Studies on Biological Nitrogen Fixation:
I. Mutagenesis of the Iron Protein of Nitrogenase
II. Investigation of the Genes *nif*VWZM

by

Thomas Anatol Collet

Vordiplom, RWTH Aachen, Federal Republic of Germany
(1985)

Submitted to the Department of Chemistry
in Partial Fulfillment of the
Requirements of the
Degree of

DOCTOR OF PHILOSOPHY
IN BIOLOGICAL CHEMISTRY

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

February 1991

Signature of Author Signature redacted
Department of Chemistry
September 10th, 1990

Certified by Signature redacted
Professor W.H. Orme-Johnson
Thesis Supervisor

Accepted by Signature redacted
Glenn A. Berchtold,
Chairman, Departmental Committee on Graduate Students

This doctoral thesis has been examined by a committee of the Department of Chemistry as follows:

Professor Joanne Stubbe____ Signature redacted _____
Chairperson

Professor William H. Orme-Johnson_ Signature redacted
Thesis Advisor

Professor D. Youvan_ Signature redacted ____

This Thesis is dedicated to the Memory

of my Father, Hans G. Collet.

Studies on Biological Nitrogen Fixation:

I. Mutagenesis of the Iron Protein of Nitrogenase

II. Investigation of the Genes *nif*VWZM

by

Thomas A. Collet

Submitted to the Department of Chemistry
on September 10, 1990 in partial fulfillment
of the requirements for the Degree of
Doctor of Philosophy in Biological Chemistry

Abstract

DNA-Sequence analysis has identified three open reading frames in a stretch of DNA previously thought to encode only one nitrogen fixation (*nif*) gene, *nif*M. The largest of these open reading frames was shown to be the gene for the *nif*M protein (mw 30.6 kd), while the other open reading frames encode for previously unknown *nif*-proteins, NifW (10.2 kd) and NifZ (16.7 kd). Deletion analysis (also see Harris et al., 1990) in a binary plasmid system in *E. coli* was used to show that the *nif*M gene is the only *nif*-gene other than the structural gene *nif*H needed for the synthesis of active, mature Fe protein in *E. coli*. Anaerobic, non-denaturing gel electrophoresis in combination with ^{55}Fe labeling was used to show that in the absence of *nif*M, only very small amounts of apoFe protein were made that did not contain any iron.

Site-directed mutagenesis was used to study the functions of five conserved cysteine residues in the Fe protein at positions 38, 85, 97, 132, and 184. Single amino acid changes of these cysteine residues to serine residues yielded mutant Fe proteins that were incompletely processed into the Fe protein-like dimer as determined by anaerobic, non-denaturing gel electrophoresis and that were unable to reduce the MoFe protein in the standard acetylene reduction assay. Two of these mutants (C38S and C184S) showed wild-type level activity in the maturation of the apoMoFe protein. C85S, the mutant proposed to be near the MgATP binding site (Hausinger & Howard, 1983), showed less than 5% activity in this function. C97S and C132S, the proposed cluster ligands, were completely inactive in this function also.

Studies directed at elucidating the function of the *nif*V gene product, a proposed homocitrate synthase, led to the overproduction of a biologically active *nif*V protein which failed to show significant activity in an assay monitoring the condensation of α-ketoglutarate and acetyl-CoA. An HPLC assay was developed that allows the separation of the open form of R,S-homocitrate from the lactone form. This assay was used to tentatively identify the open form of homocitrate as the form secreted by K. *pneumoniae* during derepression for nitrogenase.

Thesis Supervisor: Dr. W.H. Orme-Johnson
Title: Professor of Chemistry

Acknowledgements

Many people have made this thesis possible.

My work as a graduate student has been under the supervision of Professor W.H. Orme-Johnson. I thank him for giving me the opportunity to experience science in an intellectually stimulating environment and for letting me explore an unusual variety of scientific problems. His efforts in instructing me in the scientific method, as well as in improving my taste in music, will leave a long-lasting impression.

My gratitude goes to all the members and former members of the Orme-Johnson lab whom I have had the opportunity to learn from and also to all the graduate students and post-docs in MIT's biology labs who were willing to listen to the problems of a chemist doing biology. I would especially like to thank the following persons: for setting an example, Karen Howard and Georgianna Harris; for their patience in dealing with a difficult colleague, Theresa White and Anne-Frances Miller; for their encouragement and help, David Wright, Patricia Christie, and Mary Pat Reeve; and for showing that graduation is indeed possible, Alison Hickman. Contributions to individual experiments are acknowledged where appropriate in the main text.

I am grateful to my family for their unwavering support of my academic career. The knowledge that they would support any educational goal that I could set myself is the origin of this thesis.

My thanks for unique encouragement and assistance with typing go to Mr. V. Katz.

My fiancée, Jenny, deserves special mention: without her, this thesis would not have been finished. Quite apart from proof-reading it, she provided the encouragement and perspective on life when and where it was needed. I thank her for putting up with the crazy schedule of a student working at the lab bench, the lack of a personal life for the last few months, and the general impossibilities of living with a graduate student.

Table of Contents

List of Figures

Chapter IV

Appendix A

List of Tables

Chapter I:

Literature Review and Scope of This Thesis

A. Introduction

Nitrogen is the most abundant element available to man in its uncombined form. Its local concentration, however, is highest in the atmosphere (78.1% as N_2), while it is present in the litosphere at only 19 ppm in chemically bound forms metabolically readily accessible to the biosphere. Essential in biological systems, nitrogen in its "fixed" form, i.e. non-N_2, is of overwhelming importance in industrialized society as well. On a molar basis, ammonia is the chemical produced in the largest quantities (85×10^6 tons per annum) and nitric acid, urea, and ammonium nitrate are also found among the top 15 chemicals. Estimates for biological nitrogen fixation vary, but at $90\text{-}175 \times 10^6$ tons per annum it is still larger than any man-made synthetic process (Greenwood & Earnshaw, 1984, and C&E News, 1990).

Industrial production of ammonia from the elements became feasible in 1913 with the introduction of the Haber-Bosch process: a mixture of N_2 and H_2 is passed over a promoted iron oxide catalyst inside a converter tube at high pressure ($\approx$200 atm) and high temperature ($\approx$400° C). The ammonia produced is collected by condensation at subambient temperature (Holleman & Wiberg, 1976). Ammonia from this process is used mostly as fertilizer to increase agricultural yields; biological nitrogen fixation becomes limiting in modern agriculture, which is oriented towards high yields without crop rotation.

Biological systems do not have the option of employing such extreme reaction conditions to fix the nitrogen they need to grow. Instead, the enzyme nitrogenase, which performs the catalytic conversion of N_2 to NH_3 inside bacterial cells, is a multienzyme complex optimized to function at ambient temperature and pressure (Orme-Johnson, 1985). The enzymatic systems studied to date rely heavily on inorganic metal clusters to perform the chemistry of nitrogen reduction, especially on Fe/S and Mo/Fe/S clusters, although recently alternate systems using other metals have been found. Unfortunately, biological nitrogen fixation is characterized by an extreme oxygen sensitivity and a requirement for large amounts of energy in the form of MgATP.

The only organisms able to fix nitrogen from N_2 in the environment are prokaryotes and archaebacteria. Some of these organisms are free-living, others enter symbiotic relationships with eukaryotes; on the whole, it is these organisms that replenish fixed nitrogen in soils depleted by leaching or agricultural activity. The requirement for

anaerobiosis and energy determine under which conditions nitrogen fixation can take place: anaerobes, facultative anaerobes, aerobes, and symbiotic bacteria have developed in their ecological niches to meet these requirements. *Clostridium*, an obligate anaerobe, will grow only in the absence of oxygen. *Klebsiella*, a facultative anaerobe, can grow both with and without oxygen; however, it can only fix nitrogen in the absence of oxygen. *Azotobacter*, an obligate aerobe, requires oxygen for growth and nitrogen fixation; however, nitrogen fixation activity becomes sensitive to oxygen once the carbon source in the medium is depleted and the high rate of respiration needed to keep the intracellular concentration of oxygen low cannot be maintained. Symbiotic bacteria like *Rhizobia* give up part of their identity as free-living bacteria to associate with plants: the plant gains a source of fixed nitrogen while the bacterium gains both a source of carbon and protection from the harsh environment (Dixon & Wheeler, 1986). Because the physiology of free-living nitrogen fixers has been studied since the end of the last century, most of the work aimed at the elucidation of the molecular mechanism of biological nitrogen fixation has centered around *Klebsiella pneumoniae*, *Azotobacter vinelandii*, *Clostridium pasteurianum*, and related organisms.

Due to the tremendous importance of nitrogen fixation for agriculture, much effort by researchers ranging from plant physiologists to synthetic inorganic chemists has been directed at the regulation and the molecular mechanims of biological nitrogen fixation (Orme-Johnson, 1985; Coucouvanis & Kanatzidis, 1985; and Gussin et al., 1986). More, however, remains to be learned about the systems involved before significant applications such as transgenic plants with the ability to fix their own nitrogen will result.

B. The Regulation of Biological Nitrogen Fixation

Consistent with the oxygen sensitivity of the nitrogenase proteins and their extreme energy demands, expression of the nitrogen fixation (*nif*) genes is strongly repressed in the presence of fixed nitrogen. The derepression of nitrogenase when fixed nitrogen becomes limiting is a complex process which involves the activation of 20 genes in at least eight operons (J, HDKTY, ENX, USVWZ, M, F, AL, BQ; Figure I.1.) in the *nif* cluster in *Klebsiella pneumoniae* under the control of the glutamine synthase operon (*gln*).

Transcription of the *nif* operon is under the control of the *nifA*[1] gene product, a positive regulator, in conjunction with GlnF (also called NtrA), the alternate sigma factor

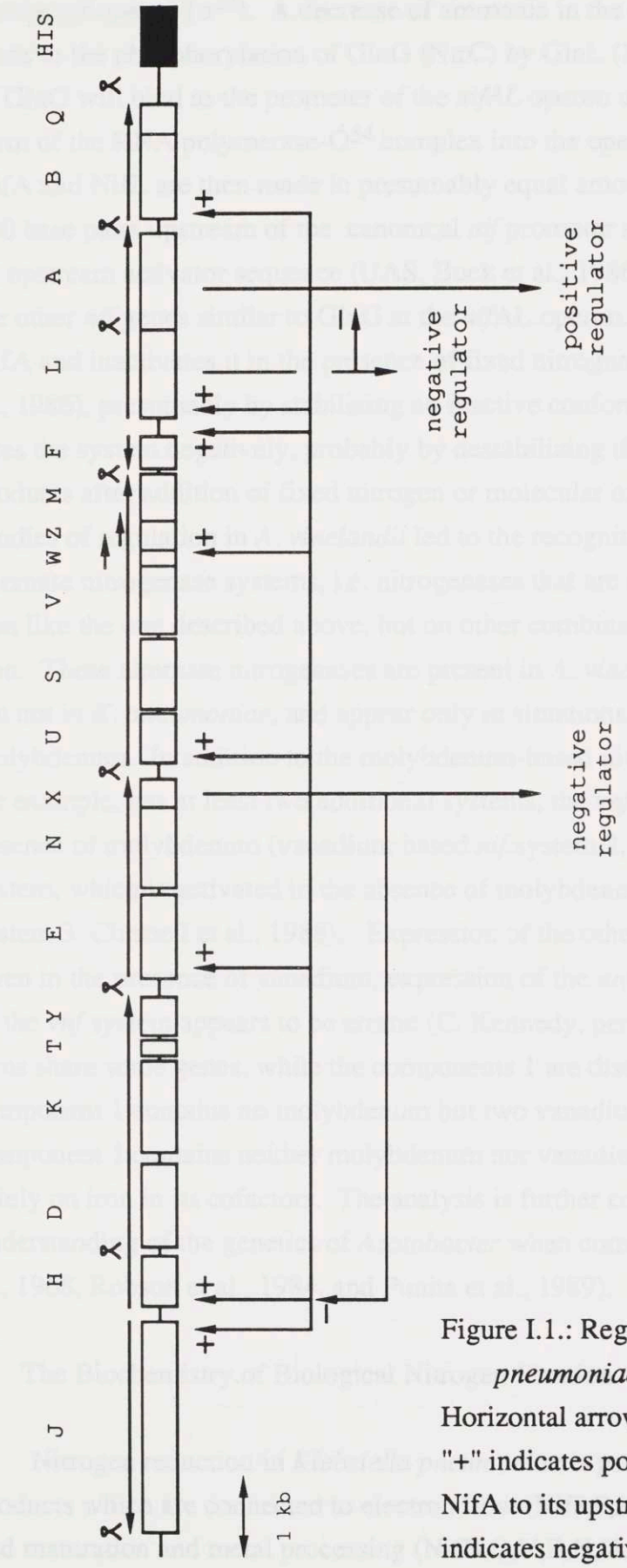


Figure I.1.: Regulation of *nif* Genes in *K. pneumoniae*

Horizontal arrows denote transcriptional units. A "+" indicates positive regulation by binding of NifA to its upstream activator sequence. A "-" indicates negative regulation by NifX or NifL.

for the *nif* operon (σ^{54}). A decrease of ammonia in the cell in the absence of oxygen leads to the phosphorylation of GlnG (NtrC) by GlnL (NtrB). The phosphorylated form of GlnG will bind to the promoter of the *nifAL* operon upstream and convert the closed form of the RNA polymerase-$Ó^{54}$ complex into the open form to initiate transcription. NifA and NifL are then made in presumably equal amounts. NifA binds approximately 100 base pairs upstream of the canonical *nif* promoter sequences (Beynon et al., 1983) to its upstream activator sequence (UAS, Buck et al., 1986) and activates transcription of the other *nif* genes similar to GlnG at the *nif*AL operon. NifL acts as an antagonist to NifA and inactivates it in the presence of fixed nitrogen and molecular oxygen (Kong et al., 1986), presumably by stabilizing an inactive conformation of NifA. NifX also modulates the system negatively, probably by destabilizing the mRNA of some of the gene products after addition of fixed nitrogen or molecular oxygen (Gosink et al., 1990). Studies of regulation in *A. vinelandii* led to the recognition (Bishop et al., 1980) of alternate nitrogenase systems, i.e. nitrogenases that are not based on molybdenum and iron like the one described above, but on other combinations of metals like vanadium and iron. These alternate nitrogenases are present in *A. vinelandii* and some other bacteria, but not in *K. pneumoniae*, and appear only in situations where cells are deprived of molybdenum. In addition to the molybdenum-based nitrogenase, *Azotobacter vinelandii*, for example, has at least two additional systems, the *vnf* system, which is turned on in the absence of molybdenum (vanadium based *nif* system 2, Bishop et al., 1980) and the *anf* system, which is activated in the absence of molybdenum and vanadium (alternate *nif* system 3, Chisnell et al., 1988). Expression of the other two systems is complex: (a) Even in the presence of vanadium, expression of the *anf* system is leaky. (b) Suppression of the *vnf* system appears to be erratic (C. Kennedy, personal communication). All systems share some genes, while the components 1 are distinct in their metal content: the *vnf* component 1 contains no molybdenum but two vanadium atoms per molecule; the *anf* component 1 contains neither molybdenum nor vanadium and is presumed to be based solely on iron in its cofactors. The analysis is further complicated by a general lack of understanding of the genetics of *Azotobacter* when compared to *Klebsiella* (Strandberg et al., 1968, Robson et al., 1984, and Punita et al., 1989).

C. The Biochemistry of Biological Nitrogen Fixation

Nitrogen reduction in *Klebsiella pneumoniae* is performed by 20 different gene products which are connected to electron flow (NifJ,F,H,D,K), regulation (NifA,L,X), and maturation and metal processing (NifB,Q,V,E,N,U,S,M). Some gene products are

not yet connected with a function (NifW,Z,T,Y). Attention has focused on electron flow (Figure I.2.) and the catalytic metal centers: two electrons produced by oxidative decarboxylation of pyruvate are funnelled from NifJ, a pyruvate-flavoprotein oxidoreductase (Wahl & Orme-Johnson, 1987), to NifF, a flavoprotein, translating the two-electron process into a one electron pathway (Nieva-Gomez et al., 1980). One electron at a time is fed into the iron protein, the specific reductant of the molybdenum-iron protein. Two iron protein dimers (*nifH* gene products), each carrying two MgATP, bind to one molybdenum-iron protein $\alpha_2ß_2$-tetramer and reduce it with coupled hydrolysis of MgATP (Orme-Johnson et al., 1972, and Hageman et al., 1980). Each eight turnovers, the fully reduced MoFe protein (*nifD,K* gene product) reduces one molecule of N_2 to 2 NH_3 and evolves one molecule of H_2 under ideal conditions. Under less than optimal conditions, the ratio of H_2 to NH_3 increases:

$$8\ e^- + 8\ H^+ + N_2 + 16\ MgATP + 16\ H_2O \rightarrow 16\ MgADP + 16\ Pi + 2\ NH_3 + H_2$$

Compounds other than N_2 containing triple bonds can also serve as substrates. Indeed, the standard laboratory assay for nitrogenase activity is the reduction of acetylene to ethylene using dithionite as in vitro reductant.

Because only two protein components are required in vitro for nitrogen reduction, attention has focused on the interaction between component 1 (molybdenum-iron protein) and the iron protein (component 2)[2]. An elaborate, multi-variable model accounts for the interactions between the components during catalysis and for product development curves for both ammonia and hydrogen (Thorneley & Lowe, 1983 and 1984). Essentially, the model consists of a cycle in which component 2 injects low potential electrons into component 1 ("Fe protein cycle") and a cycle in which products and intermediates are released from the MoFe protein at various states of reduction under various conditions ("MoFe protein cycle"). Features of this multi-variable model include the following:

- Kp2 is obliged to dissociate from Kp1 between each round of reduction; this explains the lag phase observed when product appearance is monitored at high Kp1 concentrations.
- The rates in the iron protein cycle are essentially independent of the state of reduction of the molybdenum-iron protein.

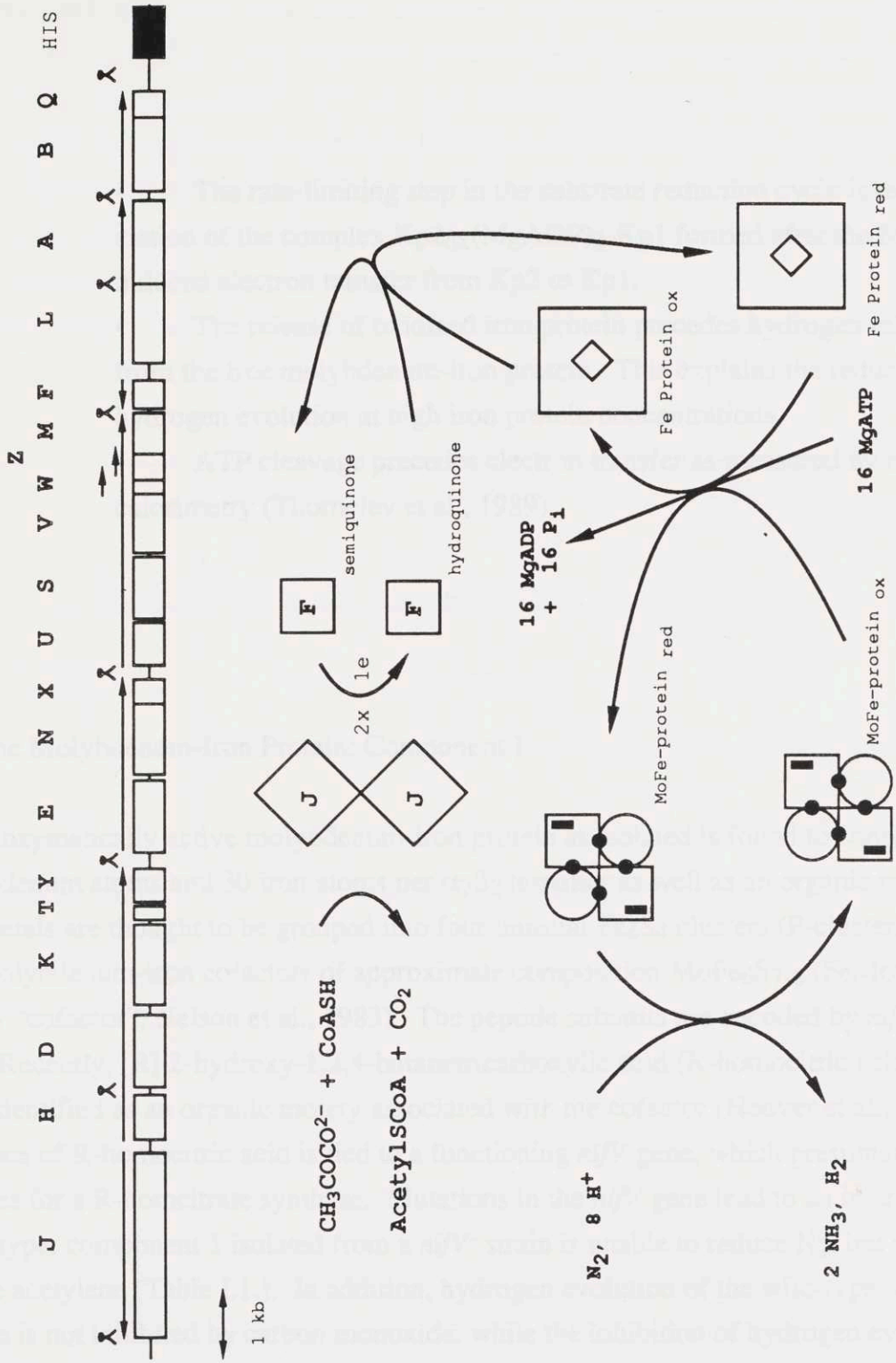


Figure I.2.: Electron Flow in Nitrogenase

Numbers near reaction arrows denote overall stoichiometry, not mechanistic detail. The open diamond symbol in the Fe protein stands for the Fe_4S_4-cluster. The filled rectangle in the MoFe protein signifies the molybdenum-iron cofactor, the filled circles the Fe_4S_4-clusters.

- The rate-limiting step in the substrate reduction cycle is the dissociation of the complex $Kp2_{ox}(MgADP)_2$-Kp1 formed after the MgATP-induced electron transfer from Kp2 to Kp1.
- The release of oxidized iron protein precedes hydrogen release from the free molybdenum-iron protein. This explains the reduced hydrogen evolution at high iron protein concentrations.
- ATP cleavage precedes electron transfer as measured by microcalorimetry (Thorneley et al., 1989).

D. The Molybdenum-Iron Protein: Component 1

Enzymatically active molybdenum-iron protein as isolated is found to contain 2 molybdenum atoms and 30 iron atoms per $\alpha_2\beta_2$ tetramer as well as an organic moiety. The metals are thought to be grouped into four unusual Fe_4S_4 clusters (P-clusters) and two molybdenum-iron cofactors of approximate composition $MoFe_6S_{7-8}$ (FeMoCo or simply "cofactor"; Nelson et al., 1983). The peptide subunits are encoded by *nifD* and *nifK*. Recently, [R] 2-hydroxy-1,2,4-butanetricarboxylic acid (R-homocitric acid) has been identified as an organic moiety associated with the cofactor (Hoover et al., 1987). Presence of R-homocitric acid is tied to a functioning *nifV* gene, which presumably encodes for a R-homcitrate synthase. Mutations in the *nifV* gene lead to an interesting phenotype: component 1 isolated from a *nifV*$^-$ strain is unable to reduce N_2, but will still reduce acetylene (Table I.1.). In addition, hydrogen evolution of the wild-type MoFe protein is not inhibited by carbon monoxide, while the inhibition of hydrogen evolution by carbon monoxide is a convenient assay for the *nifV*$^-$ MoFe protein (McLean & Dixon, 1983).

One of the most interesting features of this protein is the fact that cofactor-less component 1 (apoMoFe protein) and cofactor can be isolated separately. apoMoFe protein can be isolated from strains with mutations in genes responsible for cofactor

synthesis: *nifH*, *nifB*, *nifQ*, *nifN*, *nifE*. Extraction of the cofactor from acid-precipitated holoprotein into NMF or isolation of the cofactor using an improved chromatographic method (McLean et al., 1989, and Wink et al., 1989) allow in vitro complementation studies: typically, apoMoFe protein from a mutant in *nifB* is complemented by cofactor isolated from a different strain to yield holoMoFe protein. This pivotal experiment indicated the cofactor as site of substrate reduction: complementation using cofactor isolated from a *nifV*$^-$ strain yields holoprotein with the *nifV*$^-$ phenotype, while complementation using cofactor isolated from a *nifV*$^+$ strain yields wild-type holoprotein able to reduce N_2.

	nitrogen reduction	acetylene reduction	carbon monoxide inhibition of hydrogen evolution
wild-type component 1	yes	yes	no
nifVminus component 1	no	yes	up to 40% inhibition

Table I.1.: *nif*V Phenotype

Preliminary X-ray crystallographic analysis (Figure I.3.; Bolin, 1990) provides some limited insight into the arrangement of the clusters in component 1: at the resolution of this model, the four P-clusters cannot be resolved, but show up as two clusters of 8 iron atoms each. The distance between the eight iron site and the proximal cofactor is 19 Å, while the distance between the two cofactors is approximately 70 Å. This distance arrangement is compatible with electron transfer within each pair of cofactor/8Fe center, but not between these pairs. No information is given of the arrangement relative to α or β subunits.

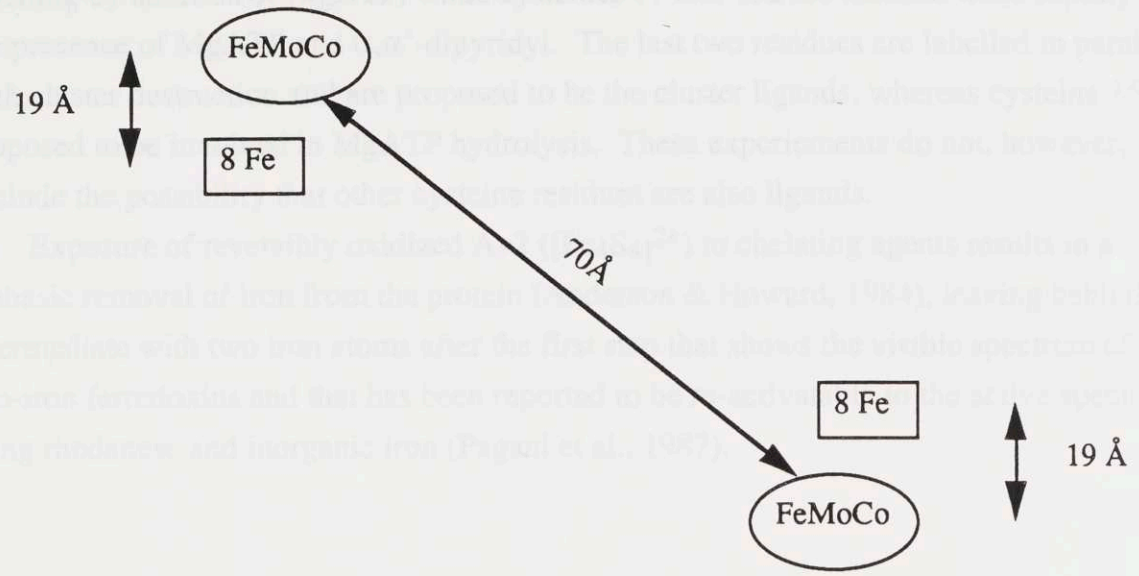


Figure I.3.: Preliminary Crystal Structure of Component 1

E. The Iron Protein: Component 2

The iron protein of nitrogenase is a dimer of molecular weight 68,000 containing one Fe_4S_4 cluster. Identical subunits are encoded by *nifH* and show a high degree of evolutionary conservation across more than 10 species sequenced at the DNA or protein level. In particular, five cysteine residues are conserved at positions 38, 85, 97, 132, and 184 (sequence positions for Kp2). The only known function for the iron protein during catalytic turnover is to provide low potential electrons to component 1, a process that is coupled to the hydrolysis of at least two MgATP per electron transferred. Another, less characterized, function for *nifH* peptide/active iron protein is its participation in the assembly of component 1.

