized to its isoelectric point with phosphate, and the FeMoco extracted into an organic solvent, typically N-methylformamide (NMF). This isolation procedure, which must be carried out anaerobically, destroys the P-clusters while extracting $\approx$ 50% of the Fe content of the MoFe protein in the form of two intact iron-molybdenum cofactors per α_2 β_2 tetramer (151, 131). FeMoco isolated in this way is able to activate extracts from nifB mutant strains of A. vinelandii (115) and K. pneumoniae (163) which are deficient in FeMoco biosynthesis (151). FeMoco is generally considered to have a composition of $Mo:Fe_{5-8}:S_{8-9}$ (57) and an apparent molecular weight of less than 800 (16). While the structure of FeMoco is unknown, X-ray absorption spectroscopy and oxidative decomposition studies suggest that it contains a 'MoS_3' core (116, 21). Isolated FeMoco has an EPR spectrum very similar to the signal observed for the native MoFe protein except that it is somewhat broadened and the g values are slightly shifted (131). These altered spectral features are interpreted as reflecting differences in the ligand environment inside and outside of the protein. Addition of thiophenol to FeMoco sharpens the EPR signal to one more closely resembling the spectrum of the native protein (17). The sharpened signal is fully developed at a

ratio of one thiophenol/molybdenum and implies the presence of a cysteinyl-thiol ligand to FeMoco within the MoFe protein. Results of electron spin echo [ESE] (168) and Fourier-transform infrared red [FT-IR] spectroscopy (178) studies on isolated FeMoco suggest a nitrogen ligand to the cofactor within the native protein. Based upon all of the above observations, cysteinyl-sulfurs, amides, and even a deprotonated peptide backbone have been suggested as possible FeMoco ligands (11, 166, 168, 178).

While there is no direct evidence linking FeMoco to the substrate-binding site, indirect evidence has come from a study of nifV mutants which produce a MoFe protein that exhibits substrate reduction properties very different from the wild type protein (55). When FeMoco isolated from these mutants is added to extracts from nifB mutants, the reconstituted MoFe protein has the substrate reduction properties of the nifV mutant strain. If wild type FeMoco is added to the nifB extracts, a MoFe protein with wild type substrate reduction properties is formed. These data strongly suggest that FeMoco is, includes, or forms part of the enzyme's substrate-binding or substrate-reducing site. Electron microscopy studies have indicated that the four MoFe protein subunits are arranged as a distorted tetrahedron and that the Fe protein is located in a cavity within the central part of the MoFe protein (176). Low resolution (8 A) X-ray data suggests that all of the metal clusters within the MoFe protein are associated with the α-subunit. Additionally, the two FeMoco clusters appear to be situated at the edges of the α-subunits separated by $\approx$ 7 A.

These data suggest that N_2 could be bound between the FeMoco clusters during the enzymatic reaction (see 32 in reference 157).

Synthesis of an Active Nitrogenase Enzyme: Maturation of the Component Proteins (nifM, nifW, nifZ) and FeMoco Biosynthesis (nifH, nifE, nifN, nifV, nifB, nifQ).

NifM: Maturation of the Fe Protein.

Mutant strains defective in nifM produce an Fe protein that is incapable of participating in substrate reduction but which is fully competent in FeMoco biosynthesis (133). Howard et al. (67) demonstrated that a catalytically active Fe protein could be synthesized in E. coli by transforming with plasmids encoding nifH, nifM, and nifZ. In a similar study, Paul and Merrick (123) unambiguously demonstrated that synthesis of active Fe protein in E. coli requires only the nif-specific products of nifH and nifM. The precise role of the nifM gene product is unclear and has been proposed to involve either assembly or insertion of the Fe-S cluster (67, 121). Conversely, the results of site-directed mutagenesis studies suggest that the role of the nifM gene product may be to modify the existing cluster within the Fe protein (65).

NifW and nifZ: Maturation of the MoFe Protein?

K. pneumoniae mutant strains defective in nifW or nifZ derepress more slowly than wild type, with nifZ mutant strains derepressing the slowest. Neither gene is essential for diazotrophic growth. Paul and Merrick (123) have suggested that the

nifW and nifZ gene products are involved in either maturation or stability of the MoFe protein or its metalloclusters, having a function analogous to the nifM gene product's role in Fe protein maturation.

nifH, nifE, nifN, nifV, nifB, and nifQ: Biosynthesis of FeMoco.

Biosynthesis of FeMoco is biochemically and genetically complex, requiring the products of at least six nif genes in K. pneumoniae. The nifQ gene product is required in the early steps of FeMoco biosynthesis and only under conditions where the availability of Mo is limited (68). When nifQ mutant strains are cultured under conditions of high Mo concentration, the effects of nifQ mutations are suppressed. These mutant strains are not deficient in Mo transport implying that nifQ does not code for a permease. Rather, the function of this gene product might be to convert Mo entering the cell into a form that can be utilized in the next step of FeMoco biosynthesis. The presence of several conserved cysteine residues in the carboxy terminus of this protein supports the idea that it may have a metal-binding function (75).

Mutants defective in nifE, nifN, or nifB produce an apo-MoFe protein that can be activated in vitro by addition of purified FeMoco (56, 133, Appendix I). The gene products of nifE and nifN share significant sequence identity to the nifD and nifK gene products, respectively (24, Appendix I). Similarities include: (1) Four of the five conserved cysteine residues within the α-subunit are conserved within the nifE gene product; (2) One of the three conserved cysteine residues within the β-

subunit is conserved within the nifN gene product; (3) The α- and β-subunits have a similar size and charge to the nifE- and nifN-encoded gene products, respectively. These observations led Dean and co-workers (24, 134, Appendix I) to suggest that the nifE and nifN gene products form a protein complex upon which FeMoco is preassembled and then donated to the apo-MoFe protein. This hypothesis is supported by Shah and co-workers (152) who have purified the nifEN protein complex and shown it to be a $\alpha_2\beta_2$ tetramer similar in structure to the MoFe protein. The function of the nifB gene product is unknown. The sequence of this protein has been deduced from the nucleotide sequence from six organisms (13, 75, 98, 119, 141) and it is characterized by the presence a unique arrangement of conserved cysteine residues thought to be involved in metal binding (13).

Recently, the nifH gene product (Fe protein) was shown to be required for FeMoco biosynthesis (35, 135). To be competent in the FeMoco biosynthetic pathway, the Fe protein need not be functional with respect to substrate reduction; nifM mutants and certain nifH mutants readily produce an active cofactor (35, 133). The role of the Fe protein in FeMoco biosynthesis may be to bind MgATP or to transfer electrons in a reaction analogous to its role in nitrogenase turnover (35, 135). Alternatively, the Fe protein may be required to only physically associate with nifEN protein complex in order to affect a conformational change.

Mutations in the nifV gene result in the formation of a MoFe protein with substrate reduction properties different from the wild type protein: (1) While this enzyme effectively reduces acetylene, it reduces N_2 at a much lower rate than the

wild type enzyme; (2) Unlike the wild type enzyme, H_2 evolution from this enzyme is strongly inhibited by CO (100, 102). Hoover et al. (63) observed that homocitric acid (V factor) accumulates only in the medium of $NifV^+$ strains of K. pneumoniae derepressed for nitrogenase. They proposed that nifV encodes a homocitrate synthase (62) and they have shown that addition of homocitrate to the medium of K. pneumoniae $NifV^-$ mutants cures the phenotype. Hoover and co-workers have also shown that citrate can replace homocitrate in an in vitro FeMoco biosynthesis system. The aberrant FeMoco that results from this synthesis exhibits altered substrate reduction properties which has led these workers to suggest that in vivo, $NifV^-$ mutants utilize citrate for FeMoco biosynthesis (60). Recently, Hoover et al (61) demonstrated that homocitrate can be released from oxidized FeMoco and that it is present in a 1:1 ratio with molybdenum. Mossbauer, EXAFS, and EPR spectroscopic analyses have been unable to detect any differences between the wild type and $nifV^-$ enzymes (152).

nifF and nifJ: Electron Transport to Nitrogenase.

K. pneumoniae is the only organism for which the electron transfer pathway to nitrogenase is genetically proven. Two gene products, a flavodoxin [nifF] (27, 58, 117) and a pyruvate:flavodoxin oxidoreductase [nifJ] (58, 153) constitute a nif-specific, nif-regulated electron transport system from pyruvate to nitrogenase (153). Strains with mutations in either nifF or nifJ have no in vivo nitrogenase activity, but extracts from these strains are highly active in vitro when dithionite is used as an

artificial electron donor (133, 117). Both gene products were purified on the basis of their ability to restore activity to crude extracts of the respective mutant strains using pyruvate as a specific electron donor (58, 153) Spectral changes have demonstrated that the pyruvate:flavodoxin oxidoreductase reduces flavodoxin to the hydroquinone (2e$^-$-reduced) whereby it is then oxidized to the semiquinone (1e$^-$-reduced) by transferring an electron to the Fe protein (153).

Not surprisingly, electron transport to nitrogenase in the obligate aerobe A. vinelandii appears to be very different from the obligately anaerobic system present in K. pneumoniae. No gene analogous to K. pneumoniae nifJ has been found in A. vinelandii. Also, unlike the K. pneumoniae nifF gene product, the A. vinelandii nifF-encoded flavodoxin is neither absolutely nif-specific nor essential for nitrogen fixation (5). While only this single species of flavodoxin has been found in A. vinelandii OP, Klugkist et al. (83) have reported finding three different flavodoxins (one of which is nif-specific) in A. vinelandii ATCC 478.

Based on experimental observations, Haaker and Klugkist (44, 42) proposed a model in which electron transport to nitrogenase in A. vinelandii is coupled with electron transport to oxygen (Fig. 4). In this model, a NADPH dehydrogenase present in the cytoplasmic membrane (46) binds NADPH which then specifically donates both of its electrons to two different redox centers. One electron is transferred to a redox center operating at -160 mv, which in turn donates its electron to a respiratory chain component with oxygen being the ultimate acceptor. The other electron can theoretically be transferred from NADPH to a redox center

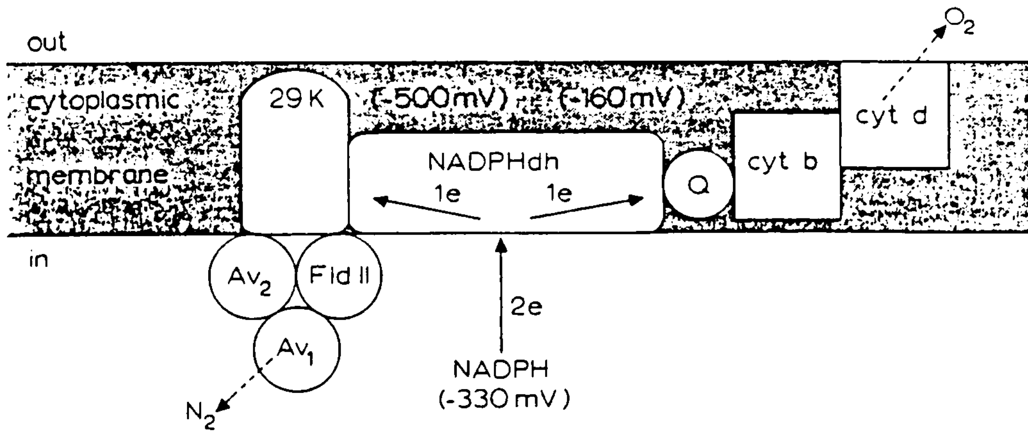


FIGURE 4. Model of the coupling of electron transport to nitrogenase with obligatory electron transport to dioxygen. Av1, MoFe protein; Av2, Fe protein; FldII, flavodoxin; NADPHdh, NADPH dehydrogenase; Q, coenzyme Q, cyt b and cyt d, cytochromes b and d, respectively; 29K, a 29 kD integral membrane protein. (From Haaker and Klugkist, 1987.)

operating at a potential of -500 mV. Two nif-specific, membrane-bound proteins (82) are thought to be involved in this latter transfer or in the subsequent transfer of the electron to a flavodoxin or ferredoxin and ultimately to N_2. This model can account for several observations. (1) It supports the observation that an energized cytoplasmic membrane is obligatory for electron transport to nitrogenase (43). (2) It supports the observation that in vivo, a structural association may exist between nitrogenase and the cytoplasmic membrane (66). (3) It explains the direct relationship between respiration and nitrogenase activity, since for each electron used in nitrogen fixation at least one electron has to be transferred through the respiratory chain to oxygen (82). (4) It offers an explanation for oxygen inhibition of in vivo nitrogenase activity. At very high oxygen concentrations all components of the respiratory chain and both redox centers within NADPH dehydrogenase are oxidized. All of the electrons from NADPH are now channeled away from flavodoxin and into the respiratory chain and ultimately to oxygen. Under this extreme condition, an oxygen tolerant, three component nitrogenase complex (136, 145, 182) can be formed and maintained. This redox-dependent, protective mechanism is termed 'conformational protection' (125). During conformational protection, the oxidized nitrogenase proteins form a oxygen tolerant complex with an Fe-S protein, called Fe/S protein II or the Shethna protein (155).

The genetic determinants of this electron transport model have not been determined. However, it has been shown that Rhizobium fixABC-like DNA is present in both A. vinelandii (30, 41) and A. chroococcum (34). These fixABC gene

products have been suggested to be involved in nitrogenase-specific electron transport unique to aerobic or microaerobic nitrogen fixation (30, 41). As such, these fix genes may represent the genetic determinants of this model.

Unknown Functions: nifU, nifS, nifT, nifY, and nifX.

A. vinelandii mutant strains defective in nifS exhibit very reduced diazotrophic growth while nifU mutant strains have a Nif$^-$ phenotype (Appendix III). Both nifU and nifS have been sequenced from A. vinelandii and K. pneumoniae and their respective gene products exhibit a high degree of interspecies sequence identity (6). Plasmids encoding the A. chroococcum nifS gene are able to complement K. pneumoniae nifS mutants for growth on N_2 which demonstrates the functional homology of this gene product between these two organisms (34).

Puhler and Klipp (128) predicted the presence of nifX and nifY based upon the production of two proteins (M_r = 18,000 and 24,000, respectively) in pulse labeled E. coli mini cells containing cloned K. pneumoniae nif fragments. Results of DNA sequence analysis confirmed the presence of these two genes in both K. pneumoniae and A. vinelandii (10, 150, Appendix III). The nifX and nifY gene products share significant sequence identity toward their respective C-terminal ends. However, the residues which are conserved between the K. pneumoniae gene products differ from those conserved between the A. vinelandii gene products (Appendix III). Both A. vinelandii nifX and nifY mutant strains are capable of diazotrophic growth (Appendix III).

The presence of nifT is predicted only on the basis of sequence analysis (1, 150). Mutations in nifT have no observable effect on diazotrophic growth (Appendix III).

nifA and nifL: Regulation.

Transcription of the nif genes in K. pneumoniae is coordinately controlled at two levels (26). The first level is a general control that is mediated by the global nitrogen regulatory elements, the gene products of ntrA, ntrB, and ntrC. The second level of control is nif-specific and is mediated by the gene products of nifA and nifL.

Promoters subject to ntr-mediated regulation lack the consensus -35 and -10 elements found in the typical prokaryotic promoter. Instead, their primary structure is characterized by the consensus sequence **GG**-N_{8}-T**TGCA** extending from -24 to -12 relative to the transcription start site (features in boldtype are invariant). Beynon et al. (7) were the first to observe that all nif-specific promoters have this characteristic consensus sequence. The ntrA gene product (NTRA) is an RNA polymerase sigma factor which confers upon the core polymerase an ability to recognize 'ntr-dependent' promoter sequences. The ntrC gene product (NTRC) is a DNA-binding, transcriptional regulatory protein which can exist in either an active or inactive state. The activational state of NTRC is regulated by the ntrB gene product (NTRB) which is believed to be a cellular nitrogen-sensing element. Under nitrogen-limiting conditions, NTRB phosphorylates and thereby activates NTRC, which in turn stimulates NTRA-mediated transcription of ntr-dependent promoters

(97). The target for NTRA-mediated transcription of the nif gene cluster is the promoter which directs synthesis of the regulatory nifA and nifL genes. Within this second level of regulation, the nifA gene product functions as a positive activator protein which is required for transcription of the remaining nif-specific promoters. This nifA-mediated transcription additionally requires participation of the ntrA-encoded sigma factor (26). The nifA gene product binds to a consensus TGT-N_{10}-ACA upstream activator sequence (UAS) that precedes most nif-specific promoters (110). Buck et al. (see 7 in reference 110) have proposed a model in which the nifA gene product bound at the UAS contacts the NTRA/RNA polymerase complex by looping out the intervening DNA. Although the nifA and nifL gene products show sequence identity to NTRC and NTRB respectively, these protein pairs do not function in a similar fashion. Arnott et al. (2, 103) have shown that while nifL mutant strains are insensitive to the presence of environmental nitrogen or oxygen, nifA activity in these strains is not impaired. They suggest that the function of the nifL gene product is to inactivate the nifA gene product in response to an increase in the nitrogen or oxygen status of the cell. Additionally, in response to the increased nitrogen status of the cell, NTRB dephosphorylates NTRC thereby preventing general NTRA-mediated transcription of the nifLA regulatory operon.

Studies on the regulation of nif gene expression in A. vinelandii are complicated by the differential expression of the two alternative nitrogen fixing systems in response to the availability of metals in the medium. In general, studies have indicated a two-tier system of nif regulation very similar to that found in K.

pneumoniae. For both organisms, the amino acid sequence of the nifA gene product has been determined from the nucleotide sequence (4, 28). A comparison of the proteins reveals that the regions of strongest interspecies sequence identity are the central domain and C-terminus. These regions contain the potential ATP-binding site and the putative helix-turn-helix DNA-binding domain, respectively. Additionally, the K. pneumoniae nifA gene product is able to activate A. vinelandii nif expression in vivo (80). Functional homology of the regulatory systems in the two organisms is also demonstrated by the ability of the A. vinelandii ntrA and ntrC gene products to complement respective K. pneumoniae mutant strains (175)

There are basic differences between the regulation of nif expression in the two organisms. While essential in K. pneumoniae, the ntrC gene product is not required for diazotrophic growth in A. vinelandii (175). Bali et al. (3) have identified a regulatory gene (nfrX) in A. vinelandii which has no homolog in K. pneumoniae. Although mutant strains defective in nfrX exhibit a Nif$^-$ phenotype, the biochemical basis for this phenotype is unknown. Bennett et al. (4) have identified an ORF that is potentially co-transcribed with nifA and which encodes a protein that shares sequence similarity to the K. pneumoniae nifL and ntrB gene products. They suggest that this ORF may encode an environmental sensory element (analogous to NTRB) or regulatory protein (analogous to the nifL gene product) that modulates the activity of the nifA gene product.

Alternative Nitrogen Fixing Systems.

The existence of alternative Mo-independent nitrogenases was first proposed by Bishop et al. (8) on the observation that Nif⁻ mutants of A. vinelandii were able to grow diazotrophically in the absence of Mo. Bishop's group later observed that nifH DNA probes hybridize to three regions on the A. vinelandii genome (72) and subsequently, two alternative nitrogen fixing complexes were isolated from A. vinelandii and one from A. chroococcum.

A vanadium (V)-containing nitrogenase, called nitrogenase-2, can be isolated from each of these organisms when grown on nitrogen- and Mo-free medium containing V. Nitrogenase-2 is composed of two component proteins, an Fe protein and a FeV protein (49, 50, 139). The nitrogenase-2 Fe protein is very similar to the Mo-containing (nitrogenase-1) Fe protein and can complement nitrogenase-1 MoFe protein for enzymatic activity. The FeV protein is an $\alpha_2\beta_2$ tetramer which contains a FeV cofactor (FeVaco) analogous to FeMoco. The FeVaco can be extracted from the FeV protein using the same treatment which is used to extract FeMoco from the MoFe protein. Addition of the isolated FeVaco to purified apo-MoFe protein (isolated from nifB mutant strains) results in activities characteristic of the VFe protein (158). While this hybrid protein is able to reduce acetylene, it is unable to reduce N_2. Smith et al. (158) suggest that reduction of N_2 requires a very specific interaction of the cofactor with polypeptide ligands which are mismatched in the reconstituted hybrid protein. The nucleotide sequence of the gene encoding the Fe protein of nitrogenase-2 (vnfH) has been determined from A. vinelandii (130) and

A. chroococcum (140), while the sequence of the genes encoding the subunits of the FeV protein (vnfD and vnfK) have been determined only from A. chroococcum (77). In each organism, a ferredoxin-like protein is encoded immediately downstream of the vnfH gene which suggests a specific role for this protein in electron transport to nitrogenase-2. In A. chroococcum an additional gene (vnfG) is located between vnfD and vnfK which is believed to code for a third type of subunit ($M_r \approx 14,000$) peculiar to nitrogenase-2 of A. chroococcum. While the function of this subunit is unknown, it is present in the purified A. chroococcum nitrogenase-2 enzyme in a 1:1:1 ratio with the α- and β-subunits (29). All eight conserved cysteine residues present in the MoFe protein subunits are conserved in the VFe protein subunits as well.

A third nitrogenase, termed nitrogenase-3, has been isolated from A. vinelandii cells grown in nitrogen free-medium in the combined absence of Mo and V (20). Nitrogenase-3 is composed of two component proteins, an Fe protein and an Fe-containing protein analogous to the MoFe protein of nitrogenase-1. The Fe-containing protein can be purified in two active configurations: an $\alpha_2\beta_2$ tetramer and an $\alpha_1\beta_2$ trimer, each of which contains Fe as its only metal. The nucleotide sequence of the genes encoding the structural proteins of nitrogenase-3 (anfH, anfD, and anfK) has been determined (77). Interestingly, a third gene (anfG) is located between anfD and anfK which encodes a protein ($M_r \approx 15,000$) that shares 39% identity with the A. chroococcum vnfG gene product. The low activities which Bishop and co-workers observe for purified nitrogenase-3 (20) suggest that the

anfG-encoded protein may be a required component in the nitrogenase-3 complex that is lost during purification. The Fe protein shares 63% identity with the Fe protein of nitrogenase-1, yet it gives less than 1.0% of the expected activity in a heterologous complementation assay with the nitrogenase-1 MoFe protein. The α- and β-subunits of the nitrogenase Fe-containing protein share only 30% identity with the respective α- and β-subunits of nitrogenase-1. Seven of the eight conserved cysteine residues in the MoFe protein subunits are conserved in α- and β-subunit proteins of nitrogenase-3; a serine is present in place of α-subunit cysteine residue 183.

The pattern of substrate reduction for the alternative enzymes is different than for nitrogenase-1, with the distinguishing characteristic being the ability of the alternative systems to reduce acetylene to ethane (74). The expression of all three nitrogenase systems appears to be regulated at the level of transcription in response to Mo, V, and fixed nitrogen (72, 74). Bishop (74) has proposed a model for the regulation of the three nitrogen fixing systems in A. vinelandii in response to metal availability. The main points of the model are as follows. (1) Mo is required for the expression of nitrogenase-1. (2) Mo suppresses the expression of both alternative nitrogenases. (3) V suppresses the expression of nitrogenase-3. (4) V is probably required for the expression of nitrogenase-2. The functions of several nif gene products appear to be shared between the conventional and the alternative nitrogen fixing systems. The nifM (Fe protein maturation) and nifB (FeMoco biosynthesis) gene products are essential for nitrogen fixation under all growth

conditions (74). The requirement for nifB indicates a role for this gene product at a common step in cofactor biosynthesis regardless of the metal composition of the respective cofactor (75). The presence of nifA is required only for expression of nitrogenase-1, but it is necessary for maximal diazotrophic growth utilizing nitrogenase-3 (76). It is unknown how nifB and nifM are expressed in the absence of nifA. Joerger et al. (76) have identified two additional nifA-like genes in A. vinelandii which are located near the structural genes coding for the alternative nitrogenases. These two genes, anfV and vnfA, are required for expression of nitrogenase-2 and nitrogenase-3, respectively. Additionally, the nfxR gene is essential for diazotrophic growth in the presence of Mo and in the combined absence of Mo and V (3).

RESEARCH GOALS

As outlined above, biological nitrogen fixation is genetically, biochemically, and catalytically complex. Biochemical and biophysical studies indicate that the nitrogenase MoFe protein contains six poorly characterized metalloclusters that function as redox centers during the reduction of N_2 to NH_3. The focus of this study was to investigate the structural and functional roles of these ill-defined metal clusters within the MoFe protein of nitrogenase. Site-directed mutagenesis was used to subtly alter the environments surrounding the metal clusters and the resulting biological, biochemical, and biophysical effects were examined. During the logical progression of this project, several additional nif genes and ORF's were discovered within the A. vinelandii nif gene cluster. Therefore, a second area of research was begun in an effort to characterize the function of these genetic elements in the process of biological nitrogen fixation.

Results from these studies generated three major reviewed publications (Appendices I, II, and III), an invited paper (Appendix IV), and five abstracts from presentations of this research at scientific meetings. In this dissertation, the initial discussion will focus on the site-directed mutagenesis project and includes all work presented in the publications in Appendices I, II, and IV. Discussion of the second major area of research will follow, and this includes work presented in the publication in Appendix III. Any additional research or side projects will be

discussed in further appendices. A description of all materials and procedures which were used throughout these studies can be found in the MATERIALS AND METHODS section of this dissertation or within the same section of the publications presented in Appendices I through IV.

MATERIALS AND METHODS

E. coli: Strains and Growth.

E. coli strain TB1 [▲lac-pro strA ara thi ø80dlacZ ▲M15 hsdR] served as the host for all plasmids and E. coli strain 71-18 [▲lac-pro thi⁻ strA supE endA sbcB15 hsdR4 (F'proAB lacI^q ZM15)] served as the host for all M13 bacteriophage vectors. Strain GM2163 [F⁻ λ⁻ ara-14 leuB6 tonA31 lacY1 tsx-78 supE44 galK2 galT22 hisG4 rpsL136 xyl-5 mtl-1 thi-1 dam-13::Tn9 dcm-6 hsdR2 mcrB⁻ mcrA⁻ Sm] was utilized as a host when unmethylated DNA was required. Strain JM105 [thi rpsL endA sbcB15 hsdR4 ▲(lac pro AB)/F' traD36 proAB lacI^{q}Z ▲M15] served as the host for vectors used to overproduce proteins. When necessary, the medium was supplemented with 100 μg/ml ampicillin or 25 μg/ml kanamycin.

A. vinelandii: Strains and Growth.

Table 4 provides descriptions of all A. vinelandii strains constructed and/or used in this study. The wild-type strain A. vinelandii OP served as the ultimate parent for all strains.

Wild-type and mutant strains were grown at 30°C on a modified Burk medium (164). When a fixed nitrogen source was required, ammonium acetate was added to a final concentration of 30 mM. When necessary, the medium was supplemented with 20 μg/ml ampicillin, 0.5 μg/ml kanamycin, or 5.0 μg/ml rifampin.

TABLE 4. A. vinelandii mutant strains constructed or used in this study.
[a] * precedes all strains constructed specifically by the author.
[b] ▲ indicates a deletion; fs indicates a frame shift; :: indicates insertion of a kanamycin resistance-encoding cartidge; z indicates insertion of an in-frame translational fusion of the lacZ gene of E. coli, and sd indicates an oligonucleotide-directed mutation.
[c] Numbers refer to the left and right terminii of deletion mutations. A dashed line indicates that the location of the terminius is unknown. A number preceded by a z or an s indicates the position of the insertion of the lacZ cartridge or the position of the nucleotide that was altered, respectively. All numbers refer to the A. vinelandii nif cluster sequence presented in Appendix XI.
[d] Proposed genes interupted by the mutation. Potential polar effects are not considered a physical disruption. Site-directed mutants are designated as follows: nif gene altered/amino acid residue altered/wild type residue/substituted residue.
[e] + indicates that the mutant strain was capable of normal diazotrophic growth, - indicates that the mutant strain was incapable of diazotrophic growth, and +/- indicates that the mutant strain was capable of only very slow diazotrophic growth.
g See Appendix III for construction of this strain.

Strain[a]	Mutation[a]	Location[c]	nif gene(s) or ORF(s) disrupted or altered[d]	Plasmid used	Parental strain	Nif phenotype[e]
*DJ250	Δ	----720	12	pDB68	DJ249	+
*DJ78	Δ::	----720	12	pDB109	WT	+
*DJ249	Δ	----1225	12 + H Promoter	pDB53	WT	-
DJ46	Δ	2034-2758	HD	pDB70	WT	-
DJ100	Δ	2546-3367	D	pDB31	WT	-
DJ200	Δ::	3665-4697	DK	pDB253	WT	-
DJ13	Δ	4223-4697	K	[g]		-
*DJ190	Δfs	4871-5961	KTY	pDB234	WT	-
*DJ206	Δ::	4871-5961	KTY	pDB261	WT	-
*DJ202	Δ::	4224-6823	KTY,1,2	pDB247	WT	-
*DJ216	Δ	5410-5450	ΔK-T intercistron	pDB260	DJ206	+
*DJ208	Δ	6176-6349	Y	pDB241	DJ202	+
DJ42	Δ	6165-13130	Y,1,2,ENX,3,4	pDB67	WT	-
DJ35	Δ	8162-8929	E	pDB35	WT	-
*DJ170	Δ	10453-11002	NX	pDB202	WT	+
DJ39	Δ	9744-11845	NX,3,4	pDB62	WT	-
*DJ166	Δ	10675-10806	X	pDB184	DJ170	+
*DJ165	Δ::	10675-10806	X	pDB187	WT	-
DJ44	Δ	10930-11565	X,3	pDB61	DJ39	+
*DJ198	Δ	11343-11483	3	pDB254	DJ39	+
*DJ213	sd	s2370,s2371	D45CA	Mn37RF	DJ46	+
*DJ156	sd	s2371	D45CS	Mn26RF	DJ46	+
*DJ273	sd	s2421,s2422	D62CA	Mn38RF	WT	-
*DJ159	sd	s2422	D62CS	Mn27RF	WT	-
*DJ181	sd	s2484	D83HN	Mn33RF	DJ46	+
*DJ231	sd	s2499,s2500	D88CA	Mn39RF	WT	-
*DJ162	sd	s2500	D88CS	Mn28RF	WT	-
*DJ62	sd	s2688	D151QE	Mn2RF	DJ100	+
*DJ258	sd	s2697,s2698	D154CA	Mn40RF	WT	-
*DJ45	sd	s2698	D154CS	Mn3RF	WT	-
*DJ101	sd	s2712	D159IV	Mn4RF	DJ100	+
*DJ103	sd	s2716	D160GA	Mn5RF	DJ100	+
*DJ63	sd	s2720	D161DE	Mn6RF	DJ100	+
*DJ240	sd	s2784,s2785	D183CA	Mn41RF	WT	±
*DJ55	sd	s2785	D183CS	Mn8RF	WT	±
*DJ64	sd	s2808	D191QE	Mn10RF	WT	-
*DJ255	sd	s2808	D191QK	Mn47RF	WT	-
*DJ178	sd	s2820	D195HN	Mn31RF	WT	-
*DJ65	sd	s3048	D271ND	Mn11RF	DJ100	+
*DJ246	sd	s3060,s3061	D275CA	Mn42RF	WT	-
*DJ56	sd	s3061	D275CS	Mn12RF	WT	-
*DJ82	sd	s3061,s3064	D275CS,D276YC	Mn22RF	WT	-
*DJ124	sd	s3064	D276YS	Mn21RF	DJ100	+
*DJ126	sd	s3067	D277RH	Mn23RF	WT	-
*DJ128	sd	s3074	D279MI	Mn24RF	DJ100	+
*DJ66	sd	s3075	D280ND	Mn13RF	DJ100	+
*DJ243	sd	s4024,s4025	K70CA	Mn43RF	WT	-
*DJ84	sd	s4025	K70CS	Mn15RF	WT	-
*DJ210	sd	s4084	K90HN	Mn34RF	DJ200	+
*DJ120	sd	s4100	K95CS	Mn16RF	WT	-
*DJ252	sd	s4273,s4274	K153CA	Mn45RF	WT	-
*DJ122	sd	s4274	K153CS	Mn17RF	DJ13	+
*DJ219	Δ	4273-4275	KΔ153C	MnΔ45RF	DJ13	+
*DJ222	sd	s2484	D83HN	pDB253	DJ181	-
	Δ::	3665-4697	DK			
*DJ324	Δz	z1985-2338	H:lacZ (ΔHD)	pDB287	DJ46	-
*DJ309	Δz	z3665-4697	D:lacZ (ΔDK)	pDB284	DJ13	-
*DJ311	Δz::	z4224-4697	K:lacZ:: (ΔK)	pDB285	DJ13	-
*DJ297	z::	z5734	T:lacZ::	pDB282	WT	+
*DJ168	Δz	z10676-11142	X:lacZ (ΔX,3)	pDB203	DJ170	+
*DJ260	Δz::	z11144-11315	3:lacZ:: (Δ3)	pDB267	DJ39	+
*DJ299	z::	z5734	T:lacZ::	pDB283	DJ216	+
	Δ	5410-5450	ΔK-T intercistron			
*DJ303	z::	z5734	T:lacZ::	pDB282	DJ166	
	Δ	10675-10806	ΔX			
*DJ305	z::	z5734	T:lacZ::	pDB282	DJ198	+
	Δ	11343-11483	Δ3			
*DJ307	z::	z5734	T:lacZ::	pDB282	DJ44	+
	Δ	10930-11565	ΔX,3			
*DJ184	Δz	z10676-11142	X:lacZ (ΔX,3)	pDB167	DJ168	-
	Δ::	---	ΔA::			
*DJ301	Δz::	z11144-11315	3:lacZ:: (Δ3)	pDB167	DJ260	-
	Δ::	---	ΔA::			

Transformation of A. vinelandii was performed in liquid culture as described by Page and von Tigerstrom (122). Placement of specific deletions or insertions into the A. vinelandii genome is detailed in Appendix III. Placement of oligonucleotide-directed mutations into the A. vinelandii genome is detailed in the following section of this dissertation and in Appendix II.

For nitrogenase derepression, 30 ml nitrogen supplemented cultures were shaken at 300 rpm in 300 ml baffled, side-arm flasks. Cells were harvested during exponential growth (Klett ≈ 65), washed with 20 ml phosphate buffer (pH 7.0), resuspended to the original volume in Burk's medium, and returned to the shaker.

Determination of Whole Cell Nitrogenase (Acetylene-reducing) Activity.

After derepressing cell cultures for three hours, 1.0 ml samples were added to stoppered 5.0 ml serum vials. After addition of 500 µl of acetylene, the vials were shaken at 300 rpm at 30 °C for 20 min. The assays were terminated by addition of 100 µl 30% TCA. Ethylene production was measured on a Perkin-Elmer Sigma 2B gas chromatograph by injecting 50 µl of the gas phase onto a Porapak R column (Alltech 8011/2) equipped with a flame ionization detector and a Varian 4290 integrator. The column was run at 82 °C with Argon as the carrier gas. All assays were performed in triplicate.

β-galactosidase Assay.

β-galactosidase-specific activities were determined using a modification of the procedure described by Miller (105). Cultures were derepressed for nitrogenase activity as described above. After cell density (OD_{600} - Bausch and Lomb Spectronic 501) was determined, 100 μl of the culture was added to 900 μl Z buffer (pH 7.0) and permeablized by the addition of 15 μl toluene. The cells were vortexed 10 s and the toluene evaporated by shaking the tubes at 37°C for 45 min. The reaction was initiated by addition of 200 μl p-nitrophenyl-β-D-galactopyranoside (4.0 mg/ml 0.1 M phosphate buffer pH 7.0) and incubated 10-15 min 30°C. The reaction was terminated by adding 500 μl 1.0 M sodium carbonate and β-galactosidase activity measured as the concentration of the cleavage products of the chromogenic substrate (OD_{410}). Sample tubes were blanked against control tubes containing cells treated identically but lacking the chromogenic substrate. Units of activity (Miller Units) were calculated by the following equation $U = OD_{410}/OD_{600}$ (t) (1000) where t = time of reaction in min. All assays were performed in triplicate.

Oligonucleotide-Directed Mutagenesis.

Oligonucleotides (primers and mutenes) were synthesized on either a Beckman System 1 Plus or an ABI Model 381A DNA synthesizer and purified by HPLC (70). Oligonucleotide-directed mutagenesis was performed using a modification of the procedure of Zoller and Smith (188) followed by a methylation step (64):

10X Kinase buffer
- 0.7 M tris-HCl (pH 8.0)
- 0.01 M $MgCl_2$
- 0.05 M DTT

10X Annealing buffer
- 0.2 M tris-HCl (pH 7.5)
- 0.1 M $MgCl_2$
- 0.5 M NaCl
- 0.01 M DTT

Solution B
- 0.2 M tris-HCl (pH 7.5)
- 0.1 M $MgCl_2$
- 0.1 M DTT

Solution N

Mix together equal volumes
- 10 mM dATP
- 10 mM dCTP
- 10 mM dGTP
- 10 mM dTTP

Methylase buffer 10X
- 0.5 M tris-HCl (pH 7.5)
- 0.05 M $MgCl_2$
- 0.01 M DTT

A. Phosphorylation of oligonucleotide

Mutene (OD = 3.4)	12.1 μl
Kinase buffer	6.0 μl
10mM ATP	2.0 μl
H_2O	38.0 μl
T4 polynucleotide kinase	10 Units

Incubate 37°C for 35 min

Incubate 65°C for 10 min

B. Annealing reaction

Phosphorylated mutene	18.0 μl
H_2O	28.0 μl
Annealing buffer	6.0 μl
Template (≈ 2 μg)	8.0 μl

Incubate 55°C for 5 min

Incubate at room temperature for 5 min

C. Extension/Ligation reaction

Annealed DNA	20.0 μl
Solution B	3.0 μl
Solution N	12.0 μl
H_20	6.0 μl
10 mM ATP	3.0 μl
T4 DNA ligase	3.0 Units
Klenow fragment	6.0 Units

Incubate at room temperature for 12 h

D. Methylation

To the overnight extension/ligation reaction add:

Methylase buffer	5.0 μl
DAM methylase	2.0 Units
S-adenosylmethionine (0.03 M)	0.5 μl

Incubate 37 °C for 1.0 h
Incubate 65 °C for 10 min

E. Precipitation/Transfection

Precipitate with 1/10th volume 5.0 M ammonium acetate (pH 5.2) and an equal volume isopropanol. Resuspend in 20 μl 0.01 M Tris/0.001 M EDTA and transfect E. coli strain 7118 with 5.0 μl.

Single-lane dideoxy sequencing (144) was used to screen for the desired mutation. All mutant phage were understreaked for single colony isolation, and the mutations confirmed by DNA sequence analysis using all four reactions. For each mutant construction, replicative-form (RF) DNA was prepared and used to transform A. vinelandii cells. Finally, single-stranded DNA was isolated from the supernatant of cell cultures from which the RF was prepared and sequenced to confirm the mutation and establish that there were no other mutations within 400 nucleotides surrounding the planned mutation.

Overproduction of Proteins in E. coli.

Expression vectors pKK223-3 and pKK233-2 were obtained from Pharmacia. Cells were grown at 37 °C to $OD_{600} \approx 0.6$ in LB medium, induced by the addition of IPTG (5mM final conc), and shaken for an additional 2.5 h. The cells were harvested and SDS-PAGE was performed essentially as described by Laemmli (88). Molecular weight markers were obtained from Sigma (SDS-7B) and LKB (1860-102).

Compared to growth in LB medium, utilization of an AB minimal salts medium had no effect on the quantity of protein expressed.

Two-Dimensional Gel Electrophoresis.

Preparation of extracts for two-dimensional gel electrophoresis was performed as described by Bishop et al. (8). Two-dimensional analysis was performed essentially as described by O'Farrell (120).

Computer Analysis

DNA sequences and deduced protein sequences were analyzed with the Pustell DNA sequence computer program (International Biotechnologies Inc.). Protein hydrophobicity plots and secondary structure analyses were performed using MSEQ 1.17 (Regents of the University of Michigan) as provided by T. D. Wilkins.

SITE-DIRECTED MUTAGENESIS OF THE MoFe PROTEIN

Nomenclature.

All site-directed mutant strains will be designated as in the following example: '**D45CS**' where 'D' designates the nif gene that was altered, '**45**' refers to the location of the altered amino acid residue within that gene product, '**C**' is the one-letter abbreviation for amino acid residue present in the wild type enzyme, and '**S**' is the one-letter abbreviation for the newly substituted amino acid residue. In discussing specific MoFe protein subunit residues, nomenclature as in the following example will be used: '**α-his-45**' where '**α**' refers to the MoFe protein α-subunit, and '**his-45**' refers to the three-letter abbreviation for the amino acid histidine which is located at residue 45 within the α-subunit. In both examples, the numbering corresponds to the A. vinelandii α- and β-subunit sequences shown in Figs. 6 and 7.

Construction of Mutant Strains.

We developed a gene replacement strategy which enabled us to reincorporate a single copy of an altered nif gene into the A. vinelandii chromosome (Appendix II). The benefits of this gene replacement strategy are twofold. (1) Strains constructed in this way do not suffer from gene dosage effects that could unbalance nif gene expression. These mutant strains are identical to the wild type strain with the exception of a single amino acid replacement within one of the MoFe protein

subunits. Therefore, any changes in the MoFe protein isolated from these mutant strains can be attributed to the specific amino acid substitution rather than arising from an indirect effect such as metallocluster processing or insertion. (2) Mutant strains can be constructed and identified regardless of their Nif phenotype.

This site-directed mutagenesis and gene replacement procedure requires three individual steps: (1) deletion of specific MoFe coding sequences from the A. vinelandii genome; (2) use of site-directed mutagenesis to produce specific single base-pair mutations in the homologous sequences carried on a recombinant vector; (3) replacement of the deleted genomic sequences with the mutated recombinant fragment. Construction of specific MoFe protein deletion strains and single site mutation vectors is described in Appendix III and MATERIAL AND METHODS, respectively. The single-site mutation vectors are used to transform competent A. vinelandii cells as shown in Fig. 5. During transformation the deleted genomic region is reincorporated by reciprocal recombination of the mutation vector with homologous sequences on the A. vinelandii genome (Panel A). Thus the recipient strain, which did not produce a complete MoFe protein α-subunit, now produces a complete subunit that carries a single specific amino acid substitution. As the recipient strain was Nif$^-$, any colonies that exhibit the Nif$^+$ or Nif± phenotype on nitrogen free medium are recombinants. Thus, this strategy provides a rapid, positive selection scheme for any amino acid substitution which does not completely eliminate nitrogenase activity. For those mutation vectors unable to restore a Nif$^+$ or Nif± phenotype to the recipient strain, a more conventional, indirect selection

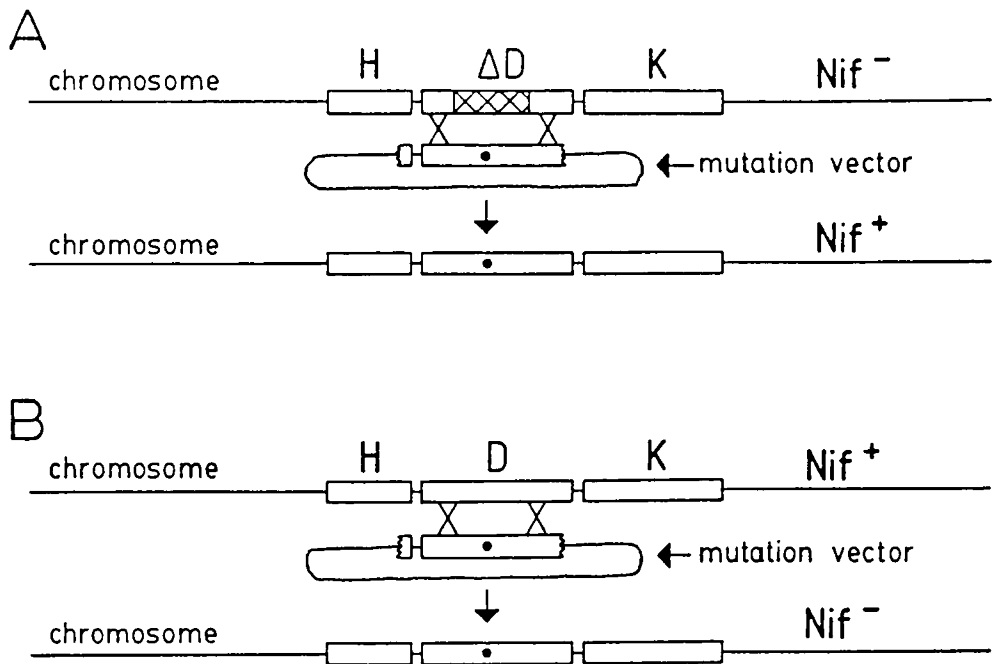


FIGURE 5. Strategy for the construction of site-directed mutant strains. Introduction of a single point mutation, indicated by a dot on the mutation vector, into the A. vinelandii nifD coding sequence leads to either a Nif$^+$ or a Nif$^-$ phenotype in the resultant mutant strain. (A) Those mutations that do not cause a Nif$^-$ phenotype are introduced into the A. vinelandii genome by transforming a Nif$^-$ strain which carries a deletion within a nif structural gene (indicated by crosshatching) to prototrophy. (B) Those mutations that result in a Nif$^-$ phenotype are introduced into the A. vinelandii genome by transforming the wild type strain to the Nif$^-$ character. Reciprocal recombination events are indicated by crosses between the chromosome segments and the mutation vector.

scheme is employed (Panel B). These mutations are introduced into the A. vinelandii chromosome by simultaneous transformation of wild type cells with both the mutation vector and purified A. vinelandii chromosomal DNA that carries a rifampicin resistance marker (congression). Rif^r transformants are scored in the absence and presence of a fixed nitrogen source to identify Nif^- cotransformants. Two-dimensional gel electrophoresis and immunological analysis were used to confirm the presence of both MoFe protein subunits in each mutant strain. As shown in Appendix II, two-dimensional gel analysis can be used to demonstrate charge shifts that are predicted to occur in the MoFe protein subunits of mutant strains. This feature permits additional proof that strains which carry phenotypically silent mutations are readily recovered by this mutagenesis scheme. For example, the α-subunits of strains D191QK and D151QE displayed a basic and an acidic shift, respectively, in the first dimension of two-dimensional gels (data not shown).

Rationale for Amino Acid Replacements.

The identification of structurally important regions in the MoFe protein and the assignment of specific amino acid residues as potential metal center ligands are based upon five criteria: (1) Consideration of the mutually exclusive extrusion requirements of the two classes of metal centers; (2) Consideration of the MoFe protein spectroscopic analyses; (3) Comparison of MoFe protein sequences from widely diverse diazotrophic species; (4) Comparison of MoFe protein α- and β-subunits to each other; (5) Comparison of the MoFe protein subunit sequences to

sequences of gene products involved in FeMoco biosynthesis. These criteria are discussed in the papers presented in Appendices I, II, and IV and are briefly reviewed below.

Extrusion Properties/Spectroscopic Analyses.

Potential cluster ligands can be targeted by assuming that the solvents required to extrude the clusters from the protein mimic the functional groups of the amino acids which serve as ligands to the clusters within the native protein. The four P-clusters are extruded from the MoFe protein by denaturing the protein in the presence of excess thiols (87). This isolation scheme is suggestive of cysteinyl thiol ligands to the P-clusters within the native protein. However, the unusual spectroscopic properties of the P-clusters suggest that one or more of the cysteinyl ligands could be replaced by other nucleophiles. Such unusual coordination could be the basis for the two classes of inequivalent P-clusters proposed by McLean et al. (101). The occurrence of histidine as a ligand to other Fe-containing prosthetic groups (e.g. [2Fe-2S] clusters and heme groups) makes it a potential ligand candidate (167, 143). The FeMoco clusters can be isolated intact from the MoFe protein by an acid/base treatment of the native protein followed by extraction of the denatured protein with NMF (151). The requirement of NMF to extract FeMoco as well as results from ESE (168) and FT-IR studies (178) on the isolated cofactor are suggestive of a nitrogen ligand (e.g. amide or imidazole) to the cofactor within the native protein. Additionally, isolated FeMoco binds thiophenol in a ratio of one

thiophenol/molybdenum indicating that the cofactor may have a cysteinyl-thiol ligand within the native protein (17). In considering these observations, two conclusions can be made in regard to potential cluster ligands. (1) The two cluster types are held within the protein by different ligand modes. Ligation of FeMoco appears to be within a region of the protein that provides both cysteinyl-thiol and nitrogen-donor features. While the P-clusters appear to be ligated predominantly via cysteinyl-thiol ligands, other nucleophilic ligands must be considered. (2) Assuming the validity of four cysteinly-thiol ligands per P-cluster and one per FeMoco cluster, a total of eighteen cysteine residues are required per $\alpha_2\beta_2$ tetramer.

Interspecies Comparisons.

Figures 6 and 7 present alignments of all deduced MoFe protein α-subunit sequences (with the exception of R. capsulatus) and of all deduced MoFe protein β-subunit sequences, respectively. With the exception of C. pasteurianum, the overall sequence conservation within the subunits is quite high and each subunit contains specific regions of very strong sequence conservation. The predicted secondary structure and charge distribution of the respective subunits is highly conserved between the different species (69, 79, 89, 180, 183), with C. pasteurianum showing the most diversity. In Fig. 8, these structural features are shown for the α- and β-subunits of A. vinelandii. Based upon interspecies comparisons of both protein sequence and predicted secondary structure, the A. vinelandii α- and β-subunits have been divided into several conserved regions. These theoretical divisions are for

FIGURE 6. Alignment of MoFe protein α-subunit sequences from *A. vinelandii* (Av), *K. pneumoniae* (Kp), *Anabaena* 7120 (An), *Br. japonicum* (Rj), *Rhizobium* sp. *Parasponia* (PR), cowpea *Rhizobium* strain IRc78 (cR), and *C. pasteurianum* (Cp). (See text for references). Amino acid residues conserved in all seven organisms or in the six organisms excluding *C. pasteurianum* are boxed. Dashed lines represent alignment adjustments to provide best fit. The five conserved cysteine residues are indicated by triangles above the sequence comparison. Amino acid residues that were substituted in this study are indicated by arrows below the sequence comparison. Numbering refer to the *A. vinelandii* sequence. Region I refers to residues 60 - 100, region II refers to residues 115 - 139, region III refers to residues 148 - 163, region IV refers to residues 180 - 206, and region V refers to residues 270 - 288.

