NOTICE

The quality of this microform is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us an inferior photocopy.

Previously copyrighted materials (journal articles, published tests, etc.) are not filmed.

Reproduction in full or in part of this microform is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30.

AVIS

La qualité de cette microforme dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de qualité inférieure.

Les documents qui font déjà l'objet d'un droit d'auteur (articles de revue, tests publiés, etc.) ne sont pas microfilmés.

La reproduction, même partielle, de cette microforme est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30.

Canada

THE UNIVERSITY OF ALBERTA

The Cloning and Sequencing of DNA From the Nitrogen Fixation Region of *Azotobacter vinelandii*.

BY

KOJI HIRATSUKA

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES AND RESEARCH

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE

OF Master of Science

DEPARTMENT OF MICROBIOLOGY

EDMONTON, ALBERTA
FALL 1987

ISBN 0-315-40945-2

THE UNIVERSITY OF ALBERTA

RELEASE FORM

NAME OF AUTHOR: Koji Hiratsuka

TITLE OF THESIS: The Cloning and Sequencing of DNA from the Nitrogen Fixation Region of *Azotobacter vinelandii*.

DEGREE: Master of Science

YEAR THIS DEGREE GRANTED: 1987

Permission is hereby granted to THE UNIVERSTIY OF ALBERTA LIBRARY to reproduce single copies of this thesis and to lend or sell such copies for private, scholarly or scienctific research purposes only.

The author reserves other publication rights, and neither the thesis nor extensive extracts from it may be printed or otherwise reproduced without the author's written permission.

Koji Hir

(Student's signature)

#20 11115 27ave

Edmonton, Alta. T6J 5N3

(Student' permanent address)

Date: October 9, 1987

THE UNIVERSITY OF ALBERTA

FACULTY OF GRADUATE STUDIES AND RESEARCH

The undersigned certify that they have read, and recommend to the Faculty of Graduate studies and Research for acceptance, a thesis entitled, The Cloning and Sequencing of DNA from the Nitrogen Fixation Region of *Azotobacter vinelandii*.

submitted by Koji Hiratsuka

in partial fulfilment of the requirements for the degree of Master of Science.

Keith L. Roy
(Supervisor)

John B. Bell

Susan Jensen

Ivy Page

Date: 6 Oct. 1987

Dedication

This thesis is dedicated to my wife, Ursula, my mother and father, whose patience, love and support gave me the incentive I needed to complete this thesis.

Abstract

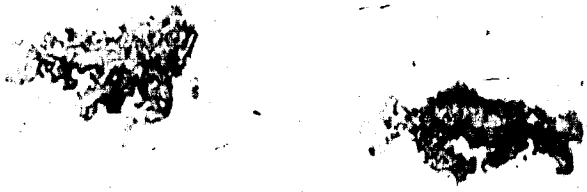
A cosmid library of clones containing genomic DNA from the free-living, nitrogen fixer *A.vinelandii* was constructed, using the cosmid vector pSa747. One clone, pAvH1, contained *A.vinelandii* DNA homologous to *K. pneumoniae nif* DNA. This was shown by hybridization of the clone with a ^{32}P-labelled fragment from the plasmid pSA30, containing the *K. pneumoniae nifH, nifD, nifK, nifY* genes as well as part of the *nifE* gene. The region on pAvH1 which showed homology to the *K. pneumoniae nif* genes was localized to a 2.4 kb *Hin* dIII- *Bgl* II fragment.

The 1.4 kb *Eco* RI fragment within the *nif* homologous region was subcloned into pUC8. The entire sequence of the 1.4 kb *Eco* RI fragment was determined. A large open reading frame (ORF) coding for a polypeptide of 409 amino acid residues was found by computer analysis of the DNA sequence. No promotor sites or termination codons were observed, however a strong ribosome binding site was found preceding the ORF, 9 bp before the first codon. Preceding the large ORF, another ORF coding for a polypeptide of 20 amino acid residues was found. This ORF was found to be in the same reading frame as the large ORF. Two termination codons occurred at the end of the smaller reading frame. The two ORFs were separated by 123 bp.

Comparison to the results of Brigle *et al.* (1985. Gene 37:37-44) showed the *Eco* RI fragment of pAvH1 to contain DNA sequences coding for the last 20 amino acids of the *nifH* gene and the first 409 amino acids of the *nifD* gene of *A. vinelandii*, as well as the intergenic region between these two genes. A substantial amount of the work reported here was conducted prior to the publication of the results of Brigle *et al.* (1985).

Comparisons of the predicted amino acid sequences of the *A. vinelandii nifD* proteins to the *nifD* protein of other nitrogen fixing organisms showed a high extent of conservation

occurring in certain regions of the protein, particularly regions thought to be involved in the binding of the 4Fe-4S centers of the nitrogenase complex.



Acknowledgements

I would like to thank Dr. Ken Roy who, as my supervisor, gave me much guidance, support and wisdom. I would also like to express my apprecation to the staff and students of the department of Microbiology. Your help and freindship has been extremly valuable to me. Speciál thanks go to the people who put up with working with me in the same lab. Namely Deborah McIntyre, Jim MacPherson, Dr. Jamie Doran, and Wan Lam. Thanks for all the technical and not so technical advice, and for elbow room in the lab.

Thanks to those who assisted me in putting this thesis together. To Rob Carmicheal and Richard Sherburne for their expertise and assistance in the photography and to William Yan for his assistance in mounting the pictures. And to the Department of Medical Microbiology and Infectious Diseases for kindly allowing me to use their coumputer and darkroom facilities.

I would also like to thank my parents for all their encouragement and advice. As always mom and dad knew what was best.

Last but not least, I would like to thank my wife, Ursula, for her patience and constant love and support. As always I coundn't have done it without you.

TABLE OF CONTENTS

	Page
I. INTRODUCTION	1
Genetic organization of *K. pneumoniae nif* genes	1
Regulation of nitrogen fixation in *K. pneumoniae*	5
Nitrogen fixation in *A. vinelandii*	9
Interspecies homology of nitrogen fixation systems	12
Comparisons of proteins and DNA sequences of the *nif* structural genes	16
Thesis objectives	17
II. MATERIAL AND METHODS	18
Bacterial strains and media	18
Cloning vectors and plasmids	18
Isolation of *A.vinelandii* DNA	21
Preparation of vector DNA	22
Preparation of *A. vinelandii* DNA	22
Ligation of pSa747 with *A. vinelandii* DNA fragments	22
Packaging and transduction	24
Screening of colonies	24
Isolation and purification of pAvH1	24
Restriction mapping of pAvH1	26
Subcloning of pAvH1 DNA	26
Sequencing of *A. vinelandii* DNA	28
Computer analysis of sequence	30
II. RESULTS	31
Preparation of *A. vinelandii* DNA for ligation	31

Preparation of pSa747 31
Isolation and identification of clones 31
Mapping of pAvH1 38
Subcloning of pAvH1 DNA 38
Sequencing of the 1.4 kb Eco RI fragment of pAvH1 54
III. DISCUSSION 84
Conclusion 95
IV. REFERENCES 97

LIST OF TABLES

	Page
Table 1. Nitrogen fixation genes of *K. pneumoniae*	3
Table 2. Media	19
Table 3. Cloning vectors and plasmids	20
Table 4. Ligations of *A. vinelandii* DNA to pSa747	25
Table 5. Relative homologies of amino acid sequences from *nifH* proteins	92
Table 6. Relative homologies of amino acid sequences from *nifD* proteins	93

LIST OF FIGURES

Page

Figure 1. Genetic organization of *nif* genes in *K. pneumoniae*..... 2

Figure 2. Genetic organization of *nif* genes in *A. vinelandii*..... 10

Figure 3. Flow diagram of cloning strategy..... 23

Figure 4. Small scale partial digests of *A. vinelandii* DNA..... 33

Figure 5. Sucrose gradient of partially digested *A. vinelandii* DNA..... 35

Figure 6. Complete digests of pSa747..... 37

Figure 7. Small scale plasmid isolation of possible cosmid clones..... 40

FIgure 8. Colony hybridization of possible clones..... 42

Figure 9. Small scale plasmid isolations of possible positive clones from colony hybridization and hybridization with the *Eco* RI fragment of pSA30..... 44

Figure 10. Example of agarose gel used for restriciton mapping of pAvH1..... 46

Figure 11. Restriction map of pAvH1..... 48

Figure 12. Small scale plasmid isolation of posslible *Eco* RI sub-clones from pAvH1and hybridization with the *Eco* RI fragment of pSA30..... 51

Figure 13. Comparison of *Eco* RI digests of pAvH1 and clone 138..... 53

Figure 14. Dot hybridizations of M13 isolates..... 56

Figure 15. Direct gel electrophoresis of M13 clones..... 58

Figure 16. Sequencing strategy..... 60

Figure 17. Complete nucleotide sequence and sequence of the 1.4 Kb *Eco* RI fragement of pAvH1 62

Figure 18. Predicted amino acid sequence from the 1.4 kb *Eco* RI fragment of pAvH1..... 65

Figure 19. Comparison of DNA sequence to that of Brigle *et al.*. 68

Figure 20. Sequencing gel showing differences in DNA sequence with sequence of Brigle *et al.* at position 9..... 71

Figure 21. Sequencing gel showing differences in DNA sequence with

sequence of Brigle *et al.* at positions 88-90..... 73

Figure 22. Sequencing gel showing differences in DNA sequence with sequence of Brigle *et al.* at position 196..... 75

Figure 23. Sequencing gel showing differences in DNA sequence with sequence of Brigle *et al.* at positions 824-825..... 77

Figure 24. Sequencing gel showing differences in DNA sequence with sequence of Brigle *et al.* at positions 962 and 970..... 79

Figure 25. Sequencing gel showing differences in DNA sequence with sequence of Brigle *et al.* at position 1356..... 81

Figure 26. Sequencing gel showing differences in DNA sequence with sequence of Brigle *et al.* at positions 1408-1410..... 83

Figure 27. Comparison of the carboxy-terminal ends of *nifH* proteins..... 87

Figure 28. Comparison of amino acid sequences from *nifD* proteins..... 89

INTRODUCTION

Nitrogen fixation is a complex process in which atmospheric nitrogen is reduced to ammonia. This is economically significant in agriculture, as crop yields are dramatically affected by the availability of fixed nitrogen in the soil. Nitrogen fixation is exclusively a prokaryotic function and can be carried out by a diverse range of bacteria. Nitrogen fixation can occur either symbiotically or non-symbiotically.

Nitrogenase is the enzyme which carries out the reduction of atmospheric nitrogen to ammonia. This process is energy-expensive, thus it is not surprising that the system is repressible in the presence of a readily utilizable nitrogen source. The nitrogen-fixing system is also usually repressed in high oxygen concentrations due to the sensitivity of the enzyme to oxygen. Along with the nitrogenase enzyme, there are usually a number of other nitrogen fixation specific proteins involved in the overall process.

Advances in the field of molecular biology have led to a greater understanding of the genetics of the nitrogen fixation systems of a wide variety of organisms. Much of what is known today about the genetics of nitrogen fixation has been elucidated from studies of the free-living nitrogen fixer *Klebsiella pneumoniae*. Although nitrogen fixation by this species is probably not economically significant, it is the most easily studied nitrogen fixing organism due to its close relatedness to *Escherichia coli* which facilitates genetic manipulations.

Genetic organization of *K. pneumoniae nif* genes.

In *K. pneumoniae* 17 *nif* (nitrogen fixation) genes are required for the production of an active nitrogenase. The genes are organized into 8 transcriptional units (operons) occupying approximately 24kb of the *K. pneumoniae* genome (Figure 1). The functions of these genes are summarized in Table 1. The nitrogenase enzyme itself is made up of 2

Figure 1. Genetic organization of *nif* genes of *K. pneumoniae*.

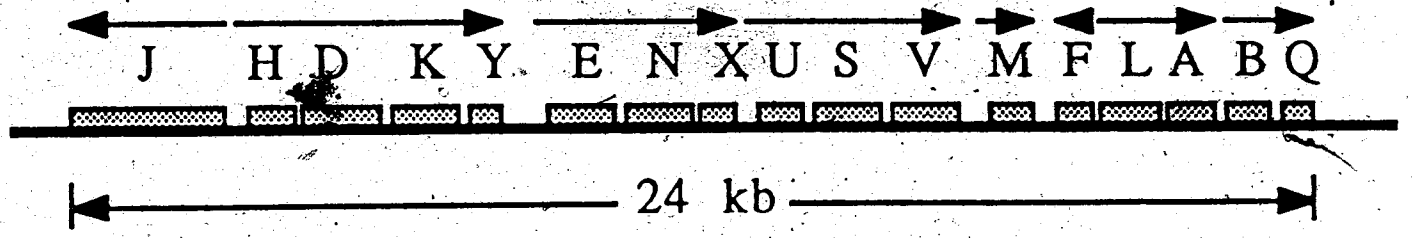


Table 1. Nitrogen fixation genes of *K. pneumoniae*.

Gene	Function	Reference
J	FeS-containing electron transport protein	*a,b*
H	Fe protein component	*c,i*
D	alpha subunit of FeMo protein component	*c*
K	beta subunit of FeMo protein component	*c*
Y	function not known	*d*
E	processing of the FeMo cofactor	*c*
N	processing of the FeMo cofactor	*c*
X	function not known	*d*
U	function not known	*d*
S	possibly involved in processing Fe protein	*c*
V	final modification of the FeMo cofactor	*e*
M	activation of Fe protein	*c*
F	flavoprotein involved in electron transport	*b,f*
L	inhibits expression of *nif* operons	*g,h*
A	activates expression of *nif* operons	*c,i,j,k*
B	processing of FeMo cofactor	*l*
Q	uptake of molybdenum	*m*

a- Shah *et al.*, 1983, *b*-Hill & Kavanagh, 1980 *c*- Roberts *et al.*, 1978, *d*- Puhler & Klipp, 1981, *e*- Hoover *et al.*, 1986, *f*- Nieva-Gomez *et al.*, 1980, *g*- Merrick *et al.*, 1982, *h* -Hill *et al.*, 1981, *i*- Dixon *et al.*, 1977, *j*- Dixon *et al.*, 1980, *k*- Buchanan-Wollaston *et al.*, 1981, *l*- Shah & Brill, 1977, *m*- Imperial *et al.*, 1984.

components, the FeMo protein and Fe protein. The FeMo protein is made up of 2 subunits, the and subunits, coded for by the *nifD* and *nifK* genes respectively. Two of each subunit form the complete FeMo protein component. Associated with the FeMo protein is the FeMo-cofactor, which is essential for the activity of the nitrogenase complex. The FeMo protein component is thought to contain both the active site and substrate binding site of the nitrogenase enzyme.

The second component of the nitrogenase enzyme is the Fe protein, or the nitrogenase reductase. This protein is a dimer which has its identical monomeric subunits coded for by the *nifH* gene. During nitrogen fixation this protein receives electrons from the nitrogenase specific electron transfer chain. The electrons are then transferred, coupled with ATP hydrolysis, to the FeMo (FeMo protein) protein. Here the dinitrogen bound by the protein is reduced to NH_3. During transcription the *nifH* gene is transcribed together with the *nifD* and the *nifK* on the same messenger RNA. The *nifY* gene is also transcribed on the same messenger, but the function of the product of this gene is not known.

Some of the *nif* genes are involved in the processing of nitrogenase to produce the active enzyme. The *nifB*, *nifE*, *nifN*, and *nifV* gene products have all been shown to be necessary for the synthesis of the FeMo-cofactor, which confers activity on the FeMo protein. The *nifV* gene is of particular interest as mutations in this gene cause an alteration of substrate specificity in the nitrogenase, suggesting that it is somehow involved with the substrate binding site of the nitrogenase complex. In fact it has been shown that the *nifV* gene product is involved in the final processing of the FeMo-cofactor (Hoover *et al.*, 1986), which could confer substrate specificity to the complex. The other three genes direct the synthesis of the FeMo-cofactor and its insertion into the FeMo protein component of the nitrogenase enzyme. Another gene which appears to be indirectly important for the production of FeMo-cofactor is the *nifQ* gene, which is apparently involved in molybdenum uptake.

The products of two genes, the *nifM* gene and the *nifS* gene, appear to be involved in the maturation of the Fe protein subunit, as mutations in these genes result in the production of inactive Fe protein subunits. The exact nature of this processing function has yet to be determined.

As mentioned earlier, the nitrogenase system of *K. pneumoniae* also codes for its own electron transfer proteins. Two such proteins have been identified, coded for by the *nifF* and *nifJ* genes. The *nifF* protein has been shown to be a flavoprotein, while the *nifJ* protein has been shown to be an iron- and sulfur- containing protein. The electron transport chain of the *K. pneumoniae nif* system starts with pyruvate as its physiological electron donor. The *nifJ* protein (a dimer of 2 identical protein subunits) acts as the first electron acceptor in the chain and passes electrons on to the *nifF* protein. The *nifF* protein in turn passes electrons on to the Fe protein subunit of the nitrogenase enzyme.

The regulation of the nitrogen fixing system in *K. pneumoniae* is fairly complex and will be discussed in greater detail below. Two genes in the *nif* cluster are directly involved with the regulation of the system. They are the *nifA* and the *nifL* genes. Products of these genes have been shown to either activate or repress the *nif* operons.

For two of the genes in the *nif* cluster, proteins have been observed in *in vitro* transcription-translation experiments, but the functions of these proteins have yet to be determined. These are the *nifX* and the *nifY* genes. Another gene which has yet to have a function assigned to its gene product is the *nifU* gene. This gene was mapped in polar mutation studies of the *nifUSVM* operon, but to date no function for this gene has been observed and no protein product for this gene has been identified.

Regulation of nitrogen fixation in *K. pneumoniae.*

As mentioned earlier, nitrogen fixation is very costly to the organism in terms of energy. Thus, in the presence of a readily usable nitrogen source, the organism shuts

down the *nif* gene functions. Also, because of the extreme sensitivity of the nitrogenase enzyme to oxygen, when high levels of oxygen occur the *nif* operons are not expressed. This suggests that a fairly complex mechanism of control must be involved.

In fact, regulation of the *nif* operons is governed by the action of at least 5 genes. Overall nitrogen metabolism in *K. pneumoniae* is mediated by a set of genes known as the *ntr* (nitrogen regulation) genes, specifically the *ntrA*, *ntrB*, and *ntrC* genes. These genes respond to availability of nitrogen by either activating or inhibiting nitrogen assimilatory functions, such as *nif* or glutamine synthetase. Two *nif* specific genes are also involved in the regulation of the *nif* operons. They are the *nifA* and the *nifL* genes.

Activation of the *nif* genes requires the products of the *nifA*, *ntrA* and *ntrC* genes. The *nifA* gene product is the positive activator of the *nif* operons and is required for the expression of the *nif* operons. In response to low levels of available nitrogen the *nifLA* operon is activated by the products of the *ntrA* and *ntrC* genes. Hirschmann *et al.* and Hunt *et al.* in 1985 looked at the transcription of the *glnA* gene of enteric bacteria *in vitro* and suggested that the *ntrA* may act in concert with the *ntrC* gene as a sigma-like factor enhancing the specific transcription of genes under the control of the *ntr* system.

Studies of the *ntrC* and *nifA* proteins have shown these proteins to have homologous functions. In fact Ow and Ausubel (1983) showed that the *nifA* protein could activate other nitrogen metabolism genes in the absence of a functional *ntrC* gene in *K. pneumoniae*. This kind of action was also observed by Drummond *et al.* (1983), who compared the promotor regions of the *nifLA* operon, the *nifHDKY* operon (which is only activated by the *nifA* protein and not the *ntrC*) and the *nifH* gene of *Rhizobium meliloti* (which is known to be activated by either the *ntrC* or the *nifA* proteins of *K. pneumoniae*). They found homologies in a region 30 bp upstream from the coding region of the *K. pneumoniae nifLA* operon and the *R. meliloti nifH* gene, which were not observed in the *K. pneumoniae nifHDKY* operon's -30 region. Thus, they hypothesized that this region

must somehow be important in the activation of these operons by the *ntrC* protein. The -10 region showed homologies for all three of the operons suggesting that this region must be important for activation by the *nifA* protein? In addition the product of the *ntrA* gene was also required, in conjunction with the *nifA* gene product, for the activation of the *nifHDKY* operon (Merrick 1983). This was similar to the *ntrC* protein which also required the action of the *ntrA* gene for activity.

In 1985 Buikema *et al.* sequenced the *ntrC* and *nifA* genes of *K. pneumoniae* as well as the *nifA* gene of *R. meliloti*. The predicted acid sequences were then compared and were found to contain regions of homologies in the central portions of the protein molecules. In 1985 Drummond *et al.* also sequenced the *K. pneumoniae nifA* and *ntrC* genes, showing that the sequence for the *nifA* gene extended 40 acids further at the carboxy-terminus than previously reported by Buikema *et al.* (1985). This region is thought to be involved in DNA binding of the protein molecule. Drummond *et al.* (1985) also found homology at the N-terminal regions of the *nifA* genes from *K. pneumoniae* and *R. meliloti*. The acid sequences of the *nifA* proteins and the *K. pneumoniae ntrC* protein were then compared with sequences of other regulatory proteins including the proteins coded for by the *ompR, dye*, and *nusA* genes of *E. coli* and the *spoOA* and *spoOF* genes of *Bacillus subtilis*. Homologies were found to exist in various regions of the protein sequences indicating that these regulatory proteins may have a evolved from a common ancestor.

In 1985 Kennedy *et al.* used cloned *nif* DNA from *K. pneumoniae* to examine the regulation of the *nif* genes in *A. vinelandii*. The study was conducted to observe similarities which occur in the regulation of nitrogen fixation between the two organisms. Plasmids carrying *K. pneumoniae nifA, ntrC*, and *nifL*, as well as *nifL-;nifH-* and *nifF-lacZ* fusions were used in the study. The results showed that the plasmid carrying the *K. pneumoniae nifA* gene could complement mutations in *A. vinelandii*, whereas the

plasmid carrying the *K. pneumoniae ntrC* could not. The results also showed that the *K. pneumoniae nifL* gene could not repress *nif* activity in *A.vinelandii*. With the *lac*- fusion plasmids the *nifL-lacZ* and *nifF-lacZ* were expressed constitutively in *A.vinelandii*, whereas the *nifH-lacZ* was not expressed except when the *nifA* of *K. pneumoniae* was also present. These results indicated that similarities exist in the regulatory mechanisms in *K. pneumoniae* and *A. vinelandii*, although the promotor regions of the *K. pneumoniae nif* genes appear to differ slightly from those of *A. vinelandii* as indicated in the *lac* - fusion studies. The plasmid carrying the *K. pneumoniae nifA* gene was able to complement regulatory mutants in *A. vinelandii*, indicating the *nifA* genes of these two organisms are similar. Parallel results have been observed by Sundaresan *et al.* (1983*a*) for *R. meliloti* , Jones *et al.* (1984) for *A. chroococcum*, Pederosa and Yates (1984) for *Azospirillum brasilense* and Alvarez-Morales and Hennecke (1985) for *R. japonicum*, in which the *nifA* gene product of *K. pneumoniae* was able to activate expression of the *nif* genes from these organisms. Activation of *nif* genes by the *ntrC* gene from *K. pneumoniae* was only observed in *R. meliloti* (Sundaresan *et al.*, 1983*b*).

Upon addition of ammonium ions to actively nitrogen fixing cultures, nitrogenase synthesis declines rapidly. This is due mostly to the repression of the *nifLA* operon by the *ntr* system and partly due to the action of the *nifL* gene product. Addition of oxygen to actively fixing cultures produces an even sharper decline in nitrogenase production. This rapid decline in nitrogenase production indicates that the repression of the *nif* genes is not occurring simply by repression of *nifA* expression and dilution of the *nifA* product. Mutation mapping experiments showed that *nifL* was involved in the repression of the *nif* genes. Collins *et al.* (1986) showed that the *nifL* gene product acts to turn off the *nif* genes by destabilizing the *nif* mRNA trancripts (except *nifLA*). This destabilizing effect appears to occur to a lesser degree in the presence of available nitrogen.

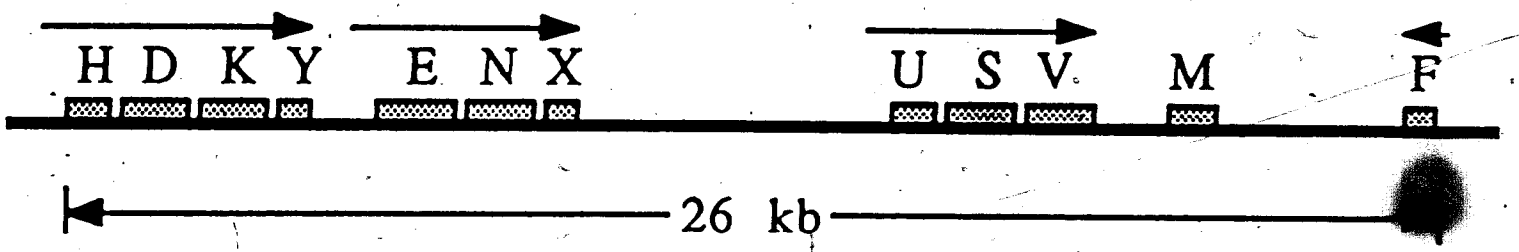
Nitrogen fixation in *Azotobacter vinelandii*

A. vinelandii is a nitrogen fixing organism capable of fixing nitrogen while growing in aerobic conditions. However the nitrogenase enzyme of this organism is still quite sensitive to oxygen and high concentrations of oxygen will inhibit the growth of *A. vinelandii* cells which are actively fixing nitrogen. To get around this problem the organism increases it's rate of respiration in order to decrease the internal concentration of oxygen thereby protecting the nitrogenase enzyme from deactivation. It has also been suggested that conformational changes in the enzyme may occur protecting it from deactivation.