Binding of MgATP has a major effect on the iron protein (Walker & Mortenson, 1974). Crystals of the protein shatter when placed in a solution of MgATP, and the EPR spectrum of the iron protein changes dramatically upon addition of MgATP, indicating that major conformational changes occur (Lindahl et al., 1987). Labelling studies (Hausinger & Howard, 1983) using iodo[^{14}C] acetic acid have shed some light on the connection between MgATP and Fe_4S_4 cluster binding: cysteine 85 is protected from

labelling by addition of MgATP, while cysteines 97 and 132 are labelled more rapidly in the presence of MgATP and α,α'-dipyridyl. The last two residues are labelled in parallel with cluster destruction and are proposed to be the cluster ligands, whereas cysteine 85 is proposed to be involved in MgATP hydrolysis. These experiements do not, however, exclude the possibility that other cysteine residues are also ligands.

Exposure of reversibly oxidized Av2 ($[Fe_4S_4]^{2+}$) to chelating agents results in a biphasic removal of iron from the protein (Anderson & Howard, 1984), leaving behind an intermediate with two iron atoms after the first step that shows the visible spectrum of two-iron ferredoxins and that has been reported to be re-activatable to the active species using rhodanese and inorganic iron (Pagani et al., 1987).

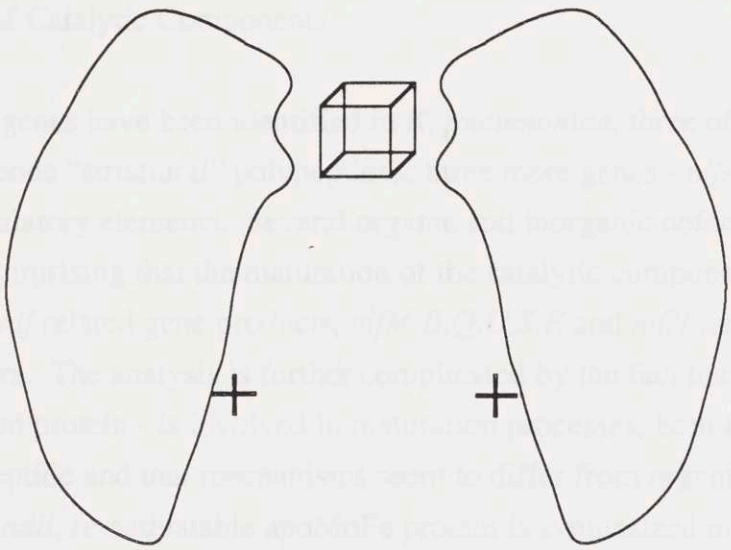


Figure I.4.: Preliminary Crystal Structure of Component 2

EPR[3] analysis of iron protein in the reduced state ($[Fe_4S_4]^{1+}$) has shown them to be similar to reduced ferredoxins, however, the signal was found to integrate to less than 1 spin/molecule. The non-stoichiometric integration of the signal has been proposed to be due to mixed spin states in the cluster: only 40% of the clusters are in the S=1/2 form exhibiting EPR near g=2, whereas 60% are presumed to be in the S≥3/2 form and show no EPR in this region. EXAFS[4] measurements indicate the presence of a Fe_4S_4-cluster; indeed, they suggest a distorted cluster with two different Fe-S distances and the pres-

ence of two conformers, one of which has no EPR signal at g=2 (Lindahl et al., 1985, and Lindahl et al., 1987).

Preliminary X-ray crystallographic studies (Georgiadis, 1990) at 3 Å resolution show the iron protein to be butterfly-shaped with the cluster sitting relatively exposed at the top of the cleft between the subunits and bound to cysteines 97 and 132. The nucleotide binding site is proposed to be inside the cleft between the subunits at the N-terminal helix approximately 20 Å away from the cluster (indicated by + in Figure I.4.). The subunits are of the single-domain type and contact between them is limited to the region immediately surrounding the cluster.

F. Maturation of Catalytic Components

Twenty *nif* genes have been identified in *K. pneumoniae*, three of which - *nifH,D,K* - are known to encode "structural" polypeptides; three more genes - *nifA,L,*X - have been identified as regulatory elements. Several organic and inorganic cofactors are known. It is therefore not surprising that the maturation of the catalytic components is complex. Apart from non-*nif* related gene products, *nifM,B,Q,U,S,E* and *nifN* are known to be maturation factors. The analysis is further complicated by the fact that the *nifH* gene product[5] - the iron protein - is involved in maturation processes, both as assembled protein and as peptide and that mechanisms seem to differ from organism to organism.

In *A. vinelandii*, re-activatable apoMoFe protein is synthesized in the cell even if *nifH* is deleted from the chromosome (Robinson et al., 1987). However, in vivo cofactor synthesis requires presence of NifH, and presumably the active iron protein. Crude extracts of *nifH*-deleted strains have no requirement for the addition of active iron protein to insert added cofactor. Insertion of the cofactor into partially purified apoMoFe protein made by a *nifH*-deleted strain, however, requires the presence of the iron protein/MgATP complex (Robinson et al., 1989).

By contrast, in *K. pneumoniae, nif*H peptide is required for the biosynthesis of the cofactor and for the biosynthesis of the apoMoFe protein (Figure 5), but not for the insertion of the cofactor into purified apoMoFe protein (Paustian et al., 1990). Extracts from *nifH*-deficient strains have been shown to be not reactivatable with cofactor, presumably because they do not contain mature apoMoFe protein (Filler et al., 1986).

In neither case does there seem to be a requirement for the presence of the structural genes *nifD* and *nifK* for cofactor synthesis (Robinson et al., 1986).

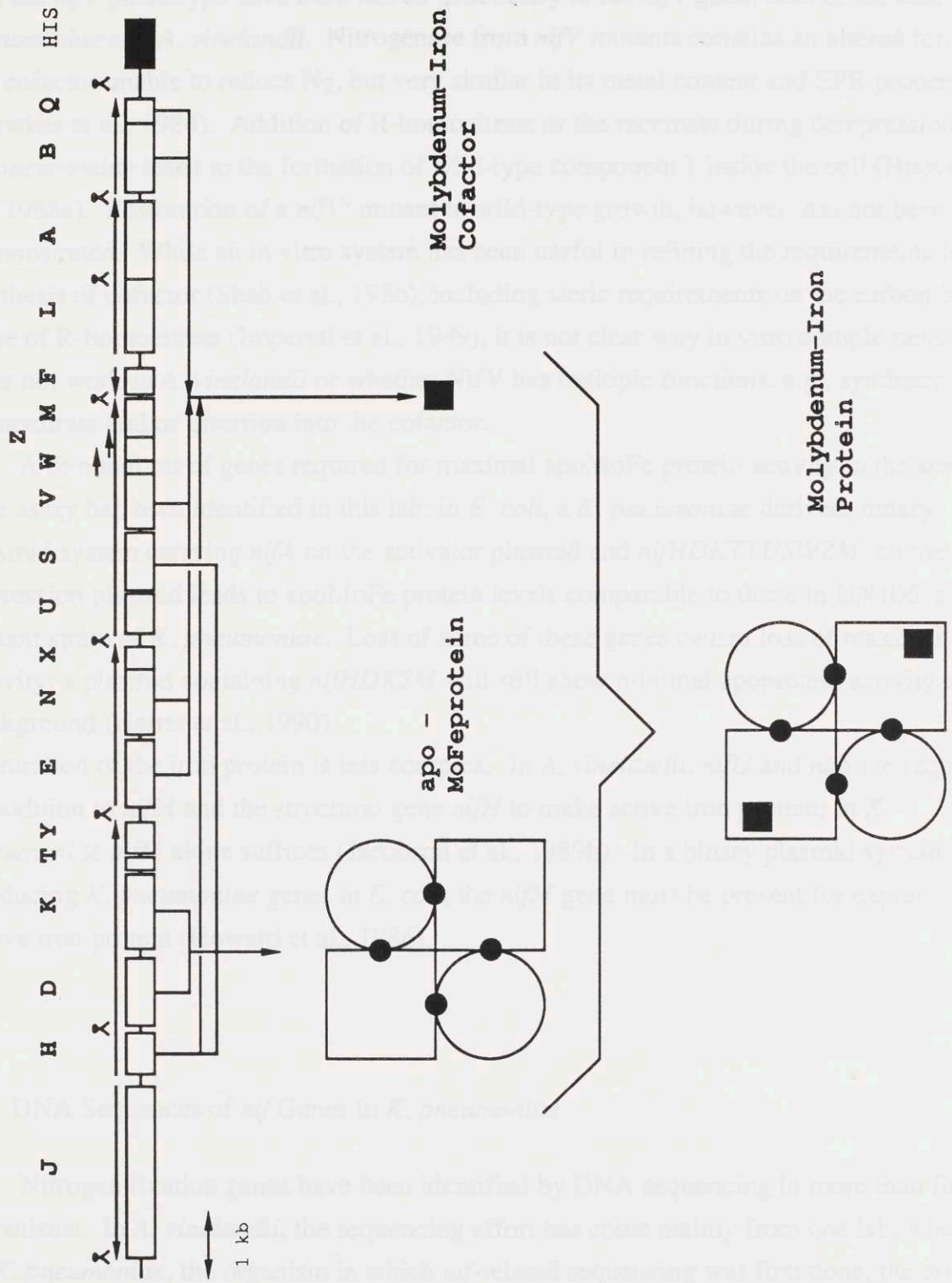


Figure I.5.: Maturation of the MoFe protein

The filled square denotes the molybenum-iron cofactor of nitrogenase, the filled circles denote the Fe_4S_4-clusters. The relative orientations of the metal cofactors to the protein subunits are not known.

R-homocitrate is an organic moiety in the cofactor (Hoover et al., 1989). Its presence and the *nifV* phenotype have been linked genetically to the *nifV* gene, both in the case of *K. pneumoniae* and *A. vinelandii*. Nitrogenase from *nifV* mutants contains an altered form of the cofactor unable to reduce N_2, but very similar in its metal content and EPR properties (Hawkes et al., 1984). Addition of R-homocitrate or the racemate during derepression of *K. pneumoniae* leads to the formation of wild-type component 1 inside the cell (Hoover et al., 1988a). Restoration of a *nifV*$^-$ mutant to wild-type growth, however, has not been demonstrated. While an in vitro system has been useful in refining the requirements for the synthesis of cofactor (Shah et al., 1986), including steric requirements on the carbon backbone of R-homocitrate (Imperial et al., 1989), it is not clear why in vitro complementation does not work in *A. vinelandii* or whether NifV has multiple functions, e.g., synthesis of R-homocitrate and/or insertion into the cofactor.

A minimal set of genes required for maximal apoMoFe protein activity in the acetylene assay has been identified in this lab: in *E. coli*, a *K. pneumoniae* derived, binary plasmid system carrying *nifA* on the activator plasmid and *nifHDKTYUSWZM* on the expression plasmid leads to apoMoFe protein levels comparable to those in UN106, a *nifB*$^-$ mutant strain of *K. pneumoniae*. Loss of some of these genes causes loss of maximal activity: a plasmid containing *nifHDKSM* will still show minimal apoprotein activity above background (Harris et al., 1990).

Maturation of the iron protein is less complex. In *A. vinelandii*, *nifU* and *nifS* are required in addition to *nifM* and the structural gene *nifH* to make active iron protein; in *K. pneumoniae nifM* alone suffices (Jacobson et al., 1989b). In a binary plasmid system producing *K. pneumoniae* genes in *E. coli*, the *nifM* gene must be present for expression of active iron protein (Howard et al., 1986).

G. DNA Sequences of *nif* Genes in *K. pneumoniae*

Nitrogen fixation genes have been identified by DNA sequencing in more than fiftenn organisms. In *A. vinelandii*, the sequencing effort has come mainly from one lab, whereas in *K. pneumoniae*, the organism in which *nif*-related sequencing was first done, the effort was shared by many laboratories. Unfortunately, this diversity is also reflected in the DNA sequences published and deposited with GenBank and the EMBL database. Unexpected problems with the cloning of the *nifV*gene of *K. pneumoniae* prompted a comprehensive

review of the published *nif*-DNA sequence data for this organism. A surprisingly large number of disagreements were found over a stretch of approximately 24,000 base pairs. This greatly reduces the value of the sequence information. As there was a fundamental question as to the extent to which the sequence disagreements reflected simple genetic divergence due to the absence of a direct evolutionary selection process, the authors of the various published sequences were contacted directly to resolve these disagreements by going back to the original data. This process culminated in the Satellite Meeting I of the 8th International Nitrogen Fixation Congress in Knoxville, TN, in which many of these discrepancies were cleared up. A "best consensus" sequence was compiled which represents the best current knowledge of the DNA sequence and reflects all remaining ambiguities. Sequence and alignment were deposited in a newly created database at the European Molecular Biology Laboratory in Heidelberg, FRG (Appendix D).

H. Scope of This Thesis

The work presented in this thesis is a continuation of work begun by others in the laboratories of W.H. Orme-Johnson at the Massachusetts Institute of Technology and of Fred Ausubel at the Massachusetts General Hospital.

Work with plasmid-expressed iron protein in the system developed by Dr. K. Howard led to the refinement of the previous model for the maturation requirement of the iron protein in genetic terms by sequence analysis and molecular terms by analysis of iron content and stability of the iron protein in absence of NIFM. This was in part a continuation of work done by Chris Earl in the lab of Fred Ausubel at MGH. These experiments are described in Chapter II.

Chapter III deals with the relevance of five evolutionarily conserved cysteine residues in the iron protein of *Klebsiella pneumoniae*. Cys-to-ser mutants created by site-directed mutagenesis are analyzed in their stability when expressed in *E. coli* and their behavior during catalytic turnover and maturation of the apoMoFe protein.

Chapter IV details experiments aimed at elucidating the role of R-homocitrate in the biosynthesis of the MoFe protein. In vivo complementation experiments with *nifV*$^-$ strains of *Klebsiella pneumoniae* and *Azotobacter vinelandii* are described, as is the overproduction of NifV from *K. pneumoniae* in *E. coli*.

I. Footnotes for Chapter I

[1] Genes will be indicated by small letters (*nif*A), gene products with capital letters (NifA).

[2] Any given component from an organism is identified by an upper-case letter for the genus and a lower-case letter for the species; a number indicates the component. Kp1 is therefore the component 1 from *Klebsiella pneumoniae*.

[3] EPR - Electron Paramagnetic Resonance

[4] EXAFS - Extended X-Ray Absorption Fine Structure

[5] Whenever the assembled, catalytically active gene product is meant, it will be called iron protein; the polypeptide unit regardless of its form in solution will be called NifH or *nif*H peptide.

Chapter II:

*nif*WZM and The Maturation of the Fe Protein

A. Introduction

One of the main challenges in understanding biological nitrogen fixation is the unraveling of its genetic complexity. On the most basic level, 20 *nif* gene products participate in the molecular mechanism of reducing nitrogen to ammonia. Although ultimately an in vitro system of purified proteins is desirable, presently the most straightforward approach to understanding the functions of various gene products is to partition the most complex functions into simpler functions that can be studied separately. For this purpose, the *nif* system in *K. pneumoniae* can be seen in terms of MoFe cofactor synthesis, apoMoFe protein synthesis, Fe protein synthesis, and catalytic activity. Traditionally, one would use the bacteriophage Mu to generate chromosomal deletions for the easiest study of *nif* gene functions (Bachhuber et al., 1976, and MacNeil et al., 1978a). Experiments based on these experiments have generated a wealth of information on the roles of various *nif* genes (Elmerich et al., 1978; Roberts et al, 1978; MacNeil et al., 1978b; and St. John et al., 1975). Modern recombinant DNA methods have allowed the study of the molecular mechanisms of holoprotein synthesis by making the complementary approach possible. Moving selected *nif* genes into a different, yet well characterized genetic background avoids most of the complications of dealing with a multitude of genes. This lab has successfully employed this approach to study the maturation requirements of the apoMoFe protein (Harris et al., 1990) and the Fe protein (Howard et al., 1986). In the work undertaken by Dr. K. Howard, the *nif*M protein (appr. mw 30 kd) located on a 1.8 kb DNA fragment was identified as the only protein factor required for the synthesis of active Fe protein other than the actual *nif*H structural gene product. Initial experiments designed to elucidate the exact role of the *nif*M gene product indicated that it might be involved in processing of the Fe_4S_4 cluster of the Fe protein.

Experiments to further study the interactions between the Fe protein and the M protein clearly required the complete analysis of the previously unsequenced DNA

fragment carrying the *nif*M gene as well as an improved experimental method to determine the iron content of the Fe protein made in the absence of *nif*M protein.

B. Methods

DNA Sequencing: This sequencing project was taken over from Chris Earl in Fred Ausubel's laboratory at the Massachusetts General Hospital, Boston. Unpublished sequence data left from the sequencing of *nif*F and *nif*V (Chris Earl, Ph.D. Thesis, 1985) was compiled and compared to the *nif*M promoter sequence (Beynon et al., 1983). This data is indicated by a thin line spanning from nucleotide 700 to nucleotide 1792 in Figure II.1. (see C. Results). Compressions in this sequence are indicated by an asterisk (*). A possible coding region for *nif*M was identified. Using already existing clones and newly constructed subclones of pVL13 in M13mp18/19 (SacI, BamHI, EcoRV, HpaI, XhoI, HpaI fragments), the sequence ambiguities were ressolved and it was found that more than one gene was located on the fragment corresponding to the EcoRI insert of pVL13 (pVL13 is pACYC184 carrying a 1.8 kb *nif* insert). Additional open reading frames (orf's) between *nif*M and *nif*V were determined using more unpublished sequence data and new sequence data derived from two synthetic primers (FA and SG) and a clone constructed by Dr. K. Howard (pKH19:EcoRV-PstI clone in M13mp18). All sequencing was done using first 2' deoxy-inosine 5'-triphosphate (dITP) and later 7-deaza-2'-deoxyguanosine 5'-triphosphate (C^7-dGTP, Mizusawa et al., 1986) to resolve the numerous compressions found.

Data Analysis: All computer algorithms used were implemented on the VAX computer at the Whitaker Computing Facility at the Massachusetts Institute of Technology as part of the University of Wisconsin Genetics Computer Group DNA and Protein Analysis Software (Version 6.0-6.2).

Synthesis of Peptides: The synthesis of the first peptide - nifZ_C - was performed using FMOC (9-fluorenylmethyloxycarbonyl) chemistry. The three other peptides - nifW_C, nifM_C, and nifH_N - were synthesized at the Biopolymers Laboratory in the Center for Cancer Research at MIT using BOC (tert-butyloxycarbonyl) chemistry.

nifH_N	TMRQCAIYGKGGIGKST	mw 1771.24
nifW_C	YQQQFQESGT	mw 1215.39
nifZ_C	AIALIEEREE	mw 1172.41
nifM_C	MISRQPGLCG	mw 1061.33

Synthesis of nifZ_C is described in detail in Appendix A.

Purification of Peptides: Samples containing cysteine and methionine residues were prepared for analysis by reducing the crude, oxidized peptide with dithiothreitol. For this purpose, 60 mg crude peptide were treated with 2 ml 20% (w/v) dithiothreitol and 8 ml 6 mol/l guanidine hydrochloride 300 mmol/l Tris (observed pH 10-11) overnight at room temperature. Before injection, particulate matter was removed by centrifugation and filtration through a 0.2 μm filter. Analysis and purification of all peptides was performed using reverse phase HPLC (Waters, Inc.). Analytical scale separations were done on a Vydac C18 reverse phase column (250 x 4.1 mm, 300 Å pore size) using an Applied Biosystems, Inc., RP-18 guard column. Preparatory scale separations were also done on a Vydac C18 reverse phase column (250 x 21 mm, 300 Å pore size). In both cases, a gradient was run from 10-70% solution B in solution A over 40 min (solution A: 0.1% TFA in H_2O, solution B: CH_3CN). Flowrates were 1 ml/min (analytical scale) and 18 ml/min (preparatory scale). Primary detection was at 214 nm; secondary detection was at 254 nm. Fractions containing the desired material were collected, lyophilized, and stored at -20° C. Purified peptides were checked by amino acid analysis and Fast Atom Bombardment - Mass Spectroscopy.

Preparation of Conjugates: Conjugates were prepared according to two procedures: nifH_N was conjugated to bovine serum albumin (BSA) with 1-(3-dimethyl-

aminopropyl)-3-ethyl-carbodiimide hydrochloride (EDC; Staros et al., 1986) and nifW_C, nifZ_C, and nifM_C were also coupled to BSA using glutaraldehyde (Van Regenmortel et al., 1988). To couple nifH_N, 3 mg BSA were dissolved in 0.3 ml H_2O and 3 mg nifH_N (30-fold molar excess) were suspended in 0.75 ml coupling buffer (100 mmol/l 4-morpholineethane-sulfonic acid pH 4.7, 900 mmol/l NaCl, 0.02% NaN_3, 4 mmol/l N-hydroxysulfo-succinimide [S-NHS]). Both solutions were mixed and added to 15 mg EDC with shaking. The reaction was allowed to proceed for 2 hours at room temperature. Precipitation was observed and the reaction was repeated twice with fresh BSA and peptide using 7.5 and 2.5 mg EDC. Supernatants and precipitates from all reactions were pooled and dialyzed extensively against buffered saline solution (PBS: 1.236 g/l Na_2PO_4, 0.18 g/l $NaH_2PO_4xH_2O$, 8.5 g/l NaCl, pH 7.4). To couple nifW_C, nifZ_C, and nifM_C, 5 mg of BSA and 2.5 mg of peptide (30-fold molar excess) were dissolved in 5 ml PBS. The solution was cooled on ice and 5 ml of a freshly prepared 2% solution of glutaraldehyde in H_2O were added dropwise with continued cooling. After 1 hour, 50 mg of $NaBH_4$ was added and the reaction was left on ice for another hour. No precipitation was observed. The 10 ml sample was dialyzed as above, concentrated by ultrafiltration over a membrane with molecular weight cut-off of 30,000 (YM30, Amicon Corp.), and stored at -80° C. Ovalbumin conjugates were prepared in a similar fashion for nifW_C, nifZ_C, and nifM_C: 1 mg of ovalbumin and 0.5 mg of peptide were suspended in 1.5 ml PBS, the solution was cooled on ice, and 0.6 ml of a freshly prepared 2% solution of glutaraldehyde in water were added dropwise. After 1 hour, 12 mg of $NaBH_4$ was added and the reaction was left on ice for another hour. No precipitation was observed. The sample was worked up by gel filtration on a G25 column and stored at -20° C.

Analysis of Conjugates: All BSA-conjugates were analyzed by amino acid analysis. Sample analysis and quantitation were done at the Biopolymers Laboratory in the Center for Cancer Research at MIT. Both 24- and 96-hour digests were performed and the average value was used in the following analysis. Starting with the experimen-

tally determined amino acid ratios for BSA, a series was calculated using Microsoft Excel, a spreadsheet program on the Apple MacIntosh. In this series, steps were calculated in which 5 moles of peptide were added to one mole of BSA. For each step, the square of the sum of the differences between the predicted and the observed individual amino acid ratios was calculated. A plot of the square versus the number of peptides coupled was plotted and allowed an approximate determination of the number of peptides coupled for each conjugate.

Preparation of Antigens: Antigens were prepared according to the instructions provided with the adjuvant (RIBI ImmunoChem Research, Inc., 1989). Two rabbit doses were prepared by mixing 1.5 mg of BSA-peptide conjugate, 0.5 purified peptide, and PBS to give a final volume of 2 ml. The solution was filter-sterilized before addition to a vial of RIBI Adjuvant. For $nifH_N$, 0.6 mg conjugate, 1 mg of purified peptide, and 0.5 mg of precipitate were mixed and added to a vial of adjuvant. Immediately before injection, the vial containing the antigen was vortexed for 1 min and then 1 ml was withdrawn into a hypodermic syringe.

Immunization and Bleeding Procedures and Schedules: All animal work was done by Chris Hewes, a veterinary technician of the Division of Comparative Medicine at MIT, under Authorization # 89-036. The animals used were white, male New Zealand rabbits. Immunization was by subcutaneous injection (5x0.2 ml). Bleeding was from the ear vein using 0.1 ml acepromezine maleate as a local anesthetic. Pre-immune serum was taken, the rabbits were injected with antigen at 0, 2, 6, and 11 weeks and were bled after 3, 6, and 12 weeks.

Preparation of Antiserum: Typically, a 10 ml bleed was collected in a stoppered test tube, left at room temperature for 1-2 hours, and then centrifuged in a serum desktop centrifuge for 20 min. Alternatively, the blood was allowed to clot overnight at 4° C. In each case, the antiserum was decanted from the blood clot and stored at -80° C.

Assay for Anti-Peptide Antibodies: Screening for antibodies against the BSA-

peptide conjugate was by SDS-PAGE, followed by Western blotting and detection of the antigen by a horse radish peroxidase-based system. Blotting was for 2 hours at 400 mA in the cold. Incubation time with the primary antibodies was for 2 hours with gentle shaking. For a detailed description of the procedures see Appendix A.

^{55}Fe Labeling of Nitrogenase Proteins in *E. coli*: The experiments in this section were done in cooperation with T. White. Bacterial cells carrying *nif*-encoding plasmids were grown, harvested, and cracked as usual with minor modifications. Media composition is as in Appendix A. ^{55}Fe was from NEN, Inc. (1 mCi/ml, $\tau_{1/2}$ = 2.6 years, decay by electron capture). The appropriate amounts of phosphates, NH_4Cl, NaCl, Casamino acids, and glucose were dissolved in water. Chelex cation exchanger (Na-form, BIORAD, Inc.) was added to 1.5% (w/v), the mixture was shaken gently for 1 hour, and the resin was allowed to settle for 1 hour. The "Chelexed" medium was steri-filtered into a nitric acid-rinsed, autoclaved 2 l flask. In this flask, the medium was complemented with the usual amounts of $MgSO_4$, $CaCl_2$, micronutrients, Na_2MoO_4 , and antibiotics. Fe-citrate was added to 2.9 μg/ml final concentration. Aliquots of 250 ml were transferred to nitric acid-rinsed, autoclaved anaerobic growth flasks. Per flask, 0.5 ml of a mid-log starter culture in LB medium and 0.5 ml of ^{55}Fe-citrate (1 mCi/ml) were added. Cultures were then grown anaerobically at 30° C overnight. Harvesting was as usual (see Appendix A), except that the cell pellet after the first centrifugation was resuspended in "cold", anaerobic buffer (0.1 mol/l HEPES, 5 mmol/l dithionite) and recentrifuged. This wash was intended to remove most of the extraneous ^{55}Fe not incorporated into nitrogenase proteins. Cells were then broken in cracking buffer: 100 mmol/l HEPES (pH 7.4), 100 mmol/l dithiothreitol, 2 mmol/l cysteamine, 5 mmol/l dithionite. After native gel electrophoresis, gels were dried down immediately and were exposed to Kodak SB5 X-ray film. Exposure time varied from overnight to 6-8 days. Development of the X-ray film was according to the manufacturer's instructions: 5 min in GBX developer, 30 s rinse, 5-8 min in GBX fixer, 10 min rinse (all at 20-25° C).

C. Results

Sequence Determination of Three *nif* Genes: The original objective of this work was to determine the nucleotide sequence encoding for what was then thought to be *nif*M on the plasmid pKH733. As the work progressed, it became apparent that additional sequence adjacent to *nif* M would have to be determined.