I

II

III

IV

V

Av-α
Kp-α
An-α
Bj-α
Pl-α
cB-α
Cp-α

FIGURE 7. Alignment of MoFe protein β-subunit sequences from *A. vinelandii* (Av), *K. pneumoniae* (Kp), *Anabaena* 7120 (An), *Br. japonicum* (Rj), *Rhizobium* sp. *Parasponia* (PR), and *C. pasteurianum* (Cp). (See text for references.) Amino acid redidues conserved in all six organisms or in the five organisms excluding *C. pasteurianum* are boxed. Dashed lines represint alignment adjustments to provide best fit. The three conserved cysteine residues are indicated by triangles above the sequence comparison. Amino acid residues that wer substituted in this study are indicated by arrows below the sequence comparison. Numbering refers to the *A. vinelandii* sequence. Regions I, II, and III refer to residues 65 - 100, 110 - 135, and 147 - 162, respectively.

```
                             20
Av-β  MSQQVDKIKASYPLFLDQDYKDMLA-KKRDGFEE-
Kp-β  MSCTIDKINSCYPLFEQDEYQELFRNKR-QL-EEA
An-β  MPQNPERTVDHVDLFKQPEYTELFENKRKN-FEGA
Rj-β  MPQSAEHVLDHVELFRGPEYQQMLA-KKK-IFENP
PR-β  MAQSADHVLDHLELFRGPEYQQMLADKK--MFENP
Cp-β  -------------------------------MLDA

           40                  60
Av-β  KYPQDKIDEVFQWTTTKEYQELNFQREALTVNPAK
Kp-β  ND-AQRVQEVFAWTTTAEYEALNFQREALTVDPAK
An-β  K-PPEEVERVSEWTKSWDYREKNFAREALTVNPAK
Rj-β  RDPAE-VERIKEWTKTAEYREKNFAREALAVNPAK
PR-β  RDPAE-VERIRAVTKTPEYREKNFA-EALAVNPAK
Cp-β  -TPKEIVER---------------K-ALRINPAK

                 80                100
Av-β  ACQPLGAVLCALGFEKTMPYVHGSQGCVAYFRSYF
Kp-β  ACQPLGAVLCSLGFANTLPYVHGSQGCVAYFRTYF
An-β  GCQPVGAMFAALGFEGTLPFVQGSQGCVAYFRTHL
Rj-β  ACQPLGAVFASVGFERTLPFVNGSQGCVAYYRSHL
PR-β  ACQPLGAVFVSVGFEGTLPFVHGSQGCVAYYRSHL
Cp-β  TCQPVGAMYAALGIHNCLPHSHGSQGCCSYHRTVL

                     120
Av-β  HRHFREPVSCVSDSMTEDAAVFGGQQNMKDGLQNC
Kp-β  NRHFKEPIACVSDSMTEDAAVFGGNNNMNLGLQNA
An-β  SRHYKEPCSAVSSSMTEDAAVFGGLNNMIEGMQVS
Rj-β  SRHFKEPSSCVSSSMTEDAAVFGGLNNMTDGLANS
PR-β  SRHFKEPSSCVSSSMTEDAAVFGGLNNMIDGLANS
Cp-β  SRHFKEPAMASTSSFTEGASVFGGGSNIKTAVKNI

      140                 160
Av-β  KATYKPDMIAVSTTCMAEVIGDDLNAFINNSKKEG
Kp-β  SALYKPEIIAVSTTCMAEVIGDDLQAFIANAKKDG
An-β  YQLYKPKMIAVCTTCMAEVIGDDLGAFITNSKNAG
Rj-β  YKMYKPKNIAVSTTCMAEVIGDDLNAFIKTSKEKG
PR-β  YNMYKPKNI-CSTTCMAEVIGDDLNAFIKTSKEKG
Cp-β  FSLYNPDIIAVHTTCLSETLGDDLPTYISQNEDAG

            180                 200
Av-β  FIPDEFPVPFANTPSFVGSHVTGWDNMFEGIARYF
Kp-β  FVDSSIAVPHANTPSFIGSHVTGWDNMFEGFAKTF
An-β  SIPQDFPVPFANTPSFVGSHITGYDNMMKGILSNL
Rj-β  SVPADFDVPFANTPAFVGSHVTGYDNALKGILENF
PR-β  SVRRSS-TPFANTPAFVGSHVTGYDNALKGILENF
Cp-β  SIPEGKLVIHTNTPSYVGSHVTGFANNVQGIVNYL

                   220
Av-β  TLKSMDDKVVGS---NKKINIVPGFETY-LGNF--
Kp-β  TADYQGQPGKLP-----KLNLVTGFETY-LGNF--
An-β  TEGKKKATS------NGKINFIPGFDTY-VGNN--
Rj-β  WDGKAGTAPKLERKPNGAINIIGGFDGYTVGNL--
PR-β  WNGKAGTAPKLERKPNEAINIIGGFDGNTVGNL--
Cp-β  SENTGAK--------NGRINVIPGF----VGPADN

        240                 260
Av-β  RVIKRMLSEMGVGYSLLSDPEEVLDTPADGQFRMY
Kp-β  RVLKRMMEQMAVPCSLLSDPSEVLDTPADGHYRMY
An-β  RELKRMMGVMGVDYTILSDSSDYFDSPNMGEYEMY
Rj-β  REIKRILELMGIQNTVLADNSEVFDTPTDGEFRMY
PR-β  REIKRILALMGIKNTILADNSEVFDTPTDGEFRMY
Cp-β  REIKRLFEAMDIPYINFPDTSGVLDGPTTGEYKMY

             280                 300
Av-β  A-GGTTQEENKDAPNALNTVLLQPWNLEKTKKFVE
Kp-β  -SGGTTQQENKEAPDAIDTLLLQPWQLLKSKKVVQ
An-β  PS-GTKLEDAADSINAKATVALQAYTTPKTREYIK
Rj-β  -DGGTTLKDAANAIHAKATISMQQWCTEKTLSFAA
PR-β  -DGGTHVEDTANAIHAKATISMQQWCTEKTLPFVS
Cp-β  PEGGTKIEDLKDTGNSDLTLSLGSYASDLGAKTLE

                     320                 340
Av-β  GTWKHEVPKLNIPMGLDWTDEFLMKVSEISGQPIP
Kp-β  ENWNQPATEVAIPLGLAATDELLMTVSQLSGKPIA
An-β  TQWKQETQVLR-PFGVKGTDEFLTAVSELTGKAIP
Rj-β  EHGQDVLSF-NYPVGCSATDDFIVALSRISGKEIP
PR-β  EHGQDVVSF-NYPVGVSATDDLLVALSRISGKEIP
Cp-β  KKCKVPFKTLRTPIGVSATDEPIMALSEATGKEVP

                           360
Av-β  ASLTKERGRLVDMMTDSHTWLHGKRFALWGDPDFV
Kp-β  DALTLERGRLVDNMLDSHTWLHGKKFGLYGDPDFV
An-β  EELEIERGRLVDAITDSYAWIHGKKFAIYGDPDLI
Rj-β  EQLARERGRLVDAIADSSAHVHGKKFAIYGDPDLC
PR-β  EQLARERGRLVDAIADSSANIHGKKFAIYGDPDLC
Cp-β  ASIEEERGQLIDLMIDAQQYLQGKKVALLGDPDEI

         380                 400
Av-β  MGLVKFLLELGCEPVHILCHNGNKRWKKAVDAILA
Kp-β  MGLTRFLLELGCEPTVILSNNANKRWQKAMNKMLD
An-β  ISITSFLLEMGAEPVHILCNNGDDTFKREMEAILA
Rj-β  YGLAAFLLELGAEPTHVLSTNGNKAWQEKMQALLA
PR-β  YGLAAFLLELGAEPTHVLSTNGNNVAGENA-TLFA
Cp-β  IALSKFIIELGAIPKYVVTGTPGMKFQKEIDAMLA

             420                 440
Av-β  ASPYGKNATVYIGKDLWHLRSLVFTDKPDFMIGNS
Kp-β  ASPYGRDSEVFINCDLWNFRSLMFTRQPDFMIGNS
An-β  ASPFGKEAKVWIQKDLWHFRSLLFTEPVDFFIGNS
Rj-β  SSPFGQGCQVYPGRDLWHMRSLLFTEPVDFLIGNT
PR-β  GSPFGELP-AYPGRDLWHMRSLLFTEPVDFLIGNT
Cp-β  EAG-IEGSKVKVEGDFFDVNQWIKNEGVDLLISNT

                   460                 480
Av-β  YGKFIQRDTLNKGKEFEVPLIRIGFPIFDRHHLHR
Kp-β  YGKFIQRDTLAKGKAFEVPLIRLGFPLFDRHHLHR
An-β  YGKYLWRDTSI-------PMVRIGYPLFDRHHLHR
Rj-β  YGKYLERDTAT-------PLIRIGFPIFDRHHHHR
PR-β  MGKYLERDTGT-------PLIRIGFPIFDRHHHHR
Cp-β  YGKFIAREENI-------PFVRFGFPIMDRYGHYY

                         500
Av-β  STTLGYEGANQILTTLVNSILERLDEETRGNQATD
Kp-β  QTTWGYEGANNIVTTLVNAVLEKLDSDTSQLGKTD
An-β  YSTLGYQGGLNILNWVVNTLLDEMDRSTNITGKTD
Rj-β  FPIWGYQGGLNVLVKILDKIFDEIDNKTNILGKTD
PR-β  FPVWGYQGGLNVLVKILDKIFDEIDKKTSVLGKTD
Cp-β  MPKVGYKGAIRLVEEITNVILDKIERRC-----TE

        520
Av-β  YNNDLVR
Kp-β  YSFDLVR
An-β  ISFDLIR
Rj-β  YSFDIIR
PR-β  YSFDIIR
Cp-β  EDFEVVR
```

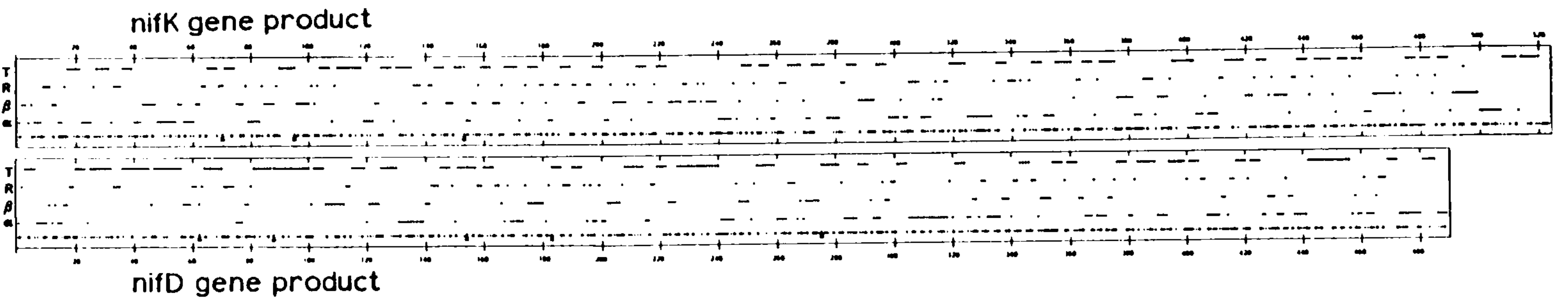


FIGURE 8. Comparison of the predicted secondary structures of the A. vinelandii MoFe protein α- and β-subunits. α = alpha helix, β = beta strand, R = random coil, and T = beta turn. Cysteine residues conserved within each individual subunit are marked with arrows. The sequences are overlayed such that the first three conserved cysteine residues in each subunit are aligned. Within each subunit, hydrophobic and charged residues are annotated: '+' designates residues which are cationic at neutral pH (i.e. Arg, His, and Lys), '-' designates residues which are anionic at neutral pH (i.e. Asp and Glu), and 'o'designates hydrophobic residues (Ile, Leu, Met, Phe, Trp, Tyr, and Met).

discussion purposes and are labeled with Roman numerals (I through V in the α-subunit and I through III in the β-subunit) in Figs. 6 and 7. There are five conserved cysteine residues among the α-subunits (**α-cys-62** and **α-cys-88** within region I, **α-cys-154** within region III, **α-cys-183** within region IV, and **α-cys-275** within region V), three among the β-subunits (**β-cys-70** and **β-cys-95** within region I, and **β-cys-153** within region III), and each is present within a highly conserved region of the respective protein. Each of these cysteine residues is located at or within a β-turn, and as noted by Wang et al. (180), this indicates a near surface location for these residues and for the prosthetic groups which they may ligate. Only **α-cys-154** is associated with an amino acid sequence similar to conserved sequences found among bacterial ferredoxins. These conserved cysteine residues account for only sixteen of the required eighteen cysteine residues that are necessary for total thiol ligation of the metalloclusters within the native protein. Assuming there are no non-conserved cysteinyl ligands to the metal clusters, there must be other types of coordinating ligands. Possible alternative nucleophilic ligands include several conserved histidine residues: **α-his-83** and **β-his-90** (both of which are located five residues preceding a conserved cysteine residue), **α-his-80**, **α-his-195**, **α-his-196**, **α-his-274**, and **β-his-106**. These conserved histidine residues are also candidates as potential nitrogen ligands to the FeMoco cluster.

Intersubunit Comparisons.

The α- and β-subunits have been shown to have a considerable degree of spatial conservation and sequence similarity within the regions surrounding the first three conserved cysteine residues of each subunit (169). Figure 9 shows a comparison of the A. vinelandii MoFe protein subunits in which these conserved regions are highlighted. The presence of this intersubunit conservation has led to the suggestion that the N-termini of the subunits have similar structural and functional roles. However, these regions within the two subunits are predicted to have very different secondary structures (Fig. 8) and it is unclear if the observed spatial and sequential similarities translate into a functional similarity. Low resolution X-ray crystallography analysis predicts structural homology between the two subunits at the three-dimensional level (184). Since three-dimensional structure is generally more conserved than the amino acid sequence (132), it is possible that these two regions share a similar role (e.g. metal cluster binding) within the native MoFe protein in spite of their limited primary identity.

Comparison of α- and β-Subunits with FeMoco Biosynthetic Gene Products.

The MoFe protein subunits are not required for FeMoco biosynthesis (134, 177). This observation implies that the cofactor is preassembled and then inserted into the apo-MoFe protein. It is likely that some of the FeMoco biosynthetic gene products bind FeMoco or some FeMoco intermediate in a manner similar to the MoFe protein. Therefore, we suggested that certain FeMoco biosynthetic gene

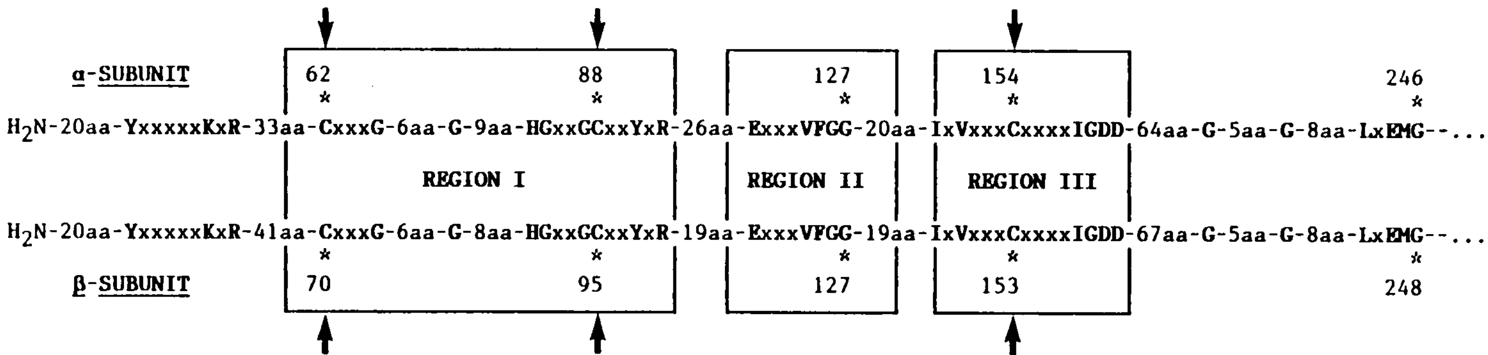


FIGURE 9. Comparison of structurally similar domains within the α- and β-subunits of the nitrogenase MoFe protein of A. vinelandii. Conserved cysteine residues are marked with arrows. 'aa' or 'x' denotes amino acid. α-subunit regions I, II, and III refer to residues 60 - 100, 115 - 139, and 148 - 163, respectively. β-subunit regions I, II, and III refer to residues 65 - 100, 110 - 135, and 147 - 162, respectively.

products would share sequence similarity when compared to the MoFe protein subunits and that these comparisons could provide information concerning the FeMoco binding domains with the MoFe protein. Indeed, a high degree of spatial conservation and sequence similarity was observed when the MoFe protein α- and β-subunits were compared to the products of the FeMoco biosynthetic nifE and nifN gene products, respectively (24, Appendix I). In Figs. 10 and 11, the nifE and nifN gene products are compared to the MoFe protein α- and β-subunits, respectively. Four of the five cysteine residues conserved within the α-subunit are conserved within the nifE gene product. Only one of the three conserved cysteine residues within the β-subunit is conserved within the nifN gene product. Interestingly, the interspecifically conserved histidine residues corresponding to **α-his-80, α-his-83, β-his-90**, and β-his-106 are also conserved within the respective nifE and nifN gene products. Additionally, the histidine residues corresponding to **α-his-83** and **β-his-90** are also conserved when the nifE and nifN gene products are compared to each other. While the nifE and nifN gene products are expected to have tertiary structures similar to the MoFe protein subunits, the predicted secondary structures of the nifE and nifN gene products are not strikingly similar to those predicted for the respective MoFe protein subunits (Figs. 12 and 13). Additionally, the distribution of charged residues (Figs. 12 and 13) and the hydrophobic indices (Figs. 14 and 15) of the compared proteins are not very similar.

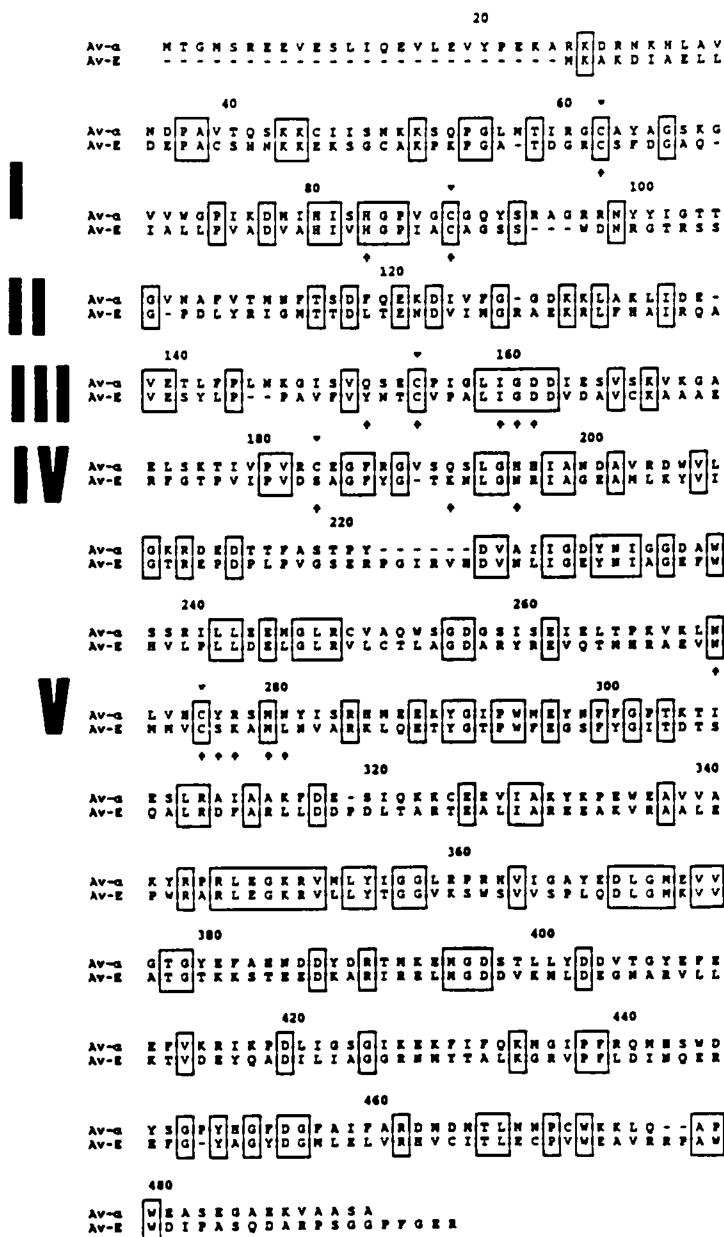


FIGURE 10. Comparison of the MoFe protein α-subunit and the nifE-encoded polypeptide. Dashed lines represent alignment adjustments to provide best fit. Numbering refers to the α-subunit sequence. The five conserved cysteine residues within the α-subunit are indicated by triangle above the sequence comparison. Amino acids that were substituted in this study are indicated by arrows below the sequence comparisons. Roman numerals refer to α-subunit regions as defined in Fig. 6.

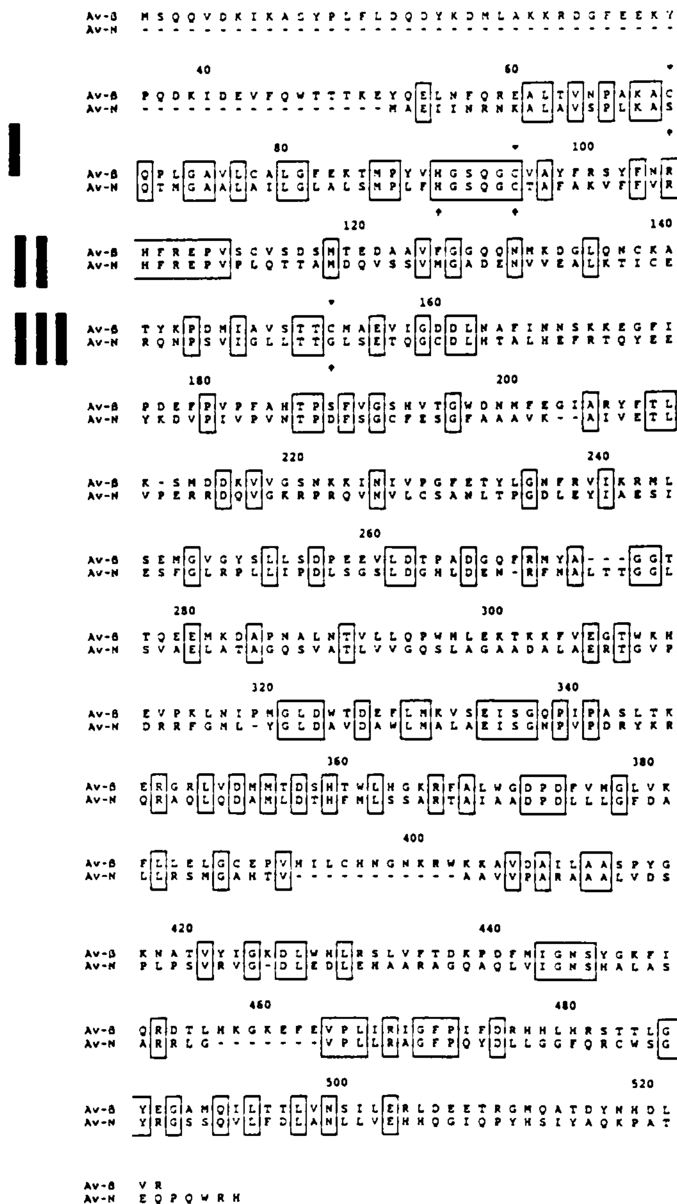


FIGURE 11. Comparison of the MoFe protein β-subunit and the nifN-encoded polypeptide. Dashed lines represent alignment adjustments to provide best fit. Numbering refers to the β-subunit sequence. The three conserved cysteine residues within the β-subunit are indicated by triangle above the sequence comparison. Amino acids that were substituted in this study are indicated by arrows below the sequence comparisons. Roman numerals refer to β-subunit regions as defined in Fig. 7.

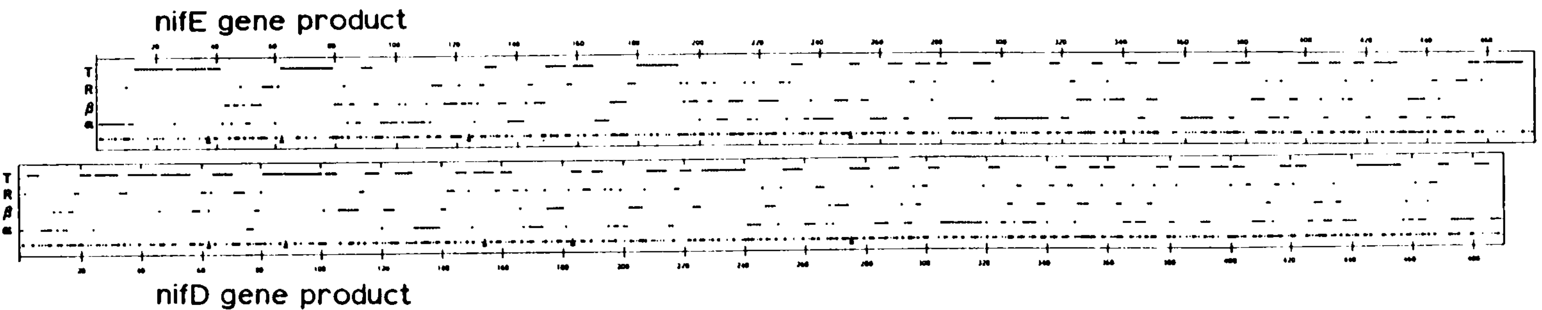


FIGURE 12. Comparison of the predicted secondary structures of the MoFe protein α-subunit and the nifE-encoded polypeptide. α = alpha helix, β = beta strand, R = random coil, and T = beta turn. The structural predictions are overlayed according to the sequence alignment in Fig. 10. Cysteine residues conserved within the α-subunit and the corresponding cysteine residues in nifE gene product are marked with arrows. Within each polypeptide, hydrophobic and charged residues are annotated: '+' designates residues which are cationic at neutral pH (i.e. Arg, His, and Lys), '-' designates residues which are anionic at neutral pH (i.e. Asp and Glu), and 'o'designates hydrophobic residues (Ile, Leu, Met, Phe, Trp, Tyr, and Met).

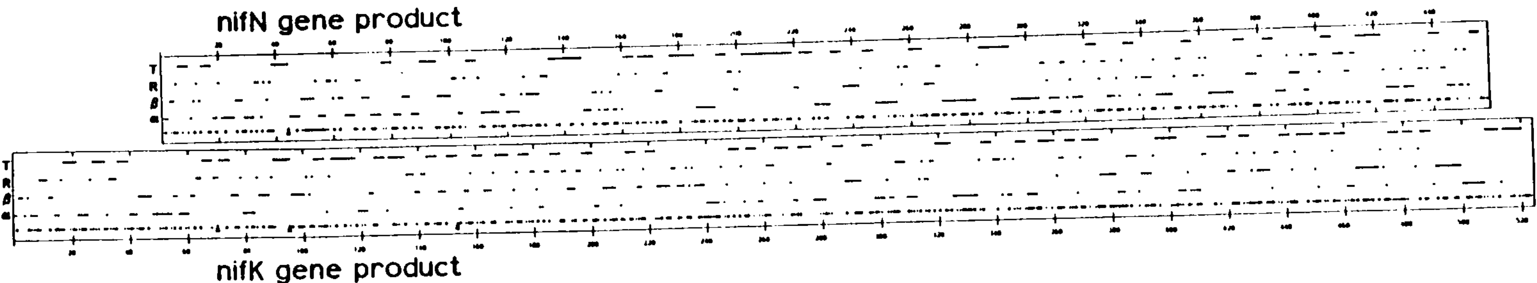


FIGURE 13. Comparison of the predicted secondary structures of the MoFe protein β-subunit and the nifN-encoded polypeptide. α = alpha helix, β = beta strand, R = random coil, and T = beta turn. The structural predictions are overlayed according to the sequence alignment in Fig. 11. Cysteine residues conserved within the β-subunit and the corresponding cysteine residues in nifN gene product are marked with arrows. Within each polypeptide, hydrophobic and charged residues are annotated: '+' designates residues which are cationic at neutral pH (i.e. Arg, His, and Lys), '-' designates residues which are anionic at neutral pH (i.e. Asp and Glu), and 'o'designates hydrophobic residues (Ile, Leu, Met, Phe, Trp, Tyr, and Met).

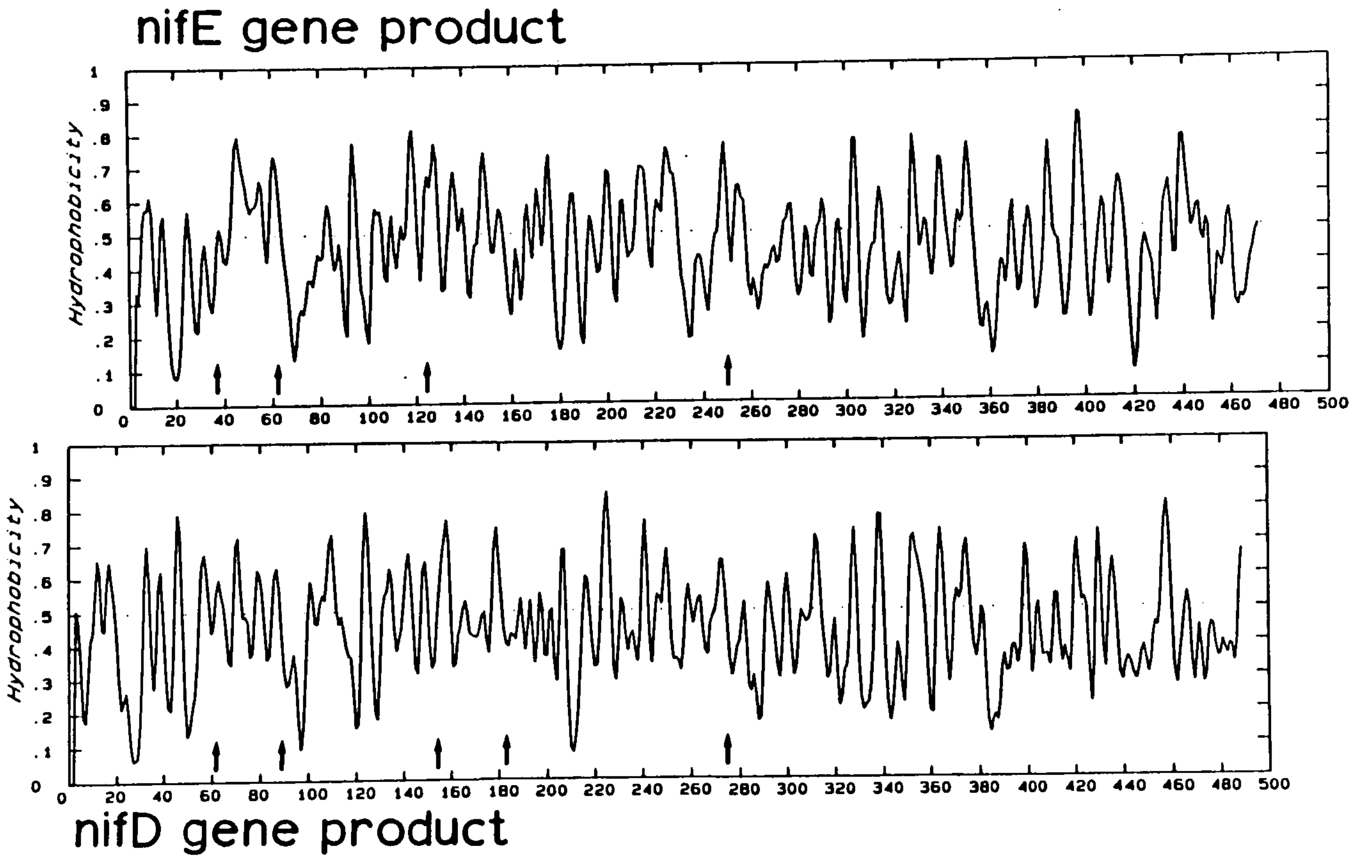


FIGURE 14. Comparison of the hydrophobicity of the MoFe protein α-subunit and the <u>nifE</u>-encoded polypeptide. The patterns are overlayed according the the sequence alignment in Fig. 10. Cysteine residues conserved within the α-subunit and the corresponding cysteine residues in <u>nifE</u> gene product are marked with arrows.

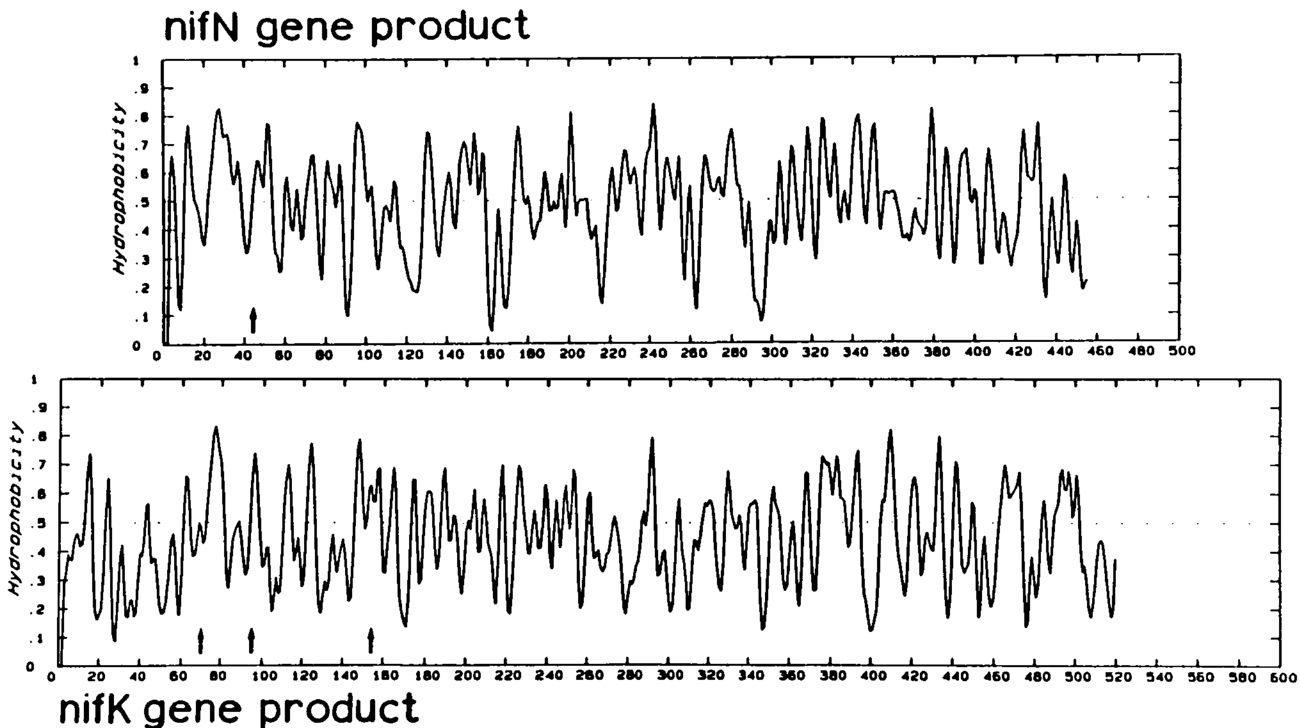


FIGURE 15. Comparison of the hydrophobicity of the MoFe protein β-subunit and the nifN-encoded polypeptide. The patterns are overlayed according the the sequence alignment in Fig. 11. Cysteine residues conserved within the β-subunit and the corresponding cysteine residues in nifN gene product are marked with arrows.

Specific Amino Acid Replacements.

The conserved cysteine residues present within the MoFe protein were targeted for replacement since these residues are most likely to function as essential thiol ligands to the metal clusters. Since non-thiol ligands are indicated for both cluster types, we also targeted highly conserved nitrogen-donor residues (e.g. histidine, asparagine, and glutamine) which were located near the conserved cysteine residues. Finally, several rather neutral changes were made within or near the sequence considered important for Fe-S cluster binding in bacterial ferredoxins. A complete list of the altered amino acids and the reason for their change is given in Table 5. Cysteine → alanine substitutions were chosen to eliminate all cluster ligand properties at that specific residue. Cysteine → serine substitutions were chosen to examine the possibility that oxygen could replace sulfur as a ligand to the various clusters. Non-cysteinyl substitutions were of two types: (1) replacement of an amino acid with a reasonably isosteric residue of different function (e.g. **asp** ⟷ **asn**, **gln** ⟷ **glu**, and **his** ⟷ **asn**) or (2) replacement of an amino acid with one of identical function but different structure (e.g. **asp** ⟷ **glu** and **gly** ⟷ **ala**). The first type of change probes function while maintaining structure and the second type of change looks loosely at the dependence of function on structure (84). Finally, several substitutions were made in order to make the α-subunit more like the nifE gene product. In all cases, substitutions were chosen in an effort to minimize structural disruption of the protein.

TABLE 5. Location and rationale for MoFe protein amino acid residues which were targeted for change.

Amino Acid Substitution	Subunit Region	Rationale for targeting
α-cys-45	α - I	Positive control: Non-conserved cys residue
α-cys-62	α - I	Potential metallocluster ligand: Conserved cys residue
α-his-83	α - I	Potential metallocluster ligand: Conserved his residue
α-cys-88	α - I	Potential metallocluster ligand: Conserved cys residue
α-gln-151	α - III	Conserved residue within ferredoxin-like protein sequence
α-cys-154	α - III	Potential metallocluster ligand: Conserved cys residue
α-ile-159	α - III	Conserved residue within ferredoxin-like protein sequence
α-gly-160	α - III	Conserved residue within ferredoxin-like protein sequence
α-asp-161	α - III	Conserved residue within ferredoxin-like protein sequence
α-cys-183	α - IV	Potential metallocluster ligand: Conserved cys residue
α-gln-191	α - IV	Potential metallocluster ligand: Conserved asn residue
α-his-195	α - IV	Potential metallocluster ligand: Conserved his residue
α-asn-271	α - V	Potential metallocluster ligand: Conserved asn residue
α-cys-275	α - V	Potential metallocluster ligand: Conserved cys residue
α-tyr-276	α - V	Conserved residue within potential FeMoco binding region
α-arg-277	α - V	Conserved residue within potential FeMoco binding region
α-met-279	α - V	Conserved residue within potential FeMoco binding region
α-asn-280	α - V	Potential metallocluster ligand: Conserved asn residue
β-cys-70	β - I	Potential metallocluster ligand: Conserved cys residue
β-his-90	β - I	Potential metallocluster ligand: Conserved his residue
β-cys-95	β - I	Potential metallocluster ligand: Conserved cys residue
β-cys-153	β - III	Potential metallocluster ligand: Conserved cys residue

Characterization of Mutants.

The site-directed mutagenesis/gene replacement strategy was used to isolate thirty-three mutant strains, twenty-six with changes in the α-subunit and seven with changes in the β-subunit. The diazotrophic growth capabilities, in vivo acetylene reducing capabilities, and crude extract component protein activities of the mutant strains are presented in Table 6. These data indicate that the diazotrophic growth capability of the individual mutant strains do not necessarily reflect the acetylene-reducing capabilities of the whole cell or of crude extracts. However, whole cell acetylene-reducing activity is a fairly good indicator of crude extract MoFe protein activity.

Substitution of nonconserved cys residue **α-cys-45** with either **ala** or **ser** had little effect on the diazotrophic growth properties (Fig. 16) or whole cell acetylene-reducing activities of the resultant strains (**D45CA** and **D45CS**). Substitution of conserved cysteine residues **α-cys-62**, **α-cys-88**, **α-cys154**, **α-cys275**, **β-cys-70**, and **β-cys-95** with either **ala** or **ser** resulted in a complete loss of diazotrophic growth and MoFe protein activity in the resultant mutant strains. Of these mutant strains, only **D88CA** and **D88CS** exhibited even a detectable amount of whole cell acetylene-reducing activity. These results strongly suggest, but do not prove, that these conserved **cys** residues provide essential thiol ligands to the metal clusters. In contrast, conserved residue **α-cys-183** does not provide an essential metallocluster ligand since mutant strains with either **ala** or **ser** substitutions at this site (**D183CA** and **D183CS**) were capable of very slow diazotrophic growth (Fig. 17) and exhibited

TABLE 6. Growth characteristics and whole cell and component protein activities of site-directed mutant and wild type strains.

			Acetylene-reduction activity		
Strain	Nif Phenotype	Doubling Time (h)	Whole Cell Activity (% Wild-type)	Crude Extract Specific Activities[#] Fe[a]	MoFe[c]
Wild type	+	2.50	100	20.0-43.0[b]	53.7
D45CA	+	2.75	87		
D45CS	+	3.25	71		
D62CA	-	-	0		
D62CS	-	-	0	37.8	0.1
D83HN	+	3.00	69		
D88CA	-	-	bd		
D88CS	-	-	bd	39.0	0.3
D151QE	+	2.75	70	33.6	38.2
D154CA	-	-	0		
D154CS	-	-	0	14.7	0.1
D159IV	+	2.75	62		
D160GA	+	3.00	46		
D161DE	+	2.50	64	18.1	41.7
D183CA	+/-	8.50	3		
D183CS	+/-	12.00	4	23.4	3.1
D191QE	-	-	12	27.3	10.0
D191QK	-	-	2		
D195HN	-	-	1		
D271ND	+	3.00	75		
D275CA	-	-	0		
D275CS	-	-	0	20.9	0.2
D275CS::D276YC	-	-	0		
D276YS	+	3.25	63		
D277RH	-	-	bd		
D279MI	+	3.25	68		
D280ND	+	4.25	25		
K70CA	-	-	0		
K70CS	-	-	0	16.2	0.0
K90HN	+	3.00	59		
K95CS	-	-	0	24.7	0.2
K153CA	-	-	1		
K153CS	+	3.50	46	26.0	25.0
K▲153C	+	3.50	38	26.0	19.0

[#]Specific activity: nmol of ethylene formed per min per mg of crude extract protein.
[a]Determined in the presence of saturating levels of purified _A. vinelandii_ MoFe protein. [b]Initial assays performed at the USDA lab (Albany, CA) gave Fe protein activities ≥ 40. Recently, Fe protein activities ≈ 20 have been routine. The reason for this variation is unknown (H. D. May, personal communication).
[c]Determined in the presence of saturating levels of purified _A. vinelandii_ Fe protein.
bd = barely detectable amount of ethylene produced.

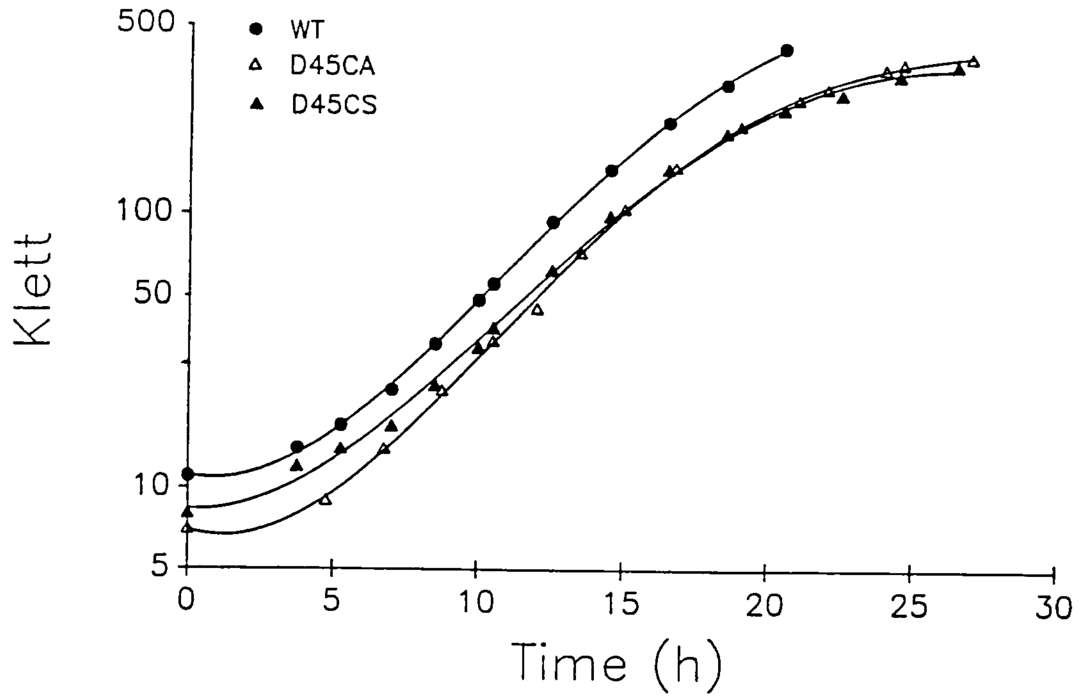


FIGURE 16. Diazotrophic growth of wild type A. vinelandii and mutant strains having substitutions for nonconserved cysteine residue α-cys-45.

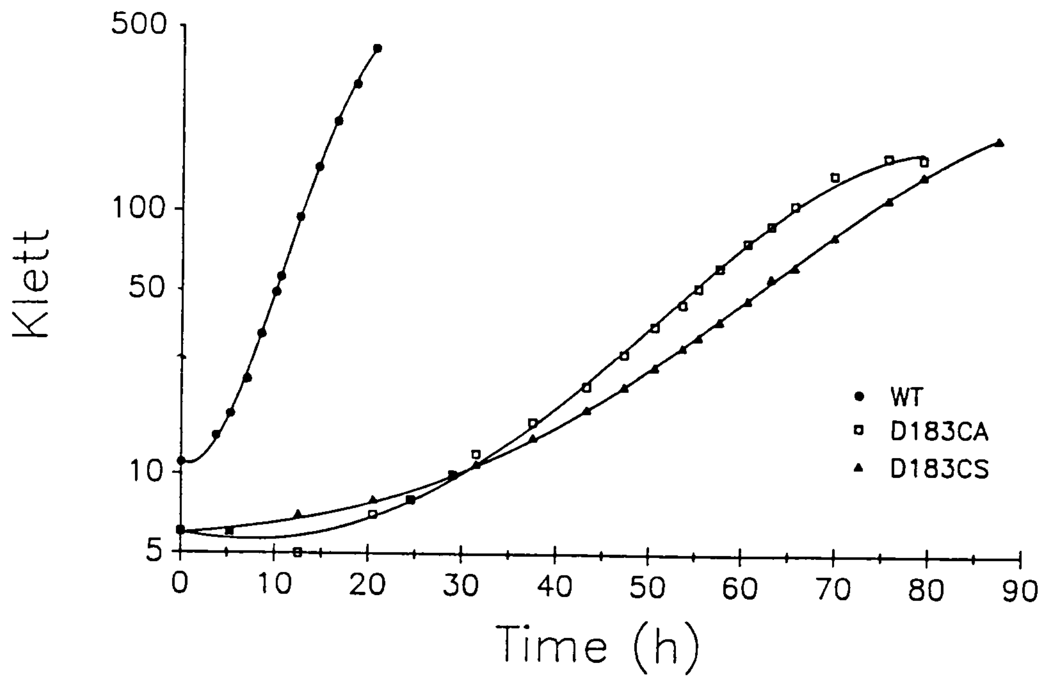


FIGURE 17. Diazotrophic growth of wild type A. vinelandii and mutant strains having substitutions for conserved cysteine residue α-cys-183.

very low but significant whole cell acetylene-reducing activities. These results suggest that α-cys-183 plays a role in maintaining the structure or stability of the MoFe protein.

The results for the substitution of conserved residue β-cys-153 were very curious. Substitution of this residue with ala resulted in a mutant strain (K153CA) which was incapable of diazotrophic growth and exhibited an extremely low level of whole-cell acetylene-reducing activity. However, replacement of β-cys-153 with ser resulted in a mutant strain (K153CS) which displayed good diazotrophic growth (Fig. 18) and a moderate level of whole cell acetylene-reducing activity. If this residue is a metallocluster ligand, these results suggest that the sulfur ligand of β-153-cys can be functionally replaced by the oxygen present in the ser residue in strain K153CS. To further analyze the function of this β-subunit residue, a mutant strain was constructed in which the β-cys-153 codon (TCG) was specifically deleted. Surprisingly, the resultant deletion strain (K▲153C) exhibited a diazotrophic growth rate (Fig. 18) and a whole cell acetylene-reducing activity which were nearly identical to those of K153CS. Therefore, if β-cys-153 does provide a ligand to one of the metalloclusters, this function is being replaced in K▲153C as well. One possible explanation is that the nucleophilic thioether group of the met residue which occupies position β-153 in K▲153C is providing the metallocluster ligand. This would be analogous to the ligand which met provides to the heme iron in cytochrome c (143). An alternative explanation is that one of the nearby oxygen-donor residues (β-ser-151, β-thr-152, β-thr-153, or β-glu-157) can provide a ligand

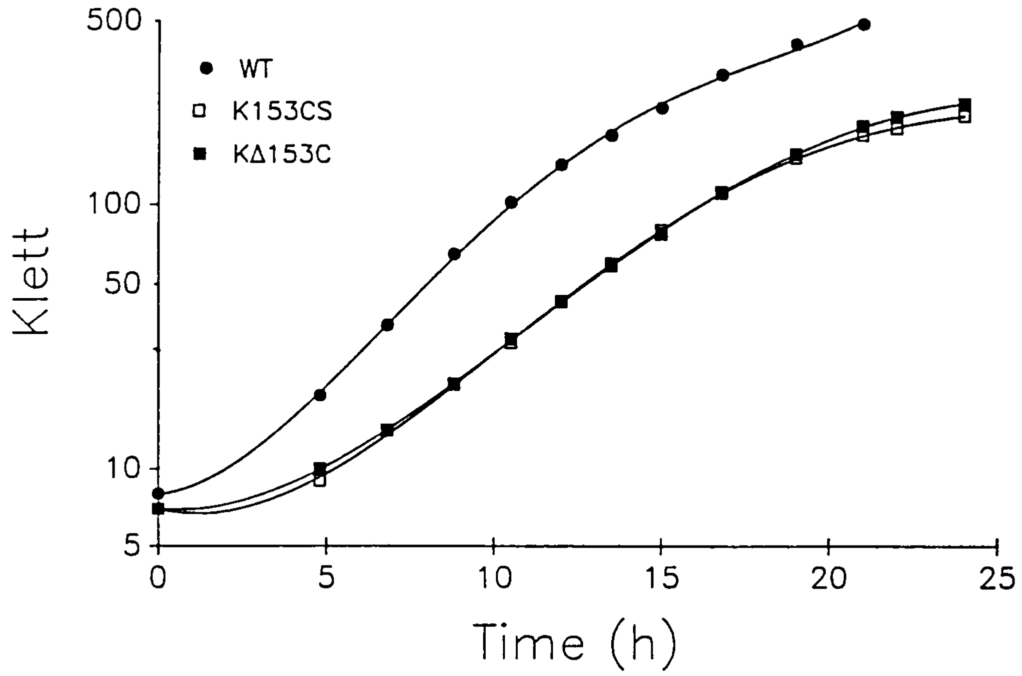


FIGURE 18. Diazotrophic growth of wild type A. vinelandii and mutant strains having substitutions for conserved cysteine residue **β-cys-153**.

to the metallocluster. If so, it appears that such a replacement is sterically hindered in K153CA. These results suggest that while there is significant spatial conservation and sequence identity surrounding **α-cys-154** and **β-cys-153** (region III from each subunit), these residues do not have strict functional equivalence within their respective subunits. Other substitutions which were made within α-subunit region III resulted in mutant strains with nearly normal diazotrophic growth capabilities (Fig. 19) and moderate to high levels of whole cell acetylene-reducing activity. The amino acid replacements in strains D151QE, D159IV, D160GA, and D161DE are essentially neutral but are within the amino acid sequence considered important for [Fe-S] cluster binding. These strains were constructed in an attempt to make very subtle alterations in the environment surrounding the metal cluster to which **α-cys-154** likely provides a thiol ligand. As such, spectroscopic analysis of these mutant strains is currently in progress.