Cloning and sequencing done on the *nif* region of *A. vinelandii* (Bishop *et al.*, 1985, Brigle *et al.*, 1985, Dean and Brigle, 1985, Brigle *et al.*, 1987, Beynon *et al.*, 1987), along with Tn5 mutagenesis (Kennedy *et al.*, 1986) have shown that the arrangement of the *nif* genes within the *A.vinelandii* genome is similar to that of *K. pneumoniae* (Figure 2). The structural genes of the *A. vinelandii nif* region, as well as the *nifY* gene are linked together in a single operon (Bishop *et al.*, 1985, Brigle *et al.*, 1985, Beynon *et al.*, 1987), as is the case in *K. pneumoniae*. The relative position of the *nifE* (Dean and Brigle, 1985, Brigle *et al.*, 1985), *nifX* (Beynon *et al.*, 1987), *nifU*, *nifS*, *nifV* (Kennedy *et al.*, 1986, Beynon *et al.*, 1987), *nifN*, and *nifM* (Kennedy *et al.*, 1986) appear to be identical to the arrangement found in *K. pneumoniae*. The *nifE*, *nifN* and *nifX* genes comprise a single transcriptional unit (Dean and Brigle, 1985, Beynon *et al.*, 1987), as do the *nifU*, *nifS* and *nifV* genes (Beynon *et al.*, 1987).

The control of the nitrogen fixing ability in *A. vinelandii* appears to be similar to that of *K. pneumoniae*. In fact Kennedy and Robson in 1983 used the cloned *nifA* regulatory gene to activate expression of the *nif* genes in both *A. vinelandii* and *A. chroococcum*. This was shown by cloning of the *nifA* gene into a wide host range multicopy plasmid,

Figure 2. Genetic organization of *nif* genes of *A. vinelandii*. (from Beynon *et al.* 1987)



with expression of the *nifA* gene occurring from a constitutive promotor. Because the *nifA* gene was expressed constitutively from this promotor it was expected that if enough homology occurred between the regulatory elements of *K. pneumoniae* and the *Azotobacter* spp., expression of the nitrogen fixation genes in *A. vinelandii* and *A. chroococcum* would also occur constitutively upon transformation of the bacteria with the plasmid. This was observed, indicating that a protein homologous to the *nifA* protein of *K. pneumoniae* exists in *A. vinelandii*. This was further supported by work done by Kennedy and Drummond in 1985 who, as mentioned earlier in this introduction, used cloned *K. pneumoniae* regulatory elements to examine the regulation of nitrogen fixation in *A. vinelandii*. In addition to the activation of the *A. vinelandii nif* genes by the *K. pneumoniae nifA* gene, they also observed that the *nifL* gene from *K. pneumoniae* was not expressed in *A. vinelandii*.

Further studies by Santero *et al.* in 1987 showed that a gene similar to the *ntr* A gene of *K. pneumoniae* may regulate overall nitrogen metabolism in *A. vinelandii*. This was shown by isolation of mutants which were simultaneously deficient in nitrogen fixation, nitrate reductase and nitrite reductase activities. All of these deficiencies were complemented by a cosmid containing a cloned *A. vinelandii* DNA fragment which also complemented *ntr* A mutants of *E.coli*. This indicated that a *ntr* A like gene exists in *A. vinelandii*.

In 1980 Bishop *et al.*, obtained some rather unusual results which led them to hypothesize the existence of a second nitrogen fixation system in *A. vinelandii*. They found Nif^+ (able to fix N_2) phenotypic revertants of Nif^- point mutations of the *nifD* and *nifK* genes. Two-dimensional protein gel analysis showed that these 'revertants' lacked the *nifD* and *nifK* proteins, but contained four new proteins which were repressed in the presence of NH_4^+. This system was also found to be expressed under conditions of

molybdenum deficiency and repressed in the presence of molybdenum or tungsten (tungsten has been shown to compete with molybdenum at the active center of the FeMo cofactor, resulting in an inactive enzyme complex).

Bishop *et al.* (1980), Page and Collinson (1982) and Premakumar *et al.* (1984) suggested that the two systems were regulated by molybdenum. That is, that the concentration of available molybdenum determined which system was expressed or repressed. Jacobson *et al.* in 1986 provided evidence that this control existed at the level of transcription. This was done by looking for *nif* homologous transcripts under varying concentrations of molybdenum. They found that with molybdenum concentrations of greater than 50nM the conventional nitrogen fixation system was expressed whereas at concentrations of less than 50 nM the alternative system was expressed.

The metal vanadium has been shown to enhance basal nitrogen fixation activity (Bishop *et al.*, 1982, Page and Collinson, 1982)). Bishop *et al.* (1982) and Page and Collinson (1982) speculated that the enhancement observed was the result of stimulation either, directly or indirectly, of the alternative system by vanadium. Exactly what role vanadium play in the alternative system was unclear, although both Bishop *et al.* (1982) and Page and Collinson (1982) suggested that the enhancement was the result of an increase in molybdenum deficiency by competition of vanadium with molybdenum transport , resulting in an increased 'derepression' of the alternative system. Another possible explanation is that vanadium is at the active center of the alternative nitrogenase. This hypothesis was supported by Robson *et al.* (1986) who showed the alternative nitrogenase of *A. chroococcum* was a vanadoprotein.

Inter-species homology of nitrogen fixation systems.

As mentioned at the beginning of this introduction, a wide variety of prokaryotic organisms are capable of fixing nitrogen. These organisms range from strict aerobes like

Azotobacter vinelandii, to strict anaerobes like *Clostridium pasteurianum*, and include organisms like the cyanobacterium *Anabaena*. Although bacteria from widely divergent taxonomic groups can fix nitrogen, it appears that the nitrogenase enzymes of these organisms are closely related. The nitrogenases of all of these organisms are composed of two components. Component 1, or the FeMo protein, contains the active site of the enzyme complex. Component 2, or the nitrogenase reductase, is the electron acceptor of the enzyme complex.

In 1968, Detroy *et al.* were the first to show the ability of components from one organism to form active nitrogen fixing complexes with the complementary components from another organism. Since that time a number of other groups have studied the complementary functioning of nitrogenase components from a wide range of nitrogen fixing organisms.

In 1978, Emerich and Burris, using highly purified components from a wide variety of nitrogen fixing organisms, showed that hybrid protein complexes with components originating from two different sources were active in fixing nitrogen. This was the first such study to utilize highly purified protein for this purpose. In the study, protein components from 8 different organisms, which fixed nitrogen under widely varying physiological conditions, were used. The protein components of the nitrogenase complexes of each of these organisms were isolated and tested for the ability to form active hybrid nitrogenase complexes. Of the 56 possible heterologous crosses, 45 generated active complexes. This suggested that the protein sequences of the nitrogenase enzyme complex components are highly conserved and possibly that the acquisition of nitrogen fixing ability has been a fairly recent evolutionary event.

In 1980 Ruvkun and Ausubel, using *nif* genes cloned from *K. pneumoniae*, showed that a significant degree of homology existed in the DNA coding for the structural genes of the nitrogenase enzymes from a number of nitrogen fixing organisms. The cloned *nif*

genes used in this study were carried on the plasmids pSA30 and pMC1. These plasmids were constructed by Cannon *et al.* in 1979 and contained all of the *nif* genes except *nifJ*. The plasmid pSA30 contains the *nifH*, *nifD*, *nifK* and *nifY* genes as well as part of the *nifE* gene. The plasmid pMC1 contains the rest of the *nif* genes not including the *nifJ* gene. From these plasmids, fragments containing the *nif* genes were first isolated and then radioactively-labelled and used as hybridization probes with Southern transfers of DNA digests from 19 nitrogen fixing organisms and 10 non-nitrogen fixing organisms. The results showed that all of the DNA samples from nitrogen fixing organisms and none of the DNA samples from the non-nitrogen fixing organisms contained homology with the *nif* structural genes (from pSA30). When the probe containing the other *nif* genes (from pMC1) was used none of the organisms tested exhibited any homology. The results of this study indicated that a high degree of homology appears to exist, at least for the structural genes of the nitrogen fixing system, and that other genes specific for the processing and functioning of the nitrogenase appear to have a much lower degree of conservation. Because of this the structural genes of *K. pneumoniae* have been used as a "universal" probe by many groups to localize or identify the structural nitrogen fixation genes of a wide variety of organisms.

The first group to use cloned *K. pneumoniae nif* DNA as an interspecies hybridization probe for *nif* genes was Nuti *et al.* (1979). In this study they were able to localize the *nif* genes of the symbiotic nitrogen fixer *Rhizobium leguminosarum* to a large indigenous plasmid. It was found that the plasmid contained DNA homologous to the *nif* structural genes of *K. pneumoniae* but not to the other *nif* genes.

Since then the 'universal' probe has been used to localize the *nif* structural genes of *Anabaena* 7120 (Mazur *et al.*, 1980,and Rice *et al.*, 1982), *Rhizobium leguminosarum*, *R. trifolii*, *R. phaseoli* (Prakash *et al.*, 1981), *R. meliloti* (Prakash *et al.*, 1981 and Corbin *et al.*, 1983), *R. japonicum* (Hennecke 1981 and Kaluza *et al.*, 1983), *Rhizobium* ORS571

(Norel *et al.*, 1985), *Rhodopseudomonas capsulata* (Avteges *et al.*, 1983), *Desulfovibrio* spp. (Postgate *et al.*, 1986), and *Thiobacillus ferrooxidans* (Pretorius *et al.*, 1986). The universal probe has also been used to identify *nif*-like DNA in organisms such as *Methylosinus* sp., *Xanthobacter* sp. H4-14 (Toukdarian and Lidstrom 1984), *Methanobacterium ivanovi, M. thermoautotrophicum, Methanococcus voltae,* and *Methanosarcina barkeri* (Sibold *et al.*, 1985), even though capability to fix nitrogen has not been demonstrated in most of these organisms.

The *nif* DNA from the cyanobacterium *Anabaena* 7120 has also been used as a hybridization probe for the mapping of *nif* DNA of a number of organisms. Kallas *et al.* in 1985 used the *nifD* and *nifK* genes of *Anabaena* 7120 to examine the organization of *nif* genes in various nonheterocystous and heterocystous cyanobacteria. Their findings suggested that differences in the arrangement of the *nif* genes between nonheterocystous and heterocystous cyanobacteria could be useful in the taxonomy of the various cyanobacteria.

It is apparent from these studies that the structural genes of the nitrogen fixing systems of all of these organisms are fairly closely related, to the extent that they can be identified by hybridization with the universal probe from *K. pneumoniae* and that nitrogen fixing activity is observed when nitrogenase protein components from heterologous sources are combined. The extent of homology between the nitrogen fixing systems of various organisms can be further examined by comparison of the actual acid sequences of the nitrogenase enzyme components themselves. The acid sequences of the proteins can be determined by two methods. The first method is to determine the sequence directly using purified protein. The other method is to determine the DNA sequence of the nitrogen fixation genes and from this sequence predict the acid sequence of the proteins for comparison. The latter method is advantageous as from the DNA sequence data, promotor sequences and gene arrangements can also be compared. Also the DNA sequence is less

likely to be in error. The sequencing of DNA is also somewhat easier to carry out in terms of time, materials and effort.

Comparisons of the acid and DNA sequences of the *nif* structural genes

The acid sequence of the *nifH* gene products of a number of organisms has been determined. The acid sequence of *Clostridium pasteurianum* Fe- protein was determined in 1977 by Tanaka *et al.* directly using the purified protein. The acid sequences of the Fe proteins of *Anabaena* 7120 (Mevarech *et al.*, 1980), *K. pneumoniae* (Scott *et al.*, 1981, Sundaresan and Ausubel, 1981), *R. meliloti* (Török and Kondorosi 1981), *R. trifolii* (Scott *et al.*, 1983*a*), *R. japonicum* (Fuhrmann and Hennecke 1984), *A. vinelandii* (Brigle *et al.*, 1985), and *Thiobacillus ferrooxidans* (Pretorius *et al.*, 1987) were all determined from the nucleotide sequences of the *nifH* genes of these organisms. From the acid sequences a great deal of homology is apparent. Regions in the protein containing cysteine residues appear to show the greatest amount of conservation, which is not unexpected since cysteine has been thought to act as ligands at the 4Fe-4S centre of the protein and may also be involved in the binding of ATP.

For the *nifD* gene, coding for the α-subunit of the FeMo protein, the complete nucleotide sequences of *A. vinelandii* (Brigle *et al.*, 1985), a *Rhizobium* sp. isolated from *Parasponia andersonii* (Scott *et al.*, 1983*b*), R. *japonicum* (Kaluza and Hennecke 1984), and *Anabaena* 7120 (Lammers and Haselkorn 1983) have been determined. Partial sequences for the *nifD* genes of *Thiobacillus ferrooxidans* (Pretorius *et al.*, 1987), *K. pneumoniae* (Scott *et al.*, 1981), *R. trifolii* (Scott *et al.*, 1983), and *R. meliloti* (Török and Khodorosi 1981) have also been determined. The acid sequences coded for by these genes was determined from the nucleotide sequence. The acid sequence of the α-subunit of the *Clostridium pasteurianum* nitrogenase (Hase *et al.*, 1984) was determined using purified protein. Again, the acid sequence showed a significant amount of conservation

especially around the cysteine residues, as was the case with the Fe protein.

The acid sequences of the protein encoded by the *nifK* gene were determined from the complete nucleotide sequences of the *nifK* genes from *Anabaena* 7120 (Mazur and Chui, 1982), *A. vinelandii* (Brigle *et al.*, 1985) and the beginning of the *nifK* gene of *R. japonicum* (Kaluza and Hennecke, 1984). The acid sequence of portions of the *nifK* gene protein from *A. vinelandii* (Lundell and Howard, 1981), and *C. pasteurianum* (Hase *et al.*, 1984) have also been determined. Since only a limited amount of data is available, protein comparisons are difficult to make, however significant degrees of homologies do exist within the sequences compared.

Thesis objectives

The project described in this thesis is the isolation and sequencing of DNA from the nitrogenase gene cluster of *Azotobacter vinelandii*. The DNA isolated was determined to be from the nitrogenase gene cluster on the basis of hybridization with the *K. pneumoniae* nitrogenase genes found in pSA30. The sequence of the isolated DNA was determined by the methods of Sanger *et al.*(1977). The objective of this project was to determine the sequence of the DNA isolated from the nitrogenase gene cluster of *A. vinelandii* and compare the nucleotide sequence as well as the predicted acid sequence with those previously determined by other groups for other nitrogen fixing organisms.

MATERIALS AND METHODS

Bacterial Strains and Media

The *Azotobacter vinelandii* strain used to isolate *nif* genes was *A. vinelandii* OP strain UW (*nif* $^{+}$, capsule^{-}) obtained from Dr. W. Page (University of Alberta). The cultures were maintained on Burk's medium slants. Compositions of the media used are listed in Table 2.

The *E. coli* strain used as the host for cosmid cloning was *E. coli* HB101 (Boyer and Roulland -Dussoix, 1969). Cultures were maintained on L-medium with the appropriate concentration of antibiotic present to maintain the cosmids of interest. The host used for pUC vectors was *E. coli* JM83 (Pharmacia, Dorval, Que), while the host used for M13 cloning was *E. coli* JM103 (Bethesda Research Laboratories, Burlington, Ont.). Again antibiotics were added to YT medium to appropriate concentrations to maintain the plasmids of interest. For the bacteriophage M13 no antibiotics were used.

Cloning Vectors and Plasmids

Cloning vectors and plasmids used are listed in Table 3. The cosmid pSa747 (Tait *et al.* 1983) was used as the initial vector to produce the *A. vinelandii* gene library. The cosmid was maintained in *E. coli* HB101 on L-medium containing 30µg/mL kanamycin.

For the purpose of sub-cloning, the high copy number plasmids pUC8 and pUC13 were used. These vectors are convenient for sub-cloning as cloned fragments can be sequenced using either the dideoxy or Maxam & Gilbert sequencing methods. Both pUC8 and pUC13 carry an ampicillin resistance marker and were thus maintained in *E. coli* JM83 on YT-medium containing 50µg/mL ampicillin.

In order to detect recombinant clones containing *nif* genes, a radio-actively labelled DNA fragment containing the *K. pneumoniae nifH,D,K,Y* and part of *nifE* from the plasmid pSA30 was used (Cannon *et al.*, 1979, obtained from F. M. Ausubel, Harvard University, Cambridge, Mass.) as hybridization probe.

Table 2. Media

Burk's Medium	K_2HPO_4	0.8 g	5mM phosphate buffer pH7.1
	KH_2PO_4	0.2 g	
	$MgSO_4 \cdot 7H_2O$	0.2 g	0.81 mM
	$CaSO_4 \cdot 2H_2O$	0.08 g	0.58 mM
	$FeSO_4 \cdot 7H_2O$	5 mg	18 μM
	$NaMoO_4$	2.5 mg	1 μM
	D-glucose	10 g	1% (w/v)
	±Difco Bacto-agar	18 g	1.8% (w/v)
	$\pm CH_3COONH_4$	1.1 g	14mM(200μg N/L)
	distilled H_2O to 1000 ml		

L-medium	Difco Bacto-tryptone	10 g
	Difco Yeast extract	5 g
	NaCl	5 g
	D-glucose	1 g
	±Difco Bacto-agar	15 g for plates 6 g for soft agar overlays
	±Antibiotics	50 μg/ml for ampicillin 30 μg/ml for kanamycin 15 μg/ml for tetracycline
	distilled H_2O to 1000 ml	
YT-medium	Same as L-medium with no D-glucose added.	

Table 3. Cloning vectors and plasmids

Plasmid/vector	Relevant genotype	Reference
pSa747	Sp^r,Km^r,*cos*	Tait *et al.*, 1983
pUC8	Ap^r	Vieira and Messing, 1982
pUC13	Ap^r	Messing, 1983
pSA30	*nif* HDK, Tc^r	Cannon *et al.*, 1979

Isolation of *A. vinelandii* DNA

Cultures of *A. vinelandii* OP strain UW were initially grown on N-free Burk's medium slants. Two litres of liquid Burk's medium were inoculated with cell suspensions obtained by washing cells off the slants. To enhance growth, ammonium acetate was added to a concentration of 1g/litre. The cultures were incubated at 30°C with vigorous aeration on a rotary shaker. After 2 days the cells were harvested by pelleting the cells and resuspending in 200 mL of TE buffer (10 mM Tris-HCl pH7.3, 1mM EDTA). SDS (sodium dodecyl sulfate) was added to a final concentration of 5% (w/v) and the suspension was incubated at 44°C for 1 hour. The suspension was then phenol extracted with one volume of phenol. The layer was removed and saved. The phenol layer was washed twice with 25 mL of TE buffer and the TE buffer washes were pooled together with the original layer. Sodium acetate was added to the solution to a final concentration of 0.3M. Two and a half volumes of were then added and the suspension was incubated at -20°C overnight. The precipitated DNA was then pelleted and redissolved in 10 mL of TE buffer. A 1µL aliquot was electrophoresed on a 0.4% agarose slab gel, followed by staining with ethidium bromide to get a rough estimate of the DNA content of the suspension. The DNA was then purified by CsCl density gradient centrifugation. This was done by adding CsCl (1.1g /mL of solution) to the DNA solution and ethidium bromide to a final concentration of 250µg/mL followed by centrifugation of the solution at 40,000 rpm for 40 hr in a Beckman Ti50 rotor using a Beckman L5-75 ultracentrifuge. The DNA formed a distinct band which was removed from the top with a 14 gauge needle. The ethidium bromide was removed by repeated extraction with 2 volumes of isoamyl alcohol. The DNA was then precipitated using 3 volumes of 95% The pellet was redissolved in TE buffer containing 0.3M sodium acetate and precipitated with 3 volumes of 95% . The pellet was redissolved in 2mL of TE buffer The concentration of the final

DNA solution was determined by measuring the absorbance at 260 nm (where 50µg/mL of DNA gives an absorbance of 1).

Cloning of *A. vinelandii nif* genes:

A flow diagram of the cloning strategy is shown on Figure 3.

Preparation of vector DNA

The vector pSa747 was linearized with the restriction endonucleases *Bgl* II (Pharmacia), *Eco* RI (New England Biolabs), *Hin* dIII (Pharmacia) and *Sst* II (New England Biolabs) under conditions described by the supplier of the enzymes. The digests were checked for completion by agarose gel electrophoresis. Linearized vectors were then treated with calf intestinal alkaline phosphatase to reduce the likelihood of vector-vector ligation.

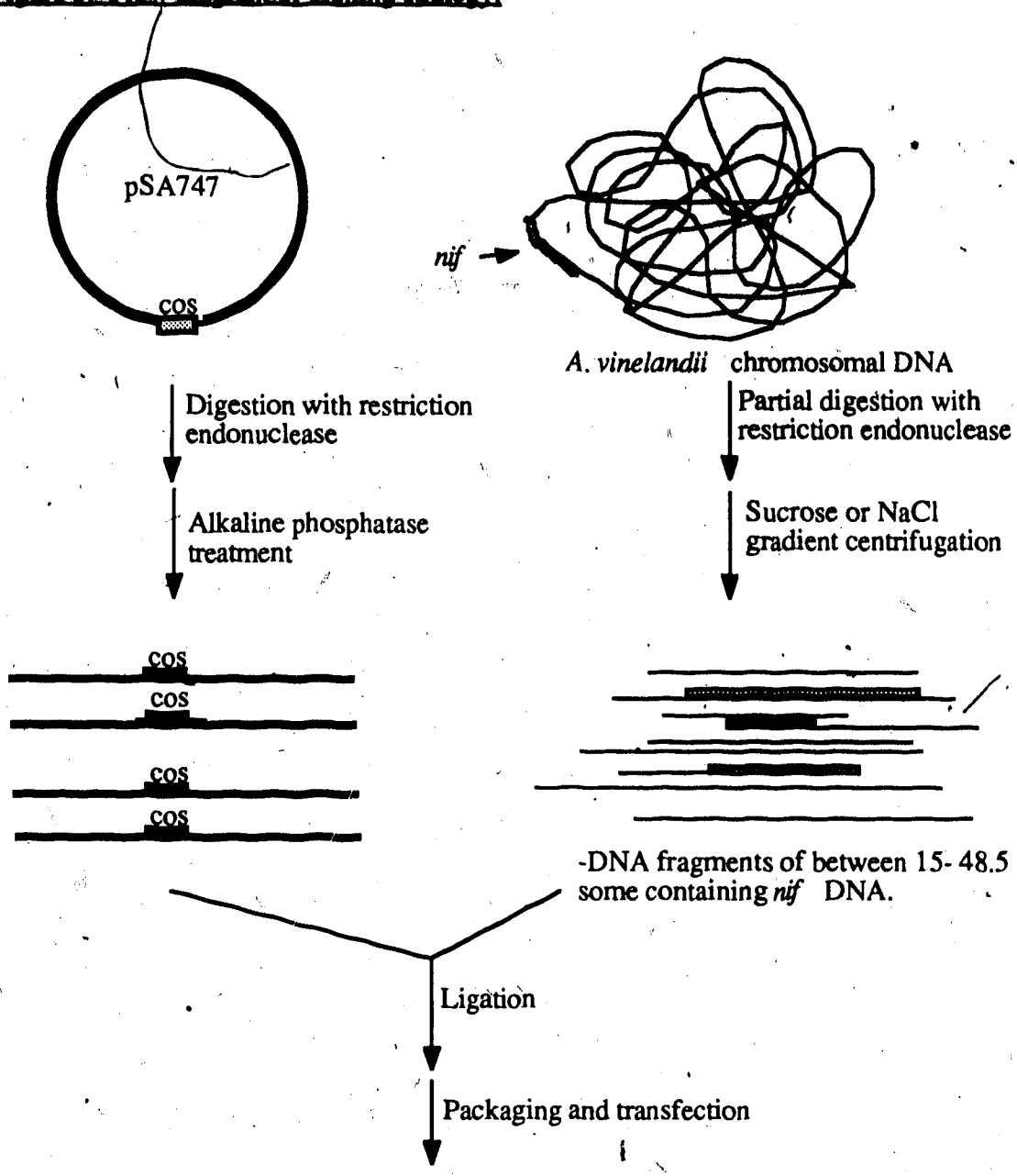
Preparation of *A. vinelandii* DNA.