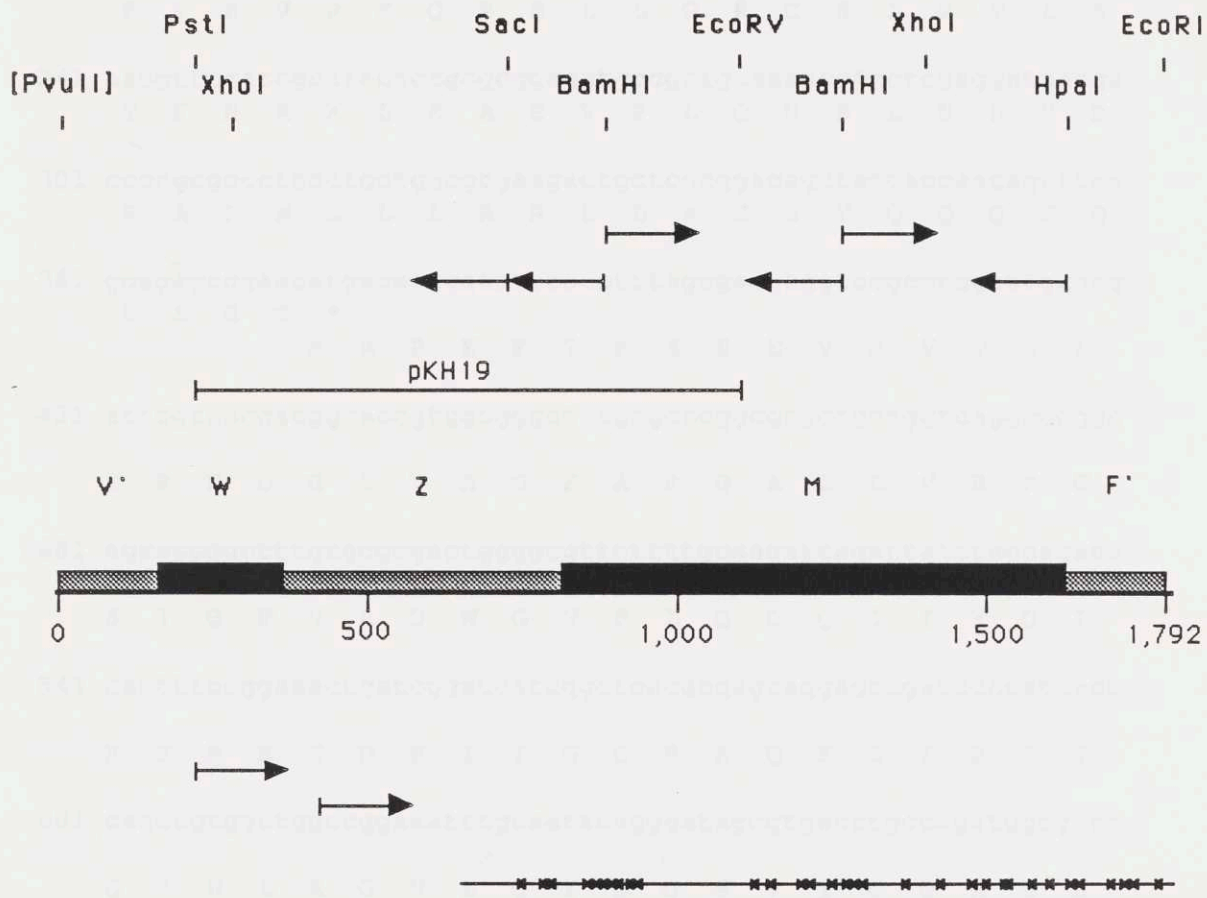


Figure II.1.: Sequencing Map of pVL13

Therefore, the scope of the sequence work was expanded to cover all of the *nif* - EcoRI insert in pVL13. This insert carries an additional 400 base pairs when compared to the *nif* M fragment in pKH733. Consequently, the complete nucleotide sequence of the EcoRI insert of pVL13 was determined (Figure II.1 and Figure II.2.).

```
  1 tgctgcccgccatccgccgcttcgccgagaactggaagcgcagcccgaaagattacgagc
     L  P  A  I  R  R  F  A  E  N  W  K  R  S  P  K  D  Y  E  L

 61 tggtggctatctacgacgagctgtgcggtgaatccgctctgcgggcgaggggggtaatgat
                                                          M  M
     V  A  I  Y  D  E  L  C  G  E  S  A  L  R  A  R  G  *

121 ggagtggttttatcaaattcccggcgtggacgaacttcgctccgccgaatctttttttca
     E  W  F  Y  Q  I  P  G  V  D  E  L  R  S  A  E  S  F  F  Q

181 gtttttcgccgtcccctatcagcccgagctgcttggccgctgcagcctgccggtgctggc
     F  F  A  V  P  Y  Q  P  E  L  L  G  R  C  S  L  P  V  L  A

241 aacgtttcatcgcaaactccgcgcggaggtgccgctgcaaaaccggctcgaggataacga
     T  F  H  R  K  L  R  A  E  V  P  L  Q  N  R  L  E  D  N  D

301 ccgcgcgccctggctgctggcgcgaagactgctcgcggagagctatcagcaacagtttca
     R  A  P  W  L  L  A  R  R  L  L  A  E  S  Y  Q  Q  Q  F  Q

361 ggagagcggaacatgagaccgaaattcacctttagcgaagaggtccgcgtcgtacgcgcg
     E  S  G  T  *
                M  R  P  K  F  T  F  S  E  E  V  R  V  V  R  A

421 attcgtaacgacggcaccgtggcgggcttcgcgcccggcgcgctgctggtcaggcgcggc
     I  R  N  D  G  T  V  A  G  F  A  P  G  A  L  L  V  R  R  G

481 agcaccggctttgtgcgcgactggggcgtttttttgcaagatcagattatctaccagatc
     S  T  G  F  V  R  D  W  G  V  F  L  Q  D  Q  I  I  Y  Q  I

541 cactttccggaaaccgatcggatcatcggctgccgcgagcaggagctgatccccatcacc
     H  F  P  E  T  D  R  I  I  G  C  R  E  Q  E  L  I  P  I  T

601 cagccgtggctggccggaaatttgcaatacagggatagcgtgacctgccagatggcgctc
     Q  P  W  L  A  G  N  L  Q  Y  R  D  S  V  T  C  Q  M  A  L

661 gcggtcaacggcgatgtggtcgtgagcgccggccagcggggacgcgttgaggctaccgat
     A  V  N  G  D  V  V  V  S  A  G  Q  R  G  R  V  E  A  T  D

721 cggggagagctcggcgacagctacaccgtcgactttagcggccgctggttcagggtcccg
     R  G  E  L  G  D  S  Y  T  V  D  F  S  G  R  W  F  R  V  P

781 gtgcaggccatcgcccttatagaggaaagagaagaatgaacccatggcaacgttttgccc
                                          M  N  P  W  Q  R  F  A  R
     V  Q  A  I  A  L  I  E  E  R  E  E  *
```

```
 841 ggcagcggctggcgcgcagccgctggaatcgcgatccggcggccctggatccggccgata
       Q  R  L  A  R  S  R  W  N  R  D  P  A  A  L  D  P  A  D  T

 901 cgccggcttttgaacaggcctggcaacgccagtgccatatggagcagacgatcgtcgcgc
       P  A  F  E  Q  A  W  Q  R  Q  C  H  M  E  Q  T  I  V  A  R

 961 gggtccctgaaggcgatattccggcggcgttgctggagaatatcgctgcctcccttgcca
       V  P  E  G  D  I  P  A  A  L  L  E  N  I  A  A  S  L  A  I

1021 tctggctcgacgagggggattttgcgccgcccgagcgcgctgccatcgtgcgccatcacg
       W  L  D  E  G  D  F  A  P  P  E  R  A  A  I  V  R  H  H  A

1081 cccggctggaactcgccttcgccgatatcgcccgccaggcgccgcagccggatctctcca
       R  L  E  L  A  F  A  D  I  A  R  Q  A  P  Q  P  D  L  S  T

1141 cggtacaggcatggtatctgcgccaccagacgcagtttatgcgcccggaacagcgtctga
       V  Q  A  W  Y  L  R  H  Q  T  Q  F  M  R  P  E  Q  R  L  T

1201 cccgccatttactgctgacggtcgataacgaccgcgaagccgtgcaccagcggatcctcg
       R  H  L  L  L  T  V  D  N  D  R  E  A  V  H  Q  R  I  L  G

1261 gcctgtatcggcaaatcaacgcctcgcgggacgctttcgcgccgctggcccagcgccatt
       L  Y  R  Q  I  N  A  S  R  D  A  F  A  P  L  A  Q  R  H  S

1321 cccactgcccgagcgcgctggaagagggtcgtttaggctggattagccgtggcctgctct
       H  C  P  S  A  L  E  E  G  R  L  G  W  I  S  R  G  L  L  Y

1381 atccgcagctcgagaccgcgctgttttcactggcggaaaacgcgctaagccttcccatcg
       P  Q  L  E  T  A  L  F  S  L  A  E  N  A  L  S  L  P  I  A

1441 ccagcgaactgggctggcatcttttatggtgcgaagcgattcgccccgccgcgcccatgg
       S  E  L  G  W  H  L  L  W  C  E  A  I  R  P  A  A  P  M  E

1501 agccgcagcaggcgctggagagcgcgcgcgattatctttggcagcagagccagcagcgcc
       P  Q  Q  A  L  E  S  A  R  D  Y  L  W  Q  Q  S  Q  Q  R  H

1561 atcagcgccagtggctggaacagatgatttcccgtcagccgggactgtgcgggtagcctc
       Q  R  Q  W  L  E  Q  M  I  S  R  Q  P  G  L  C  G  *

1621 ggcggctacccgttaacgcctacagcacggtgcgtttaatctcctcaagccagctcgcca
                          *  L  V  T  R  K  I  E  E  L  W  S  A

1681 gacgcgcttcggtctggtcgaactggttatcctgatccagcaccagcccaacaaagcggt
         R  A  E  T  Q  D  F  Q  N  D  Q  D  L  V  L  G  V  F  R

1741 cgccttccagcgccgaggacgcgctgaattc
     *  G  E  L  A  S  S  A  S  F  E
```

Figure II.2.: DNA Sequence and Translation of *nif*V'WZMF'
Deposited as KPNIF01 at GenEmbl under accession number M24106

Analysis of the codon usage using CODONPREFERENCE (Gribskov et al., 1984) revealed the existence of three open reading frames with high coding probabilities between the previously sequenced *nif* genes *nif* F and *nif* V (Figure II.3.).

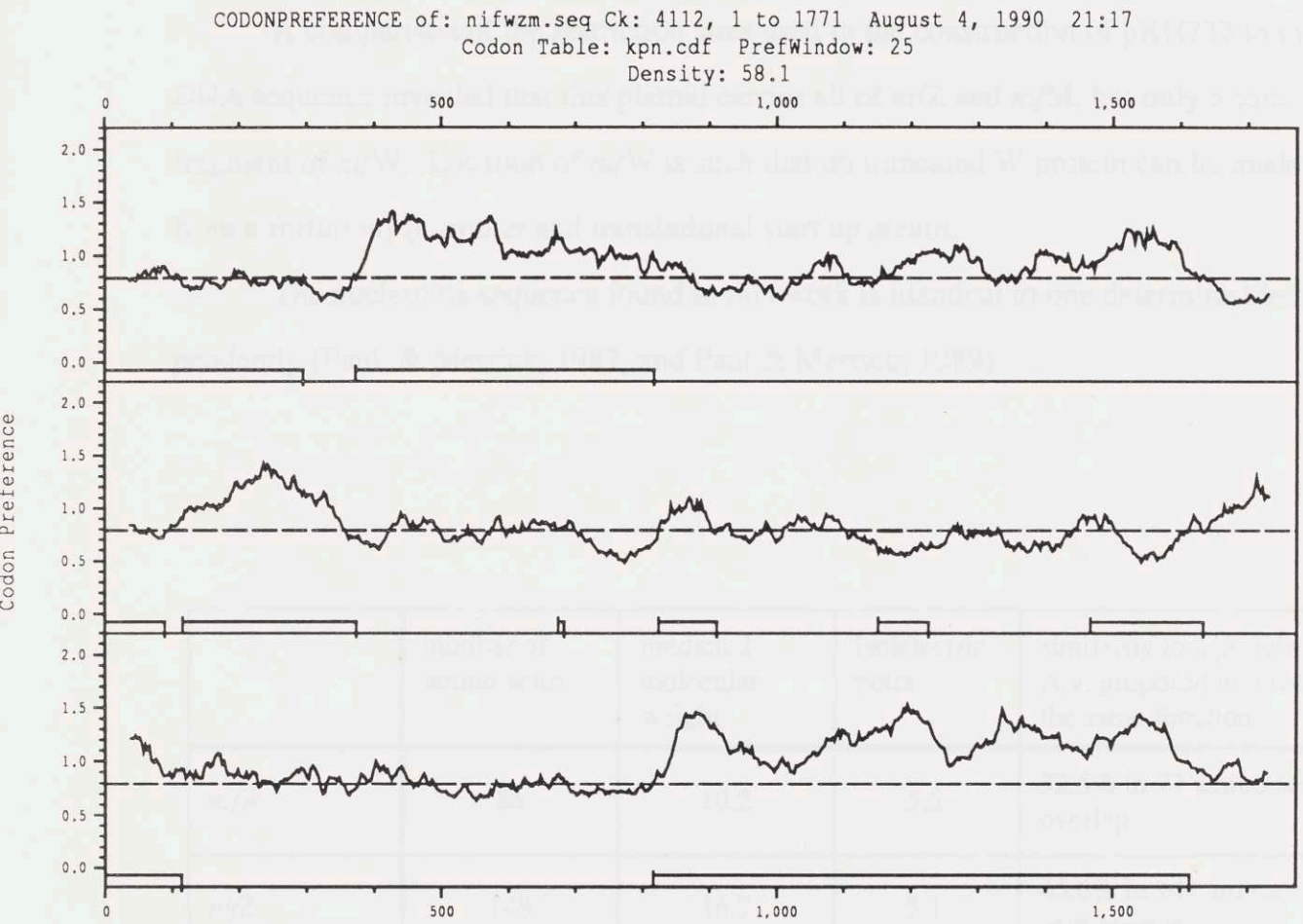


Figure II.3.: Codon Frequency Analysis of pVL13 Insert

Unexpectedly, use of a codon frequency file derived from genes in *E. coli* did not result in any significant pattern: a codon frequency file had to be constructed from the known sequences of *nif*H, *nif*D, and *nif*K in *K. pneumoniae* for use in CODONPREFERENCE in order to obtain interpretable results. In accord with the SDS-

PAGE-derived size data, the largest open reading frame was identified as *nif*M; the two smaller ones were named n*if*Z and *nif*W. In addition to the *nif*M promoter identified by S1 mapping (Beynon, et al., 1983), another promoter agreeing with the CTGGYAYR-N_4-TTGCA consensus pattern was identified in *nif*W, presumably activating transcription of *nif*Z (Ausubel, 1983). Both promoters are thought to act in conjunction with the strong *nif*U promoter.

A comparison of the restriction sites used in the construction of pKH733 to the DNA sequence revealed that this plamid carries all of *nif*Z and *nif*M, but only a small fragment of *nif*W. Location of *nif*W is such that no truncated W protein can be made from a fortuitous promoter and translational start upstream.

The nucleotide sequence found in this work is identical to one determined independently (Paul, & Merrick, 1987, and Paul & Merrick, 1989).

	number of amino acids	predicted molecular weight	isolelectric point	similarity to a protein in A.v. proposed to have the same function
nifW	86	10.2	5.2	32.5% in 77 amino acid overlap
nifZ	148	16.7	5.1	45.6% in 147 amino acid overlap
nifM	266	30.6	7.1	30.2% in 199 amino acid overlap
Calculations were performed using the UWGCG Program Package, Version 6.				

Table II.1.: Protein Data Summary for NifW, NifZ, and NifM

Analysis of The Open Reading Frames: The three open reading frames are thought to encode three proteins named NifW, NifZ, and NifM; all three proteins show significant sequence similarity to the predicted amino acid sequences of three *nif* proteins in *A. vinelandii* (Jacobson et al., 1989a, Table II.1.). The calculated molecular weight of the *nif*M protein (30.6 kd) is in reasonable agreement with experimental data from two-dimensional electrophoresis (27 kd; Roberts & Brill, 1980) and SDS-PAGE (28 kd; Pühler & Klipp, 1982). No N-terminal sequence data is available to determine whether the small difference between calculated and experimentally determined molecular weight is due to post-translational modification. Interestingly, the *nif*M protein shows some similarity to the predicted amino acid seqence of the positive regulator of the maltose operon, the *mal*T protein (Cole & Raibaud, 1986). The region showing strongest similarity, however, is not the one thought to constitute the DNA binding site (Figure II.4.).

```
27.4% identity in 106 aa overlap
           80        90       100       110       120       130
Kpnifm GDFAPPERAAIVRHHARLELAFADIARQAPQPDLSTVQAWYLRHQTQFMRPEQRLTRHLL
                                        | |   |||    |         | |
Ecomal LNHAWSLFNHSELSLLEESLKALPWDSLLENPQLVLLQAWLMQSQHRYGEVNTLLARAEH
       380       390       400       410       420       430

           140       150       160       170       180       190
Kpnifm LTVDNDREAVHQRILGLYRQINASRDAFAPLAQRHSHCPSALEEGRLGWI-SRGLLYPQL
         |       |     | |   |  |     | |       ||||   ||  ||       |
Ecomal EIKDIREDTMHAEFNAL-RAQVAINDGNPDEAERLAK-LALEELPPGWFYSRIVATSVL
       440       450       460       470         480       490

           200       210       220       230       240       250
Kpnifm ETALFSLAENALSLPIASELGWHLLWCEAIRPAAPMEPQQALESARDYLWQQSQQRHQRQ
         |    |   ||
Ecomal GEVLHCKGELTRSLALMQQTEQMARQHDVWHYALWSLIQQSEILFAQGFLQTAWETQEKA
         500       510       520       530       540       550
```

Figure II.4.: Alignment of NifM and MalT

Functions of *nif*W, *nif*Z, and *nif*M: The functions of *nif*W, *nif*Z, and *nif*M are also the object of T. White's doctoral thesis and are described elsewhere, with the excep-

tion of the ^{55}Fe experiments, which are described below (White, T., Ph.D. Thesis, 1990, and Harris et al., 1990). Briefly, her experiments confirm Dr. K. Howard's conclusions and show that, of the *nif* proteins, only the product of the *nif*M coding region is needed for Fe protein synthesis in addition to the structural gene, *nif*H (Table II.2.).

Preparation of Anti-Peptide Antibodies: In order to determine whether the open reading frames found by DNA sequencing actually encode for proteins, polyclonal

Plasmid (+ nifA)	nif genes	specific activity for Kp2 (nmol/[minxmg])
UN	all	71
pKH733	HZM	43.5/2.8
pKH733∂Z	HM	60.3/3.9
pKH733∂M	HZ	ND
ND.....≤ 0.02 Numbers behind / are standard deviations.		

Table II.2.: Specific Activities of *nif*Z and *nif*M Deletions in pKH733

antibodies were raised in white New Zealand rabbits against the C-terminus of each predicted protein sequence: nifW_C, nifZ_C, and nifM_C. As a positive control, antibodies were also raised against the N-terminal heptadecapeptide of the iron protein of nitrogenase, nifH_N. The choice of peptide was influenced by the following considerations: the termini of proteins are usually exposed to the solvent and show increased flexibility (Tainer et al., 1984); coupling to carrier protein is straight-forward if the N-terminal amino group is the only one in the peptide; and processing is infrequent at the C-terminus. For the iron protein it is known that there is no N-terminal processing. Hydrophilicity, flexibility, and antigenic index for the proteins encoded by *nif*W, *nif*Z, *nif*M, and *nif*H

were calculated using PLOTSTRUCTURE of the UWGCG Program Package and confirmed, within the limits of the computational techniques, that the sequences chosen are suitable epitopes according to the above criteria (Kyte & Doolittle, 1982, and Jameson & Wolf, 1988, and Wolf et al., 1988).

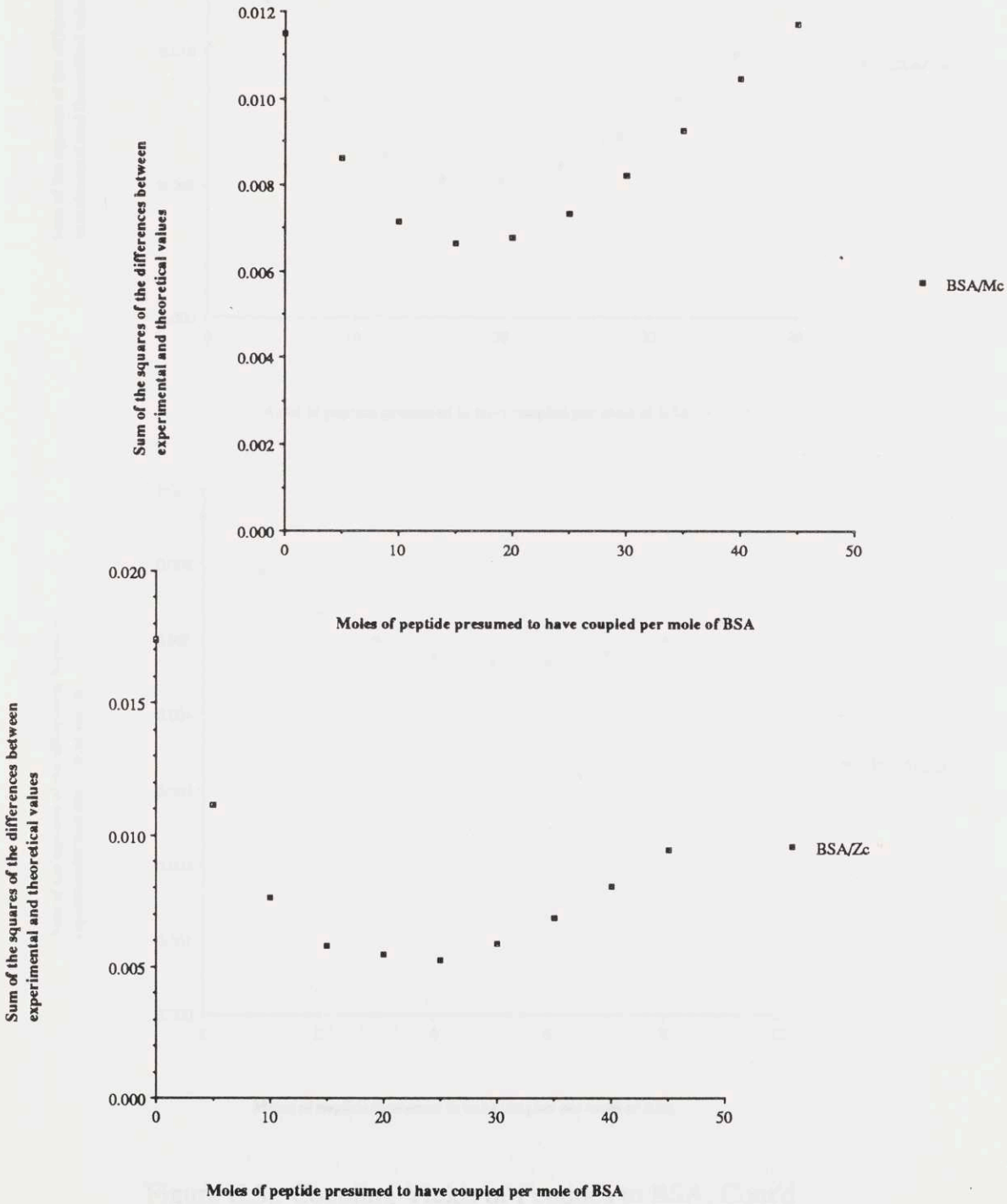


Figure II.5.: Coupling Yields of Peptides to BSA

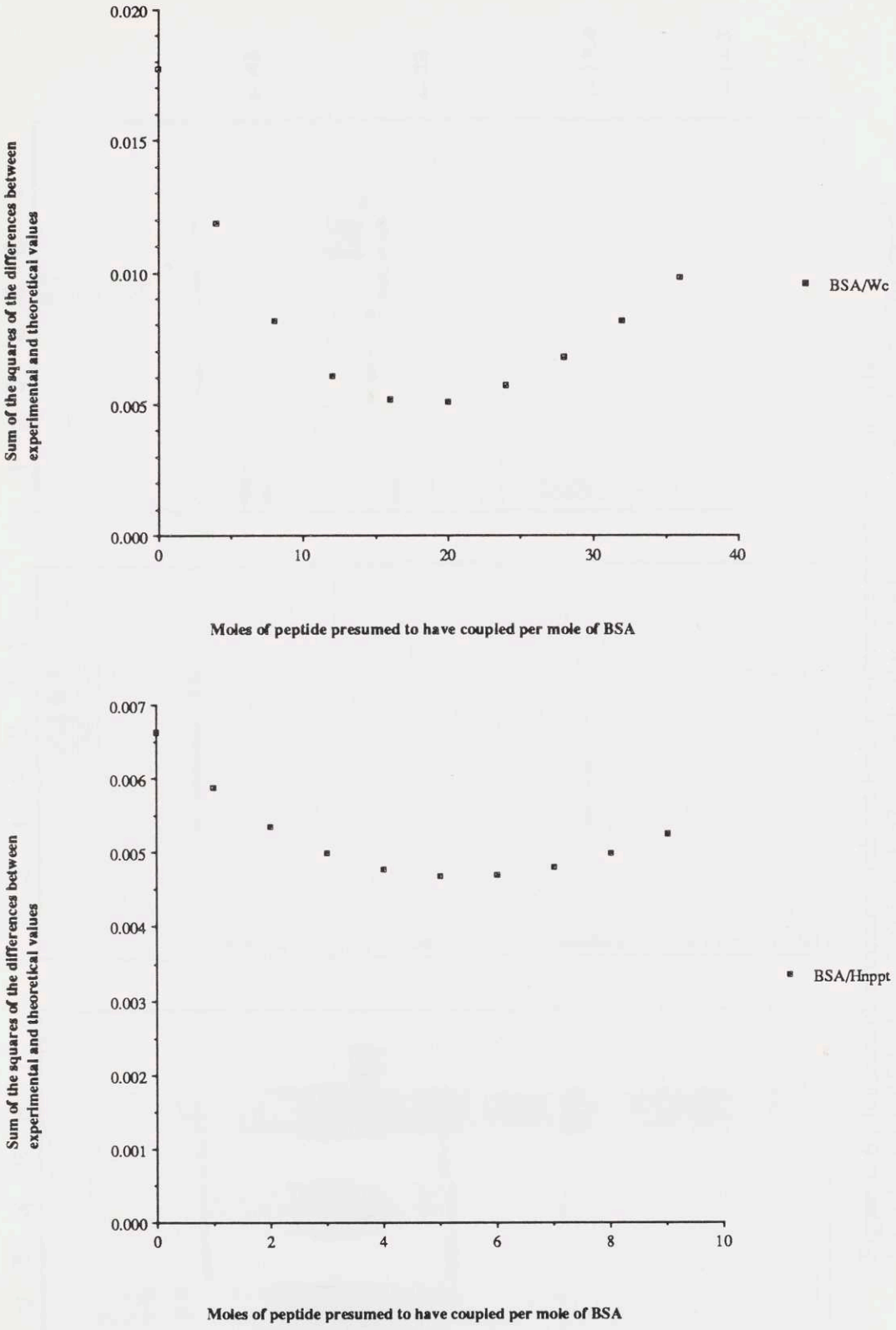


Figure II.5.: Coupling Yields of Peptides to BSA, Cont'd

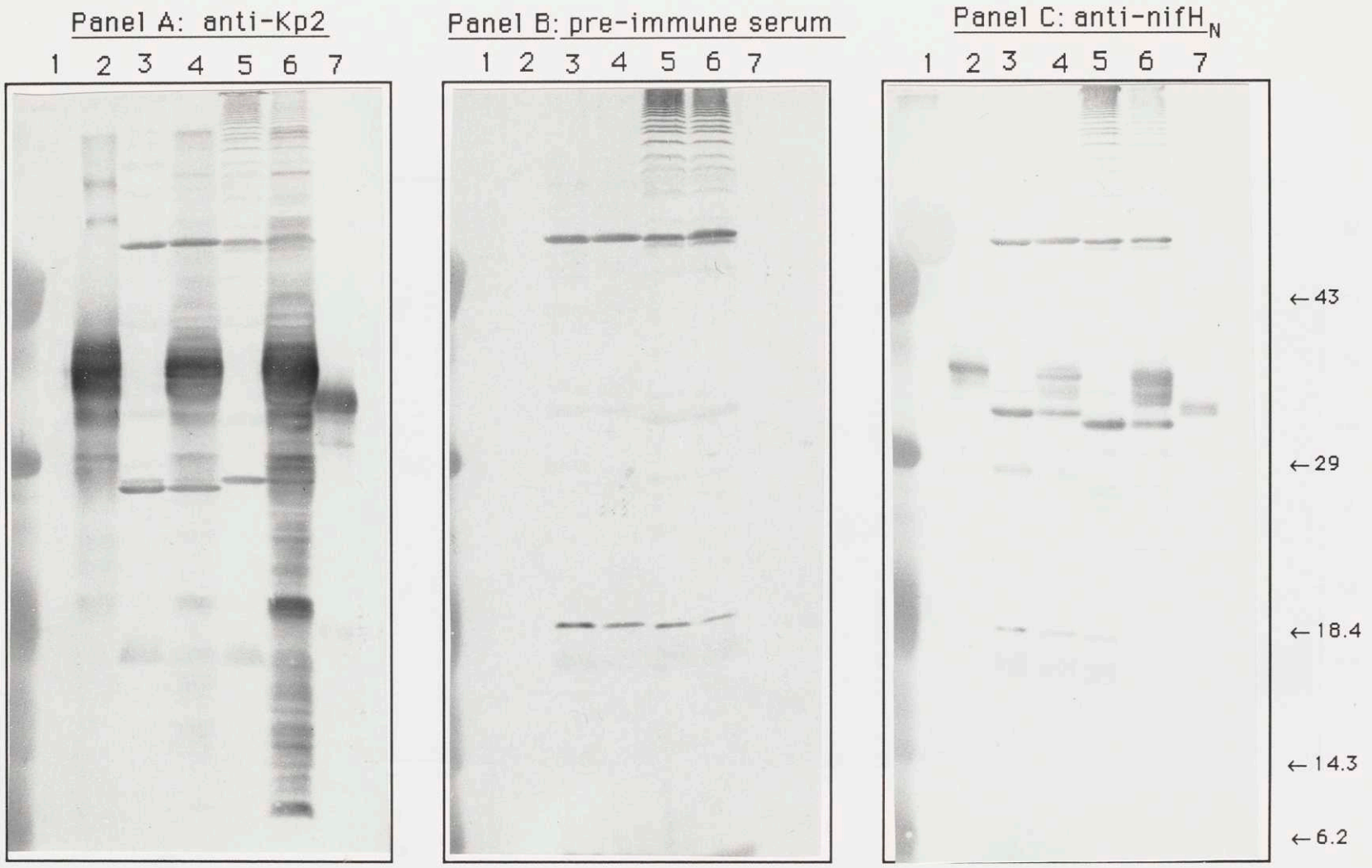


Figure II.6.: Immunoblot probing with antiKp2, pre-immune serum and anti-nifHN
Lane 1: 20 µg BSA, Lane 2: 0.5 µg Kp2, Lane 3: 20 µg UN[repr.], Lane 4: 20 µg UN[derepr.],
Lane 5: 20 µg pACYC184/pVL15, Lane 6: 20 µg pGH1/pVL15, Lane 7: 0.5 µg Av2