Residues α-his-83 and **β-his-90** are conserved at the interspecies level (with the exception of the Anabaena β-subunit), conserved at the intersubunit level, conserved when the α- and β-subunits are compared to the respective nifE and nifN gene products, and conserved when the nifE and nifN gene products are compared to each other. Additionally, each **his** residue is located five residues preceding a conserved **cys** residue and each is located within a highly conserved region (region I) of its respective subunit. This conservation suggests that these **his** residues may have similar functional roles within their respective subunits, possibly as metallocluster ligands. Surprisingly, substitution of these **his** residues with **asn**

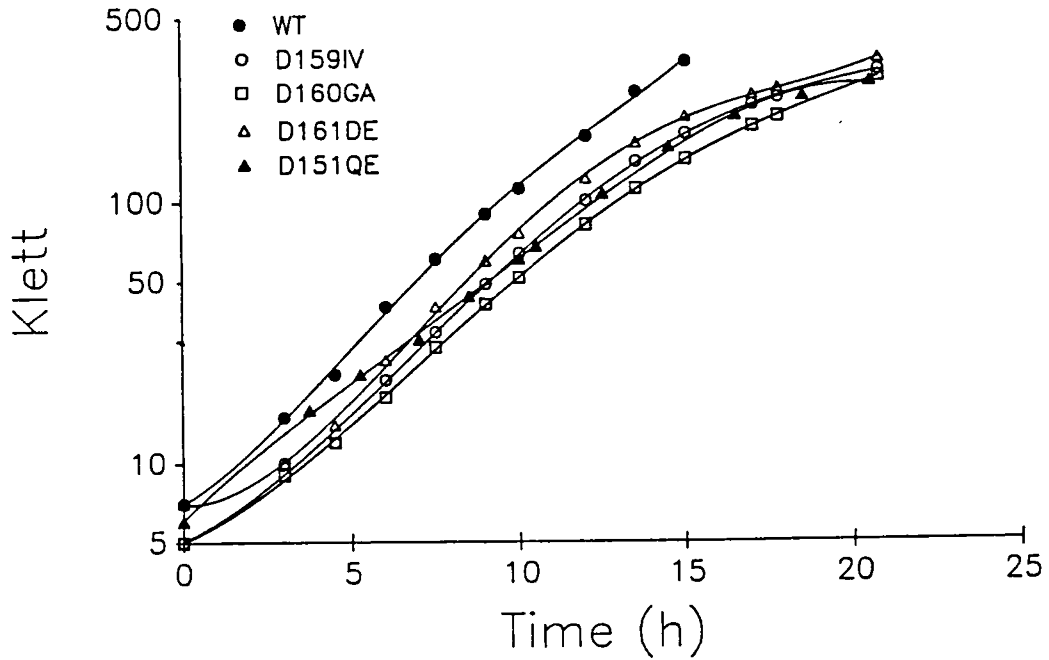


FIGURE 19. Diazotrophic growth of wild type A. vinelandii and mutant strains having substitutions for amino acid residues within α-subunit region III. The altered residues surround conserved cysteine residue **α-cys-154** and are within the sequence considered important for Fe-S cluster binding in bacterial ferredoxins.

resulted in mutant strains (D88HN and K90HN) with fairly high levels of whole cell acetylene-reducing activity and only slightly diminished diazotrophic growth capabilities (Fig. 20). In general, these mutant strains exhibited longer lag times but nearly identical growth rates compared to the wild type strain. These results suggest that either these **his** residues are not metallocluster ligands or that **asn** (or a nearby residue) can functionally replace either of the residues. Since each of these residues is predicted to be within a β-turn and therefore at the protein surface, it is possible that these charged **his** residues may be involved in protein-protein interaction. As such, the effects of eliminating one charged **his** residue may be offset by the presence of the other. To examine this possibility, we attempted to construct a double mutant in which both of these **his** residues are replaced by **asn**. To do so, a mutant strain which contains the **α-asn-83** substitution and which is deleted for the N-terminal region of nifK was constructed (DJ222). We attempted to rescue the Nif$^-$ phenotype of this strain by transforming with the DNA (plasmid Mn34RF - see Table 4) that was used to construct mutant strain K90HN. While a positive control (transformation with wild type DNA) yielded Nif$^+$ transformants, transformation with plasmid Mn34RF produced none. This result shows that concurrent substitution of both **α-83-his** and **β-90-his** with **asn** results in a strict Nif$^-$ phenotype. Understanding the physical basis for this result will depend upon biochemical analyses of the individual **asn** substituted mutant strains.

In addition to substitutions for **α-183-cys**, other amino acid changes within α-subunit region IV produced interesting results. Substitution of **α-191-gln** with **glu**

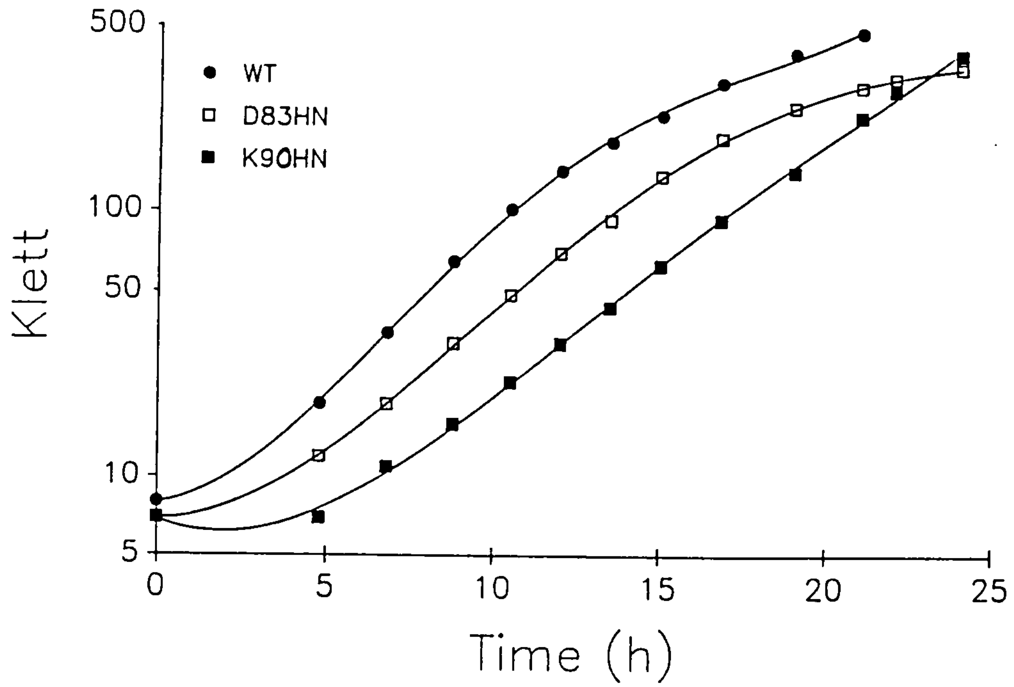


FIGURE 20. Diazotrophic growth of wild type A. vinelandii and mutant strains having substitutions for highly conserved histidine residues present within region I of both subunits.

resulted in a mutant strain (D191QE) which was incapable of diazotrophic growth yet retained a substantial level of whole cell acetylene-reducing activity. In contrast, strains D183CA and D183CS were capable of diazotrophic growth but exhibited much lower whole cell acetylene-reducing activities. Substitution of **α-191-gln** with **lys** (the residue present in the nifE gene product) resulted in a mutant strain (D191QK) which was also incapable of diazotrophic growth and exhibited whole cell acetylene-reducing activity similar to strains D183CA and D183CS. Such results indicate that these four mutant strains have different substrate binding or substrate reducing capabilities which implicates α-subunit region IV as a potential site of FeMoco interaction. The importance of this region is emphasized by the results obtained by substituting **α-195-his** with **asn**. The resultant mutant strain (D195HN) was incapable of diazotrophic growth and displayed an extremely low level of whole cell acetylene-reducing activity.

We had initially suggested that α-subunit region V (**α-cys-275** and the residues which surround it) contained residues likely to be important for binding FeMoco (11). This suggestion was based on the observation that several conserved nitrogen-donor residues (α-asn-271, **α-his-274**, **α-asn-280**) which might serve as ligands to FeMoco were present around **α-cys-275**, itself a potential FeMoco thiol ligand. However, mutagenesis of this region yielded results which do not fully support this original idea. Substitution of potential nitrogen-donor residues **α-asn-271** and **α-asn-280** with **asp** resulted in mutant strains (D271ND and D280ND) fully capable of diazotrophic growth (Fig. 21). While strain D271ND displayed growth

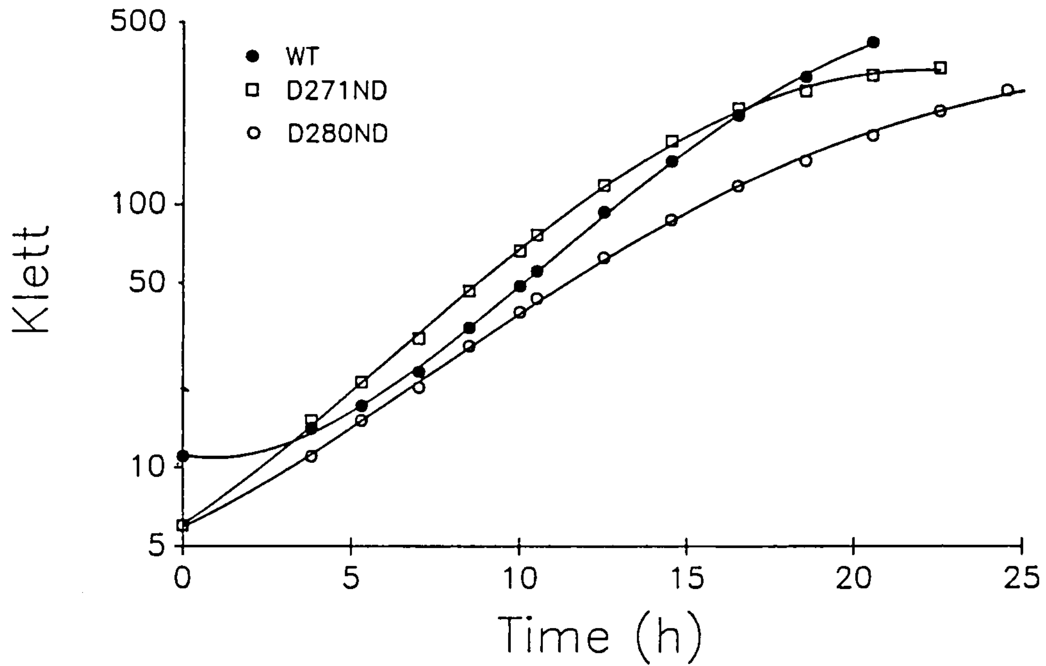


FIGURE 21. Diazotrophic growth of wild type A. vinelandii and mutant strains having substitutions in conserved nitrogen-donor residues within α-subunit region V. The altered residues flank conserved cysteine residue α-cys-275.

characteristics nearly identical to wild type, strain D280ND exhibited a slower log phase growth rate. This observation correlates with the lower level of whole cell acetylene-reducing activity displayed by strain D280ND. While **α-asn-280** appears to have some role in the proper functioning of nitrogenase, it is clear that neither of these **asn** residues provides an essential nitrogen ligand to the FeMoco cluster. Two other substitutions in this region, **ser** for **α-tyr-276** and **ile** for **α-met-279**, resulted in mutant strains (D276YS and D279MI) with nearly normal diazotrophic growth characteristics (Fig. 22) and fairly high levels of whole cell acetylene-reducing activity. Since α-tyr-276 was nonessential, we attempted to rescue the Nif⁻ phenotype of strain D275CS by replacing **α-tyr-276** with **cys**. However, the resulting double mutant was Nif⁻ and displayed no whole cell acetylene-reducing activity implying that the steric requirements of this potential thiol ligand are quite stringent. With the exception of **α-cys-275**, the only essential, targeted residue in this region was α-arg-277. Substitution of this **arg** with **his** resulted in a mutant strain (D277RH) that was incapable of diazotrophic growth and exhibited only a barely detectable level of whole cell acetylene-reducing activity.

Working with MoFe protein **cys → ala** site-directed mutants of K. pneumoniae, Smith et al. (157) have provided indirect evidence that **α-cys-275** is a ligand to FeMoco. Compared to crude extracts of other **cys**-substituted mutant strains, a crude extract of strain D275CA has a significantly greater pool of FeMoco which is available to reconstitute apo-MoFe protein isolated from NifB mutants. They

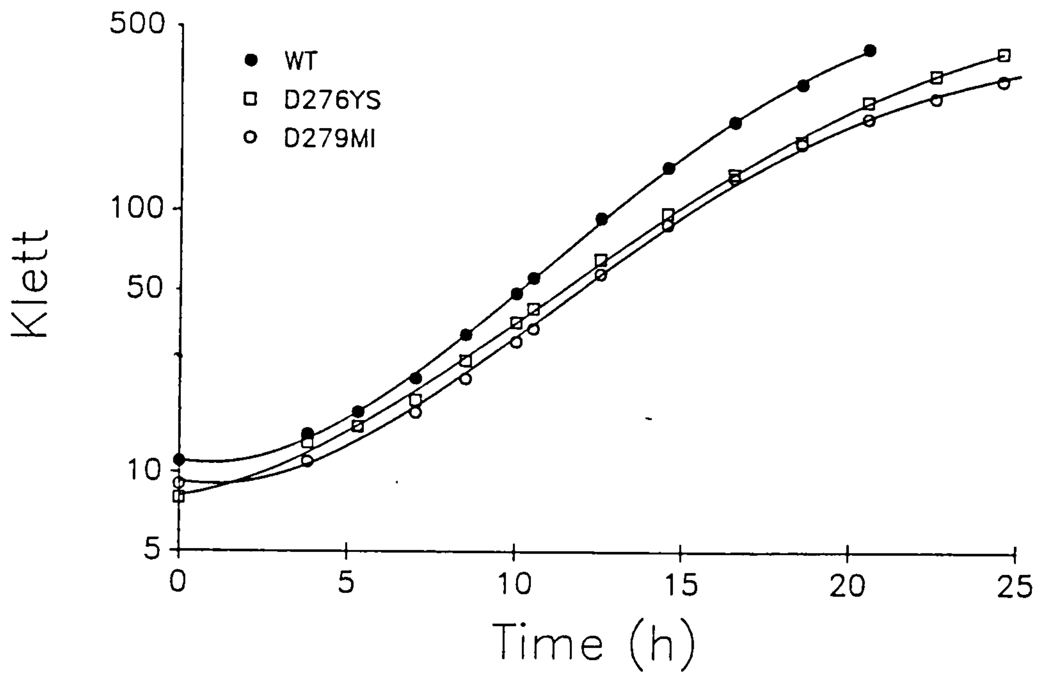


FIGURE 22. Diazotrophic growth of wild type A. vinelandii and mutant strains having substitutions for amino acid residues within α-subunit region V near conserved cysteine residue α-cys-275.

suggest that the FeMoco in mutant strain D275CA is less tightly bound to the MoFe protein due to the missing **α-cys-275** thiol ligand.

A Model for Sites of Metallocluster Binding within Nitrogenase.

Based on protein sequence data, cluster extrusion requirements, spectroscopic analyses, and the results of site-directed mutagenesis experiments, we have formulated a preliminary model (Fig. 23) for the binding of the metal clusters within the MoFe protein. The basic features of this model are: (1) Regions I and III from both MoFe protein subunits form the P-cluster binding domains; (2) Regions IV and V from the α-subunit form the FeMoco binding domain; (3) The P-clusters are located near the MoFe protein surface and the cofactor is contained entirely within the α-subunit.

Since spectroscopic data suggests that the P-clusters are located within similar protein environments, it is reasonable to expect them to be contained within regions of the MoFe protein subunits which have similar sequence and structure. In this respect, α- subunit regions I, II, and III share similar structural features with these respective regions from the β-subunit. Combined, these regions contain six conserved cysteine residues (**α-cys-62, α-cys-88, α-cys-154, β-cys-70, β-cys-95,** and **β-cys-153**) making it possible to assign four similar cluster-binding domains. The cysteine residues account for twelve thiol ligands in the native MoFe protein and therefore, each cluster is assumed to be coordinated by one non-cysteinyl ligand. This assumption is supported by the unusual spectroscopic features of the P-clusters.

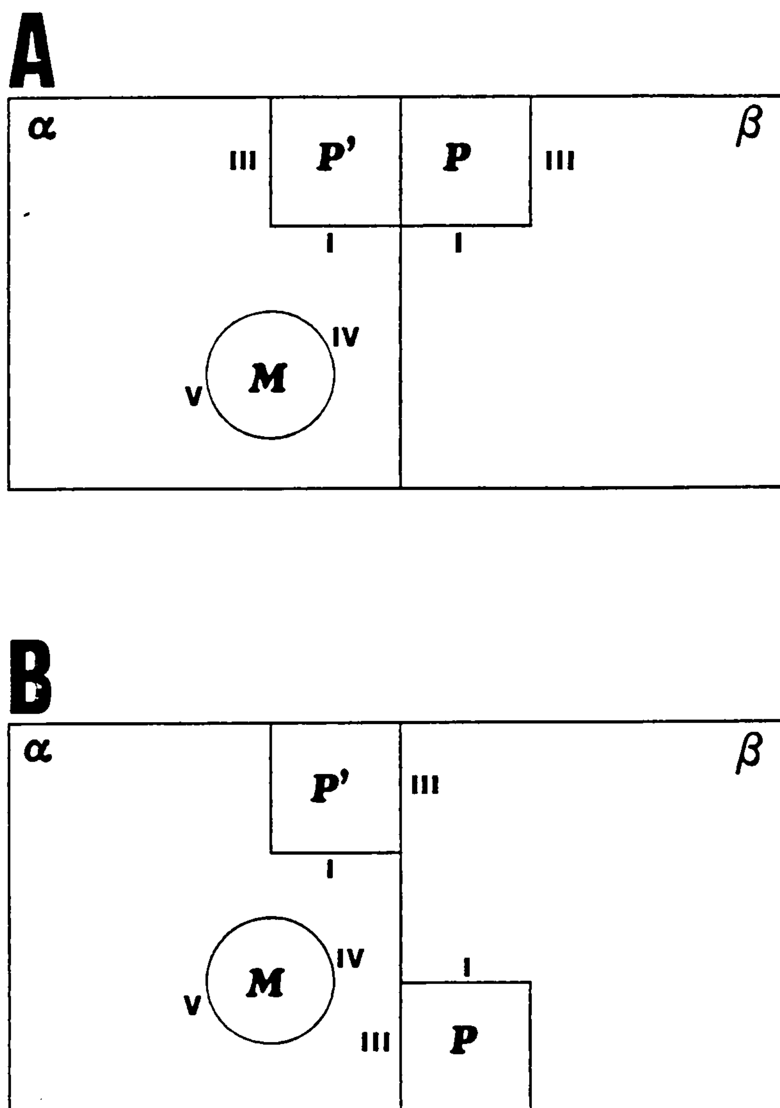


FIGURE 23. Proposed models for the sites of metallocluster binding within nitrogenase. Each model represents only one half of the nitrogenase enzyme. Roman numerals designate MoFe protein α- and β-subunit regions as defined in Figs. 6 and 7. P and P' designate two slightly inequivalent [4Fe-4S] clusters, and M designates a FeMoco cluster. In Model A, each P-cluster is coordinated only by residues from within a single subunit. In Model B, each P-cluster is bridged between the two subunits with ligands to a single cluster being provided by residues from both subunits.

Highly conserved residues **α-his-83** and **β-his-90**, both located five residues proximal to a conserved cysteine residue, were originally considered likely candidates to provide cluster ligands. However, mutant strains **D83HN** and **K90HN** remain fully capable of diazotrophic growth (Table 6 and Figure 20) suggesting an alternative role for these residues. The predicted near surface location of these charged residues may indicate a role in protein-protein interaction, possibly for proper orientation of a nearby metallocluster. The normal diazotrophic growth capability of strain K153CS (Fig. 18) suggests that the P-clusters could be coordinated by an oxygen-donor ligand. Additionally, it is interesting that a displaced methionine residue in strain K▲153 may be acting to provide a metallocluster ligand. As such, methionine residues α-met-245 and β-met-247 can be considered possible cluster ligands. These residues are conserved at the interspecies level and can be considered conserved when the α- and β-subunits are compared (Fig. 9). There are two possible configurations for the assigned P-cluster binding domains (Fig. 23): (A) A cluster could be contained entirely within a single subunit, coordinated only by residues from that specific subunit; (B) A cluster could be bridged between the two MoFe protein subunits, with ligands to a single cluster being provided by residues from both subunits. Each of these arrangements predicts a surface location for the clusters and allows for the clusters to be in close proximity and capable of interaction. Additionally, these configurations accommodate the suggestion by Hagen et al. (48) that the four P-clusters are actually arranged as two larger eight-iron clusters.

By designating the conserved cysteine residues in α- and β-subunit regions I and III as P-cluster ligands, only **α-cys-183** and **α-cys-275** remain as potential thiol ligands to FeMoco. Of these two, located in regions IV and V respectively, only **α-cys-275** is absolutely essential for nitrogenase activity. This result, plus the observation that crude extracts of K. pneumoniae strain D275CA have a large pool of unbound FeMoco (157), supports the idea that **α-cys-275** provides a thiol ligand to FeMoco. However, mutagenesis of the amino acids surrounding **α-cys-275** provides results which indicate that several nitrogen-donor residues within this region are not essential cofactor ligands as we had previously suggested (11). Mutagenesis of residues in α-subunit region IV gives results which indicate that residue **α-cys-183** and surrounding amino acids are within the immediate environment of the cofactor in the native protein. Particularly interesting are the altered substrate utilization patterns exhibited when comparing mutant strains D183CA, D183CS, D191QE, D191QK, and D195HN. Residues in this region may be providing the nitrogen-donor ligands which are thought to coordinate FeMoco within the native protein. Possible candidates include conserved residues **α-191-gln**, **α-195-his**, **α-196-his**, and **α-199-asn**. While none of these four residues is conserved when the α-subunit and the nifE gene product are compared, this is not unexpected. While there would be similarities between the FeMoco binding domains of these two proteins, the cofactor must eventually 'escape' the nifEN-encoded complex and

therefore would not require as many coordinating ligands as compared to the MoFe protein.

The configurations of this preliminary model are difficult to reconcile with the low resolution (8 A) crystallographic data which predicts all of the metal clusters to be within the MoFe protein α-subunits (see 32 in reference 157). A higher resolution crystallographic structure would provide valuable information regarding potential metallocluster ligands and would enable us to fine tune our site-directed mutagenesis studies. In lieu of such information however, we will continue to use our working model as a basis for targeting additional residues likely to be important in nitrogenase function. Our collaborators are presently examining the catalytic, redox, and spectroscopic properties of MoFe protein purified from mutant strains which we have already constructed. These studies should provide valuable insight into the catalytic and redox functions of the individual metalloclusters during biological nitrogen reduction.

Additionally, the site-directed mutagenesis and gene replacement strategy developed during the course of this study is now routinely used in our laboratory to place specific substitutions, deletions, or insertions into any gene product encoded within the A. vinelandii nif-gene cluster.

GENETIC ANALYSIS OF REGIONS FLANKING THE nifE AND nifN GENES

After determining the nucleotide sequence of the A. vinelandii FeMoco biosynthetic genes nifE and nifN, it was a logical progression to continue sequence analysis of the nif cluster in an effort to complete the region between the nifK and nifE genes (Fig. 2). In addition, we continued our sequence analysis in the region immediately downstream of the nifE and nifN genes. In the K. pneumoniae nif cluster, these two regions had been proposed to encode two nif specific genes, nifY and nifX, respectively. Puhler and Klipp (128) predicted the presence of these genes based upon the production of two proteins (M_r = 24,000 and 18,000, respectively) in pulse labeled E. coli mini cells containing cloned K. pneumoniae nif fragments. These studies indicated that nifY was part of the nifHDK transcriptional unit and that nifX was part the nifEN transcriptional unit.

In this study, I completed the sequence analysis of these regions (nucleotides 5520-7700 and 10580-12567 in Appendix XI) from the A. vinelandii genome. This analysis confirmed the presence of nifX and nifY within the A. vinelandii nif cluster and predicted the existence of a new nif gene, nifT, located immediately preceding nifY (Appendix III). In addition, further analysis of these regions revealed the presence of four potential ORF's (ORF1, ORF2, ORF3, and ORF4) which have no homologs to products encoded within the K. pneumoniae nif gene cluster (Appendix III). I genetically analyzed several of these genes by introducing specific deletions

and lac fusions into the A. vinelandii chromosome. Finally, I was able to overproduce the ORF3 gene product in E. coli.

Sequence Comparisons.

In Appendix III (Fig.2), the gene products of the A. vinelandii and K. pneumoniae nifT, nifY, and nifX genes are aligned. The nifT encoded gene products show significant sequence identity (42%) dispersed throughout the length of the polypeptides. The nifY encoded gene products share a sequence identity of only 19% with the majority concentrated in the C-terminal half of the polypeptides. Comparison of the nifX-encoded sequences reveals a sequence identity of only 29% with the majority confined to the initial two thirds of the polypeptides. While the molecular weights of the compared proteins are very similar, their calculated pI values are quite different (Table 7). In addition, the predicted secondary structures and hydrophobic indices of the compared proteins are also dissimilar (data not shown). The different physical characteristics (limited sequence identity, pI) observed for the A. vinelandii nifT, nifY and nifX gene products when compared to the K. pneumoniae nifT, nifY and nifX gene products may indicate that the respective functions of these proteins are different within the two organisms. Alternatively, the different physical characteristics may have no effect on the function of these proteins if compensatory changes have occurred within the cellular components with which these gene products interact.

TABLE 7. Features of the nifT, nifY, and nifX gene products of A. vinelandii and K. pneumoniae.

nif gene	Product molecular mass (daltons) Av	Kp	Calculated product pI[a] Av	Kp	% identity
T	8,038	8,245	4.61	9.44	42
Y	26,683	24,671	10.60	8.60	19
X	17,270	18,209	4.44	7.41	29

[a] Due to an updated Pustell program which uses a different algorithm, pI values for A. vinelandii gene products are slightly different than those given in Appendix III - Table 2.

In each organism, the nifX encoded gene products share significant sequence identity with its respective nifY encoded gene products (Appendix III - Fig. 5). However, the residues which are conserved between the K. pneumoniae gene products differ from those conserved between the A. vinelandii gene products. Therefore, although it appears that nifY and nifX could have common evolutionary origins (as proposed for nifDK and nifEN), their present functions are probably very different.

While none of the ORF's identified from this region have a structural counterpart in the K. pneumoniae nif cluster, two of them (ORF2 and ORF3) have been shown to be very similar to nif-associated genes identified in R. capsulatus (98, 109). Interestingly, ORF2 encodes a periodic protein which contains a twenty-four amino acid motif that is repeated seven times (71). Jacobson et al. (71) have suggested that this protein is involved in molybdenum sensing or processing. The conservation of ORF2 and ORF3 between two diverse nitrogen fixing organisms and their presence within the nif cluster suggests that the products of these genes may indeed have a physiological role in nitrogen fixation.

Gene Expression.

Because there are no nif-consensus or typical E. coli-type promoters immediately preceding nifX or ORF3, transcription of these genes is likely directed by the nifEN promoter. As such, nifX and ORF3 expression should respond to a physiological demand for fixed nitrogen. To assay for their expression, nifX:lacZ (DJ168) and

ORF3:lacZ (DJ260) translational fusion strains were constructed (Fig. 24) and assayed for β-galactosidase activity under both nitrogenase repressing and derepressing conditions. Derepressed cultures of strain DJ168 displayed a four fold increase in β-galactosidase activity compared to repressed cultures, while derepressed cultures of strain DJ260 displayed a fourteen fold increase in β-galactosidase activity. A kanamycin-resistance-encoding cartridge was inserted into the nifA gene in both of these lac fusion strains. β-galactosidase activity in the resultant mutant strains (DJ184 and DJ301) was identical in both repressed and derepressed cultures. Thus, the expression of both genes is under control of the nifA gene product (positive activator of nif gene expression) and accumulation of their gene products is elevated in response to a demand for fixed nitrogen. Note that the increase in nifX expression was 3.5 times less than that for **ORF3** gene expression. While transcription of both genes is apparently directed by same promoter, translational coupling of nifX and nifN may be responsible for its decreased expression relative to the expression of ORF3. Results obtained from attempts to overproduce the **ORF3** gene product support this idea (see below).

Mutagenesis of nifY, nifX, and ORF3.

As shown in Fig. 24, deletion strains were constructed in which specific in-frame deletions were placed within nifX (DJ166), **ORF3** (DJ198), and a double deletion of both nifX and **ORF3** (DJ44). A specific in-frame deletion of nifY (DJ208) was constructed by oligonucleotide-directed mutagenesis in which an

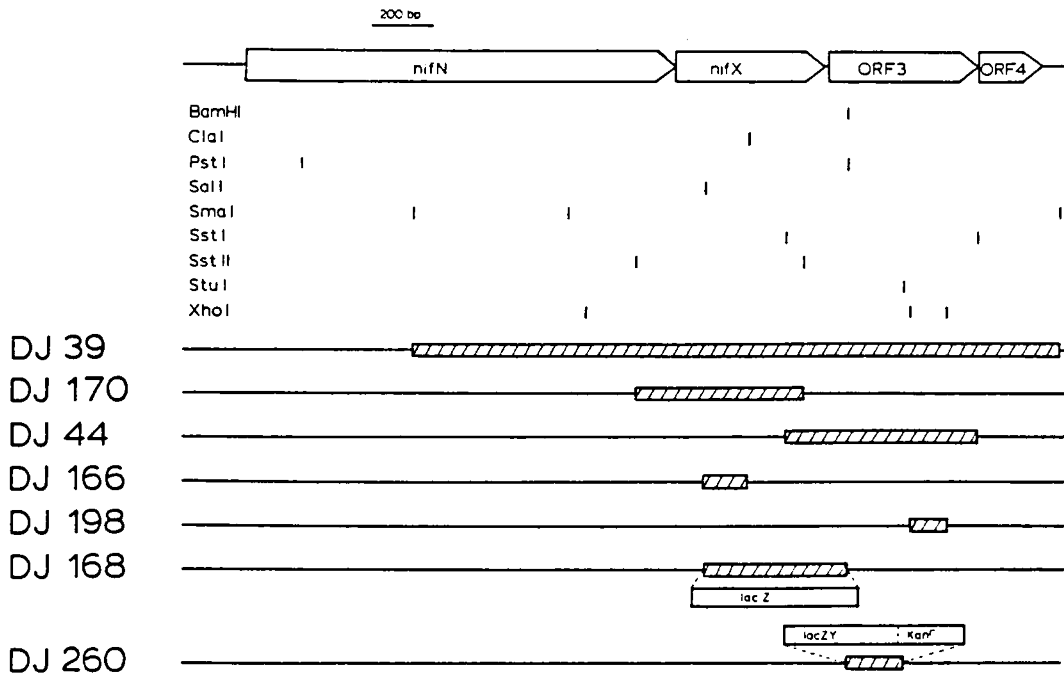


FIGURE 24. Physical map of the A. vinelandii nif cluster surrounding nifX and construction of A. vinelandii strains used to analyze genes within this region. Strain DJ44 contains an in-frame SstI deletion which removes the terminal 42 amino acids of the nifX gene product and all of ORF3. Strain DJ166 contains an in-frame HincII-ClaI deletion within the nifX gene which removes amino acid residues 33 - 76. Strain DJ198 contains an in-frame XhoI within ORF3 which removes amino acid residues 86 - 132. Strain DJ168 carries an in-frame translational fusion of the lacZ gene of E. coli to the 32nd codon of the nifX gene. Strain DJ260 carries an in-frame translational fusion of the lacZ gene of E. coli to the 19th codon of ORF3. Strains DJ44, DJ198, and DJ260 were constructed by transforming strain DJ39 (Nif^-) to the Nif^+ phenotype. Strains DJ166 and DJ168 were constructed by transforming strain DJ170 (Nif^-) to the Nif^+ phenotype. In the figure, thin lines represent A. vinelandii DNA sequences and hatched boxes indicate deleted regions. See references 88 and 154 for descriptions of the lacZ fusion cassette vectors.

additional Sma site was added to the nifY DNA sequence (by changing nucleotide 6177 C to 6177 G). This allowed for the removal of a nifY specific Sma fragment encoding amino acid residues 145 through 202. Mutant strain DJ208 (▲Y) had growth characteristics identical to wild type, while mutant strains DJ166 (▲X), DJ198 (▲3), and DJ44 (▲X3) displayed only slightly diminished diazotrophic growth capabilities (Figure 25 and Table 8).

Biochemical Analysis.

Because none of these mutant strains exhibited any remarkable outward phenotypic differences from wild type, we examined crude extracts of these strains in an attempt to detect any subtle differences in their respective nitrogenase component proteins. Upon addition of increasing amounts of Fe protein to crude extracts, the specific activity of the MoFe protein of DJ44 (▲X3) was elevated significantly over that of wild type (Fig. 26). With the addition of increasing amounts of MoFe protein to the crude extracts, the specific activity of the Fe protein of DJ44 (▲X3) decreased more rapidly than that of wild type. These results suggest that DJ44 has accumulated an excess of MoFe protein in comparison to wild type. As shown by the MoFe protein activities in Table 8, this effect of overproducing MoFe protein appears to be specific to strains having deletions in ORF3. In general, the growth rates exhibited by these strains are inversely proportional to their MoFe protein activities. This correlation is not unexpected based on the observation that

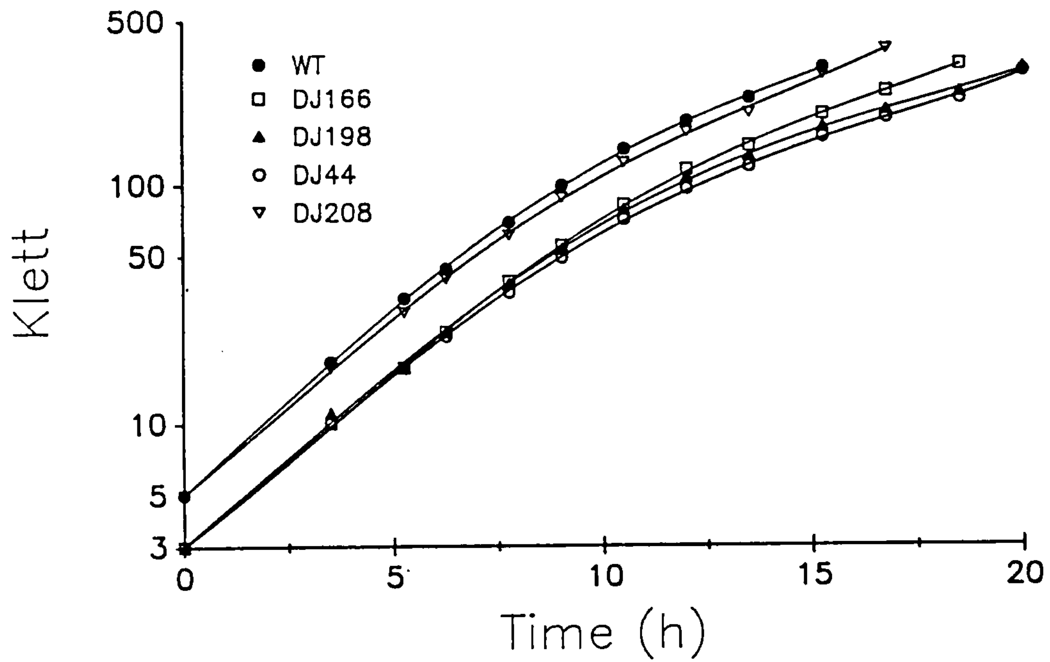


FIGURE 25. Diazotrophic growth of wild type and mutant strains having deletions in nifY, nifX, and ORF3.

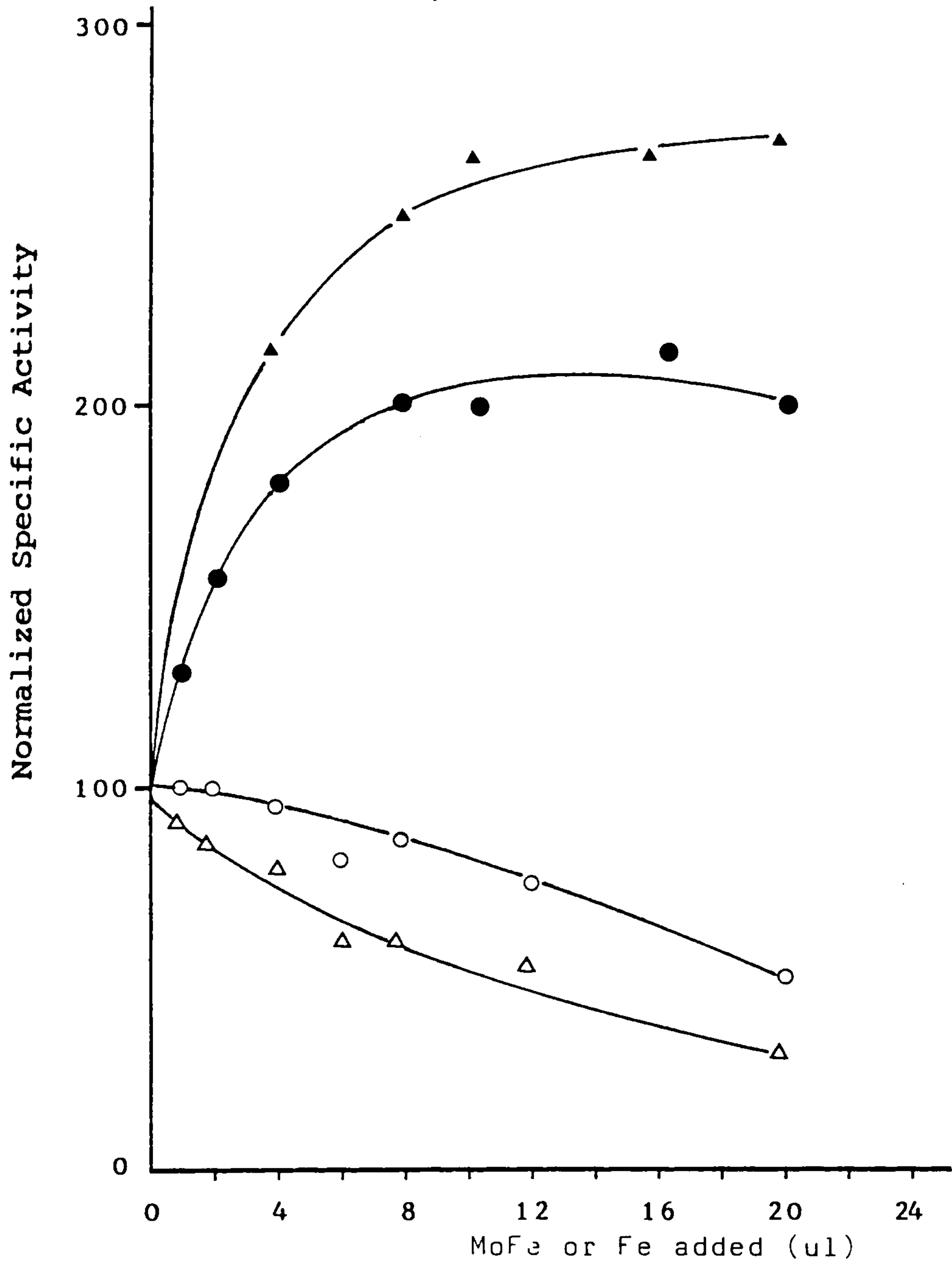


FIGURE 26. Normalized specific activity (nmol ethylene formed per min per mg of crude extract) of MoFe protein [filled symbols] and of Fe protein [open symbols] in wild type and mutant strain DJ44 (▲X3).
●, wild type - MoFe protein; ▲, DJ44 - MoFe protein; o, wild type - Fe protein; △, DJ44 - Fe protein.

TABLE 8. Growth characteristics and component protein activities of wild type and mutant strains having deletions in nifY, nifX, and **ORF3.**

Strain	Doubling Time (h)	Acetylene-reduction activity Crude Extract Specific Activities[#] Fe[a]	MoFe[b]
Wild type	2.50	20	52
DJ166 (▲X)	2.75	19	50
DJ198 (▲3)	3.00	18	79
DJ44 (▲X3)	3.25	nd	63
DJ208 (▲Y)	2.50	nd	nd

[#]Specific activity: nmol of ethylene formed per min per mg of crude extract protein.
[a]Determined in the presence of saturating levels of purified A. vinelandii MoFe protein.
[b]Determined in the presence of saturating levels of purified A. vinelandii Fe protein.
nd = not determined.

in vitro, nitrogenase turnover decreases as the MoFe/Fe protein ratio increases (184). Therefore, under the growth conditions used in this study, it appears that the ORF3 gene product plays a role in regulating expression of the nifD and nifK genes. It is unknown if this is a direct effect of the ORF3 gene product or one which is mediated through the presence of another (nif) gene product. An attractive suggestion involves a direct binding of the ORF3 gene product to the potential secondary structure predicted to form in the nifH - nifD intercistronic region (11).

Overexpression of the ORF3 Gene Product in E. coli

Obtaining large quantities of the ORF3 encoded gene product is an important step in the purification and characterization of the protein. As shown in Fig. 27, a series of plasmids was constructed in an attempt to specifically overproduce the ORF3 gene product using the IPTG inducible tac promoter. Overexpression of the ORF3 gene product was complicated by the presence of the adjacent nifX gene which encodes a product with a molecular mass nearly identical to the ORF3 encoded protein (Appendix III - Table 2). As shown by the constructs in Fig. 27 and PAGE analysis of induced, transformed cells in Fig. 28 (Panels A and C), only those plasmids which contain an intact ORF3 gene (pDB91 and pDB246) were capable of expressing the 18 kDa protein. Cells transformed with pDB240, which contains nifX but not ORF3, did not express this protein (Fig. 28 - Panel A). Additionally, the absence of the 18 kDa protein from cells transformed with pDB92

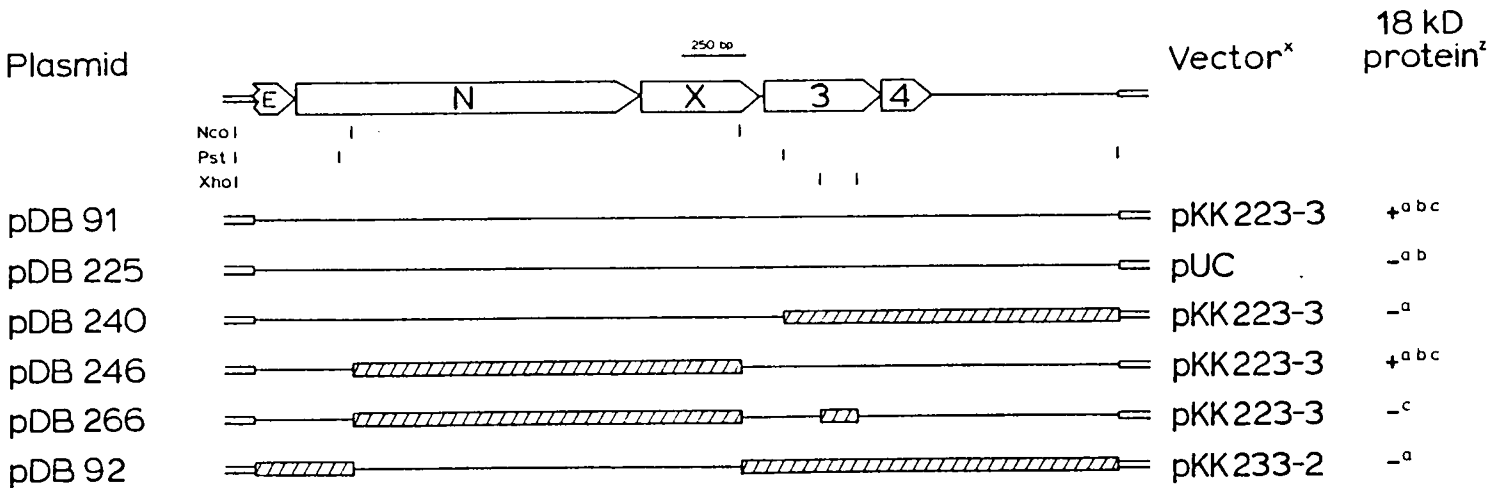
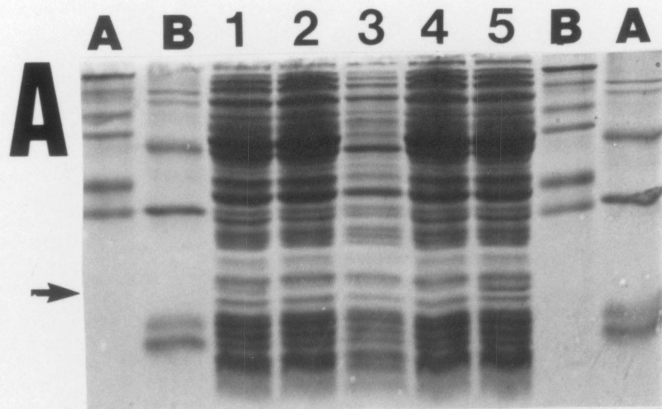
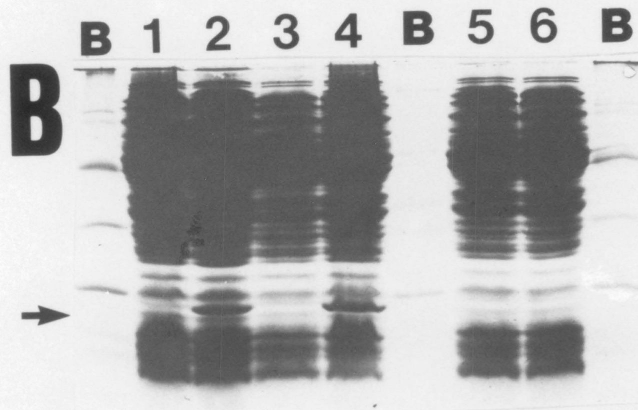
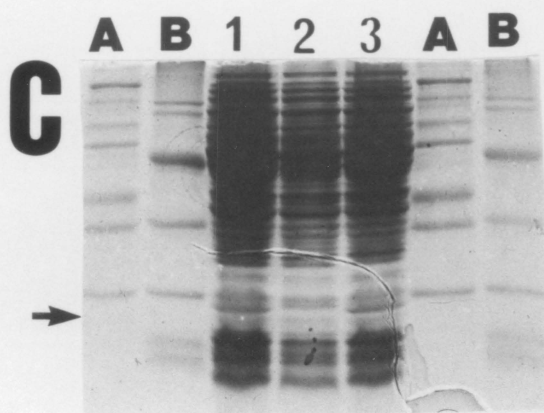


FIGURE 27. Physical map of the A. vinelandii nif cluster surrounding nifX and construction of plasmids used to overproduce the **ORF3** gene product in E. coli. Thin lines represent A. vinelandii DNA sequences and hatched boxes indicate deleted regions.
X: pKK223-3 and pKK233-2 utilize the tac promoter and pUC utilizes the lac promoter.
Z: letters refer to the Panels in Fig. 28.

FIGURE 28. Expression of the ORF3 gene product in E. coli JM105. Cells were transformed with plasmids shown in Fig. 29, induced with IPTG for 2.5 h and electrophoresed through 12% SDS gels. The arrow points to the 18 kd protein. PANEL A: lane 1, pDB91; lane 2, pDB92; lane 3, pDB225; lane 4, pDB240; lane 5, pDB246. PANEL B: lane 1, pDB91 uninduced; lane 2, pDB91 induced; lane 3, pDB246 uninduced; lane 4, pDB246 induced; lane 5, pDB225 uninduced; lane 6, pDB225 induced. PANEL C: lane 1, pDB91; lane 2, pDB266; lane 3, pDB246. In all Panels, lanes A and B are the following molecular weight markers: A: 180,000; 116,000; 84,000; 58,000; 48,500; 36,500; 26,600 and B: 77,000; 66,250; 43,000; 30,000; 17,200; 12,300. In Panels A and B, molecular weight markers B also contain A. vinelandii flavodoxin (19,600).


A B 1 2 3 4 5 B A
A


B 1 2 3 4 B 5 6 B
B


A B 1 2 3 A B
C

or pDB240 indicates that this protein is not a spurious nifN or nifN:nifX encoded polypeptide (Fig. 28 - Panel A). Finally, cells transformed with pDB266, an ORF3 specific deletion derivative of pDB246, did not express the 18 kDa protein (Fig. 28 - Panel C). These results strongly suggest that the overproduced 18 kDa protein is indeed the ORF3 gene product. [Plasmid pDB225, which is identical to pDB91 except that it is driven by the lac promoter, was unable to direct synthesis of the ORF3 gene product (Fig. 28 - Panels A and B). This difference may be due to the fact that lac promoter is weaker than the tac promoter.] Expression of the ORF3 gene product (pDB91 and pDB246) was dependent upon induction with IPTG (Fig. 28 - Panel B) which indicates that this region contains no recognizable E. coli-type promoter sequences. It is interesting that cells transformed with pDB240 are unable to express the nifX gene product (Fig. 28 - Panel A). In this plasmid construction, an out-of-frame initiation codon prevents expression of the nifN gene product. Taken together, this suggests that translation of the nifX gene product is coupled to and dependent upon translation of the nifN gene product.

Summary.

I have determined the nucleotide sequence of the regions flanking the nifE and nifN genes of A. vinelandii and have identified three nif genes (nifT, nifY, and nifX) and four open reading frames (ORF1, ORF2, ORF3, and ORF4) from this analysis. Both nifX and ORF3 gene expression were shown to be under nif regulation and synthesis of their gene products elevated in response to a demand for fixed nitrogen.

Strains with deletions in ORF3 appeared to accumulate excess amounts of MoFe protein when compared to wild type. Future studies utilizing wild type and ORF3 mutant strains containing nifD:lacZ and nifK:lacZ transcriptional and translational constructs (see Appendix VI) should help define the mechanism by which this occurs. Additionally, the ORF3 gene product has been overproduced in E. coli. This provides an important first step in characterizing the protein and affords an additional approach toward elucidating the molecular basis for its control of nifDK gene expression.

LITERATURE CITED

1. Arnold, W., A. Rump, W. Klipp, U. B. Priefer, and A. Puhler. 1988. Nucleotide sequence of a 24,206-base-pair DNA fragment carrying the entire nitrogen fixation gene cluster of Klebsiella pneumoniae. J. Mol. Biol. 203:715-738.

2. Arnott, M., C. Sidoti, S. Hill, and M. Merrick. 1989. Deletion analysis of the nitrogen fixation regulatory gene, nifL of Klebsiella pneumoniae. Arch. Microbiol. 151:180-182.

3. Bali, A., S. Hill, E. Santero, A. Toukdarian, J. Walmsley, and C. Kennedy. 1988. NifA and NtrC activate separate nitrogen fixation pathways in Azotobacter. In Nitrogen Fixation: Hundred Years After. H. Bothe, F. J. de Bruijn, and W. E. Newton (eds) 316.

4. 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.

5. 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.

6. 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.

7. 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.

8. Bishop, P. E., D. M. L. Jarlenski, and D. R. Hetherington. 1980. Evidence for an alternative nitrogen fixation system in Azotobacter vinelandii. Proc. Natl. Acad. Sci. USA 77:7342-7346.

9. Bishop, P. E., R. Premakumar, D. R. Dean, M. R. Jacobson, J. R. Chisnell, T. M. Rizzo, and J. Kopczynski. 1986. Nitrogen fixation by Azotobacter vinelandii strains having deletions in structural genes for nitrogenase. Science 232:92-94.

10. Brigle, K. E., J. Beynon, W. Newton, H. May, and D. Dean. 1988. The isolation, sequence analysis, and mutagenesis of the nifX region from Azotobacter vinelandii and Klebsiella pneumoniae. In Nitrogen Fixation: Hundred Years After. H. Bothe, F. J. de Bruijn, and W. E. Newton (eds) 320.

11. Brigle, K. E., W. E. Newton, and D. R. Dean. 1985. Complete nucleotide sequence of the Azotobacter vinelandii structural gene cluster. Gene 37:37-44.

12. Bruschi, M. and F. Guerlesquin. 1988. Structure, function and evolution of bacterial ferredoxins. FEMS Microbiol. Rev. 54:155-176.

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. Bulen, W. A. and J. R. LeComte. 1966. The nitrogenase system from Azotobacter: Two-enzyme requirement for N_2 reduction, ATP-dependent H_2 evolution, and ATP hydrolysis. Proc. Natl. Acad. Sci. USA 56:979-986.

15. Burgess, B. K. 1985. Nitrogenase mechanism: An overview. In Current Plant Science and Biotechnology in Agriculture: Nitrogen Fixation Research Progress. H. Evans, P. Bottomley, and W. E. Newton (eds) 543-549.

16. Burgess, B. K. and W. E. Newton. 1983. Iron-molybdenum cofactor and its complementary protein from mutant organisms. In Nitrogen Fixation. A. Muller and W. E. Newton (eds) 83-110.

17. 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.

18. Burris, R. H. 1988. 100 years of discoveries in biological N_2 fixation. In Nitrogen Fixation: Hundred Years After. H. Bothe, F. J. de Bruijn, and W. E. Newton (eds) 21-30.

19. Chen, K. C.-K., J.-S. Chen, and J. L. Johnson. 1986. Structural features of multiple nifH-like sequences and very biased codon usage in nitrogenase genes of Clostridium pasteurianum. J. Bacteriol. 166:162-172.

20. 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.

21. Conradson, S. D., B. K. Burgess, W. E. Newton, K. O. Hodgson, J. W. McDonald, J. F. Rubinson, S. F. Gheller, L. E. Mortenson, M. W. W. Adams, P. K. Mascharak, W. A. Armstrong, and R. H. Holm. 1985. Structural insights from the Mo K-edge X-ray absorption near edge structure of the iron-molybdenum protein of nitrogenase and its iron-molybdenum cofactor by comparison with synthetic Fe-Mo-S clusters. J. Am. Chem. Soc. 107:7935-7940.