In order to obtain fragments of suitable size for cloning into pSa747, partial digests were done on the *A. vinelandii* DNA. The enzymes used for the partial digests were *Bam* HI(New England Biolabs), *Bgl* II, *Eco* RI , *Hin* dIII, *Mbo* I (New England Biolabs), and *Sst* II. The conditions for partial digestion which produced the best range of fragments were determined by small scale digestion of the DNA. Once conditions were established large scale digestions were carried out. For each large scale digestion approximately 50µg of DNA were used. These digests were then size fractionated by centrifugation in either sucrose (10-40%) or NaCl (5-29%) linear gradients. The gradients were fractionated and the fragment sizes of the DNA contained in the fractions were determined by agarose gel electrophoresis with suitable size markers. Fractions containing fragments in the 21-36 kb range were pooled and the DNA was concentrated by precipitation and redissolution in 200µL of TE buffer.

Ligation of pSa747 with *A. vinelandii* DNA fragments.

Approximately 1µg of linearized pSa747 was ligated to equal amounts of *A. vinelandii*

Figure 3. Flow diagram of cloning strategy.



Clones are selected by resistance to kanamycin and screened by colony hybridiztion with *nif* genes of *K. pneumoniae*, carried on pSA30, for the presence of *nif* DNA.

DNA in a final volume of 10 µL using 3 units of T4 DNA ligase (New England Biolabs) using conditions described by the supplier of the enzyme. Table 4 lists the vector sites into which the partial digest fragments were cloned. The ligations were incubated overnight at 12°C and stopped with 5 µL of 0.5M EDTA pH 8.

Packaging and Transduction

The ligated DNA was packaged *in vitro* into lambda phage heads using procedures described by Collins (1979) and Hohn (1979). The packaged DNA was then used for transduction of *E. coli* HB101 cells, which were subsequently plated on L-agar containing 30 µg/mL kanamycin. Enough packaged DNA was used to produce between 30 and 300 colonies per plate.

Screening of colonies

In order to get a rough estimate of the number of possible recombinant colonies which were present with each ligation, a number of colonies were randomly selected and "quick plasmid assays" were done using the methods of McCormick *et al.*, 1981. From this the ligations producing the highest proportion of recombinants can be selected.

Following this, colony hybridizations were carried out using the methods of Gergen *et al.* (1979). Colonies were screened with a ^{32}P labelled fragment containing cloned *K. pneumoniae nif* genes (*nifH*, *nifD*, and *nifK*), from pSA30. Any possible recombinant cosmids were then isolated using the methods of Birnboim and Doly (1979) and were then electrophoresed on agarose. The DNA was transferred to nitrocellulose using the methods of Southern (1975) and the filters were probed with a radioactively labelled sample of the *Eco* RI insert of pSA30 to confirm that the cosmids contained DNA homologous to the *K. pneumoniae* DNA.

Isolation and Purification of pAvH1

Large scale isolation of pAvH1 DNA was done using the cleared lysate method

Table 4. Ligation of *A. vinelandii* DNA to pSa747

Vector site	*A. vinelandii* fragments
Bgl II	*Bam* HI, *Bgl* II, *Mbo* I
Eco RI	*Eco* RI
Hin dIII	*Hin* dIII
Sst II	*Sst* II

developed by Clewell & Helinski (1969). The DNA was purified by cesium chloride gradient centrifugation and the purity of the sample was checked by electrophoresis on 0.35% agarose slab gels. The concentration of the DNA solution was determined by measuring the absorbance at 260 nm.

Restriction Mapping of pAvH1

A restriction map of pAvH1 was constructed by analysis of digestion patterns of various restriction endonucleases. Single and double digests were carried out. Fragments were fractionated by electrophoresis on 0.35% agarose slab gels and were visualized by ethidium bromide staining. Then, from photographs of the gels, fragment sizes could be determined (Southern,1979). The enzyme sites mapped were *Bam* HI (New England Biolabs), *Bgl* II (Pharmacia), *Cla* I (Boehringer-Mannheim, Munich, West Germany), *Eco* RI (New England Biolabs), *Eco* RV (New England Biolabs), *Hin* dIII (Pharmacia), Hpa I (New England Biolabs), *Pvu* II (Pharmacia), *Sma* I (Bethesda Research Laboratories), and *Xho* I (New England Biolabs). Generally, 1 µg of DNA was digested for 2 hours with 2 to 5 units of enzyme in conditions as described by the suppliers. The DNA fragments from the agarose gels were then transferred to nitrocellulose filters by the methods of Southern (1975), and the filters were probed with radioactively labelled Eco RI insert of pSA30 to determine which fragments had homology to the *K. pneumoniae nif* genes.

Subcloning of pAvH1 DNA.

The pAvH1 DNA was prepared by digestion with the appropriate enzymes under conditions as described by the enzyme suppliers. The digests carried out were an *Eco* RI single enzyme digestion and a *Hin* dIII -*Bgl* II double digest. The digests were checked by electrophoresis of small aliquots on 0.75% agarose gels. The fragments of interest for subcloning were isolated using the DEAE-membrane band intercept method (Lizardi *et al.*, 1984). The method followed was described by the manufacturer of the membrane (Schleicher & Schuell Inc., Keene, N. H.). Enough pAvH1 was digested (between 60 to

70 µg) to produce approximately 2 µg of the final fragments.

The pUC vectors were digested with appropriate enzymes as described by the suppliers. Approximately 3 µg of pUC8 vector DNA was digested with *Eco* RI. An aliquot of the digest containing approximately 0.1 µg of DNA was electrophoresed on a 0.75% agarose gel to confirm that digestion was complete. The concentration of the completely digested DNA was adjusted to approximately 1 µg/µL by dissolving it in the appropriate volume of TE buffer after precipitation of the DNA. For the pUC13 vector DNA, three µg were digested first with *Hind* III, followed by *Bgl* II. After each digestion, aliquots containing approximately 0.1 µg were electrophoresed on 0.75% agarose gels to confirm that the digestions were complete. Once digestion was completed, the DNA was purified by DEAE membrane band interception and redissolved to give a final concentration of approximately 1 µg/µL.

Approximately 1 µg of linearized vector DNA was ligated to approximately 1 µg of appropriate fragment in a final volume of 10 µL using 3 units of T4 DNA ligase (New England Biolabs). The incubations were carried out overnight at 12°C and stopped with 5 µL of 0.5M EDTA pH8. The ligated DNA was used to transform competent *E. coli* JM83 cells. The pUC plasmids all contain part of the ß-galactosidase gene from *E. coli* and are able to complement certain mutations in the ß-galactosidase gene. Cloning is done into a region containing a number of unique restriction endonuclease sites within the partial ß-galactosidase gene, which does not alter the reading frame of the DNA. Recombinants were selected for on L-agar + ampicillin (50µg/mL) plates containing 5-Bromo-4-chloro, 3-indolyl-ß-D-galactopyranoside (X-gal) and isopropyl- ß-D-thiogalactopyranoside (IPTG). With X-gal and IPTG the recombinant plasmids could be selected on the basis of the presence or absence of active ß-galactosidase. Cells producing active ß-galactosidase would break down the X-gal to produce a blue colored pigment, thus producing blue

colonies. On the other hand, if fragments were cloned into the vector, the ß-galactosidase gene sequence would be interrupted, preventing complementation of the ß-galctosidase gene in the host cells, resulting in white colonies. The IPTG was added as an inducer of the ß-galactosidase gene, but is not itself a substrate of the enzyme. White colonies were screened by colony hybridization using the methods of Grunstein and Wallis (1979), with the purified 1.4 kb *Eco* RI fragment of pAvH1 as probe.

Sequencing of *A. vinelandii* DNA

The sequence of the 1.4 kb Eco RI fragment was determined using the chain termination method of Sanger *et al.*. (1977). For this method, the DNA was first cloned into M13 bacteriophage. The vector used was M13-mp9. The M13-mp9 DNA (replicative double stranded form) was digested with *Acc* I using conditions specified by the manufacturer of the enzyme. The concentration of the digested DNA was adjusted to a final concentration of 10ng/µL by dissolving the DNA in an appropriate volume of TE buffer after precipitation of the DNA.

The pAvH1-1 DNA was digested with *Hin* PI, *Msp* I, *Nar* I, or *Taq* I, which can all be used for cloning into the Acc I site. The DNA was precipitated and redissolved in TE buffer to a final concentration of approximately 10ng/µL. One µL (10ng/µL) of appropriately digested vector (M13-mp9) was ligated to 1µL (10ng/µL) of the corresponding pAvH1-1 DNA with approximately 3 units of T4 ligase in a volume of 5 µL under conditions described by the manufacturer of the enzyme. The ligations were carried out at 12°C overnight.

Ligated DNA was used to tranform *E. coli* JM103 cells made competent by the methods of Morrison (1979) and plated out to give between 10-300 plaques per plate on L-medium containing X-gal and IPTG. Plaques containing possible recombinants were identified by the inability of the phage to break down X-gal. White plaques were

transferred using sterile toothpicks to 3mL of YT-broth and were then allowed to grow for six hours at 37°C and the cultures were transferred to 4°C. The cells were allowed to settle and 20 µL of the cleared culture media was then spotted onto nitrocellulose. The bacteriophage were then lysed on the nitrocellulose filters by floating the filters for 5 min. on a solution containing 0.1M NaOH and 1.5M NaCl. The filters were washed by floating them on a solution containing 0.5M Tris-HCl pH7.3, 3M NaCl for 5 minutes. The filters were dried and then baked at 80°C for 2 hours. These filters were then probed with the 1.4 kb *Eco* RI fragment of pAvH1-1.

Another method used to screen the possible recombinant plaques was direct gel electrophoresis. For this method 20 µL of the cleared culture medium was added to 1 µL of 2% (w/v) SDS and 3 µL electrophoresis loading dye (20% Ficoll, 0.3% Orange G dye, 50mM EDTA pH8). These samples were loaded directly onto 0.75% agarose gels and electrophoresed. The DNA was visualized by ethidium bromide staining and UV illumination. The DNA in these gels was transferred (without alkaline denaturation) to nitrocellulose by the methods of Southern (1975). Again, these filters were probed with the 1.4 kb Eco RI fragment of pAvH1-1.

Since it is possible for insert fragments to be inserted in two orientations, it was necessary to conduct tests on those recombinants which appeared to be of similar size. This was done by taking 20µL from each of the samples in question and mixing them with 1 µL of 2% (w/v) SDS and 6 µL electrophoresis loading dye. The mixture was overlayed with mineral oil and incubated at 65°C for one hour. The samples were then loaded onto 0.75% agarose gels and electrophoresed. DNA from phage containing opposite strands will anneal together causing extra bands to appear on the gel. Wherever possible, clones for each orientation were isolated.

The M13 template DNA was then isolated from screened samples as follows. First the

phage particles were separated from the whole cells by centrifugation at low speed. The supernatant was saved (a little less than 3 mL) and 600µL of a polyethylene glycol (PEG) solution was added (20% w/v PEG, 2.5M NaCl) to precipitate the phage, and the suspension was incubated at room temperature for 15 min. This solution was centrifuged (5000 rpm, 10min) and the pellet was saved. The pellet was then resuspended in 300 µL of TE buffer and transferred to a 1.5 mL Eppendorf tube. An equal volume of phenol (300 µL) was added and the tube was vortexed for 15 sec. The tube was incubated at room temperature for 5 min. and vortexed again. The tube was centrifuged for 4 min. and the layer was removed to a new tube. Sodium acetate (3M) was added to a final concentration of 0.3M and 3 volumes of 95% were added. The tube was incubated at -20° overnight. The solution was then centrifuged to pellet the DNA and the supernatant was removed. One mL of 70% was then added to the pellet without disturbing it and the tube was centrifuged again for 2 min. The supernatant was removed and the pellet was redissolved in 20 µL of TE buffer. This DNA was then used as template DNA in the sequencing reactions. Sequencing reactions were carried out using α-[^{32}P]-dATP.

Samples were run on polyacrylamide gels of varying concentration for varying times dependant on the sample being sequenced. Generally, gels of higher concentration were run for shorter times and were used to obtain sequence in regions where the fragment sizes were small. Lower concentration gels were normally run for longer periods of time to obtain sequence in regions where fragments were large.

Computer Analysis of Sequence

The DNA sequences obtained were analyzed using the programs of Staden (1980). This system allows one to search the DNA for restriction sites and to translate the DNA in all 3 reading frames into a predicted acid sequence.

RESULTS

Preparation of *A. vinelandii* DNA for ligation

Partial digests of the *A. vinelandii* DNA were used to obtain fragments for cloning into the cosmid vector pSa747. To establish the optimum conditions for the partial digests small scale digestions using approximately 1 µg of *A. vinelandii* DNA were done. The partial digests were checked by gel electrophoresis with suitable size markers (lambda DNA , which is 48.5 kb, + lambda DNA digested with *Xho* I, which gives fragments of 33.5 kb and 15 kb) to determine which conditions gave the best range of fragments. Partial digestions were carried out two ways. The first method was to use enough enzyme to completely digest the DNA after about two hours and to remove aliquots at time intervals during the digestion to give partial digestion. This method was used with *Bam* HI , *Bgl* II, *Eco* RI and *Mbo* I. The second method was to use a low concentration of enzyme to give partial digestion after one hour of incubation. This method was used to obtain partial digests with *Hin* dIII and *Sst* II. Figure 4 shows agarose gels from one of the small scale digests from which conditions for large scale digestion were determined. Large scale digests were done using approximately 50 µg of DNA. The large scale digests were then fractionated using sucrose or sodium chloride gradients. Fractions were electrophoresed on agarose gels to determine which fraction(s) contained DNA of the desired size. Figure 5 is an example of agarose gels used for this purpose.

Preparation of pSa747

Approximately 10 µg of cosmid DNA was digested to completion with the appropriate restriction enzymes. The digested DNA was electrophoresed on agarose to check for complete digestion (Figure 6).

Isolation and identification of clones

The *A. vinelandii* DNA was ligated to the cosmid DNA and packaged into lambda phage heads as described in Methods. The packaged DNA was used to transfect *E. coli*

Figure 4. Small scale digests of *A. vinelandii* DNA. Lane 1: λDNA (48.5 kb) + λDNA digested with *Xho* I (33.5 kb and 15 kb), Lanes 2-5: One μg of *A.vinelandii* DNA digested with 0.8, 0.4, 0.2, and 0.1 units of *Hin* dIII respectively, for a 2 hour incubation period at 37°C. Numbers at left indicate size in kb.

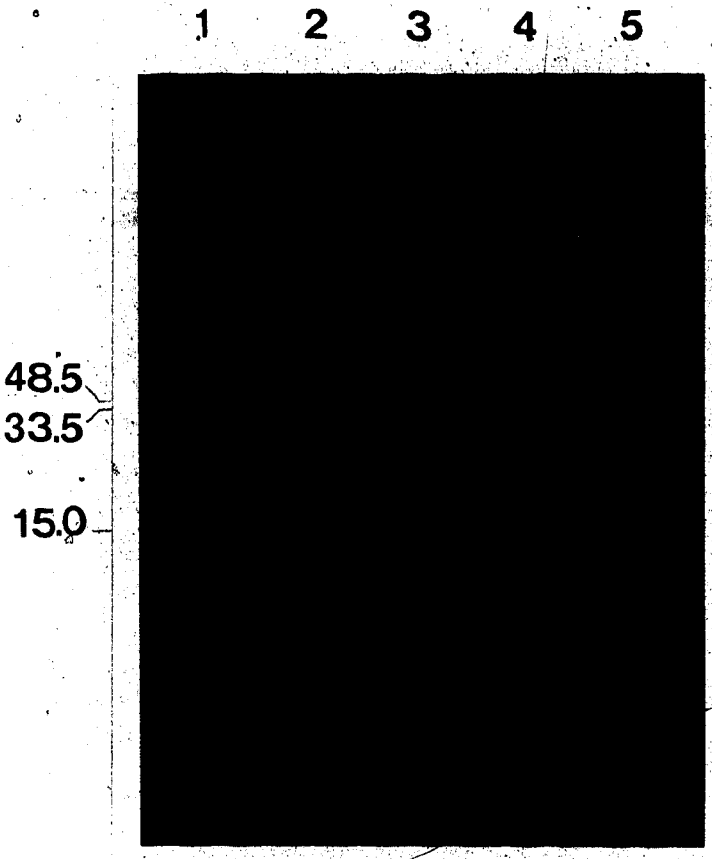

1
2
3
4
5
48.5
33.5
15.0

Figure 5. Sucrose gradient of partially digested *A. vinelandii* DNA. Lanes 1 and 9 are λ DNA (48.5 kb)+λ DNA digested with *Xho* I (33.5 kb and 15 kb). Lanes 2-8 are aliquots from fractions (from the top to bottom of the gradient respectively) of the sucrose gradient for *A. vinelandii* DNA partially digested with *Hin* dIII. Numbers at right indicate size in kb.

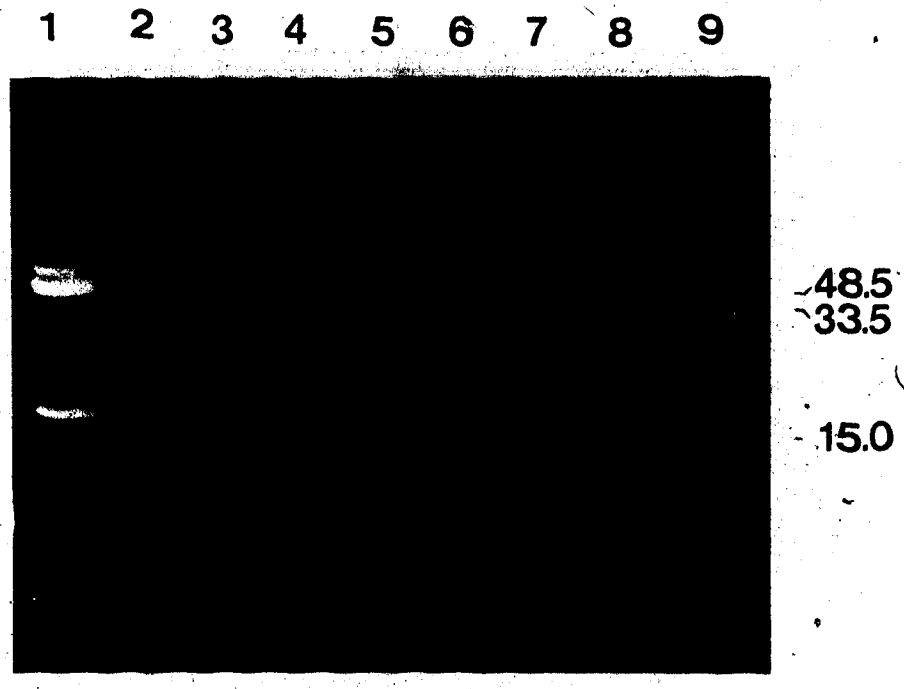
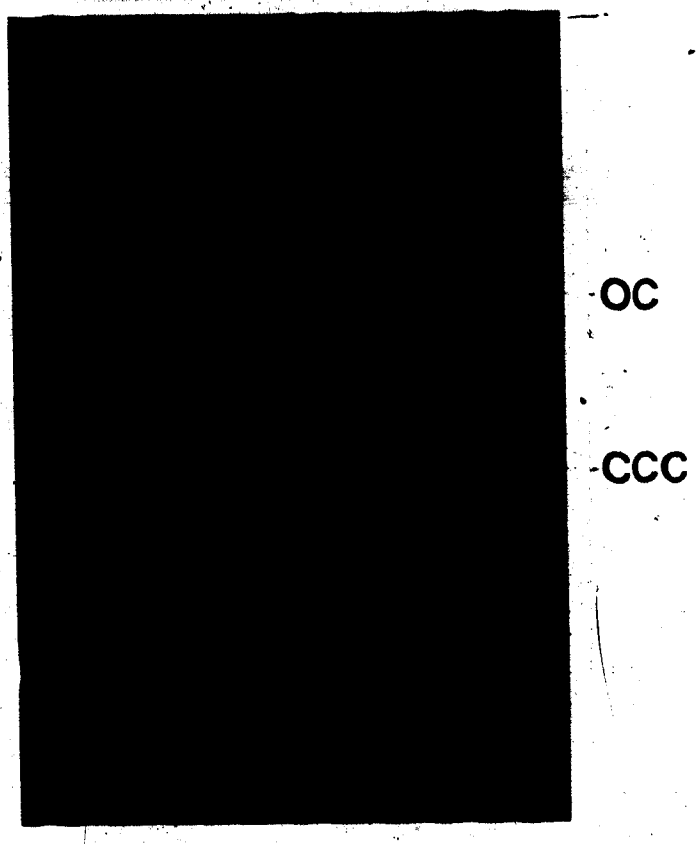

1 2 3 4 5 6 7 8 9
48.5
33.5
15.0

Figure 6. Complete digests of pSa747. Lane 1: pSa747 digested with *Sst* II, Lane 2: pSa747 digested with *Hin* dIII, Lane 3: pSa747 digested with *Eco* RI, Lane 4: pSa747 digested with *Bgl* II, Lane 5: undigested pSa747. In lane 1 a faint band of uncut DNA in the open circular (OC) form is visible above the linearized plasmid band. The *Hin* dIII cut vector in lane 3 appears smaller than the other linearized vectors as pSa747 contains two *Hin* dIII sites, resulting in two fragments. The second fragment is not visible on this gel as it was too small and was run off the bottom. In the uncut pSa747 lane the covalently closed circular form (CCC) and the open circular form (OC) are visible, as well as a faint smear of chromosomal DNA (uppermost band).


1
2
3
4
5
OC
CCC

HB101. The transfected cells were plated and between 10 and 200 colonies per plate were isolated. Significant numbers of colonies were observed only where *Bgl* II-, *Hin* dIII-, and *Mbo* I- digested *A. vinelandii* DNA were cloned into the vector. Since the vector in this case was not treated with alkaline phosphatase, the extent of vector:vector ligation could have been significant. Thus, four colonies from each plate were "pre-screened" by doing small scale plasmid isolations as described in Methods. Figure 7 shows the results of this test. The results showed that the *Hin* dIII clones had the greatest number of putative recombinants, thus efforts were concentrated on further screening using the *Hin* dIII clones.

Two hundred colonies were screened as described in Methods. The screening method used gave poor results (Figure 8). However 12 of the colonies appeared to hybridize with the pSA30 probe and were isolated as described in Methods. These isolates were then electrophoresed on an agarose gel and transferred to nitrocellulose (Southern, 1979), and probed with the *Eco* RI insert of pSA30 (Figure 9). Isolate number 123 was shown to contain DNA homologous to the pSA30 DNA and this isolate was renamed pAvH1. Large scale isolation of pAvH1 was done as described in Methods.

Mapping of pAvH1

An example of an agarose gel used for restriction mapping is shown in Figure 10. The restriction map of pAvH1 is shown on Figure 11. The map was constructed from data obtained from agarose gel electrophoresis of single and multiple digests of the pAvH1 DNA as described in Methods. The fragments of pAvH1 containing DNA homologous to the *K. pneumoniae nif* genes were identified by transferring the DNA from the agarose gels onto nitrocellulose by the method of Southern (1979), followed by hybridization with the pSA30 probe. The region showing homology is indicated on the map (Figure 11).

Subcloning of pAvH1 DNA

From the mapping of the cosmid, the region which showed homology to the *K.*

Figure 7. Small scale isolation of possible clones. A: Lanes 1, 6, 11, and 16: pSa747 DNA, Lanes 2-5: *Hin* dIII isolates, Lanes 7-10: *Mbo* I isolates, Lanes 12-15: *Bgl* II isolates. B: Hybridization of DNA from above gel with ^{32}P-labelled *A. vinelandii* DNA. Blot shows DNA from lanes 4, 5, and 15 to hybridize with the *A.vinelandii* DNA probe.