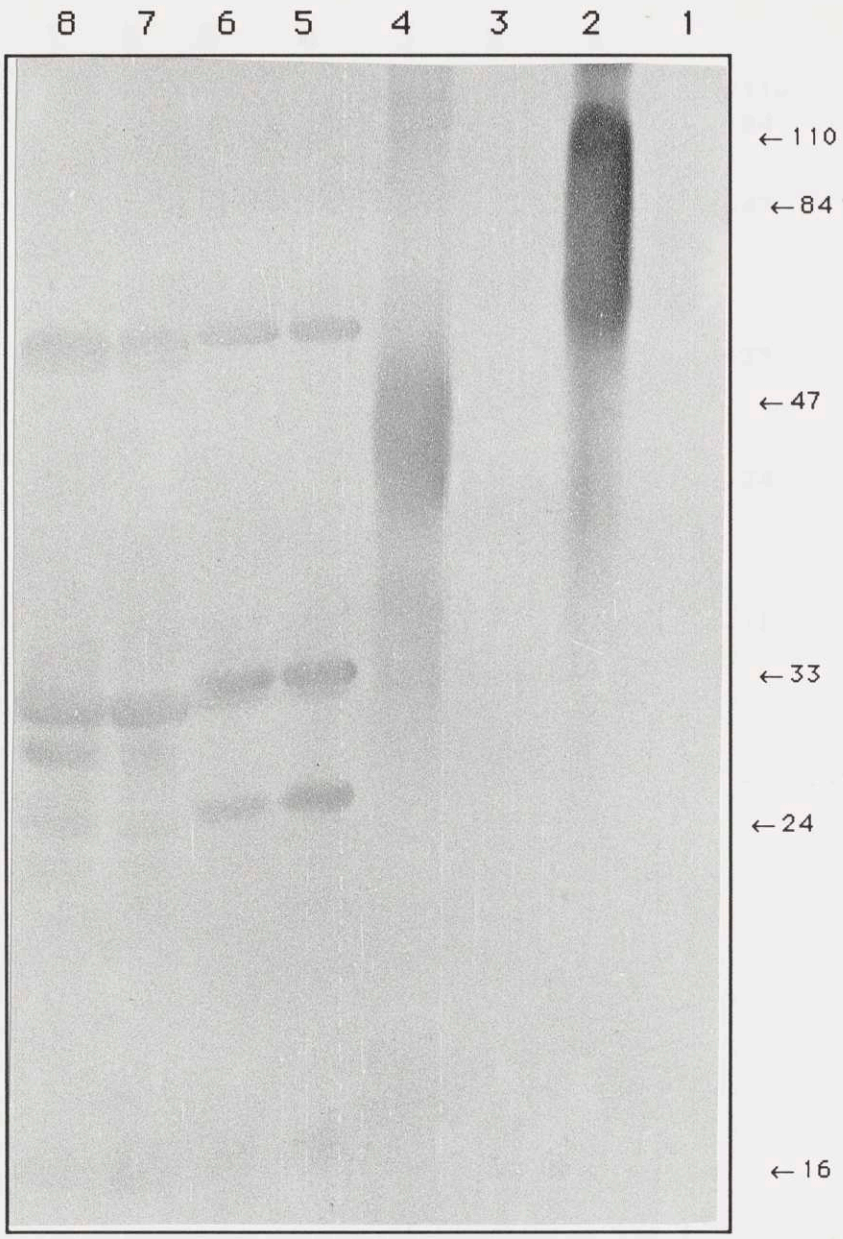


Figure II.7.: Immunoblot probing with anti-nifMc
Lane 1: 5 µg BSA, Lane 2: 5 µg BSA-nifMc conjugate, Lane 3: 5 µg ovalbumin, Lane 4: 5 µg ovalbumin-nifMc conjugate, Lane 5: 40 µg UN[repr.], Lane 6: 40 µg UN[derepr.], Lane 7: 40 µg pACYC184/pVL15, Lane 8: 40 µg pGH1/pVL15.

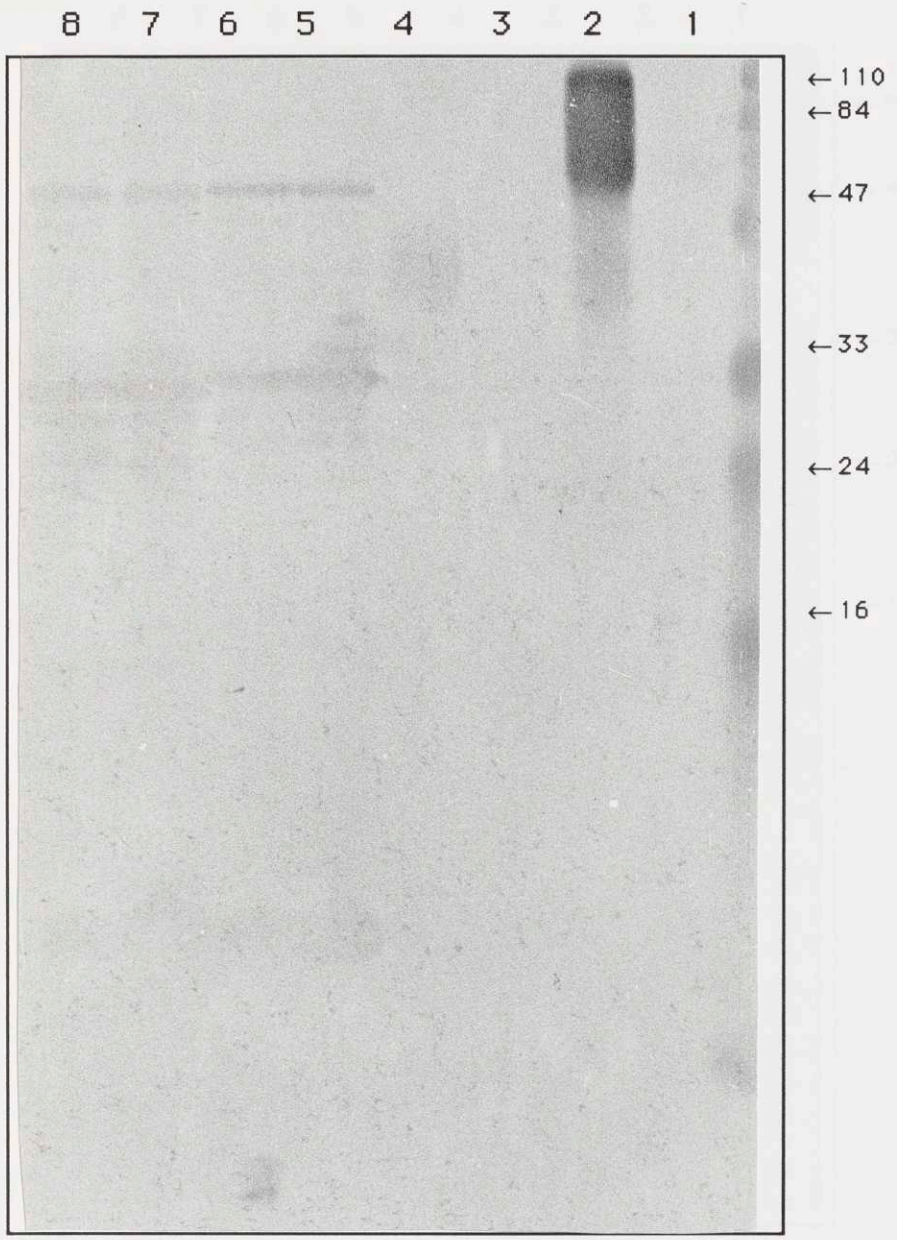


Figure II.8.:Immunoblot probing with anti-nifZc
Lane 1: 5 µg BSA, Lane 2: 5 µg BSA-nifZc conjugate, Lane 3: 5 µg ovalbumin, Lane 4: 5 µg ovalbumin-nifZc conjugate, Lane 5: 40 µg UN[repr.], Lane 6: 40 µg UN[derepr.], Lane 7: 40 µg pACYC184/pVL15, Lane 8: 40 µg pGH1/pVL15.

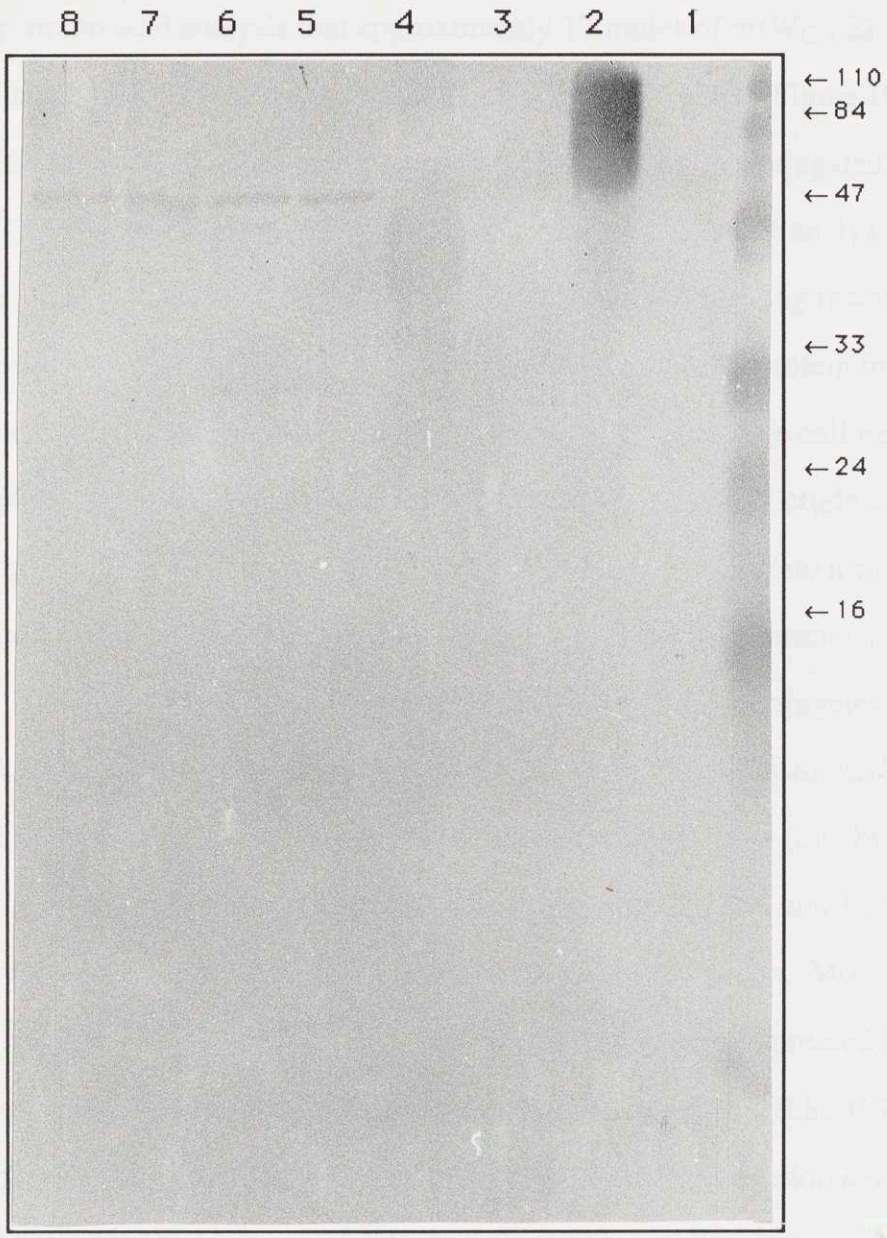


Figure II.9.: Immunoblot probing with anti-nifWc
Lane 1: 5 µg BSA, Lane 2: 5 µg BSA-nifWc conjugate, Lane 3: 5 µg ovalbumin, Lane 4: 5 µg ovalbumin-nifWc conjugate, Lane 5: 40 µg UN[repr.], Lane 6: 40 µg UN[derepr.], Lane 7: 40 µg pACYC184/pVL15, Lane 8: 40 µg pGH1/pVL15.

Peptides were synthesized, purified, and coupled to BSA as carrier protein. It was determined by amino acid analysis that approximately 17 moles of nifW_C , 22 moles of nifZ_C, and 18 moles of nifM_C had been coupled to one mole of BSA (Figure II.5.). Polyclonal antibodies were successfully raised against the peptides conjugated to the carrier proteins as determined by SDS-PAGE followed by immunoblot analysis (Figures II.6., II.7., II.8., II.9.). Reactivity of the anti-nifH_N antiserum was strong towards purified iron proteins from *K. pneumoniae* and *A. vinelandii*, towards Fe protein in crude cell extracts of derepressed *K. pneumoniae*, and towards Fe protein in crude cell extracts of *E. coli* carrying pGH1/pVL15. Repressed extracts of *K. pneumoniae* and crude cell extracts *E. coli* carrying pACYC184/pVL15, a plasmid combination producing no iron protein, showed no reactivity other than the reactivity found also with the pre-immune serum (Figure II.5.). For antisera raised against nifM_C, nifW_C, and nifZ_C conjugates the results were identical, although negative: all antisera showed strong reactivity towards the conjugate and somewhat less reactivity towards a secondary conjugate (ovalbumin-peptide). The reactivity of the antisera towards crude cell extracts UN and *E. coli* systems carrying *nif* plasmids is no different than that of pre-immune sera. More telling, there is no additional reactivity in the lanes containing extracts of derepressed UN cultures when compared to those of repressed UN cultures (Figures II.7., II.8., II.9.).

Analysis of ^{55}Fe Labelled Cultures: Cultures of various deletion derivatives of pKH733 were grown in the presence of ^{55}Fe to follow up on preliminary results by Howard et al. suggesting the presence of a clusterless Fe protein in the absence of the *nif*M protein. The labeled *nif* proteins were separated by native gel electrophoresis and analyzed by immunoblotting and autoradiography (Figures II.10. and II.11.). It became apparent that the absence of *nif*M on pKH733 reduced the steady-state levels of assembled Fe protein inside the cell.

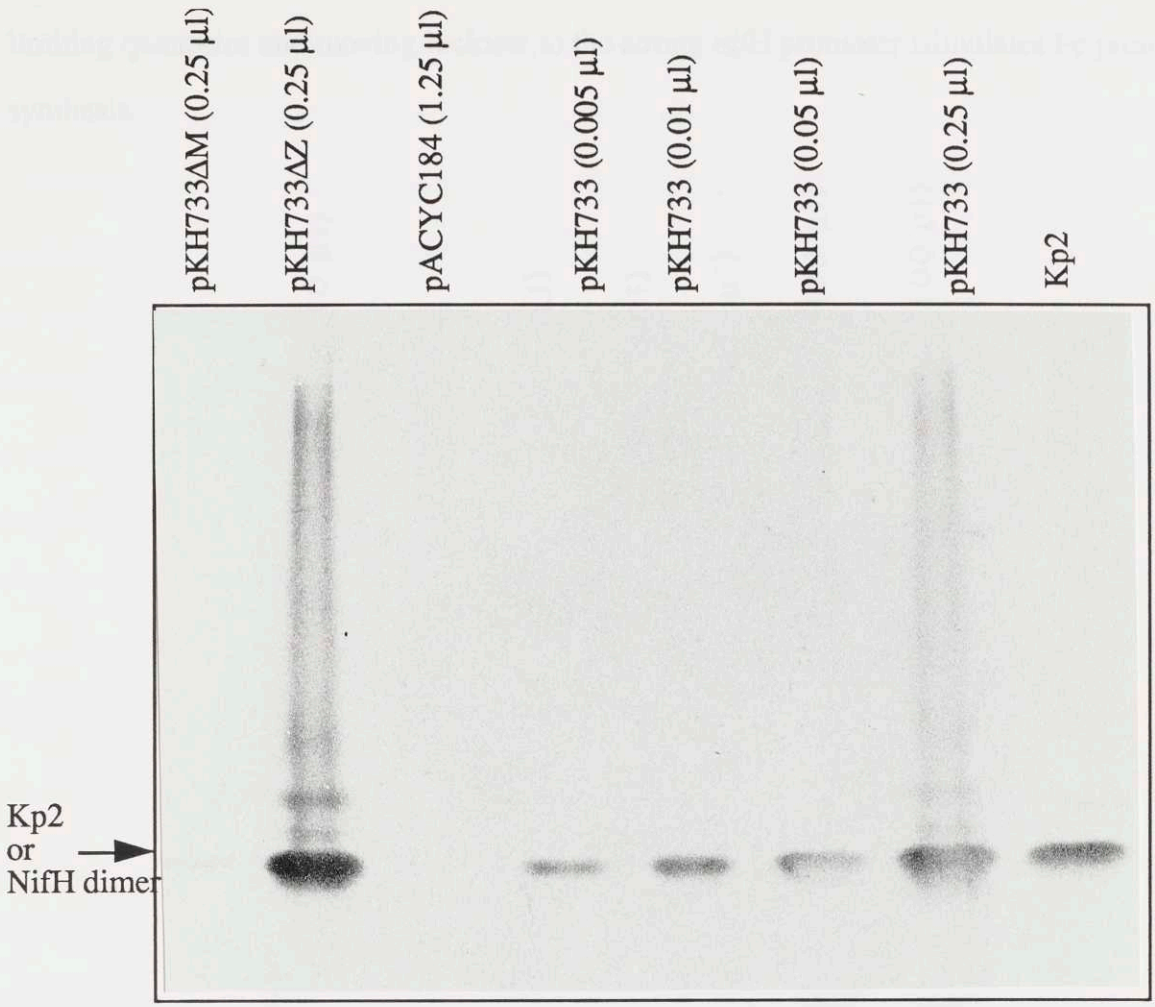


Figure II.10.: Western Blot Analysis of ^{55}Fe Labeled *E. coli* Extracts
Extracts contain the activator plasmid pVL15 in addition to the expression plasmids. Volumes, in µl, are in parentheses.

Therefore, efforts were made to evaluate data from autoradiograms only in connection with dilution results from immunoblotted native gels. SDS-PAGE gels were found to be misleading in that they consistently showed larger amounts of peptide than native gels (as always, in comparison to the corresponding parent construct). The deletion of *nif*Z, however, has no detrimental effect on the Fe protein activity. Slight increases in activity and steady-state levels of Fe protein can be accounted for by the distance effect on pKH733ΔZ: in this construct, the *nif*M protein is apparently made in

limiting quantities and moving it closer to the strong *nif*H promoter stimulates Fe protein synthesis.

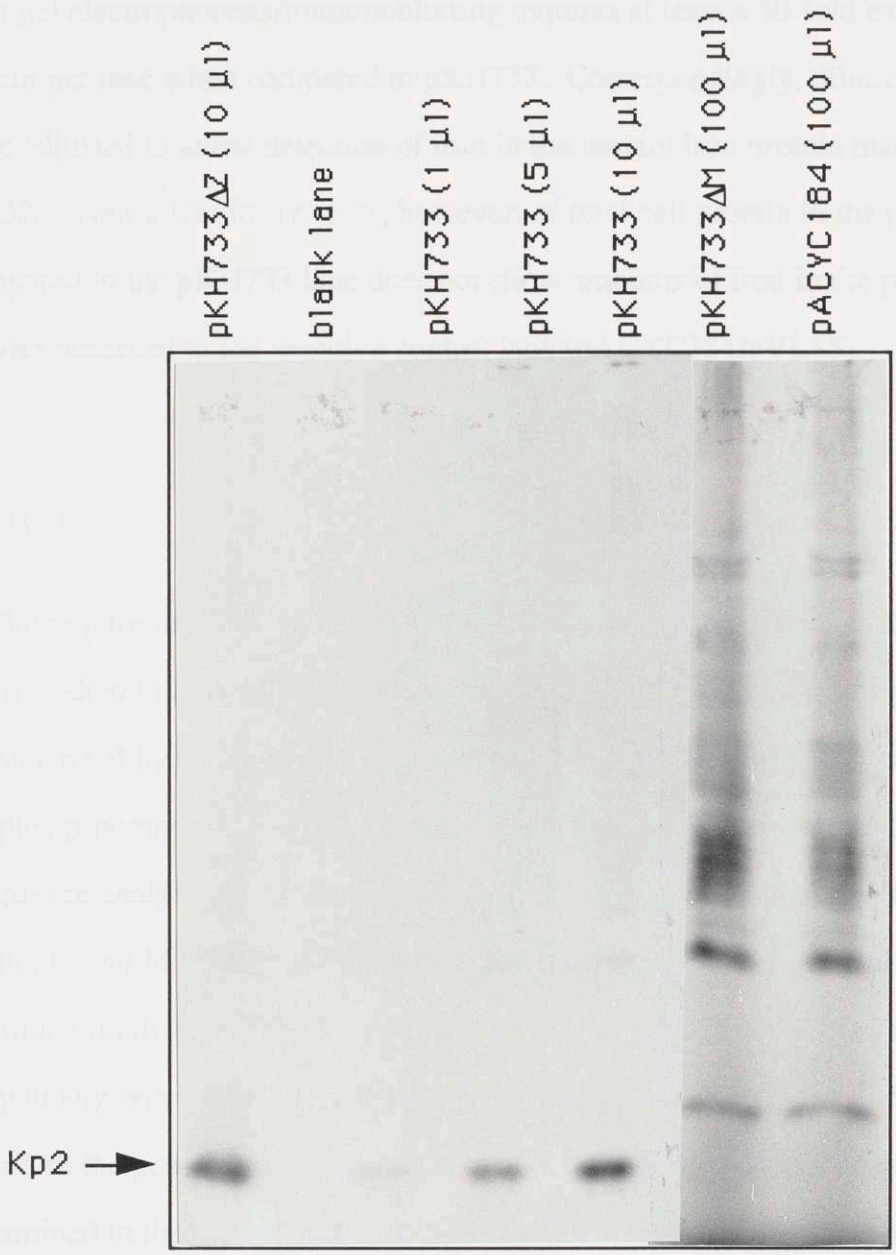


Figure II.11.: ^{55}Fe Autoradiograph of *E. coli* Extracts
Extracts containing activator plasmid pVL15 with expression plasmids were analyzed by anaerobic, native gel electrophoresis. Exposure was for 6 days. Volumes loaded, in µl, are shown in parentheses.

The effect of deleting *nif*M, however, is dramatic: the steady-state level of the Fe protein falls precipitously. Detection of the Fe protein dimer derived from pKH733ΔM by native gel electrophoresis/immunoblotting requires at least a 50-fold excess of total cell protein per lane when compared to pKH733. Correspondingly, dilutions on native gels were adjusted to allow detection of iron in the mutant iron protein made off pKH733ΔM: even a 100-fold excess, however, of total cell protein in the pKH733ΔM lane compared to the pKH733 lane does not show amounts of iron in the position for Kp2 above those observed in the negative control lane (pACYC184/pVL15).

D. Discussion

The requirements for the synthesis and maturation of the Fe protein of nitrogenase have been studied in *E. coli* by expressing a minimal set of *nif* genes on a plasmid that was characterized by DNA sequencing. It was found that the maturation of the Fe protein is a complex process involving protein factors other than its own polypeptide, NifH. DNA sequence analysis, in combination with deletion mutagenesis, has identified a 30.6 kd protein , the *nif* M protein, as the maturation factor necessary in *E. coli* to process the *nif* H peptide into the holo Fe protein. The *nif* M protein does not display significant similarity to any previously sequenced protein or gene other than to a poorly characterized region of the positive regulator of the maltose operon in *E. coli*, the *mal*T protein. It was determined in these experiments that the level of *nif* H peptide inside the cell as judged by SDS-PAGE is a poor measure for the actual levels of properly folded and assembled Fe protein. Quantitation of mutant Fe proteins that are inactive in the standard acetylene reduction assay therefore needs to take into account that these proteins may or may not be present inside the cell in a form amenable to the usual biochemical analysis. Analysis of crude cell extracts labeled with ^{55}Fe taking into account these quantification

problems showed that in the absence of the *nif* M protein a clusterless apoFe protein is synthesized that migrates like holoFe protein on anaerobic native gels. Since the *nif*M protein is only made in vanishingly small amounts, i.e., in amounts much smaller than the Fe protein, that it is therefore probably acting catalytically, and that iron is present in the Fe protein only in the presence of the *nif* M protein, it seems reasonable to assume that it is the *nif* M protein that is responsible for the insertion of the Fe_4S_4-cluster into the apoFe protein (Figure II.12.).