22. Cordewener, J., H. Haaker, and C. Veeger. 1983. Binding of MgATP to the nitrogenase proteins from Azotobacter vinelandii. Eur. J. Biochem. 132:47-54.

23. Cordewener, J., M. Kruse-Wolters, H. Wassink, H. Haaker, and C. Veeger. 1988. The role of MgATP hydrolysis in nitrogenase catalysis. Eur. J. Biochem. 172:739-745.

24. Dean, D. R. and K. E. Brigle. 1985. Azotobacter vinelandii nifD- and nifE-encoded polypeptides share structural homology. Proc. Natl. Acad. Sci. USA 82:5720-5723.

25. Dingler, C., J. Kuhla, H. Wassink, and J. Oelze. 1988. Levels and activities of nitrogenase proteins in Azotobacter vinelandii grown at different dissolved oxygen concentrations. J. Bacteriol. 170:2148-2152.

26. Dixon, R. A., S. Austin, M. Buck, M. Drummond, S. Hill, A. Holtel, S. MacFarlane, M. Merrick, and S. Minchin. 1987. Genetics and regulation of nif and related genes in Klebsiella pneumoniae. Phil. Trans. R. Soc. London B 317:147-158.

27. 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.

28. Drummond, M., P. Whitty, and J. Wootton. 1986. Sequence and domain relationships of ntrC and nifA from Klebsiella pneumoniae: Homologies to other regulatory proteins. EMBO J. 5:441-447.

29. Eady, R. R., R. L. Robson, T. H. Richardson, R. W. Miller, and M. Hawkins. 1987. The vanadium nitrogenase of Azotobacter chroococcum. Purification and properties of the VFe protein. Biochem. J. 244:197-207.

30. Earl, C. D., C. W. Ronson, and F. M. Ausubel. 1987. Genetic and structural analysis of the Rhizobium meliloti fixA, fixB, fixC, and fixX genes. J. Bacteriol. 169:1127-1136.

31. Emerich, D. W. and R. H. Burris. 1976. Interactions of heterologous nitrogenase components that generate catalytically inactive complexes. Proc. Natl. Acad. Sci. USA 73:4369-4373.

32. Emerich, D. W. and R. H. Burris. 1978. Complementary functioning of the component proteins of nitrogenase from several bacteria. J. Bacteriol. 134:936-943.

33. Emerich, D. W., T. Ljones, and R. H. Burris. 1978. Nitrogenase: Properties of the catalytically inactive complex between the Azotobacter vinelandii MoFe protein and the Clostridium pasteurianum Fe protein. Biochim. Biophys. Acta 527:359-369.

34. Evans, D., R. Jones, P. Woodley, and R. Robson. 1988. Further analysis of nitrogen fixation (nif) genes in Azotobacter chroococcum: Identification and expression in Klebsiella pneumoniae of nifS, nifV, nifM and nifB genes and localization of nifE/N-, nifU-, nifA- and fixABC-like genes. J. Gen. Microbiol. 134:931-942.

35. Filler, W. A., R. M. Kemp, J. C. Ng, T. R. Hawkes, R. A. Dixon, and B. E. Smith. 1986. The nifH gene product is required for the synthesis or stability of the iron-molybdenum cofactor of nitrogenase from Klebsiella pneumoniae. Eur. J. Biochem. 160:371-377.

36. Fox, G. E., K. R. Pechman, and C. R. Woese. 1977. Comparative cataloging of 16S ribosomal ribonucleic acid: Molecular approach to procaryotic systematics. Int. J. Syst. Bacteriol. 27:44-57.

37. Fuhrmann, M., H. Hennecke. 1984. Rhizobium japonicum nitrogenase Fe protein gene (nifH). J. Bacteriol. 158:1005-1011.

38. Galfre, G. and C. Milstein. 1981. Preparation of monoclonal antibodies: strategies and procedures. Methods Enzymol. 73:3-46.

39. Golden, J. W., S. J. Robinson, and R. Haselkorn. 1985. Rearrangement of nitrogen fixation genes during heterocyst differentiation in the cyanobacterium Anabaena. Nature 314:419-423.

40. Gotto, J. W. and D. C. Yoch. 1985. Regulation of nitrogenase activity by covalent modification in Chromatium vinosum. Arch. Microbiol. 141:40-43.

41. Gubler, M. and H. Hennecke. 1986. fixA, B and C genes are essential for symbiotic and free-living, microaerobic nitrogen fixation. FEBS Lett. 200:186-192.

42. Haaker, H. 1988. Biochemistry and physiology of nitrogen fixation. BioEssays 9:112-117.

43. Haaker, H., A. DeKok, and C. Veeger. 1974. Regulation of dinitrogen fixation in intact Azotobacter vinelandii. Biochim. Biophys. Acta 357:344-357.

44. Haaker, H. and J. Klugkist. 1987. The bioenergetics of electron transport to nitrogenase. FEMS Microbiol. Rev. 46:57-71.

45. Haaker, H., C. Laane, K. Hellingwerf, B. Houwer, W. N. Konings, and C. Veeger. 1982. Short-term regulation of the nitrogenase activity in Rhodopseudomonas sphaeroides. Eur. J. Biochem. 127:639-645.

46. Haaker, H. and C. Veeger. 1977. Involvement of the cytoplasmic membrane in nitrogen fixation by Azotobacter vinelandii. Eur. J. Biochem. 77:1-10.

47. 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.

48. Hagen, W. R., H. Wassink, R. R. Eady, B. E. Smith, and H. Haaker. 1987. Quantitative EPR of an S = 7/2 system in thionine-oxidized MoFe proteins of nitrogenase. A redefinition of the P-cluster concept. Eur. J. Biochem. 169:457-465.

49. 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.

50. Hales, B. J., D. J. Langosch, and E. E. Case. 1986. Isolation and characterization of a second nitrogenase Fe-protein from Azotobacter vinelandii. J. Biol. Chem. 261:15301-15306.

51. Hartmann, A., H. Fu, and R. H. Burris. 1986. Regulation of nitrogenase activity by ammonium chloride in Azospirillum sp. J. Bacteriol. 165:864-870.

52. Hausinger, R. P. and J. B. Howard. 1980. Comparison of the iron proteins from the nitrogen fixation complexes of Azotobacter vinelandii, Clostridium pasteurianum, and Klebsiella pneumoniae. Proc. Natl. Acad. Sci. USA 77:3826-3830.

53. Hausinger, R. P. and J. B. Howard. 1982. The amino acid sequence of the nitrogenase iron protein from Azotobacter vinelandii. J. Biol. Chem. 257:2483-2490.

54. Hausinger, R. P. and J. B. Howard. 1983. Thiol reactivity of the nitrogenase Fe-protein from Azotobacter vinelandii. J. Biol. Chem. 258:13486-13492.

55. 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.

56. Hawkes, T. R. and B. E. Smith. 1983. Purification and characterization of the inactive MoFe protein (NifB$^-$ Kp1) of the nitrogenase from nifB mutants of Klebsiella pneumoniae. Biochem. J. 209:43-50.

57. Hedman, B., P. Frank, S. F. Gheller, A. L. Roe, W. E. Newton, and K. O. Hodgson. 1988. New structural insights into the iron-molybdenum cofactor from Azotobacter vinelandii nitrogenase through sulfur K and molybdenum L X-ray absorption edge studies. J. Am. Chem. Soc. 110:3798-3805.

58. Hill, S. and E. P. Kavanagh. 1980. Roles of nifF and nifJ gene products in electron transport to nitrogenase in Klebsiella pneumoniae. J. Bacteriol. 141:470-475.

59. 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.

60. Hoover, T. R., J. Imperial, J. Liang, P. W. Ludden, and V. K. Shah. 1988. Dinitrogenase with altered substrate specificity results from the use of homocitrate analogues for in vitro synthesis of the iron-molybdenum cofactor. Biochemistry 27:3647-3652.

61. Hoover, T. R., J. Imperial, P. W. Ludden, and V. K. Shah. 1989. Homocitrate is a component of the iron-molybdenum cofactor of nitrogenase. Biochemistry 28:2768-2771.

62. Hoover, T. R., A. D. Robertson, R. L. 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 329:855-857.

63. Hoover, T. R., V. K. Shah, G. P. Roberts, and P. W. Ludden. 1986. nifV-dependent, low-molecular weight factor required for in vitro synthesis of iron-molybdenum cofactor of nitrogenase. J. Bacteriol. 167:999-1003.

64. Horton, R. D. and S. T. Lord. 1986. In vitro methylation doubles the efficiency of oligonucleotide-directed mutagenesis. Nucl. Acids Res. 14:5112.

65. Howard, J. B., R. Davis, B. Moldenhauer, V. L. Cash, and D. Dean. 1989. Fe:S cluster ligands are the only cysteines required for nitrogenase Fe-protein activity. J. Biol. Chem. In press.

66. Howard, K. S, B. J. Hales, and M. D. Socolofsky. 1985. In vivo interaction between nitrogenase molybdenum-iron protein and membrane in Azotobacter vinelandii and Rhodospirillum rubrum. Biochim. Biophys. Acta 812:575-585.

67. Howard, K. S., P. A. McLean, F. B. Hansen, P. V. Lemley, K. S. Koblan, and W. H. Orme-Johnson. 1986. Klebsiella pneumoniae nifM gene product is required for stabilization and activation of nitrogenase iron protein in Escherichia coli. J. Biol. Chem. 261:772-778.

68. Imperial, J., R. A. Ugalde, V. K. Shah, and W. J. Brill. 1984. Role of the nifQ gene product in the incorporation of molybdenum into nitrogenase in Klebsiella pneumoniae. J. Bacteriol. 158:187-194.

69. 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.

70. Itakura, K., J. J. Rossi, and R. B. Wallace. 1984. Synthesis and use of synthetic oliogonucleotides. Ann. Rev. Biochem. 53:323-356.

71. Jacobson, M. R., D. R. Dean, B. Masepohl, and W. Klipp. 1989. A nif-associated gene common to Azotobacter vinelandii and Rhodobacter capsulatus encodes a periodic protein. Mol. Gen. Genet. In press.

72. 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.

73. Jensen, B. B. and R. H. Burris. 1985. Effect of high pN_2 and high pD_2 on NH_3 production, H_2 evolution, and HD formation by nitrogenase. Biochemistry 24:1141-1147.

74. Joerger, R. D. and P. E. Bishop. 1988a. Bacterial alternative nitrogen fixation systems. Crit. Rev. Microbiol. 16:1-14.

75. Joerger, R. D. and P. E. Bishop. 1988b. Nucleotide sequence and genetic analysis of the nifB-nifQ region from Azotobacter vinelandii. J. Bacteriol. 170:1475-1487.

76. Joerger, R. D., M. R. Jacobson, and P. E. Bishop. 1989. Two nifA-like genes required for the expression of alternative nitrogenases by Azotobacter vinelandii. J. Bacteriol. IN PRESS.

77. Joerger, R. D., M. R. Jacobson, R. Premakumar, E. D. Wolfinger, and P. E. Bishop. 1989. Nucleotide sequence and mutational analysis of the structural genes (anfHDGK) for the second alternative nitrogenase from Azotobacter vinelandii. J. Bacteriol. 171:1075-1086.

78. Jones, R. and R. Haselkorn. 1988. The DNA sequence of the Rhodobacter capsulatus nifH gene. Nucl. Acids Res. 16:8735.

79. Kaluza, K. and H. Hennecke. 1984. Fine structure analysis of the nifDK operon encoding the α and β subunits of dinitrogenase from Rhizobium japonicum. Mol. Gen. Genet. 196:35-42.

80. Kennedy, C. and R. L. Robson. 1983. Activation of nif gene expression in Azotobacter by the nifA gene product of Klebsiella pneumoniae. Nature 301:626-628.

81. Klugkist, J. and H. Haaker. 1984. Inhibition of nitrogenase activity by ammonium chloride in Azotobacter vinelandii. J. Bacteriol. 157:148-151.

82. Klugkist, J., H. Haaker, and C. Veeger. 1986. Studies on the mechanism of electron transport to nitrogenase in Azotobacter vinelandii. Eur. J. Biochem. 155:41-46.

83. Klugkist, J., J. Voorberg, H. Haaker, and C. Veeger. 1986. Characterization of three different flavodoxins from Azotobacter vinelandii. Eur. J. Biochem. 155:33-40.

84. Knowles, J. R. 1987. Tinkering with enzymes: What are we learning? Science 236:1252-1258.

85. Kratzsch, G. 1987. Hellriegel's work at Bernberg. Phil. Trans. R. Soc. Lond. B 317:107-109.

86. Kuhla, J. and J. Oelze. 1988. Dependence of nitrogenase switch-off upon oxygen stress on the nitrogenase activity in Azotobacter vinelandii. J. Bacteriol. 170:5325-5329.

87. Kurtz, D. M., R. S. McMillan, B. K. Burgess, L. E. Mortenson, and R. H. Holm. 1979. Identification of iron-sulfur centers in the iron-molybdenum proteins of nitrogenase. 76:4986-4989.

88. Laemmli, U. K. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 227:680-685.

89. Lammers, P. J. and R. Haselkorn. 1983. Sequence of the nifD gene coding for the α subunit of dinitrogenase from the cyanobacterium Anabaena. Proc. Natl. Acad. Sci. USA 80:4723-4727.

90. Lowe, D. J. and R. N. F. Thorneley. 1984a. The mechanism of Klebsiella pneumoniae nitrogenase action. Pre-steady-state kinetics of H_2 formation. Biochem. J. 224:877-866.

91. Lowe, D. J. and R. N. F. Thorneley. 1984b. The mechanism of Klebsiella pneumoniae nitrogenase action. The determination of rate constants required for the simulation of the kinetics of N_2 reduction and H_2 evolution. Biochem. J. 224:895-901.

92. Lowery, R. G., C. L. Chang, L. C. Davis, M.-C. McKenna, P. J. Stephens, and P. W. Ludden. 1989. Substitution of histidine for arginine-101 of dinitrogenase reductase disrupts electron transfer to dinitrogenase. Biochemistry 28:1206-1212.

93. Lowery, R. G. and P. W. Ludden. 1988. Purification and properties of dinitrogenase reductase ADP-ribosyltransferase from the photosynthetic bacterium Rhodospirillum rubrum. J. Biol. Chem. 263:16714-16719.

94. Lowery, R. G., L. L. Saari, and P. W. Ludden. 1986. Reversible regulation of the nitrogenase iron protein from Rhodospirillum rubrum by ADP-ribosylation in vitro. J. Bacteriol. 166:513-518.

95. Ludden, P. W., G. P. Roberts, R. G. Lowery, W. P. Fitzmaurice, L. L. Saari, L. Lehman, D. Lies, D. Woehle, H. Wirt, S. A. Murrell, M. R. Pope, and R. H. Kanemoto. 1988. Regulation of nitrogenase activity by reversible ADP-ribosylation of dinitrogenase reductase. In Nitrogen Fixation: Hundred Years After. H. Bothe, F. J. de Bruijn, and W. E. Newton (eds) 157-162.

96. Lundell, D. J. and J. B. Howard. 1981. Isolation and sequences of the cysteinyl tryptic peptides from the MoFe-protein of Azotobacter vinelandii nitrogenase. J. Biol. Chem. 12:6385-6391.

97. Magasanik, B. 1988. Reversible phosphorylation of an enhancer binding protein regulates the transcription of bacterial nitrogen utilization genes. Trends in Biochem. Sci. 13:475-479.

98. Masepohl, B., W. Klipp, and A. Puhler. 1988. Genetic characterization and sequence analysis of the duplicated nifA/nifB gene region of Rhodobacter capsulatus. Mol. Gen. Genet. 212:27-37.

99. Mazur, B. J. and C.-F. Chui. 1982. Sequence of the gene coding for the β-subunit from the blue-green alga Anabaena. Proc. Natl. Acad. Sci. USA 79:6782-6786.

100. McLean, P. A. and R. A. Dixon. 1981. Requirement of nifV gene for production of wild-type nitrogenase enzyme in Klebsiella pneumoniae. Nature 292:655-656.

101. McLean, P. A., V. Papaefthymiou, W. H. Orme-Johnson, and Eckard Munck. 1987. Isotopic hybrids of nitrogenase. Mossbauer study of MoFe protein with selective ^{57}Fe enrichment of the P-cluster. J. Biol. Chem. 262:12900-12903.

102. McLean, P. A., B. E. Smith, and R. A. Dixon. 1983. Nitrogenase of Klebsiella pneumoniae nifV mutants. Biochem. J. 211:589-597.

103. Merrick, M., S. Hill, H. Hennecke, M. Hahn, R. Dixon, and C. Kennedy. 1982. Repressor properties of the nifL gene product in Klebsiella pneumoniae. Mol. Gen. Genet. 185:75-81.

104. Mevarech, M., D. Rice, and R. Haselkorn. 1980. Nucleotide sequence of a cyanobacterial nifH gene coding for nitrogenase reductase. Proc. Natl. Acad. Sci. USA 77:6476-6480.

105. Miller, J. H. 1972. Assays of β-galactosidase. Experiments in Molecular Genetics 352-355.

106. Miller, R. W. and R. R. Eady. 1988a. Cyanamide: a new substrate for nitrogenase. Biochim. Biophys. Acta 952:290-296.

107. Miller, R. W. and R. R. Eady. 1988b. Tight binding interactions of molybdenum and vanadium nitrogenase proteins with adenine nucleotides. In Nitrogen Fixation: Hundred Years After. H. Bothe, F. J. de Bruijn, and W. E. Newton (eds) 90.

108. Miller, R. W., R. L. Robson, M. G. Yates, and R. R. Eady. 1980. Catalysis of exchange of terminal phosphate groups of ATP and ADP by purified nitrogenase proteins. Can. J. Biochem. 58:542-548.

109. Moreno-Vivian, C., S. Hennecke, A. Puhler, and W. Klipp. 1989. Open reading frame 5 (ORF5), encoding a ferredoxinlike protein, and nifQ are cotranscribed with nifE, nifN, nifX, and ORF4 in Rhodobacter capsulatus. J. Bacteriol. 171:2591-2598.

110. Morett, E. and M. Buck. 1988. NifA-dependent in vivo protection demonstrates that the upstream activator sequence of nif promoters is a protein binding site. Proc. Natl. Acad. Sci. USA 85:9401-9405.

111. Morett, E., S. Moreno, and G. Espin. 1988. Transcription analysis of the three nifH genes of Rhizobium phaseoli with gene fusions. Mol. Gen. Genet. 213:499-504.

112. Munck, E., H. Rhodes, W. H. Orme-Johnson, L. C. Davis, W. J. Brill, and V. K. Shah. 1975. Nitrogenase. VIII. Mossbauer and EPR spectroscopy. The MoFe protein component from Azotobacter vinelandii OP. Biochim. Biophys. Acta 400:32-53.

113. Murrell, J. C. 1988. The rapid switch-off of nitrogenase activity in obligate methane-oxidizing bacteria. Arch. Microbiol. 150:489-495.

114. Murrell, S. A., R. G. Lowery, and P. W. Ludden. 1988. ADP-ribosylation of dinitrogenase reductase from Clostridium pasteurianum prevents its inhibition of nitrogenase from Azotobacter vinelandii. Biochem. J. 251:609-612.

115. Nagatani, H. H., V. K. Shah, and W. J. Brill. 1974. Activation of inactive nitrogenase by acid-treated component I. J. Bacteriol. 120:697-701.

116. Newton, W. E., S. F. Gheller, B. Hedman, K. O. Hodgson, S. M. Lough, and J. W. McDonald. 1986. Elicitation of thiomolybdates from the iron-molybdenum cofactor of nitrogenase. Comparison with synthetic Fe-Mo-S complexes. Eur. J. Biochem. 159:111-115.

117. 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.

118. Norel, F. and C. Elmerich. 1987. Nucleotide sequence and functional analysis of the two nifH copies of Rhizobium ORS571. J. Gen. Microbiol. 133:1563-1576.

119. Noti, J. D., O. Folkerts, A. N. Turken, and A. A. Szalay. 1986. Organization and characterization of genes essential for symbiotic nitrogen fixation from Bradyrhizobium japonicum I110. J. Bacteriol. 167:774-783.

120. O'Farrel, P. H. 1975. High resolution two-dimensional electrophoresis of proteins. J. Biol. Chem. 250:4007-4021.

121. Pagani, S., M. Eldridge, and R. R. Eady. 1987. Nitrogenase of Klebsiella pneumoniae: rhodanese-catalysed restoration of activity of the inactive 2Fe species of the Fe protein. Biochem. J. 244:485-488.

122. Page, W. J. and M. von Tigerstrom. 1979. Optimal conditions for transformation of Azotobacter vinelandii. J. Bacteriol. 139:1058-1061.

123. Paul, W. and M. Merrick. 1989. The roles of the nifW, nifZ, and nifM genes of Klebsiella pneumoniae in nitrogenase biosynthesis. Eur. J. Biochem. 178:675-682.

124. Pope, M. R., S. A. Murrell, and P. W. Ludden. 1985. Covalent modification of the iron protein of nitrogenase from Rhodospirillum rubrum by adenosine diphosphoribosylation of a specific arginine residue. Proc. Natl. Acad. Sci. USA 82:3173-3177.

125. Postgate, J. R. 1970. Biological nitrogen fixation. Nature 226:25-27.

126. Pretorius, I.-M., D. E. Rawlings, E. G. O'Neill, W. A. Jones, R. Kirby, and D. R. Woods. 1987. Nucleotide sequence of the gene encoding the nitrogenase iron protein of Thiobacillus ferrooxidans. J. Bacteriol. 169:367-370.

127. Prosser, J., L. Graham, and R. J. Maier. 1988. Hydrogen-mediated enhancement of hydrogenase expression in Azotobacter vinelandii. J. Bacteriol. 170:1990-1993.

128. Puhler, A. and W. Klipp. 1981. Fine structure analysis of the gene region for N_2-fixation (nif) of Klebsiella pneumoniae. In Biology of Inorganic Nitrogen and Sulfur. H. Bothe and A. Trebst (eds) 276-286.

129. Quinto, C., H. de la Vega, M. Flores, J. Leemans, M. A. Cevallos, M. A. Pardo, R. Azpiroz, M. de Lourdes Girard, E. Calva, and R. Palacios. 1985. Nitrogenase reductase: a functional multigene family in Rhizobium phaseoli. Proc. Natl. Acad. Sci. USA 82:1170-1174.

130. Raina, R., M. A. Reddy, D. Ghosal, and H. K. Das. 1988. Characterization of the gene for the Fe-protein of the vanadium dependent alternative nitrogenase of Azotobacter vinelandii and construction of a Tn5 mutant. Mol. Gen. Genet. 214:121-127.

131. Rawlings, J., V. K. Shah, J. R. Chisnell, W. J. Brill, R. Zimmermann, E. Munck, and W. H. Orme-Johnson. 1978. Novel metal cluster in the iron-molybdenum cofactor of nitrogenase. J. Biol. Chem. 253:1001-1004.

132. Richardson, J. S., D. C. Richardson, and K. A. Thomas. 1976. Similarity of three-dimensional structure between the immunoglobulin domain and the copper, zinc superoxide dismutase subunit. J. Mol. Biol. 102:221-235.

133. 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.

134. 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.

135. 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.

136. Robson, R. L. 1979. Characterization of an oxygen-stable nitrogenase complex isolated from Azotobacter chroococcum. Biochem. J. 181:569-575.

137. Robson, R. L. 1984. Identification of possible adenine nucleotide-binding sites in nitrogenase Fe- and MoFe-proteins by amino acid sequence comparison. FEBS Lett. 173:394-398.

138. Robson, R. L. 1986. Nitrogen fixation in strains of Azotobacter chroococcum bearing deletions of a cluster of genes coding for nitrogenase. Arch. Microbiol. 146:74-79.

139. Robson, R. L., R. R. Eady, T. H. Richardson, R. W. Miller, M. Hawkins, and J. R. Postgate. 1986. The alternative nitrogenase of Azotobacter chroococcum is a vanadium enzyme. Nature 322:388-390.

140. Robson, R., P. Woodley, P., and R. Jones. 1986. Second gene (nifH*) coding for a nitrogenase iron protein in Azotobacter chroococcum is adjacent to a gene coding for a ferredoxin-like protein. EMBO J. 5:1159-1163.

141. Rossen, L, Q.-S. Ma, E. A. Mudd, A. W. B. Johnston, and J. A. Downie. 1984. Identification and DNA sequence of fixZ, a nifB-like gene from Rhizobium leguminosarum. Nucl. Acids Res. 12:7123-7134.

142. Saari, L. L., E. W. Triplett, and P. W. Ludden. 1984. Purification and properties of the activating enzyme for iron protein of nitrogenase from the photosynthetic bacterium Rhodospirillum rubrum. J. Biol. Chem. 259:15502-15508.

143. Salemme, F. R. 1977. Structure and function of cytochromes C. Ann. Rev. Biochem. 46:299-329.

144. Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA 74:5463-5467.

145. Scherings, G., H. Haaker, H. Wassink, and C. Veeger. 1983. On the formation of an oxygen-tolerant three-component nitrogenase complex from Azotobacter vinelandii. Eur. J. Biochem. 135:591-599.

146. Schumann, J. P., G. M. Waitches, and P. A. Scolnik. 1986. A DNA fragment hybridizing to a nif probe in Rhodobacter capsulatus is homologous to a 16S rRNA gene. Gene 48:81-92.

147. 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.

148. Scott, K. F., B. G. Rolfe, and J. Shine. 1983. Biological nitrogen fixation: primary structure of the Rhizobium trifolii iron protein gene. DNA 2:149-155.

149. Scott, K. F., B. G. Rolfe, and J. Shine. 1983. Nitrogenase structural genes are unlinked in the nonlegume symbiont Parasponia Rhizobium. DNA 2:141-148.

150. Setterquist, R., K. Brigle, J. Beynon, M. Cannon, F. Cannon, and D. Dean. 1988. Nucleotide sequence of the nifTYEN regions from Azotobacter vinelandii and Klebsiella pneumoniae. In Nitrogen Fixation: Hundred Years After. H. Bothe, F. J. de Bruijn, and W. E. Newton (eds) 324.

151. Shah, V. K. and W. J. Brill. 1977. Isolation of an iron-molybdenum cofactor from nitrogenase. Proc. Natl. Acad. Sci. USA 74:3249-3253.

152. Shah, V. K., T. R. Hoover, J. Imperial, T. D. Paustian, G. P. Roberts, and P. W. Ludden. 1988. Role of nif gene products and homocitrate in the biosynthesis of iron-molybdenum cofactor. In Nitrogen Fixation: Hundred Years After. H. Bothe, F. J. de Bruijn, and W. E. Newton (eds) 115-120.

153. Shah, V. K., G. Stacey, and W. J. Brill. 1983. Electron transport to nitrogenase: Purification and characterization of pyruvate:flavodoxin oxidoreductase, the nifJ gene product. J. Biol. Chem. 258:12064-12068.

154. Shapira, S. K., J. Chou, F. V. Richaud, and M. J. Casadaban. 1983. New versatile plasmid vectors for expression of hybrid proteins coded by a cloned gene fused to lacZ gene sequences encoding an enzymatically active carboxy-terminal portion of β-galactosidase. Gene 25:71-82.

155. Shethna, Y. I., D. V. DerVartanian, and H. Beinert. 1968. Non-heme (iron-sulfer) proteins of Azotobacter vinelandii. Biochem. Biophys. Res. Commun. 31:862-868.

156. Simpson, F. B. and R. H. Burris. 1984. A nitrogen pressure of 50 atmospheres does not prevent evolution of hydrogen by nitrogenase. Science 224:1095-1097.

157. 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. In Nitrogen Fixation: Hundred Years After. H. Bothe, F. J. de Bruijn, and W. E. Newton (eds) 91-100.

158. Smith, B. E., R. R. Eady, D. J. Lowe, and C. Gormal. 1988. The vanadium-iron protein of vanadium nitrogenase from Azotobacter chroococcum contains an iron-vanadium cofactor. Biochem. J. 250:299-302.

159. Souillard, N., M. Magot, O. Possot, and L. Sibold. 1988. Nucleotide sequence of regions homologous to nifH (nitrogenase Fe protein) from the nitrogen-fixing archaebacteria Methanococcus thermolithotrophicus and Methanobacterium ivanovii: Evolutionary implications. J. Mol. Evol. 27:65-76.

160. Souillard, N. and L. Sibold. 1986. Primary structure and expression of a gene homologous to nifH (nitrogenase Fe protein) from the archaebacterium Methanococcus voltae. Mol. Gen. Genet. 203:21-28.

161. Stam, H., A. H. Stouthamer, and H. W. van Verseveld. 1987. Hydrogen metabolism and energy costs of nitrogen fixation. FEMS Microbiol. Rev. 46:73-92.

162. Steinbauer, J., W. Wenzel, and D. Hess. 1988. Nucleotide and deduced amino acid sequences of the Klebsiella pneumoniae nifK gene coding for the β-subunit of nitrogenase MoFe protein. Nucl. Acids Res. 16:7199.

163. St. John, R. T., M. Johnston, C. Seidman, D. Garfinkel, J. K. Gordon, V. K. Shah, and W. J. Brill. 1975. Biochemistry and genetics of Klebsiella pneumoniae mutant strains unable to fix N_2. J. Bacteriol. 121:759-765.

164. Strandberg, G. W. and P. W. Wilson. 1968. Formation of the nitrogen-fixing enzyme system in Azotobacter vinelandii. Can. J. Microbiol. 14:25-31.

165. Sundaresan, V. and F. M. Ausubel. 1981. Nucleotide sequence of the gene coding for the nitrogenase iron protein from Klebsiella pneumoniae. J. Biol. Chem. 256:2808-2812.

166. Syrtsova, L. A., A. V. Kulikov, S. Y. Druzhinin, and G. I. Likhtenshtein. 1985. Determination of nature of protein ligands of Fe-Mo-cofactor of nitrogenase by spin-labeling technique. Biokhimiya 50:634-638.

167. Telser, J., B. M. Hoffman, R. LoBrutto, T. Ohnishi, A.-L. Tsai, D. Simpkin, and G. Palmer. 1987. Evidence for N coordination to Fe in the [2Fe-2S] center in yeast mitochondrial complex II. FEBS Letters 214:117-121.

168. Thomann, H., T. V. Morgan, H. Jin, S. J. N. Burgmayer, R. E. Bare, and E. I. Stiefel. 1987. Protein nitrogen coordination to the FeMo center of nitrogenase from Clostridium pasteurianum. J. Am. Chem. Soc. 109:7913-7914.

169. 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.

170. Thorneley, R. N. F. and D. J. Lowe. 1983. Nitrogenase of Klebsiella pneumoniae. Kinetics of the dissociation of oxidized iron protein from molybdenum-iron protein: Identification of the rate-limiting step for substrate reduction. Biochem. J. 215:393-403.

171. Thorneley, R. N. F. and D. J. Lowe. 1984a. The mechanism of Klebsiella pneumoniae nitrogenase action. Pre-steady-state kinetics of an enzyme-bound intermediate in N_2 reduction and of NH_3 formation. Biochem. J. 224:887-984.

172. Thorneley, R. N. F. and D. J. Lowe. 1984b. The mechanism of Klebsiella pneumoniae nitrogenase action. Simulation of the dependencies of H_2-evolution rate on component-protein concentration and ratio and sodium dithionite concentration. Biochem. J. 224:903-909.

173. Tiedeman, A. A. and J. M. Smith. 1988. lacZY gene fusion cassettes with Kan^R resistance. Nucl. Acids Res. 16:3587.

174. Torok, I. and A. Kondorosi. 1981. Nucleotide sequence of the R. meliloti nitrogenase reductase (nifH) gene. Nucl. Acids Res. 9:5711-5723.

175. 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. 5:399-407.

176. Tsuprun, V. L., E. V. Orlova, N. A. Kiselev, I. Z. Mitsova, I. S. Blazshcuk, and R. I. Gvozdev. 1986. Electron microscopy of the nitrogenase molecule from Azotobacter vinelandii. J. Inorg. Biochem. 27:141-146.

177. 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.

178. 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.

179. Wang, S.-Z., J.-S. Chen, and J. L. Johnson. 1987. Nucleotide and deduced amino sequences of nifD encoding the α-subunit of nitrogenase MoFe protein of Clostridium pasteurianum. Nucl. Acids Res. 15:3935.

180. Wang, S.-Z., J.-S. Chen, and J. L. Johnson. 1988a. Distinct structural features of the α and β subunits of nitrogenase molybdenum-iron protein of Clostridium pasteurianum: an analysis of amino acid sequences. Biochemistry 27:2800-2810.

181. Wang, S.-Z., J.-S. Chen, and J. L. Johnson. 1988b. The presence of five nifH-like sequences in Clostridium pasteurianum: sequence divergence and transcription properties. Nucl. Acids Res. 16:439-454.

182. Wang, Z.-C., A. Burns, and G. D. Watt. 1985. Complex formation and O_2 sensitivity of Azotobacter vinelandii nitrogenase and its component proteins. Biochemistry 24:214-221.

183. Weinman, J. J., F. F. Fellows, P. M. Gresshoff, J. Shine, and K. F. Scott. 1984. Structural analysis of the genes encoding the molybdenum-iron protein of nitrogenase in the Parasponia Rhizobium strain ANU289. Nucl. Acids Res. 12:8329-8344.

184. Wherland, S., B. K. Burgess, E. I. Stiefel, and W. E. Newton. 1981. Nitrogenase reactivity: Effects of component ratio on electron flow and distribution during nitrogen fixation. Biochemistry 20:5132-5140.

185. Yamane, T., M. S. Weininger, L. E. Mortenson, and M. G. Rossmann. 1982. Molecular symmetry of the MoFe protein of nitrogenase. J. Biol. Chem. 257:1221-1223.

186. Yun, A. C. and A. A. Szalay. 1984. Structural genes of dinitrogenase and dinitrogenase reductase are transcribed from two separate promoters in the broad host range cowpea Rhizobium strain IRc78. Proc. Natl. Acad. Sci. USA 81:7358-7362.

187. Zimmermann, R., E. Munck, W. J. Brill, V. K. Shah, M. T. Henzl, J. Rawlings, and W. H. Orme-Johnson. 1978. Nitrogenase X: Mossbauer and EPR studies on reversibly oxidized MoFe protein from Azotobacter vinelandii OP. Nature of the iron centers. Biochim. Biophys. Acta 537:185-207.

188. Zoller, M. J. and M. Smith. 1983. Oligonucleotide-directed mutagenesis of DNA fragments cloned into M13 vectors. Methods Enzymol. 100B:468-500.

189. Zumft, W. G. and F. Castillo. 1978. Regulatory properties of the nitrogenase from Rhodopseudomonas palustris. Arch. Microbiol. 117:53-60.

APPENDIX I

JOURNAL OF BACTERIOLOGY, Apr. 1987, p. 1547–1553
0021-9193/87/041547-07$02.00/0
Copyright © 1987, American Society for Microbiology
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.

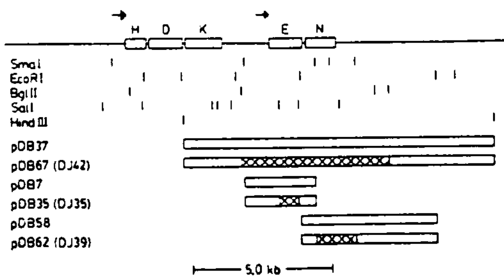


FIG. 1. Physical map of the *A. vinelandii* genome surrounding the *nif* structural gene cluster and plasmids used for deletion strain constructions. Plasmid pDB37 is a *nif*-pKT230 (1) hybrid. The horizontal arrows above the *nifHDK* and *nifEN* gene clusters indicate the direction of transcription. The indicated *nif-HindIII* fragment was cloned into the *HindIII* site of pKT230. pDB67 is a deletion derivative of pDB37. The internal *BglII* fragments contained within pDB37 are deleted in pDB67. pDB6 contains a *SmaI* fragment cloned into the *SmaI* site of pUC8 (12). For the construction of pDB35, an internal *SalI* fragment was deleted from pDB7. pDB58 contains an *EcoRI* fragment cloned into the *EcoRI* site of pBR322. For the construction of pDB62, the internal *SmaI* 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, *RsaI*, *HaeIII*, *Sau3A*, *EcoRI*, *SmaI*, *XhoI*, *PstI*, *SalI*, and *HincII*, 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