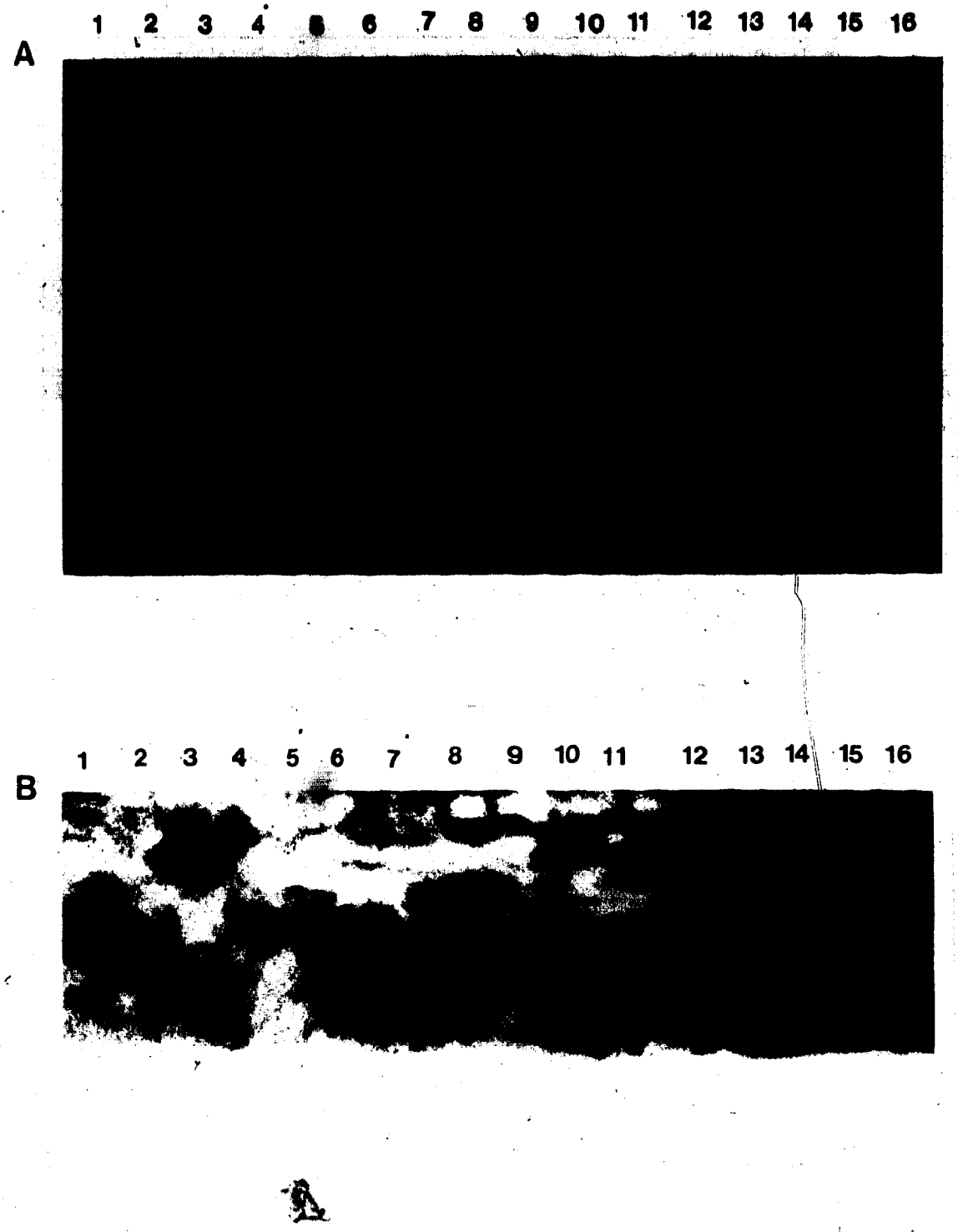
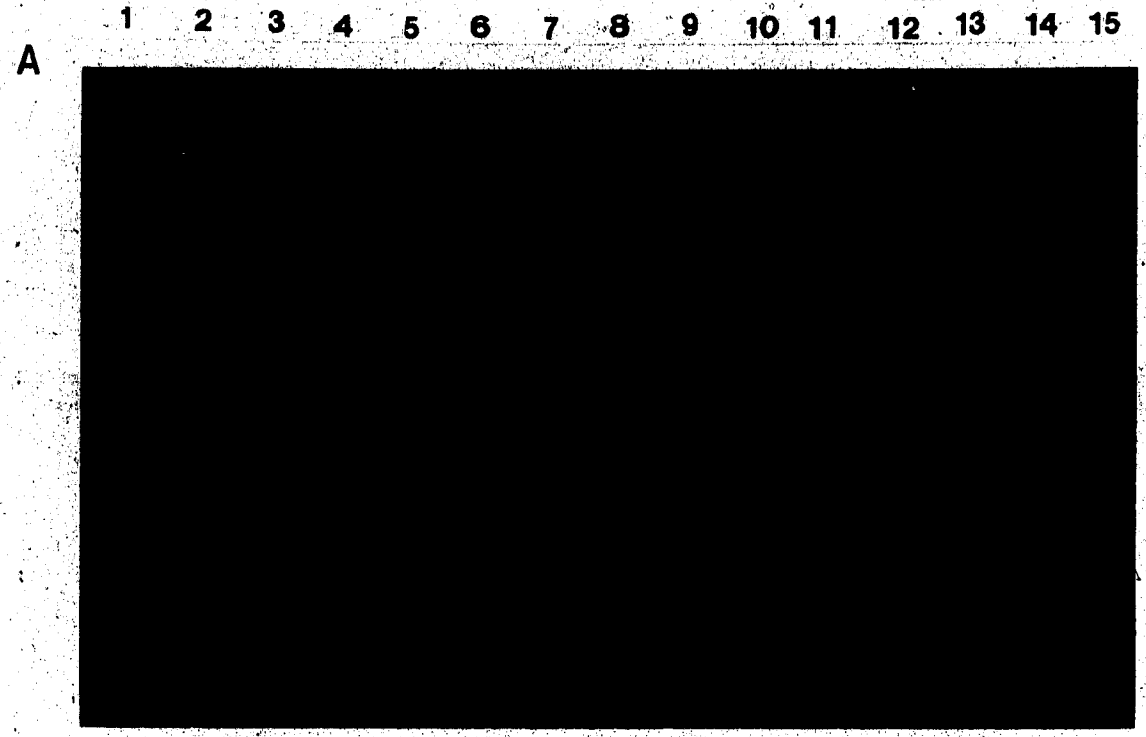

A
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16
B
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16

Figure 8. Colony hybridization of possible clones. Colonies were isolated from master plate from areas corresponding to areas indicated on blot. The probe used was ^{32}P-labelled *Eco* RI fragment of pSA30.

Figure 9. Small scale plasmid isolation of possible positive clones from colony hybridization and hybridization with the *Eco* RI fragment of pSA30. A: Lanes 1, 8, and 15: pSa747 DNA, Lanes 2-7 and 9-14: DNA from isolates. B: Hybridization of DNA from above gel with ^{32}P-labelled *Eco* RI fragment of pSA30. DNA from lane 14 hybridizes with the probe.


A
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15

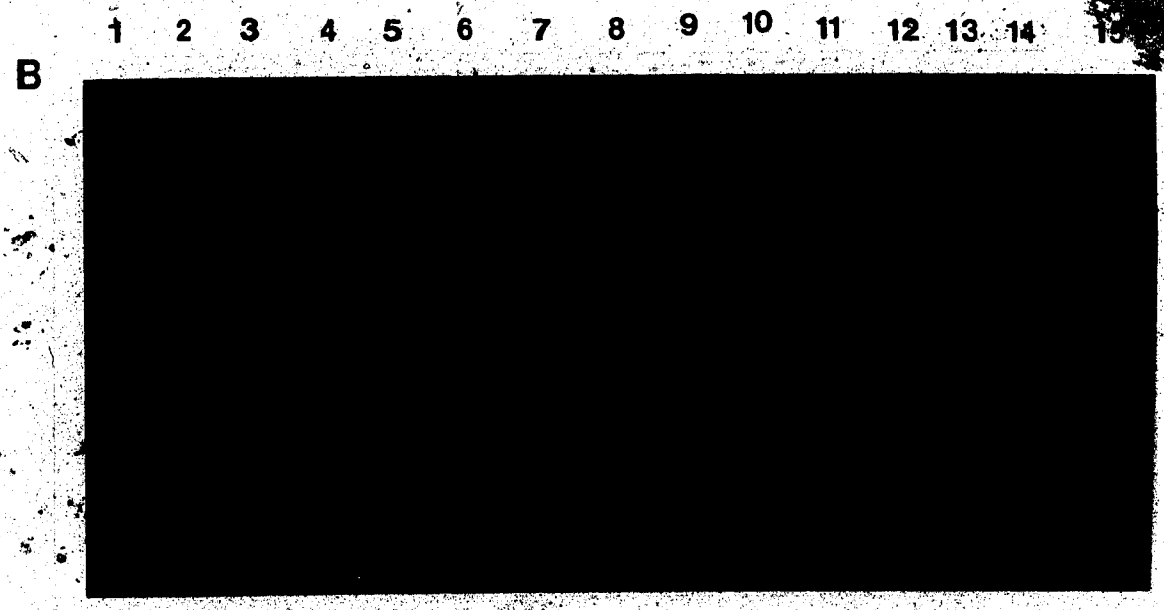

B
1 2 3 4 5 6 7 8 9 10 11 12 13 14

Figure 10. Example of agarose gel (0.35%) used for restriction mapping pAvH1. M-marker, U- uncut pAvH1, Ba- *Bam* HI, Bg- *Bgl* II, Cl- *Cla* I, EcI- *Eco* RI, EcV- *Eco* RV, Hi- *Hin* dIII, Hp- *Hpa* I, Ps- *Pst* I, Pv- *Pvu* II, Sa- *Sal* I, Sm- *Sma* I, Ss- *Sst* I, Xh- *Xho* I. The sizes of the bands in the marker lanes from top to bottom are (in kb) 33.5, 23.1, 15.0, 9.4, 6.6, 4.4, 2.3, and 2.0.

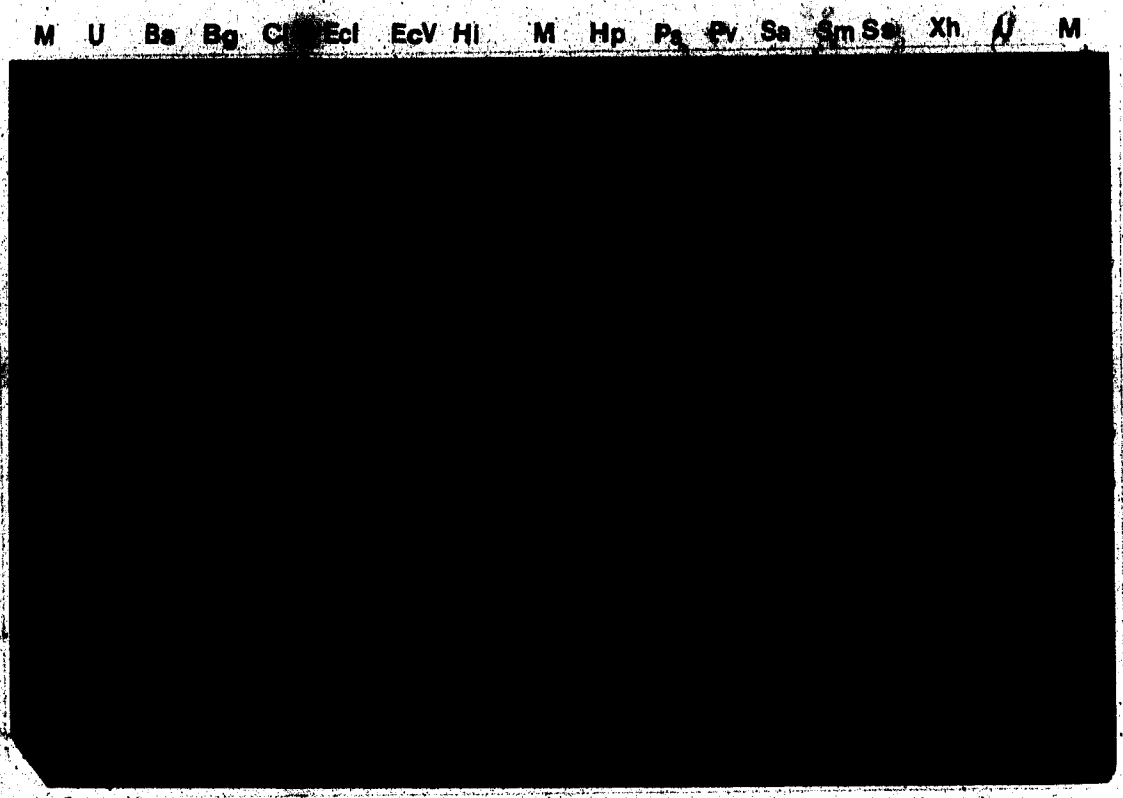
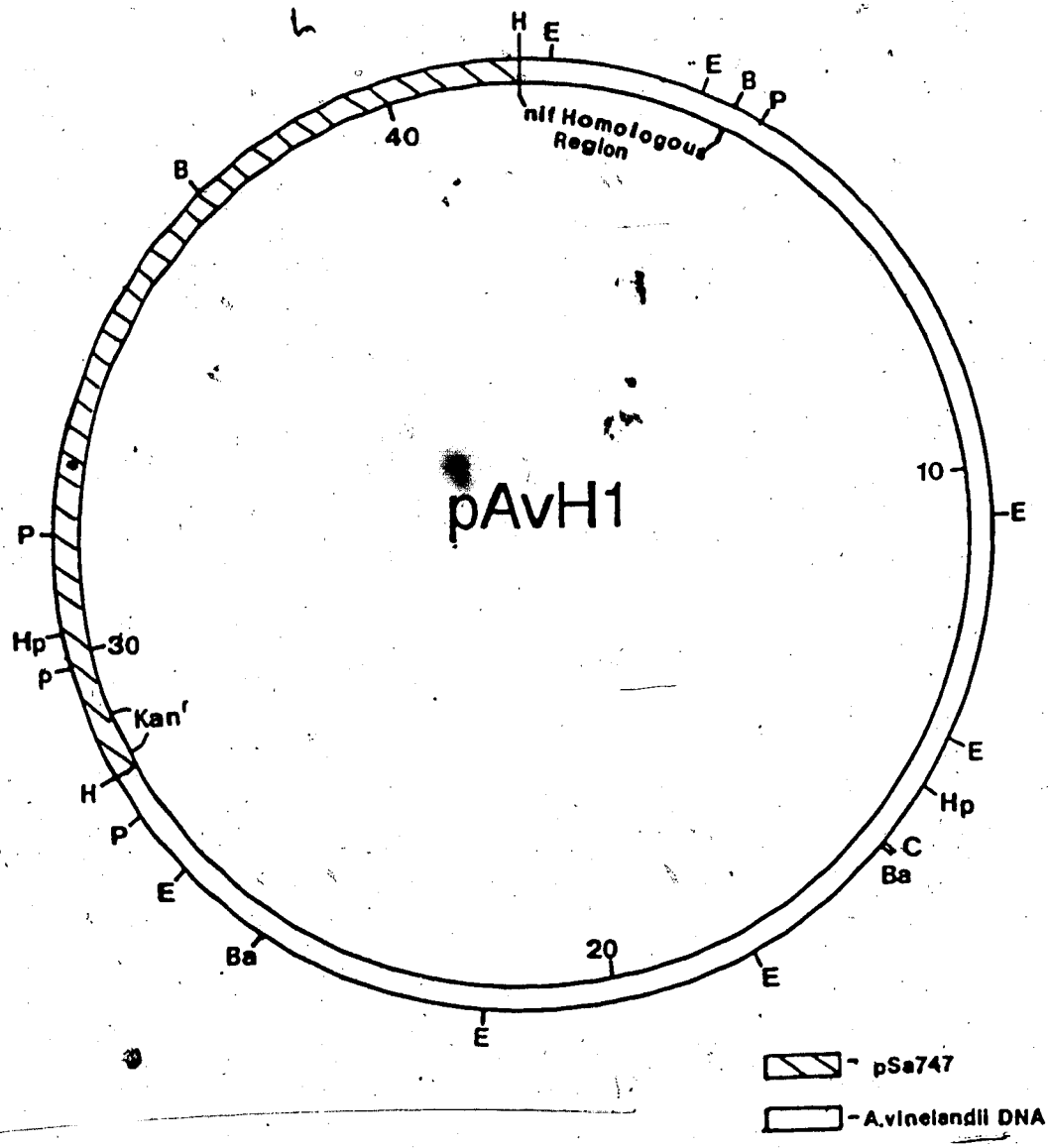

M U Ba Bg Cl Ecl EcV Hi M Hp Pa Pv Sa Sm Se Xh U M

Figure 11. Restriction map of pAvH1. Ba-*Bam* HI, B-*Bgl* II, C- *Cla* I, E- *Eco* RI, H- *Hin* dIII, Hp- *Hpa* I, P- *Pvu* II. The cosmid is 42.6 kb. The vector DNA (pSa747) represents 14.0 kb of the total size of the cosmid. The remaining 28.5 kb is DNA from *A. vinelandii* . The region of the cosmid which hybridizes with the ^{32}P-labelled *Eco* RI fragment from pSA30, containing the *K. pneumoniae nifH,D,K,Y*, and part of *E*, is indicated on the map.

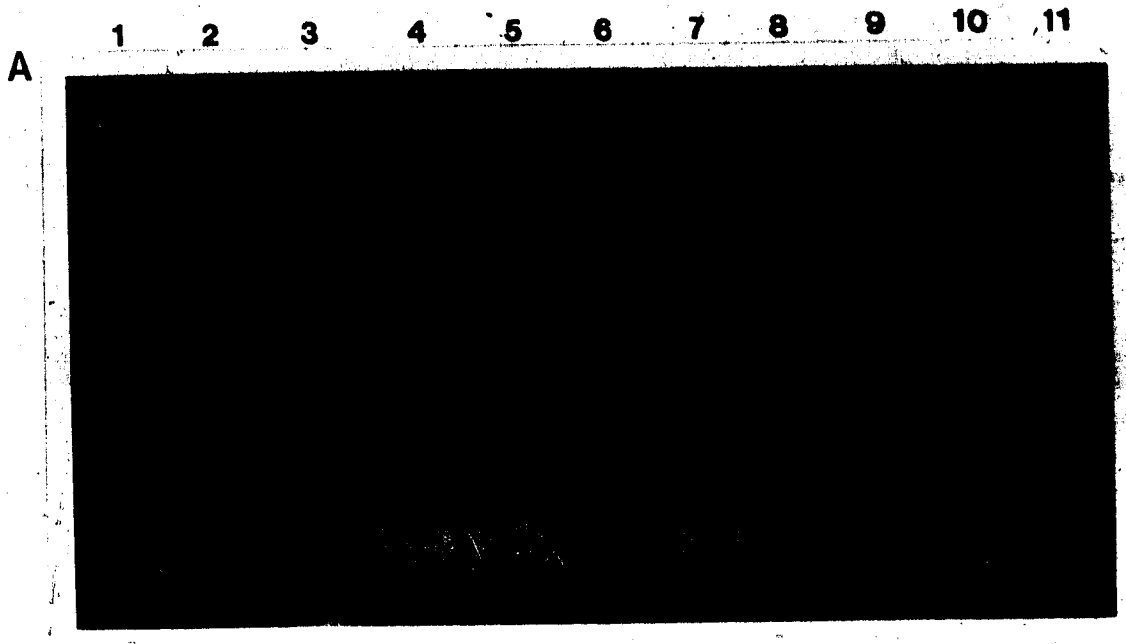

pAvH1
H
E
E
B
P
nif Homologous Region
40
B
10
E
P
Hp
30
P
Kan^r
H
P
E
Ba
20
E
E
Hp
C
Ba
pSa747
A.vinelandii DNA

pneumoniae nif genes was found to lie on a 2.4 kb *Hin* dIII- *Bgl* II fragment. Within this fragment is a 1.4 kb *Eco* RI fragment. These two fragments were chosen for subcloning into pUC vectors (pUC8 for the 1.4 kb *Eco* RI fragment and pUC 13 for the 2.4 kb *Hin* dIII-*Bgl* II fragment).

Nine white colonies were isolated from the 1.4 kb *Eco* RI fragment -pUC8 clones. Quick plasmid assays of these clones showed 3 of them to contain inserts of appropriate size (Figure 12). The DNA was transferred to nitrocellulose and the DNA was hybridized with the pSA30 probe. Figure 12 also shows the results of the hybridization. All clones hybridized with the probe to some extent, including DNA isolated from a blue colony presumably carrying only pUC8 DNA. This result is unexpected as there should be no homology between pUC8 and pSA30. The homology observed may have been a result of weak homology between the vector pUC8 with pACYC184, the vector used for the construction of pSA30. However, the DNA from three clones, clones 36, 108 and 138, hybridized more intensely than the others. One of these clones, clone 138, which hybridized the best with the pSA30 probe, was further tested by digesting the DNA with *Eco* RI and comparing it to *Eco* RI digests of pAvH1 by agarose gel electrophoresis. Figure 13 shows the gel which clearly shows clone 138 to contain the 1.4 kb *Eco* RI fragment of pAvH1. This clone was renamed pAvH1-1. Large-scale isolation of the DNA from this clone was then performed using the cleared lysate method of Clewell & Helinski (1969). The concentration of the plasmid preparation was measured by absorbance at 260 nm and was adjusted to give a final concentration of 1 µg/µl.

Attempts to clone the 2.4 kb *Hin* dIII- *Bgl* II fragment proved unsuccessful. Clones were isolated which gave white colonies with X -gal and IPTG. Quick plasmid preparations were done as described in methods on three colonies, but upon digestion of these clones with *Eco* RI , more bands than expected were observed leading to the conclusion that the 2.4 kb fragment had not been cloned.

Figure 12. Small scale isolation of possible *Eco* RI sub-clones from pAvH1 and hybridization with the *Eco* RI fragment of pSA30. A: Lanes 1 and 11: pUC8 DNA, Lane 2: isolate 25, Lane 3: isolate 36, Lane 4: isolate 108, Lane 5: isolate 109, Lane 6: isolate 114, Lane 7: isolate 138, Lane 8: isolate 139, Lane 9: isolate 148, Lane 10 isolate 152. B: Hybridization of DNA from above gel with ^{32}P-labelled pSA30. All samples including pUC8 hybridize to some extent, indicating pUC8 may share some homology with the vector pACYC184 used as the vector for pSA30. Lanes 3,4,and 7 hybridize the strongest.


A
1 2 3 4 5 6 7 8 9 10 11

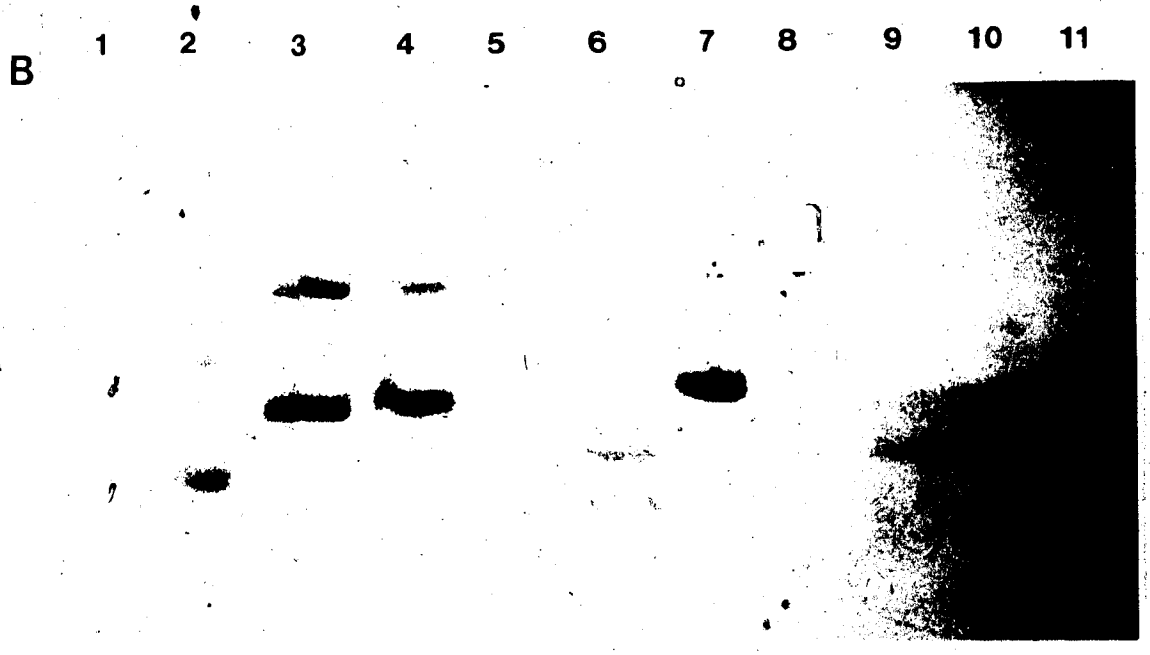
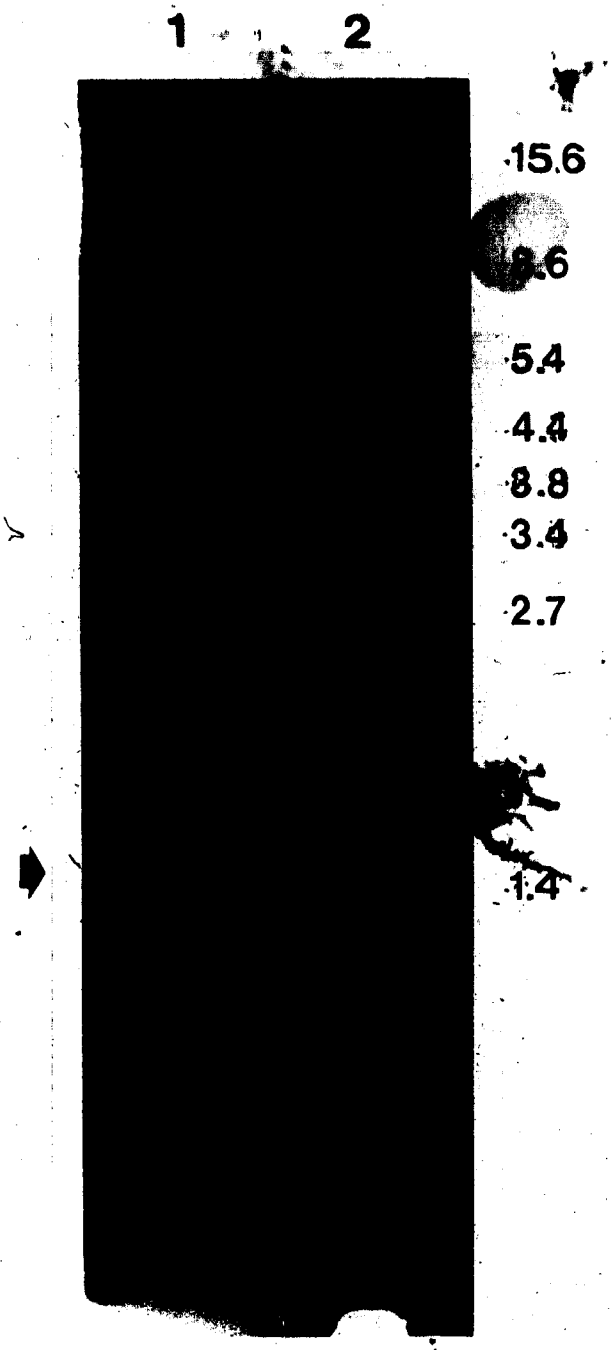

B
1 2 3 4 5 6 7 8 9 10 11

Figure 13. Comparison of *Eco* RI digests of pAvH1 and clone 138. Lane 1: isolate 138 digested with *Eco* RI, Lane 2: pAvH1 digested with *Eco* RI. Vector fragment (pUC8) and insert fragment (arrow) show clearly in Lane 1 and insert band corresponds to the 1.4 Kb fragment of pAvH1. Numbers at right indicate sizes in kb.