The role of the *nif*M protein may well not be restricted to the maturation of the Fe protein. Work done by Dr. G. Harris and T. White in *E. coli* has shown that while a deletion in *nif*H completely abolishes apoMoFe protein activity, a deletion in *nif*M only leads to a 50% loss in activity. This indicates that it is not the active Fe protein that is essential for apoMoFe protein synthesis, but the *nif*H peptide (which is present in the ΔM but not in the ΔH system). From these experiments, however, it cannot be excluded that the effect of the deletion of *nif*M is not only through the Fe protein/*nif*H peptide, but that a deletion in *nif*M may also have a direct effect on apoMoFe protein synthesis.

During the early stages of this work, two new, previously unmapped open reading frames, *nif*W and *nif*Z, were identified by DNA sequencing. Located between *nif*V and *nif*M, these potential genes show strong similarity to open reading frames in *A. vinelandii* on the protein level. In *A.v.*, the corresponding open reading frames are also located near the *nif*USV operon; *nif*W is part of this operon, as no additional promoter is found up-stream. *nif*M may well be the only gene in its operon: a functional *nif* promoter has been identified by S1-mapping (Beynon et al., 1983). *nif*Z has a *nif*-consensus promoter upstream, however, the S1-mapping experiments did not pick it up. Therefore it is not clear, whether this is indeed a functional promoter.

Deletion analysis in a plasmid system has shown that these orf's are not essential for Fe protein synthesis in *E. coli*, but that they have an ancillary function in apoMoFe synthesis. Transcriptionally, *nif*W, *nif*Z, and *nif*M are coupled to the strong promoter of

the *nif*USV operon by overlap of the upstream start and stop codons. Additional promoters in *nif*W and *nif*Z were identified as *nif* consensus promoters and presumably direct additional transcription of *nif*Z and *nif*M.

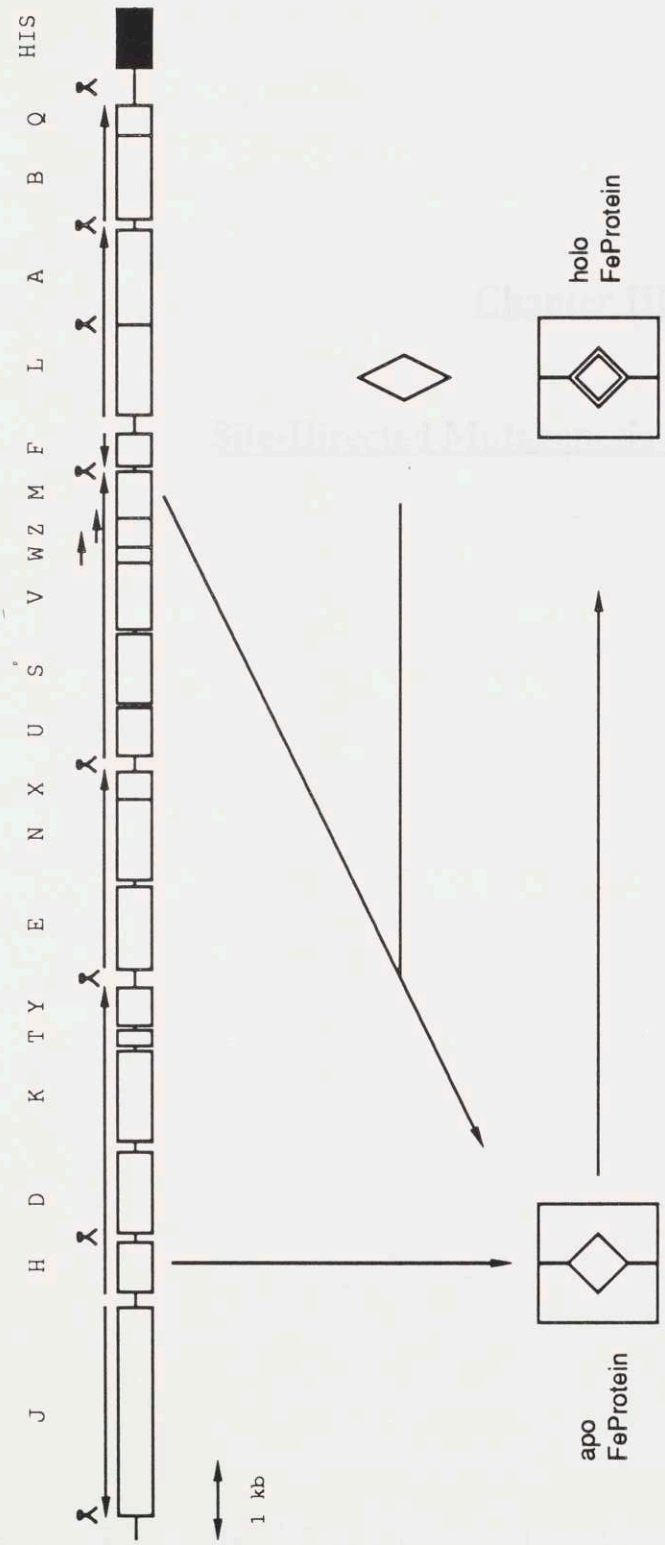


Figure II.12.: Fe Protein Synthesis

The ◊ signifies the ill-defined precursor of the Fe_4S_4-cluster before it is acted on by NifM and converted to the Fe_4S_4-cluster in the Fe protein of nitrogenase. The apoFe protein is the Fe protein-like species as detected by native gel electrophoresis.

Chapter III:

Site-Directed Mutagenesis of the Fe Protein

A. Introduction

The Fe protein of nitrogenase has for some time been known to be the specific reductant of the MoFe protein. Therefore, study of the Fe protein has focused on its properties related to catalysis, i.e., its ability to hydrolyze MgATP, its redox potential, electron transfer, and spectroscopy of its Fe_4S_4 cluster. Comparison of various Fe proteins from different species has shown that these properties place tight restrictions on changes in the amino acid sequence. Among the many conserved amino acid residues in this protein, particular attention has been paid to the five cysteine residues conserved in all species sequenced to date, due to their ability to act as potential Fe_4S_4 ligands and to participate in MgATP hydrolysis. In the absence of a completed crystal structure, [^{14}C]-iodoacetic acid labeling has been used to study the roles of the various cysteine residues (Hausinger & Howard, 1983). Cysteines 97 and 132 (numbering scheme from *K. pneumoniae*) were tentatively identified as thiol cluster ligands based on the observations that the rate of labeling of these residues is increased by the presence of MgATP and chelators and that labeling procedes in parallel with cluster destruction.

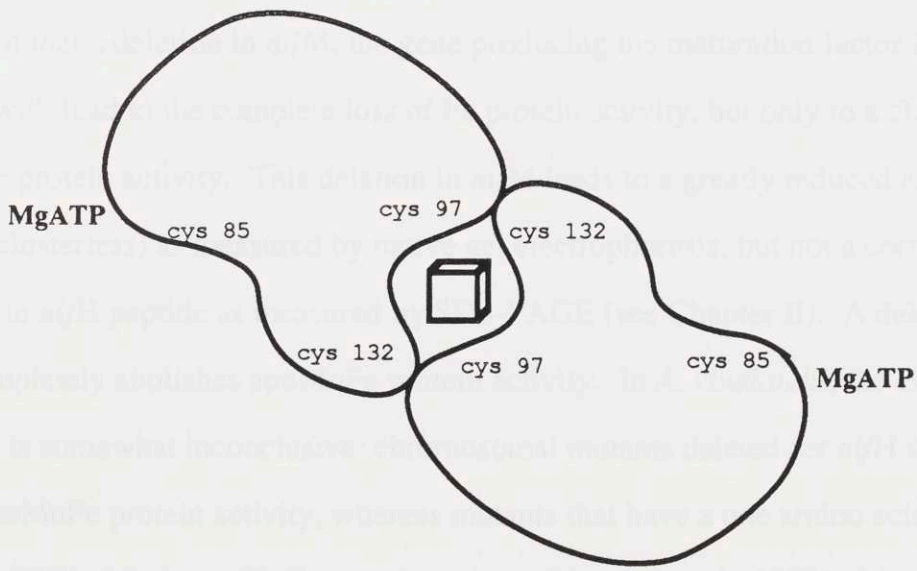


Figure III.1.: Thiol Reactivity Model of Fe protein (Hausinger & Howard, 1983)

Cysteine 85 was partially protected by MgATP from labeling and was therefore proposed to be near or in the site of MgATP hydrolysis. Based on these reactivities, a simple model was proposed (Figure III.1.). Site-directed mutagenesis experiments in which one cysteine at a time was changed to serine in the chromosome of *A. vinelandii* confirmed these assignments partially: only the Fe proteins in which the cluster-binding cysteines were changed to serines showed no activity; the others had at least moderate activity in the standard acetylene reduction assay (Howard et al., 1989).

In recent years, however, it has become obvious that the role of the Fe protein and its non-native peptide, nifH, is much more complex and includes not only electron transfer in catalysis, but also maturation of the other catalytic component, the MoFe protein. These maturation processes have been studied mainly in *A. vinelandii*, *K. pneumoniae*, and in *E. coli* using binary plasmid systems carrying *K. pneumoniae* genes (Harris et al., 1990). It was found that the Fe protein is absolutely required for the MoFe cofactor synthesis in these systems (except in *E. coli*, where MoFe cofactor synthesis is inefficient at best), whereas the requirement for Fe protein for apoMoFe protein synthesis varies from system from system. Both in the *E. coli* system and in *K. pneumoniae*, it is not the Fe protein that is absolutely required, but the *nif*H peptide. Work in the plasmid system has shown that a deletion in *nif*M, the gene producing the maturation factor for the Fe protein, will lead to the complete loss of Fe protein activity, but only to a 50% loss in apoMoFe protein activity. This deletion in *nif*M leads to a greatly reduced amount of Fe protein (clusterless) as measured by native gel electrophoresis, but not a corresponding decrease in *nif*H peptide as measured by SDS-PAGE (see Chapter II). A deletion in *nif*H itself completely abolishes apoMoFe protein activity. In *A. vinelandii*, the experimental evidence is somewhat inconclusive: chromosomal mutants deleted for *nif*H show considerable apoMoFe protein activity, whereas mutants that have a one amino acid change in *nif*H lose 90% of their apoMoFe protein activity (Howard et al., 1989). Moreover, there is evidence that in *A. vinelandii*, the Fe protein plays a catalytic role in MoFe cofactor

insertion into the apoMoFe protein (Robinson et al., 1989). The evaluation of such experiments in *A. vinelandii* is also complicated by a lack of understanding of its genetics and the presence of at least two alternate *nif* systems.

B. Methods

Construction of Plasmids: pTAC01 was constructed from pKH733 by removing the EcoRI site downstream of *nifH*. pKH733 was digested partially with EcoRI, the singly-cut band was isolated by agarose gel electrophoresis, a fill-in reaction using Klenow enzyme and all four deoxynucleotides was performed, and the plasmid was ligated back on itself in a blunt-end ligation reaction. M13mp19#4 was constructed by inserting the 0.7 kb EcoRI/KpnI fragment from pKH733 into the M13 phage derivative M13mp19 (Messing et al., 1977), making use of the EcoRI and KpnI sites present in its multiple cloning site. The orientation of the fragment was such that the non-coding strand of the insert is packaged for export. Mutations generated in the M13 system were transferred back into pTAC01 by exchanging the wild-type EcoRI/KpnI fragment against the near-identical fragment carrying the one base pair mutation. pTAC01-derived plasmids carrying a mutation were named pTACXX where XX stands for the number of the cysteine residue changed to serine. The transfer of the mutation into the expression system was confirmed by double-stranded sequencing using the upstream mutagenic primer as sequencing primer. Mutations were transferred from pTAC01-based plasmids into the pGH1 system by exchange of the 2.2 kb KpnI/XbaI fragment. pGH1-based plasmids carrying a mutation were named pTACXXG where XX stands for the number of the cysteine residue changed to serine.

Site-Directed Mutagenesis: Specific changes in the coding sequence of *nifH* located on pKH733 were achieved using oligonucleotide-directed mutagenesis (Zoller &

Smith, 1983, and Smith, 1985) with some modifications to favor survival of the mutant strand (Taylor et al., 1985). The oligonucleotides for this project were designed by Charles Eads, a postdoctoral associate, and provided by Professor F. Ausubel at the Massachachusetts General Hospital as part of an ongoing collaboration on the functions of various *nif* gene products. Design of the oligonucleotides was such that a single mismatch located near the center of the sequence would cause a change in the coding sequence from cysteine to serine (mismatches are underlined):

cys38 → ser38: 5'-GTCGGCTCCGATCCG-3'

cys85 → ser85: 5'-GTGCGCTCCGCGGAATCC-3'

cys97 → ser97: 5'-GTCGGCTCCGCGGGACG-3'

cys132 → cys132: 5'-GTGGTCTCCGGCGGCTTC-3'

cys184 → ser184: 5'-CTGATCTCTAACTCA-3'

These oligonucleotides - together with a primer (5'-ATGACCATGCGTCAATG-3') annealing at the beginning of the gene - were also used in the sequence confirmation of the mutant genes. All sequence information was based in published data (Scott et al., 1981).

Analysis of Mutant Proteins: Analysis of mutant proteins was by SDS-PAGE, native, anaerobic gel electrophoresis, ^{55}Fe labeling, and by in vitro nitrogenase assays. All methods are described in detail in Appendix A. MoFe cofactor samples with specific activities from 110 -190 nmol C_2H_2/[minxnmol molybdenum] (240 nmol C_2H_2/ [minxnmol molybdenum] theoretical maximum) were kindly provided by Dr. A. Hickman, D.W. Wright, and P. Christie. Proteins samples (Kp1, Kp2) in various stages of purification were a generous gift of Dr. G. Harris and Dr. R. Wahl. Kp1 and Kp2 were further purified to specific activities of greater than 1000 nmol C_2H_2/[minxmg] and 1400 nmol C_2H_2/[minxmg], respectively.

^{55}Fe Labeling of Nitrogenase Proteins in *E. coli*: Bacterial cells carrying *nif*-encoding plasmids were grown, harvested, and cracked as usual with minor modifica-

tions. Media composition is as in Appendix A. ^{55}Fe was from NEN, Inc. (1 mCi/ml, $\tau_{1/2}$ = 2.6 years, decay by electron capture). The appropriate amount of phosphates, NH_4Cl, NaCl, Casamino acids, and glucose were dissolved in water. Chelex cation exchanger (Na-form; BIORAD, Inc.) was added to 1.5% (w/v), the mixture was shaken gently for 1 hour, and the resin was allowed to settle for 1 hour. The "Chelexed" medium was steri-filtered into a nitric acid-rinsed, autoclaved 2 l flask. In this flask, the medium was complemented with the usual amounts of $MgSO_4$, $CaCl_2$, micronutrients, Na_2MoO_4, and antibiotics. Fe-citrate was added to 2.9 µg/ml final concentration. Aliquots of 250 ml were transferred to nitric acid-rinsed, autoclaved anaerobic growth flasks. Per flask, 0.5 ml of a mid-log starter culture in LB medium and 0.5 ml of ^{55}Fe-citrate (1 mCi/ml) were added. Cultures were then grown anaerobically at 30° C overnight. Harvesting was as usual except that the cell pellet after the first centrifugation was resuspended in "cold", anaerobic buffer (0.1 mol/l HEPES, 5 mmol/l dithionite) and recentrifuged. This wash was intended to remove most of the extraneous ^{55}Fe not incorporated into nitrogenase proteins. After native gel electrophoresis, gels were dried down immediately and were exposed to Kodak SB5 X-ray film. Exposure time was from overnight to 6-8 days. Development was according to the manufacturer's instructions: 5 min in GBX developer, 30 s rinse, 5-8 min in GBX fixer, 10 min rinse (all at 20-25° C).

C. Results and Discussion

In order to change cysteine residues 38, 85, 97, 132, and 184 of the Fe protein of *K. pneumoniae* to serine residues, the KpnI/EcoRI fragment of pKH733 spanning the stretch of DNA encoding all these residues was cloned into M13mp19, where the mutations were performed using oligonucleotide-directed mutagenesis. Using the method of

Eckstein (Taylor et al., 1985), mutation efficiencies greater than 90% were achieved, i.e., of 5 colonies screened for the desired mutation, typically four or five carried the desired mutation. The presence of the desired mutations and the absence of unwanted, secondary mutations were confirmed using dideoxy-sequencing: no secondary mutations were found. The KpnI/EcoRI fragments carrying the mutations were cloned back into the pKH733-derivative pTAC01, the expression vector. pTAC01-derived plasmids carrying a mutation were named pTACXX where XX stands for the number of the cysteine residue changed to serine. Thus, pTAC97 and pTAC132 represent mutations in the proposed cluster ligands , while pTAC85 is carries a mutation near or in the proposed MgATP-binding site. These plasmids were used together with the *nif*A-activator plasmid to co-transform *E. coli* strain W3110. The standard acetylene-reduction assay for the Fe protein of nitrogenase gave no detectable activity for any of the mutant proteins (Table III.1.). In order to ascertain that the absence of activity was not due to problems connected to the expression system, extracts were analyzed by SDS-PAGE/immunoblot (Figure III.2.). The presence of the *nif*H peptide was confirmed, thereby establishing that

expression plasmid + pVL15 (nifA)	Fe protein, specific activity (nmol/[minxmg])	relative amount of nifH peptide (SDS-PAGE) in %	relative amount of Fe protein like species (native gel electrophoresis) in %
pTAC01	46.1	100	100
pTAC38	ND	~ 25	< 5
pTAC85	ND	~ 25	< 5
pTAC97	ND	~ 50 - 75	< 5
pTAC132	ND	~ 50 - 75	< 5
pTAC184	ND	~ 25 - 50	< 5
ND...none detected (≤ 0.02)			

Table III.1.: Summary of Results in pTAC01-System

the specific activities are due to inactive mutant Fe protein and not to the absence of mutant Fe protein (Table III.1.).

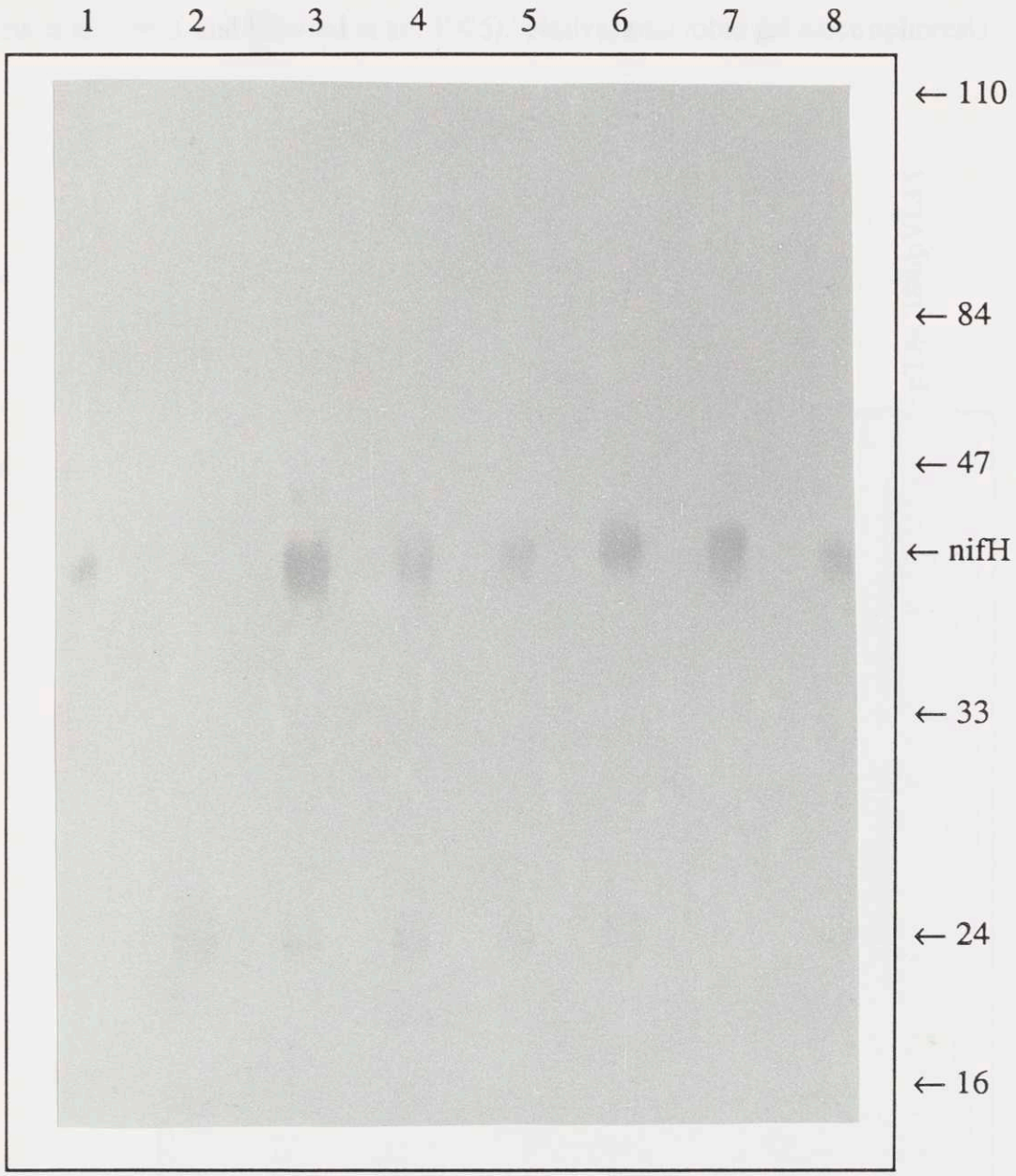


Figure III.2.: SDS-PAGE Analysis of pTAC01-Derivatives

The same extracts as in Figure III.3 were analyzed by SDS-PAGE on a 12% gel to determine whether the absence of Fe protein-like material during native gel electrophoresis is due to the absence of the NifH peptide. Detection of Fe proteins was with anti-Kp2 antiserum following Western blotting. All lanes contained 2 µg total cell protein with the exception of lane 1. Lane 1: 0.1 µg Kp2, Lane 2: 2 µg pACYC184/ pVL15, Lane 3: 2 µg pTAC01/pVL15, Lane 4: 2 µg pTAC38/pVL15, Lane 5: 2 µg pTAC85/pVL15, Lane 6: 2 µg pTAC97/pVL15, Lane 7: 2 µg pTAC132/pVL15, Lane 8: 2 µg pTAC184/pVL15.

The electrophoretic behavior of nitrogenase proteins in non-denaturing ("native") gels has been studied extensively and has been found to be a sensitive indicatior of the native structure of the Fe protein and the MoFe protein (apoprotein and holoproteins; Harris et al., 1990, and Howard et al., 1986). Native, anaerobic gel electrophoresis

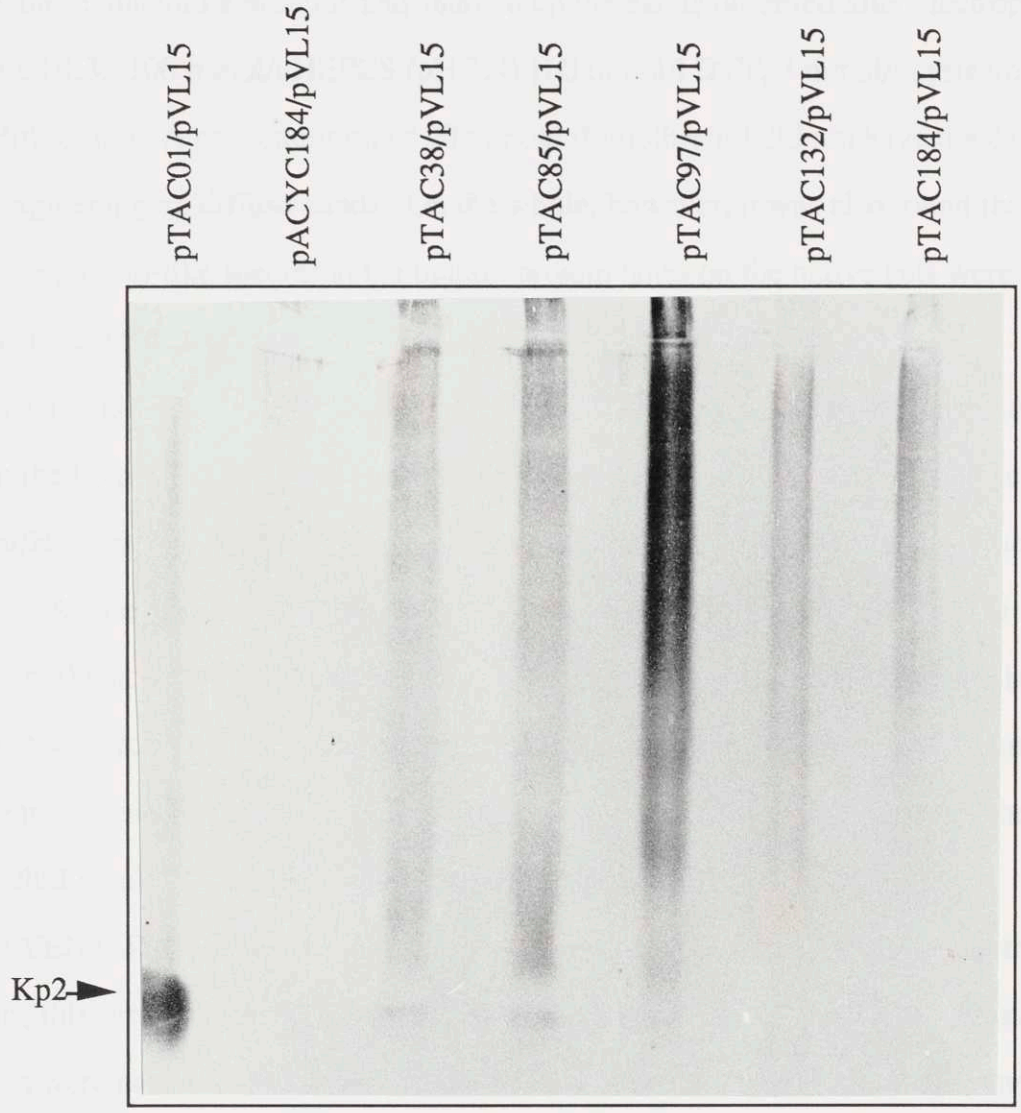


Figure III.3.: Native Gel Electrophoresis of pTAC01-Derivatives
Cells were broken in 100 mmol/l HEPES (pH 7.4) 100 mmol/l DTT 2 mmol/l cysteamine and 5 mmol/l dithionite. Native gel electrophoresis was followed by Western blotting and immunoreactive protein was visualized using horse radish peroxidase-conjugated secondary antibodies. Lanes contain equal amounts of extract.

followed by immunoblotting was therefore chosen to analyze the structural aspects of the mutant Fe proteins. Surprisingly, however, no distinct bands could be obtained at the position where Fe protein species were expected when the cells were broken in the standard cell disruption buffer: 100 mmol/l Tris (pH 7.4) 5 mmol/l dithionite. Therefore, the same cell disruption buffer as in the ΔM project (Chapter II) was eventually used to stablilize these mutant Fe proteins and sharpen up the bands observed after electrophoresis (Figure III.3.; 100 mmol/l HEPES (pH 7.4) 100 mmol/l DTT, 2 mmol/l cysteamine, 5 mmol/l dithionite). A concentration of 20 mmol/l dithiothreitol did not have the desired effect of tightening up diffuse bands. On the whole, however, it was also found that the levels of Fe protein-like species in the mutant protein lanes on the native gels were far below the level in the parent plasmid control lane, pTAC01 (Table III.1.).