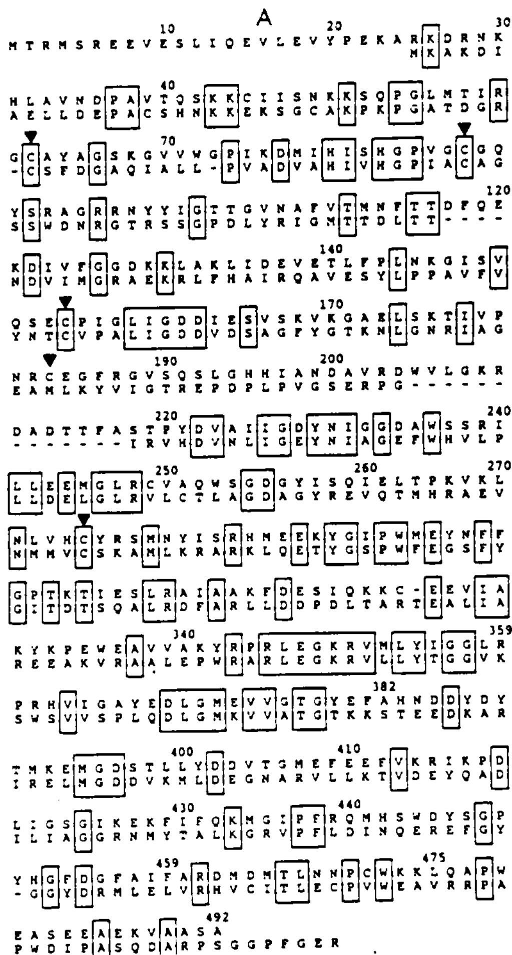
```
GAA TTC GGC TAT GGC GGC TAC GAC CGC ATG CTG GAA CTG GTG CGT
GLU PHE GLY TYR GLY GLY TYR ASP ARG MET LEU GLU LEU VAL ARG

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 TGAGCGATGTCCAGCCCGACCCGACAATTGCAGGTATTGCATA
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



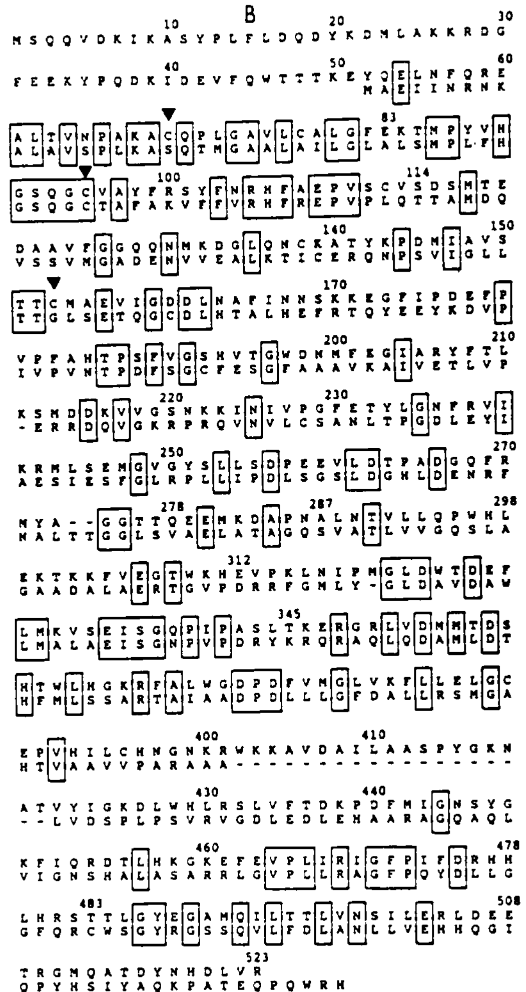


FIG. 3. Comparison of the *nifD*- and *nifE*-encoded (A) and the *nifK*- and *nifN*-encoded (B) polypeptides. The upper sequences represent the *nifD*- or the *nifK*-encoded polypeptides, and lower sequences represent the *nifE*- or *nifN*-encoded polypeptides. Perfect homologies are boxed. Alignment 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.

APPENDIX II

Proc. Natl. Acad. Sci. USA
Vol. 84, pp. 7066–7069, October 1987
Biochemistry

Site-directed mutagenesis of the nitrogenase MoFe protein of *Azotobacter vinelandii*

(iron–molybdenum cofactor/nitrogen fixation/*nifD* gene/protein engineering)

KEVIN E. BRIGLE*, ROBERT A. SETTERQUIST*, DENNIS R. DEAN*, JOHN S. CANTWELL†, MARY C. WEISS‡, AND WILLIAM E. NEWTON†‡§

*Department of Anaerobic Microbiology, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061; †Western Regional Research Center, U.S. Department of Agriculture-Agricultural Research Service, Albany, CA 94710; and ‡Department of Biochemistry and Biophysics, University of California, Davis, CA 95616

Communicated by Martin D. Kamen, July 8, 1987

ABSTRACT **A strategy has been formulated for the site-directed mutagenesis of the *Azotobacter vinelandii nifDK* genes. These genes encode the α and β subunits of the MoFe protein of nitrogenase, respectively. Six mutant strains, which produce MoFe proteins altered in their α subunit by known single amino acid substitutions, have been produced. Three of these transversion mutations involve cysteine-to-serine changes (at residues 154, 183, and 275), two involve glutamine-to-glutamic acid changes (at residues 151 and 191), and one involves an aspartic acid-to-glutamic acid change (at residue 161). All three possible phenotypic responses are observed within this group—i.e., normal, slow, and no growth in the absence of a fixed-nitrogen source. Two-dimensional gel electrophoresis indicates that all mutants accumulate normal levels of the subunits of both nitrogenase component proteins. Whole-cell and crude-extract acetylene-reduction activities indicate substantial levels of Fe protein activity in all strains. In contrast, MoFe protein activities do not parallel the diazotrophic growth capability for all strains. Two strains appear to exhibit altered substrate discrimination. Such analyses should aid in the identification of metallocluster-binding sites and subunit–subunit interaction domains of the MoFe protein and also provide insight into the mechanistic roles of the various prosthetic groups in catalysis.**

Biological nitrogen fixation is catalyzed by nitrogenase (EC 1.18.6.1), a complex metalloenzyme composed of two separately purifiable component proteins called the Fe protein and the MoFe protein. The Fe protein is encoded by the *nifH* gene and acts as a specific ATP-binding, one-electron reductant of the MoFe protein, which contains the site for substrate binding and reduction (1–3). The MoFe protein is an $\alpha_2\beta_2$ protein of approximately 220,000 daltons and it contains 2 Mo atoms and approximately 32 Fe and 32 S^{2-} atoms per molecule. The α and β subunits are encoded by the *nifD* and *nifK* genes, respectively. Mossbauer spectroscopy (4–6) and quantitative extrusion studies indicate that the metal atoms contained within the MoFe protein are organized into at least six metal-containing prosthetic groups of at least two distinct types. Approximately 16 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 (7, 8). All or most of the remaining Fe and both Mo atoms constitute the two identical iron–molybdenum cofactors (FeMoco), and these can be isolated from the native MoFe protein by anaerobic acid/base treatment (which destroys the Fe-S clusters) followed by extraction with *N*-methylformamide (9). The mutual exclusivity of these two cluster-releasing processes indicates different binding modes for the two cluster types.

Mossbauer and electron paramagnetic resonance spectroscopies (2–6), as well as biochemical and genetic studies (10), have been used to demonstrate that the metal centers located within the MoFe protein are likely to participate in the binding and catalytic reduction of dinitrogen. However, little direct information is available concerning the structures, redox properties, and functions of the individual metal centers within the MoFe protein. Furthermore, there is no information regarding the spatial distribution of the metal centers among or between the four subunits of the MoFe protein. One approach toward addressing these issues is to alter specifically the polypeptide environments of the individual clusters and to determine the spectroscopic, redox, and catalytic consequences that result from such alterations. For example, specific amino acid substitutions at key metal-center ligation points within the MoFe protein should produce altered proteins with strengthened, weakened, or eliminated metallocenter interactions. Observations of these effects could indicate appropriate cluster-binding amino acid residues and also provide information on the mechanistic roles of the individual clusters. For this purpose, we isolated the *nifHDK* gene cluster from *Azotobacter vinelandii* and determined its complete nucleotide sequence (11). Here, we describe the formulation of a strategy for the site-directed mutagenesis of the genes that encode the MoFe protein subunits and describe the isolation and characterization of six mutant *A. vinelandii* strains that produce altered MoFe α-subunit proteins with known amino acid substitutions.

MATERIALS AND METHODS

Oligonucleotide-Directed Mutagenesis. Oligonucleotides used for mutagenesis were synthesized on a Beckman System 1 Plus DNA synthesizer. After deprotection, the oligonucleotides were purified by HPLC (12). Oligonucleotide-directed mutagenesis was performed by the method of Zoller and Smith (13). Template DNA used for mutagenesis was a 1.4-kilobase *A. vinelandii nifD*-specific *Eco*RI fragment (11) cloned into the *Eco*RI site of bacteriophage vector M13mp18 (14). Single-lane dideoxy sequencing (15) was used to screen for the desired mutations. Further DNA sequence analysis using all four reactions confirmed the mutations and established that there were no other mutations within 400 nucleotides surrounding the planned mutation. For each mutant construction, replicative-form DNA carrying a known single-base-pair transversion mutation within the *nifD* coding region was prepared and used for transformation of *A. vinelandii* cells. Such DNA, which carries a single transversion muta-

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.

Abbreviation: FeMoco, iron–molybdenum cofactor.
§To whom reprint requests should be addressed at: Western Regional Research Center, USDA-ARS, Albany, CA 94710.

tion within the *nifD* coding sequence, will hereafter be called mutation-vector DNA.

Isolation of Mutant Strains. Transversion mutations carried on mutation-vector DNA were transferred to the *A. vinelandii* chromosome in either one or two steps (Fig. 1), depending on the resultant Nif phenotype of the mutant strain. Mutations resulting in a single amino acid substitution within the *nifD*-encoded polypeptide may exhibit one of three Nif phenotypes. Namely, such mutant strains will grow normally, at a reduced rate, or not at all in the absence of a fixed-nitrogen source. Mutations that cause slow growth or that have no effect on diazotrophic growth can be introduced into the *A. vinelandii* genome by a marker-rescue procedure where the mutation-vector DNA is used to transform a strain deleted for *nifD* to prototrophy (see Fig. 1). Mutations that result in a strictly Nif$^-$ phenotype can be introduced into the *A. vinelandii* chromosome by the simultaneous transformation of wild-type cells with mutation-vector DNA and purified *A. vinelandii* chromosomal DNA that carries an uncharacterized rifampicin-resistance marker. Rif^r transformants are recovered by selection on rifampicin-containing medium, and the coincident transfer (congression) of the Nif$^-$ marker is identified by scoring transformants in the absence or presence of a fixed-nitrogen source (see ref. 16 for a detailed description of this procedure).

Cell Growth and Nitrogenase Derepression. Wild-type and mutant strains of *A. vinelandii* were cultured on a modified Burk medium (17) that was supplemented with filter-sterilized urea to a final concentration of 10 mM when a fixed-nitrogen source was required. For derepression of nitrogenase synthesis, 11 liters of urea-supplemented Burk medium was inoculated with a 500-ml culture (≈200 Klett units, no. 54 filter) in a New Brunswick Microgen SF116 fermentor of 12-liter working capacity. All cultures were stirred at 300 rpm at 30°C with an aeration rate of 12 liters/min. When the culture density reached ≈110 Klett units, cells were concentrated to 1 liter with a Millipore tangential-flow ultrafiltration apparatus. Two liters of sterile nitrogen-free Burk medium were then added to the fermentor, and the culture was concentrated again. A similar second wash was taken to 500-ml final volume. Eleven liters of sterile, nitrogen-free Burk medium was used to transfer cells back to the fermentor and incubated as above for 2.5 hr. Derepressed cells were harvested as above except that the cells were washed with chilled 0.05 M Tris (pH 8.0), then centrifuged at 10,000 × *g* for 10 min and stored at −80°C until needed.

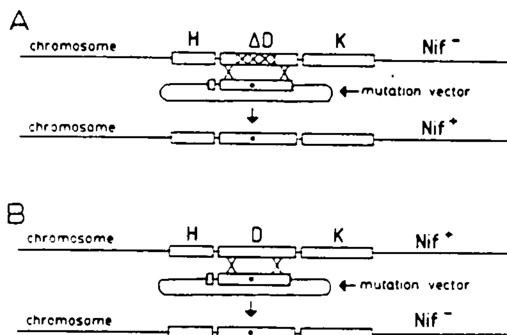


FIG. 1. Strategy for site-directed mutagenesis. Introduction of a single point mutation, indicated by a dot on the mutation vector, into the *A. vinelandii nifD* coding sequence leads to either a Nif$^+$ or a Nif$^-$ phenotype in the resultant mutant strain. (*A*) Those mutations that do not cause a Nif$^-$ phenotype are introduced into the *A. vinelandii* genome by transformation of strain DJ100 (16), which carries a defined deletion within the *nifD* gene (indicated by cross-hatching), to prototrophy. (*B*) Those mutations that result in a Nif$^-$ phenotype are introduced into the *A. vinelandii* genome by transforming the wild-type strain to the Nif$^-$ character. Reciprocal recombination events are indicated by crosses between the chromosome segments and the mutation vector.

Two-Dimensional Gel Electrophoresis. Preparation of extracts for two-dimensional gel electrophoresis was performed as described by Bishop *et al.* (18). Two-dimensional gel analysis was performed as described by O'Farrell (19).

Enzyme Assay. Frozen cells (≈13 g) were suspended in 39 ml of cold 0.05 M Tris (pH 8.0) containing 1 mM $Na_2S_2O_4$ and passed through a chilled, argon-flushed French pressure cell at 12,000 psi (82.7 MPa), followed by centrifugation at 17,000 × *g* for 20 min at 4°C. The extracts were then degassed under argon and the MoFe and Fe protein activities were assayed by acetylene reduction (20), with and without saturating amounts of the purified complementary protein. Specific activities of these purified MoFe and Fe proteins were 2020 and 1700 nmol of acetylene reduced per min per mg of protein, respectively. Protein concentrations were determined by the biuret method (21).

RESULTS AND DISCUSSION

Rationale for Amino Acid Replacements. Several indirect approaches have been used in attempts to identify structurally important regions within the MoFe protein. These include (*i*) comparison of MoFe protein sequences to other metal-center-containing proteins of known structure, in particular ferredoxins (11); (*ii*) comparison of MoFe protein sequences from widely diverse diazotrophic species (see for example refs. 11 and 22); (*iii*) comparison of the MoFe protein α and β subunits (22); and (*iv*) comparison of MoFe protein subunit sequences to sequences of polypeptides that are required for metal-center assembly [for example, the *nifE* and *nifN* gene products, which are required for FeMoco biogenesis (20)].

The above comparisons have shown that there are no typical ferredoxin-like sequences located within the *A. vinelandii* MoFe protein, an observation that is consistent with the unusual redox and spectroscopic properties of the MoFe protein Fe-S centers when they are compared to ferredoxin Fe-S clusters. Interspecies comparisons of MoFe protein subunits reveal homologies that are greatest within the amino-terminal half of homologous subunit polypeptides, and these homologies are largely centered around cysteine residues. There are five cysteine residues that can be considered conserved among all known α-subunit sequences (residues 62, 88, 154, 183, and 275, using the *A. vinelandii* sequence in ref. 11) and three conserved cysteine residues among all known β-subunit sequences (residues 70, 95, and 153, using the *A. vinelandii* sequence in ref. 11). Significant sequence homology is also found surrounding α-subunit cysteine residues 62, 88, and 154 when they are respectively compared to β-subunit cysteine residues 70, 95, and 153. These latter homologies have led to the suggestion that there are structural prosthetic group-binding features common to both subunits of the MoFe protein (22). Finally, the products of the FeMoco-specific biosynthetic genes, *nifE* and *nifN*, share structural homology when compared to the MoFe protein α and β subunits, respectively (20). Importantly, these homologies are also centered around those cysteine residues that exhibit interspecies and, in some cases, intersubunit sequence conservation.

Insights regarding potential metal-center ligands within the MoFe protein can also be gained by considering the extrusion requirements of the metal centers. The Fe-S clusters, which are extruded by thiol treatment, are likely to exhibit cysteine-thiol-type ligation within the MoFe protein. However, the unusual redox and spectroscopic properties of the Fe-S

centers within the MoFe protein raise the possibility that there are also other ligand modes. Thiol reactivity experiments using purified FeMoco have also indicated the potential for a single cysteine-thiol ligand for each FeMoco species (23). In addition, the requirement for *N*-methyformamide or formamide for the extraction of FeMoco indicates a potential role for amide groups in binding FeMoco to the MoFe protein.

Construction and Expression of Altered MoFe Proteins. In *Klebsiella pneumoniae*, the *nif* genes are clustered and are organized into at least seven transcriptional units (24). The promoter regions for these *nif* gene clusters share a characteristic structure and their activation requires the presence of a positive regulatory element, the product of the *nifA* gene (24). It has been shown that elevation of the copy number of the *nifH* gene promoter unbalances *nif* gene expression (25), presumably by sequestering the available activator molecules. Recent studies (20, 26) indicate that the genetic organization, as well as the mechanisms for regulating *nif* gene expression in *A. vinelandii*, is similar to that determined for *K. pneumoniae*. Thus, to avoid potential deleterious effects that could unbalance the synthesis of *nif*-specific components, it is important that altered *nif* genes are reincorporated into the chromosome in single copy rather than residing in the form of multicopy plasmids. Such gene-replacement strategies are described in *Materials and Methods* and are depicted in Fig. 1. Strains constructed in this way are particularly useful, since all of the ancillary *nif*-specific genes and their products are unaffected by this procedure. Consequently, any alterations in the properties of the MoFe proteins prepared from the various mutant strains can be attributed to the introduced mutation rather than as arising from indirect effects on processing or metallocenter-assembly functions.

Characterization of Mutants. The site-directed mutagenesis strategy described above was used to isolate six mutant strains, each of which contains a known transversion mutation resulting in an altered MoFe protein α subunit containing a known single amino acid substitution (Table 1). All three of the anticipated Nif phenotypes are represented in this group. Namely, mutant strains were isolated that are altered within the MoFe protein α subunit and that individually exhibit a Nif^- phenotype (DJ45, DJ64, and DJ56), a Nif^+ phenotype (DJ62 and DJ63), or slow growth in the absence of a fixed-nitrogen source (DJ55). The diazotrophic growth properties of these mutants are shown in Fig. 2. Strain DJ62 was useful because it exhibits normal diazotrophic growth yet carries a mutation that should lead to a charge change within the MoFe protein α subunit (Glu-191 substituted for Gln-191, Table 1). This feature permitted the proof that strains that carry phenotypically silent mutations are readily recovered by our mutagenesis scheme. This proof was accomplished by demonstrating the appropriate charge shift in the MoFe protein α-subunit polypeptide as judged by two-dimensional gel electrophoresis (Fig. 3). Two-dimensional-gel and/or immunological analysis of crude extracts prepared from nitrogenase derepressed cultures was also used to demonstrate that each of the mutant strains accumulates normal amounts of Fe protein and MoFe protein subunits (data not shown). Thus, active MoFe protein is not required for the full expression of nitrogenase components in *A. vinelandii*. These results further indicate that inactive MoFe protein subunits

Table 1. Mutant strains isolated and characterized in this study

Strain	*nifD* codon altered*	Codon change†	Amino acid substitution‡
DJ45	154	TGC→TCC	Cys→Ser
DJ55	183	TGC→TCC	Cys→Ser
DJ56	275	TGC→TCC	Cys→Ser
DJ62	151	CAG→GAG	Gln→Glu
DJ63	161	GAC→GAG	Asp→Glu
DJ64	191	CAG→GAG	Gln→Glu

*Indicates the position of the altered codon within *nifD* in the mutant strain. The initiation ATG codon is number 1.
†Indicates the specific nucleotide change (wild type→mutant) within the altered codon.
‡Indicates the amino acid substitution resulting from the transversion mutation.

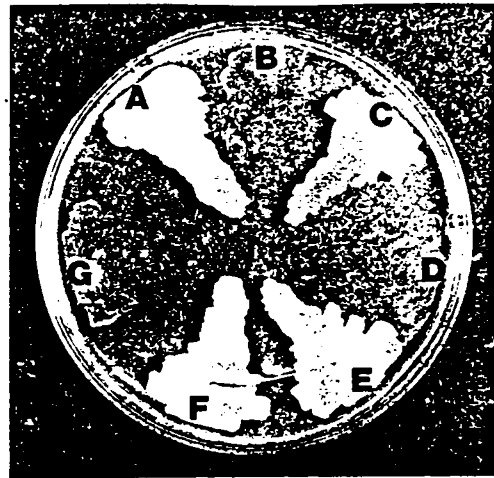


FIG. 2. Diazotrophic growth characteristics of the mutant strains. Cells were streaked onto Burk's nitrogen-free medium and incubated at 30°C for 5 days. A, wild type; B, DJ45; C, DJ55; D, DJ56; E, DJ62; F, DJ63; G, DJ64. Note growth of DJ55; this growth was not evident after 2 days, although good growth was recognized after 2 days for the other Nif^+ strains.

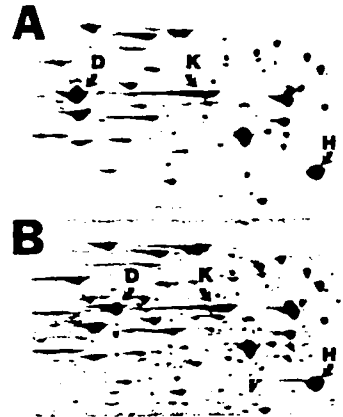


FIG. 3. Two-dimensional gel electrophoresis of protein extracts of wild-type strain and mutant strain DJ62. For each gel the first dimension is the isoelectric-focusing dimension (left to right, basic to acidic) and the second dimension is the size dimension (top to bottom, large to small). Positions of the Fe protein (*nifH* product), the MoFe protein α subunit (*nifD* product), and the MoFe protein β subunit (*nifK* product) have been determined previously (16) and are indicated. (*A*) Wild type. (*B*) DJ62. Notice a shift in the mobility of the MoFe α subunit toward the acidic region in strain DJ62 when compared to wild type. Entire gels are not shown.

Table 2. Component protein activities in mutant and wild-type strains

Strain	Acetylene-reduction activity		
	Whole cell, % wild type	Crude extract, spec. act.*	
		Fe protein†	MoFe protein‡
Wild type	100	43.3	53.7
DJ45	0	14.7	0.1
DJ55	4	23.4	3.1
DJ56	0	20.9	0.2
DJ62	70	33.6	38.2
DJ63	64	18.1	41.7
DJ64	12	27.3	10.0

All values represent the average of at least three determinations.
*Specific activity: nmol of ethylene formed per min per mg of crude extract protein.
†Determined in the presence of saturation levels of purified *A. vinelandii* MoFe protein.
‡Determined in the presence of saturation levels of purified *A. vinelandii* Fe protein.

are not rapidly degraded in *A. vinelandii* as was suggested for *K. pneumoniae* (27).

Whole-cell and crude-extract activity measurements (Table 2) show that alteration of the interspecifically conserved α-subunit cysteine residues 154 and 275 results in a complete loss of both MoFe protein activity and the capacity for diazotrophic growth. In contrast, alteration of α-subunit Cys-183 results in substantially lowered MoFe protein activity, yet the mutant strain retains the capacity for diazotrophic growth. These comparisons are in line with the notion that conserved cysteine residues 154 and 275 play an essential structural or functional role in the MoFe protein, whereas conserved cysteine residue 183 does not. These results do not, however, prove a proposed role for cysteine residues 154 and 275 as metallocluster-binding ligands within the MoFe protein. Rather, the data provide a basis for the further biophysical analyses of MoFe protein prepared from these mutant strains, as well as provide some insight regarding other potential targets for site-directed mutagenesis.

Comparison of mutant strains DJ55 and DJ64 is of particular interest. Although strain DJ64 is incapable of diazotrophic growth (Fig. 2), it retains substantial levels of MoFe protein activity as judged by its acetylene-reduction activity (Table 2). In contrast, DJ55 exhibits a much lower MoFe protein acetylene-reduction activity (Table 2), yet it is still capable of diazotrophic growth (Fig. 2). The difference in substrate utilization shown by these two strains must reside in their ability to recognize or reduce different substrates. Consequently, the further characterization of these strains and the isolation of other mutants with similar phenotypes should prove useful in elucidating the substrate-discrimination and -reduction properties of nitrogenase. In this regard, analytical measurements of the Fe and Mo content and spectroscopic characterization of MoFe protein prepared from these mutant organisms should be particularly useful.

In summary, the above results demonstrate the potential for introducing any site-specific mutation within the genes encoding nitrogenase components. Such mutant organisms are readily constructed irrespective of their Nif phenotype. Importantly, mutants constructed by the procedure described here do not alter the expression of any of the ancillary *nif*-specific components required for metallocenter assembly or component maturation. Another important observation is that normal levels of both MoFe protein subunits accumulate in strains altered within the α subunit, indicating that inactive subunit polypeptides are not rapidly degraded in *A. vinelandii*. Purification and exhaustive biochemical characterization of MoFe protein purified from mutant strains described here should help to identify the role each mutation has in altering enzyme activity. Specifically, we believe that such analyses will provide insights into the roles that each cluster type plays in the catalytic activity of this extraordinarily complex metalloenzyme. Finally, studies using the mutagenesis strategy described here could identify catalytically important residues within the MoFe protein β subunit and the Fe protein, and in those gene products involved in FeMoco biogenesis.

This work was supported by the National Science Foundation (Grant DMB 85-10029).

1. Hageman, R. V. & Burris, R. H. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 2699–2702.
2. Smith, B. E., Lowe, D. J. & Bray, R. C. (1973) *Biochem. J.* **135**, 331–341.
3. Orme-Johnson, W. H., Hamilton, W. D., Jones, T. L., Tso, M.-Y., Burris, R. H., Shah, V. K. & Brill, W. J. (1972) *Proc. Natl. Acad. Sci. USA* **69**, 3142–3145.
4. Smith, B. E. & Lang, G. (1974) *Biochem. J.* **137**, 169–180.
5. Zimmerman, R., Munck, E., Brill, W. J., Shah, V. K., Henzl, M. T., Rawlings, J. & Orme-Johnson, W. H. (1978) *Biochim. Biophys. Acta* **537**, 185–207.
6. Dunham, W. H., Hagen, W. R., Braaksma, A., Grande, H. J. & Haaker, H. (1985) *Eur. J. Biochem.* **146**, 497–501.
7. Kurtz, D. M., McMillan, R. S., Burgess, B. K., Mortensen, L. E. & Holm, R. H. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 4986–4989.
8. Watt, G. D., Burns, A. & Tennent, D. L. (1981) *Biochemistry* **20**, 7272–7277.
9. Shah, V. K. & Brill, W. J. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 3249–3253.
10. Hawkes, T. R., McLean, P. A. & Smith, B. E. (1984) *Biochem. J.* **217**, 317–321.
11. Brigle, K. E., Newton, W. E. & Dean, D. R. (1985) *Gene* **37**, 37–44.
12. Itakura, K., Rossi, J. J. & Wallace, R. B. (1984) *Annu. Rev. Biochem.* **53**, 323–356.
13. Zoller, M. J. & Smith, M. (1983) *Methods Enzymol.* **100B**, 468–500.
14. Messing, J. & Vieira, J. (1982) *Gene* **19**, 269–276.
15. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
16. Robinson, A. C., Burgess, B. K. & Dean, D. R. (1986) *J. Bacteriol.* **166**, 180–186.
17. Strandberg, G. W. & Wilson, P. W. (1968) *Can. J. Microbiol.* **14**, 25–31.
18. Bishop, P. E., Jarlenski, D. M. L. & Hetherington, D. R. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 7342–7346.
19. O'Farrell, P. H. (1975) *J. Biol. Chem.* **250**, 4007–4021.
20. Brigle, K. E., Weiss, M. C., Newton, W. E. & Dean, D. R. (1987) *J. Bacteriol.* **169**, 1547–1553.
21. Gornall, A. G., Bardawall, C. J., & David, M. M. (1948) *J. Biol. Chem.* **177**, 751–766.
22. Thony, B., Kaluza, K. & Hennecke, H. (1985) *Mol. Gen. Genet.* **198**, 441–448.
23. Burgess, B. K., Stiefel, E. I. & Newton, W. E. (1980) *J. Biol. Chem.* **255**, 353–356.
24. Bevnon, J., Cannon, M., Buchanan-Wollaston, V. & Cannon, F. (1983) *Cell* **34**, 665–671.
25. Buchanan-Wollaston, V., Cannon, M. C. & Cannon, F. C. (1981) *Mol. Gen. Genet.* **184**, 102–106.
26. Kennedy, C. & Robson, R. (1983) *Nature (London)* **301**, 626–628.
27. Roberts, G. P., MacNeil, T., MacNeil, D. & Brill, W. J. (1978) *J. Bacteriol.* **136**, 267–279.

APPENDIX III

JOURNAL OF BACTERIOLOGY, Feb. 1989, p. 1017-1027 Vol. 171, No. 2
0021-9193/89/021017-11$02.00/0
Copyright © 1989, American Society for Microbiology

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.

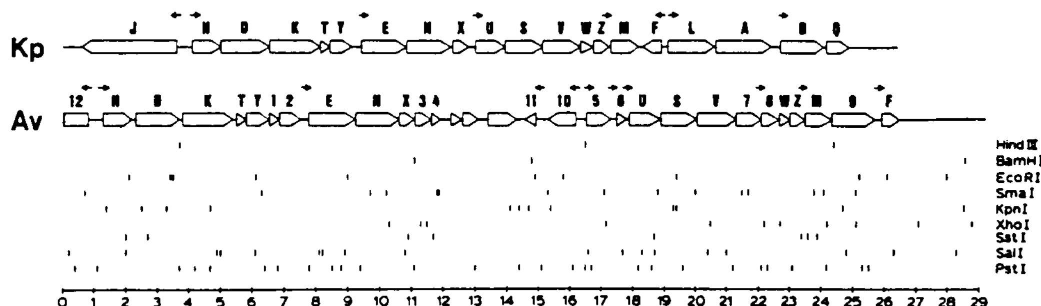


FIG. 1. Physical map of the major *nif* cluster from *A. vinelandii* and comparison with that of *K. pneumoniae*. The upper portion corresponds to the organization of *nif*-specific genes from *K. pneumoniae* (Kp), and the lower portion corresponds to the physical map for the major *nif* cluster from *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

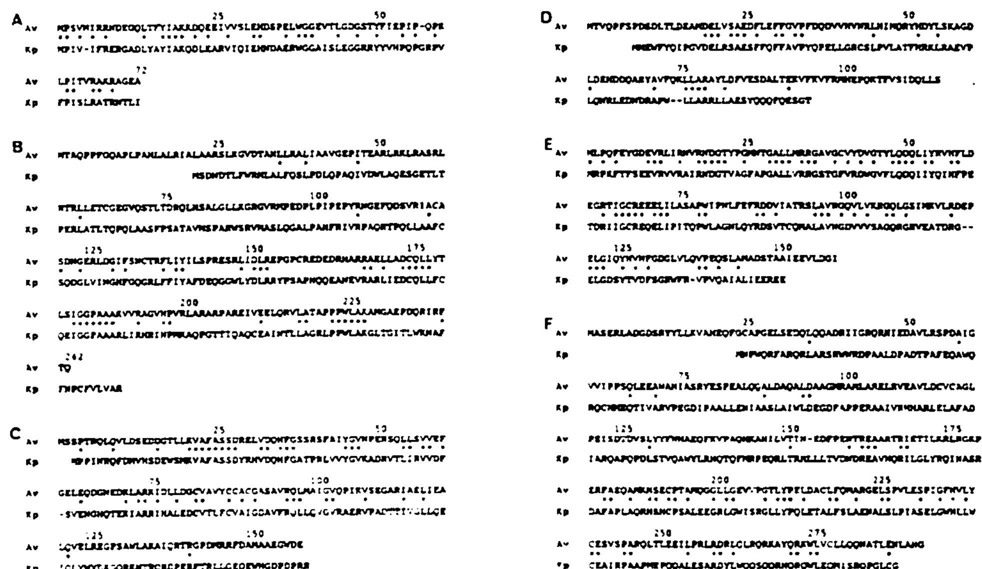


FIG. 2. Comparison of *A. vinelandii* (Av) and *K. pneumoniae* (Kp) *nifT* (A), *nifY* (B), *nifX* (C), *nifW* (D), *nifZ* (E), and *nifM* (F) gene products. Perfect identities are indicated by asterisks. Alignment adjustments made to provide the best fit are indicated by hyphens. Numbers refer to the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

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.

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*

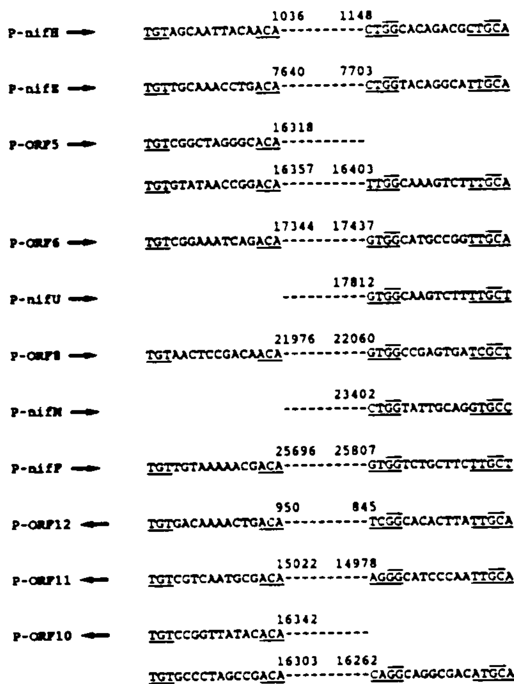


FIG. 3. Comparison of potential promoter sequences and upstream activator sequences identified within the major *nif* cluster from *A. vinelandii*. Arrows indicate the directions of transcription of the proposed promoter sequences as shown in Fig. 1. Numbers refer to the positions of the proposed upstream 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.

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

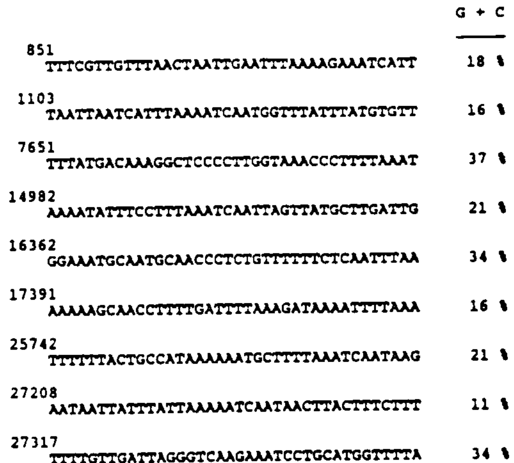


FIG. 4. Positions of A+T-rich regions within and following the major *nif* cluster from *A. vinelandii*. Numbers refer to the location of each A+T-rich region located within the sequenced portion of the *A. vinelandii* genome and can be correlated with the numbering shown on the physical map in Fig.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

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

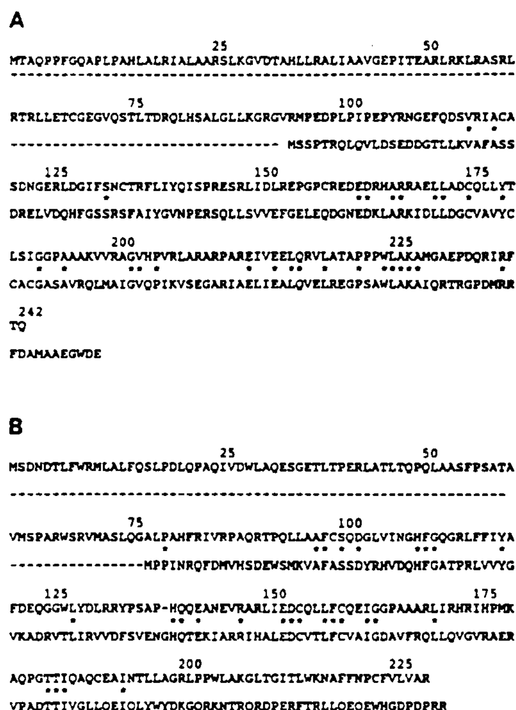


FIG. 5. Comparison of the *A. vinelandii nifY* and *nifX* gene products (A) and the *K. pneumoniae nifY* and *nifX* gene products (B). For each comparison, the upper sequence represents the *nifY* gene product and the lower sequence represents the *nifX* gene product. Perfect identities are indicated 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 AI07873-01) 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 *EcoRI* 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 Tn5 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.

APPENDIX IV

Bothe/de Bruijn/Newton (eds.), Nitrogen Fixation: Hundred Years After
Gustav Fischer · Stuttgart · New York · 1988

Site-directed mutagenesis of the nitrogenase MoFe protein

D. R. Dean*, K. E. Brigle*, H. D. May and W. E. Newton

The Virginia Polytechnic Institute and State University, Blacksburg Virginia, 24061*, and The USDA-ARS, Western Regional Research Center, Albany, CA 94710

INTRODUCTION

The nitrogenase MoFe protein is believed to contain the substrate-binding and reduction site for nitrogen fixation. It has an $\alpha2\beta2$ composition and contains 2 Mo atoms and approximately 32 Fe and 32 $S^{=}$ atoms per native tetramer. Mossbauer spectroscopy and quantitative extrusion studies indicate that the metal atoms contained within the MoFe protein are probably organized into six prosthetic groups of at least two distinct types: Fe-S centers (also called P centers) and FeMo-cofactors (also called M centers). The available biochemical and biophysical evidence indicate that the metalloclusters located within the MoFe protein are likely to function as redox centers which participate in the catalytic reduction of nitrogen (see ref. 6 for the most recent review). However, little direct information is available concerning the structures, redox properties, and functions of the individual metal-containing centers that are located within the MoFe protein. We have initiated a site-directed mutagenesis approach for addressing these issues. Here we discuss our rationale for targeting specific amino acid residues as potential metallocenter ligands within the MoFe protein and describe a gene replacement strategy for the site-directed mutagenesis of _nif_-specific components.

TARGETING POTENTIAL METAL CENTER LIGANDS

Extrusion requirements. Approximately 16 Fe atoms can be extruded from the MoFe protein in the form of [4Fe-4S] clusters by treatment of the native protein with thiols in a denaturing organic solvent. These extruded [4Fe-4S] clusters are presumed to represent the Fe-S clusters characterized by spectroscopic and electrochemical techniques. However, it cannot be concluded that the Fe-S centers within the MoFe protein are indeed [4Fe-4S] clusters since the potential rearrangement of the clusters upon their extrusion has not been eliminated. Typical [4Fe-4S] clusters, such as those found in ferredoxins, are bonded to polypeptides by cysteine mercaptide ligands. The unusual spectroscopic properties of the MoFe protein Fe-S clusters as compared with the [4Fe-4S] ferredoxin-type centers suggests either a novel cluster arrangement or that there are coordinating ligands other than the cysteinyl mercaptide type.

All or most of the remaining Fe and both Mo atoms contained within the MoFe protein constitute the two identical FeMo-cofactors, and these can be isolated intact from the native MoFe protein by anaerobic acid/base treatment (which destroys the Fe-S clusters) followed by extraction with N-methylformamide (9). The extrusion of FeMo-cofactor into NMF, as well

as recent electron spin echo (11) and FT-IR analyses (14), indicate a probability for one or more deprotonated N ligands to FeMo-cofactor. The observation that isolated FeMo-cofactor can react with a single thiolate per Mo atom also indicates the potential for at least one cysteinyl mercaptide ligand to the FeMo-cofactor.

Interspecies Sequence Comparisons. Complete protein sequences, deduced from nucleotide sequence data, are now available for MoFe protein α- and β-subunits from at least seven different organisms. Comparisons of the respective MoFe protein subunits reveal a high level of interspecies primary sequence identity among all of them (see ref. 15 for the latest compilation of sequence information). There are five cysteine residues that can be considered conserved among all known α-subunit sequences (residues 62, 88, 154, and 275, using the A. vinelandii sequence as a reference, see Fig. 1) and three conserved cysteine residues among all known β-subunit sequences (residues 70, 95, and 153, using the A. vinelandii sequence as a reference, see Fig. 1). Based on a thiol requirement for extrusion of the Fe-S centers and the reaction of isolated FeMo-cofactor with a single thiolate, some cysteinyl mercaptide type ligation to both of the prosthetic group types is expected. It is, therefore, probably significant that there is a high degree of interspecies sequence conservation surrounding the conserved cysteine residues from both subunits. However, as there are only sixteen potential cysteine residues available that could provide thiol ligands to the metalloclusters (excluding non-conserved cysteines) there must be other coordinating ligands. This conclusion is in line with the unusual spectroscopic properties of the Fe-S clusters mentioned above.

Intersubunit Sequence Comparisons. Recent MoFe protein Mossbauer spectrum data have indicated that the Fe-S centers can be differentiated into two non-identical subsets (5). MoFe protein subunit sequence data also indicate that there must be at least two different protein environments for the Fe-S clusters because the MoFe protein is composed of two pairs of non-identical subunits and there are no stretches of identically repeated sequences within or between the two different subunit types. However, on the basis of previous Mossbauer results, we have assumed that the coordinating ligand arrangements for the individual Fe-S clusters are likely to be similar. From this perspective it is interesting that there are three interspecifically conserved cysteine residues within the α-subunit that are located within regions having significant amino acid sequence identity when compared with the β-subunit, which in turn has three analogous interspecifically conserved cysteine residues (12, 15). These cysteines (α-subunit residues 62, 88, and 154 and β-subunit residues 70, 95, and 153) are therefore candidates for one or more of the potential cysteine mercaptide ligands to the Fe-S clusters. These sequence comparisons also reveal that there is an interspecies and intersubunit conservation of a histidine residue at positions five residues preceding the conserved cysteinyls 88 (α-subunit) and 95 (β-subunit). The one exception is that glutamine is substituted for histidine in the Anabaena β-subunit sequence. Histidine ligands to a [2Fe- 2S] cluster have recently been suggested (10). Thus, conserved histidine residues 83 (α-subunit) and 90 (β-subunit) can be considered as potential Fe-S cluster ligands within the MoFe protein. We consider this a particularly attractive hypothesis owing to the high level of sequence conservation at

both the interspecies and intersubunit level surrounding regions flanking cysteine residues 88 (α-subunit) and 95 (β-subunit). N ligands to FeMo-cofactor have also been suggested (3, 11, 14) and consequently, these conserved histidine residues must also be considered as potential candidates for providing that coordinating group.

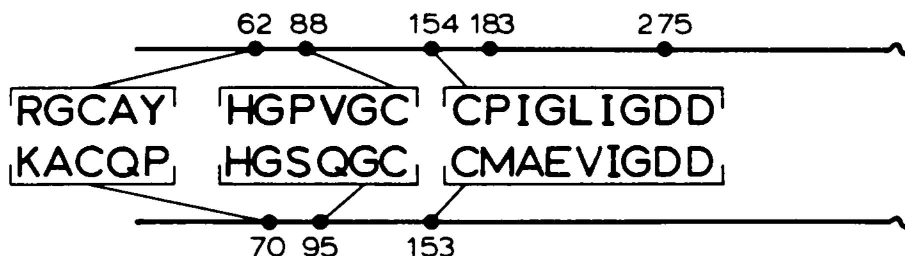


Figure 1. Schematic comparison and spatial distribution of the interspecifically conserved cysteine residues located within the α (upper line) and β (lower line) subunits of the MoFe protein from A. vinelandii. The relevant portion of each subunit is represented as a line with the N-terminus located at the leftward end. The location of the interspecifically conserved cysteine residues (numbered) are indicated by dots on the respective lines. Sequences flanking those cysteine residues which can be considered conserved in both subunits are indicated by the single letter amino acid code.

Comparison of MoFe Protein Subunits and FeMo-cofactor Specific Gene Products. It has been clearly demonstrated that the MoFe protein subunits are not required for FeMo-cofactor biosynthesis (4, 7, 13). These results imply that FeMo-cofactor is synthesized and then inserted into a cofactorless MoFe protein species. Consequently, it was suggested that certain FeMo-cofactor biosynthetic gene products might share primary sequence identity when compared with the MoFe protein subunits (2) and that such sequence comparisons could provide clues regarding the identification of FeMo-cofactor binding domains within the MoFe protein. This suggestion was made on the prediction that FeMo-cofactor or FeMo-cofactor-intermediate binding sites would be conserved among these protein species. Indeed, there is a high degree of primary sequence conservation when the A. vinelandii MoFe protein α-and β-subunits are compared with the products of the FeMo-cofactor biosynthetic genes nifE and nifN, respectively (2). We have also determined the nucleotide sequence of the K. pneumoniae nifE and nifN genes (8) and these results support the above findings. The interspecifically conserved sequences that are shared between the MoFe protein α-subunit and the nifE gene product and between the MoFe protein β-subunit and the nifN gene product are centered around those cysteine residues that exhibit interspecies and, in some cases, intersubunit sequence conservation (2, 8). Based on these extensive sequence identities, we have suggested that the nifEN gene products form an FeMo-cofactor biosynthetic complex (2). It also appears that, in addition to FeMo-cofactor binding sites, there could be one or more Fe-S

centers located within this proposed complex. Such Fe-S centers could be involved in redox reactions used for the biosynthesis of FeMo-cofactor. Another possibility is that these conserved regions represent subunit-subunit interface sites. These two possibilities are not necessarily exclusive. There is also a region within the nifE gene product which shares sequence identity with the MoFe protein α-subunit sequence surrounding interspecifically conserved cysteine residue 275. This sequence conservation within the nifE product includes an analogously conserved cysteine residue. This sequence identity is consistent with our previous suggestion that the region surrounding conserved α-subunit cysteine residue 275 is a potential candidate for an FeMo-cofactor-binding domain (2, 3).

SITE-DIRECTED MUTAGENESIS

Metallocluster extrusion requirements, interspecies and intersubunit sequence comparisons, and comparison of the respective MoFe protein subunit sequences with FeMo-cofactor biosynthetic gene product sequences all strongly indicate important structural or functional roles for the conserved cysteine residues that are found within the MoFe protein subunits. An obvious conjecture is that all or most of these conserved residues could provide essential thiol ligands to the various individual MoFe protein metalloclusters. Other metallocluster ligand modes, however, are also clearly indicated. In this regard, we have targeted interspecificically conserved histidine residues as potentially providing N-donor ligands. These considerations now provide a basis for initiating a strategy for the site-directed mutagenesis of the MoFe protein. Our initial approach involves the systematic substitution of serine residues for the interspecifically conserved cysteine residues that are found within the MoFe protein subunits. Diazotrophic growth properties and MoFe protein activities should indicate which of the conserved cysteines have an essential structural and/or functional role. Although these experiments cannot establish whether or not a particular conserved cysteine residue provides an essential thiol ligand to a metallocenter, the results can provide a basis for the further biophysical analysis of MoFe protein prepared from the mutant strains. Perhaps more importantly, these experiments can also provide insight regarding other potential targets for site-directed mutagenesis.

Construction of mutant strains. Azotobacter vinelandii is the current organism of choice for performing site-directed mutagenesis experiments for the purpose of the biochemical analysis of nitrogenase and associated nif-specific gene products. The reasons for this are: a) pure and highly active nitrogenase is readily obtained from this organism, b) A. vinelandii has an efficient DNA transformation system, and c) nearly all of the nif-specific genes from A. vinelandii have been isolated, sequenced, and mutagenized. We have previously described a chromosomal gene replacement strategy that can be used to incorporate nif-specific mutations into the A. vinelandii genome (1, 7). The importance of the gene replacement approach is that strains constructed in this way do not suffer from gene dosage effects that could unbalance nif gene expression. Consequently, any changes in the properties of altered MoFe proteins prepared from such mutant strains can be attributed to the introduced mutation rather than as arising from indirect effects on processing or

metallocenter-assembly functions. Another important aspect of this approach is that mutant strains can be identified regardless of their diazotrophic growth capability.

Using established site-directed mutagenesis procedures and the gene replacement technique we have constructed 30 individual mutant strains that are altered within the MoFe protein. Each of these mutants has a transversion mutation within the MoFe protein coding sequence and each mutation results in a single amino acid substitution within one or the other of the MoFe protein subunits. Some of these amino acid substitutions and the Nif phenotypes of the resultant mutant strains are depicted in Figure 2.

Analysis of mutant strains. Mutant strains of *A. vinelandii* have now been isolated where each of the interspecifically conserved cysteine residues from the MoFe protein α-subunit (residues 62, 88, 154, 183, and 275) has been individually substituted with serine. Similarly, mutant strains where each of the interspecifically conserved cysteines residues from the MoFe protein β-subunit (residues 70, 95, and 153) has been individually substituted with serine have also been constructed. Substitution of α-subunit cysteine residues 62, 88, 154, or 275 results in a nearly complete loss of the diazotrophic growth capability of the respective mutant strains. Similarly, substitution of β-subunit cysteine residue 70 or 95 also results in loss of diazotrophic growth. In contrast, substitution of α-subunit cysteine residue 183 results only in a severe reduction in diazotrophic growth while substitution of β-subunit cysteine residue 153 has no adverse affect on diazotrophic growth. This latter observation was surprising since there is significant sequence conservation when the β-subunit cysteine residue 153 region is compared with the analogous α-subunit cysteine residue 154 region (see above). Since α-subunit cysteine residue 154 is essential for diazotrophic growth (but β-subunit cysteine residue 153 is not), it is unlikely that this residue has the same structural or functional role as conserved β-subunit cysteine residue 153. This result illustrates the inherent problems involved in relying on sequence comparisons to draw functional conclusions in the absence of biochemical data. However, this finding does not necessarily rule out the possibility that β-subunit residue 153 provides a ligand to a metallocluster, since oxygen could potentially replace sulfur as a cluster ligand.

Substitution of asparagine for conserved α-subunit histidine residue 83 also does not affect the diazotrophic growth capability of the mutant strain. This result was also very surprising since this histidine residue (and the cysteine located five residues downstream) can be considered conserved within both of the MoFe protein subunits, as well as conserved within the *nifE* and *nifN* gene products. A detailed spectroscopic analysis of purified protein is necessary to determine if there is any affect of this substitution upon the redox properties of the altered MoFe protein from this mutant strain. These experiments and the introduction of other substitutions for α-subunit histidine residue 83 are now in progress.

Many of the mutant strains shown in Figure 1 have been examined for their whole-cell and crude extract acetylene reduction activities, whole-cell and crude extract EPR spectra, and the thermal lability of their MoFe protein activities. These results have shown that quantitative EPR spectra from these strains closely parallel their MoFe protein activities.

However, the diazotrophic growth properties of the individual mutant strains do not necessarily reflect their acetylene reduction capabilities in whole-cells or in crude extracts. For example, substitution of glutamate for glutamine residue 191 results in complete loss of diazotrophic growth, yet this strain has substantial levels of acetylene reduction activity in whole-cells or crude extracts (1). This result must indicate a difference in the substrate recognition or substrate reduction capability of this strain when compared with wild type. Thus, the further analysis of this and other mutant strains having a similar phenotype should prove useful in elucidating the substrate discrimination and substrate reduction properties of nitrogenase.

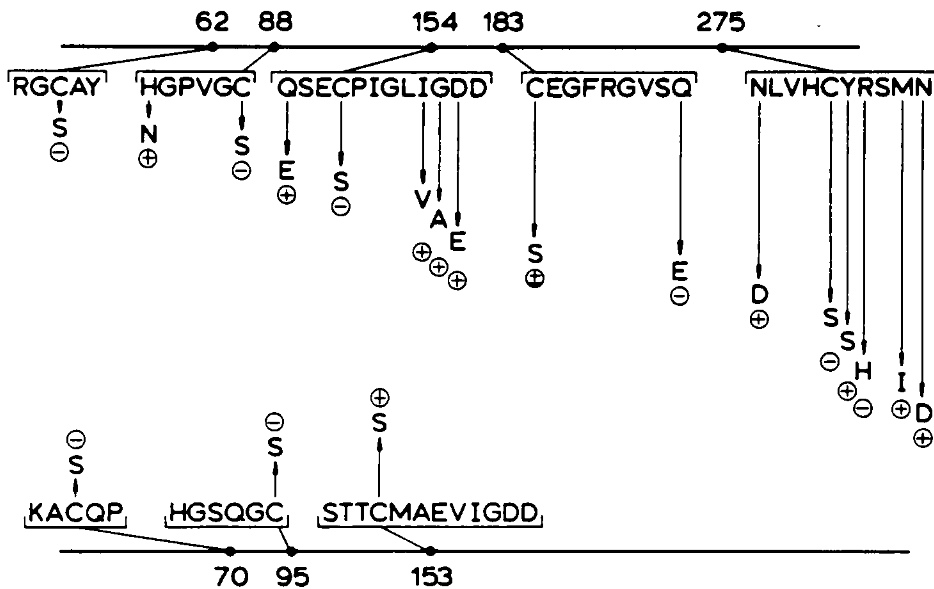


Figure 2. Schematic representation of amino acid substitutions incorporated into the α (top figure) or β (bottom figure) subunit of the MoFe protein from A. vinelandii. Each subunit is represented as a line with the N-terminus located at the leftward end. Interspecifically conserved cysteine residues are indicated as dots. Flanking sequences for each of the interspecifically conserved cysteine residues are indicated by the single letter amino acid code. Specific substitutions for an individual residue are indicated by arrows. The Nif phenotype for each mutant strain is shown below each of the indicated substitutions.

DISCUSSION

We have developed a site-directed mutagenesis and gene replacement strategy that can be used to place specific amino acid substitutions, deletions or insertions within the nitrogenase MoFe protein. This technique can also be applied to any nif-specific gene from A. vinelandii. The strategy relies upon a knowledge of the nucleotide sequence of the

particular nif gene that one wishes to mutagenize, as well as the availability of a mutant strain that has a defined deletion within the coding region for that gene. Such information and mutant strains are now available for the A. vinelandii nifH, nifD, nifK, nifT, nifY, nifE, nifN, nifX, nifU, nifS, nifV, nifW, nifZ, nifM, nifF, nifA, and nifB genes. The mutagenesis procedure has now been successfully used to introduce 30 amino acid substitutions within the MoFe protein subunits, 5 amino acid substitutions within the Fe protein (Cash, Dean & Howard, unpublished), 3 amino acid substitutions within the nifE gene product (Wilson, Newton & Dean, unpublished) and 1 amino acid substitution within the nifM gene product (Jacobson & Dean, unpublished). We believe that this site-directed mutagenesis approach, together with the appropriate biochemical analyses of component proteins prepared from specifically-constructed mutant organisms, will contribute towards the elucidation of the specific functions of the nitrogenase structural components and the associated nif-specific gene products.

Our work is supported by the NSF and the USDA

REFERENCES

1. Brigle, K.E., Setterquist, R.A., Dean, D.R., Cantwell, J.S., Weiss, M.C. and W.E. Newton. (1988) Proc. Natl. Acad. Sci. U.S.A. 84, 7066-7069.
2. Brigle, K.E., Weiss, M.C., Newton, W.E. and D.R. Dean. (1987) J. Bacteriol. 169, 1547-1533.
3. Brigle, K.E., Newton, W.E. and D.R. Dean. (1985) Gene 37, 37-44.Dean, D.R. and K.E. Brigle. (1985) Proc. Natl. Acad. Sci. U.S.A. 56, 5720-5723.
4. Filler, W.A., Kemp, R.M., Ng, J.C., Hawkes, T.R., Dixon, R.A. and B.E. Smith. (1986) Eur. J. Biochem. 160, 371-377.
5. McLean, P.A., Papaefthymiou, V., Orme-Johnson, W.H. and E. Munck. (1987) J. Biol. Chem. 262, 12900-12903.
6. Orme-Johnson, W. H. (1985) Ann. Rev. Biophys. Biophys. Chem. 14, 419-459.
7. Robinson, A.C., Burgess, B.K. and D.R. Dean. (1986) J. Bacteriol. 166, 180-186.
8. Setterquist, R.A., Brigle, K.E., Beynon, J, Ally, A., Cannon, M., Cannon, F. and D.R. Dean. (1988) Nucl. Acids. Res. in Press.
9. Shah, V.K. and W.J. Brill. (1977) Proc. Natl. Acad. Sci. U.S.A. 74, 3249-3253
10. Telser, J., Hoffman, B.M., LoBruttoo, R., Ohnishi, T., Tsai, A.-L, Simpkin, S. and G. Palmer. (1987) FEBS. Lett. 214, 117-121.
11. Thomann, H., Morgan, V.T., Jin, H., Burgmayer, S.J.N., Bare, R.E. and E.I. Stiefel. (1987) J.A.C.S. 109, 7913.
12. Thony, B., Kaluza, K. and H. Hennecke. (1985) Mol. Gen. Genet. 198, 441-448.
13. Ugalde, R.A., Imperial, J., Shah, V.K. and W.J. Brill. (1984) J. Bacteriol. 159, 888-893.
14. Walters, M.A., Chapman, S.K. and W.H. OrmeJohnson. (1986) Polyhedron 5, 561-565.
15. Wang, S.-Z., Chen, J.-S, and J.L. Johnson. (1988) in Press.

APPENDIX V

Project: Analysis of the nifK - nifT Intercistronic Region.

Because there are no nif-consensus or typical E. coli-type promoters preceding nifT or nifY, transcription of these genes is likely directed by the nifHDK promoter. Further, there is no obvious rho-independent transcription termination structure in the intercistronic region following nifK (11). However, as shown in Fig. 29, potential secondary structure is predicted to form within this region. It is possible that these secondary structures are regulatory in nature and may serve to attenuate expression of nifT, nifY, and the downstream ORF's. As such, removal of these potential structures might allow for an increase in the expression of these genes. Such a difference in expression can be monitored utilizing nifT:lacZ fusion strains of A. vinelandii.

Due to the presence of two unique EagI sites within the nifK - nifT intercistronic region (Fig. 29), it was possible to construct a strain (DJ216) in which much of the predicted secondary structure was removed (Fig. 30). As shown in Fig. 31, removal of this region had no effect on the diazotrophic growth properties of the resultant mutant strain.

In order to monitor nifT expression, two nifT:lacZ fusion strains (DJ297 and DJ299) were constructed (Fig. 30). In addition to the nifT:lacZ fusion, mutant strain DJ299 also carries a deletion of the potential secondary structure identical to

the deletion in strain DJ216. When derepressed for nitrogenase activity, cultures of strain DJ299 displayed higher levels of β-galactosidase activity over the first three hours when compared to derepressed cultures of strain DJ297 (Fig.32). This indicates that deletion of DNA from within the nifK - nifT intercistronic region results in an elevated level of nifT gene expression. Further, this result suggests that the secondary structure which is removed by the deletion may responsible for regulating the expression of the downstream genes. This may be either a direct effect that is due only to formation of the proper secondary structure or one which requires the presence of another (nif) gene product (perhaps to bind to this structure). By the end of the six hour derepression, the β-galactosidase activity exhibited by strain DJ299 was always lower than that of strain DJ297 (Fig.32). The cause for this decrease in activity has not been examined. It should be noted that interpretation of these experiments is complicated by the possible polar effects which might result from insertion of the lacZ cartridge into the A. vinelandii genome. For example, the nifT:lacZ constructs that are present in strains DJ297 and DJ299 likely disrupt expression of nifY, ORF1, and ORF2. If the function(s) of any of these gene products is regulatory, their disruption could indirectly affect the β-galactosidase activities which are observed.

In summary, it appears that the nifK - nifT intercistronic region plays a role in regulating expression of downstream genes. While the molecular basis of this regulation is unknown, it may involve the potential secondary structure predicted to form within this region.

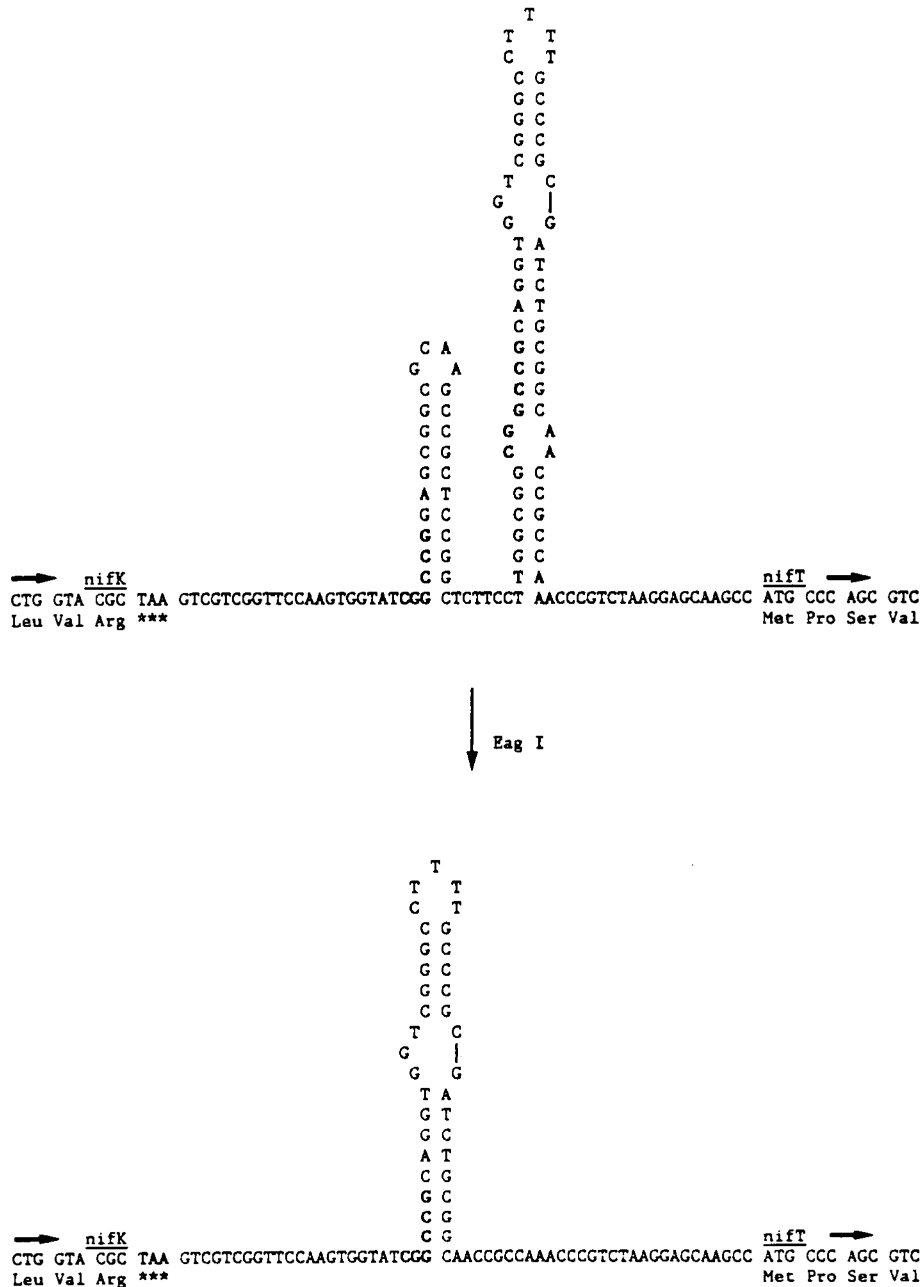


FIGURE 29. Potential secondary structure predicted to form in the <u>nifK</u> - <u>nifT</u> intercistronic region (upper panel) and potential secondary structure predicted to form after removal of the 40-bp EagI (CGGCCG) fragment (lower panel). This is the deletion carried in strains DJ216 and DJ299 (see Fig. 30).

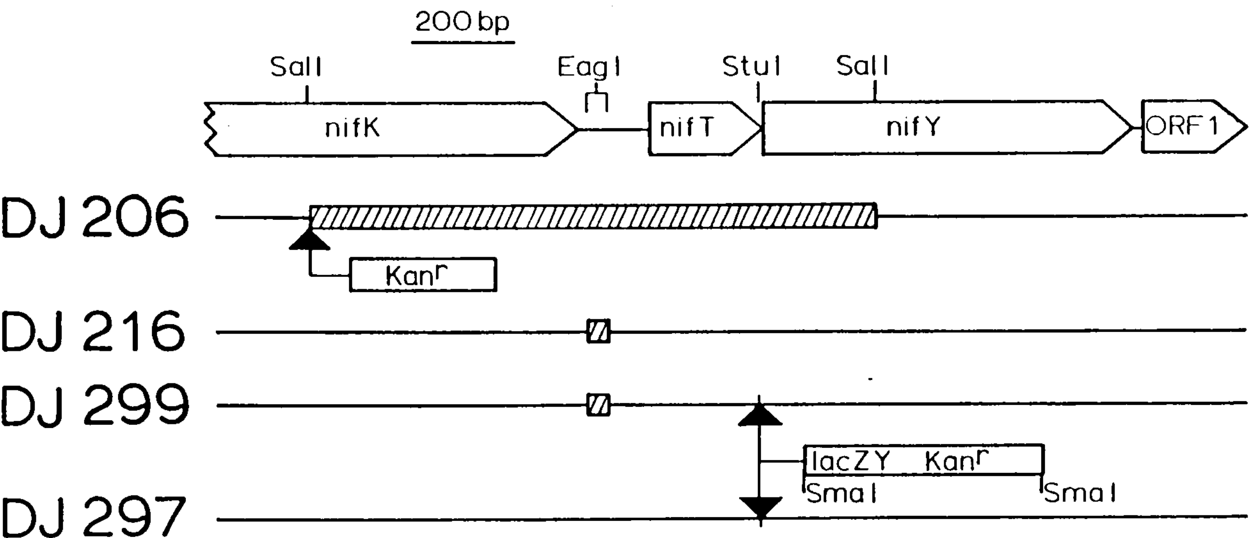


FIGURE 30. Physical map of the _A. vinelandii nif_ cluster surrounding _nifT_ and construction of _A. vinelandii_ strains used to analyze genes within this region. Thin lines represent _A. vinelandii_ DNA sequences and hatched boxes indicate deleted regions. Strain DJ216 contains a 40-bp EagI deletion which removes part of the secondary structure predicted to form in this region (see Fig. 29). This strain was constructed by transforming strain DJ206 (Nif^-) to the Nif^+ phenotype. Strain DJ299 is identical to strain DJ216 except that it also carries an in-frame translational fusion of the _lacZ_ gene of _E. coli_ to the final codon of _nifT_. Strain DJ297 carries the same _lac_ fusion as strain DJ299 however, it does not carry the 40-bp EagI deletion.

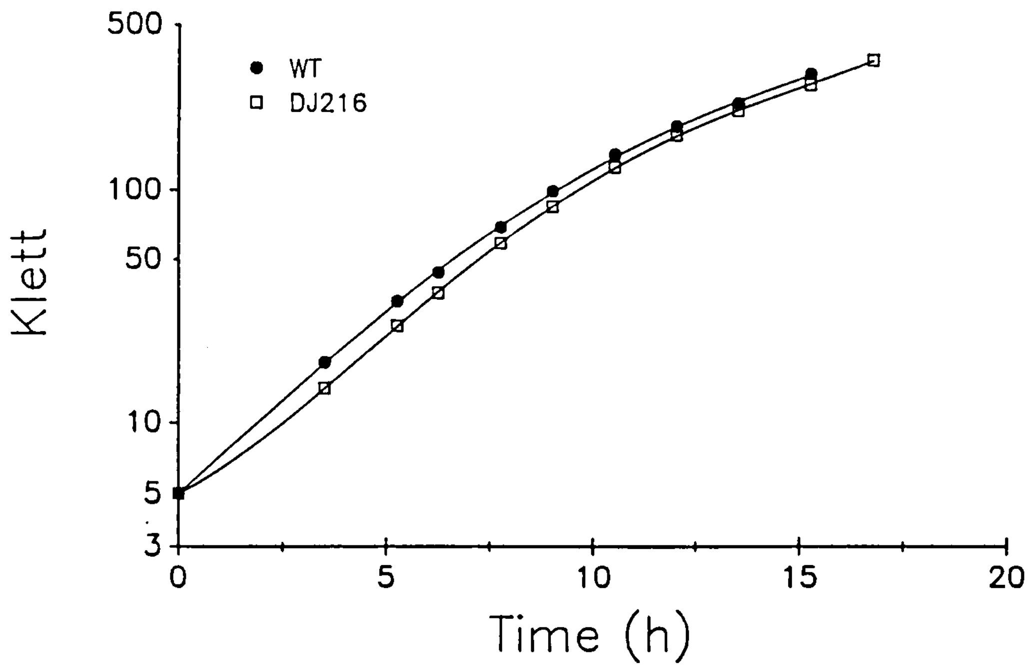


FIGURE 31. Diazotrophic growth of wild type and a mutant strain which carries a deletion in the nifK - nifT intercistronic region.

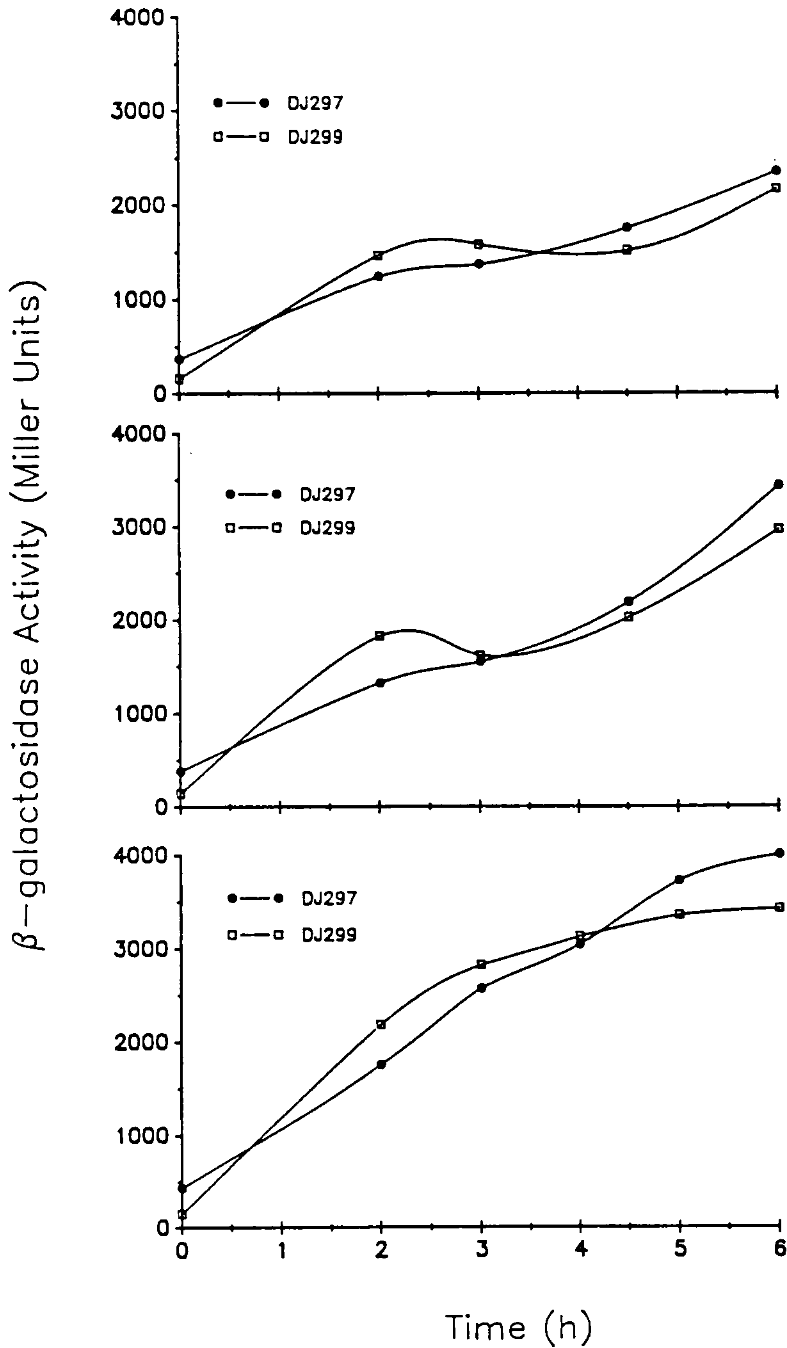


FIGURE 32. Time course of β-galactosidase activity of strains DJ297 and DJ299 upon derepressing for nitrogenase activity. At time = 0 the cells were washed free of their fixed nitrogen source and resuspended in Burk's nitrogen free medium.

APPENDIX VI

Project: Construction of nifH:-, nifD:-, and nifK:lacZ Plasmids and Fusion Strains.

As reported in the body of this dissertation and in Appendix V, lacZ fusion constructs can be utilized to study the expression of the nif genes under a variety of environmental conditions and in wide range of genetic backgrounds. In spite of the potential polar effects which can result from insertion of the lacZ cartridge into the A. vinelandii genome, the use of nif:lacZ fusion constructs provides an excellent means of determining the interactions and functional relationships of the gene products encoded within the nif gene cluster. Especially important are the roles which these other gene products have in regulating, balancing, and maintaining proper levels of expression of the three nif structural genes. While the situation in vivo is unclear, in vitro studies have shown nitrogenase turnover to be dependent upon the Fe/MoFe protein ratio (184). Besides features which are present in the primary nucleotide sequence (tandem stop codons and regions of potential secondary structure), other (nif) gene products may be essential for such differential expression of the individual structural gene products. To this effect, a series of lacZ fusion plasmids were constructed in which the lacZ gene of E. coli was translationally fused to each of the nif structural genes. These lacZ fusion plasmids can be used to transform A. vinelandii strains having almost any genotypic background. Upon derepression, β-galactosidase activity of resultant mutant strains

can be monitored and compared to the activity present in the 'wild type' fusion strains shown in Fig. 33.

All six of the lacZ fusion strains that I constructed in these studies (nifH, nifD, nifK, nifT, nifX, and ORF3) have displayed a significant level of β-galactosidase activity when grown on medium containing a fixed nitrogen source. For example, each of these fusion strains turned blue when streaked on Burk's + nitrogen agar plates containing the chromogenic indicator XGal (5-bromo-4-chloro-3-indoyl β-D-galactoside). It is likely that this low basal level of β-galactosidase activity represents normal background expression that is peculiar to these constructs. Alternatively, this activity may indicate that ntr- and nifA-controlled expression of the nif genes is not as tight as we might have suspected.

Additionally, all six of the lacZ fusion plasmids and strains constructed in this study are translational fusions. As such, expression of the lacZ gene is dependent upon both the transcription and translation of the gene to which it is fused. While this type of construction identifies regulatory mechanisms acting at the level of translation, it is unable to distinguish between these mechanisms of control and those functioning at the level of transcription. To resolve this problem, future studies concerning regulation of the nif genes should include construction of complementary transcriptional lacZ fusions.

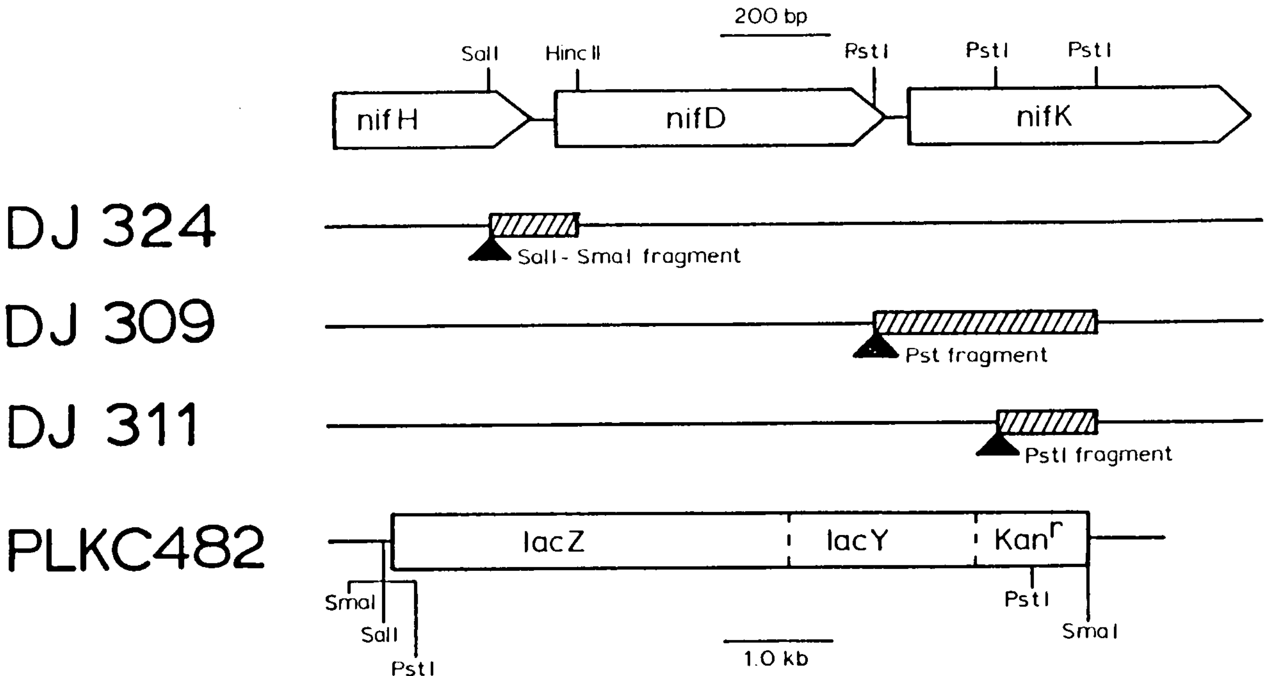


FIGURE 33. Physical map of the A. vinelandii nif cluster surrounding the structural genes and construction of A. vinelandii nif:lacZ fusion strains used to analyze expression of these genes. Thin lines represent A. vinelandii DNA sequences and hatched boxes indicate deleted regions. Strain DJ324 carries an in-frame translational fusion of the lacZ gene of E. coli to the 249th codon of nifH. Strain DJ309 carries an in-frame translational fusion of the lacZ gene of E. coli to the 476th codon of nifD. Strain DJ311 carries an in-frame translational fusion of the lacZ gene of E. coli to the 136th codon of nifK. See references 88 and 154 for a description of the lacZ fusion cassette vectors.

APPENDIX VII

Project: Production of Polyclonal and Monoclonal Antibodies.

Polyclonal Antibody.

A one year old, castrated Nubian goat was initially immunized intramuscularly with 1.0 mg MoFe protein mixed 1:1 with complete Freund adjuvant. Within the next 10 days, the animal received two subcutaneous injections of 1.0 mg MoFe protein mixed 1:1 with incomplete Freund adjuvant. The animal then received 1.0 mg MoFe protein without adjuvant subcutaneously every seven days for the next six weeks. The animal was killed and the blood was collected and incubated at 37 °C for 1 h followed by a 12 h incubation at 4 °C. The sera was slowly brought to 50% saturation with ammonium sulfate and the precipitate collected by centrifugation. The precipitate was dissolved in 0.85% NaCl to the original sample volume, dialyzed three days against six changes of 5 mM borate buffer/0.85% saline (pH 8.5), and two days against 20 mM phosphate buffer (pH 7.0)/0.85% saline (PBS). Finally, the sera was successively filtered through 5.0 μm and 0.45 μm membranes and stored at -80 °C.

An ELISA procedure was developed to aid in the purification of the altered MoFe proteins from the oligonucleotide-directed mutant strains and to roughly quantitate the amount of MoFe protein present in extracts of mutant strains. Wells of Immulon type 1 microtiter plates (Dynatech Laboratories) were coated with 100

μl sample in carbonate buffer (pH 9.6) and incubated overnight at 37 °C. Each well was washed three times with PBST [NaCl, 8.0 g; KH_2PO_4, 0.2 g; $Na_2HPO_4 \cdot 12 H_2O$, 2.9 g; KCl, 0.2 g; Triton X-100, 0.5 ml/liter (pH 7.4)] and blocked with 200 μl PBST + 0.1% BSA for 1 h at 37 °C. The wells were washed five times with PBST and then incubated with 100 μl 10^{-4} anti-MoFe protein antibody in PBST for 1 h at 37 °C. The wells were washed five times with PBST and incubated with 100 μl 10^{-3} rabbit anti-goat IgG-alkaline phosphatase conjugate (Sigma A-7650) for 1 h at 37 °C. The wells were washed five times with PBST and incubated with 100 μl p-nitrophenylphosphate (1.0 mg/ml diethanolamine buffer) [Biorad 172-1063] for 15 min at room temperature. The reaction was terminated by the addition of 100 μl 5 N NaOH. Absorbance due to phosphatase activity was measured at OD_{415} on a Bio-Tek EL-308 microtiter plate reader.

Monoclonal Antibody.

BALB/c mice were subcutaneously immunized with 0.1 mg MoFe protein without adjuvant every two weeks for a period of fourteen weeks. Each animal received 0.1 mg MoFe protein without adjuvant intraperitoneally seven and three days before splenectomy. Spleen cells from immunized mice were fused with CHO myeloma cells by using the polyethylene glycol (Kodak 1450) fusion method (38). Hybrid cells were selected in RPMI medium containing hypoxanthine-aminopterin-thymidine, and their culture supernatant fluids were analyzed for MoFe protein antibody by ELISA screening (described above). Hybrids which were positive for

MoFe protein antibody were subcloned. ELISA screening and western analysis was used to select a subclone that was producing an antibody exhibiting optimal specificity for the MoFe protein. Approximately 1 x 10^6 MoFe protein antibody-producing cells of this subclone were injected intraperitoneally into twenty BALB/c mice primed with pristane. Ascites fluid was clarified by centrifugation and brought to 50% saturation with ammonium sulfate. The precipitate was collected by centrifugation, dissolved in PBS to the original sample volume, and dialyzed against several changes of PBS. The preparation was passed through a 0.45 μm membrane and stored at 0 °C. Preliminary western analysis indicated that the monoclonal antibody was specific to the α-subunit of the MoFe protein (data not shown).

Using both the monoclonal and polyclonal antibodies, it is now possible to devise a sandwich ELISA which can be used to accurately quantitate the amount of MoFe protein in any given sample. In addition, the monoclonal antibody may also be useful as a tool to complement the structural/functional data provided by the site-directed mutagenesis studies of the MoFe protein.

APPENDIX VIII

The following is a list of mutagenic primers (mutenes) used in this study. The letter in parentheses refers to the nif gene for which the mutene is specific. A '*' designates the specific nucleotide(s) residue which was changed. The individual change specified by each mutene and the single-stranded DNA template that was used is given to the right of each primer sequence. Only those mutenes which are shown in bold type were successfully used to construct mutant strains.