1
2
15.6
8.6
5.4
4.4
3.8
3.4
2.7
1.4

Sequencing of the 1.4 *Eco* RI fragment of *A. vinelandii*

Before sequencing was carried out, a restriction enzyme survey of pAvH1-1 was done to find enzymes which would facilitate sequencing using M13 by cutting the DNA to more manageable sizes. DNA from the gels was then transferred to nitrocellulose and the DNA was hybridized with the pSA30 probe. From these results enzymes chosen for use in cloning the *A. vinelandii* DNA into M13 were *Hin* PI, *Msp* I, *Nar* I and *Taq* I. Digests using these enzymes were shotgun cloned into M13 mp9. Figures 14 and 15 show some of the results of dot hybridiztions and direct agarose gel electrophoresis respectively. Following this, C-tests were conducted as described in Methods to obtain recombinants with the DNA cloned representing both strands of DNA. Sequencing reactions were performed as described in Methods. The sequencing gels were then analyzed to give the complete sequence of the 1.4 kb *Eco* RI fragment. Figure 16 shows the sequencing strategy used to obtain the sequence of the fragment. The complete sequence is shown in Figure 17.

The sequence was then entered into the computer for analysis as described in Methods. Using the programs available, restriction endonuclease sites were located and translational analysis of the DNA was carried out to give possible amino acid sequences in all three reading frames from both strands. An open reading frame (ORF) of 1227 bp was found in the sequence beginning 189 bases into the nucleotide sequence. This corresponded to a polypeptide of 409 amino acids in length. No termination signal was found in the sequence. Another ORF of 60 bases starting from the first base was found in the same reading frame as the large ORF. This corresponded to a polypeptide of 20 amino acids. The two open reading frames were separated by 120 bases. Within this region a possible ribosome binding site occurs 14 bases before the first codon of the larger ORF. No

Figure 14. Dot blot hybridization of M13 isolates. ^{32}P-labelled *Eco* RI fragment from pSA30 was used as probe. Position A7 and F7 were negative controls using samples from blue colonies.

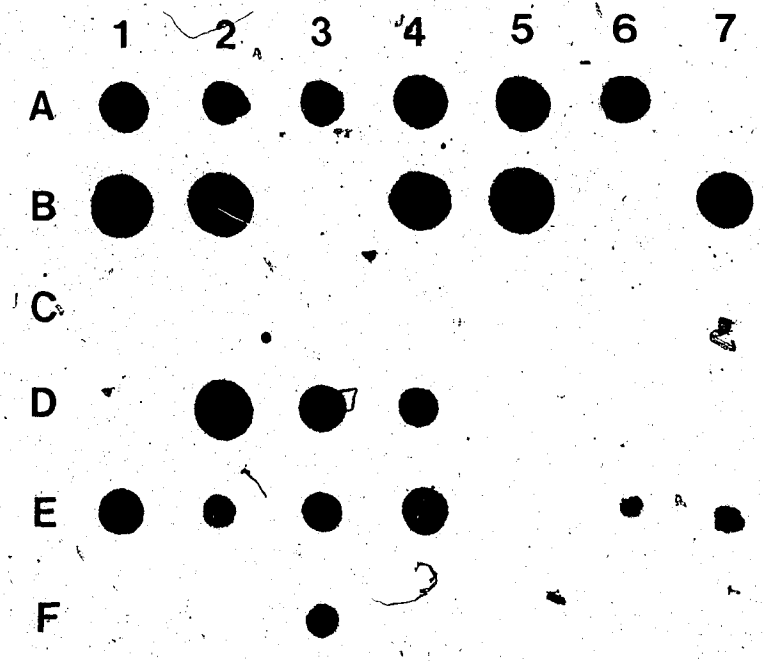
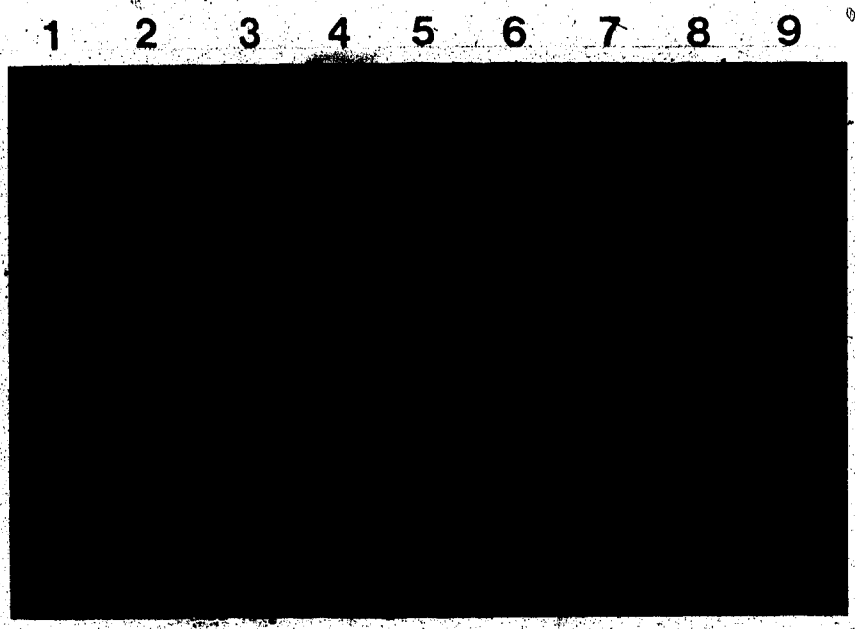

1
2
3
4
5
6
7
A
B
C
D
E
F

Figure 15. Direct[redacted]trophoresis of M13 samples. Lanes 1-8 are from white plaques. Lane 9 [redacted]solated from blue plaque. It is apparent that the samples in lanes 1-8 sho[redacted]obility through the gel than the control (lane 9), indicating that DNA fragments may have been cloned into the M13 vector in these samples. Smaller bands faintly visible in lanes 4, 5 and 7 are spontaneous deletion mutants of the M13 bacteriophage and do not effect the preparations in any way.


1 2 3 4 5 6 7 8 9

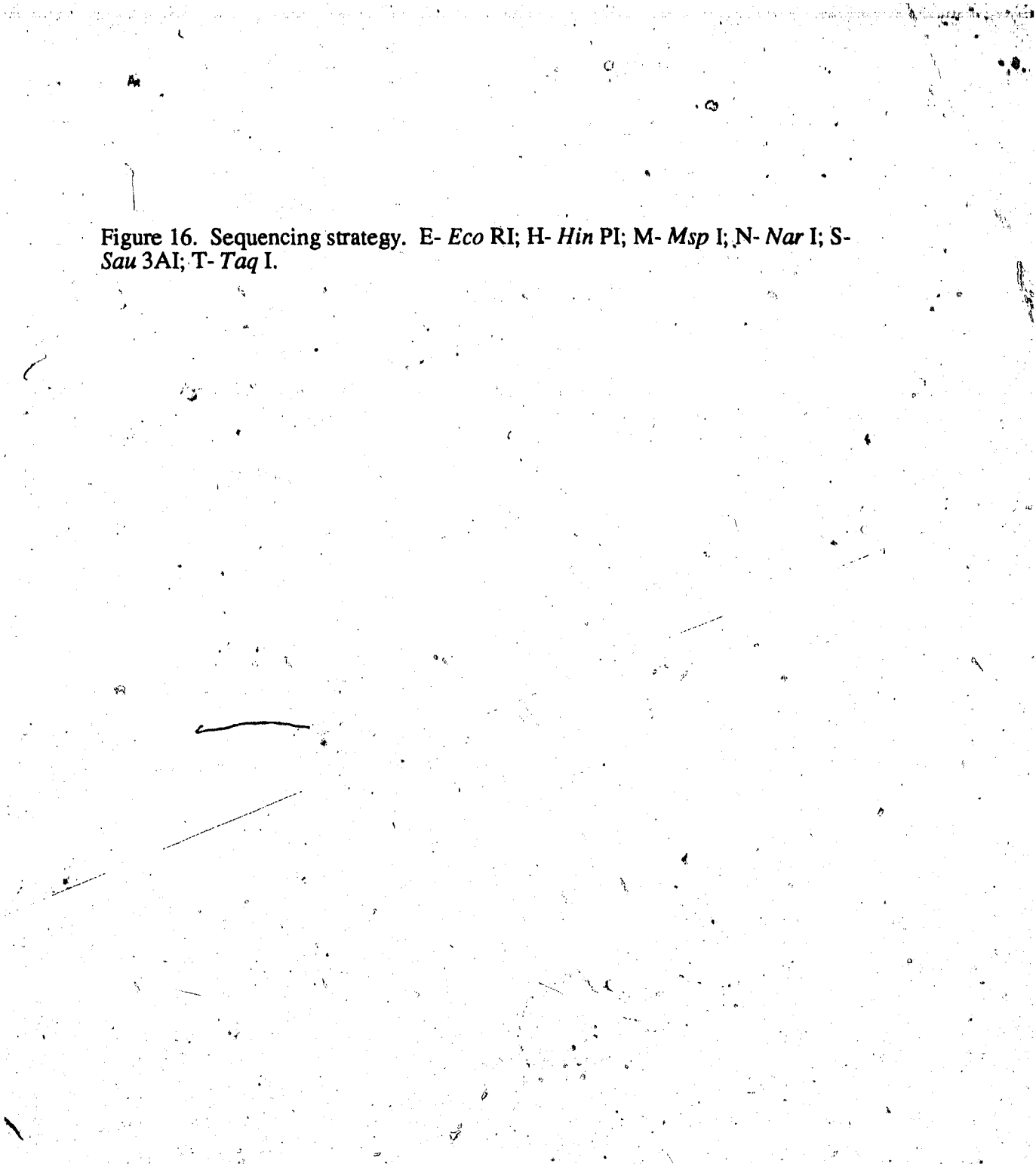


Figure 16. Sequencing strategy. E- *Eco* RI; H- *Hin* PI; M- *Msp* I; N- *Nar* I; S- *Sau* 3AI; T- *Taq* I.

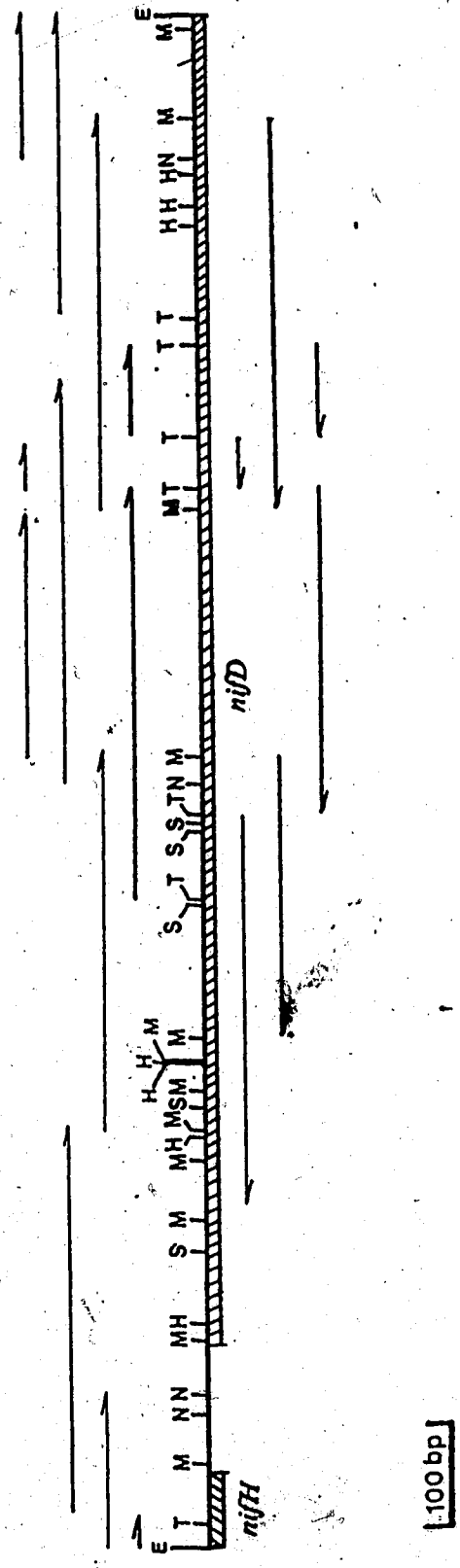

E
T
M
N N
MH
S M
MH
MSM
H H M
M
S
T S S TN M
nifD
M T
T
T T
H H HN M
E M
nifH
100 bp

Figure 17. Complete nucleotide sequence of the 1.4 Kb *Eco* RI fragment of pAvH1.

```
        10          20          30          40          50          60
GAATTCGGCA TCATGGAAGT CGAAGACGAA TCCATCGTCG GCAAAACCGC CGAAGAAGTC

        70          80          90         100         110         120
TGATAGCCGC TCCGGTTTCA GAAGGACGGG ACAGGGCAGA TTGGCTCTGT CGGGGTGGCG

       130         140         150         160         170         180
CCCCCCGCAT TGGGCGGGCG CCCACCCGTT ACCCGCATTA TGAACGCTAA GGCAAGAGGA

       190         200         210         220         230         240
GTCATACCCA TGACCGGTAT GTCGCGCGAA GAGGTTGAAT CCCTCATCCA GGAAGTTCTG

       250         260         270         280         290         300
GAAGTTTATC CCGAGAAGGC TCGCAAGGAT CGTAACAAGC ACCTGGCCGT CAACGACCCG

       310         320         330         340         350         360
GCGGTTACCC AGTCCAAGAA GTGCATCATC TCCAACAAGA AGTCCCAGCC CGGTCTGATG

       370         380         390         400         410         420
ACCATCCGCG GCTGCGCCTA CGCCGGTTCC AAAGGCGTGG TCTGGGGCCC CATCAAGGAC

       430         440         450         460         470         480
ATGATCCACA TCTCCCACGG TCCGGTAGGC TGCGGCCAGT ATTCGCGCGC CGGCCGTCGT

       490         500         510         520         530         540
AACTACTACA TCGGTACCAC CGGTGTGAAC GCCTTCGTCA CCATGAACTT CACCTCGGAC

       550         560         570         580         590         600
TTCCAGGAGA AGGACATCGT GTTCGGTGGC GACAAGAAGC TCGCCAAACT GATCGACGAA

       610         620         630         640         650         660
GTGGAAACCC TGTTCCCGCT GAACAAGGGT ATCTCCGTCC AGTCCGAGTG CCCGATCGGC

       670         680         690         700         710         720
CTGATCGGCG ACGACATCGA ATCCGTGTCC AAGGTCAAGG GCGCCGAGCT CAGCAAGACC

       730         740         750         760         770         780
ATCGTACCGG TCCGTTGCGA AGGCTTCCGC GGCGTTTCCC AGTCCCTGGG CCACCACATC

       790         800         810         820         830         840
GCCAACGACG CAGTCCGCGA CTGGGTCCTG GGCAAGCGTG ACGAAGACAC CACCTTCGCC

       850         860         870         880         890         900
AGCACTCCTT ACGATGTGGC CATCATCGGC GACTACAACA TCGGCGGCGA CGCCTGGTCT

       910         920         930         940         950         960
TCCCGCATCC TGCTGGAAGA AATGGGCCTG CGTTGCGTAG CCCAGTGGTC CGGCGACGGC

       970         980         990        1000        1010        1020
TCCATCTCCG AAATCGAGCT GACCCCGAAG GTCAAGCTGA ACCTGGTTCA CTGCTACCGC

      1030        1040        1050        1060        1070        1080
TCGATGAACT ACATCTCCCG TCACATGGAA GAGAAGTACG GTATCCCATG GATGGAGTAC

      1090        1100        1110        1120        1130        1140
AACTTCTTCG GCCCGACCAA GACCATCGAG TCGCTGCGTG CCATCGCCGC CAAGTTCGAC

      1150        1160        1170        1180        1190        1200
GAGAGCATCC AGAAGAAGTG CGAAGAGGTC ATCGCCAAGT ACAAGCCCGA GTGGGAAGCG

      1210        1220        1230        1240        1250        1260
GTGGTCGCCA AGTACCGTCC GCGCCTGGAA GGCAAGCGCG TCATGCTCTA CATCGGTGGC

      1270        1280        1290        1300        1310        1320
CTGCGTCCGC GCCACGTGAT CGGCGCCTAC GAAGACCTGG GCATGGAAGT GGTGGGTACC

      1330        1340        1350        1360        1370        1380
GGCTACGAGT TCGCCCACAA CGACGACTAT GACCGCACCA TGAAAGAAAT GGGTGACTCC

      1390        1400        1410
ACCCTGCTGT ACGATGACGT GACCGGCTAC GAATTC
```

prom[redacted]r sequences were observed in this region. Figure 18 shows the two open reading frames as well as the intergenic space.

The DNA sequence was compared to the sequence obtained by Brigle *et al.* (1985) and was found to correspond to the last sixty bases of the *nifH* gene and the first 1227 bases of the *nifD* gene of *A. vinelandii*, as well as the intergenic sequence separating the two genes. The sequence differed from that of Brigle *et al.* at 13 bases (Figure 19). No changes in amino acid sequence were observed for the 20 amino acids of the *nifH* gene, however, five changes in amino acid sequence occurred as a result of the differences in the *nifD* DNA sequence. These differences are shown in Figure 20-26.

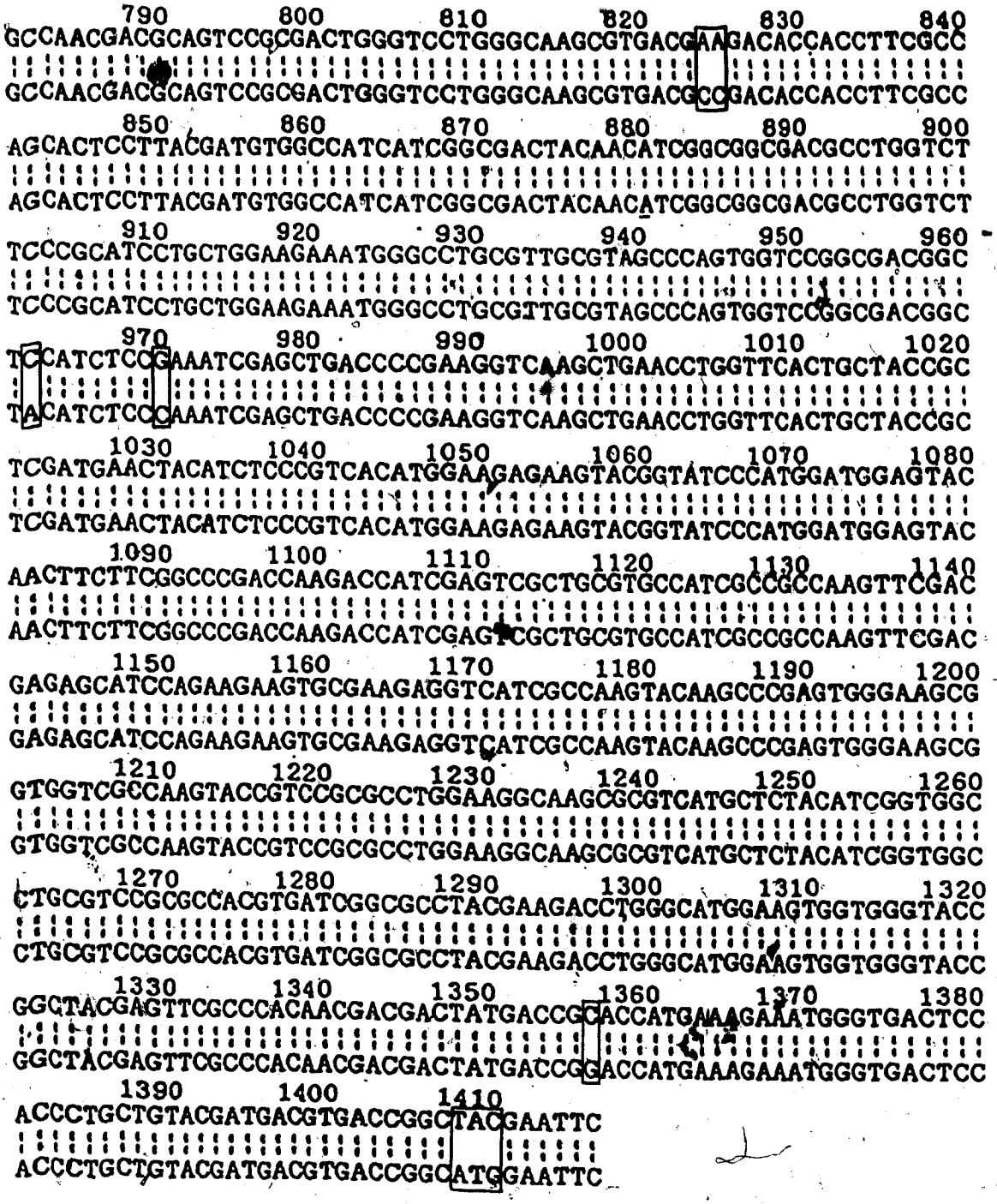
Figure 18. Predicted acid sequence from the 1.4 kb *Eco* RI fragment of pAvH1. Possible ribosome binding site is indicated by underlined DNA sequence.

```
        10         20         30         40         50         60
GAATTCGGCATCATGGAAGTCGAAGACGAATCCATCGTCGGCAAAACCGCCGAAGAAGTC
GluPheGlyIleMetGluValGluAspGluSerIleValGlyLysThrAlaGluGluVal

        70         80         90        100        110        120
TGATAGCCGCTCCGGTTTCAGAAGGACGGGACAGGGCAGATTGGCTCTGTCGGGGTGGCG
EndEnd

       130        140        150        160        170        180
CCCCCCGCATTGGGCGGGCGCCCACCCGTTACCCGCATTATGAACGCTAAGGCAAGAGGA


       190        200        210        220        230        240
GTCATACCCATGACCGGTATGTCGCGCGAAGAGGTTGAATCCCTCATCCAGGAAGTTCTG
          MetThrGlyMetSerArgGluGluValGluSerLeuIleGlnGluValLeu

       250        260        270        280        290        300
GAAGTTTATCCCGAGAAGGCTCGCAAGGATCGTAACAAGCACCTGGCCGTCAACGACCCG
GluValTyrProGluLysAlaArgLysAspArgAsnLysHisLeuAlaValAsnAspPro

       310        320        330        340        350        360
GCGGTTACCCAGTCCAAGAAGTGCATCATCTCCAACAAGAAGTCCCAGCCCGGTCTGATG
AlaValThrGlnSerLysLysCysIleIleSerAsnLysLysSerGlnProGlyLeuMet

       370        380        390        400        410        420
ACCATCCGCGGCTGCGCCTACGCCGGTTCCAAAGGCGTGGTCTGGGGCCCCATCAAGGAC
ThrIleArgGlyCysAlaTyrAlaGlySerLysGlyValValTrpGlyProIleLysAsp

       430        440        450        460        470        480
ATGATCCACATCTCCCACGGTCCGGTAGGCTGCGGCCAGTATTCGCGCGCCGGCCGTCGT
MetIleHisIleSerHisGlyProValGlyCysGlyGlnTyrSerArgAlaGlyArgArg

       490        500        510        520        530        540
AACTACTACATCGGTACCACCGGTGTGAACGCCTTCGTCACCATGAACTTCACCTCGGAC
AsnTyrTyrIleGlyThrThrGlyValAsnAlaPheValThrMetAsnPheThrSerAsp

       550        560        570        580        590        600
TTCCAGGAGAAGGACATCGTGTTCGGTGGCGACAAGAAGCTCGCCAAACTGATCGACGAA
PheGlnGluLysAspIleValPheGlyGlyAspLysLysLeuAlaLysLeuIleAspGlu

       610        620        630        640        650        660
GTGGAAACCCTGTTCCCGCTGAACAAGGGTATCTCCGTCCAGTCCGAGTGCCCGATCGGC
ValGluThrLeuPheProLeuAsnLysGlyIleSerValGlnSerGluCysProIleGly

       670        680        690        700        710        720
CTGATCGGCGACGACATCGAATCCGTGTCCAAGGTCAAGGGCGCCGAGCTCAGCAAGACC
LeuIleGlyAspAspIleGluSerValSerLysValLysGlyAlaGluLeuSerLysThr

       730        740        750        760        770        780
ATCGTACCGGTCCGTTGCGAAGGCTTCCGCGGCGTTTCCCAGTCCCTGGGCCACCACATC
IleValProValArgCysGluGlyPheArgGlyValSerGlnSerLeuGlyHisHisIle
```

```
       790        800        810        820        830        840
GCCAACGACGCAGTCCGCGACTGGGTCCTGGGCAAGCGTGACGAAGACACCACCTTCGCC
AlaAsnAspAlaValArgAspTrpValLeuGlyLysArgAspGluAspThrThrPheAla

       850        860        870        880        890        900
AGCACTCCTTACGATGTGGCCATCATCGGCGACTACAACATCGGCGGCGACGCCTGGTCT
SerThrProTyrAspValAlaIleIleGlyAspTyrAsnIleGlyGlyAspAlaTrpSer

       910        920        930        940        950        960
TCCCGCATCCTGCTGGAAGAAATGGGCCTGCGTTGCGTAGCCCAGTGGTCCGGCGACGGC
SerArgIleLeuLeuGluGluMetGlyLeuArgCysValAlaGlnTrpSerGlyAspGly

       970        980        990       1000       1010       1020
TCCATCTCCGAAATCGAGCTGACCCCGAAGGTCAAGCTGAACCTGGTTCACTGCTACCGC
SerIleSerGluIleGluLeuThrProLysValLysLeuAsnLeuValHisCysTyrArg

      1030       1040       1050       1060       1070       1080
TCGATGAACTACATCTCCCGTCACATGGAAGAGAAGTACGGTATCCCATGGATGGAGTAC
SerMetAsnTyrIleSerArgHisMetGluGluLysTyrGlyIleProTrpMetGluTyr

      1090       1100       1110       1120       1130       1140
AACTTCTTCGGCCCGACCAAGACCATCGAGTCGCTGCGTGCCATCGCCGCCAAGTTCGAC
AsnPhePheGlyProThrLysThrIleGluSerLeuArgAlaIleAlaAlaLysPheAsp

      1150       1160       1170       1180       1190       1200
GAGAGCATCCAGAAGAAGTGCGAAGAGGTCATCGCCAAGTACAAGCCCGAGTGGGAAGCG
GluSerIleGlnLysLysCysGluGluValIleAlaLysTyrLysProGluTrpGluAla

      1210       1220       1230       1240       1250       1260
GTGGTCGCCAAGTACCGTCCGCGCCTGGAAGGCAAGCGCGTCATGCTCTACATCGGTGGC
ValValAlaLysTyrArgProArgLeuGluGlyLysArgValMetLeuTyrIleGlyGly

      1270       1280       1290       1300       1310       1320
CTGCGTCCGCGCCACGTGATCGGCGCCTACGAAGACCTGGGCATGGAAGTGGTGGGTACC
LeuArgProArgHisValIleGlyAlaTyrGluAspLeuGlyMetGluValValGlyThr

      1330       1340       1350       1360       1370       1380
GGCTACGAGTTCGCCCACAACGACGACTATGACCGCACCATGAAAGAAATGGGTGACTCC
GlyTyrGluPheAlaHisAsnAspAspTyrAspArgThrMetLysGluMetGlyAspSer

      1390       1400       1410
ACCCTGCTGTACGATGACGTGACCGGCTACGAATTC
ThrLeuLeuTyrAspAspValThrGlyTyrGluPhe
```