After finding that all five mutants are inactive in the function most closely associated with the Fe protein, it became of immediate interest whether the activity of the Fe protein/*nif*H peptide in apoMoFe protein synthesis was affected by the single amino acid mutations. As mentioned in Chapter II (also see T. White's Ph.D. thesis, 1990), in the *E. coli* system, the non-processed *nif*H peptide made in the absence of the *nif*M protein is quite capable of acting in the maturation of the apoMoFe protein. The system chosen for subsequent analysis was pGH1, a pACYC184-based plasmid containing most of the *nif* genes, including all genes needed for apoMoFe protein synthesis: *nif*HDKTYENXUSVWZM. Again, the mutation was introduced by swapping restriction fragments, this time unique 2.2 kb KpnI/XbaI fragments. pGH1-based plasmids carrying a mutation were named pTACXXG where XX stands for the number of the cysteine residue changed to serine. Specific Fe protein and apoMoFe protein activities were determined (Table III.2.). The results obtained from the pTAC01 system were confirmed: no Fe protein activity was detected. The results for the apoMoFe protein activities showed that extracts with Fe protein (C38S) and Fe protein (C184S) contained reactivatable apoMoFe protein at normal levels when compared to the parent plasmid

pGH1. Extracts with Fe protein (C85S), believed to be near the MgATP-hydrolysis site, showed partial activity, while extracts with the proteins carrying the mutation at the proposed cluster ligand site, Fe protein (C97S) and Fe protein (C132S), showed no

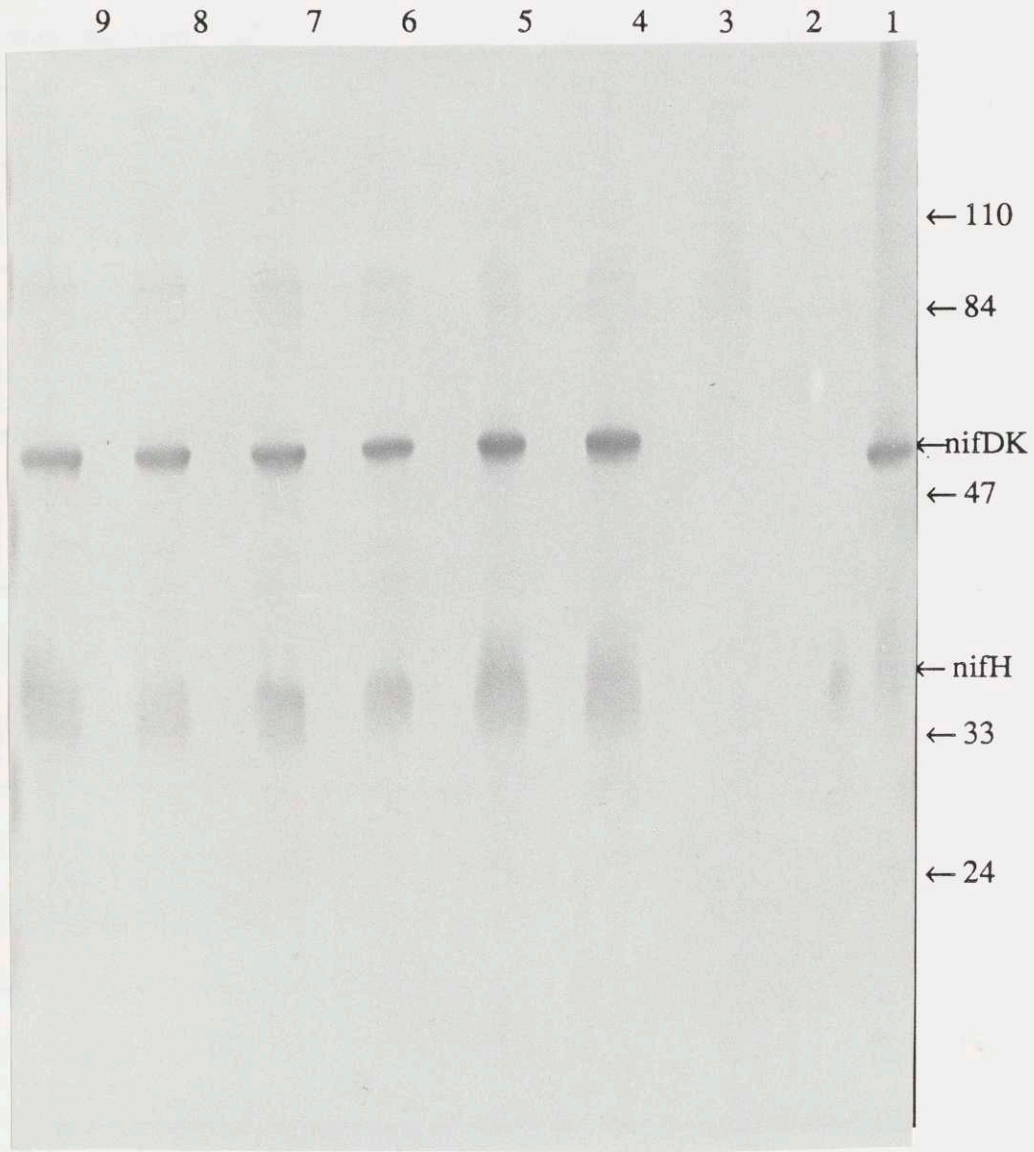


Figure III.4.: SDS-PAGE Analysis of pGH1-Derivatives

The same extracts as in Figure III.5. were analyzed by SDS-PAGE on a 12% polyacrylamide gel to determine whether the absence of Fe protein activity, apoMoFe protein activity, and the absence of properly assembled apoMoFe protein were due to the absence of the structural peptides nifH and nifDK. Detection was with anti-Kp2 and anti-Kp1 antisera following Western blotting. All lanes contained 10 µg total cell protein with the exception of lanes 1 and 2. Lane 1: 0.5 µg Kp1, Lane 2: 0.1 µg Kp2, Lane 3: 10 µg pACYC184/pVL15, Lane 4: 10 µg pGH1/pVL15, Lane 5: 10 µg pTAC38G/pVL15, Lane 6: 10 µg pTAC85G/pVL15, Lane 7: 10 µg pTAC97G/pVL15, Lane 8: 10 µg pTAC132G/pVL15, Lane 9: 10 µg pTAC184G/pVL15.

activity above background. In comparison, apoMoFe protein made from a ΔnifH plasmid is also not re-activatable. As in the pTAC01 system, the level of protein expression was monitored by SDS-PAGE followed by immunoblotting with the levels of the *nif*D and *nif*K peptides as internal standards (Figure III.4.). Again, some decrease in *nif*H peptide levels was observed.

Table III.2.: Summary of Data from pGH1 System

Expression plasmid + pVL15 (nifA)	Specific Activities (nmol/[minxmg])		relative amount of nifD,K peptide (SDS-PAGE) in %	relative amount of nifH peptide (SDS-PAGE) in %	relative amount of apoMoFe protein in %	relative amount of Fe protein like species in %	one-step or two-step activation ?
	Fe protein	apoMoFe protein					
pGH1	106.3	22.3	100	100	100	100	two
pTAC38G	ND	20.7	75 - 100	75 - 100	75 - 100	≤ 1	two
pTAC85G	ND	1.2	75 - 100	50 - 75	≤ 1	<< 1	two
pTAC97G	ND	ND	75 - 100	50 - 75	≤ 1	<< 1	two
pTAC132G	ND	ND	75 - 100	50 - 75	≤ 1	<< 1	one
pTAC184G	ND	21.3	75 - 100	75 - 100	75 - 100	≤ 1	two

ND...none detected

Currently, the only way to follow the activation pathway from apoMoFe protein to holoMoFe protein is to analyze the various species by native gel electrophoresis. Typically, upon activation of apoMoFe protein with cofactor, it is possible to distinguish three bands in the process: the apoMoFe protein which runs slowest, the fully activated MoFe protein containing two MoFe cofactors per tetramer, and an intermediate believed to contain one MoFe cofactor. Extensive work with various combinations of genes in the binary plasmid system has shown that a deletion in *nif*H will result in reduced levels of inactive apoMoFe protein which contains no iron and undergoes a one-step mobility shift (Harris et al., 1990). In order to track the cause of inactivity of the apoMoFe proteins made in the presence of the mutant Fe proteins and in order to determine whether these apoMoFe proteins undergo the two-step mobility shift observed for the mature apoMoFe protein, small aliquots (~ 0.1 ml) of the same extracts analyzed by SDS-PAGE were "activated" with purified MoFe cofactor, analyzed by native gel electrophoresis, and compared to non-activated extracts (Figure III.5.). For all the mutant Fe proteins, the levels of Fe protein-like species are reduced compared to expression in the pTAC01 system. However, in agreement with the apoMoFe protein activity data for pTAC38G and pTAC184G, the apoMoFe protein species in these extracts and their shifts in mobility upon addition of MoFe cofactor are identical to those in the pGH1 extracts. In the other extracts (pTAC85G, pTAC97G, pTAC132G), the levels of apoMoFe proteins are reduced greatly (Table III.2.). In addition, for pTAC132G, a proposed cluster ligand mutant, the intermediate in the two-step activation could not be observed, thus resulting in a Δ*nif*H-like apoMoFe protein.

Knowing the metal content of holo- and apo-nitrogenase proteins is central to the understanding of protein maturation. The *nif*M protein has been shown to be central to the insertion of the Fe_4S_4 cluster of the Fe protein (Chapter II) and is suspected to take part in the maturation of the MoFe protein as well (T. White's Ph.D. thesis, 1990). The *nif*H peptide is known to be necessary for apoMoFe protein synthesis, i.e., for the matura

Figure III.5.: Activation of pGH1 Extracts with MoFe Cofactor

tion of a $\alpha_2\beta_2$ dimer containing four unusual Fe_4S_4 clusters. Starting with the working hypothesis that the function of the *nif*H peptide might be participation in cluster assembly on the apoMoFe protein, the iron content of the various apoMoFe proteins was determined by native gel electrophoresis of ^{55}Fe labeled cell extracts. Again, as in the ΔM project (Chapter II), the same extracts that were analyzed by native gel electrophoresis/ autoradiography were, in a parallel experiment, analyzed by native gel electrophoresis/ immunoblotting to confirm that the absence of an ^{55}Fe signal at a certain position was due to the absence of iron in the protein band and not due to the absence of protein. Cells were broken in the presence of 2.5 μg/ml leupeptin, 3.4 μg/ml pepstatin, and 0.1 mg/ml phenylmethylsulfonylfluoride as peptidase inhibitors. Since it was suspected that some of the mutant proteins might be freeze-thaw sensitive, all extracts were prepared and loaded on the native gels the same day (Figure III.6.), except for one set of experiments where the same extracts were taken through one freeze-thaw cycle in liquid nitrogen before analysis (Figure III.7. and Figure III.8.). The ^{55}Fe experiments confirmed the expectations for the apoMoFe proteins in the pTAC38G and pTAC184G extracts: they contained near-normal levels of iron. The apoMoFe proteins present in pTAC85G, pTAC97G, and pTAC132G extracts contained little iron, and that disappeared completely after freeze-thawing the extracts once. Of the mutant Fe protein species, only Fe protein (C184S) contained some iron whose presence was not affected by freeze-thawing. The Fe protein (C38S) which is present at a level comparable to that of Fe protein (C184S) did not contain any iron detectable even with a two-week exposure time.

In concluding, it can be said that all five evolutionarily conserved cysteine residues are essential to the Fe protein: replacing any of these residues with a serine, the amino acid closest to cysteine in structural and chemical terms, has a profound effect both on the maturation of the Fe protein and on its catalytic properties. While each of the mutant Fe proteins is made in significant amounts as determined by SDS-PAGE, only Fe protein (C38S) and Fe protein (C184S) are processed into "mature", Fe protein-like species in

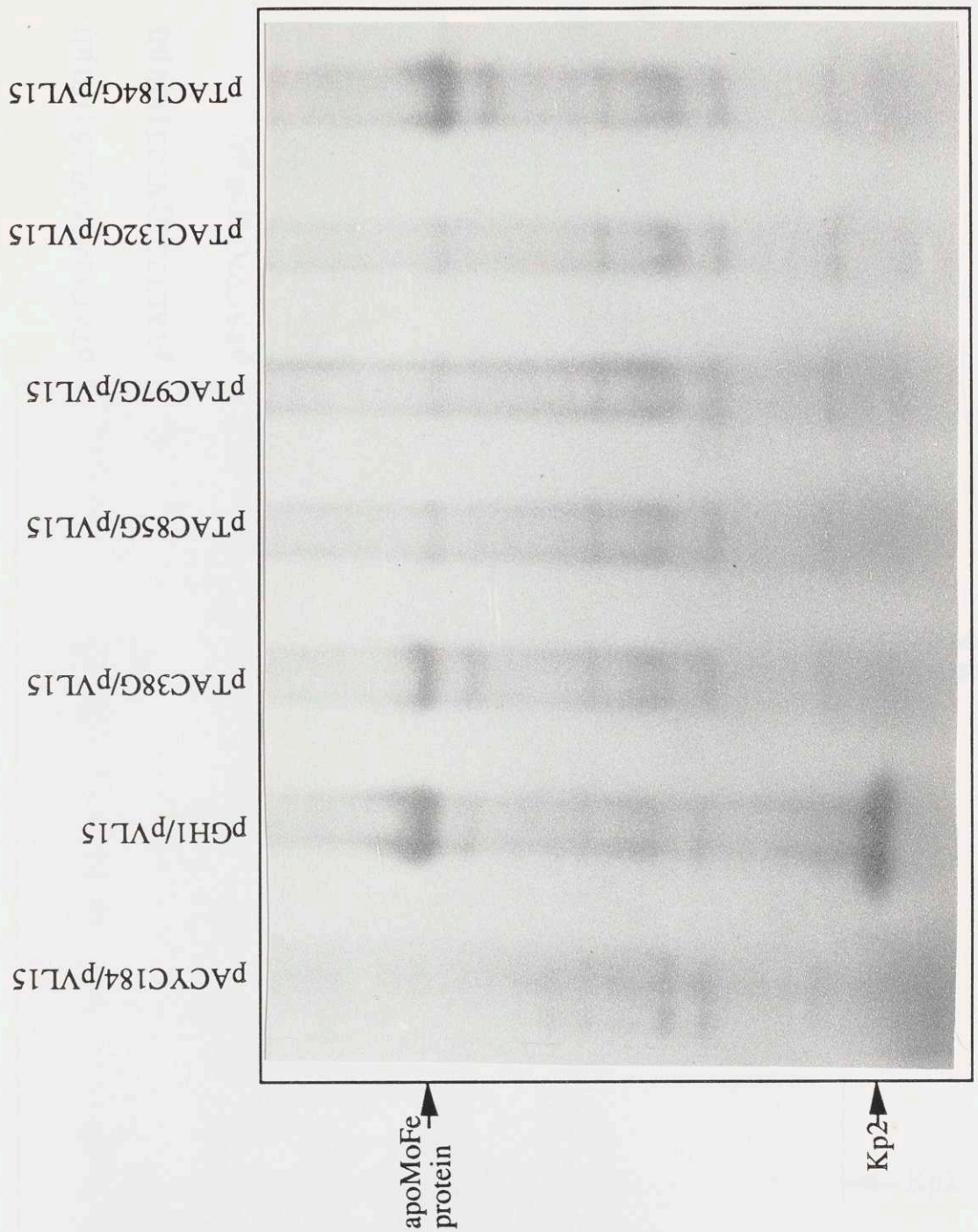


Figure III.6.: ^{55} Fe Labeling of Extracts in the pGH1 System
The iron content of the various Fe proteins and apoMoFe proteins was analyzed by non-denaturing, anaerobic gel electrophoresis of the ^{55}Fe labeled crude cell extracts followed by autoradiography. Equal amounts of extract were loaded per lane (150 µl, appr. 750 µg of total protein).
Equal amounts of extracts (150 µl) were loaded per lane, analyzed by native gel electrophoresis, and exposed to X-ray film for two weeks.

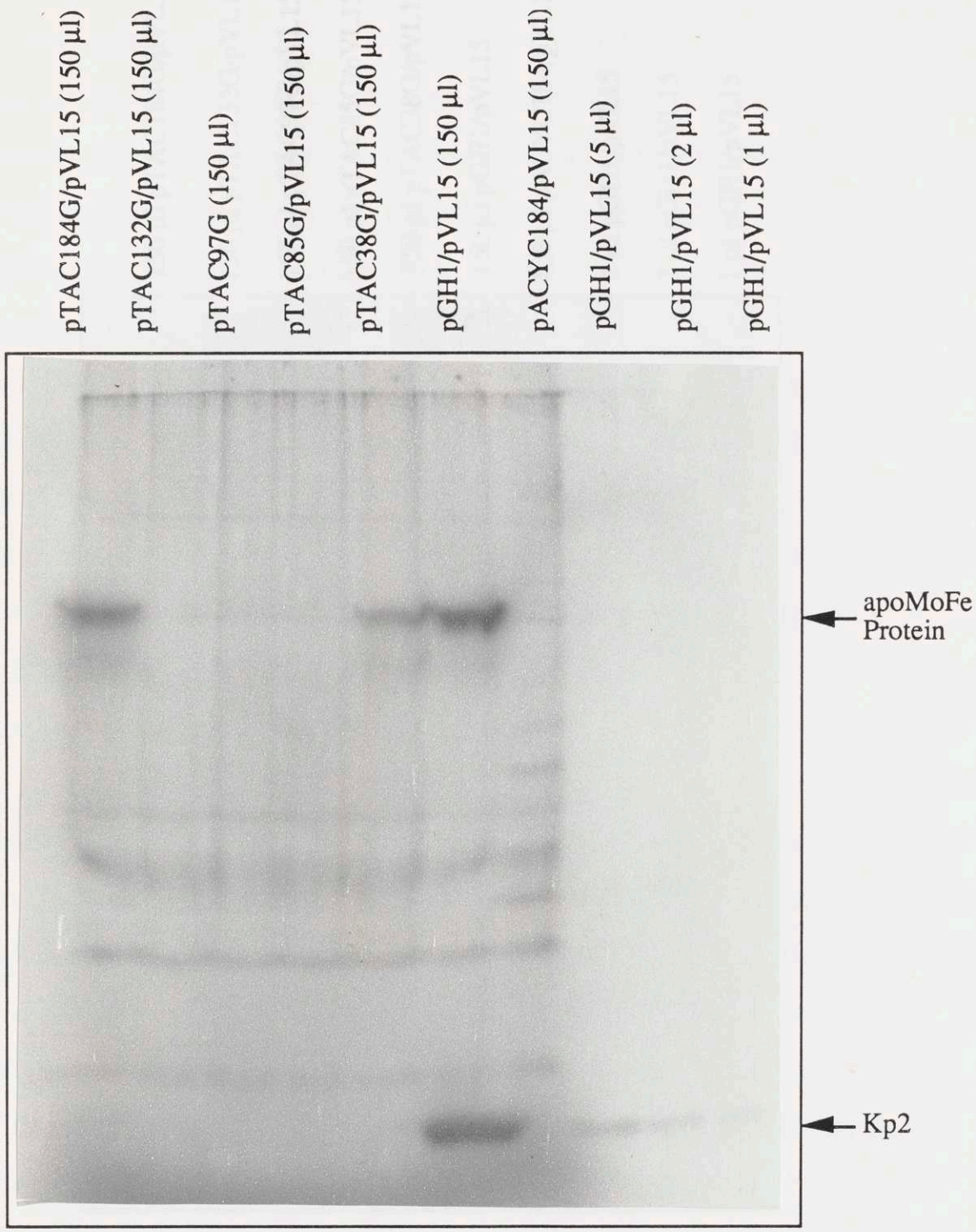


Figure III.7.: ^{55}Fe Analysis of Mutants in pGH1 System after Freeze-Thaw
To examine the dependence of the iron content of the various Fe proteins and apoMoFe proteins, the same extracts that were used in Figure III.6. were subjected to a single, anaerobic freeze-thaw cycle in liquid nitrogen. Equal amounts of extract were loaded per lane (150 µl, appr. 750 µg of total protein) except for lanes 8-10.

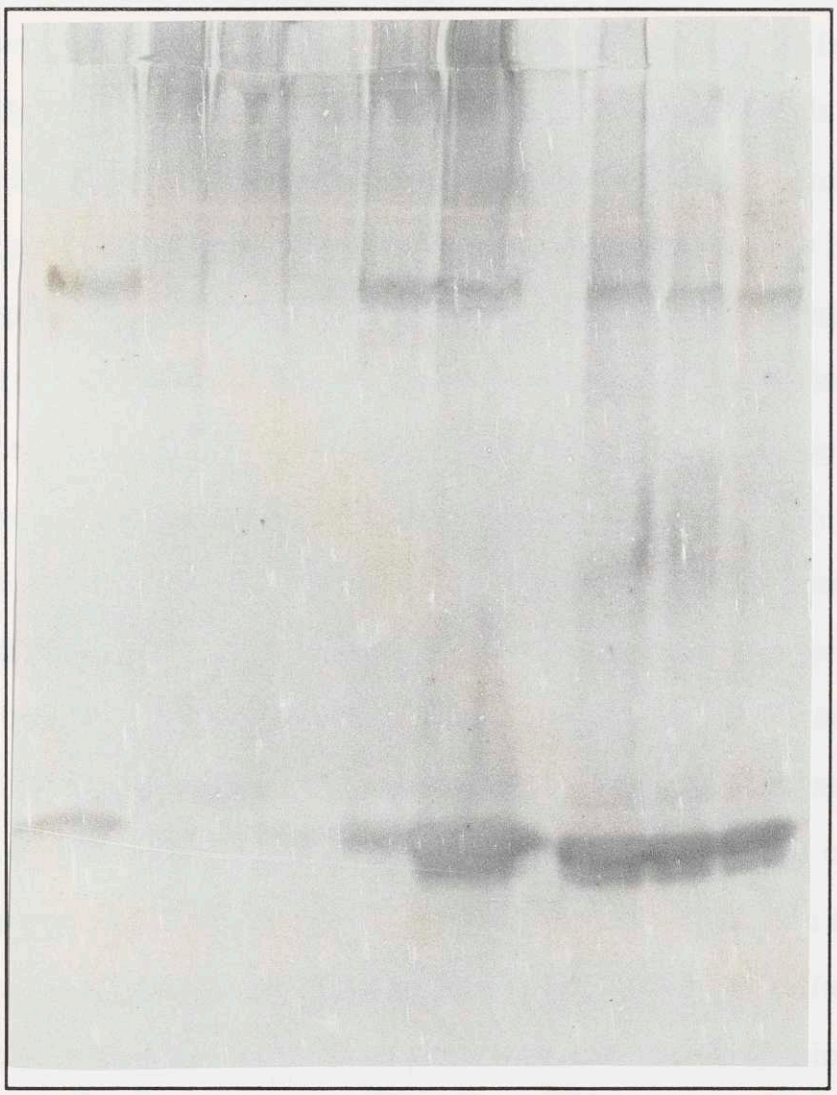


Figure III.8.: Immunoblot of ^{55}Fe Samples in Figure III.7.
In a control experiment to Figure III.7. performed in parallel, identical volumes of the same ^{55}Fe labeled crude cell extracts were analyzed by non-denaturing, anaerobic gel electrophoresis followed by Western blotting. Fe proteins and apoMoFe proteins were detected with anti-Kp1 and anti-Kp2 antisera. The pGH1+pVL15 lane was overloaded deliberately to permit exact duplication of running conditions.

amounts sufficient to determine definitively that only the Fe protein (C184S) contains iron, while the Fe protein (C38S) shows no associated iron. Overall, these proteins are similar to the Fe protein made in the absence of the *nif*M protein. As to the effect of the mutations on apoMoFe protein synthesis is concerned, it can be said conclusively that cys38 and cys184 are either not required in this process, or that a serine residue in this position performs the same function with comparable efficiency. The apoMoFe proteins made in the presence of these mutant Fe proteins are indistinguishable from the apoMoFe protein made in the presence of the wild-type Fe protein by the methods employed in this work. The maturation of the apoMoFe protein involves - at least - assembly of the α- and β-subunits and synthesis and insertion of the P-centers, presumed to be Fe_4S_4-clusters. One of the possible roles of the Fe protein in this process has been thought to be the insertion of these cluster. The data from the Fe protein (C38S) precludes this role: although some of this Fe protein mutant assembles into a dimer-like species, it does not contain iron; yet, it aids in the synthesis of fully re-activatable apoMoFe protein.

One caveat that applies to the cysteine Fe protein mutants that make re-activatable apoMoFe protein or contain iron (C184S) is posed by non-conserved cysteine residues that might be able to "jump in" once a conserved cysteine residue is changed to a serine. Such behavior was observed in the site-directed mutagenesis of the ferredoxin in *A. vinelandii*, where a non-conserved cysteine 4 amino acids away from a cysteine-to-alanine mutation assumed the role of the cluster ligand (Martin, A.E. et al., 1990). A final interpretation of this mutagenesis experiment was only possible after determining the X-ray crystal structure of the mutant protein. Non-conserved cysteine residues that could complicate the analysis of the results in this work are found at positions 5, 151, 234, and 259 of the Fe protein.

The apoMoFe proteins made by the other three mutant Fe proteins fall into two categories. Although in both cases the levels of apoMoFe proteins as determined by native gel electrophoresis are low, only the apoMoFe protein made in the presence of the

Fe protein (C85S) has some activity. Both other apoMoFe proteins, made in the presence of the proposed cluster ligand mutants, Fe protein (C97S) and Fe protein (C132S), are not re-activatable with MoFe cofactor, although the former shows the two-step transition associated with activation. Of all mutations, the change of cysteine 132 to serine has the most significant effect on apoMoFe protein synthesis; in fact, this apoMoFe protein is similar to the one made in the absence of the *nif*H gene. Not only is the apoMoFe protein not re-activatable, but it only contains little, freeze-thaw-sensitive iron and undergoes a one-step conversion upon additon of cofactor.

With this set of five mutations within one protein, the maturation process of the apoMoFe protein - as seen from the point of view of the Fe protein - can be divided into two functions. The first function is associated with the two-step shift as opposed to the one-step shift in the ΔnifH apoMoFe protein. The second function is associated with conferring an additional property on the apoMoFe protein showing a two-step shift that then allows re-activation. A two-step shift is therefore necessary but not sufficient for activation (e.g. C97S). In this respect, the apoMoFe proteins made in the presence of Fe proteins (C97S) and Fe protein (C85S) are the most intriguing: both apoMoFe proteins show a two-step "activation", thus indicating that some processing took place compared to the Δ*nif*H apoMoFe protein with its one-step "activation", yet only one of them is fully re-activatable.

Further investigation into the nature of the processing of the apoMoFe protein, i.e., Fe_4S_4 cluster insertion, two-step transition, and re-activation by MoFe cofactor, is needed, to elucidate the molecular basis of apoMoFe protein maturation.