```
        MUTENE                  ALTERATION        TEMPLATE
```

Mn 2 (D)

```
            *
ATC TCC GTC GAG TCC GAG     151Q --> 151E     Eco1A
```

Mn 3 (D)

```
             *
CAG TCC GAG TCC CCG ATC     154C --> 154S     Eco1A
```

Mn 4 (D)

```
        *
GGC CTG GTC GGC GAC GAC     159I --> 159V     Eco1A
```

Mn 5 (D)

```
             *
GGC CTG ATC GCC GAC GAC     160G --> 160A     Eco1A
```

Mn 6 (D)

```
              *
CTG ATC GGC GAG GAC ATC     161D --> 161E     Eco1A
```

Mn 8 (D)

```
             *
CCG GTC CGT TCC GAA GGC     183C --> 183S     Eco1A
```

Mn 9 (D)

*

GTC CGT TGC GAC GGC TTC 184E --> 184D

Mn 10 (D)

*

GGC GTT TCC GAG TCC CTG 191Q --> 191E Eco1A

Mn 11 (D)

*

GTC AAG CTG GAC CTG GTT 271N --> 271D Eco1A

Mn 12 (D)

*

CTG GTT CAC TCC TAC CGC 275C --> 275S Eco1A

Mn 13 (D)

*

CGC TCG ATG GAC TAC ATC 280N --> 280D Eco1A

Mn 14 (D)

*

CGT CAC ATG GAC GAG AAG 287E --> 287D

Mn 15 (K)

*

CCG GCC AAG GCT TCC CAG CCG 70C --> 70S Eco2B

Mn 16 (K)

*

GGT TCC CAG GGT TCC GTC GCC 95C --> 95S Eco2B

Mn 17 (K)

*

GTG TCC ACC ACC TCC ATG GCC 153C --> S153 Eco2B

Mn 18 (D)

*

GTA GCC CAG TGG GCC GGC GAC 254S --> 254A

Mn 19 (D)

*

CAG TGG TCC GGC AAC GGC TCC 256D --> 256N

Mn 20 (D)

*

CTG AAC CTG GTT CGC TGC TAC 274H --> 274R

Mn 21 (D)

*

CTG GTT CAC TGC TCC CGC TCG 276Y --> 276S Eco1A

Mn 22 (D)

* *

CTG GTT CAC TCC TGC CGC TCG 275S276Y --> 275S276C Mn12ss

Mn 23 (D)

*

GTT CAC TGC TAC CAC TCG ATG 277R --> 277H Eco1A

Mn 24 (D)

*

TGC TAC CGC TCG ATC AAC TAC 279M --> 279I Eco1A

Mn 25 (D)

*

CGC TCG ATG AAC TCC ATC TCC 281Y --> 281S

Mn 26 (D)

*

CAG TCC AAG AAG TCC ATC ATC 45C --> 45S 6Pst3

Mn 27 (D)

*

ACC ATC CGC GGC TCC GCC TAC 62C --> 62C 6Pst3

Mn 28 (D)

*

GGT CCG GTA GGC TCC GGC CAG 88C --> 88S 6Pst3

Mn 29 (D)

*

ATG GGC CTG CGT TCC GTA GCC 249C --> 249S

Mn 30 (D)

*

ATC CAG AAG AAG TCC GAA GAG 324C --> 324S

Mn 31 (D)

*

CAG TCC CTG GGC AAC CAC ATC 195H --> 195N Eco1A

Mn 32 (D)

*

AAC TAC ATC TCC CAT CAC ATG 284R --> 284H

Mn 33 (D)

*

ATC CAC ATC TCC AAC GGT CCG 83H --> 83N 6Pst3

Mn 34 (K)

*

ATG CCC TAC GTG AAC GGT TCC 9OH --> 90N Eco2B

Mn 35 (D)

*

ATC GGC GAC TAC GAC ATC GGC 230N --> 230D

Mn 36 (D)

*

CTG AAC CTG GTT AAC TGC TAC 274H --> 274N

Mn 37 (D)

**

CAG TCC AAG AAG GCC ATC ATC 45C --> 45A Mn26ss

Mn 38 (D)

**

ACC ATC CGC GGC GCC GCC TAC 62C --> 62A Mn27ss

Mn 39 (D)

* *

GGT CCG GTA GGC GCC GGC CAG 88C --> 88A Mn28ss

Mn 40 (D)

* *

GTC CAG TCC GAG GCC CCG ATC 154C --> 154A Mn3ss

Mn 41 (D)

**

GTA CCG GTC CGT GCC GAA GGC 183C --> 183A Mn8ss

Mn 42 (D)

* *

AAC CTG GTT CAC GCC TAC CGC 275C --> 275A Mn12ss

Mn 43 (K)

* *

CCG GCC AAG GCT GCC CAG CCG 70C --> 70A Mn15ss

Mn 44 (K)

* *

GGT TCC CAG GGT GCC GTC GCC 95C --> 95A

Mn 45 (K)

**

GTG TCC ACC ACC GCC ATG GCC 153C --> 153A Mn17ss

Mn 46 (D)

*

CAC ATC TCC CAG GGT CCG GTA 83H --> 83Q

Mn 47 (D)

*

CGC GGC GTT TCC AAG TCC CTG 191Q --> 191K Eco1A

Mn 48 (K)

*

CCC TAC GTG CAG GGT TCC CAG 90H --> 90Q

APPENDIX IX

The following is a list of sequencing primers used in this study. The letter in parentheses refers to the nif gene for which the primer is specific. The number above each primer sequence refers to the amino acid residue for which that codon specifies.

Pn 1 (D) 140
GAA GTG GAA ACC CTG TTC

Pn 7 (D) 180
ACC ATC GTA CCG GTC CGT

Pn 15 (D) 20
CTG GAA GTT TAT CCC GAG

Pn 16 (D) 60
ATG ACC ATC CGC GGC TGC

Pn 17 (D) 100
CGT AAC TAC TAC ATC GGT

Pn 18 (D) 220
GCC AGC ACT CCT TAC GAT

Pn 19 (D) 260
GGC TCC ATC TCC GAA ATC

Pn 20 (D) 300
TAC AAC TTC TTC GGC CCG

Pn 21 (D) 340
GCG GTG CTC GCC AAG TAC

Pn 22 (D) 380
ACC GGC TAC GAG TTC GCC

Pn 23 (D) 420
AAG CCC GAC CTG ATC GGC

Pn 24 (D) 460
GCC ATC TTC GCC CGT GAC

Pn 25 (K) 20
GAT CAG GAC TAC AAG GAC

Pn 26 (K) 60
TTC CAG CGC GAA GCC CTG

Pn 27 (K) 100
GCC TAC TTC CGC TCC TAC

Pn 28 (K) 140
AAC TGT AAG GCT ACC TAC

Pn 29 (K) 180
GAC GAG TTC CCG GTG CCG

Pn 30 (K) 220
GTG GTT GGC AGC AAC AAG

Pn 31 (K) 260
CCG GAA GAA GTG CTG GAC

Pn 32 (K) 300
CAC CTG GAG AAG ACC AAG

Pn 33 (K) 340
GGC CAG CCG ATT CCG GCG

Pn 34 (K) 380
ATG GGC CTG GTC AAG TTC

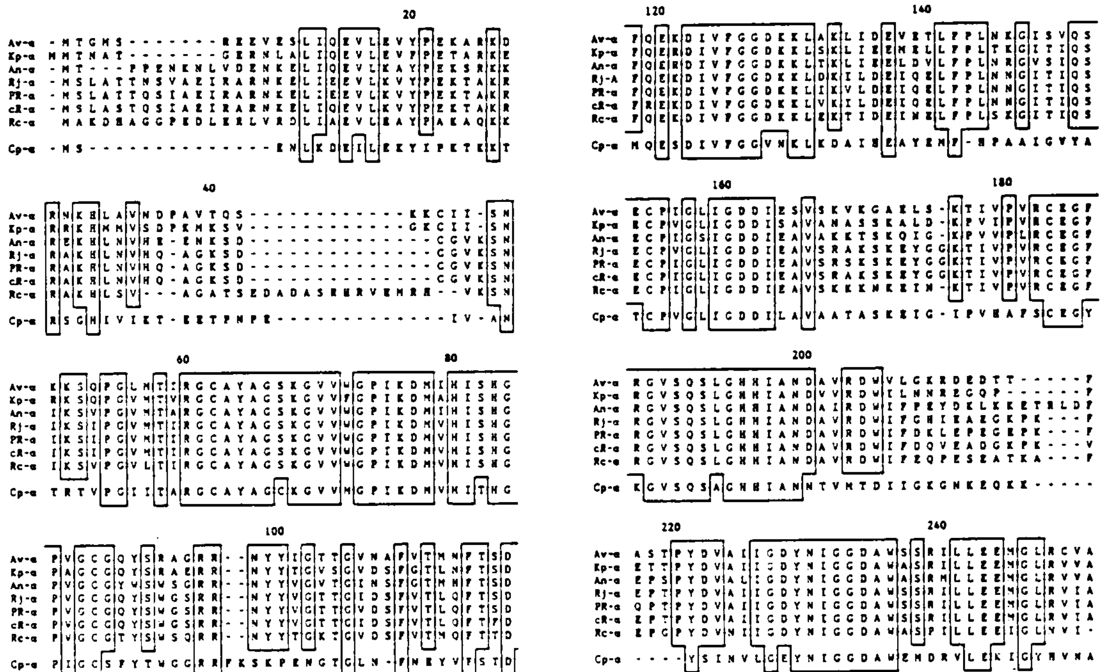
Pn 35 (K) 420
AAG AAT GCT ACC GTC TAC

Pn 36 (K) 460
CAC AAG GGC AAG GAG TTC

Pn 37 (K) 500
CTG GTG AAC TCG ATC CTG

APPENDIX X

Alignment of all MoFe protein α-subunit sequences including the questionable R. capsulatus sequence: A. vinelandii (Av), K. pneumoniae (Kp), Anabaena 7120 (An), Br. japonicum (Rj), Rhizobium sp. Parasponia (PR), cowpea Rhizobium strain IRc78 (cR), C. pasteurianum (Cp), and R. capsulatus (Rc). See text for references. Amino acid residues conserved in all eight organisms or in the seven organisms excluding C. pasteurianum are boxed. Dashed lines represent alignment adjustments to provide best fit.



```
                  260                 280
Av-α  QWSGDGSISEIELTP-KVKLNLVHCYRSMNYISRH
Kp-α  QWSGDGTLVEMENTP-FVKLNLVHCYRSMNYIARH
An-α  QWSGDGTLNELIQGP-AAKLVLIHCYRSMNYICRS
Rj-α  QWSGDGSLAELEATP-KLKLNILHCYRSMNYISRH
PR-α  QWSGDGSLAELEATP-KAKLNILHCYRSMNYISRH
cR-α  QWSGDGSLAELEANV-EGKLNILHCYRSMNYISRH
Rc-α  -WSGDATLAEMERAP-KAKLNLIHCYRSMNYICRY

Cp-α  TLTGDAT-YEKVQNADKADLNLVQCHRSINYIAEM

                        300
Av-α  MEEKYGIPWMEYNFFGPTKTIESLRAIAAKFDES-
Kp-α  MEEKHQIPWMEYNFFGPTKIAESLRKIADQFDDT-
An-α  LEEQYGMPWMEFNFFGPTKIAASLREIAAKFD-SK
Rj-α  MEEKFGIPWCEYNFFGPSKIADSLRRIAGYFDD-K
PR-α  MEEKFGIPWCEYNFFGPSKIAESLRKIAGYFDD-K
cR-α  MEEKFGIPWCEYNFFGPSKIAESLRRIAGYFDD-K
Rc-α  MEEKYSIGWMEYNFFGPTQIEASLR-I-GKFDET-

Cp-α  METKYGIPWIKCNFIGVDGIVETLRDMAKCFDDPE

     320                  340
Av-α  IQKKCEEVIAKYKPEWEAVVAKYRPRLEGKRVMLY
Kp-α  IRANAEAVIARYEGQMAAIIAKYRPRLEGRKVLLY
An-α  IQENAEKVIAKYTPVMNAVLDKYRPRLEGNTVMLY
Rj-α  IKEGAERVIEKYQPLVDAVIAKYRPRLEGKTVMLY
PR-α  IKEGAERVIEKYQPLVDAVIAKYRPRLEGKTVMLY
cR-α  IKEGAERVIEKYQPLVNAVIAKYRPRLEGKTVMLY
Rc-α  IQANVEKVIAKYRPLVDGILAKYKPRLEGKSVMLY

Cp-α  LTKRTEEVIAEEIAAIQDDLDYFKEKLCGKTACLY

          360                  380
Av-α  IGGLRPRHVIGAYEDLGMEVVGTGYEFAHNDDYD-
Kp-α  IGGLRPRHVIGAYEDLGMEIIAAGYEFAHNDDYD-
An-α  VGGLRPRHVVPAFEDLGIKVVGTGYEFAHNDDYK-
Rj-α  VGGLRPRHVIGAYEDLGMDVIGTGYEFGHNDDYQ-
PR-α  VGGLRPRHVIGAYEDLGMEVVGTGYEFGHNDDYQ-
cR-α  VGGLRSRHVIGAYEDLGMEVIGTGYEFGHNDDYQ-
Rc-α  V-RPRSAPRRHAYDDLGMVIAGTGYEFAHNDDYK-

Cp-α  VGGSRSHTYMNMLKSFGVDSLVAGFEFAHRDDYEG

Av-α  RT---------------------------------
Kp-α  RT---------------------------------
An-α  RT---------------------------------
Rj-α  RT---------------------------------
PR-α  RT---------------------------------
cR-α  RT---------------------------------
Rc-α  RT---------------------------------

Cp-α  REVIPTIKIDADSKNIPEITVTPDEQKYRVVIPED

                               400
Av-α  ------------------MKEMGDSTLLYDDVTGY
Kp-α  ------------------LPDLKEGTLLFDDASSY
An-α  ------------------THYIDNATIIYDDVTAY
Rj-α  -----------------AQEYVKDSTLIYDDVNGY
PR-α  -----------------AQEYVKDSTLIYDDVNGY
cR-α  -----------------AQEYVKDGTLIHDDVNGY
Rc-α  ------------------GHYVKEGT-IYDDVTGY

Cp-α  KVEELKKAGVPLSSYGGMMKEMKDGTILIDDMNEH

                420                440
Av-α  EFEEFVKRIKPDLIGSGIKE-KFIFQKMGIPFRQM
Kp-α  ELEAFVKALKPDLIGSGIKE-KYIFQKMGVPFRQM
An-α  EFEEFVKAKKPDLIASGIKE-KYVFQKMGLPFRQM
Rj-α  EFERFVERLQPDLVGSGIKE-KYVFQKMSVPFRQM
PR-α  EFERFVEKVQPDLVGSGIKE-KYVFQKMGVPFPEM
cR-α  EFERFVEKLQPDLVGSGIKE-KYVFQKMGGPFRQM
Rc-α  ELEKFIEKIRPDLVASGIKEEKYPVQKMGIPFRQM

Cp-α  DMEVVLEKLKPDMPFAGIKE-KFVIQKGGVLSKQL

                           460
Av-α  HSWDYSGPYHGFDGFAIFARDM-DMTLNNPCWKKL
Kp-α  HSWDYSGPYHGYDGFAIFARDM-DMTLNNPAWNEL
An-α  HSWDYSGPYHGYDGFAIFARDM-DLSLNSPTWSLI
Rj-α  HSWDYSGPYHGYDGFAIFARDM-DMAVNSPIWKET
PR-α  HSWDYSGPYHGYDGFAIFARDM-DMAVNSPIWKET
cR-α  HSWDYSGPYHGYDGFAIFARDM-DMAINSPVWKET
Rc-α  HSWDYSGPYHGYWASPILLDAPWDKPHRASSWKIT

Cp-α  HSYDYNGPYAGFRGVVNFGREL-VNGIYTPAWKMI

         480
Av-α  Q-APWEASEGAEKVAASA
Kp-α  T-APWLKSA
An-α  G-APWKKAA-AKAKAAS
Rj-α  K-APWKDAERQDSRLQNNATELALEESPGIPI
PR-α  K-APWKEAAKPKLLAAE
cR-α  K-APWKEASRAKLLAAE
Rc-α  PAAPWKAASEA

Cp-α  T-PPWKKASSESKVVVGGEA
```

APPENDIX XI

The complete nucleotide sequence and deduced amino acid sequence of a 28,793 base-pair region from the A. vinelandii genome carrying the nitrogenase structural gene region. The transcriptional direction of the individual nif genes and ORF's are indicated by arrows. Potential nif promoter sequences (CTGG-N_8-TTGCA) and upstream activator sequences (TGT-N_{10}-ACA) are under- and overscored. ORF12 initiates with a valine residue. The author was directly responsible for determining the nucleotide sequence between residues 1 - 12,567.

```
                                                  50
          *         *         *         *         *         *         *         *         *
GCGCTCGCAGACCTTGCGTTCTTTGTCGCTCGGCTGATCGATCAGGACCCAGCCGGCCGGTTCGGCGGCATCGTAGATCATGTCGCTCAT
<ArgGluCysValLysArgGluLysAspSerProGlnAspIleLeuValTrpGlyAlaProGluAlaAlaAspTyrIleMetAspSerMet

         100                                       150
          *         *         *         *         *         *         *         *         *
CACCATGCGTTCGGTGTCGCGGTTGAAGCGCAGGCCCAGGAACAGGTATCGCTTGTTCTTGCGCAGCAGCTTGACCCAGTCGGGAATGGC
<ValMetArgGluThrAspArgAsnPheArgLeuGlyLeuPheLeuTyrArgLysAsnLysArgLeuLeuLysValTrpAspProIleAla

                    200                                       250
          *         *         *         *         *         *         *         *         *
GAAGCCGCCCATCAGCTCGGTGATGTAGTCGACGAAGTCGGCATCCGAGGCCACGTAGCCCGGCTTGGGCAGCGGCGTGCCGAGCGGTTT
<PheGlyGlyMetLeuGluThrIleTyrAspValPheAspAlaAspSerAlaValTyrGlyProLysProLeuProThrGlyLeuProLys

                              300                                       350
          *         *         *         *         *         *         *         *         *
GAACAGTACCGGCAGGGCGCTATCGACCTCTTCCAGCGCGACCTTGCGGTACTTGCCGTCCTCGAAGTGGTAGAGGTCGAAGCGATAGGG
<PheLeuValProLeuAlaSerAspValGluGluLeuAlaValLysArgTyrLysGlyAspGluPheHisTyrLeuAspPheArgTyrPro

                                        400                                       450
          *         *         *         *         *         *         *         *         *
CGTCGCGGCGAGGCGCGCGGCGCCGACCACCAGGGTGTGCCGGCGCCCGGCGTCGATGTAGCAACGCTGCAGCTGGGTGTCGCGGTTGCT
<ThrAlaAlaLeuArgAlaAlaGlyValValLeuThrHisArgArgGlyAlaAspIleTyrCysArgGlnLeuGlnThrAspArgAsnSer

                                                  500
          *         *         *         *         *         *         *         *         *
GTCGATGACGTAGGGCAGGTCGAGTCCGGCCAGCCACTGGTGGACCGGAGATGCGCTCCAGTTGTCGCCACCGTAGGTCTGGGTCAGGAA
<AspIleValTyrProLeuAspLeuGlyAlaLeuTrpGlnHisValProSerAlaSerTrpAsnAspGlyGlyTyrThrGlnThrLeuPhe

         550                                       600
          *         *         *         *         *         *         *         *         *
GCGTTCGATGAAGCTGCGGCCTTTCTTGTTCTCCACGTGCATGGCCGCACGCGGGAACTCGTACATGAGACGCGGCGACATCGGCCGACC
<ArgGluIlePheSerArgGlyLysLysAsnGluValHisMetAlaAlaArgProPheGluTyrMetLeuArgProSerMetProArgGly

                    650                                       700
          *         *         *         *         *         *         *         *         *
GCCGGTCATGGCGAGAATCAGGCTGTCGCTGTCGGCGGGAATCGGCTTGCCCGTGTCGCGATCGACGACTCCGCGCAGGACGCCCGGGCC
<GlyThrMetAlaLeuIleLeuSerAspSerAspAlaProIleProLysGlyThrAspArgAspValValGlyArgLeuValGlyProGly

                              750                          <—— 12          800
          *         *         *         *         *         *         *         *         *
CAGATAGGGAACGATGTCGCCCGAGCCGAGCTGGGCGAGGATTTCCTTTAATAAGCTGTCGGTCACTGAACTCTCCTGCTGAGGGAAGGG
<LeuTyrProValIleAspGlySerGlyLeuGlnAlaLeuIleGluLysLeuLeuSerAspThrVal
```

```
                                850                                                  900
          *         *         *         *         *         *         *         *         *
CAAGAATCGACAACCTATTGCAATAAGTGTGCCGAAAGGATTTCGTTGTTTAACTAATTGAATTTAAAAGAAATCATTGGTGATTTCGGA

                                          950
          *         *         *         *         *         *         *         *         *
ATGGCTTGTCGTATCCGTGGGCCAGGATGGGGCGTGGCTTCACGACAATTGTCAGTTTTGTCACAGGGGGCCGGACCAGGATGGTGGACG

     1000                                          1050
          *         *         *         *         *         *         *         *         *
CTCGATGGGGATGTCGGGCCATTGTTCGGTTGTAGCAATTACAACAGTCGGAGTAGGGGGATTGTAGGGGGATTGTTGTGTATCAGACCG

               1100                                          1150
          *         *         *         *         *         *         *         *         *
CCCTGCAGCTCCCGTCGATGGATAATTAATCATTTAAAATCAATGGTTTATTTATGTGTTGCGGGTGCTGGCACAGACGCTGCATTACCT

                         1200                              H──►  1250
          *         *         *         *         *         *         *         *         *
TTGGTGCGCGGAGTTGTTCGGGCTTACGGCCGAACGTTCAAGTGGAAATGCAACCTGAGGAAATTAACTATGGCTATGCGTCAATGCGCC
                                                                        MetAlaMetArgGlnCysAla>

                                1300                                                1350
          *         *         *         *         *         *         *         *         *
ATCTACGGCAAAGGTGGTATCGGTAAGTCCACCACTACTCAGAACCTGGTGGCAGCCCTGGCTGAGATGGCAAGAAGGTCATGATCGTT
IleTyrGlyLysGlyGlyIleGlyLysSerThrThrThrGlnAsnLeuValAlaAlaLeuAlaGluMetGlyLysLysValMetIleVal>

                                          1400
          *         *         *         *         *         *         *         *         *
GGTTGTGACCCGAAAGCTGACTCCACCCGCCTGATCCTGCACTCCAAGGCCCAGAACACCATCATGGAAATGGCTGCCGAAGCCGGTACC
GlyCysAspProLysAlaAspSerThrArgLeuIleLeuHisSerLysAlaGlnAsnThrIleMetGluMetAlaAlaGluAlaGlyThr>

     1450                                          1500
          *         *         *         *         *         *         *         *         *
GTGGAAGATCTGGAGCTGGAAGACGTGCTGAAGGCTGGCTACGGCGGCGTCAAGTGCGTTGAGTCCGGTGGTCCGGAGCCGGGCGTTGGC
ValGluAspLeuGluLeuGluAspValLeuLysAlaGlyTyrGlyGlyValLysCysValGluSerGlyGlyProGluProGlyValGly>

               1550                                          1600
          *         *         *         *         *         *         *         *         *
TGCGCCGGCCGTGGTGTTATCACCGCCATCAACTTCCTGGAAGAGGAAGGCGCCTACGAAGACGATCTGGACTTCGTATTCTACGACGTG
CysAlaGlyArgGlyValIleThrAlaIleAsnPheLeuGluGluGluGlyAlaTyrGluAspAspLeuAspPheValPheTyrAspVal>

                         1650                                          1700
          *         *         *         *         *         *         *         *         *
CTGGGCGACGTGGTGTGTGGCGGCTTCGCCATGCCGATCCGCGAGAACAAGGCCCAGGAAATCTACATCGTCTGCTCCGGTGAGATGATG
LeuGlyAspValValCysGlyGlyPheAlaMetProIleArgGluAsnLysAlaGlnGluIleTyrIleValCysSerGlyGluMetMet>

                                1750                                                1800
          *         *         *         *         *         *         *         *         *
GCCATGTACGCCGCCAACAACATCTCCAAGGGCATCGTGAAGTATGCCAACTCCGGCAGCGTGCGTCTGGGCGGCCTGATCTGCAACAGC
AlaMetTyrAlaAlaAsnAsnIleSerLysGlyIleValLysTyrAlaAsnSerGlySerValArgLeuGlyGlyLeuIleCysAsnSer>
```

```
                                                  1850
         *         *         *         *         *         *         *         *         *
CGTAACACCGACCGCGAAGACGAGCTGATCATCGCTCTGGCCAACAAGCTGGGCACCCAGATGATCCACTTCGTGCCGCGTGACAACGTC
ArgAsnThrAspArgGluAspGluLeuIleIleAlaLeuAlaAsnLysLeuGlyThrGlnMetIleHisPheValProArgAspAsnVal>

       1900                                          1950
         *         *         *         *         *         *         *         *         *
GTGCAGCGCGCCGAAATCCGCCGCATGACCGTGATCGAATACGATCCGAAAGCCAAGCAAGCCGACGAATACCGCGCTCTGGCCCGCAAG
ValGlnArgAlaGluIleArgArgMetThrValIleGluTyrAspProLysAlaLysGlnAlaAspGluTyrArgAlaLeuAlaArgLys>

                 2000                                          2050
         *         *         *         *         *         *         *         *         *
GTCGTCGACAACAAACTGCTGGTCATCCCGAACCCGATCACCATGGACGAGCTCGAAGAGCTGCTGATGGAATTCGGCATCATGGAAGTC
ValValAspAsnLysLeuLeuValIleProAsnProIleThrMetAspGluLeuGluGluLeuLeuMetGluPheGlyIleMetGluVal>

                           2100                                          2150
         *         *         *         *         *         *         *         *         *
GAAGACGAATCCATCGTCGGCAAAACCGCCGAAGAAGTCTGATAGCCGCTCCGGTTTCAGAAGGACGGGACAGGGCAGATTGGCTCTGTC
GluAspGluSerIleValGlyLysThrAlaGluGluValEndEnd>

                                     2200                                  D ->      2250
         *         *         *         *         *         *         *         *         *
GGGGTGGCGCCCCCCGCATTGGGCGGGCGCCCACCCGTTACCCGCATATGAACGCTAAGGCAAGAGGAGTCATACCCATGACCGGTATGT
                                                                              MetThrGlyMet>

                                                  2300
         *         *         *         *         *         *         *         *         *
CGCGCGAAGAGGTTGAATCCCTCATCCAGGAAGTTCTGGAAGTTTATCCCGAGAAGGCTCGCAAGGATCGTAACAAGCACCTGGCCGTCA
SerArgGluGluValGluSerLeuIleGlnGluValLeuGluValTyrProGluLysAlaArgLysAspArgAsnLysHisLeuAlaVal>

       2350                                          2400
         *         *         *         *         *         *         *         *         *
ACGACCCGGCGGTTACCCAGTCCAAGAAGTGCATCATCTCCAACAAGAAGTCCCAGCCCGGTCTGATGACCATCCGCGGCTGCGCCTACG
AsnAspProAlaValThrGlnSerLysLysCysIleIleSerAsnLysLysSerGlnProGlyLeuMetThrIleArgGlyCysAlaTyr>

                 2450                                          2500
         *         *         *         *         *         *         *         *         *
CCGGTTCCAAAGGCGTGGTCTGGGGCCCCATCAAGGACATGATCCACATCTCCCACGGTCCGGTAGGCTGCGGCCAGTATTCGCGCGCCG
AlaGlySerLysGlyValValTrpGlyProIleLysAspMetIleHisIleSerHisGlyProValGlyCysGlyGlnTyrSerArgAla>

                           2550                                          2600
         *         *         *         *         *         *         *         *         *
GCCGTCGTAACTACTACATCGGTACCACCGGTGTGAACGCCTTCGTCACCATGAACTTCACCTCGGACTTCCAGGAGAAGGACATCGTGT
GlyArgArgAsnTyrTyrIleGlyThrThrGlyValAsnAlaPheValThrMetAsnPheThrSerAspPheGlnGluLysAspIleVal>

                                     2650                                                2700
         *         *         *         *         *         *         *         *         *
TCGGTGGCGACAAGAAGCTCGCCAAACTGATCGACGAAGTGGAAACCCTGTTCCCGCTGAACAAGGGTATCTCCGTCCAGTCCGAGTGCC
PheGlyGlyAspLysLysLeuAlaLysLeuIleAspGluValGluThrLeuPheProLeuAsnLysGlyIleSerValGlnSerGluCys>
```

```
                                              2750
      *         *         *         *         *         *         *         *         *
CGATCGGCCTGATCGGCGACGACATCGAATCCGTGTCCAAGGTCAAGGGCGCCGAGCTCAGCAAGACCATCGTACCGGTCCGTTGCGAAG
ProIleGlyLeuIleGlyAspAspIleGluSerValSerLysValLysGlyAlaGluLeuSerLysThrIleValProValArgCysGlu>

    2800                                          2850
      *         *         *         *         *         *         *         *         *
GCTTCCGCGGCGTTTCCCAGTCCCTGGGCCACCACATCGCCAACGACGCAGTCCGCGACTGGGTCCTGGGCAAGCGTGACGAAGACACCA
GlyPheArgGlyValSerGlnSerLeuGlyHisHisIleAlaAsnAspAlaValArgAspTrpValLeuGlyLysArgAspGluAspThr>

              2900                                          2950
      *         *         *         *         *         *         *         *         *
CCTTCGCCAGCACTCCTTACGATGTGGCCATCATCGGCGACTACAACATCGGCGGCGACGCCTGGTCTTCCCGCATCCTGCTGGAAGAAA
ThrPheAlaSerThrProTyrAspValAlaIleIleGlyAspTyrAsnIleGlyGlyAspAlaTrpSerSerArgIleLeuLeuGluGlu>

                        3000                                          3050
      *         *         *         *         *         *         *         *         *
TGGGCCTGCGTTGCGTAGCCCAGTGGTCCGGCGACGGCTCCATCTCCGAAATCGAGCTGACCCCGAAGGTCAAGCTGAACCTGGTTCACT
MetGlyLeuArgCysValAlaGlnTrpSerGlyAspGlySerIleSerGluIleGluLeuThrProLysValLysLeuAsnLeuValHis>

                                  3100                                          3150
      *         *         *         *         *         *         *         *         *
GCTACCGCTCGATGAACTACATCTCCCGTCACATGGAAGAGAAGTACGGTATCCCATGGATGGAGTACAACTTCTTCGGCCCGACCAAGA
CysTyrArgSerMetAsnTyrIleSerArgHisMetGluGluLysTyrGlyIleProTrpMetGluTyrAsnPhePheGlyProThrLys>

                                            3200
      *         *         *         *         *         *         *         *         *
CCATCGAGTCGCTGCGTGCCATCGCCGCCAAGTTCGACGAGAGCATCCAGAAGAAGTGCGAAGAGGTCATCGCCAAGTACAAGCCCGAGT
ThrIleGluSerLeuArgAlaIleAlaAlaLysPheAspGluSerIleGlnLysLysCysGluGluValIleAlaLysTyrLysProGlu>

    3250                                          3300
      *         *         *         *         *         *         *         *         *
GGGAAGCGGTGGTCGCCAAGTACCGTCCGCGCCTGGAAGGCAAGCGCGTCATGCTCTACATCGGTGGCCTGCGTCCGCGCCACGTGATCG
TrpGluAlaValValAlaLysTyrArgProArgLeuGluGlyLysArgValMetLeuTyrIleGlyGlyLeuArgProArgHisValIle>

              3350                                          3400
      *         *         *         *         *         *         *         *         *
GCGCCTACGAAGACCTGGGCATGGAAGTGGTGGGTACCGGCTACGAGTTCGCCCACAACGACGACTATGACCGCACCATGAAAGAAATGG
GlyAlaTyrGluAspLeuGlyMetGluValValGlyThrGlyTyrGluPheAlaHisAsnAspAspTyrAspArgThrMetLysGluMet>

                        3450                                          3500
      *         *         *         *         *         *         *         *         *
GTGACTCCACCCTGCTGTACGATGACGTGACCGGCTACGAATTCGAAGAATTCGTCAAGCGCATCAAGCCCGACCTGATCGGCTCCGGTA
GlyAspSerThrLeuLeuTyrAspAspValThrGlyTyrGluPheGluGluPheValLysArgIleLysProAspLeuIleGlySerGly>

                                  3550                                          3600
      *         *         *         *         *         *         *         *         *
TCAAGGAGAAGTTCATCTTCCAGAAGATGGGCATCCCCTTCCGTGAAATGCACTCCTGGGATTATTCCGGCCCCTACCACGGCTTCGATG
IleLysGluLysPheIlePheGlnLysMetGlyIleProPheArgGluMetHisSerTrpAspTyrSerGlyProTyrHisGlyPheAsp>
```

```
                                          3650
       *         *         *         *         *         *         *         *         *
GCTTCGCCATCTTCGCCCGTGACATGGACATGACCCTGAACAATCCGTGCTGGAAGAAACTGCAGGCTCCCTGGGAAGCTTCCGAAGGCG
GlyPheAlaIlePheAlaArgAspMetAspMetThrLeuAsnAsnProCysTrpLysLysLeuGlnAlaProTrpGluAlaSerGluGly>

     3700                                         3750
       *         *         *         *         *         *         *         *         *
CCGAGAAAGTCGCCGCCAGCGCCTGATAGCAGAGCAATCGTACGCAACGTCCGCTGCGGGCGGTTTCCGCCCGCGGCCGACATCCGCTAA
AlaGluLysValAlaAlaSerAlaEndEnd>

               3800                K ──►                     3850
       *         *         *         *         *         *         *         *         *
CGCCGTTCACAGATGAGTGAGGCGTAGGAGAGAGTCATGAGCCAGCAAGTCGATAAAATCAAAGCCAGCTACCCGCTGTTCCTCGATCAG
                                    MetSerGlnGlnValAspLysIleLysAlaSerTyrProLeuPheLeuAspGln>

                         3900                                                  3950
       *         *         *         *         *         *         *         *         *
GACTACAAGGACATGCTTGCCAAGAAGCGCGACGGCTTCGAGGAAAAGTATCCGCAGGACAAGATCGACGAAGTATTCCAGTGGACCACC
AspTyrLysAspMetLeuAlaLysLysArgAspGlyPheGluGluLysTyrProGlnAspLysIleAspGluValPheGlnTrpThrThr>

                                   4000                                                  4050
       *         *         *         *         *         *         *         *         *
ACCAAGGAATACCAGGAGCTGAACTTCCAGCGCGAAGCCCTGACCGTCAACCCGGCCAAGGCTTGCCAGCCGCTGGGCGCCGTTCTCTGC
ThrLysGluTyrGlnGluLeuAsnPheGlnArgGluAlaLeuThrValAsnProAlaLysAlaCysGlnProLeuGlyAlaValLeuCys>

                                             4100
       *         *         *         *         *         *         *         *         *
GCCCTCGGTTTCGAGAAGACCATGCCCTACGTGCACGGTTCCCAGGGTTGCGTCGCCTACTTCCGCTCCTACTTCAACCGTCATTTCCGC
AlaLeuGlyPheGluLysThrMetProTyrValHisGlySerGlnGlyCysValAlaTyrPheArgSerTyrPheAsnArgHisPheArg>

     4150                                         4200
       *         *         *         *         *         *         *         *         *
GAGCCGGTTTCCTGCGTTTCCGACTCCATGACCGAAGACGCGGCAGTGTTCGGCGGCCAGCAGAACATGAAGGACGGTCTGCAGAACTGT
GluProValSerCysValSerAspSerMetThrGluAspAlaAlaValPheGlyGlyGlnGlnAsnMetLysAspGlyLeuGlnAsnCys>

               4250                                                4300
       *         *         *         *         *         *         *         *         *
AAGGCTACCTACAAGCCCGACATGATCGCAGTGTCCACCACCTGCATGGCCGAGGTCATCGGTGACGACCTCAACGCCTTCATCAACAAC
LysAlaThrTyrLysProAspMetIleAlaValSerThrThrCysMetAlaGluValIleGlyAspAspLeuAsnAlaPheIleAsnAsn>

                         4350                                                  4400
       *         *         *         *         *         *         *         *         *
TCGAAGAAGGAAGGTTTCATTCCTGACGAGTTCCCGGTGCCGTTCGCCCATACCCCGAGCTTCGTGGGCAGCCACGTGACCGGCTGGGAC
SerLysLysGluGlyPheIleProAspGluPheProValProPheAlaHisThrProSerPheValGlySerHisValThrGlyTrpAsp>

                                   4450                                                  4500
       *         *         *         *         *         *         *         *         *
AACATGTTCGAAGGCATTGCTCGCTACTTCACCCTGAAGTCCATGGACGACAAGGTGGTTGGCAGCAACAAGAAGATCAACATCGTCCCC
AsnMetPheGluGlyIleAlaArgTyrPheThrLeuLysSerMetAspAspLysValValGlySerAsnLysLysIleAsnIleValPro>
```