Figure 19. Comparison of DNA sequence to that of Brigle *et al.* Sequence obtained in this study is shown on the upper line while sequence obtained by Brigle *et al.* is shown on the bottom line. Places where sequences differ are indicated by boxes.

```
          10         20         30         40         50         60
GAATTCGGCATCATGGAAGTCGAAGACGAATCCATCGTCGGCAAAACCGCCGAAGAAGTC
||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||||
GAATTCGGTATCATGGAAGTCGAAGACGAATCCATCGTCGGCAAAACCGCCGAAGAAGTC
          70         80         90        100        110        120
TGATAGCCGCTCCGGTTTCAGAAGGACGGGACAGGGCAGATTGCCTCTGTCGGGGTGGCG
|||||||||||||||||||||||||||| |||||||||||||||||||||||||||||||
TGATAGCCGCTCCGGTTTCAGAAGGACTTTACAGGGCAGATTGCCTCTGTCGGGGTGGCG
         130        140        150        160        170        180
CCCCCCGCATTGGGCGGGCGCCCACCCGTTACCCGCATTATGAACGCTAAGGCAAGAGGA
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
CCCCCCGCATTGGGCGGGCGCCCACCCGTTACCCGCATTATGAACGCTAAGGCAAGAGGA
         190        200        210        220        230        240
GTCATACCCATGACCGTATGTCGCGCGAAGAGGTTGAATCCCTCATCCAGGAAGTTCTG
|||||||||||||| |||||||||||||||||||||||||||||||||||||||||||||
GTCATACCCATGACCGTATGTCGCGCGAAGAGGTTGAATCCCTCATCCAGGAAGTTCTG
         250        260        270        280        290        300
GAAGTTTATCCCGAGAAGGCTCGCAAGGATCGTAACAAGCACCTGGCCGTCAACGACCCG
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
GAAGTTTATCCCGAGAAGGCTCGCAAGGATCGTAACAAGCACCTGGCCGTCAACGACCCG
         310        320        330        340        350        360
GCGGTTACCCAGTCCAAGAAGTGCATCATCTCCAACAAGAAGTCCCAGCCCGGTCTGATG
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
GCGGTTACCCAGTCCAAGAAGTGCATCATCTCCAACAAGAAGTCCCAGCCCGGTCTGATG
         370        380        390        400        410        420
ACCATCCGCGGCTGCGCCTACGCCGGTTCCAAAGGCGTGGTCTGGGGCCCCATCAAGGAC
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
ACCATCCGCGGCTGCGCCTACGCCGGTTCCAAAGGCGTGGTCTGGGGCCCCATCAAGGAC
         430        440        450        460        470        480
ATGATCCACATCTCCCACGGTCCGGTAGGCTGCGGCCAGTATTCGCGCGCCGGCCGTCGT
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
ATGATCCACATCTCCCACGGTCCGGTAGGCTGCGGCCAGTATTCGCGCGCCGGCCGTCGT
         490        500        510        520        530        540
AACTACTACATCGGTACCACCGGTGTGAACGCCTTCGTCACCATGAACTTCACCTCGGAC
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
AACTACTACATCGGTACCACCGGTGTGAACGCCTTCGTCACCATGAACTTCACCTCGGAC
         550        560        570        580        590        600
TTCCAGGAGAAGGACATCGTGTTCGGTGGCGACAAGAAGCTCGCCAAACTGATCGACGAA
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
TTCCAGGAGAAGGACATCGTGTTCGGTGGCGACAAGAAGCTCGCCAAACTGATCGACGAA
         610        620        630        640        650        660
GTGGAAACCCTGTTCCCGCTGAACAAGGGTATCTCCGTCCAGTCCGAGTGCCCGATCGGC
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
GTGGAAACCCTGTTCCCGCTGAACAAGGGTATCTCCGTCCAGTCCGAGTGCCCGATCGGC
         670        680        690        700        710        720
CTGATCGGCGACGACATCGAATCCGTGTCCAAGGTCAAGGGCGCCGAGCTCAGCAAGACC
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
CTGATCGGCGACGACATCGAATCCGTGTCCAAGGTCAAGGGCGCCGAGCTCAGCAAGACC
         730        740        750        760        770        780
ATCGTACCGGTCCGTTGCGAAGGCTTCCGCGGCGTTTCCCAGTCCCTGGGCCACCACATC
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
ATCGTACCGGTCCGTTGCGAAGGCTTCCGCGGCGTTTCCCAGTCCCTGGGCCACCACATC
```


790 800 810 820 830 840
GCCAACGACGCAGTCCGCGACTGGGTCCTGGGCAAGCGTGACGAAGACACCACCTTCGCC
GCCAACGACGCAGTCCGCGACTGGGTCCTGGGCAAGCGTGACGCCGACACCACCTTCGCC
850 860 870 880 890 900
AGCACTCCTTACGATGTGGCCATCATCGGCGACTACAACATCGGCGGCGACGCCTGGTCT
AGCACTCCTTACGATGTGGCCATCATCGGCGACTACAACATCGGCGGCGACGCCTGGTCT
910 920 930 940 950 960
TCCCGCATCCTGCTGGAAGAAATGGGCCTGCGTTGCGTAGCCCAGTGGTCCGGCGACGGC
TCCCGCATCCTGCTGGAAGAAATGGGCCTGCGTTGCGTAGCCCAGTGGTCCGGCGACGGC
970 980 990 1000 1010 1020
TCCATCTCCGAAATCGAGCTGACCCCGAAGGTCAAGCTGAACCTGGTTCACTGCTACCGC
TACATCTCCCAAATCGAGCTGACCCCGAAGGTCAAGCTGAACCTGGTTCACTGCTACCGC
1030 1040 1050 1060 1070 1080
TCGATGAACTACATCTCCCGTCACATGGAAGAGAAGTACGGTATCCCATGGATGGAGTAC
TCGATGAACTACATCTCCCGTCACATGGAAGAGAAGTACGGTATCCCATGGATGGAGTAC
1090 1100 1110 1120 1130 1140
AACTTCTTCGGCCCGACCAAGACCATCGAGTCGCTGCGTGCCATCGCCGCCAAGTTCGAC
AACTTCTTCGGCCCGACCAAGACCATCGAGTCGCTGCGTGCCATCGCCGCCAAGTTCGAC
1150 1160 1170 1180 1190 1200
GAGAGCATCCAGAAGAAGTGCGAAGAGGTCATCGCCAAGTACAAGCCCGAGTGGGAAGCG
GAGAGCATCCAGAAGAAGTGCGAAGAGGTCATCGCCAAGTACAAGCCCGAGTGGGAAGCG
1210 1220 1230 1240 1250 1260
GTGGTCGCCAAGTACCGTCCGCGCCTGGAAGGCAAGCGCGTCATGCTCTACATCGGTGGC
GTGGTCGCCAAGTACCGTCCGCGCCTGGAAGGCAAGCGCGTCATGCTCTACATCGGTGGC
1270 1280 1290 1300 1310 1320
CTGCGTCCGCGCCACGTGATCGGCGCCTACGAAGACCTGGGCATGGAAGTGGTGGGTACC
CTGCGTCCGCGCCACGTGATCGGCGCCTACGAAGACCTGGGCATGGAAGTGGTGGGTACC
1330 1340 1350 1360 1370 1380
GGCTACGAGTTCGCCCACAACGACGACTATGACCGCACCATGAAAGAAATGGGTGACTCC
GGCTACGAGTTCGCCCACAACGACGACTATGACCGGACCATGAAAGAAATGGGTGACTCC
1390 1400 1410
ACCCTGCTGTACGATGACGTGACCGGCTACGAATTC
ACCCTGCTGTACGATGACGTGACCGGCATGGAATTC

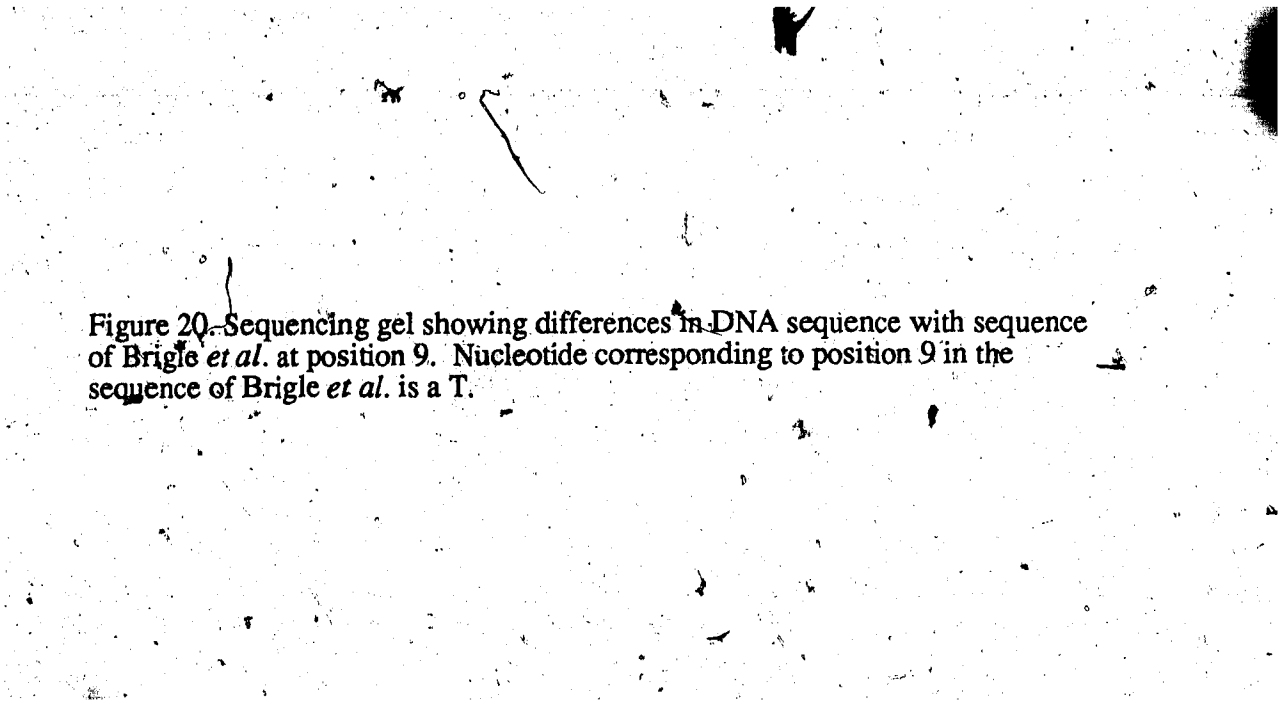


Figure 20. Sequencing gel showing differences in DNA sequence with sequence of Brigle *et al.* at position 9. Nucleotide corresponding to position 9 in the sequence of Brigle *et al.* is a T.

Figure 21. Sequencing gel showing differences in DNA sequence with sequence of Brigle *et al.* at positions 88-90. Nucleotides corresponding to position 88-90 in the sequence of Brigle *et al.* are TTT.

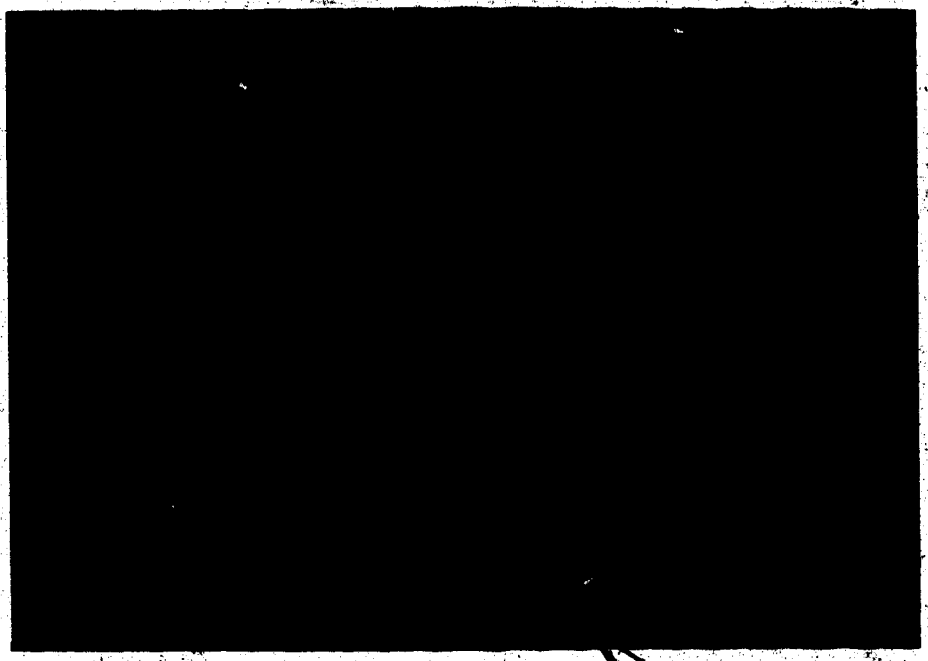
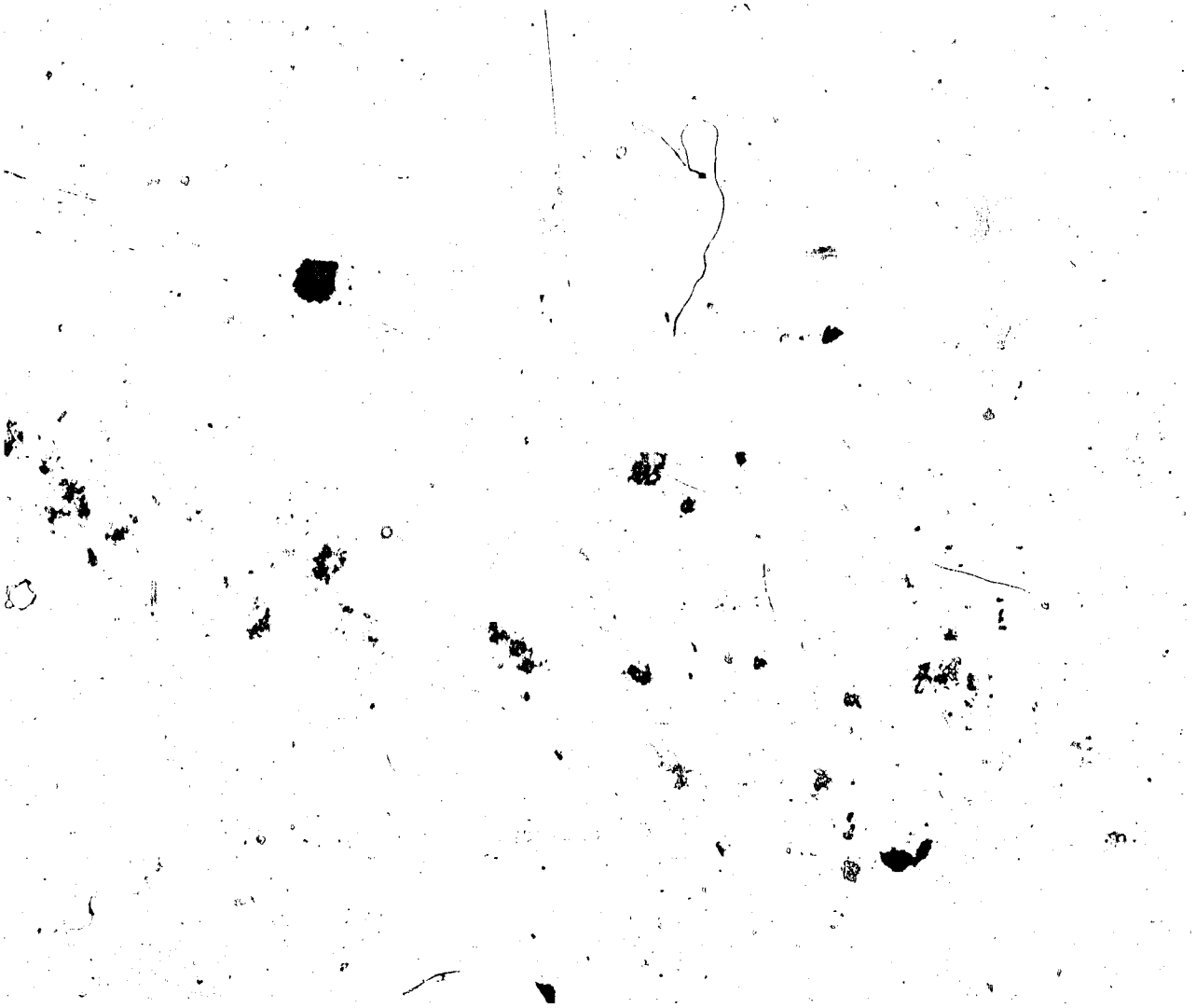
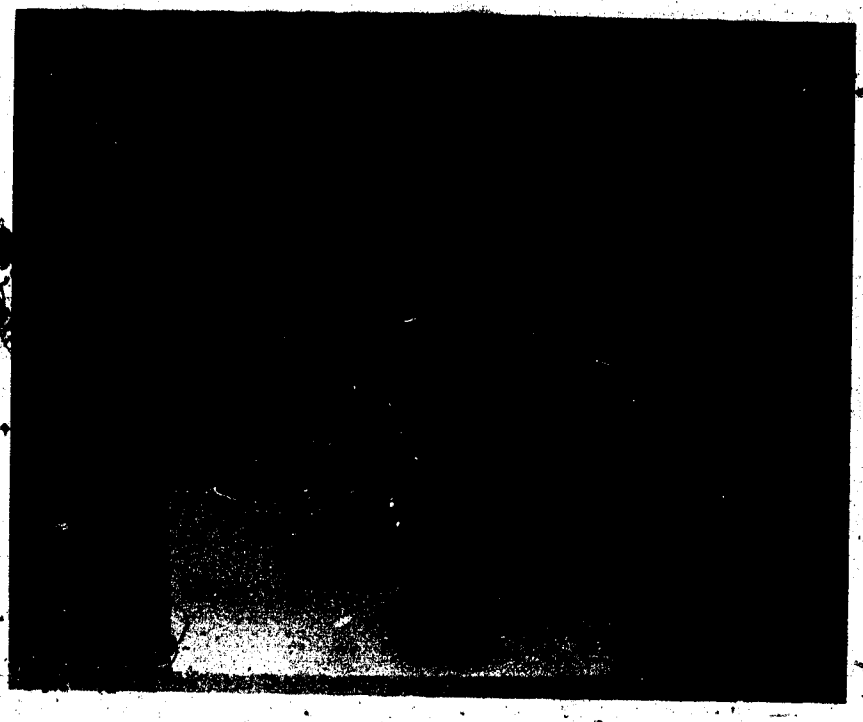


Figure 22. Sequencing gel showing differences in DNA sequence with sequence of Brigle *et al.* at position 196. Nucleotide corresponding to position 196 in the sequence of Brigle *et al.* is a C. Base 200 is indicated with an arrow.