Chapter IV:

nifV and Homocitrate

A. Introduction

R-homocitric acid is an organic component of the MoFe cofactor of nitrogenase (Hoover et al., 1987) whose isolation and characterization is of longstanding interest to this lab. It is important to an understanding of this chapter that the reader appreciate that typical preparations of this cofactor, freed of protein, contain numerous, incompletely evaluated organic and inorganic components, at least some of which are added during the preparation of cofactor from MoFe protein (McLean et al., 1989, and Wink et al., 1989). In *K. pneumoniae,* the occurrence of homocitric acid in the MoFe cofactor and in the culture medium results from the presence of a functioning *nif*V gene. Point mutations in this gene lead to a fascinating phenotype: the MoFe protein loses its ability to reduce nitrogen, while it is still able to form ethylene from acetylene (McLean & Dixon, 1982). Moreover, hydrogen evolution of the mutant enzyme is partially inhibited by carbon

	nitrogen reduction	acetylene reduction	carbon monoxide inhibition of hydrogen evolution
wild-type component 1	yes	yes	no
nifVminus component 1	no	yes	up to 40% inhibition

Table IV.1.: *nif*V Phenotype

monoxide (up to 40% at "infinite" CO concentration), while hydrogen evolution of the wild-type enzyme is apparently totally unaffected by carbon monoxide (Erbes & Orme-Johnson, 1975), a fact remarkable in itself considering that all other known hydrogenases

are completely inhibited by carbon monoxide (Table IV.1.). Both acetylene reduction and nitrogen reduction of the wild-type enzyme are inhibited by carbon monoxide: EPR analysis has shown that CO binds to two sites of the enzyme (Davis et al., 1979).

Interestingly, these mutations can be cured by adding homocitric acid to the medium of derepressed cultures of *K. pneumoniae nif*V$^-$ mutants (Hoover et al., 1988a). The development of an in vitro system for the synthesis of the MoFe cofactor using mixtures of extracts from various mutant strains has shown that organic acids can take the place of homocitric acid on the cofactor, resulting in altered substrate specificities like the ones mentioned above (Hoover et al., 1988b). The initial assumption that citric acid replaces homocitric acid in the *nif*V$^-$ MoFe protein was borne out in these in vitro experiments: citric acid was the only small, organic acid sufficiently competent in the in vitro system to be incorporated in significant amounts leading to *nif*V$^-$ MoFe protein.

Homocitric acid has only been connected to nitrogen fixation in *K. pneumoniae*; attempts to find it in *A. vinelandii* have failed (D. Dean, personal communication, 1989). It is, however, known to be the first intermediate in the biosynthesis of lysine in *Saccharomyces cerevisiae* (Gray & Bhattacharjee, 1976).

Purification and analysis of MoFe cofactor, crystallization of the MoFe cofactor , and determination of the requirements for holo MoFe protein synthesis are of longstanding interest to this lab. The presence of homocitric acid on the cofactor or as a bridge from peptide backbone to cofactor greatly complicates the evaluation of purification procedures and attempts at synthesizing model compounds (D. Coucouvanis, personal communication, 1989). Models for the biosynthesis of cofactor must address the question of how R-homocitric acid is incorporated into the MoFe protein: by protein-directed assembly or by simple mass action.

Due to the change in substrate specificity upon substitution of citric acid for homocitric acid on the cofactor, research has focussed on the role of homocitric acid during catalysis. Not much attention has been paid to the fact that R-homocitric acid

readily crosses the cell membrane both ways and is found in significant amounts in the culture medium of derepressed cells. This must represent a more than negligible expenditure of cellular resources. One must examine the hypothesis that, in *K. pneumoniae* at least, there may be a function for homocitric acid in addition to binding to the MoFe cofactor, which is synthesized at much lower molar levels than homocitrate. Any such secondary function would have gone undetected in most experiments looking at the effect of mutations on *nif* functions: the loss of R-homocitric acid incorporation alone would have sufficed to give a "non-leaking" *nif* phenotype. One possible secondary function, use of R-homocitric acid as a metal carrier to import iron or molybdenum, would have been suppressed under growth conditions employing high metal concentrations, anyway.

Despite a lack of direct experimental evidence up to now, the obvious chelating potential of homocitrate has to be considered in any account of its peculiar properties - particular the observation that a critical and specific enzyme component is spilled into the extracellular medium.

B. Methods

Materials: Most materials and suppliers are listed in Appendix A in the appropriate section. All chemicals were at least reagent grade and nitrogen gas was prepurified grade from Airco Co. The plasmid pKK223-3 was obtained from Pharmacia/LKB, Inc. All computer algorithms used were implemented on the VAX computer at the Whitaker Computing Facility at the Massachusetts Institute of Technology as part of the University of Wisconsin Genetics Computer Group DNA and Protein Analysis Software (Version 6.0-6.2).

Plasmid Constructions: pAT10 was constructed by ligating the XhoI/

Eco47III fragment of pCF3 carrying *nifV* to the larger SalI/PvuII fragment of pBR322. pAT11 was derived from pAT10 by removing the MscI fragment, thus deleting most of *nifU* and *nifS*: pAT10 was opened with MscI and closed up on itself after removal of the smaller MscI fragment by agarose gel electrophoresis. pKK#6 was constructed by replacing the small SmaI/HindIII fragment in the polylinker of pKK223-3 with the SmaI/HindIII piece carrying *nifV* from pAT11. pKOV was constructed by Amy Anderson from pKK#6 by generating uni-directional deletions away from the tac promoter into the remainder of the *nifS* gene (and - inadvertently - also into the *nifV* gene). Uni-directional deletions were achieved using the Erase-A-Base kit from Promega Corp. and the reaction conditions listed in the manual (Promega Corp., 1987). pKK#6 was opened and the tac promoter was protected from ExonucleaseIII activity by complete digestion with EcoRI, followed by fill-in with thionucleotides. Digestion with SmaI generated a blunt end susceptible to attack by ExonucleaseIII. ExonucleaseIII was added and time points were taken to cover a range of deletions. S1 nuclease digestion, Klenow repair, ligation with T4 Phage DNA Ligase, and transformation into *E. coli* strain JM105 yielded colonies that were analyzed for protein content. pV#2/3 was constructed by opening up pKK#6 at the EcoRI and MluI sites; ligation with a custom-made "sticky" linker restored the first two codons of *nifV* and introduced a new NcoI site together with a ribosome-binding site modelled after the one found for *lacZ*. Custom-made linkers (5' - AATTCAGGAAACAGACCATGGAA - 3' and 5'- CGCGTTCCATGGTCTGTTTCCTG - 3') were desalted on a NAP-10 column (Pharmacia/LKB Corp., 1989), phosphorylated using T4 DNA Kinase, annealed at 65° C for 15 min, and used directly in a sticky end ligation (50-fold excess of linker over plasmid). In order to prevent the formation of concatenated inserts, the ligation mix was cut with NcoI and religated. This last ligation mix was then used to transform *E. coli* strain XL1 Blue. Finally, transformants were analyzed for protein content. pV#2/3 was also analyzed by DNA sequencing using a commercially avail-

able primer reading of the tac promoter into the multiple cloning site of pKK223-3.

Screening for Overproducers: Overnight cultures grown in LB medium with appropriate antibiotics were diluted 1:100 into fresh medium and grown at 37° C to an approximate A_{660nm} of 0.7. Isopropyl-ß-D-thiogalacto-pyranoside (IPTG) was added to 2.5 mmol/l and the culture was grown for another 3-4 hours. The bacterial cells were harvested by centrifugation and resuspended in 1x Laemmli buffer (10% of the original volume). Typically, 15-25 µl were analyzed by SDS-PAGE.

N-terminal Analysis of Recombinant Proteins: Samples were analyzed by SDS-PAGE and then electro-blotted onto Immobilon membranes (see Appendix A). Bands of interest were cut out and sent to the Biopolymers Laboratory at the Center for Cancer Research at MIT for N-terminal gas phase sequencing by Edman degradation.

Quantitation by Scanning Densitometry: Samples were run on an SDS-PAGE gel, stained by Coomassie, and destained until sufficient resolution was obtained. Lanes were scanned at 6.5 cm/min using the Hoefer GS 300 Scanning Densitometer and analyzed using GS 370 Data System software (both Hoefer Scientific Instruments, Inc.). For a 20 cm gel, 1,500 data points were collected. Each scan was corrected to give a uniform left edge, and the baseline was adjusted with a handdrawn curve. The data were smoothed once and expanded vertically to allow quantitative comparison between scans. Peaks were identified automatically at medium sensitivity. When necessary, peaks were marked manually to maintain consistency between scans. Integration was done assuming Gaussian line shapes.

Complementation Assays by Diazotrophic Growth: *K. pneumoniae* was grown as described in Appendix A in the absence of fixed nitrogen. Cultures of the strains and transformants of interest were grown in LB medium with the appropriate antibiotics for 10 hours at 37° C. To 2.4 ml of starter culture, 8.3 ml nitrogen-free MM9 medium were added (2% sucrose and antibiotics, if appropriate) and growth was continued overnight at 30° C. In the morning, 250 ml of nitrogen-free MM9 medium were inoculated with 2 ml

of the overnight culture in a 500 ml anaerobic culture flask, the headspace was evacuated and backfilled with argon gas twice, and growth was started at 30° C. Nitrogen to satisfy growth requirements was supplied to the anaerobic cultures as a slow stream of N_2, leading to growth of only the *nif*V$^+$ strains. Growth was followed using a Klett spectrophotometer with a # 54 filter (green).

Sample Preparation for Homocitrate Synthase Assay: *K. pneumoniae* was grown as described in Appendix A in the presence of L-serine to allow growth of *nif*V$^+$ and *nif*V$^-$ strains. Overnight cultures of the strains and transformants of interest were grown in LB medium with the appropriate antibiotics at 30° C. Of this overnight culture, 1 ml was added to 250 ml of MM9 medium containing L-serine at 50 µg/ml and L-arginine at 0.02% (w/v). Flasks were cycled as above and the headspace of the cultures was flushed continuously with nitrogen during growth to maintain anaerobicity. After 11 hours, IPTG was added to 1.5 mmol/l using anaerobic techniques. After an additional 8 hours, in vivo nitrogenase activities were determined and cells were harvested. All further manipulations were also done anaerobically. Cells were harvested by centrifugation and resuspended in 0.1 mol/l HEPES (pH 7.4) containing leupeptin at 1.9 µg/ml and pepstatin at 3.5 µg/ml (1% of original culture volume). Typically, 0.1 ml of extract were used per assay.

Homocitrate Synthase Assay: Cell extracts of various strains and transformants of *K. pneumoniae* were tested for homocitrate synthase activity using the procedure described for the enzyme in *Saccharomyces cerevisiae* (Gray & Bhattacharjee, 1976) with only one modification: anaerobic techniques were used throughout the assay. An acetyl-CoA stock solution (pH 7.0) was prepared from the lithium salt, while an α-ketoglutarate stock solution was prepared by neutralization of the free acid with KOH. A solution containing Tris (pH 8.0) at 100 mmol/l, $MgCl_2$ at 5 mmol/l, and 2,6-dichlorophenol-indophenol (DCPIP) at 11.3 µg/ml was prepared fresh daily. To 2.5 ml of this last solution, acetyl-CoA was added to 27.8 µg/ml and α-ketoglutarate was added

to 23 µg/ml when required. All solutions required in the assay were then added to an anaerobic cuvette at 32° C; the reaction was initiated with the addition of cell extract and followed at 600nm for 6.7 min. To test whether low molecular weight compounds affect the reduction of the dye DCPIP, 1 ml of the extract was desalted on a G25 Sephadex minicolumn in an anaerobic glove box and tested in the assay. Specific activities were expressed in ΔA_{600nm}/(minxmg).

Sample Preparation for Detection of R-homocitric Acid: Cultures were grown as in "Sample Preparation for Homocitrate Synthase Assay". Samples were taken before and after induction with IPTG, bacterial cells were removed by centrifugation, and the supernatant was sterilized by filtration through a 0.2 µm membrane. Samples were stored at 4° C prior to use.

Detection of R-homocitric Acid by HPLC: R,S-Homocitric acid lactone was obtained from Sigma Chemical Corp. Chromatographic analysis was carried out isocratically on an analytical ION-300 ion exchange column (Interaction Chemicals, Inc.) with 0.01 N H_2SO_4 as eluent. Typically, a flow rate of 0.3 ml/min was employed. Detection was at 214 nm with occasional use of a refractive index detector. Sample size was from 10 - 100 µl. Cycle time per injection was 45-60 min. Reactions were started by dissolving the lactone form in the appropriate buffer maintained at 23° C (75 mmol/l phosphate at pH 2.1, 7.4, 11.0).

C. Results and Discussion

Computer Analysis of the *nifV* Gene Product: A database search of Genbank Version 64.0 and of the European Molecular Biology Laboratory databank using FASTA and TFASTA found several proteins of interest with significant similarities to the *nifV* gene product of *K. pneumoniae*. α-isopropylmalate synthase (*leu*A) from *Salmonella*

typhimurium shows 30.6% identity in a 320 amino acid overlap, 3-hydroxy-3-methylglutaryl-CoA reductase (*mva*B) from *Pseudomonas mevalonii* shows 24.2% identity in a 285 amino acid overlap, and oxalacetate decarboxylase from *Klebsiella pneumoniae* shows 19.6% identity in a 240 amino acid overlap.

α-Isopropylmalate synthase (Ricca& Calvo, 1990, and Kohlhaw et al., 1969) catalyzes the first committed step in the biosynthesis of leucine, the conversion of α-ketoisovalerate to α-isopropylmalate (Figure IV.1.), and 3-hydroxy-3-methylglutaryl-CoA reductase (Beach & Rodwell, 1989, and Anderson, & Rodwell, 1989) catalyzes the oxidation of R-mevalonate to S-3-hydroxy-3-methylglutaryl coenzyme A (HMG-CoA).

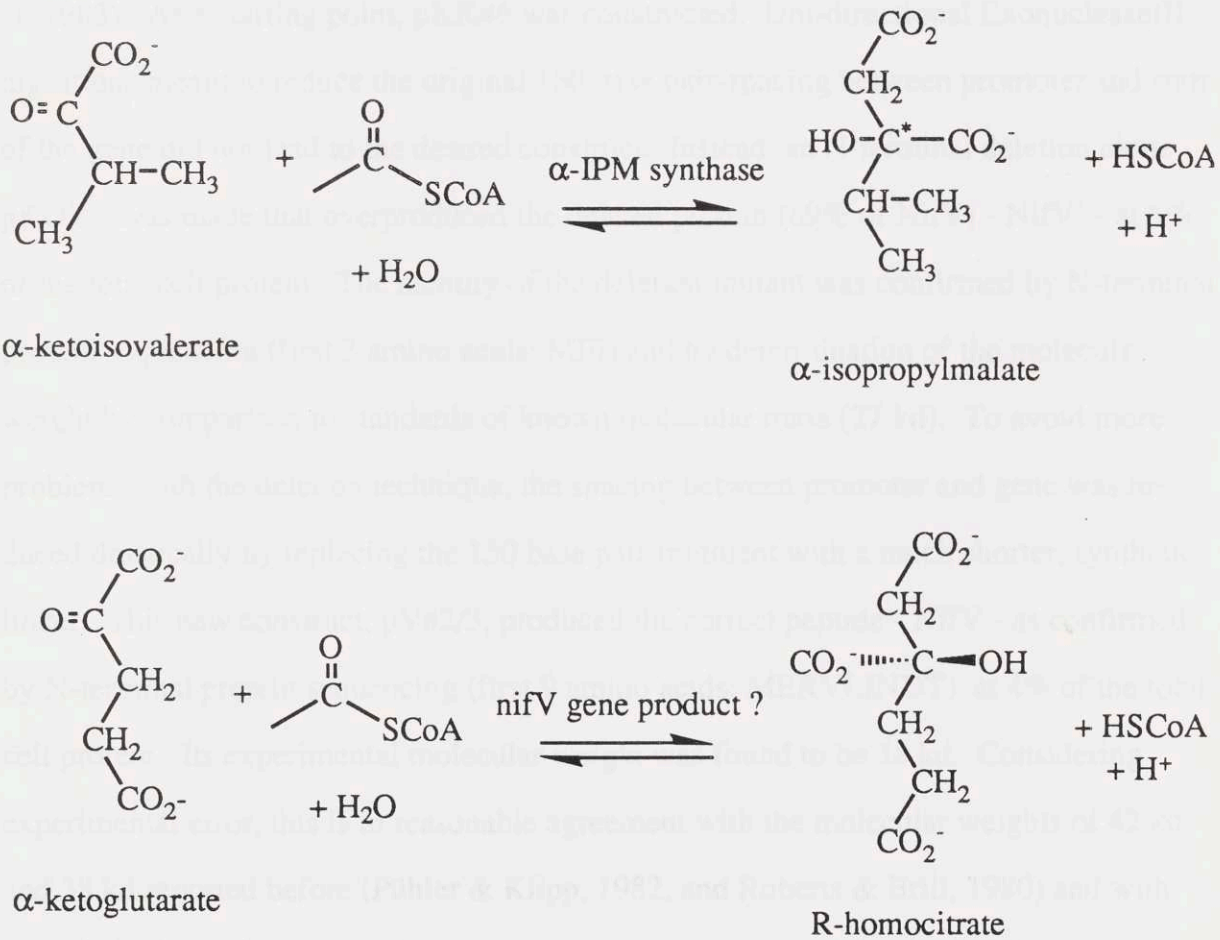


Figure IV.1.: Comparison of α-IPM Synthase and Homocitrate Synthase

Oxalacetate decarboxylase pumps Na^+ ions out of the bacterial cell at the expense of the free energy of the decarboxylation reaction during the import of oxalacetate into the cell (Schwarz et al., 1988). Interestingly, no significant similarity was found to any of at least five citric acid synthases in the database. Sequence similarities and a comparison of the chemistries of α-IPM synthase and homocitrate synthase suggest that R-homocitrate might be synthesized in *K. pneumoniae* the same way as in yeast, i.e., by the condensation of α-ketoglutarate and acetyl-CoA (Figure IV.1).

Overexpression of the *nifV* Gene Product in *E. coli* strain XL1 Blue: In order to obtain large quantities of the *nifV* gene product for characterization, several plasmids were constructed in which *nifV* was located downstream of the tac promoter (Amann et al.,1983). As a starting point, pKK#6 was constructed. Uni-directional ExonucleaseIII digestions meant to reduce the original 150 base pair-spacing between promoter and start of the gene did not lead to the desired construct. Instead, an N-terminal deletion clone - pKOV - was made that overproduced the deleted protein (69% of NifV) - NifV' - at 5% of the total cell protein. The identity of the deletion mutant was confirmed by N-terminal protein sequencing (first 3 amino acids: MFI) and by determination of the molecular weight by comparison to standards of known molecular mass (27 kd). To avoid more problems with the deletion technique, the spacing between promoter and gene was reduced drastically by replacing the 150 base pair fragment with a much shorter, synthetic linker. This new construct, pV#2/3, produced the correct peptide - NifV - as confirmed by N-terminal protein sequencing (first 9 amino acids: MERVLINDT) at 4% of the total cell protein. Its experimental molecular weight was found to be 38 kd. Considering experimental error, this is in reasonable agreement with the molecular weights of 42 kd and 38 kd reported before (Pühler & Klipp, 1982, and Roberts & Brill, 1980) and with the value of 41.1 kd predicted from the DNA sequence (Arnold et al., 1988).

5' - CTG TTGA CAATTAATCATCGGCTCG TATAA TGT -

GTGGAATTGTGAGCGGATAACAATTTCACAC**AGGA**AACAGAATTC -

AGGAAACAGA**CCATGG**AACGCGTGCT - 3'

Figure IV.2.: The Promoter Construct in pV#2/3
Shine-Dalgarno sequences are printed bold. Restriction sites are underlined: EcoRI, NcoI, MluI from left to right. The re-introduced start codon is printed bold and underlined. The RNA polymerase recognition sites are underlined and separated by three spaces from the other sequence.

The dependence of expression of NifV' and NifV on the presence of IPTG was demonstrated in a time course experiment, in which samples were taken from the cultures grown in LB medium at regular intervals after addition of 2.5 mmol/l IPTG and were analyzed by SDS-PAGE (Figures IV.3. and IV.4.). This experiment confirmed that the expression of NifV and NifV' in these constructs was under proper control of the tac promoter and that, as judged by the steady-state level of the peptides, induction of both proteins was maximal after 3 hours.

Overexpression of the *nifV* Gene Product in *K. pneumoniae*: In order to compare the level of expression in *E. coli* to that in *K. pneumoniae*, UN1990 - a *nifV*$^-$ strain - was transformed with pKK#6, pKOV, and pV#2/3. pKK223-3 served as the negative control. Induction with IPTG of cultures grown in LB medium lead to an additional band - NifV - in the pV#2/3 lane that was determined to constitute 7% of the total cell protein under non-optimized induction conditions. Its molecular weight was found to be 37 kd. Surprisingly, no nifV or nifV' protein was observed in any of the other lanes (Figure IV.5.). While the absence of an additional band in the pKK#6-lane is expected (due to the non-optimized promoter-gene distance), the absence of NifV' in the pKOV-lane seems to indicate that this deletion mutant does not accumulate to significant levels in *K.*

pneumoniae as opposed to *E. coli*. Alternatively, there may be a degradation pathway for "improper" proteins in *K. pneumoniae* that does not exist in *E. coli*.

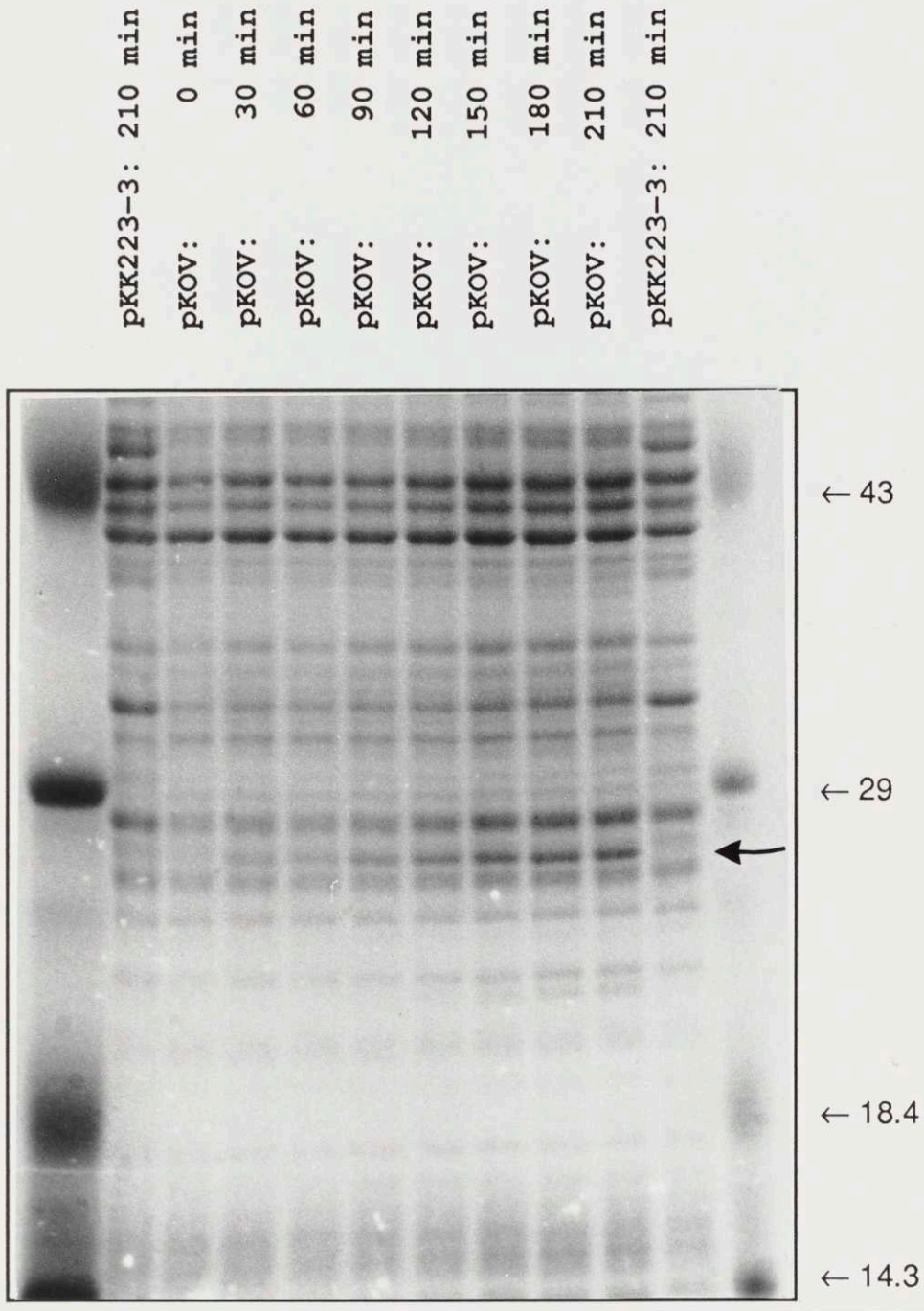


Figure IV.3.: Induction of pKOV in *E. coli*

E. coli strain Xl1 Blue was transformed with pKOV and induced with 2.5 mmol/l IPTG. Samples were taken at regular intervals and analyzed by SDS-PAGE.

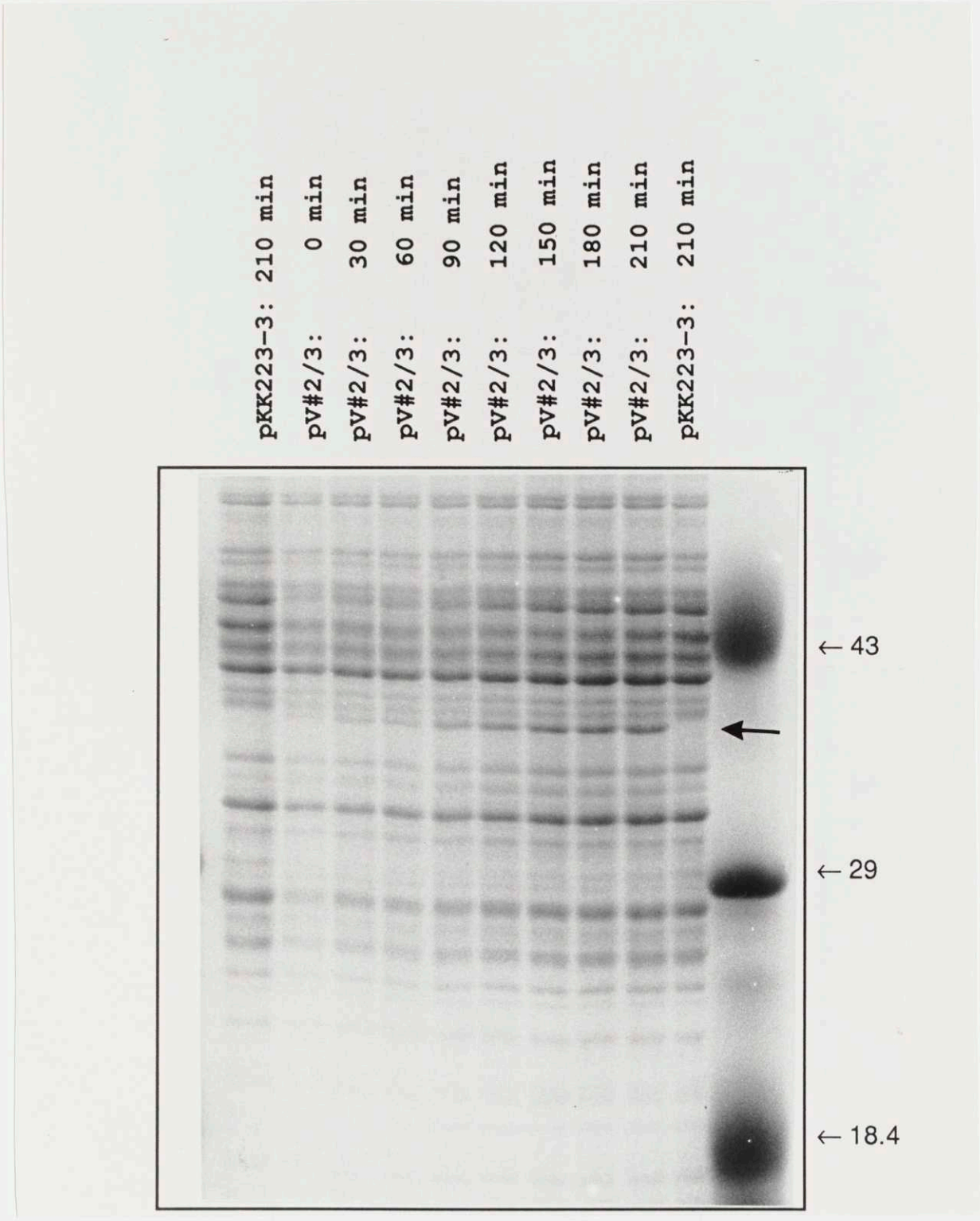


Figure IV.4.: Induction of pV#2/3 in *E. coli*

E. coli strain Xl1 Blue was transformed with pV#2/3 and induced with 2.5 mmol/l IPTG. Samples were taken at regular intervals and analyzed by SDS-PAGE.