```
                                                                    4550
          *         *         *         *         *         *         *         *         *
GGCTTCGAGACCTACCTGGGCAACTTCCGCGTGATCAAGCGCATGCTTTCGGAAATGGGCGTGGGCTACAGCCTGCTCTCCGATCCGGAA
GlyPheGluThrTyrLeuGlyAsnPheArgValIleLysArgMetLeuSerGluMetGlyValGlyTyrSerLeuLeuSerAspProGlu>

      4600                                                4650
          *         *         *         *         *         *         *         *         *
GAAGTGCTGGACACCCCGGCTGACGGCCAGTTCCGCATGTACGCGGGCGGCACCACTCAGGAAGAGATGAAGGACGCTCCGAACGCCCTC
GluValLeuAspThrProAlaAspGlyGlnPheArgMetTyrAlaGlyGlyThrThrGlnGluGluMetLysAspAlaProAsnAlaLeu>

                4700                                                4750
          *         *         *         *         *         *         *         *         *
AACACCGTCCTGCTGCAGCCGTGGCACCTGGAGAAGACCAAGAAGTTCGTCGAGGGTACCTGGAAGCACGAAGTACCGAAGCTGAACATC
AsnThrValLeuLeuGlnProTrpHisLeuGluLysThrLysLysPheValGluGlyThrTrpLysHisGluValProLysLeuAsnIle>

                          4800                                                4850
          *         *         *         *         *         *         *         *         *
CCGATGGGCCTGGACTGGACCGACGAGTTCCTGATGAAAGTCAGCGAAATCAGCGGCCAGCCGATTCCGGCGAGCCTGACCAAGGAGCGT
ProMetGlyLeuAspTrpThrAspGluPheLeuMetLysValSerGluIleSerGlyGlnProIleProAlaSerLeuThrLysGluArg>

                                    4900                                                4950
          *         *         *         *         *         *         *         *         *
GGCCGTCTGGTCGACATGATGACCGACTCCCACACCTGGCTGCACGGCAAGCGTTTCGCCCTGTGGGGTGATCCGGACTTCGTGATGGGC
GlyArgLeuValAspMetMetThrAspSerHisThrTrpLeuHisGlyLysArgPheAlaLeuTrpGlyAspProAspPheValMetGly>

                                              5000
          *         *         *         *         *         *         *         *         *
CTGGTCAAGTTCCTGCTGGAACTGGGTTGCGAGCCGGTACACATTCTCTGCCACAACGGCAACAAGCGTTGGAAGAAGGCGGTCGACGCC
LeuValLysPheLeuLeuGluLeuGlyCysGluProValHisIleLeuCysHisAsnGlyAsnLysArgTrpLysLysAlaValAspAla>

      5050                                                5100
          *         *         *         *         *         *         *         *         *
ATCCTCGCCGCTTCGCCCTACGGCAAGAATGCTACCGTCTACATCGGCAAGGACCTGTGGCACCTGCGTTCGCTGGTCTTCACCGACAAG
IleLeuAlaAlaSerProTyrGlyLysAsnAlaThrValTyrIleGlyLysAspLeuTrpHisLeuArgSerLeuValPheThrAspLys>

                5150                                                5200
          *         *         *         *         *         *         *         *         *
CCGGACTTCATGATCGGCAACAGCTACGGTAAGTTCATCCAGCGCGACACCCTGCACAAGGGCAAGGAGTTCGAGGTTCCGCTGATCCGT
ProAspPheMetIleGlyAsnSerTyrGlyLysPheIleGlnArgAspThrLeuHisLysGlyLysGluPheGluValProLeuIleArg>

                          5250                                                5300
          *         *         *         *         *         *         *         *         *
ATCGGCTTCCCGATCTTCGACCGTCATCACCTGCATCGCTCCACCACCCTGGGTTACGAGGGCGCCATGCAGATCCTGACCACCCTGGTG
IleGlyPheProIlePheAspArgHisHisLeuHisArgSerThrThrLeuGlyTyrGluGlyAlaMetGlnIleLeuThrThrLeuVal>

                                    5350                                                5400
          *         *         *         *         *         *         *         *         *
AACTCGATCCTGGAACGTCTGGACGAGGAAACCCGCGGTATGCAGGCCACCGACTACAACCACGACCTGGTACGCTAAGTCGTCGGTTCA
AsnSerIleLeuGluArgLeuAspGluGluThrArgGlyMetGlnAlaThrAspTyrAsnHisAspLeuValArgEnd>
```

```
                                         5450
        *         *         *         *         *         *         *         *         *
AGTGGTATCGGCCGGAGCGGCGCAAGCCGCTCCGGCTCTTCCTTGGCGGCGGCCGCAGGTGGTCGGGCCTTTTGCCCGCGATCTGCGGCA

     5500               T -->                   5550
        *         *         *         *         *         *         *         *         *
ACCGCCAAACCCGTCTAAGGAGCAAGCCCATGCCCAGCGTCATGATTCGCCGCAACGACGAAGGCCAACTGACCTTCTATATCGCCAAGA
                             MetProSerValMetIleArgArgAsnAspGluGlyGlnLeuThrPheTyrIleAlaLys>

              5600                                        5650
        *         *         *         *         *         *         *         *         *
AAGACCAGGAAGAGATCGTGGTGTCCCTGGAGCATGACAGCCCCGAACTCTGGGGTGGCGAAGTCACCCTCGGCGACGGTTCGACCTATT
LysAspGlnGluGluIleValValSerLeuGluHisAspSerProGluLeuTrpGlyGlyGluValThrLeuGlyAspGlySerThrTyr>

                        5700                                      Y --> 5750
        *         *         *         *         *         *         *         *         *
TCATCGAGCCGATACCGCAACCCAAGCTGCCGATCACCGTGCGCGCCAAGCGAGCCGGCGAGGCCTGATTCCATGACTGCGCAACCGCCA
PheIleGluProIleProGlnProLysLeuProIleThrValArgAlaLysArgAlaGlyGluAlaEnd    MetThrAlaGlnProPro>

                                  5800                                                 5850
        *         *         *         *         *         *         *         *         *
TTCGGCCAGGCCCCCCTGCCGGCCCATCTGGCCCTGCGCATCGCTCTGGCGGCACGCTCGCTCAAGGGGGTGGATACCGCCCACCTGCTG
PheGlyGlnAlaProLeuProAlaHisLeuAlaLeuArgIleAlaLeuAlaAlaArgSerLeuLysGlyValAspThrAlaHisLeuLeu>

                                            5900
        *         *         *         *         *         *         *         *         *
CGCGCCCTGATTGCCGCCGTTGGAGAACCCATCACGGAAGCGCGTCTACGCAAATTGCGTGCTTCCCGCCTGCGTACCCGCCTGCTCGAA
ArgAlaLeuIleAlaAlaValGlyGluProIleThrGluAlaArgLeuArgLysLeuArgAlaSerArgLeuArgThrArgLeuLeuGlu>

     5950                                                6000
        *         *         *         *         *         *         *         *         *
ACGTGCGGAGAGGGTGTGCAGTCGACCCTGACCGATCGCCAGCTGCACAGCGCGCTCGGTCTGCTCAAAGGCCGTGGCGTGCGCATGCCG
ThrCysGlyGluGlyValGlnSerThrLeuThrAspArgGlnLeuHisSerAlaLeuGlyLeuLeuLysGlyArgGlyValArgMetPro>

              6050                                        6100
        *         *         *         *         *         *         *         *         *
GAAGATCCCCTGCCGATCCCGGAGCCCTATCGCAACGGTGAATTCCAGGACTCGGTGCGCATCGCCTGCGCTTCCGACAACGGCGAGCGC
GluAspProLeuProIleProGluProTyrArgAsnGlyGluPheGlnAspSerValArgIleAlaCysAlaSerAspAsnGlyGluArg>

                        6150                                                6200
        *         *         *         *         *         *         *         *         *
CTGGACGGCATCTTCAGCAACTGTACACGCTTTCTCATCTACCAGATCTCGCCCCGCGAGAGCCGCCTGATCGATCTACGCGAACCCGGC
LeuAspGlyIlePheSerAsnCysThrArgPheLeuIleTyrGlnIleSerProArgGluSerArgLeuIleAspLeuArgGluProGly>

                                  6250                                                 6300
        *         *         *         *         *         *         *         *         *
CCCTGCCGCGAGGACGAAGACCGTCATGCGCGTCGGGCAGAATTGCTCGCCGACTGCCAGTTGCTCTACACCCTGAGCATCGGTGGCCCG
ProCysArgGluAspGluAspArgHisAlaArgArgAlaGluLeuLeuAlaAspCysGlnLeuLeuTyrThrLeuSerIleGlyGlyPro>
```

```
                                            6350
          *         *         *         *         *         *         *         *         *
GCGGCGGCCAAGGTGGTGCGTGCCGGGGTCCATCCCGTGCGCCTGGCCCGGGCCAGGCCGGCCCGCGAGATCGTCGAGGAGCTGCAGCGG
AlaAlaAlaLysValValArgAlaGlyValHisProValArgLeuAlaArgAlaArgProAlaArgGluIleValGluGluLeuGlnArg>

        6400                                          6450
          *         *         *         *         *         *         *         *         *
GTGCTGGCCACCGCGCCGCCCCCCTGGCTGGCCAAGGCCATGGGTGCCGAGCCAGACCAGCGCATACGTTTCACCCAGTAACCGCTCCGG
ValLeuAlaThrAlaProProProTrpLeuAlaLysAlaMetGlyAlaGluProAspGlnArgIleArgPheThrGlnEnd>

              1  6500                                                  6550
          *   ----->*         *         *         *         *         *         *         *
ATGAGGACACAGCCATGATCAGCCAAACACTTCTCGACGATGTCGTTCGCCAGGCCGAACAGGGAGTGCTGGGCGAAAGCCTACTCGCCA
              MetIleSerGlnThrLeuLeuAspAspValValArgGlnAlaGluGlnGlyValLeuGlyGluSerLeuLeuAla>

                              6600                                                  6650
          *         *         *         *         *         *         *         *         *
GCCTGCGCACGGCCCATCCGGGGGTGCACTTCACCTGCTGCATGGACGACGACGTCGTGGTCAACGCCAAGCCGGTCGTCGAGCGTCCGG
SerLeuArgThrAlaHisProGlyValHisPheThrCysCysMetAspAspAspValValValAsnAlaLysProValValGluArgPro>

                                        6700                                                6750
          *         *         *         *         *         *         *         *         *
GCTTCAACGTCTATCTGGTCAACTCCTGCGACCACTGCTCGGTGCTGACCAACGATCTGGACTCCGCTTCCGGCATTGTCCTGGCGGAAA
GlyPheAsnValTyrLeuValAsnSerCysAspHisCysSerValLeuThrAsnAspLeuAspSerAlaSerGlyIleValLeuAlaGlu>

                      2 ----->                     6800
          *         *         *         *         *         *         *         *         *
TCATTGAGGATTGAGGTACGCATGGCCGACGATCGAGGCGACATCACCCCCATCGGCGATTGCCGGGTCTGCAGCTTCCGCATGAGCCTG
IleIleGluAspEnd       MetAlaAspAspArgGlyAspIleThrProIleGlyAspCysArgValCysSerPheArgMetSerLeu>

        6850                                          6900
          *         *         *         *         *         *.        *         *         *
CTGCTGACCGGCCGCTGTACGCCGGGCGACGCCTGCGTGGCGGTGGAGAGCGGACGGCAGATCGACCGTTTCTTCCGCAATAACCCGCAC
LeuLeuThrGlyArgCysThrProGlyAspAlaCysValAlaValGluSerGlyArgGlnIleAspArgPhePheArgAsnAsnProHis>

                6950                                                  7000
          *         *         *         *         *         *         *         *         *
TTGGCCGTGCAGTACCTGGCCGATCCCTTCTGGGAGCGCCGCGCCATCGCCGTGCGCTATTCGCCGGTGGAGGCGCTGACCCCGTTGATT
LeuAlaValGlnTyrLeuAlaAspProPheTrpGluArgArgAlaIleAlaValArgTyrSerProValGluAlaLeuThrProLeuIle>

                              7050                                                  7100
          *         *         *         *         *         *         *         *         *
CGCGACAGCGACGAGGTGGTGCGTCGGGCCGTCGCCTACCGGCTGCCCCGCGAGCAGTTGAGTGCGCTGATGTTCGACGAGGACCGCGAG
ArgAspSerAspGluValValArgArgAlaValAlaTyrArgLeuProArgGluGlnLeuSerAlaLeuMetPheAspGluAspArgGlu>

                                        7150                                                7200
          *         *         *         *         *         *         *         *         *
GTGCGCATCACCGTGGCCGATCGCCTGCCGCTTGAGCAACTGGAGCAGATGGCCGCCGACAGGGATTATCTGGTGCGCGCCTATGTAGTC
ValArgIleThrValAlaAspArgLeuProLeuGluGlnLeuGluGlnMetAlaAlaAspArgAspTyrLeuValArgAlaTyrValVal>
```

```
                                                   7250
     *       *       *       *       *       *       *       *       *
CAGCGTATCCCTCCGGGGCGCCTGTTCCGTTTCATGCGCGACGAGGACCGCCAGGTGCGCAAGCTGGTGGCCAAGCGCCTGCCCGAGGAA
GlnArgIleProProGlyArgLeuPheArgPheMetArgAspGluAspArgGlnValArgLysLeuValAlaLysArgLeuProGluGlu>

   7300                                                    7350
     *       *       *       *       *       *       *       *       *
AGCCTGGGTCTCATGACCCAGGACCCGGAGCCCGAAGTCCGGCGCATCGTCGCCTCGCGCCTGCGCGGCGACGATTTGCTGGAGCTGTTG
SerLeuGlyLeuMetThrGlnAspProGluProGluValArgArgIleValAlaSerArgLeuArgGlyAspAspLeuLeuGluLeuLeu>

           7400                                                    7450
     *       *       *       *       *       *       *       *       *
CACGACCCGGATTGGACGGTACGCCTGGCCGCCGTGGAGCATGCGTCGCTGGAGGCGCTGCGTGAACTCGACGAGCCCGATCCCGAGGTC
HisAspProAspTrpThrValArgLeuAlaAlaValGluHisAlaSerLeuGluAlaLeuArgGluLeuAspGluProAspProGluVal>

                   7500                                                    7550
     *       *       *       *       *       *       *       *       *
CGCCTGGCGATCGCCGGGCGGCTCGGCATCGCCTGATCCCCGCGTCTGACGGGCGGAACGGTTTCGCCGGCCGCCCCGCCCCCCGCATTG
ArgLeuAlaIleAlaGlyArgLeuGlyIleAlaEnd>

                           7600                                                    7650
     *       *       *       *       *       *       *       *       *
CCCTTCGTACCCTTGTCACCTTTCGCCCAGCGTTGACGTGAATGCCCGATGCCGGACGCCTGCTTGTTGCAAACCTGACAGGAAGGCGCG

                                   7700
     *       *       *       *       *       *       *       *       *
TTTATGACAAAGGCTCCCCTTGGTAAACCCTTTTAAATCAGGCGCTTGCCTTCTGGTACAGGCATTGCAATGATCCGTTGCAACATGCTG

   7750         E ——>                                7800
     *       *       *       *       *       *       *       *       *
CATCAACGACCAGAGGGGTACGCGATGAAAGCCAAGGATATTGCCGAACTGCTCGACGAGCCCGCCTGCAGTCACAACAAGAAGGAAAAG
                        MetLysAlaLysAspIleAlaGluLeuLeuAspGluProAlaCysSerHisAsnLysLysGluLys>

           7850                                                    7900
     *       *       *       *       *       *       *       *       *
TCCGGCTGCGCCAAGCCCAAGCCGGGCGCCACCGACGGTCGGTGCTCCTTCGACGGCGCGCAGATCGCCCTGCTGCCCGTCGCCGACGTG
SerGlyCysAlaLysProLysProGlyAlaThrAspGlyArgCysSerPheAspGlyAlaGlnIleAlaLeuLeuProValAlaAspVal>

                   7950                                                    8000
     *       *       *       *       *       *       *       *       *
GCGCATATCGTTCACGGGCCGATCGCTTGCGCCGGCAGTTCCTGGGACAACCGCGGCACCCGCTCCAGCGGGCCGGACCTGTACCGCATC
AlaHisIleValHisGlyProIleAlaCysAlaGlySerSerTrpAspAsnArgGlyThrArgSerSerGlyProAspLeuTyrArgIle>

                           8050                                                    8100
     *       *       *       *       *       *       *       *       *
GGCATGACCACCGATCTCACCGAGAACGACGTGATCATGGGGCGCGCCGAGAAGCGCCTGTTCCATGCCATCCGCCAGGCGGTGGAAAGC
GlyMetThrThrAspLeuThrGluAsnAspValIleMetGlyArgAlaGluLysArgLeuPheHisAlaIleArgGlnAlaValGluSer>
```

```
                                        8150
        *         *         *         *         *         *         *         *         *
TATCTGCCGCCGGCGGTGTTCGTCTACAACACCTGCGTGCCGGCGCTGATCGGCGACGACGTCGACGCAGTGTGCAAAGCCGCCGCCGAG
TyrLeuProProAlaValPheValTyrAsnThrCysValProAlaLeuIleGlyAspAspValAspAlaValCysLysAlaAlaAlaGlu>

      8200                                        8250
        *         *         *         *         *         *         *         *         *
CGCTTCGGCACCCCGGTCATCCCGGTCGACTCGGCCGGCTTCTACGGCACCAAGAACCTCGGCAACCGCATCGCCGGTGAGGCCATGCTC
ArgPheGlyThrProValIleProValAspSerAlaGlyPheTyrGlyThrLysAsnLeuGlyAsnArgIleAlaGlyGluAlaMetLeu>

                8300                                        8350
        *         *         *         *         *         *         *         *         *
AAGTACGTGATCGGCACCCGCGAGCCCGATCCGCTGCCCGTCGGCAGCGAGCGTCCGGGCATCCGCGTGCACGACGTCAACCTGATCGGC
LysTyrValIleGlyThrArgGluProAspProLeuProValGlySerGluArgProGlyIleArgValHisAspValAsnLeuIleGly>

                          8400                                        8450
        *         *         *         *         *         *         *         *         *
GAGTACAACATCGCCGGCGAGTTCTGGCATGTCCTGCCGCTGCTCGACGAACTGGGCCTGCGGGTGCTCTGCACCCTGGCCGGCGATGCG
GluTyrAsnIleAlaGlyGluPheTrpHisValLeuProLeuLeuAspGluLeuGlyLeuArgValLeuCysThrLeuAlaGlyAspAla>

                                    8500                                        8550
        *         *         *         *         *         *         *         *         *
CGCTACCGCGAGGTGCAGACCATGCACCGCGCCGAAGTGAACATGATGGTCTGCTCCAAGGCCATGCTCAATGTCGCTCGCAAGCTGCAG
ArgTyrArgGluValGlnThrMetHisArgAlaGluValAsnMetMetValCysSerLysAlaMetLeuAsnValAlaArgLysLeuGln>

                                              8600
        *         *         *         *         *         *         *         *         *
GAAACCTACGGCACGCCCTGGTTCGAGGGCAGCTTCTACGGCATCACCGACACCTCCCAGGCGCTGCGCGACTTCGCCCGGCTGCTCGAT
GluThrTyrGlyThrProTrpPheGluGlySerPheTyrGlyIleThrAspThrSerGlnAlaLeuArgAspPheAlaArgLeuLeuAsp>

      8650                                        8700
        *         *         *         *         *         *         *         *         *
GATCCCGACCTGACCGCCCGCACCGAGGCGCTGATCGCGCGCGAGGAGGCCAAGGTCCGCGCCGCCCTCGAACCCTGGCGTGCGCGTCTG
AspProAspLeuThrAlaArgThrGluAlaLeuIleAlaArgGluGluAlaLysValArgAlaAlaLeuGluProTrpArgAlaArgLeu>

                8750                                        8800
        *         *         *         *         *         *         *         *         *
GAGGGCAAGCGCGTGCTGCTCTACACCGGCGGCGTGAAGTCCTGGTCGGTGGTTTCCCCCCTGCAGGACCTGGGCATGAAGGTGGTCGCC
GluGlyLysArgValLeuLeuTyrThrGlyGlyValLysSerTrpSerValValSerProLeuGlnAspLeuGlyMetLysValValAla>

                          8850                                        8900
        *         *         *         *         *         *         *         *         *
ACCGGCACCAAGAAGTCCACCGAGGAAGACAAGGCACGCATCCGCGAACTGATGGGCGACGACGTCAAGATGCTCGACGAGGGCAATGCG
ThrGlyThrLysLysSerThrGluGluAspLysAlaArgIleArgGluLeuMetGlyAspAspValLysMetLeuAspGluGlyAsnAla>

                                    8950                                        9000
        *         *         *         *         *         *         *         *         *
CGGGTGCTGCTGAAGACCGTCGACGAGTACCAGGCCGACATCCTCATCGCCGGCGGACGCAACATGTACACCGCGCTCAAGGGCCGCGTG
ArgValLeuLeuLysThrValAspGluTyrGlnAlaAspIleLeuIleAlaGlyGlyArgAsnMetTyrThrAlaLeuLysGlyArgVal>
```

```
                                            9050
     *         *         *         *         *         *         *         *         *
CCCTTCCTCGACATCAACCAGGAGCGCGAATTCGGCTATGCCGGCTACGACGGCATGCTGGAACTGGTGCGTCACGTCTGCATCACCCTG
ProPheLeuAspIleAsnGlnGluArgGluPheGlyTyrAlaGlyTyrAspGlyMetLeuGluLeuValArgHisValCysIleThrLeu>

    9100                                            9150
     *         *         *         *         *         *         *         *         *
GAATGCCCGGTGTGGGAGGCGGTGCGCCGCCCCGCGCCCTGGGACATCCCCGCCAGCCAGGACGCCCGCCCGAGCGGCGGCCCGTTCGGC
GluCysProValTrpGluAlaValArgArgProAlaProTrpAspIleProAlaSerGlnAspAlaArgProSerGlyGlyProPheGly>

              9200 N -->                                      9250
     *         *         *         *         *         *         *         *         *
GAACGCTGAGGAGAACGCATGGCCGAGATCATCAATCGCAACAAGGCCCTGGCCGTCAGCCCGCTGAAGGCCAGTCAGACCATGGGCGCC
GluArgEnd         MetAlaGluIleIleAsnArgAsnLysAlaLeuAlaValSerProLeuLysAlaSerGlnThrMetGlyAla>

                        9300                                                9350
     *         *         *         *         *         *         *         *         *
GCGCTGGCCATCCTCGGCCTGGCTCTCAGCATGCCGCTGTTTCACGGCTCGCAGGGTTGCACCGCGTTCGCCAAGGTGTTCTTCGTTCGC
AlaLeuAlaIleLeuGlyLeuAlaLeuSerMetProLeuPheHisGlySerGlnGlyCysThrAlaPheAlaLysValPhePheValArg>

                                  9400                                                9450
     *         *         *         *         *         *         *         *         *
CATTTCCGCGAGCCGGTCCCGCTGCAGACCACCGCCATGGACCAGGTCAGTTCGGTGATGGGCGCCGACGAGAACGTCGTCGAGGCGCTG
HisPheArgGluProValProLeuGlnThrThrAlaMetAspGlnValSerSerValMetGlyAlaAspGluAsnValValGluAlaLeu>

                                            9500
     *         *         *         *         *         *         *         *         *
AAGACCATCTGCGAACGGCAGAATCCCTCGGTCATCGGCCTGCTCACCACCGGCCTGTCGGAAACCCAGGGCTGCGACCTGCACACCGCC
LysThrIleCysGluArgGlnAsnProSerValIleGlyLeuLeuThrThrGlyLeuSerGluThrGlnGlyCysAspLeuHisThrAla>

    9550                                            9600
     *         *         *         *         *         *         *         *         *
CTGCACGAGTTCCGTACCCAGTACGAGGAGTACAAGGACGTGCCGATCGTTCCGGTGAACACGCCGGACTTCAGCGGCTGCTTCGAGAGC
LeuHisGluPheArgThrGlnTyrGluGluTyrLysAspValProIleValProValAsnThrProAspPheSerGlyCysPheGluSer>

              9650                                            9700
     *         *         *         *         *         *         *         *         *
GGCTTCGCCGCGGCGGTCAAGGCGATCGTCGAGACGCTGGTGCCCGAACGGCGCGATCAGGTGGGCAAGCGTCCGCGCCAGGTCAACGTG
GlyPheAlaAlaAlaValLysAlaIleValGluThrLeuValProGluArgArgAspGlnValGlyLysArgProArgGlnValAsnVal>

                        9750                                                9800
     *         *         *         *         *         *         *         *         *
CTGTGCTCGGCCAATCTCACCCCGGGCGATCTGGAGTACATCGCCGAAAGCATCGAGAGCTTCGGTCTGCGCCCCTTGCTGATCCCCGAC
LeuCysSerAlaAsnLeuThrProGlyAspLeuGluTyrIleAlaGluSerIleGluSerPheGlyLeuArgProLeuLeuIleProAsp>

                                  9850                                                9900
     *         *         *         *         *         *         *         *         *
CTGTCCGGCTCGCTCGACGGCCACCTGGACGAGAACCGTTTCAATGCCCTGACCACCGGCGGACTGAGCGTGGCCGAACTGGCCACCGCC
LeuSerGlySerLeuAspGlyHisLeuAspGluAsnArgPheAsnAlaLeuThrThrGlyGlyLeuSerValAlaGluLeuAlaThrAla>
```

9950
* * * * * * * * *
GGACAGAGCGTCGCCACTCTGGTGGTCGGGCAGAGCCTGGCGGGTGCGGCCGACGCCCTGGCCGAGCGCACCGGCGTGCCCGACCGGCGC
GlyGlnSerValAlaThrLeuValValGlyGlnSerLeuAlaGlyAlaAlaAspAlaLeuAlaGluArgThrGlyValProAspArgArg>

10000 10050
* * * * * * * * *
TTCGGCATGCTCTACGGTCTGGATGCGGTCGATGCCTGGCTGATGGCGCTGGCCGAGATCAGCGGCAATCCGGTGCCCGACCGCTACAAG
PheGlyMetLeuTyrGlyLeuAspAlaValAspAlaTrpLeuMetAlaLeuAlaGluIleSerGlyAsnProValProAspArgTyrLys>

10100 10150
* * * * * * * * *
CGCCAGCGTGCCCAATTGCAGGACGCCATGCTCGACACCCACTTCATGCTCAGTTCCGCACGCACGGCCATCGCCGCCGATCCCGATCTG
ArgGlnArgAlaGlnLeuGlnAspAlaMetLeuAspThrHisPheMetLeuSerSerAlaArgThrAlaIleAlaAlaAspProAspLeu>

10200 10250
* * * * * * * * *
CTGCTCGGTTTCGATGCCCTGCTGCGCAGCATGGGCGCGCACACGGTAGCCGCCGTGGTGCCGGCCCGGGCCGCCGCGCTGGTCGATTCG
LeuLeuGlyPheAspAlaLeuLeuArgSerMetGlyAlaHisThrValAlaAlaValValProAlaArgAlaAlaAlaLeuValAspSer>

10300 10350
* * * * * * * * *
CCTCTGCCCTCCGTGCGGGTCGGCGACCTGGAGGACCTCGAGCATGCCGCCCGCGCCGGCCAGGCCCAACTGGTGATCGGCAACAGCCAC
ProLeuProSerValArgValGlyAspLeuGluAspLeuGluHisAlaAlaArgAlaGlyGlnAlaGlnLeuValIleGlyAsnSerHis>

10400
* * * * * * * * *
GCCCTGGCCAGCGCCCGTCGCCTCGGTGTGCCACTGTTGCGTGCCGGCTTCCCGCAGTACGATCTGCTGGGCGGTTTCCAACGCTGCTGG
AlaLeuAlaSerAlaArgArgLeuGlyValProLeuLeuArgAlaGlyPheProGlnTyrAspLeuLeuGlyGlyPheGlnArgCysTrp>

10450 10500
* * * * * * * * *
TCCGGCTACCGCGGCAGCAGTCAGGTGCTGTTCGATCTGGCCAACCTGCTGGTCGAACACCACCAGGGTATCCAGCCCTATCATTCGATC
SerGlyTyrArgGlySerSerGlnValLeuPheAspLeuAlaAsnLeuLeuValGluHisHisGlnGlyIleGlnProTyrHisSerIle>

10550 X ⟶ 10600
* * * * * * * * *
TATGCGCAAAAACCGGCAACCGAACAGCCGCAATGGAGGCACTGAAGCGATGTCCAGCCCGACCCGACAATTGCAGGTATTGGATAGCGA
TyrAlaGlnLysProAlaThrGluGlnProGlnTrpArgHisEnd MetSerSerProThrArgGlnLeuGlnValLeuAspSerGlu>

10650 10700
* * * * * * * * *
GGACGACGGCACCCTGCTGAAGGTCGCCTTCGCCTCGTCCGACCGCGAACTGGTCGACCAGCATTTCGGCTCCTCGCGCTCCTTCGCCAT
AspAspGlyThrLeuLeuLysValAlaPheAlaSerSerAspArgGluLeuValAspGlnHisPheGlySerSerArgSerPheAlaIle>

10750 10800
* * * * * * * * *
CTACGGGGTCAATCCCGAGCGCTCCCAGCTCCTCTCGGTCGTCGAGTTCGGCGAGCTGGAGCAGGACGGCAACGAGGACAAGCTGGCCAG
TyrGlyValAsnProGluArgSerGlnLeuLeuSerValValGluPheGlyGluLeuGluGlnAspGlyAsnGluAspLysLeuAlaArg>

```
                                                  10850
         *         *         *         *         *         *         *         *         *
GAAGATCGATCTGCTCGACGGCTGCGTTGCCGTGTACTGTTGCGCCTGCGGCGCCTCGGCGGTGCGCCAGTTGATGGCCATCGGCGTGCA
 LysIleAspLeuLeuAspGlyCysValAlaValTyrCysCysAlaCysGlyAlaSerAlaValArgGlnLeuMetAlaIleGlyValGln>

       10900                                                10950
         *         *         *         *         *         *         *         *         *
GCCGATCAAGGTGAGCGAAGGCGCGCGCATCGCCGAGCTCATCGAAGCCCTCCAGGTCGAGTTGCGCGAAGGTCCTTCGGCCTGGCTGGC
 ProIleLysValSerGluGlyAlaArgIleAlaGluLeuIleGluAlaLeuGlnValGluLeuArgGluGlyProSerAlaTrpLeuAla>

                 11000                                                11050
         *         *         *         *         *         *         *         *         *
CAAGGCCATCCAGCGTACCCGCGGCCCCGACATGAGGCGCTTCGACGCCATGGCCGCCGAAGGCTGGGACGAATAGGTCCGACCACTTTT
 LysAlaIleGlnArgThrArgGlyProAspMetArgArgPheAspAlaMetAlaAlaGluGlyTrpAspGluEnd>

               3 ──►     11100                                                11150
         *         *         *         *         *         *         *         *         *
TCAGCGAGGAAGCGACATGTACTACGAAGAGCAACAAGAGCCCGTGGTGCAGGAAGACGACAAGTTCCTGCAGGATCCGATCATTCGCCA
                MetTyrTyrGluGluGlnGlnGluProValValGlnGluAspAspLysPheLeuGlnAspProIleIleArgGln>

                                   11200                                                11250
         *         *         *         *         *         *         *         *         *
GATGGTGGTGCAACTGCGGGCCGTGGACAGCTACGGCACCTACGACACCTGGAGCGATGCCCGCGTGGTCGATCCGCTAGTGCTGACCAA
 MetValValGlnLeuArgAlaValAspSerTyrGlyThrTyrAspThrTrpSerAspAlaArgValValAspProLeuValLeuThrLys>

                                                  11300
         *         *         *         *         *         *         *         *         *
GGAGCGCCGCAGGGCCATCCCGGTGGTCGGCGATCCGGACGAAACCACCATCTCCCGGATCAAGGCCTACTACAACACCCTCGCCCAGTT
 GluArgArgArgAlaIleProValValGlyAspProAspGluThrThrIleSerArgIleLysAlaTyrTyrAsnThrLeuAlaGlnLeu>

       11350                                                11400
         *         *         *         *         *         *         *         *         *
GCTCGAGCGGGAAACCGGCCTGCTCGCCGTACCGGTGATCAACATCACCCACGAAGGTTTCGGGCGCGCCCTGATCCTGGTCGGCAAGCT
 LeuGluArgGluThrGlyLeuLeuAlaValProValIleAsnIleThrHisGluGlyPheGlyArgAlaLeuIleLeuValGlyLysLeu>

                 11450                                                11500
         *         *         *         *         *         *         *         *         *
GGTCGCCCTGGACAAGACCCTGCGCGACGTTCACCGTTTCGGTTTCGAATCCCTCGAGGCGCTGGTCGCCGAAGCCAACAAGCAGCTGGG
 ValAlaLeuAspLysThrLeuArgAspValHisArgPheGlyPheGluSerLeuGluAlaLeuValAlaGluAlaAsnLysGlnLeuGly>

                           11550                                   4 ──► 11600
         *         *         *         *         *         *         *         *         *
CAAGGCCGCAACACTGGTCAACGAGCACCGTACGGTCGCCGAGCTCTGATCGATCGGGAGAGACACAGATGACCGAAGACGAAATCAAGG
 LysAlaAlaThrLeuValAsnGluHisArgThrValAlaGluLeuEnd                       MetThrGluAspGluIleLys>

                                   11650                                                11700
         *         *         *         *         *         *         *         *         *
CCCTGAAGAAAGAGGTAAGCCAGAAGAAGCGCATCGCGACCGAATGGGCTTCGCAGATCCACGATCTGGTCGAGGATCGCCTGTTCAACG
AlaLeuLysLysGluValSerGlnLysLysArgIleAlaThrGluTrpAlaSerGlnIleHisAspLeuValGluAspArgLeuPheAsn>
```

```
                                      11750
       *         *         *         *         *         *         *         *         *
 ATTACGAATCGCTGCCGGAGCTGGCCCGCCAGGCCCGTCAAGCCTGTGTCGAGTGGGCCGAAGCCAAGGCGCGTCTCGACGCGACCGGCG
AspTyrGluSerLeuProGluLeuAlaArgGlnAlaArgGlnAlaCysValGluTrpAlaGluAlaLysAlaArgLeuAspAlaThrGly>

     11800                                        11850
       *         *         *         *         *         *         *         *         *
 CCGCCTAGCCGGGCGGCGGCTTTCGCCGGGCACGGCACGCCCATGTGGCGTGCCCGGGCCCGGGCCCACGAAGCCCGCGGCTCCCCTGTA
AlaAlaEnd>

               11900                                        11950
       *         *         *         *         *         *         *         *         *
 CCGGGGCCTTGTCGGCCCGAATCCTTGCAGCCCGTCCCGATGGCGGGCAGCTGCCGCCCTGGCTCACGAGGCCGGGCGGTGCCCGACCCG

                         12000                                        12050
       *         *         *         *         *         *         *         *         *
 GTCGGGTAGCGCTGATACATCCATTGAGAGCGGGCTCGCGAACCGCCTTTGTCGATTTCGACAACCGTTTGGCCTCCGCACCGCAAGGGG

                                   12100                                        12150
       *         *         *         *         *         *         *         *         *
 CGGGCAGGAGAAAAAGGTAGATGGCAATGATTGTCGGGCGTACCCGTGGCGGCTCCGAGTGGATTCCGCAGTTCATCACCGCATTGAGTC

                                             12200
       *         *         *         *         *         *         *         *         *
 CGCAGGCCCGCGTCGGCTGCGGTCGCTGCTACAAGGTCTGTCCCAAACAGTGCCATAGCCACGAAGCCGCAGCAGCGGCAGCTTGAGGAG

     12250                                        12300
       *         *         *         *         *         *         *         *         *
 GAAGATTCATGGCCAAACCCGAGTTCCATATCTTTATTTGCGCTCAGAACCGTCCCGCCGGCCATCCGCGCGGCAGTTGCGGCGCAAAGG

               12350                                        12400
       *         *         *         *         *         *         *         *         *
 GTGCCGAAGGCGTCTATAACGCCTTCGCCCAGGTACTGATCCAGAAGAACCTGACCAATCGCATCGCCCTGACCACCACCGGCTGCCTCG

                         12450                                        12500
       *         *         *         *         *         *         *         *         *
 GGCCCTGCCAGGCCGGCGCCAACGTGCTGATCTATCCCGGCGCGGTCATGTATAGCTGGGTGGAGCCCGCCGACGCCGCGATCATCGTCG

                                   12550                                        12600
       *         *         *         *         *         *         *         *         *
 AACAGCACCTGCTCGGCGGCGAACCCTACGCGGACAAGCTCACCCCTGCCGAGATCTGGTAGCCGACATGGCCGATGTGCTGCTGTTCGC

                                             12650
       *         *         *         *         *         *         *         *         *
 CGGCGAACGGGCGTTTTTTTCCGACAGCACGCCCCAGGCGCTGCGCTATCTGCTCAACAACGCTCTCTGCGCCGGGACTCCCGATGCCGT
```

```
     12700                                                  12750
         *          *          *          *          *          *          *          *          *
CGAGGCGTTGCTGCTGGAGGCGCGCCGGCGCTGGCCGCAGGAGCCGGACGCGCACATCGGCCTGTACAAATTCTACTTCGTCAGCAAGCG

               12800                                                  12850
         *          *          *          *          *          *          *          *          *
CTACGAAGAGGCCGAGGCTGCGGCCTGGGCGGCGCTGCGCGCTGCCGCCCAGGCCGCCGGCTTCGACCGCAACTATCGCCGCCTGCAACC

                          12900                                                  12950
         *          *          *          *          *          *          *          *          *
GGGCAGCGCCGACTGGAGCCGACGCCAGGGGGCCGAGCGGTTGTATCTGTTCAGCCTCAAGGCGCTGGGCGTGATCCGCCTGCGTCGCGC

                                     13000                                                  13050
         *          *          *          *          *          *          *          *          *
CAAGGTGGCGATGGCTCGGAAAGTGCTGGAGAAACTGCTCGAACTCGATCCCGGCGATGAGATCGGCGGCGGTGCCTTCCTGCAGATCGC

                                                13100
         *          *          *          *          *          *          *          *          *
CCGTTCCTTCGAGGAGGATACCGCTTGACCAATGCGATGCTTGAGAGGCGTCTGGACGACGCCCGACCGCTGTTCTCGGAGATCTGGCAG

     13150                                                  13200
         *          *          *          *          *          *          *          *          *
GCGCTGGCCAGCAGCCTGAGCGCCGCCGGAGTCGAGCAGGGCCTGGCGGTGTCCGGTACGCCGCACACCCGCGGCGAGTTGCGCGAGGAC

               13250                                                  13300
         *          *          *          *          *          *          *          *          *
CCCTACGATCACAGCCAGACCCTGTATGCCGAGTGGCGCACGCCGGACGGCGGCTATCTGGGCAGCGTACTGGTGCATGGAGACGGCCAG

                          13350                                                  13400
         *          *          *          *          *          *          *          *          *
GCCTTCGCCGAGTTCGACGTGTTGCTGCCGCATCCGCGCAAGTCGGAATGGGTGATCGAGGCCGCCACGGCATGGGGTTACCGCGGTGCG

                                     13450                                                  13500
         *          *          *          *          *          *          *          *          *
CTGAAGAGCGAACTCCGCCTGCTGCCGGCGCTCGGCTCATGAACGGCCTGTACGACTGGCTGCTGGAGCATGCCGCCGGCGTCACCCGCG

                                                13550
         *          *          *          *          *          *          *          *          *
ACGACCCTGTCGAACGGGTCTGTCTCGGCCTCAGCTGGACCCTGGCGGAAGTCGGCGACAGCCTCGGTTTCGCTTTCGGCCCGCGCCAGG

     13600                                                  13650
         *          *          *          *          *          *          *          *          *
TGCCACGCACCCTGAGCTGGGCCGGCACCTTGGCCGGCCAGGGTGCCGAACGCCTGTACCCCTGGCTGCTGTCCTGGAACGGCGCCGAGG

               13700                                                  13750
         *          *          *          *          *          *          *          *          *
CGGCCGTGGGGCTGGCGGTGGTCAATGCCGTGGTCAATCGCGCCTCTCCCTGCCTGCGCCAGGCCCAGCCGTTGCACGGAGCCGTGCCCG
```

```
                                   13800                                                     13850
            *           *           *           *           *           *           *           *           *
GCAACCTGAGCGTCTTCGAGCATTTCCGGCCGCGCCTGGAGGGACAGAAGGTCGCGGTGATCGGCCATTATCCGGGGCTCGCGCAGCTCT

                                               13900                                                     13950
            *           *           *           *           *           *           *           *           *
GGAAGGACCTGCCGTACCGCTGCTTCGAGCGCGAGCCGCAGGAGGGCGATCTTCCCGATACGGCGGTCGAGCGGCTGTTGCCCGAGGCCG

                                                           14000
            *           *           *           *           *           *           *           *           *
ATTGGGTATTCATCACTGCCAGCAGCCTGGCCAACAAGACCCTGCCGCGCCTGCTGGAGCTGTCGTCGCACGCCCAGGTGGTGCTGATGG

       14050                                                      14100
            *           *           *           *           *           *           *           *           *
GGCCGAGTCTGCCATGGATCTACGACTGGCGACGGTTCGGGGTGAATTACCTGGCGGGCGTACGGGTGCGGGATGCCCATGCCGCACGCC

                       14150                                                      14200
            *           *           *           *           *           *           *           *           *
AGGTGGTCGCCGAAGGAGGCGGTACCCGGCTGTTCGCCGGGCCGGTGGAATATGCCCTGTTGGGTCTGGACTCATGAACGAACAGCAAAC

                                   14250                                                      14300
            *           *           *           *           *           *           *           *           *
CAGTGAAGAGTTGTCATGGCGGATTCTGCTCTGCCACAAGCATCCGGTGAGCGCGCGCCTGCGCTTTCTGATCCCGACCGGGGGCGGCGT

                                               14350                                                     14400
            *           *           *           *           *           *           *           *           *
GGTCTTGCCACAGACTTTGCCCAGGCTGGCGGTGATCGCCGAGGATCAGGAGGCGCCCGTGCAATGCCATCCGGCCAGCGCGCTGCGTGC

                                                           14450
            *           *           *           *           *           *           *           *           *
CCTGCAGGAAACGATGGCGCTCGGCTGGCAGTTGGAACTGATCGGCGAGTTCCGTCTGAACATGGAGGTACCGGGGCAGATCATGCCGAT

       14500                                                      14550
            *           *           *           *           *           *           *           *           *
CTATCTGGCGGCCCTGGCCGGGCATGAACTCCCGCCGCCGCCGGAAGGCACCCGCTGGATCGAGCTGACGCAGAGCATCGGCATGCCCTG

                       14600                                                      14650
            *           *           *           *           *           *           *           *           *
GCTCGACCGCGAACTGCTGCGGCGGGTCTACGAGGAACTGATCGGGTTCGGCTGCTGAGTGTGGAGCGGTTGCCGGGGCGGTGCGCCCCG

                                   14700                                                      14750
            *           *           *           *           *           *           *           *           *
GCGAGCGGGTTCAGGCGCGCAGCAGGGCCTTGGCCATCTTCGGCGACAACTGGTCCTCGGTGAACTGCGGCTCATTCGGCGGATACAGCA
          <EndAlaArgLeuLeuAlaLysAlaMetLysProSerLeuGlnAspGluThrPheGlnProGluAsnProProTyrLeuVal

                                               14800                                                     14850
            *           *           *           *           *           *           *           *           *
CATCTTCGATGATGCTGAAGCCGGACTCCTTGGCCAGGGCGAGGATGTCGCGGATCTTTTCGCAGGCTTTCAGTTTCTCGGCCAGAGCCG
<AspGluIleIleSerPheGlySerGluLysAlaLeuAlaLeuIleAspArgIleLysGluCysAlaLysLeuLysGluAlaLeuAlaPro
```

14900 ◄— 11

* * * * * * * * *

GATCCGCCTTGGCCTTCTCGGCGAAGGTCTTGATTTCCTTGATGGACATGCACTTCTCTCTTTGCTGGTTGCTCAAACCCCGGAACGTCG
<AspAlaLysAlaLysGluAlaPheThrLysIleGluLysIleSerMet

14950 15000

* * * * * * * * *

AATTCGCGGGTACCGGCCATATGCAATTGGGATGCCCTTGGAAAATATTTCCTTTAAATCAATTAGTTATGCTTGATTGTTTGTCGCATT

15050 15100

* * * * * * * * *

GACGACAATCGCCAGTATTTCTTGGCGCAATCGCGACAGTGGTTTACTTCACTCCATCAGCGCCACCGACTTGATCTGTGCATATACCTG
<EndGluMetLeuAlaValSerLysIleGlnAlaTyrValGln

15150 15200

* * * * * * * * *

CAGGCCCGGCGCCAGTCCCAGCCGGTCGGCCGACAGGCGGGTTATCCGTGCCAGGATGCAGGCGCTGTCCAGATCCAGGCGCACCATGAC
<LeuGlyProAlaLeuGlyLeuArgAspAlaSerLeuArgThrIleArgAlaLeuIleCysAlaSerAspLeuAspLeuArgValMetVal

15250 15300

* * * * * * * * *

CCGTGCCGAATCCTGTTCGTGGAACAGGTCGGCGATACGGACGCGAAGAATGTTGAGGATGCTGCTGTTGTGCGGCGGGTCGAGGGAAAG
<ArgAlaSerAspGlnGluHisPheLeuAspAlaIleArgValArgLeuIleAsnLeuIleSerSerAsnHisProProAspLeuSerLeu

15350

* * * * * * * * *

GCTCACGTCCCGGGCGAAGATCCGTACTCGCGTCTCCGCGCCCGGAGGCAGGCGGCTCAGGGAGACGCTGAGGGTTCCACCGGTACCCG
<SerValAspArgAlaPheIleArgValArgThrGluAlaGlyProProLeuArgSerLeuSerValSerLeuThrGlyGlyProValArg

15400 15450

* * * * * * * * *

GACGGTGCTGAGCCGGTAATGGGGGTCGTGGTGTTCCACCTGGCCGATGAGTACCACTGCGGCTTCGTCCGGATGGTTGAGCGGTAGACG
<ValThrSerLeuArgTyrHisProAspHisHisGluValGlnGlyIleLeuValValAlaAlaGluAspProHisAsnLeuProLeuArg

15500 15550

* * * * * * * * *

CGGGTCGCTCAGCAAGCGTCCCGGCGGGCCACTGGCGAACGTCTTGCCCTTGTCCAGCAGAATCAGATGGTCGGCCAGCCGGGTGACCTC
<ProAspSerLeuLeuArgGlyProProGlySerAlaPheThrLysGlyLysAspLeuLeuIleLeuHisAspAlaLeuArgThrValGlu

15600 15650

* * * * * * * * *

GTCCTGGCTGTGAGTGACGTAGAGCAGGGGAATCCGCAGTTGCCCGCGCAACTGTTCGAGGGCCGGCAGGATCTGCGCCCGGCTGTGACG
<AspGlnSerHisThrValTyrLeuLeuProIleArgLeuGlnGlyArgLeuGlnGluLeuAlaProLeuIleGlnAlaArgSerHisArg

15700 15750

* * * * * * * * *

GTCCAGGCAGATCAGCGGGGCATCCAGCAACAGCAGGTCGGGGCTGGTCAGTAATGCGCAGGCGATCGCGACACGCTGGCGCGGACCGTT
<AspLeuCysIleLeuProAlaAspLeuLeuLeuLeuAspProSerThrLeuLeuAlaCysAlaIleAlaValArgGlnArgProGlyAsn

15800

GGGGAGCCATTCGGCGCGTTGGTCGAGCAGGTCTTCCAATCCGAACAGGGCGATCACCTCATCGAGGCCGAGGCGCCTCCGGGCGCGGGG
<ProLeuTrpGluAlaArgGlnAspLeuLeuAspGluLeuGlyPheLeuAlaIleValGluAspLeuGlyLeuArgArgArgAlaArgPro

15850 15900

AATGCGCCGATAGCCGAATTCGAGGTTGGCCCGCACGTCCAGATGCGGGAGCAGTTCGCTGGCCTGGCGGACATAGCCCAGCGAGCGCTG
<IleArgArgTyrGlyPheGluLeuAsnAlaArgValAspLeuHisProLeuLeuGluSerAlaGlnArgValTyrGlyLeuSerArgGln

15950 16000

GTGGGGAGGACGGAAGCAGGTGGCGTCCTGCCAGAGCTGCCCGCCAACCTGAAGGAAACCCTCGGCCCGTTCCAGCCCGGCGATGGCGCG
<HisProProArgPheCysThrAlaAspGlnTrpLeuGlnGlyGlyValGlnLeuPheGlyGluAlaArgGluLeuGlyAlaIleAlaArg

16050 16100

CAACAGCGTGGTCTTGCCGCTGCCGGAGCGACCGAGCAGCACACTGATGCCGTTGGCCGGCAGTTGCAGGTGGGCATCGAGCCGGAAGGC
<LeuLeuThrThrLysGlySerGlySerArgGlyLeuLeuValSerIleGlyAsnAlaProLeuGlnLeuHisAlaAspLeuArgPheAla

16150 ◄— 10 16200

TCCCCGCTGCAGGCGCAACTGCGCCTCTATTCCCTGTTCTGGCATCTGAAATCTCCCGTGCCCAGCGGGCGGAAAAGATAAGGGAGGCAG
<GlyArgGlnLeuArgLeuGlnAlaGluIleGlyGlnGluProMet

16250

GGCGCCAGCGGTCGCCCGAGCCGTCGCGCCGGGCAGGGCCGCTCGTGCATGTCGCCTGCCTGCGCCTTGCGGCCCCGCTTGCGCCGGCCT

16300 16350

TCCCGGAGGGCTTGTCGGCTAGGGCACAAGGCTGATTTATGTGTGTAGCTTTGTGTATAACCGGACAAGCGGAAATGCAATGCAACCCTC

16400 16450

TGTTTTTTCTCAATTTAACGCATTGGCAAAGTCTTTGCACTCCAGATGCTCAGCGATTATTTGTATTCATTTCTATATAGCGCATGCCTG

16500 5 —► 16550

TTCCCGGCCCGCCCGGTGCCGGTCCATCATGCCGCTGCTGCAGGAGGTTCAGGATCATGCTGGAAGCAAGCTTTGCAACCGCACCGTTGC
MetLeuGluAlaSerPheAlaThrAlaProLeu>

16600 16650

CCGACAGCTTCGCATTCGCCCGCGATGGCTTCGTCCTGCTGCGTGGCTGTCTTCCCGAAAGCCTGGTCGACCGGCATGCCGCCGCGGTGG
ProAspSerPheAlaPheAlaArgAspGlyPheValLeuLeuArgGlyCysLeuProGluSerLeuValAspArgHisAlaAlaAlaVal>

```
                                                16700
         *         *         *         *         *         *         *         *         *
CGGCCTATCGTCTGCGCTGCGAGCCGGCCGCCAGGCAGCGTCAGGCCCATCCGCAGGTATTGCGCCAGCGCAACACCTTCCATCTGCAGG
AlaAlaTyrArgLeuArgCysGluProAlaAlaArgGlnArgGlnAlaHisProGlnValLeuArgGlnArgAsnThrPheHisLeuGln>

     16750                                        16800
         *         *         *         *         *         *         *         *         *
CGGCCGGCATCCTCGGTCTGGTACTGCACGAACGGCTGCTCGCACCGCTGAACATGCTGGCCGGCGAGAGAGCCTACCTGCTCGGCATCG
AlaAlaGlyIleLeuGlyLeuValLeuHisGluArgLeuLeuAlaProLeuAsnMetLeuAlaGlyGluArgAlaTyrLeuLeuGlyIle>

              16850                                       16900
         *         *         *         *         *         *         *         *         *
ATGCCCTGGAAAGTGCGCAGGTGTTGCTCGCTTGCAGCGCCTGCGCTGGCAGCGTGCCCCATCTGAGCGCGTTGATCGCCCTGCGGCCGA
AspAlaLeuGluSerAlaGlnValLeuLeuAlaCysSerAlaCysAlaGlyGlnValProHisLeuSerAlaLeuIleAlaLeuArgPro>