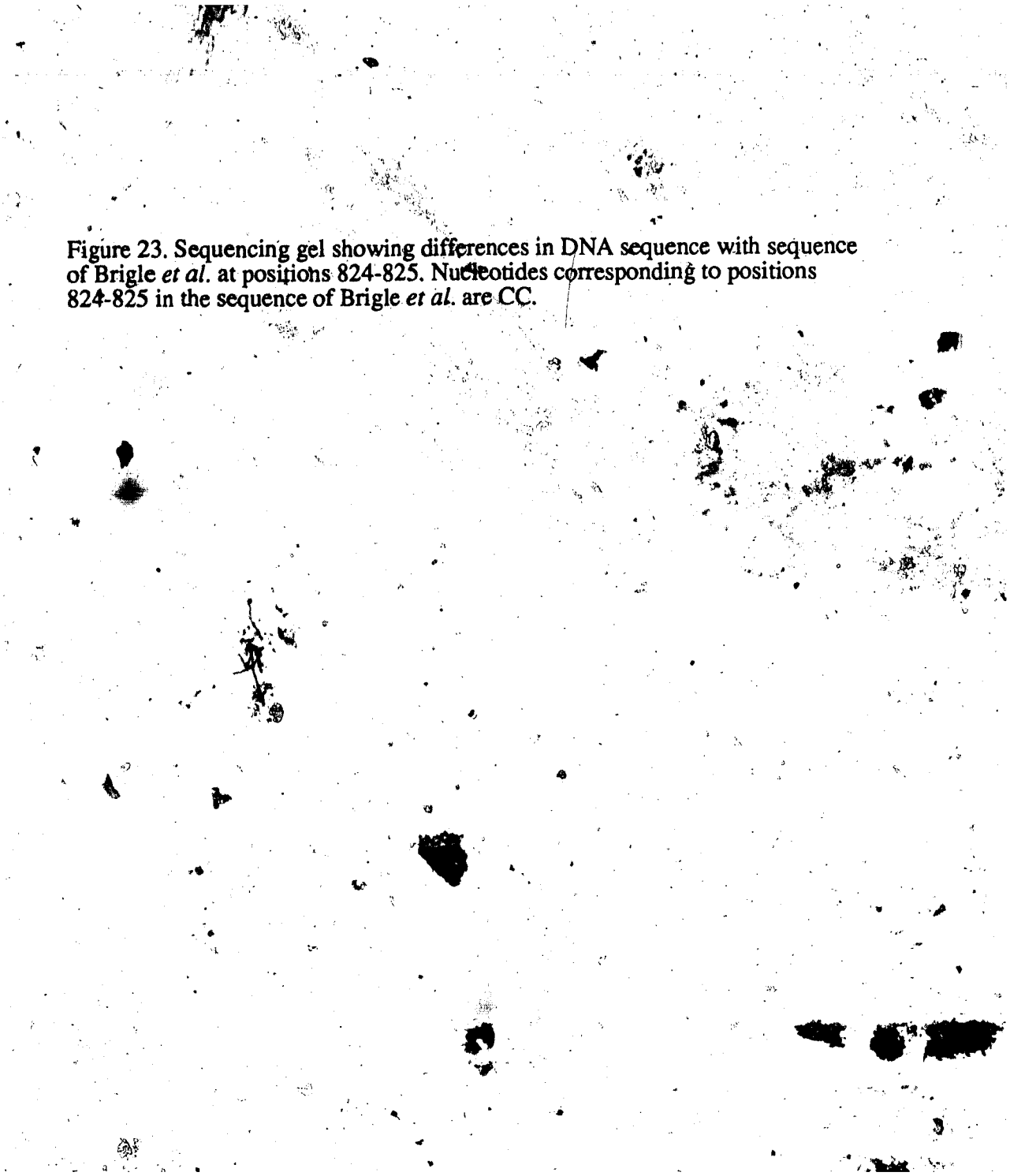


Figure 23. Sequencing gel showing differences in DNA sequence with sequence of Brigle *et al.* at positions 824-825. Nucleotides corresponding to positions 824-825 in the sequence of Brigle *et al.* are CC.

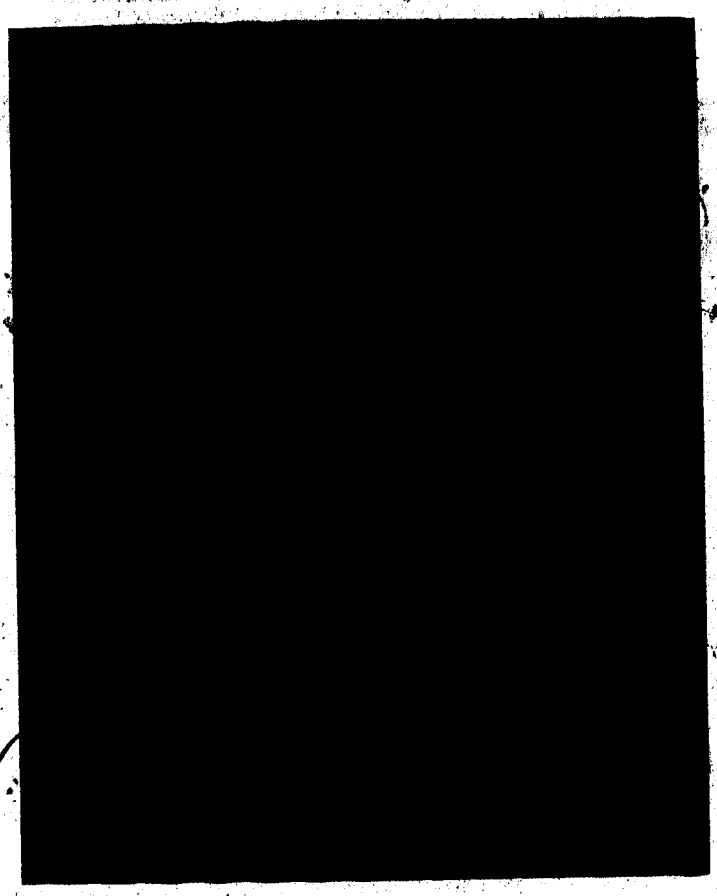


Figure 24. Sequencing gel showing differences in DNA sequence with sequence of Brigle *et al.* at positions 962 and 970. Nucleotides corresponding to positions 962 and 970 in the sequence of Brigle *et al.* are A and C respectively.

Figure 25. Sequencing gel showing differences in DNA sequence with sequence of Brigle *et al.* at position 1356. Nucleotide corresponding tp position 1356 in the sequence of Brigle *et al.* is a G. Base 1350 is indicated by an arrow.

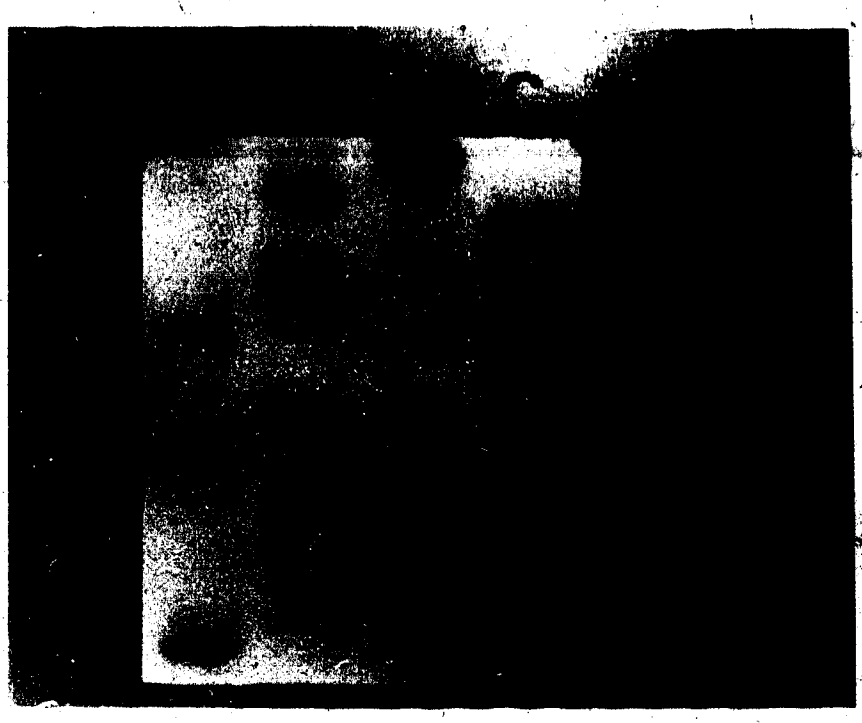
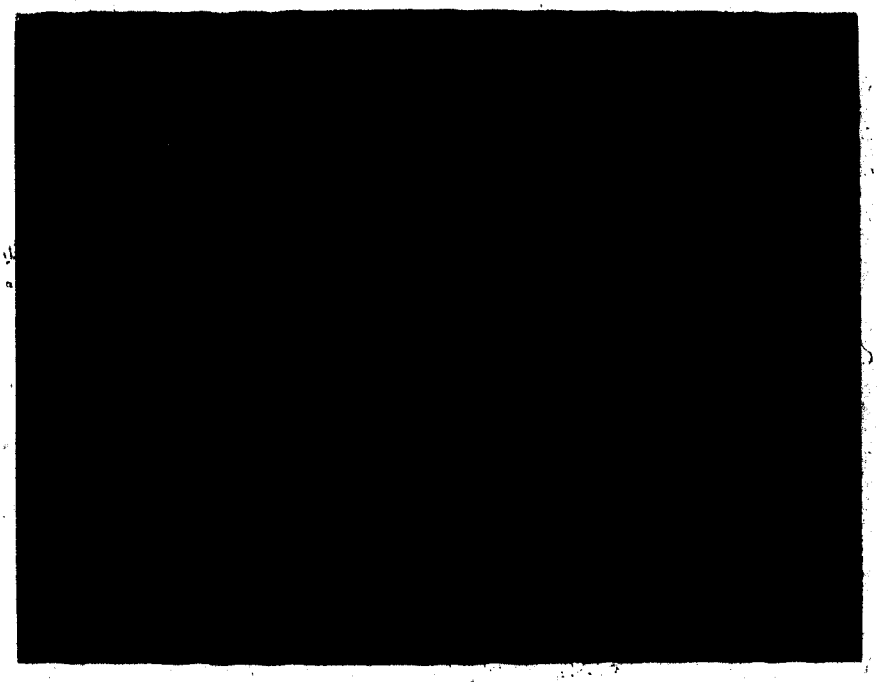


Figure 26. Sequencing gel showing differences in DNA sequence with sequence of Brigle *et al.* at positions 1408-1410. Nucleotides corresponding to position 1408-1410 in the sequence of Brigle *et al.* are ATG.



DISCUSSION

A genomic library of DNA fragments from the *A. vinelandii* genome was successfully produced using the cosmid vector pSa747. One of the clones, pAvH1, contained DNA homologous to the *nif* genes of *K. pneumoniae* contained in the plasmid pSA30. The region of homology to the *nif* probe was narrowed down to a 2.4 kb *Hin* dIII-*Bgl* II fragment. The 1.4 kb *Eco* RI fragment within this region was subcloned into the plasmid pUC8 and the complete nucleotide sequence of this fragment was determined. This subclone was shown by Doran *et al.*(1987) to complement mutations in the *nifD* gene of *A. vinelandii* strain UW10. From the nucleotide sequence one large open reading frame (ORF) was found coding for a polypeptide of 409 amino acids. No termination codon was found indicating the ORF extended beyond the Eco RI fragment. Another small ORF coding for a polypeptide of 20 amino acids in length was found preceding the large ORF in the same reading frame. Two termination codons occurred at the end of the smaller ORF.

The nucleotide and the predicted amino acid sequences obtained in this thesis were compared to the sequences obtained by Brigle *et al.*(1985). A substantial amount of the work reported here was conducted prior to the publication of the results of Brigle *et al.*(1985). From this comparison, it appears that the sequence obtained in this thesis corresponds to a region coding for the last 20 amino acids of the *A. vinelandii nifH* gene and the first 409 amino acids of the *nifD* gene as well as the intergenic space between these two genes.

By comparison of the DNA sequence of Brigle *et al.*(1985) with the restriction map of the cosmid clone pAvH1, it appears that pAvH1 contains the entire *nifH* gene, as well as most of the *nifD* gene. If the arrangement of the *nif* region in *A. vinelandii* is similar to the arrangement in *K. pneumoniae* it is possible that the *nifJ* gene of *A. vinelandii* is also contained in pAvH1.

The nucleotide sequence obtained in this thesis was found to differ from the sequence

obtained by Brigle *et al.*(1985) at 13 nucleotides. This was confirmed by sequencing gels at these points (Figures 20-26 in Results). Of these differences in nucleotide sequence, one falls within the *nifH* sequence, three fall within the intergenic region and ten fall within the *nifD* sequence.

The amino acid sequence of the *nifH* sequence is identical to the last 20 amino acids in the sequence obtained by Brigle *et al.*(1985), while there are 5 amino acids which are different in the *nifD* amino acid sequence as a result of the differences in the nucleotide sequence.

The predicted amino acid sequence obtained in this thesis for the *nifD* gene product of *A. vinelandii* was then compared with the amino acid sequences from the *nifH* and *nifD* genes of a number of other nitrogen fixing organisms. Figures 27 and 28 show the comparison of the *A. vinelandii* amino acid sequences predicted here with the sequences from other organisms for the *nifH* and *nifD* genes respectively. It is interesting to note that the amino acid sequence obtained in this thesis for the *A.vinelandii* *nifD* gene shows a greater degree of conservation of amino acid sequences than the sequence obtained by Brigle *et al.*(1985) at the points where these two sequences differ (Figure 28).

In comparing the amino acid sequences of these proteins, it is expected that some regions involved in the function of the proteins would be conserved. Other regions, such as the regions involved in the interaction between subunits in the nitrogenase complex, may not be as well conserved. The regions most likely to be conserved are the regions involved in the binding of [4Fe-4S] clusters and, in the case of the *nifD* protein, the binding of the FeMo cofactor. By looking at the homologies which exist between analogous proteins from different nitrogen fixing organisms some of the functions of the proteins may be localized on the protein molecules. The regions of interaction between subunits may also be localized by correlating the regions of homology with the results of the experiments done by Emerich and Burris (1978), who observed the activity of

Figure 27. Comparison of Carboxy-terminal ends of *nifH* proteins. A.v.-*Azotobacter vinelandii* ; this thesis, Brigle *et al.* (1985); K.p.-*Klebsiella pneumoniae* , Scott *et al.* (1981); An. -*Anabaena* 7120, Mevarech et al (1980); R.m.-*Rhizobium meliloti*, Török and Kondorosi (1981); R. j. -*Rhizobium japonicum* , Fuhrmann and Hennecke (1984); R. T.-*Rhizobium trifolii* , Scott *et al.* (1983*a*); T.f. -*Thiobacillus ferrooxidans,* Pretorius *et al.* (1987); M.v. *Methanococcus voltae*, Souillard and Sibold (1986); C.p.-*Clostridium pasteurianum* ; Tanaka *et al.* (1977). Hyphens have been inserted to maximize homology.

A.v. -EFGIME VEDES - - - - IV GKT - - - A EE V-COOH
K.p. -EFGI ME EEDTS - - - - I I GKT - - A A EE N A A-COOH
An. -EYGLLD - DDT KHS - EI I GKP - - - A E A T N R S C R N-COOH
R.m. -DFGI MKS - DEQMLAELHAKE - - - A K - V I A P H-COOH
R.j. -EHGI KAV - DES - - - - I I GKT - - A A EL - A A S-COOH
R.t. -DFGI MKS - DEQMLEELL AKEVQAA - - V - A P-COOH
M.v. -ENGLD ELT - ES - I EELVR RKYE-COOH
T.f. -DFGI MQKEDTS - - - - I I GKT - - A A E L A A A G M-COOH
C.p. -Q Y G L M D L-COOH

Figure 28. Comparison of acid sequences from *nifD* proteins. Conserved cysteine residues are indicated with an arrow. Residues which differ between the acid sequence obtained in this thesis and the results of Brigle *et al.* (1985) are indicated in boxes. A.v.- *Azotobacter vinelandii*, This thesis; A.v.b- *Azotobacter vinelandii*, Brigle *et al.* (1985); R.p.- *Rhizobium Parasponia*, Scott *et al.* (1983*b*); R.j.- *Rhizobium japonicum* , Kaluza & Hennecke (1984); An.- *Anabaena*, Lammers & Haselkorn (1983); C.p.- *Clostridium pasteurianum*, Hase *et al.* (1984); T.f.-*Thiobacillus ferrooxidans*, Pretorius *et al.*; (1987); K.p.-*Klebsiella pneumoniae*, Scott *et al.* (1981); R.t.- *Rhizobium trifolii*, Scott *et al.* (1983*a*); R.m.-*Rhizobium meliloti*,Török & Kondorosi (1981). Hyphens have been added to maximize homology.

```
                                        10                     20          30
A.v.                         MT[G]MSREE VESLIQEVLE VYPEKAR KDRNK
A.v.b                        MT[R]MSREE VESLIQEVLE VYPEKAR KDRNK
R.p.          MSLATTQS I AEI RARNKELIEEVLKVYPEKTA KRVRK
R.j.          MSLATTNS V AEI RARNKELIEEVLKVYPEKTA KRRAK
An                  MT PPENKNLVDENKELIQEMLKAYPEKS RKKREK
C.p.                            SENLKDEI LEKYI PKTKKT RSG
T.f.   MSMISAEDLSTQPQRR KLPE I AELIDE TLKAYPEKPAKRRAK
K.p.                    MMTNATGERNLALIQEVLEVFPET ARK ERRK
R.t.                 MSLDYENDGDLHARLIDEVLSQYPDKT AKRRKK
R.m.                 MSLDYENDNALHEKLIEEA I AL- SRQGAKRRKK

                 40                          50             60
A.v.   HL AVNDPA VTQSKK- - - - - -   CI I SNKKSQPG LMTIRGCA
A.v.b  HL AVNDPA VTQSKK- - - - - -   CI I SNKKSQPG LMTIRGCA
R.p.   HL NVHQAAGKSDCG- - - - - - - -VKSN I KSI PGVMT IRGCA
R.j.   HL NVHQ -AGKSDCG- - - - - - - -VKSNI KS IPGVMT IRGCA
An.    HL NVH - EENKSDCG- - - - - - - -VKSNI KSVPGVMTARGCA
C.p.   H I V I KT EETPNPE - - - - - - - - - I VANTRTVPGI I TARGCA
T.f.   HLN VY - EEGKSECD - - - - - - -C -KSNI KSVPGVMT IRGCA
K.p.   HMMVSDPKMKSVGK- - - - - - -C I I SNRKSQPGVMTVRGCA
R.t.   H LGVAKGRE ALEQGSDALCETGVKSN IKS-I PGVMTVRGCA
R.m.   H LS VAKNKQETAEEGQVVSE -CDK SNI

           70           80              90              100
A.v.   YAGSKGVVW GPIKDM IHISHGPVGCGQYS RAG RR - -NYY I
A.v.b  YAGSKGVVW GPIKDM IHISHGPVGCGQYS RAG RR - -NYY I
R.p.   YAGSKGVVW GPIKDMVHISHGPVGCGQYSWGS RR - -NYYV
R.j.   YAGSKGVVW GPIKDMVHISHGPVGCGQYSWGS R - - -NYYV
An.    YAGSKGVVW GPIKDMVHISHGPVGCGYWSWSGRR - -NYYV
C.p.   YAGSKGVVM GPIKDMVHITHGP I GCSF YTWGGRRFKSKPE
T.f.   YAGSYGVVW SPVKDMI HISHGPVGCGHYARA GRR - -AYYI
K.p.   YAGSKGVV F GPIKDMAHISHGPAGCGQYS RA ERR - -NYYT
R.t.   YAGSKGVVW GPIKDMAHISHGPAGCGHYSWS QRR - -NYYV

            110             120              130              140
A.v.   GTTGVNAFV - TMNFTSDFQEKDIVFGGDKKLAKLIDEVET LF
A.v.b  GTTGVNAFV - TMNFTSDFQEKDIVFGGDKKLAKLIDEVET LF
R.p.   GTTGVDS FVCT LQFTSDFQEKDIVFGGDKKL I KVLDEI QE LF
R.j.   GTTG IDS FV - T LQFTSDFQEKDIVFGGDKKLDKI LDEI QE LF
An.    GVTG INS FG -TMHFTSDFQERDIVFGGDKKLT KLI EELDV LF
C.p.   DGTGLN -FN - EYVFSTDMQESDIVFGGVNKLKDAI HEAYEMF
T.f.   GTTGVDT YT -TMHFTSDFQVKDIVFGGDKKLALMDELE -E LP
K.p.   GVSGVDS FG - TLNFTSDF QERDIVFGGDKKLSKLIEEMELLF
R.t.   GLTGVDA FV - TMQFTSDGQEKDIVFGGDKKLL

            150            160             170                180
A.v.   PLNKGISVQSECPI GLIGDDIES VSKVKGAELS - KT IV PVRCE
A.v.b  PLNKGISVQSECPI GLIGDDIES VSKVKGAELS - KT IV PVRCE
R.p.   PLNNGIT IQSE CPI GLIGDDIEAVSRSKSKEYGGKT IV PVRCE
R.j.   PLNNGIT IQSGCPVGLIGDDIEAVSRAKSKEYGGKT IV PVRCE
An.    PLNRGVSIQS ECRI GSIGDDIEAVAKKTS KQI G - - PVV PLRCE
C.p.   -HPA AIGVYATCPVGLIGDDILAVAATAS KEI G -I PVHAFS CE
T.f.   PMSKGITVQSE CPI GLIGDDIEAVFKKKAA EPG -KPVVPNRCE
K.p.   PLTKG IT IQSGCPVGLIGDDISAVANASS KGLD -KPV I PVRCE
```

```
          190                 200                 210                 220
A.v. GFRGVSQSLGHHIANDAVRDWVLGKR-DEDTTFAS-----TP
A.v.bGFRGVSQSLGHHIANDAVRDWVLGKR-DADTTFAS-----TP
R.p. GFRGVSRSLGHHIANDAVRL-IF---DKLE-PEGPKFQPTP
R.j. GFRGVSQSLGHHIANDAVRDWIFGHIEAEGK----PKFGPTP
An.  GFRGVSQSLGHHIANDAIRKWIFPEY-DKLKKETRLDFEPSP
C.p. GYKGVSQSAGHHIANNTVMTDIIGKGNKEEKK----------
T.f. GFRGVSQSLGHHIANDSIRDWVLDPAADK
K.p. GFRGVSQSLGHHIANDVVRDWI

                    230                 240                 250                 260
A.v. YDVAIIGDYNIGGDAWSSRILLEEMGLRCVAQWSGDGSIS
A.v.b YDVAIIGDYNIGGDAWSSRILLEEMGLRCVAQWSGDGMIS
R.p. YDVAIIGDYNIGGDAWSSRILLEEMGLRVIAQWSGDGSLA
R.j. YDVAIIGDYNIGGDAWSSRILLEEMGLRVIAQWSGDGSLA
An.  YDVALIGDYNIGGDAWASRMLLEEMGLRVVAQWSGDGTLN
C.p. YSINVLGEYNIGGDAWEMDRVLEKIGYHVNATLTGDATY-

                    270                 280                 290                 300
A.v. EIELTP-KVKLNLVHCYRSMNYISRHMEEKYGIPWMEYNFF
A.v.bQIELTP-KVKLNLVHCYRSMNYISRHMEEKYGIPWMEYNFF
R.p. ELEATP-KAKLNILHCYRSMNYISRHMEEKFGIPWCEYNFF
R.j. ELEATP-KAKLNILHCYRSMNYISRHMEEKFGIPWCEYNFF
An.  ELIQGP-AAKLVLIHCYRSMNYICRSLEEQYGMPWMEFNFF
C.p. EKVQNADKADLNLVQCHRSINYIAEMMETKYGIPWIKCNFI

                    310                 320                 330                 340
A.v. GPTKTIESLRAIAAKFDESIQKKC-EEVIAKYKPEWEAVVAK
A.v.b GPTKTIESLRAIAAKFDESIQKKC-EEVIAKYKPEWEAVVAK
R.p. GPSKIAESLRKIAGYFDD-KIKEGAERVIEKYQPLVDAVIAK
R.j. GPSKI-DSLRRIAGYFDD-KIKEGAERVIQKYQPLVDAVIAK
An.  GPTKIAASLREIAAKFDS-KIQENAEKVIAKYTPVMNAVLDK
C.p. GVNGIVETLRDMAKCFDDPELTKRTEEVIAEEIAAIQDDLDY

                    350                 360                 370                 380
A.v. YRPRLEGKRVMLYIGGLRPRHVIGAYEDLGMEVVGTGYEFA
A.v.b YRPRLEGKRVMLYIGGLRPRHVIGAYEDLGMEVVGTGYEFA
R.p. YRPRLEGKTVMLYVGGLRPRHVIGAYEDLGMEVVGTGYEFG
R.j. YRPRLEGKTVMLYVGGLRPRHVIGAYEDLGMDVIGTGYEFG
An.  YRPRLEGNTVMLYVGGLRPRHVVPAFEDLGIKVVGTGYEFA
C.p. FKEKLQGKTACLYVGGSRS-HTYM-LKSFGVDSLVAGFEFA

          390
A.v. HNDDYDRTMKEMG----------------------------
A.v.b HNDDYDRTMKEMG----------------------------
R.p. HNDDYQRTAQHYVK----------------------------
R.j. HNDDYQRTAEHYVK----------------------------
An.  HNDDYKRTTHYIDNA--------------------------
C.p. HRDDYEGREVIPTIKIDADSKNIPEITVTPDEQKRVVIPEDKV

                              400
A.v. ----------------------DSTLLYDDVTGYEF...
A.v.b ----------------------DSTLLYDDVTGMEF...
R.p. ----------------------DSTLIYDDVNGYEF...
R.j. ----------------------DSTLIYDDVNGYEF...
An.  ----------------------TIIYDDVTAYEF...
C.p. EELKKAGVPLSSTGGMMKEMHDGTILIDDMNHMEV...
```

nitrogenase complexes constructed from *nif* proteins from heterologous sources.The relative amount of homology found between amino acid sequences compared was assigned a numerical value and was calculated as :

$$\text{Relative Homology} = \frac{\text{\# of amino acids matching}}{\text{\# of amino acids matching + \# of amino acid mismatching}}$$

Tables 5 and [illegible] calculated values for the relative homologies calculated for the *nifH* and *nifD* genes respectively.