                        16950                                       17000
         *         *         *         *         *         *         *         *         *
TGCCCTTGGAAGCCGGACTCAAAGTCCAGCCGGGCAGCCACCTGGACAGTCGCCGCTCGCTGCGGCGGCTGCTGGCCGGCAACCCGGCGC
MetProLeuGluAlaGlyLeuLysValGlnProGlySerHisLeuAspSerArgArgSerLeuArgArgLeuLeuAlaGlyAsnProAla>

                                  17050                                       17100
         *         *         *         *         *         *         *         *         *
TGGCCGCCCAGTTGCAGCGGATGCACGAGGAGGGCGCGAGCCTGGAGCAGTGGCGCCAGCTCGAGATGCGTTTGCAACGGGCGCAGGCCC
LeuAlaAlaGlnLeuGlnArgMetHisGluGluGlyAlaSerLeuGluGlnTrpArgGlnLeuGluMetArgLeuGlnArgAlaGlnAla>

                                            17150
         *         *         *         *         *         *         *         *         *
GGGAACCGTCGGCCGGGCAGGCCATGCGCCTGCTGGCCCTGCGCAAGGGGGACGTGCTGATCCAGCAGCAGGGACTGCTCGGGGCGGCCG
ArgGluProSerAlaGlyGlnAlaMetArgLeuLeuAlaLeuArgLysGlyAspValLeuIleGlnGlnGlnGlyLeuLeuGlyAlaAla>

     17200                                        17250
         *         *         *         *         *         *         *         *         *
GGCGGGTGGAGCGCAACAGTTGCCTGATGGTTCGCTATGGCGCGGCCCAATTGCATCGTACCCCTTATTTCAGCGAGGCGGCCCTGGCCG
GlyArgValGluArgAsnSerCysLeuMetValArgTyrGlyAlaAlaGlnLeuHisArgThrProTyrPheSerGluAlaAlaLeuAla>

              17300                                       17350
         *         *         *         *         *         *         *         *         *
GCGATTGACCGTCCCGACCGTGCCGTTGCGTATTAGGCCGACAAGCCTTGTCGGAAATCAGACATTCATGGCCACAGGTATTGTTCGCTA
GlyAspEnd>

                        17400                                       17450
         *         *         *         *         *         *         *         *         *
CAGCCACGGGCTAGTTTGTCAAAAAGCAACCTTTTGATTTTAAAGATAAAATTTTAAACATATGGCGTGGCATGCCGGTTGCAGCAATAC

                                  17500       6 -->                                   17550
         *         *         *         *         *         *         *         *         *
CTTGCGACAGGTTGAGCAAGCAACCGCCAAGTTTTGTAGGAGACGAGCCGATGATTACCCTGACAGAGAGCGCCAAGAGCGCGGTTACCC
                                                  MetIleThrLeuThrGluSerAlaLysSerAlaValThr>
```

```
                                                   17600
         *         *         *         *         *         *         *         *         *
GCTTCATCAGCAGCACTGGCAAGCCGATCGCCGGTTTGCGCATCCGGGTCGAGGGTGGCGGTTGTTCCGGCCTGAAATACAGCCTGAAGC
ArgPheIleSerSerThrGlyLysProIleAlaGlyLeuArgIleArgValGluGlyGlyGlyCysSerGlyLeuLysTyrSerLeuLys>

      17650                                                17700
         *         *         *         *         *         *         *         *         *
TGGAAGAAGCCGGTGCGGAAGACGATCAACTGGTCGACTGCGACGGCATCACCCTGCTGATCGACAGCGCCAGCGCTCCTCTGCTCGATG
LeuGluGluAlaGlyAlaGluAspAspGlnLeuValAspCysAspGlyIleThrLeuLeuIleAspSerAlaSerAlaProLeuLeuAsp>

               17750                                                17800
         *         *         *         *         *         *         *         *         *
GCGTGACCATGGACTTCGTGGAAAGCATGGAAGGTAGCGGTTTCACCTTCGTCAATCCGAATGCCACCAACAGTTGTGGTTGTGGCAAGT
GlyValThrMetAspPheValGluSerMetGluGlySerGlyPheThrPheValAsnProAsnAlaThrAsnSerCysGlyCysGlyLys>

                         17850                                                17900U --->
         *         *         *         *         *         *         *         *         *
CTTTTGCTTGTTGAATAAGCAACCCTGAGGTCGCCGGCCGGGCCGGCCCCCAAAGACTCACCGGGAGATGAAGCCGACATGTGGGATTAT
SerPheAlaCysEnd                                                               MetTrpAspTyr>

                                   17950                                                18000
         *         *         *         *         *         *         *         *         *
TCGGAAAAAGTCAAAGAGCACTTCTACAACCCCAAGAATGCTGGAGCCGTGGAAGGCGCCAACGCCATCGGCGACGTCGGATCGCTGAGT
SerGluLysValLysGluHisPheTyrAsnProLysAsnAlaGlyAlaValGluGlyAlaAsnAlaIleGlyAspValGlySerLeuSer>

                                             18050
         *         *         *         *         *         *         *         *         *
TGCGGTGATGCGCTGCGCCTGACCCTGAAGGTGGACCCGGAAACCGACGTGATTCTGGATGCCGGCTTCCAGACCTTCGGCTGTGGTTCC
CysGlyAspAlaLeuArgLeuThrLeuLysValAspProGluThrAspValIleLeuAspAlaGlyPheGlnThrPheGlyCysGlySer>

      18100                                                18150
         *         *         *         *         *         *         *         *         *
GCCATCGCTTCCTCCTCGGCGCTGACCGAGATGGTCAAGGGCCTGACCCTGGACGAGGCGCTGAAGATCAGTAACCAGGACATCGCCGAC
AlaIleAlaSerSerSerAlaLeuThrGluMetValLysGlyLeuThrLeuAspGluAlaLeuLysIleSerAsnGlnAspIleAlaAsp>

               18200                                                18250
         *         *         *         *         *         *         *         *         *
TACCTCGATGGCCTGCCGCCGGAGAAGATGCACTGCTCGGTGATGGGCCGCGAAGCCCTGCAGGCCGCGGTGGCCAACTACCGTGGCGAG
TyrLeuAspGlyLeuProProGluLysMetHisCysSerValMetGlyArgGluAlaLeuGlnAlaAlaValAlaAsnTyrArgGlyGlu>

                         18300                                                18350
         *         *         *         *         *         *         *         *         *
ACGATCGAGGACGACCACGAAGAGGGCGCGCTGATCTGCAAGTGCTTCGCCGTCGACGAGGTGATGGTCCGCGATACCATCCGTGCCAAC
ThrIleGluAspAspHisGluGluGlyAlaLeuIleCysLysCysPheAlaValAspGluValMetValArgAspThrIleArgAlaAsn>

                                   18400                                                18450
         *         *         *         *         *         *         *         *         *
AAGCTGTCTACCGTTGAGGACGTGACCAATTACACCAAGGCCGGCGGTGGCTGCTCCGCCTGCCACGAGGCTATCGAGCGCGTGCTGACC
LysLeuSerThrValGluAspValThrAsnTyrThrLysAlaGlyGlyGlyCysSerAlaCysHisGluAlaIleGluArgValLeuThr>
```

```
                                         18500
          *         *         *         *         *         *         *         *         *
GAAGAGCTGGCCGCTCGCGGTGAAGTCTTCGTCGCGGCCCCGATAAAGGCCAAGAAGAAGGTCAAGGTGCTCGCCCCCGAGCCGGCTCCC
GluGluLeuAlaAlaArgGlyGluValPheValAlaAlaProIleLysAlaLysLysLysValLysValLeuAlaProGluProAlaPro>

      18550                                         18600
          *         *         *         *         *         *         *         *         *
GCCCCGGTGGCCGAAGCCCCGGCGGCTGCCCCGAAGCTGAGCAACCTGCAGCGCATCCGTCGCATCGAGACCGTGCTGGCGGCGATCCGT
AlaProValAlaGluAlaProAlaAlaAlaProLysLeuSerAsnLeuGlnArgIleArgArgIleGluThrValLeuAlaAlaIleArg>

               18650                                         18700
          *         *         *         *         *         *         *         *         *
CCGACCTTGCAGCGCGACAAGGGCGACGTCGAACTGATCGATGTCGACGGCAAGAACGTTTATGTCAAGCTCACCGGCGCCTGCACCGGC
ProThrLeuGlnArgAspLysGlyAspValGluLeuIleAspValAspGlyLysAsnValTyrValLysLeuThrGlyAlaCysThrGly>

                         18750                                         18800
          *         *         *         *         *         *         *         *         *
TGCCAGATGGCCAGCATGACCCTCGGCGGCATCCAGCAGCGCCTGATCGAGGAGCTCGGCGAGTTCGTCAAGGTGATTCCGGTCAGCGCT
CysGlnMetAlaSerMetThrLeuGlyGlyIleGlnGlnArgLeuIleGluGluLeuGlyGluPheValLysValIleProValSerAla>

                      S ——> 18850                                                     18900
          *         *         *         *         *         *         *         *         *
GCGGCTCACGCGCAGATGGAGGTCTGAAATGGCTGACGTCTATCTCGATAACAACGCCACCACCCGGGTGGACGACGAAATCGTCCAGGC
AlaAlaHisAlaGlnMetGluValEnd MetAlaAspValTyrLeuAspAsnAsnAlaThrThrArgValAspAspGluIleValGlnAla>

                                         18950
          *         *         *         *         *         *         *         *         *
CATGCTGCCGTTCTTCACCGAACAGTTCGGCAACCCCTCGTCGCTGCACAGCTTCGGCAACCAGGTCGGCATGGCGCTGAAGAAGGCCCG
 MetLeuProPhePheThrGluGlnPheGlyAsnProSerSerLeuHisSerPheGlyAsnGlnValGlyMetAlaLeuLysLysAlaArg>

      19000                                         19050
          *         *         *         *         *         *         *         *         *
CCAGAGCGTGCAGAAACTGCTCGGTGCCGAACACGATTCGGAAATCCTGTTCACCTCCTGCGGCACCGAGTCGGACTCCACCGCGATCCT
 GlnSerValGlnLysLeuLeuGlyAlaGluHisAspSerGluIleLeuPheThrSerCysGlyThrGluSerAspSerThrAlaIleLeu>

               19100                                         19150
          *         *         *         *         *         *         *         *         *
CTCGGCGCTCAAGGCCCAGCCCGAACGCAAGACCGTCATCACCACCGTGGTCGAACACCCGGCCGTCCTCAGCCTGTGCGATTACCTCGC
 SerAlaLeuLysAlaGlnProGluArgLysThrValIleThrThrValValGluHisProAlaValLeuSerLeuCysAspTyrLeuAla>

                         19200                                         19250
          *         *         *         *         *         *         *         *         *
CAGTGAGGGCTACACCGTGCACAAGCTGCCGGTGGACAAGAAAGGCCGTCTGGATCTGGAGCATTACGCCAGCCTGCTGACCGACGATGT
 SerGluGlyTyrThrValHisLysLeuProValAspLysLysGlyArgLeuAspLeuGluHisTyrAlaSerLeuLeuThrAspAspVal>

                                   19300                                               19350
          *         *         *         *         *         *         *         *         *
CGCCGTGGTGTCGGTGATGTGGGCCAACAACGAGACCGGTACCCTGTTCCCGATCGAAGAGATGGCGCGCCTGGCCGACGACGCCGGCAT
 AlaValValSerValMetTrpAlaAsnAsnGluThrGlyThrLeuPheProIleGluGluMetAlaArgLeuAlaAspAspAlaGlyIle>
```

19400
CATGTTTCACACCGATGCCGTGCAGGCCGTGGGCAAGGTACCGATCGACCTGAAGAATTCGTCGATCCACATGCTTTCGCTGTGCGGCCA
MetPheHisThrAspAlaValGlnAlaValGlyLysValProIleAspLeuLysAsnSerSerIleHisMetLeuSerLeuCysGlyHis>

19450 19500
CAAGCTGCACGCGCCCAAAGGTGTCGGCGTGCTCTATCTGCGCCGCGGCACCCGCTTCCGTCCGCTGCTGCGCGGTGGTCACCAGGAGCG
LysLeuHisAlaProLysGlyValGlyValLeuTyrLeuArgArgGlyThrArgPheArgProLeuLeuArgGlyGlyHisGlnGluArg>

19550 19600
CGGCCGTCGTGCCGGCACCGAGAACGCCGCCTCCATCATCGGCCTGGGTGTCGCCGCCGAACGCGCGCTGCAGTTCATGGAGCACGAGAA
GlyArgArgAlaGlyThrGluAsnAlaAlaSerIleIleGlyLeuGlyValAlaAlaGluArgAlaLeuGlnPheMetGluHisGluAsn>

19650 19700
CACCGAGGTCAACGCCCTGCGCGACAAGCTGGAGGCCGGTATCCTCGCCGTCGTGCCCCATGCCTTCGTCACCGGCGATCCGGACAACCG
ThrGluValAsnAlaLeuArgAspLysLeuGluAlaGlyIleLeuAlaValValProHisAlaPheValThrGlyAspProAspAsnArg>

19750 19800
CTTGCCCAACACCGCCAACATCGCGTTCGAGTACATCGAGGGCGAGGCCATCCTGCTGCTGCTGAACAAGGTCGGCATCGCCGCCTCCAG
LeuProAsnThrAlaAsnIleAlaPheGluTyrIleGluGlyGluAlaIleLeuLeuLeuLeuAsnLysValGlyIleAlaAlaSerSer>

19850
CGGTTCGGCCTGCACCTCCGGCTCCTTGGAGCCCTCCCACGTGATGCGCGCCATGGACATTCCCTATACTGCCGCCCACGGCACCGTGCG
GlySerAlaCysThrSerGlySerLeuGluProSerHisValMetArgAlaMetAspIleProTyrThrAlaAlaHisGlyThrValArg>

19900 19950
CTTCTCCCTGTCGCGCTACACCACCGAGGAGGAGATCGACCGGGTGATCCGCGAGGTTCCGCCGATCGTGGCCCAGTTGCGCAACGTGTC
PheSerLeuSerArgTyrThrThrGluGluGluIleAspArgValIleArgGluValProProIleValAlaGlnLeuArgAsnValSer>

20000 20050
GCCCTACTGGAGCGGCAACGGTCCGGTGGAGGACCCGGGCAAGGCCTTCGCTCCGGTCTACGGCTGAGTCGCCGCCGTCCTGCGGGAGCG
ProTyrTrpSerGlyAsnGlyProValGluAspProGlyLysAlaPheAlaProValTyrGlyEnd>

20100 V ⟶ 20150
TATCCCGCGGGAAATCGTCCCCACCCCGCCCGAGTTGTTGGAGAAAGCCATGGCTAGCGTGATCATCGACGACACTACCCTGCGTGACGG
MetAlaSerValIleIleAspAspThrThrLeuArgAspGly>

20200 20250
TGAACAGAGTGCCGGGGTCGCCTTCAATGCCGACGAGAAGATCGCTATCGCCCGCGCGCTCGCCGAACTGGGCGTGCCGGAGTTGGAGAT
GluGlnSerAlaGlyValAlaPheAsnAlaAspGluLysIleAlaIleAlaArgAlaLeuAlaGluLeuGlyValProGluLeuGluIle>

```
                                                      20300
         *         *         *         *         *         *         *         *         *
CGGCATTCCCAGCATGGGCGAGGAAGAGCGCGAGGTGATGCACGCCATCGCCGGTCTCGGCCTGTCGTCTCGCCTGCTGGCCTGGTGCCG
 GlyIleProSerMetGlyGluGluGluArgGluValMetHisAlaIleAlaGlyLeuGlyLeuSerSerArgLeuLeuAlaTrpCysArg>

     20350                                                 20400
         *         *         *         *         *         *         *         *         *
GCTATGCGACGTCGATCTCGCGGCGGCGCGCTCCACCGGGGTGACCATGGTCGACCTTTCGCTGCCGGTCTCCGACCTGATGCTGCACCA
 LeuCysAspValAspLeuAlaAlaAlaArgSerThrGlyValThrMetValAspLeuSerLeuProValSerAspLeuMetLeuHisHis>

              20450                                                 20500
         *         *         *         *         *         *         *         *         *
CAAGCTCAATCGCGATCGCGACTGGGCCTTGCGCGAAGTGGCCAGGCTGGTCGGCGAAGCGCGCATGGCCGGGCTCGAGGTGTGCCTGGG
 LysLeuAsnArgAspArgAspTrpAlaLeuArgGluValAlaArgLeuValGlyGluAlaArgMetAlaGlyLeuGluValCysLeuGly>

                        20550                                                 20600
         *         *         *         *         *         *         *         *         *
CTGCGAGGACGCCTCGCGGGCGGATCTGGAGTTCGTCGTGCAGGTGGGCGAAGTGGCGCAGGCCGCCGGCGCCCGTCGGCTGCGCTTCGC
 CysGluAspAlaSerArgAlaAspLeuGluPheValValGlnValGlyGluValAlaGlnAlaAlaGlyAlaArgArgLeuArgPheAla>

                                  20650                                                 20700
         *         *         *         *         *         *         *         *         *
CGACACCGTCGGGGTCATGGAGCCCTTCGGCATGCTCGACCGCTTCCGTTTCCTCAGCCGGCGCCTGGACATGGAGCTGGAAGTGCACGC
 AspThrValGlyValMetGluProPheGlyMetLeuAspArgPheArgPheLeuSerArgArgLeuAspMetGluLeuGluValHisAla>

                                                      20750
         *         *         *         *         *         *         *         *         *
CCACGATGATTTCGGGCTGGCCACGGCCAACACCCTGGCCGCGGTGATGGGCGGGGCGACTCATATCAACACCACGGTCAACGGGCTCGG
 HisAspAspPheGlyLeuAlaThrAlaAsnThrLeuAlaAlaValMetGlyGlyAlaThrHisIleAsnThrThrValAsnGlyLeuGly>

     20800                                                 20850
         *         *         *         *         *         *         *         *         *
CGAGCGTGCCGGCAACGCCGCGCTGGAAGAGTGCGTGCTGGCGCTCAAGAACCTCCACGGTATCGACACCGGTATCGATACCCGCGGCAT
 GluArgAlaGlyAsnAlaAlaLeuGluGluCysValLeuAlaLeuLysAsnLeuHisGlyIleAspThrGlyIleAspThrArgGlyIle>

              20900                                                 20950
         *         *         *         *         *         *         *         *         *
CCCGGCCATCTCCGCGCTGGTCGAGCGGGCCTCGGGGCGCCAGGTGGCCTGGCAGAAGAGCGTGGTCGGCGCCGGGGTGTTCACTCACGA
 ProAlaIleSerAlaLeuValGluArgAlaSerGlyArgGlnValAlaTrpGlnLysSerValValGlyAlaGlyValPheThrHisGlu>

                        21000                                                 21050
         *         *         *         *         *         *         *         *         *
GGCCGGTATCCACGTCGACGGACTGCTCAAGCATCGGCGCAACTACGAGGGGCTGAATCCCGACGAACTCGGTCGCAGCCACAGTCTGGT
 AlaGlyIleHisValAspGlyLeuLeuLysHisArgArgAsnTyrGluGlyLeuAsnProAspGluLeuGlyArgSerHisSerLeuVal>

                                  21100                                                 21150
         *         *         *         *         *         *         *         *         *
GCTGGGCAAGCATTCCGGGGCGCACATGGTGCGCAACACGTACCGCGATCTGGGTATCGAGCTGGCGGACTGGCAGAGCCAAGCGCTGCT
 LeuGlyLysHisSerGlyAlaHisMetValArgAsnThrTyrArgAspLeuGlyIleGluLeuAlaAspTrpGlnSerGlnAlaLeuLeu>
```

```
                                         21200
      *         *         *         *         *         *         *         *
CGGCCGCATCCGTGCCTTCTCCACCAGGACCAAGCGCCGCAGCCCGCAGCCTGCCGAGCTGCAGGATTTCTATCGGCAGTTGTGCGAGCA
 GlyArgIleArgAlaPheSerThrArgThrLysArgArgSerProGlnProAlaGluLeuGlnAspPheTyrArgGlnLeuCysGluGln>

   21250              7 -->               21300
      *         *         *         *         *         *         *         *
AGGCAATCCCGAACTGGCCGCAGGAGGAATGGCATGAGTCTGCTTACGCAATGGCGTGATGACATCCGCTGCGTCTTCGAGCGCGACCCG
 GlyAsnProGluLeuAlaAlaGlyGlyMetAlaEnd  LeuLeuThrGlnTrpArgAspAspIleArgCysValPheGluArgAspPro>
                                MetSer

          21350                                    21400
      *         *         *         *         *         *         *         *
GCGGCACGCACCACCTTCGAGGTGCTCACCACCTATCCGGGGGTGCACGCGATCATCGGCTACCGGCTGGCCAATCGCCTGTGGCGTGCC
AlaAlaArgThrThrPheGluValLeuThrThrTyrProGlyValHisAlaIleIleGlyTyrArgLeuAlaAsnArgLeuTrpArgAla>

                    21450                                    21500
      *         *         *         *         *         *         *         *
GCGTGGCGCTATCCGGCACGTCTGCTGTCGTTCGTCGCCCGCATGCTGAGCAACGTGGACATTCACCCGGGGGCGACCATCGGCGAGCGT
AlaTrpArgTyrProAlaArgLeuLeuSerPheValAlaArgMetLeuSerAsnValAspIleHisProGlyAlaThrIleGlyGluArg>

                              21550                                    21600
      *         *         *         *         *         *         *         *
TTCTTCATCGATCACGGCGCCTGCGTGGTGATCGGCGAAACCGCCGAGATCGGCAACGACGTGACCCTCTACCATGGCGTGACTCTCGGC
PhePheIleAspHisGlyAlaCysValValIleGlyGluThrAlaGluIleGlyAsnAspValThrLeuTyrHisGlyValThrLeuGly>

                                        21650
      *         *         *         *         *         *         *         *
GGCACCAGTTGGAACAAGGGCAAGCGCCATCCGACGCTGAGGGACGGCGTGCTGGTGGGCGCGGGAGCCAAGATTCTCGGTCCCATCACC
GlyThrSerTrpAsnLysGlyLysArgHisProThrLeuArgAspGlyValLeuValGlyAlaGlyAlaLysIleLeuGlyProIleThr>

   21700                                     21750
      *         *         *         *         *         *         *         *
GTGGGCGCGGGCGCCCGGGTGGGGGCCAACTCCGTGGTGGTCCAGGATGTGCCGGACGGTTGCACCGTGGTCGGCATCCCCGGCAAGGTA
ValGlyAlaGlyAlaArgValGlyAlaAsnSerValValValGlnAspValProAspGlyCysThrValValGlyIleProGlyLysVal>

          21800                                    21850
      *         *         *         *         *         *         *         *
GTCAAGGTGCGCGAGGCCGGCCGCCCGAATCCCTATGGCATCGACCTCGACCATCACCTGATTCCGGACCCGGTGGGCAAGGCCATCGCC
ValLysValArgGluAlaGlyArgProAsnProTyrGlyIleAspLeuAspHisHisLeuIleProAspProValGlyLysAlaIleAla>

                    21900                                    21950
      *         *         *         *         *         *         *         *
TGCCTGCTGGAGCGCATCGATGCGCTGGAGAAGCAGGTCGAGGCGGGTGGCTGGTCACCGTCGGCAGCCAGCAGCAGTTCTACCAGGTC
CysLeuLeuGluArgIleAspAlaLeuGluLysGlnValGluAlaGlyGlyLeuValThrValGlySerGlnGlnGlnPheTyrGlnVal>

                              22000                                    22050
      *         *         *         *         *         *         *         *
TGTAACTCCGACAACAGCATTTGCGAGAGCGACTGCGCCGGTGGCACCACTGCGCAGGCTCAACAATCGGCCGGTCGCCGTCGGGCCGTA
CysAsnSerAspAsnSerIleCysGluSerAspCysAlaGlyGlyThrThrAlaGlnAlaGlnGlnSerAlaGlyArgArgArgAlaVal>
```

```
                                                22100 8 ──►
         *         *         *         *         *         *         *         *         *
CCCACTCCGGTGGCCGAGTGATCGCTGGAACCACGGGAGACGACTGAGCATGGACCTGCAGGACTTCGACAGCGACAGCCTGTACTACGA
ProThrProValAlaGluEnd                          MetAspLeuGlnAspPheAspSerAspSerLeuTyrTyrAsp>

      22150                                               22200
         *         *         *         *         *         *         *         *         *
CGAGCCGCTCGCCCCGGAGGTAGCGGCCCGTCTGGCCGCGGCCGCCGAACAGTATGCCGAGGGCACCGCCGAGCAGCCGCTGCTCGAGGC
 GluProLeuAlaProGluValAlaAlaArgLeuAlaAlaAlaAlaGluGlnTyrAlaGluGlyThrAlaGluGlnProLeuLeuGluAla>

                22250                                               22300
         *         *         *         *         *         *         *         *         *
ACAGGCCCTCGCCCCGGACGATCTGACGGTGCTGGTCGGCCTGTATCGCTTCTATTTCTACCAGCACCGCTATCACGACGCTCTGCAGAT
 GlnAlaLeuAlaProAspAspLeuThrValLeuValGlyLeuTyrArgPheTyrPheTyrGlnHisArgTyrHisAspAlaLeuGlnIle>

                          22350                                               22400
         *         *         *         *         *         *         *         *         *
CGCCTGGCGGGTACTGGAGGTGGTCGGTCCCCGGCTGCTGCTGCCCGCGGACTGGCGCGCGATCGATGCCGGGTATCTCGCGGCGGCATC
 AlaTrpArgValLeuGluValValGlyProArgLeuLeuLeuProAlaAspTrpArgAlaIleAspAlaGlyTyrLeuAlaAlaAlaSer>

                                    22450                                               22500
         *         *         *         *         *         *         *         *         *
CGAGCGGGGCATCGCGCTCCTGCGTTTCTATCTGCTGGCGCTCAAGGGTGCCGGCTATCTGAACCTGCGTCTGGGCCGGTTCGAGCAGGG
 GluArgGlyIleAlaLeuLeuArgPheTyrLeuLeuAlaLeuLysGlyAlaGlyTyrLeuAsnLeuArgLeuGlyArgPheGluGlnGly>

                                              22550
         *         *         *         *         *         *         *         *         *
CAAGGCCATGCTCGCCAAGGTCGTCGAGCTGGACACGGAAAACCGCCTGGGAGCCAGGCTGTTGTTGGACGTGCTGGCCGAACACAGCGC
 LysAlaMetLeuAlaLysValValGluLeuAspThrGluAsnArgLeuGlyAlaArgLeuLeuLeuAspValLeuAlaGluHisSerAla>

      22600                               W ──►   22650
         *         *         *         *         *         *         *         *         *
TGAAATCCTGATTTTCCCCACTGCGGCCAACGCGGAGATAAGACCATGACCGTGCAACCCTTTTCCCCGGATTCCGACCTGACCCTGGAC
 GluIleLeuIlePheProThrAlaAlaAsnAlaGluIleArgProEnd  ValGlnProPheSerProAspSerAspLeuThrLeuAsp>
                                             MetThr

                22700                                               22750
         *         *         *         *         *         *         *         *         *
GAGGCCATGGACGAGCTGGTCTCGGCCGAGGACTTCCTCGAGTTCTTCGGTGTGCCCTTCGATCAGGATGTCGTGCATGTCAACCGGCTG
GluAlaMetAspGluLeuValSerAlaGluAspPheLeuGluPhePheGlyValProPheAspGlnAspValValHisValAsnArgLeu>

                          22800                                               22850
         *         *         *         *         *         *         *         *         *
CACATCATGCAACGCTACCACGACTACCTGAGCAAGGCCGGCGATCTGGACGAGCACGACGACCAGGCCCGCTACGCGGTGTTCCAGAAG
HisIleMetGlnArgTyrHisAspTyrLeuSerLysAlaGlyAspLeuAspGluHisAspAspGlnAlaArgTyrAlaValPheGlnLys>

                                    22900                                               22950
         *         *         *         *         *         *         *         *         *
CTGCTGGCGCGCGCCTATCTGGACTTCGTCGAGTCCGATGCCCTGACCGAGAAGGTCTTCAAGGTGTTTCGCATGCACGAACCGCAGAAG
LeuLeuAlaArgAlaTyrLeuAspPheValGluSerAspAlaLeuThrGluLysValPheLysValPheArgMetHisGluProGlnLys>
```

```
                                                 23000  Z ──►
        *         *         *         *         *         *         *         *         *
ACCTTCGTTTCCATCGACCAACTGTTGAGCTGAGCTTAAGAGGGGAGGCGCCATGCTGCCGCAGTTTGAATATGGTGATGAGGTTCGCCT
ThrPheValSerIleAspGlnLeuLeuSerEnd                     MetLeuProGlnPheGluTyrGlyAspGluValArgLeu>

     23050                                      23100
        *         *         *         *         *         *         *         *         *
GATTCGCAACGTGCGTAACGACGGCACCTATCCGGGCATGAATACCGGCGCCCTGCTGATGCGCCGTGGCGCGGTGGTTGCGTCTATGA
 IleArgAsnValArgAsnAspGlyThrTyrProGlyMetAsnThrGlyAlaLeuLeuMetArgArgGlyAlaValGlyCysValTyrAsp>

               23150                                      23200
        *         *         *         *         *         *         *         *         *
CGTCGGTACTTATCTGCAGGATCAATTGATCTATCGCGTGCACTTCCTCGACGAAGGACGCACCATCGGCTGCCGCGAGGAGGAGCTGAT
 ValGlyThrTyrLeuGlnAspGlnLeuIleTyrArgValHisPheLeuAspGluGlyArgThrIleGlyCysArgGluGluGluLeuIle>

                         23250                                      23300
        *         *         *         *         *         *         *         *         *
CCTCGCCTCGGCTCCCTGGATTCCCAACCTTTTCGAGTTCCGCGACGATGTGATCGCCACCCGCAGTCTTGCCGTGCGCGGCCAGGTGCT
 LeuAlaSerAlaProTrpIleProAsnLeuPheGluPheArgAspAspValIleAlaThrArgSerLeuAlaValArgGlyGlnValLeu>

                                   23350                                      23400
        *         *         *         *         *         *         *         *         *
GGTGAAGCGTGGCCAGCTCGGCAGCATCATGAAGGTGCTGCGCGACGAGCCCGAGCTCGGCATCCAGTACCATGTCCACTTCGGCGACGG
 ValLysArgGlyGlnLeuGlySerIleMetLysValLeuArgAspGluProGluLeuGlyIleGlnTyrHisValHisPheGlyAspGly>

                                             23450                M ──►
        *         *         *         *         *         *         *         *         *
TCTGGTATTGCAGGTGCCCGAGCAGAGCCTGGCGATGGCCGACTCCACGGCGGCCATCGAGGAGGTGCTCGATGGCATCTGAGCGTCTCG
 LeuValLeuGlnValProGluGlnSerLeuAlaMetAlaAspSerThrAlaAlaIleGluGluValLeuAspGlyIleEnd ArgLeu>
                                                                          MetAlaSerGlu

     23500                                      23550
        *         *         *         *         *         *         *         *         *
CCGACGGCGACAGCCGCTATTACTTGCTGAAGGTCGCCCACGAGCAGTTCGGCTGCGCGCCCGGCGAGCTCAGCGAGGATCAGCTCCAGC
AlaAspGlyAspSerArgTyrTyrLeuLeuLysValAlaHisGluGlnPheGlyCysAlaProGlyGluLeuSerGluAspGlnLeuGln>

               23600                                      23650
        *         *         *         *         *         *         *         *         *
AGGCCGACCGCATCATCGGCCGGCAGAGGCATATCGAGGACGCCGTGTTGCGCAGCCCCGATGCGATCGGTGTGGTGATCCCGCCCTCCC
GlnAlaAspArgIleIleGlyArgGlnArgHisIleGluAspAlaValLeuArgSerProAspAlaIleGlyValValIleProProSer>

                         23700                                      23750
        *         *         *         *         *         *         *         *         *
AGCTCGAAGAGGCCTGGGCACACATCGCCAGCCGCTACGAGAGTCCCGAGGCGCTACAGCAGGCCCTCGACGCGCAGGCGCTGGATGCTG
GlnLeuGluGluAlaTrpAlaHisIleAlaSerArgTyrGluSerProGluAlaLeuGlnGlnAlaLeuAspAlaGlnAlaLeuAspAla>

                                   23800                                      23850
        *         *         *         *         *         *         *         *         *
CCGGCATGCGCGCCATGCTGGCCCGGGAGCTCAGGGTCGAGGCCGTTCTCGACTGCGTCTGCGCCGGCCTGCCGGAGATCAGCGACACCG
AlaGlyMetArgAlaMetLeuAlaArgGluLeuArgValGluAlaValLeuAspCysValCysAlaGlyLeuProGluIleSerAspThr>
```

```
                                                                 23900
         *            *            *            *            *            *            *            *            *
ATGTGAGCCTCTACTACTTCAACCACGCCGAGCAATTCAAGGTGCCCGCCCAGCACAAGGCCCACATCCTGGTCACCATCAACGAGGATT
AspValSerLeuTyrTyrPheAsnHisAlaGluGlnPheLysValProAlaGlnHisLysAlaHisIleLeuValThrIleAsnGluAsp>

  23950                                                          24000
         *            *            *            *            *            *            *            *            *
TTCCGGAAAACACCCGCGAAGCCGCCCGGACGCGCATCGAGACCATCCTCAAGCGGCTGCGCGGCAAGCCGGAGCGCTTCGCCGAGCAGG
PheProGluAsnThrArgGluAlaAlaArgThrArgIleGluThrIleLeuLysArgLeuArgGlyLysProGluArgPheAlaGluGln>

                24050                                                          24100
         *            *            *            *            *            *            *            *            *
CGATGAAGCACTCCGAATGTCCCACCGCGATGCAGGGCGGATTGCTCGGCGAGGTGGTCCCGGGAACGCTTTATCCCGAACTGGACGCCT
AlaMetLysHisSerGluCysProThrAlaMetGlnGlyGlyLeuLeuGlyGluValValProGlyThrLeuTyrProGluLeuAspAla>

                              24150                                                          24200
         *            *            *            *            *            *            *            *            *
GCCTGTTCCAGATGGCGCGGGGAGAACTGAGTCCGGTACTGGAATCGCCGATCGGTTTTCACGTGCTGTACTGCGAAAGCGTGAGCCCCG
CysLeuPheGlnMetAlaArgGlyGluLeuSerProValLeuGluSerProIleGlyPheHisValLeuTyrCysGluSerValSerPro>

                                            24250                                                          24300
         *            *            *            *            *            *            *            *            *
CCCGGCAGCTCACCCTCGAGGAGATCCTGCCGCGTCTGCGCGACAGGCTGCAGCTCCGGCAGCGCAAGGCGTATCAGCGCAAATGGCTGG
AlaArgGlnLeuThrLeuGluGluIleLeuProArgLeuArgAspArgLeuGlnLeuArgGlnArgLysAlaTyrGlnArgLysTrpLeu>

                                          9 ——► 24350
         *            *            *            *            *            *            *            *            *
TTTGCCTGCTCCAACAAAACGCTACTTTGGAGAACCTCGCCCATGGATAAAACGGACAAACCCTGCTGCTCGTTCTGCGGCGCGGAGAAG
ValCysLeuLeuGlnGlnAsnAlaThrLeuGluAsnLeuAlaHisGlyEnd ThrAspLysProCysCysSerPheCysGlyAlaGluLys>
                                                  MetAspLys

  24400                                                          24450
         *            *            *            *            *            *            *            *            *
AGCCCGACCGTTCCGCTGATCGCCGGCAACGACGGGCGAATCTGCGAAGCCTGCGTCAAGCTTGCCCATCAGGTGGTCAGCAGCTGGGG
SerProThrValProLeuIleAlaGlyAsnAspGlyArgIleCysGluAlaCysValLysLeuAlaHisGlnValValSerSerTrpGly>

                24500                                                          24550
         *            *            *            *            *            *            *            *            *
CAGCGGCGCAAGAACCAGCAGCTGGCTCCGCAGCTACGCACCCCCGTGGAATACAAGAAGCACCTGGACGAGTCGGTGATCGGTCAGGAG
GlnArgArgLysAsnGlnGlnLeuAlaProGlnLeuArgThrProValGluTyrLysLysHisLeuAspGluSerValIleGlyGlnGlu>

                              24600                                                          24650
         *            *            *            *            *            *            *            *            *
GCAGCCAAGGAGACCCTGTCGGTCGCGGTCTACAACCACTACCTGCGCTTGCTCAACTGCGATCGCGAACCGGTCTGCCAACTGGGCGAG
AlaAlaLysGluThrLeuSerValAlaValTyrAsnHisTyrLeuArgLeuLeuAsnCysAspArgGluProValCysGlnLeuGlyGlu>

                                            24700                                                          24750
         *            *            *            *            *            *            *            *            *
CAGGTCGAACTGGAGAAATCCAACATCCTCATGGCCGGCCCTTCGGGTACCGGCAAGACCCTGCTGGTGCGTACCCTGGCGCGCATTCTC
GlnValGluLeuGluLysSerAsnIleLeuMetAlaGlyProSerGlyThrGlyLysThrLeuLeuValArgThrLeuAlaArgIleLeu>
```

```
                                                  24800
          *         *         *         *         *         *         *         *         *
GGGGTTCCCTTCGCCATGGCCGATGCCACCACCCTGACCCAGGCCGGCTATGTCGGCGATGACGTCGACAGCATCATCACCCGTCTGCTC
GlyValProPheAlaMetAlaAspAlaThrThrLeuThrGlnAlaGlyTyrValGlyAspAspValAspSerIleIleThrArgLeuLeu>

     24850                                        24900
          *         *         *         *         *         *         *         *         *
GATGCGGCCGGCGGCGACGTGCAGCAGGCGCAGTGGGGCATCGTCTATATCGACGAGGTGGACAAGCTGGCCAAGCGCAGCGGGGGCGGT
AspAlaAlaGlyGlyAspValGlnGlnAlaGlnTrpGlyIleValTyrIleAspGluValAspLysLeuAlaLysArgSerGlyGlyGly>

              24950                                       25000
          *         *         *         *         *         *         *         *         *
ACGGCGGTGCGCGACATCTCCGGCGAAGGGGTACAGCAGGCCCTGCTGAAGATGGTCGAAGGCACCGAGGTGCGCATCAGCAAGTCCGGG
ThrAlaValArgAspIleSerGlyGluGlyValGlnGlnAlaLeuLeuLysMetValGluGlyThrGluValArgIleSerLysSerGly>

                        25050                                       25100
          *         *         *         *         *         *         *         *         *
CGACGCAACGAGCACAGCGAGGAGCAGGTGGTGGACACCCGCAATATCCTGTTCATCGCCGGCGGTGCCTTCCCCGGGTTGGAAGCACTG
ArgArgAsnGluHisSerGluGluGlnValValAspThrArgAsnIleLeuPheIleAlaGlyGlyAlaPheProGlyLeuGluAlaLeu>

                                  25150                                       25200
          *         *         *         *         *         *         *         *         *
GTCTCGAGCCGCATCCAGCCGAAGAACACCGGTATCGGTTTCCATGCCCAGCCGCGTCGGGAGGCCCCTTCGATCAACGAGCTGATGGCC
ValSerSerArgIleGlnProLysAsnThrGlyIleGlyPheHisAlaGlnProArgArgGluAlaProSerIleAsnGluLeuMetAla>

                                            25250
          *         *         *         *         *         *         *         *         *
TCGCTGCTACCGGACGACCTGCACGAGTTCGGCCTGATTCCCGAATTCATCGGTCGTTTCCCGATCATCACCTTCCTGCAGGAACTCGAC
SerLeuLeuProAspAspLeuHisGluPheGlyLeuIleProGluPheIleGlyArgPheProIleIleThrPheLeuGlnGluLeuAsp>

     25300                                        25350
          *         *         *         *         *         *         *         *         *
CACGCCACCCTGCTGCGTATCCTCACCGAGCCGCGCAATGCACTGGTCAAACAGTACAAGCAGTTGTTCGCCTATCAGGGCGTCGAACTG
HisAlaThrLeuLeuArgIleLeuThrGluProArgAsnAlaLeuValLysGlnTyrLysGlnLeuPheAlaTyrGlnGlyValGluLeu>

              25400                                       25450
          *         *         *         *         *         *         *         *         *
GTGATCACCGATGCGGCCCTCAACTACATCGCCGACCAGGCGCTGATTCGCAAGACCGGTGCCCGCGGATTGCGCGCTGTACTGGAAGCC
ValIleThrAspAlaAlaLeuAsnTyrIleAlaAspGlnAlaLeuIleArgLysThrGlyAlaArgGlyLeuArgAlaValLeuGluAla>

                        25500                                       25550
          *         *         *         *         *         *         *         *         *
GCCCTGCAGCAGACCATGTTCAACATGCCCTCGCAGCCACAACTGCGCGGTTGTACGCTGGATCTGGTCGAGCATGAGGATGGCGGCCGC
AlaLeuGlnGlnThrMetPheAsnMetProSerGlnProGlnLeuArgGlyCysThrLeuAspLeuValGluHisGluAspGlyGlyArg>

                                  25600                                       25650
          *         *         *         *         *         *         *         *         *
TCTCTGGAGGTTCTGACCCGACTGGCCGAAGATGGCAGCGGACGGATTCAGCCCGATCCATCTCCGGTAGTGGAGGAAAAATCCGCGCTG
SerLeuGluValLeuThrArgLeuAlaGluAspGlySerGlyArgIleGlnProAspProSerProValValGluGluLysSerAlaLeu>
```

```
                                                   25700
          *          *          *          *          *          *          *          *          *
TCGGCCGACCTCTGAGGAGGTCGTAAGAACTGTTGTAAAAACGACAAGGCCTGCTGCCCGCCCCATTCTTATCGTCGCATTTGCGACAGG
SerAlaAspLeuEnd>

   25750                                                   25800
          *          *          *          *          *          *          *          *          *
ATTTTTTACTGCCATAAAAAATGCTTTTAAATCAATAAGTTGGTAAATGTGTTCGGGGTTTGGGGAGTGGTCTGCTTCTTGCTGTTACTC

               25850                                                   25900       F -->
          *          *          *          *          *          *          *          *          *
ACCGGCCAGCCAGGTGTACAAGCCCGGAGACTGGCGCTCCGACCACGAATCTCATGCATGCAGCCAGAGGTTAAGTTATGGCCAAGATTG
                                                                                 MetAlaLysIle>

                          25950                                                   26000
          *          *          *          *          *          *          *          *          *
GACTCTTCTTCGGTAGCAACACCGGTAAAACCCGCAAGGTCGCCAAGTCGATCAAGAAGCGTTTCGACGACGAAACCATGTCCGATGCGC
GlyLeuPhePheGlySerAsnThrGlyLysThrArgLysValAlaLysSerIleLysLysArgPheAspAspGluThrMetSerAspAla>

                                     26050                                                   26100
          *          *          *          *          *          *          *          *          *
TGAACGTCAACCGTGTTTCCGCGGAAGACTTCGCCCAGTACCAGTTCCTGATTCTGGGTACTCCGACCCTTGGCGAAGGCGAACTCCCCG
LeuAsnValAsnArgValSerAlaGluAspPheAlaGlnTyrGlnPheLeuIleLeuGlyThrProThrLeuGlyGluGlyGluLeuPro>

                                                26150
          *          *          *          *          *          *          *          *          *
GTCTGTCCTCCGACTGCGAAAACGAAAGCTGGGAAGAATTCCTGCCGAAGATCGAGGGCCTGGATTTCAGCGGCAAGACCGTGGCGCTGT
GlyLeuSerSerAspCysGluAsnGluSerTrpGluGluPheLeuProLysIleGluGlyLeuAspPheSerGlyLysThrValAlaLeu>

   26200                                                   26250
          *          *          *          *          *          *          *          *          *
TCGGCCTGGGCGATCAGGTTGGCTATCCCGAGAATTACCTGGATGCCCTGGGCGAACTGTATTCCTTCTTCAAGGACCGTGGCGCCAAGA
PheGlyLeuGlyAspGlnValGlyTyrProGluAsnTyrLeuAspAlaLeuGlyGluLeuTyrSerPhePheLysAspArgGlyAlaLys>

               26300                                                   26350
          *          *          *          *          *          *          *          *          *
TCGTAGGCTCCTGGTCGACCGACGGCTACGAGTTCGAAAGCTCCGAGGCCGTGGTTGACGGCAAGTTCGTCGGCCTGGCGCTGGATCTGG
IleValGlySerTrpSerThrAspGlyTyrGluPheGluSerSerGluAlaValValAspGlyLysPheValGlyLeuAlaLeuAspLeu>

                          26400                                                   26450
          *          *          *          *          *          *          *          *          *
ACAATCAGAGCGGCAAGACCGACGAGCGCGTTGCTGCCTGGCTGGCACAGATCGCTCCCGAGTTCGGGCTGTCCCTGTAAAAGTCGATCC
AspAsnGlnSerGlyLysThrAspGluArgValAlaAlaTrpLeuAlaGlnIleAlaProGluPheGlyLeuSerLeuEnd>

                                     26500                                                   26550
          *          *          *          *          *          *          *          *          *
CCGGTCATGCGGTCTCCGCTGCATGGCCGGGTGACTTGGCATGTTTTCTTTTCGTTTCTCTGCGGATCGCGCCGCTGGGCGCACACTCTC

                                                26600
          *          *          *          *          *          *          *          *          *
TCGCAAGGCACGACCGACTCCGATGGCCGGATCTCTTGCGCTTTGTTTCACCGTCATATGGTGTCGCTGTGCGGACAGCCCGCACGACT
```

```
  26650                                         26700
    *         *         *         *         *         *         *         *         *
CGTCTTCTGTCCGACAAATCGCTCTGTCGGCGATGCCTTGCCCTGAAGGGCGAGGTCGATGGGAAGAAATGCGCCAGGCCTGCCCGGATG

              26750                                         26800
    *         *         *         *         *         *         *         *         *
GGAGTACGGACAGACCCGGCGACTTGAACGGGTCACTTAGTGATAGCTGATGCCGGAGTCGGCCTGCGCGGTGGCTTCGCGGTAGGCGCG

                        26850                                         26900
    *         *         *         *         *         *         *         *         *
GTTGCCACGGATCTCGACCGCCTCGCCAATGACCAGTTTGTCATTGGGTGCGCGTAATACCTTTTCCTCGCCGCCGGCCGTGCCGGCCTT

                                  26950                                         27000
    *         *         *         *         *         *         *         *         *
GTAGGCTCGCTCGCTGAGGCCGGAGGCATCTTGGGCCTCGCCACCGAACAGGGCGCCGACACCTGCACCGACCAACATGCCGAGCGGGCC

                                            27050
    *         *         *         *         *         *         *         *         *
ACCGATGGCGCCGATCATCATGCCGCTCATGCCACCGACCGACTTGCCGGAAGTCTGGTCCTTGGTTTCCTCGAGGACTTCGGCGGCATG

  27100                                         27150
    *         *         *         *         *         *         *         *         *
CACACTACCTGTGGCTATCAAGACGAGAGTCAGAGCCAGGGTGCCGAGCTTTTTCATGGCGAATCCTCCGAGGAAGTCTTCAATAGTCTT

              27200                                         27250
    *         *         *         *         *         *         *         *         *
TGCTCTTTATCCATGAAGCGTGCCAAGAATAATTATTTATTAAAATCAATAACTTACTTTCTTTGGCAGGCGATGTGAGCGACAAAGTGT

                        27300                                         27350
    *         *         *         *         *         *         *         *         *
GGTGATGCCGACAAAATATTGTCTGGTTTCGCACAGAAATCTGCGATTTTGTTGATTAGGGTCAAGAAATCCTGCATGGTTTTATGGCTG

                                  27400                                         27450
    *         *         *         *         *         *         *         *         *
AGGTGCTTGTGTAGGGCCGGTCGCGGCGGGATAGGGTGTGCGACACTCGGGACTCCCGATGTCGTGGCCGGGCGCTGGAAATTCGGCGGA

                                            27500
    *         *         *         *         *         *         *         *         *
GTGCCTTTCAGCATGGAGATGCATGCACGAAGACCATTGTTGGCGACGAGCTTCGGGCTGTTCGCTGCCGATGCGGATCGGGACGCCTGG

  27550                                         27600
    *         *         *         *         *         *         *         *         *
GCTTGCTGCGCTCGGGGGAAGGTCCGGCGGCCGTCGCTGAGATTCTCGCAGGAGGGTTCGAGATAACGGGTATTGTTGGCAGCGCCATGG

              27650                                         27700
    *         *         *         *         *         *         *         *         *
TGATGGATACCCTGCCGGTAAGGGATGCCCTCGATTCGGGTCTGGTGGCAGATGTCGAAGCTGCCTTCGGGACAGTGCTCGGTGAAGTTC
```

```
                 27750                                        27800
    *      *      *      *      *      *      *      *      *
ATTTCCACCTGCTGGCCGGGCAAGGCGAGCGACAGGAATGGCGGGCGCATGGAGAGTTCTGTCGAGAGGACGCCCGACGTCTAACGTGAC

                        27850                                      27900
    *      *      *      *      *      *      *      *      *
ATGCCAGCTCGTCTCTTCCTCCGAGGCCTTCTGCCGCGACGGGCTTCCCGCCCCGCGCAAAAAGCCTCACCATGCCGCCAATGTTGAAGA

                               27950
    *      *      *      *      *      *      *      *      *
ACATTCCCACCCACCTGATCGCCGGCCCCCTCGGGGCGGGCAAGACCAGCCTGATCCGCCAGTTGCTGGCACAGAAGCCGGCCACCGAGC

  28000                                  28050
    *      *      *      *      *      *      *      *      *
GCTGGGCGGTGCTGATCAACGAATTCGGCCGGATCGGCCTGGATGCCGCCCTGCTGGATAGCGGCACAGAGGGCGTCGCTTTCGCCGAAG

         28100                                  28150
    *      *      *      *      *      *      *      *      *
TCGCCGGTGGCTGCCTCTGTTGCGTCAACGGTGTCCCCTTTCAGGTCGCTCTGACGCGGCTGCTGCGCAAGGTCCGTCCCTGCCGCCTGC

                28200                                   28250
    *      *      *      *      *      *      *      *      *
TGATCGAGCCTTCCGGACTCGGCCATCCGGTCGGGCTGTGGATCGGCTCATGGCGCCTCCCTGGTTCGAAGTGCTGGCACTGCGGCCCG

                        28300                                      28350
    *      *      *      *      *      *      *      *      *
CCGTACTGGTTTTGGATGCTGCCGACCTGGTTGCCGGTGCGGCACTGCCGTCGACTCAGGCGGAGATGTTGGCCATGGCCGGGCTGCTGG

                               28400
    *      *      *      *      *      *      *      *      *
TCCTGAACAAGGCGGAGTCGCTGGATGGGGCAGCACGTGCGCAATTGGCGGAAAGGCTGCCGCGACGTCCGTTGTACTGGACCAGCCAGG

  28450                                  28500
    *      *      *      *      *      *      *      *      *
GAGCGTTGCCTCTGGCGCTGCTGCCGGAGCTGGCCGTCCGCACGGACGAGGGCGGCGGGCTTCCGGGCGCTGAATTGCCGGCTGTTGCGG

         28550                                  28600
    *      *      *      *      *      *      *      *      *
CGGCGCTGGAAGAGGTCTGGGCGGATCCGCGGCTTCCGATCTGCGTCAGCCAGGCTCGGGCCGATGGCTGGAGCATCGGCTGGCGTTGGC

                28650                                   28700
    *      *      *      *      *      *      *      *      *
ATGCCAGTCAGCGTTTCGCGCTGACGCGCATCCGGCATTGGCTGTCCGGCCTGCATTGGCGGCGCGCCAAGCTGGTTCTGCATGCCGGGG

                        28750
    *      *      *      *      *      *      *      *
ACGGCTGGCTGTCCGGCAACGCGCTAGATCGCCAGCCGATCCGCTGGCAGCCCAGCGAATGGCGCAGGGATTCGCGGCTCGAG
```

The vita has been removed from the scanned document