The relative homologies for the *nifH* gene proteins calculated were in general quite low. However, given that only the last 20 amino acids were compared, this is likely not an accurate indication of the relative homologies of the entire proteins, and other studies of the *nifH* gene protein have shown a high degree of conservation.

A more reasonable estimation of relative homologies can be made from the data obtained for the *nifD* protein sequence. For this protein it appears that a great deal of conservation of amino acid sequence occurs, as most comparisons gave values of 0.6 of greater. The *nifD* protein sequence of *C. pasteurianum* showed the least amount of conservation of sequence compared with the other protein sequences. It was also considerably larger than the other proteins compared and contained a unique string of approximately 50 amino acids not found in the proteins of the other organisms compared. It is possible that this region could be sensitive to oxygen and thus was not conserved during the course of evolution as nitrogen fixation moved out into aerobic and microaerophillic environments.

It is also possible that this region maybe involved in the subunit interactions within the nitrogenase complex of *C. pasteurianum*. Results of Emerich and Burris (1978) showed when nitrogenase components from *C. pasteurianum* were combined with components from other nitrogen fixing organisms very little activity was observed. This indicated that little homology existed between nitrogen fixing proteins of *C. pasteurianum* and other

Table 5. Relative Homology of *nif* H protein sequences

	A.v.	K.p.	An.	R.m.	R.j.	M.v.	T.f	R.t.	C.p.
A.v.	-	0.667	0.276	0.310	0.565	0.305	0.520	0.321	0.286
K.p.		-	0.300	0.296	0.609	0.222	0.720	0.241	0.286
An.			-	0.185	0.391	0.133	0.286	0.207	0.286
R.m.				-	0.259	0.148	0.276	0.724	0.286
R.j.					-	0.148	0.520	0.250	0.143
M.v.						-	0.067	0.200	0.100
T.f.							-	0.400	0.286
R.t.								-	0.286

Values were calculated as: $\frac{\text{\#amino acids matching}}{\text{\# amino acids matching + \# amino acids mismatching}}$

A.v. -*Azotobacter vinelandii* ; K.p. -*Klebsiella pneumoniae* ; An. -*Anabaena* 7120; R.m. -*Rhizobium meliloti* ; R.j. -*Rhizobium japonicum* M.v. -*Methanococcus voltae* ; T.f.-*Thiobacillus ferrooxidans* ; R.t. -*Rhizobium trifolii*.

Table 6. Relative Homology of *nif* D protein sequences

	A.v.	R.p.	R.j.	An.	C.p.	T.f.	K.p.	R.t.	R.m.
A.v.	-	0.702	0.695	0.623	0.350	0.647	0.694	0.583	0.250
R.p.		-	0.928	0.675	0.363	0.602	0.643	0.618	0.254
R.j.			-	0.688	0.360	0.614	0.650	0.618	0.266
An.				-	0.331	0.637	0.651	0.610	0.237
C.p.					-	0.382	0.441	0.387	0.136
T.f.						-	0.541	0.544	0.232
K.p.							-	0.566	0.193
R.t.								-	0.483

Values were calculated as: $\frac{\text{\#amino acids matching}}{\text{\# amino acids matching + \# amino acids mismatching}}$

A.v. -*Azotobacter vinelandii* ; R.p. -*Rhizobium* sp. isolated from *Paraspoonia andersonii* ; R.j. - *Rhizobium japonicum* ; An. - *Anabaena* 7120; C.p. -*Clostridium pasteurianum* ; T.f.-*Thiobacillus ferrooxidans* ; K.p. -*Klebsiella pneumoniae* ; R.t. - *Rhizobium trifolii* ; R.m. - *Rhizobium meliloti*

nitrogen fixers, at least where subunit interactions were concerned. From this we may speculate that the extra 50 amino acids in the *C. pasteurianum nifD* protein are involved in the interaction of the subunits in the nitrogenase complex of this organism, or that this region at least interferes with the formation of an active complex when components from other organisms were used.

The greatest amount of homology was observed between a *Rhizobium* sp. isolated from *Parasponia andersonii* and *R. japonicum*.. The relative homology observed between these two organisms was 0.928. The *nifD* proteins for these two organisms did not show significant homology with the proteins of the other two *Rhizobium* sp. compared. However, the data for these others were incomplete.

The amino acid sequence obtained in this thesis for *A. vinelandii* exhibited significant homology with the amino acid sequences of *K. pneumoniae*, a *Rhizobium* sp. isolated from *Parasponia andersonii* , and *R. japonicum* (relative homologies of around 0.7 for all three). A slightly lesser amount of homology was observed with the amino acid sequences of *R. trifolii* and *Anabaena* 7120. The least homology was observed with the amino acid sequences of *C. pasteurianum and R. meliloti*.. The latter amino acid sequence was incomplete, thus the comparison is likely not accurate.

Overall some regions definitely show a greater degree of homology than others. As mentioned earlier regions which were expected to be conserved were regions involved with the binding of [4Fe-4S] clusters and the binding of the FeMo cofactor. The amino acid cysteine is one of the most likely candidates in the role of [4Fe-4S] binding. For this reason cysteine residues and the regions containing them would be predicted to be highly conserved if in fact the cysteine residues were involved in this function. Five cysteine residues occur in highly conserved regions in the amino acid sequences compared. These residues are indicated on Figure 28. The conservation of sequence around the cysteine residues is expected as the positioning of the residues in the protein molecule is important

if it is to function as a ligand. It also appears that the cysteine residues occur in the amino terminal half of the protein molecule.

The role which these residues play in the function of the protein is not clear. However, in the case of *A. vinelandii*, Brigle *et al.* (1985) suggest that three of these residues (62, 88,and 183; see Figure 28) may be involved in the binding of the [4Fe-4S] clusters as they showed similarities to the proposed [4Fe-4S] binding sites of the *nifH* protein of *A.vinelandii*. Brigle *et al.* (1985) also suggest that the region from amino acid 264 to 289 may be involved in the binding of the FeMo cofactor . This region shows significant differences from regions thought to be involved in the binding of [4Fe-4S] clusters while at the same time exhibiting characteristics favourable for the binding of the FeMo cofactor . A high degree of conservation was also observed in analogous regions of the other *nifD* proteins compared.

Conclusion

The results obtained in this thesis show the sequence of a 1.4 Kb region of DNA from the nitrogen-fixation gene cluster of *A. vinelandii*. This fragment contains the carboxy-terminal portion of the *nifH* gene and the first 1227 bases of the *nifD* gene, as well as the intergenic space between these two genes. The sequence obtained here agrees with the sequence obtained by Brigle *et al.* (1985) with the exception of 13bases. These 13 differences result in 5 amino acid differences in the *nifD* protein sequence predicted from the DNA sequence. These differences appear to be minor, although the sequence obtained here shows a greater degree of homology, at the positions where these two sequences differ, with the protein sequences determined for the *nifD* genes of a number of other organisms (Figure 28).

The predicted amino acid sequences determined in this thesis were also compared with the protein sequences of a number of other organisms and showed a significant amount of homology. Certain regions implicated in the binding of the [4Fe-4S] clusters showed the

greatest degree of homology between all of the proteins compared. This would indicate that the acquisition of the ability to fix atomospheric nitrogen occurred fairly recently in the course of evolution. *C. pasteurianum* showed the least amount of conservation of sequence and may represent an older 'version' of the nitrogen fixation ability. This is logical owing to the sensitivity of the nitrogenase complex to oxygen. One would thus expect the nitrogen fixing system to develop in anaerobic conditions.

The objectives of this thesis have been achieved and it has been shown here that the nitrogen fixation system of *A. vinelandii* is fairly closely related to the systems occurring in other nitrogen fixing organisms.

References

Alveraz-Morales, A., and H. Hennecke. 1985. Expression of *Rhizobium japonicum nifH* and *nifDK* operons can be activated by the *Klebsiella pneumoniae nifA* protein but not by the product of the *ntrC*. Mol. Gen. Genet. **199**:306-314.

Avtges, P., P.A. Scolnik, and R. Haselkorn. 1983. Genetic and physical map of the structural genes (*nifH,D,K*) coding for the nitrogenase complex of *Rhodopseudomonas capsulata*. J. Bacteriol. **156**:251-256.

Beynon, J., A. Ally, M.Cannion, F. Cannon, M. Jacobson, V. Cash, and D. Dean. 1987. Comparative organization of nitrogen fixation-specific genes from*Azotobacter vinelandii* and *K. pneumoniae*: DNA sequence of the *nifUSV* genes. J. Bacteriol. **169**:4024-4029.

Birnboim, H.C. and J. Doly. 1979. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucl. Acids Res. **7**:1513-1523.

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.

Bishop, P.E., T.M. Rizzo, and K.F. Bott. 1985. Molecular cloning of *nif* DNA from *Azotobacter vinelandii*. J. Bacteriol.**162**:21-28

Boyer, H.W. and D. Roulland-Dussoix. 1969. A complementation analysis of restriction and modification of DNA in *Escherichia coli*. J. Mol. Biol. **41**:459.

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.

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.

deBruijn, F.J. and F. M. Ausubel. 1983. The cloning and characterization of the *glnF* (*ntrA*) gene of *Klebsiella pneumoniae* : role of *glnF* in the regulation of nitrogen fixation (*nif*) and other nitrogen assimilation genes. Mol. Gen. Genet. **192**:342-353.

Buchanen-Wollaston, V., M.C. Cannon, J.L. Beynon, & F.C. Cannon. 1981*a*. Role of *nifA* gene product in the regulation of *nif* expression in *Klebsiella pneumoniae*. Nature (London) **294**:776-778.

Buchanen-Wollaston, V., M.C. Cannon, and F.C. Cannon. 1981*b*. Isolation of symbiotically defective mutants in *Rhizobium leguminosarum* by insertion of the transposon Tn5 into a transmissible plasmid. Mol. Gen. Genet. **184**:102-106.

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 *R. meliloti* share homology with the general nitrogen regulatory gene *ntrC* of *K. pneumoniae*. Nucl. Acids Res. **13**:4539-4555.

Cannon, F.C., G.E. Riedel, and F.M. Ausubel. 1979. Overlapping sequences of *Klebsiella pneumoniae nif* DNA cloned and characterized. Mol. Gen.Genet. **174**:59-66.

Clewell, D.B., and D.R. Helinski. 1969. Supercoiled DNA-protein complexes in *Escherichia coli* : Purification and induced conversion to an open circular DNA form. Proc. Natl. Acad. Sci. USA. **62**:1159-1166.

Collins, J. 1979. *Escherichia coli* plasmids packageable *in vitro* in lambda bacteriophage particles. Methods Enzymol. **68**:309-325.

Collins, J.J., G.P. Roberts, and W.J. Brill. 1986. Posttranscriptional control of *Klebsiella pneumoniae nif* mRNA stability by the *nifL* product. J. Bacteriol. **168**:173-178.

Corbin, D., G. Ditta, and D.R. Helinski. 1982. Clustering of nitrogen fixation (*nif*) genes in *Rhizobium meliloti*. J. Bacteriol. **149**:221-228.

Dean, D. R. and K.E. Brigle. 1985. *Azotobacter vinelandii nifD*- and *nifE*-encoded polypeptides share structural homology. Proc. Natl. Acad. Sci. USA. **82**:7520-5723.

Detroy, R.W., D.F. Witz, R.A. Parejko, and P.W. Wilson. 1968. Reduction of nitrogen by complementary functioning of two components from nitrogen-fixing bacteria. Proc. Natl. Acad. Sci. USA. **61**: 537-541.

Dixon, R.A., C. Kennedy, A. Kondorosi, V. Krishnapillai, & M. Merrick. 1977. Complementation analysis of *Klebsiella pneumoniae* mutants defective in nitrogen fixation. Mol. Gen. Genet. **157**:189-198.

Dixon, R.A., R.R. Eady, G. Espin, S. Hill, M. Iaccarino, D. Kahn, & M. Merrick.1980. Analysis of regulation of the *Klebsiella pneumoniae* nitrogen fixation (*nif*) gene cluster with gene fusions. Nature (London) **286**:128-132.

Doran., J.L., W.H. Bingle, K.L. Roy, K. Hiratsuka, and W.J. Page. 1987. Plasmid transformation of *Azotobacter vinelandii* OP. J. Gen. Microbiol. **133**:2059-2072.

Drummond, M., J. Clements, M. Merrick, and R. Dixon. 1983. Positive control and autogenous regulation of the *nifLA* promotor in *Klebsiella pneumoniae*. Nature (London). **301**:302-307.

Drummond, M., P. Whitty and J. Wootton. 1985. Sequence and domain relationships of *ntrC* and *nifA* from *K. pneumoniae:* homologies to other regulatory proteins. EMBO J. **5**:441-447.

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

Elmerich, C. 1984. Molecular biology and ecology of diazotrophs associated with non-leguminous plants. Biotech. **2**:967-978.

Espin, G., A. Alvarez-Morales, and M. Merrick. 1981. Complementation analysis of *glnA* linked mutations which affect nitrogen fixation in *Klebsiella pneumoniae*. Mol. Gen. Genet. **184**:213-217.

Franche, C. and Cohen-Bazire. 1985. The structural *nif* genes of four symbiotic *Anabaena azollae* show a highly conserved physical arrangement. Plant Sci. **39**:125-131.

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

Gergen, J.P., R.H. Stern, P.C. Wensink. 1979. Filter replicas and permanent collections of recombinant DNA plasmids. Nucl. Acids Res. **7**:2115-2136.

Gronenborn, B. and J. Messing. 1978. Methylation of single-stranded DNA *in vitro* introduces new restriction endonuclease cleavage sites. Nature (London) **272**:375-377.

Grunstein, M. and J. Wallis. 1979. Colony hybridization. Methods Enzymol. **68**:379-388.

Hase,T., T. Nakano, H. Matsubara, and G. Zumft. 1981. Correspondence of the larger subunit of the MoFe-protein in clostridial nitrogenase to the *nifD* gene products of other nitrogen fixing organisms. J. Biochem. **90**:295-298.

Hase, T., S. Wakabayashi, T. Nakano, G. Zumft, and H. Matsubara. 1984. Structural homologies between the amino acid sequence of *Clostridium pasteurianum* MoFe protein and the DNA sequences of *nifD* and *K* genes of phylogenetically diverse bacteria. FEBS Lett. **166**:39-43.

Haselkorn, R. 1984. Organization and regulation of nitrogen fixation genes of *Rhodopseudomonas capsulata*. Microbiol. Sci. **1**:194-199.

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.

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.

Hennecke, H. 1981. Recombinant plasmids carrying nitrogen fixation genes from *Rhizobium japonicum*. Nature (London). **291**:354-355.

Hill, S., C. Kennedy, E. Kavanagh, R. Goldberg, & R. Hanau. 1981. Nitrogen-fixation

gene (*nifL*) involved in oxygen regulation of nitrogenase synthesis in *Klebsiella pneumoniae*. Nature (London)**290**:424-426.

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

Hirschmann, J., P.-K. Wong, K. Sei, J. Kreener, and S. Kustu. 1985. Products of nitrogen regulatory genes *ntrA* and *ntrC* of enteric bacteria activate *glnA* transcription *in vitro*: evidence that the *ntrA* product is a sigma factor. Proc. Natl. Acad. Sci. USA. **82**:7527-7529

Hohn, B. 1979. *In vitro* packaging of lambda and cosmid DNA. Methods Enzymol. **68**.299-308.

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

Hunt, T.P., and B. Magasanik. 1985. Transcription of *glnA* by purified *Escherichia coli* components: core RNA polymerase and the products of *glnF*, *glnG*, and *glnL*. Proc. Natl. Acad. Sci. USA. **82**:8453-8457.

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

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.

Jones, R., P. Woodley,and R. Robson. 1984. Cloning and organization of some genes for nitrogen fixation from *Azotobacter chroococcum* and their expression in *Klebsiella pneumoniae*. Mol. Gen. Genet. **197**:318-327.

Kallas, T., T. Coursin and R. Rippka. 1985. Different organization of *nif* genes in nonheterocystous and heterocystous cyanobacteria. Plant Mol. Biol. **5**:321-329.

Kaluza, K., M. Fuhrmann, M. Hahn, B. Regensburger, and H. Hennecke. 1983. In *Rhizobium japonicum* the nitrogenase genes *nifH* and *nifDK* are separated. J. Bacteriol. **155**:915-918.

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

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

Kennedy, C., and M.H. Drummond. 1985. The use of cloned *nif* regulatory elements

from *Klebsiella pneumoniae* to examine *nif* regulation in *Azotobacter vinelandii*. J. Gen. Microbiol. **131**:1787-1795.

Kennedy, C., R. Gamal, R. Humphrey, J. Ramos, K. Brigle, and D. Dean. 1986. The *nifH*, *nifM*, and *nifN* genes of *Azotobacter vinelandii*: Characterization by Tn5 mutagenesis and isolation from pLAFR1 gene banks. Mol. Gen. Genet. **205**:318-325.

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

Leonardo, J.M. and R.B. Goldberg. 1980. Regulation of nitrogen metabolism in glutamine auxotrophs of *Klebsiella pneumoniae*. J. Bacteriol **142**:99-110.

Lizardi, P.M., R. Binder and S.A. Short. 1984. Preparative isolation of DNA and biologically active mRNA from diethylaminoethyl membrane. Gene Anal. Tech. **1**:33-39

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. **256**:6385-6391.

MacNeil, D., M.M. Howe, and W.J. Brill. 1980. Isolation and characterization of lambda specialized transducing bacteriophages carrying *Klebsiella pneumoniae nif* genes. J. Bacteriol. **141**:1264-1271.

Maxam, A. and W. Gilbert. 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol. **65**:499-559.

Mazur, B.J., D. Rice, and R. Haselkorn. 1980. Identification of blue-green algal nitrogen fixation genes by using heterologous DNA hybridization probes. Proc. Natl. Acad. Sci. USA. **77**:186-190.

Mazur B.J., and C. Chui. 1982. Sequence of the gene coding for the beta-subunit of dinitrogenase from the blue-green algae *Anabaena*. Proc. Natl. Acad. Sci. USA. **79**:6782-6786.

McCormick, M., W. Wisehart, H, Ohtsubo, F. Heffron, and E. Ohtsubo. 1981. Plasmid cointegrates and their resolution mediated by transposon-Tn3 mutants. Gene **15**:103-118.

Mclean, P. and R.A. Dixon. 1981. Requirement of *nifV* for production of wild-type nitrogenase enzyme in *Klebsiella pneumoniae*. Nature (London) **292**:655-656.

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

Merrick, M.S. 1983. Nitrogen control of the *nif* regulon in *Klebsiella pneumoniae*:

involvement of the *ntrA* gene and analogies between *ntrC* and *nifA*. EMBO. J. **2**:29-44.

Messing, J., B. Gronenborn, B. Müller-Hill, P.H. Hofschneider. 1977. Filamentous coliphage M13 as cloning vector: insertion of a Hind II fragment of *lac* regulatory region in M13 replicative form *in vitro*. Proc. Natl. Acad. Sci. USA. **74**:3642-3646.

Messing, J. 1981. A system for shotgun DNA sequencing. Nucl. Acids Res. **9**:309-322.

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.

Messing, J. 1983. New M13 vectors for cloning. Methods Enzymol. **101**:20-76.

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

Morrison, D.A. 1979. Transformation and preservation of competent bacterial cells by freezing. Methods Enzymol. **68**:326-330.

Mortenson, L.E., and R.N.F. Thorneley. 1979. Structure and function of nitrogenase. Ann. Rev. Biochem. **48**:387-418.

Niera-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**:3346-3350.

Norel, F., N. Desnoues, and C. Elmerich. 1985. Characterization of DNA sequences homologous to *Klebsiella pneumoniae nifH, D, K* and *E* in tropical *Rhizobium* ORS571. Mol. Gen. Genet. **199**:352-356.

Nuti, M.P., A.A. Lepidi, R.K. Prakash, R.A. Schilperoort, and F.C. Cannon. 1979. Evidence for nitrogen fixation (*nif*) genes on indigenous *Rhizobium* plasmids. Nature (London) **282**:533-535.

Ow, D.W. and F.M. Ausubel. 1983. Regulation of nitrogen metabolism genes by *nifA* gene product in *Klebsiella pneumoniae*. Nature (London) **301**:307-313.

Page, W.J., and S.K. Collinson. 1982. Molybdenum enhancement of nitrogen fixation in a Mo- starved *Azotobacter vinelandii* Nif⁻ mutant. Can. J. Microbiol. **28**:1173-1180.

Pederosa, F.O., and M.G. Yates. 1984 Regulation of nitrogen fixation (*nif*) genes of *Azospirillum brasilense* by *nifA* and *ntr* (*gln*) type gene products. FEMS Microbiol. Lett. **23**:95-101.

Postgate, J., H.M. Kent and R.L. Robson. 1986. DNA from diazotrophic *Desulfovibrio*

strains is homologous to *Klebsiella pneumoniae* structural *nif* DNA and can be chromosomal or plasmid-borne. FEMS Microbiol. Lett. **33**:159-163.

Prakash, R.K., R.A. Schilperoort, and M.P. Nuti. 1981. Large plasmids of fast-growing rhizobia: homology studies and location of structural nitrogen fixation (*nif*) genes. J. Bacteriol. **145**:1129-1136.

Premakumar, R., E.M. Lemos, and P.E. Bishop. 1984. Evidence for two dinitrogenase reductases under regulatory control by molybdenum in *Azotobacter vinelandii*. Biochem. Biophys. Acta. **797**:64-70.

Pretorius, I., D.E. Rawlings and D.R. Woods. 1986. Identification and cloning of *Thiobacillus ferrooxidans* structural *nif* genes in *Escherichia coli* . Gene **45**:59-65.

Pretorius, I., 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.

Pühler, A., and W. Klipp. 1981. Fine structure analysis of the gene region for nitrogen fixation (*nif*) of *Klebsiella pneumoniae*., pp. 276-286. *In* H. Bothe & A. Trebst (eds.), Biology of Inorganic Nitrogen and Sulfur. Springer-Verlag, New York.

Rice, D., B.J. Mazur, and R. Haselkorn. 1982. Isolation and physical mapping of nitrogen fixation genes from the cyanobacterium *Anabaena* 7120. J. Biol. Chem. **57**:13157-13163.

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.

Roberts, G.P. and W.J. Brill. 1980. Gene-product relationships of the *nif* regulon of *Klebsiella pneumoniae*. J. Bacteriol. **144**:210-216.

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

Rüther, U. 1980. Construction and properties of a new cloning vector allowing direct screening for recombinant plasmids. Mol. Gen. Genet. **178**:475-477.

Ruvkun, G.B., and F.M. Ausubel. 1980. Interspecies homology of nitrogenase genes. Proc. Natl. Acad. Sci. USA. **77**:191-195.

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

Santero, E., F. Luque, J.R. Medina, and M. Tortolero. 1986 Isolation of *ntrA*-like mutants of *Azotobacter vinelandii*. J. Bacteriol. **166**:541-544.

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.

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

———————————— 1983*b*. Nitrogenasse structural genes are unlinked in the nonlegume symbiont *Parasponia rhizobium*. DNA. **2**:141-148.

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

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

Sibold, L., D. Pariot, L. Bhatnagar, M. Henriquet, and J. Aubert. 1985. Hybridization of DNA from methanogenic bacteria with nitrogenase structural genes (*nifHDK*). Mol. Gen. Genet. **200**:40-46.

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.

Southern, E.M. 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. **98**:503-517.

Southern, E.M. 1979. Gel electrophoresis of restriction fragments. Methods Enzymol. **68**:152-175.

Staden, R. 1980. A new computer method for storage and manipulation of DNA gel reading data. Nucl. Acids Res. **9**:6129-6137.

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.

Sundaresan, V., J.D.G. Jones, D.W. Ow, and F.M. Ausubel. 1983*a*. *Klebsiella pneumoniae nifA* product activates the *Rhizobium meliloti* nitrogenase promotor. Nature **301**:728-732.

Sundaresan V., D.W. Ow and F.M. Ausubel. 1983*b*. Activation of *Klebsiella pneumoniae* and *Rhizobium meliloti* nitrogenase promotors by *gln* (*ntr*) regulatory proteins. Proc. Natl. Acad. Sci. USA. **80**:4030-4034.

Tait, R.,C., T.J. Close, R.C. Lundquist, M. Hagiya, R. L. Rodriguez & C.I. Kado. 1983. Construction and characterization of a versatile broad host range DNA cloning system for gram-negative bacteria. Biotech. **1**:269-275.

Tanaka, M., M. Haniu, and K.T. Yasunobu. 1977. The amino acid sequence of *Clostridium pasteurianum* iron protein, a component of nitrogenase. J. Biol. Chem. **252:**7093-7100.

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

Toukdarian, A.E., and M.E. Lidstrom. 1984. DNA hybridiztion analysis of the *nif* region of two methylotrophs and molecular cloning of *nif* -specific DNA. J. Bacteriol. **157:**925-930.

Vieira, J. & J. Messing. 1982. The pUC plasmids, an M13mp7 derived system for insertion mutagenesis and sequencing with synthetic universal primer. Gene**19:**259.