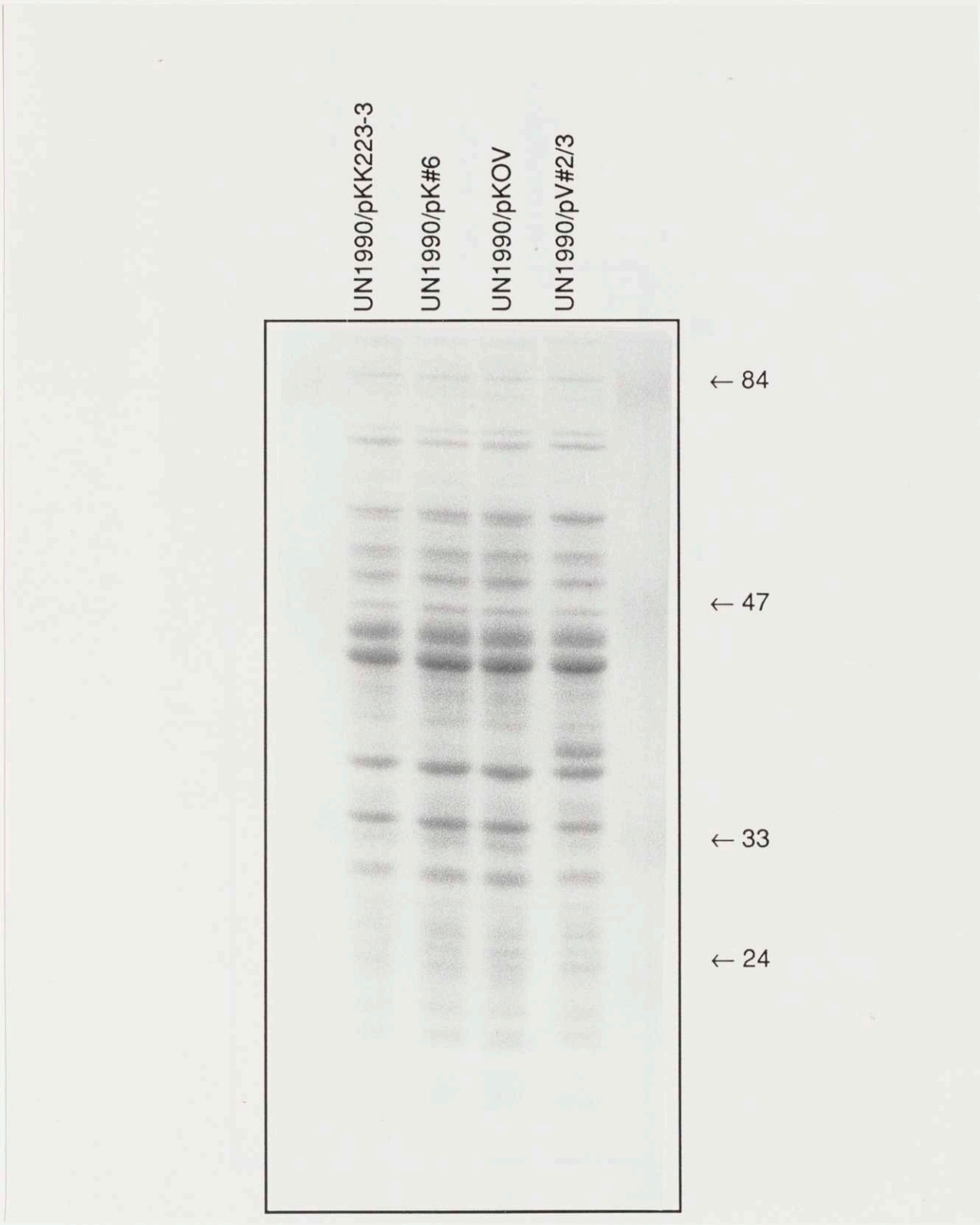


Figure IV.5.: Induction of *nif*V-Encoding Plasmids in UN1990

Judging from the SDS-PAGE data, the proteins made from pV#2/3 in *E. coli* and *K. pneumoniae* are identical in size within experimental error. Analysis of the deletion mutant NifV' in its native genetic background (*K. pneumoniae*) may be impossible due to stability problems.

Complementation of UN1190 with a *nifV* Overproducer: In the absence of a known enzymatic assay for nifV, the activity of the recombinant constructs was determined in a complementation assay. UN1990, a *nifV*⁻ mutant of *K. pneumoniae* was transformed with the plasmids of interest.

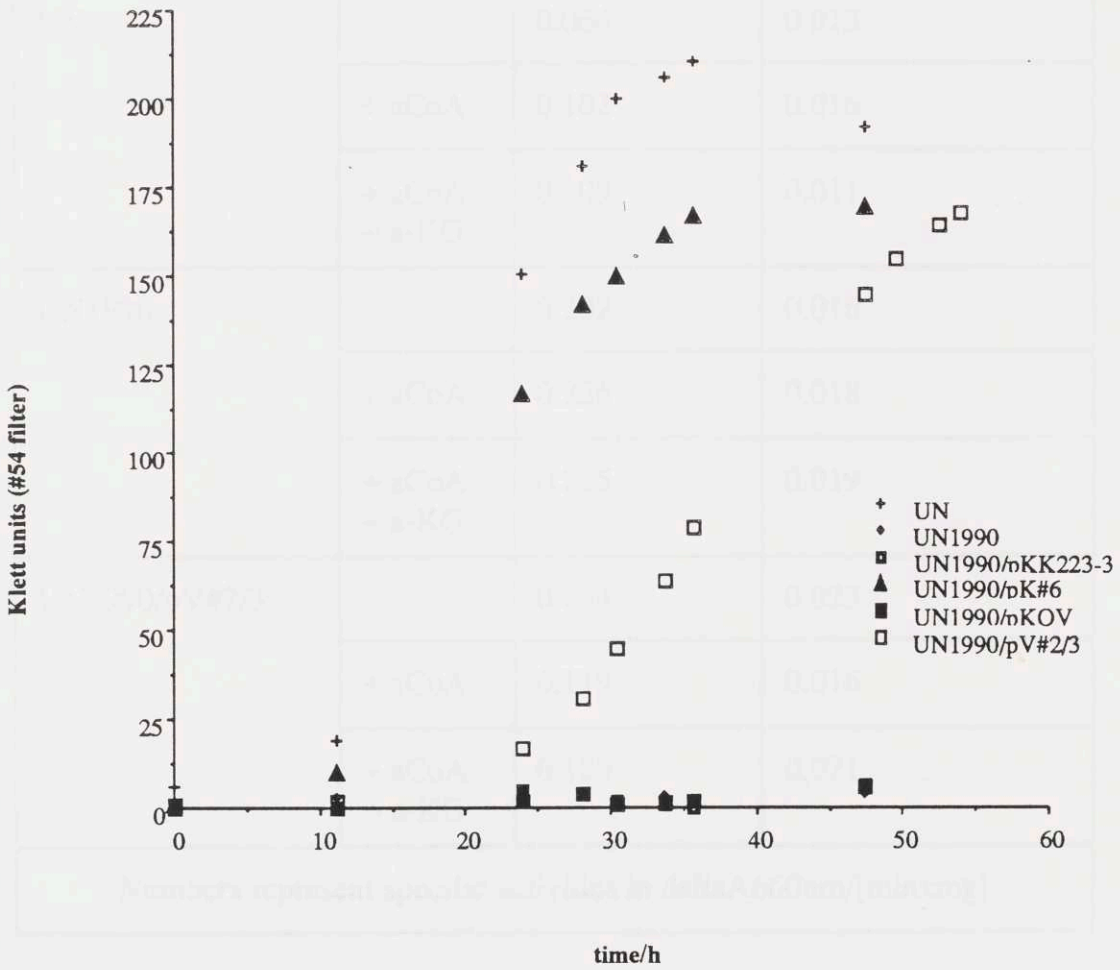


Figure IV.6.: Complementation of UN1990 (*nif*V⁻) with *nif*V-Encoding Plasmids

The transformants were tested to determine whether the presence of the plasmid would restore the ability to grow diazotrophically, i.e., the ability to grow on N_2 in the absence of fixed nitrogen (Figure IV.6.). It was found that pV#2/3 and pKK#6 produced amounts of active *nifV* gene product sufficient to restore the wild-type growth phenotype, while pKOV showed no activity above background.

Homocitrate Synthase Assay: In view of the identification of the low molecular weight factor found in the presence of a functioning *nif*V gene as R-homocitrate (Hoover

		crude cell extract	desalted crude cell extract (G25)
UN		0.064	0.013
	+ aCoA	0.102	0.016
	+ aCoA + a-KG	0.109	0.011
UN1990		0.202	0.016
	+ aCoA	0.226	0.018
	+ aCoA + a-KG	0.215	0.019
UN1990/pV#2/3		0.094	0.023
	+ aCoA	0.119	0.016
	+ aCoA + a-KG	0.129	0.021
Numbers represent specific activities in deltaA660nm/[minxmg]			

Table IV.2.: Results from Homocitrate Synthase Assay

Cultures were grown under nitrogen flow, derepressed, and induced with IPTG. Cell harvesting, assay, and all other manipulations were done using anaerobic techniques.

et al., 1987) and in view of the results of the database searches, an enzymatic assay was attempted for a possible activity of the *nifV* gene product as a homocitrate synthase. The procedure used was that for the homocitrate synthase in *Saccharomyces cerevisiae* (Gray & Bhattacharjee, 1976). Specific acitvities (ΔA_{660nm}/[minxmg]) were determined for crude cell extracts and for crude cell extracts desalted by sizing chromatography. Activities were determined in the absence and presence of α-ketoglutarate (α-KG) and acetyl-CoA (aCoA). Unfortunately, although some increase in activity was observed upon addition of α-ketoglutarate and acetyl-CoA, no significant difference was observed between extracts of UN (wild-type), UN1990 (*nif*V$^-$ strain), and UN1990/pV#2/3 (*nif*V-plasmid in *nif*V$^-$ strain). It was determined that most of the reduction of the dye DCPIP was due to low molecular weight compounds that were removed by the gel filtration step. Desalted extracts showed no stimulation of activity due to α-ketoglutarate and acetyl-CoA (see Table IV.2.). It was concluded that the activity of the *nif*Vgene product is either not the one assayed for or that the background resulting from other reductive processes in the crude cell extracts was obscuring the activity of the *nif*V gene product.

Study of the Solution Behavior of R,S-Homocitric Acid: In collaboration with C. Klabbers and D. Wright, the behavior of R,S-homocitric acid in aqueous solution was studied in order to determine in how far its solution behavior corresponds to possible functions like metal binding (Figure IV.7.).

R - homocitrate R - homocitrate lactone

Figure IV.7.: Lactonization of R-Homocitric Acid

This included finding out how likely it is to occur in the closed or in the open form and in what form it is secreted into the medium. It was found that the lactone form and the open form could be readily distinguished by HPLC on an ION-300 ion exchange column. By employing a flow rate of 0.3 ml/min at variable back pressures, retention times of 21 and 26 minutes were obtained for the open and closed form, respectively, with a peak separation of 4 to 5 minutes (Figures IV.9. and IV.8.). Detection was at 214 nm with a response of 0.00196 A_{214} per µg of homocitric acid lactone (0.0024 A_{214} per µg open form).

Using this analytical technique, equilibrium constants were determined for the lactonization reaction at 23°C: K_{eq} = [closed]/[open] was found to be 10.8 at pH 7.4 and 0.074 at pH 11.0. On the time scale of a typical derepression experiment (< 10 hours), less than 5% ring opening was observed if the lactone was incubated in pH 11.0 or in pH 7.4 buffer. The open form of homocitric acid - generated by prolonged incubation of the lactone in pH 11.0 buffer - showed no significant ring closing during the same time frame (6-10 hours). This analysis shows that although the closed form is thermodynamically favored under biological conditions, either form can have biological significance because the kinetics of ring opening and closing are slow.

Preliminary experiments with culture supernatants of *nif*V$^+$ and *nif*V$^-$ strains of *K. pneumoniae* indicate that it is the open form that is secreted. The analysis is made difficult, however, by the complexity of the chromatogram due to metabolites and composition of the media. Overall, the experimental data on the interconversion of the open and closed form show that, at physiological pH in the time frame of a derepression experiment, the open form - once secreted - exists long enough to be biologically relevant. This kinetic stability of the open form, together with the fact that homocitric acid is made in large amounts and readily crosses the cell membrane, suggests that there is an important physiological function for the open form of homocitric acid, possibly in metal binding and import.

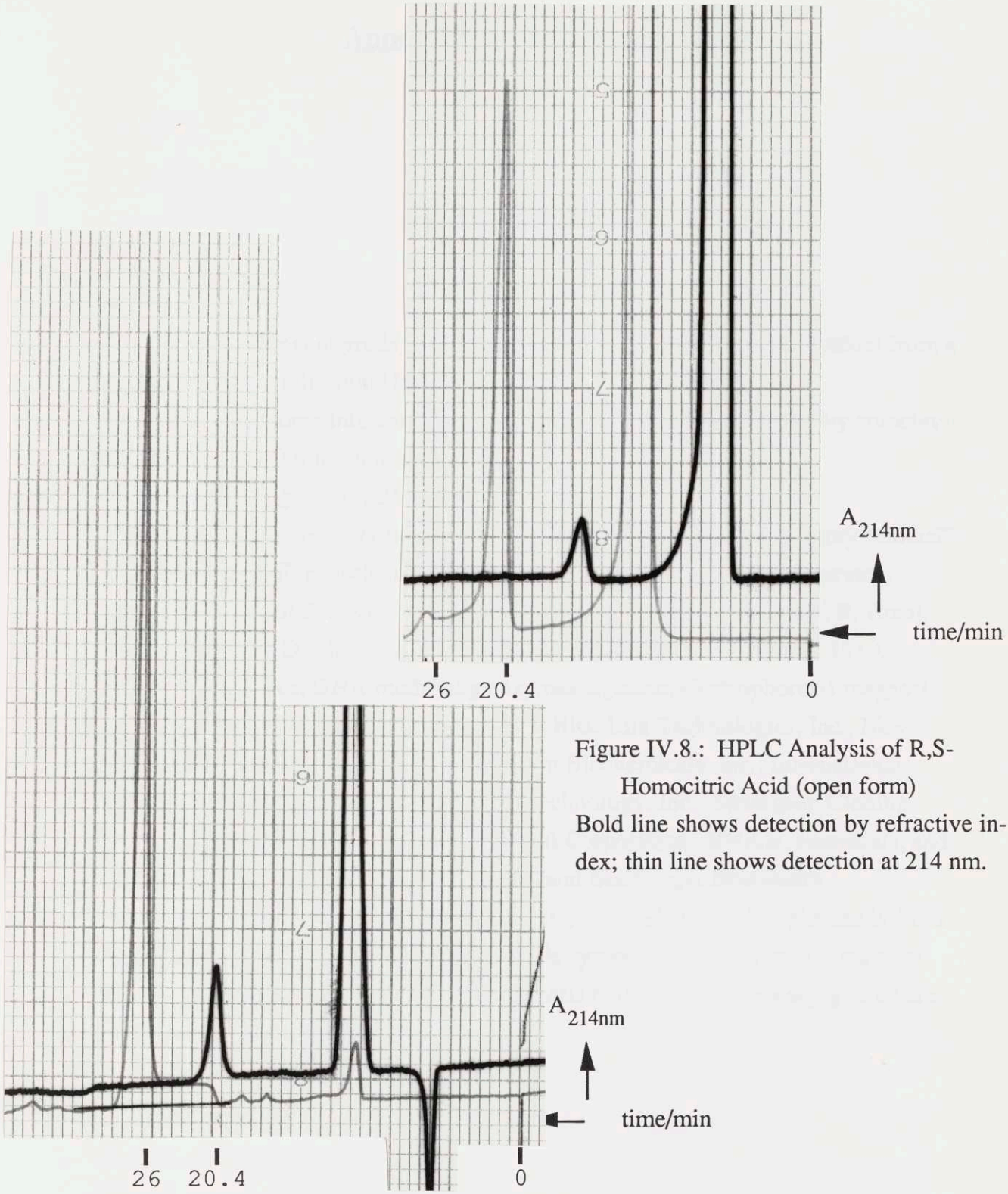


Figure IV.8.: HPLC Analysis of R,S-Homocitric Acid (open form)
Bold line shows detection by refractive index; thin line shows detection at 214 nm.

Figure IV.9.: HPLC Analysis of R,S-Homocitric Acid (lactone form)
Bold line shows detection by refractive index; thin line shows detection at 214 nm.

Appendix A: THE COOKBOOK

DNA Techniques

All reagents are reagent grade unless otherwise mentioned. All H_2O is taken from a Millipore SuperQ Filtration Unit.

All reagents that come into contact with bacteria or DNA are sterilized by autoclaving (121° C, 20-30 min) unless otherwise mentioned.

All pH values are given for 25° C.

Many procedures given rely heavily on "Molecular Cloning: A Laboratory Manual" by T. Maniatis, E.F. Fritsch, and J. Sambrook, Cold Spring Harbor Laboratory, 1982, and "Current Protocols in Molecular Biology" by F. M. Ausubel, R. Brent, R.E. Kingston, D. D. Moore, J. A. Smith, J. G. Seidman, and K. Struhl, 1989.

Restriction enzymes, DNA modifying enzymes, agarose, electrophoresis reagents, and sequencing reagents are purchased from BRL Life Technologies, Inc., New England Biolabs, Inc.; Boehringer Mannheim Biochemicals, Inc.; International Biotechnologies, Inc.; Pharmacia LKB Biotechnology, Inc.; Stratagene Cloning Systems, Inc.; and United States Biochemical Corporation. RNAse, chemicals, and buffers are from Sigma Chemical Company and Boehringer Mannheim Biochemicals, Inc. Glassmilk is from BIO101, Inc. and HK™ Phosphatase is from Epicentre Technologies, Inc. Oligonucleotide synthesis and purification reagents are from Applied Biosystems, Inc. Mutagenesis reagents and sequencing label are from Amersham Corporation.

Restriction Digests

Procedure

1. To a microfuge tube add: 8 µl DNA solution
   2 µl 10X buffer
   10 µl H_2O
2. Heat at 65° C for 10 min, spin down liquid, and add restriction enzyme. In general, 1 unit of enzyme will digest 1 µg of DNA with one restriction site in question in 1 h at its optimum temperature.
3. Incubate at the optimum temperature of the enzyme for 1 h.
4. Heat at 65° C for 10 min and run on a 1% agarose gel for analysis.

If small fragments are to be visualized, RNAse must be used; add 0.1 µl per digest. Enzyme digests are typically done in the restriction buffer supplied by the manufacturer of the enzyme. Double digests can be done in two ways: if the salt concentrations recommended for the two enzymes are compatible, both enzymes can be used at the same time. If not, the lower salt digest is done first and the salt concentration is then increased for the second enzyme. Standard buffers are given below.

Reagents

The values given refer to the concentration in the digest.

no salt buffer:	25 mmol/l Tris-HCl[1] (pH 7.8)
	10 mmol/l $MgCl_2$
	100 mg/ml BSA[2]
	2 mmol/l ß-mercaptoethanol
low-salt buffer:	25 mmol/l Tris-HCl (pH 7.8)
	50 mmol/l NaCl
	10 mmol/l $MgCl_2$
	100 mg/ml BSA
	2 mmol/l ß-mercaptoethanol
medium-salt buffer:	25 mmol/l Tris-HCl (pH 7.8)
	100 mmol/l NaCl
	10 mmol/l $MgCl_2$
	100 mg/ml BSA
	2 mmol/l ß-mercaptoethanol

high-salt buffer:	25 mmol/l Tris-HCl (pH 7.8)
	150 mmol/l NaCl
	10 mmol/l $MgCl_2$
	100 mg/ml BSA
	2 mmol/l ß-mercaptoethanol
TA buffer:	33 mmol/l Tris-acetate (pH 7.8)
	66 mmol/l $KAcO^3$
	10 mmol/l $MgAcO^4$
	0.5 mmol/l DTT^5
	100 µg/ml of BSA

TA buffer works for nearly every restriction enzyme, regardless of the manufacturer's recommendation for NaCl concentration (O'Farrell et al., 1980). Several enzymes have unique buffers. Refer to individual enzyme listing in the manufacturer's catalogue for buffer description.

RNAse:	10 mg/ml pancreatic RNAse A
	10 mmol/l Tris (pH 7.5)
	15 mmol/l NaCl
	Boil for 20 min on an open waterbath and allow to cool slowly; freeze in small aliquots at -20° C.

DNA-Agarose Gel Electrophoresis

The useful range of separation of this technique is 0.3-25 kb. For the range 0.6 - 23 kb (λHindIII-digest), a 0.8 - 1.2% gel is appropriate. To increase resolution for small fragments, increase the agarose concentration up to 2%; for large fragments (>10 kb), decrease the concentration down to 0.7%.

Procedure

1. Assemble gel frame according to manufacturer's instructions. Cool to -20° C.
2. Dissolve agarose in electrophoresis buffer (1% w/v). Heat over burner until all agarose is dissolved. Add 0.01% $EtBr^6$ and mix.

3. When gel mix is still hot to the touch, pour it into frame.
4. Allow to solidify, remove comb carefully, and place in gel box.
5. Add 10-12% loading buffer (FSUDS) to samples. Load wells with standard and samples (max. to 4/5th capacity), and run at constant voltage.

For analytical gels, TBE is recommended as bands appear somewhat sharper; SEB (TAE) should be used if DNA is to be isolated from the gel.

running conditions: 50-80 V for uncooled minigel box
75-130 V for uncooled gel box
100-250 V for cooled gel box

Reagents

TBE:	Tris (pH 8.3)	100.0 mmol/l	12.11 g/l
	HBO_3[7]	89.0 mmol/l	5.50 g/l
	$Na_2EDTAx2H_2O$[8]	2.2 mmol/l	0.81 g/l
	Make as 10x stock and store at room temperature.		
SEB (TAE):	Tris (pH 8.0)	40.0 mmol/l	4.85 g/l
	AcOH	18.0 mmol/l	1.14 ml/l
	$Na_2EDTAx2H_2O$	1.2 mmol/l	0.45 g/l
	Make as 50x and store at room temperature.		
FSUDS:	bromophenol blue		0.25%
	xylene cyanol		0.25%
	Ficoll 400		15.00%

$\lambda_{HindIII}$ standards[9] :		low range standards:	
	23.1 kb	(BIORAD # 170-3465)	1746 bp
	9.4 kb		1434 bp
	6.6 kb		800 bp
	4.4 kb		634 bp
	2.3 kb		303 bp
	2.0 kb		279 bp
	0.6 kb		249 bp
	0.1 kb		222 bp
			88 bp

DNA-Polyacrylamide Gel Electrophoresis

Polyacrylamide gels are used to analyze fragments of DNA smaller than 1 kb. Polyacrylamide concentrations are chosen as follows:

Acrylamide (%)	Range of Separation (bp)
3.5	100-1000
5.0	80-500
8.0	60-400
12.0	40-200
20.0	10-100

Procedure

1. Set up vertical slab gel according to manufacturer's instructions.
2. Make up gel mix and degas for 15 min using water aspirator.
3. Add 30 µl TEMED[10] per 100 ml mix and pour gel. Keep remaining mix on ice to slow down polymerization; this leaves some mix in case of a leak. Place comb immediately without catching bubbles.
4. Allow to polymerize for 1 h, remove comb, flush wells with buffer, and assemble gel set-up.
5. Add FSUDS to 10% to the DNA samples and load. Up to 1 µg DNA per band can be loaded in a well per 0.1 cm^2 horizontal surface.
6. Run at constant voltage (25 - 100 V depending on gel box).
7. At the end of the run, soak gel in staining solution for 45 min. Blot off excess liquid and visualize DNA on a transilluminator (sensitive to 10 ng DNA per band).

The following table shows the approximate size of DNA fragments with which the dyes co-migrate:

Percentage of gel	Bromophenol Blue	Xylene cyanol
3.5	100	460
5.0	65	260
8.0	45	160
12.0	20	70
20.0	12	45

Reagents

acrylamide stock:	acrylamide	29 g
	BIS[11]	1 g
	H_2O to	100 ml
	Store in the dark at 4° C for up to 1 month.	
APS:	$(NH_4)_2S_2O_8$	3.0 g
	H_2O to	10.0 ml
	Store at 4° C for up to 1 week.	

gel mixes:

Reagents/ml	3.5%	5.0%	8.0%	12.0%	20.0%
30% acrylamide	11.6	16.6	26.6	40.0	66.6
H_2O	76.3	71.3	61.3	47.9	21.3
3% APS	2.1	2.1	2.1	2.1	2.1
10xTBE	10.0	10.0	10.0	10.0	10.0
Total	100.0	100.0	100.0	100.0	100.0

Staining solution: 0.5 mg EtBr per liter buffer

Isolation of DNA from Gels

In both methods, yields vary significantly with size of the DNA and between experiments. Therefore, quantitation after isolation is recommended.

Agarose Gels - Phenol Method

Procedure

1. Run gel in SEB buffer.
2. Cut band on the transilluminator using a sterile razor blade. Use long wavelength setting.

3. Chop band into small pieces on Saranwrap.
4. Transfer to microfuge tube, add 1 volume phenol, and vortex.
5. Leave at -20° C for 20 min.
6. Centrifuge for 10 min and transfer aqueous layer to fresh tube.
7. Extract with phenol and back-extract with $CHCl_3$.
8. Add NH_4AcO to 2 mol/l and precipitate DNA with 2.5 volumes of cold EtOH. Leave at -20° C for 15 min and centrifuge for 20 min. Discard supernatant, add 1 ml 80% EtOH, vortex, and re-precipitate.
9. Spin for 20 min, briefly invert microfuge tubes on a paper towel, and dry under vacuum.
   Resuspend in suitable volume of 0.1xTE.

Reagents

NH_4OAc: Use stock solution at 10.5 mol/l NH_4AcO (pH 8).

Agarose Gels - Glass Milk™ Method

This method is based on proprietary reagents (BIO101 Inc., 1986). All solutions come in a kit. Use of SEB as electrophoresis buffer is recommended. Keep DNA away from light while it is in contact with ethidium bromide.

Procedure

1. Make up working solutions according to the manual.
2. Excise DNA band from EtBr stained SEB-agarose gel using a new razor blade. Cut into smaller pieces and transfer to pre-weighed centrifuge tube
3. Weigh slice and add 2-3 volumes NaI solution to reduce melting point of gel and avoid denaturation of DNA (Hamaguchi & Geiduschek, 1962). Place on 45-55° C bath until agarose is just melted (3-5 min).
4. Vortex insoluble silica matrix until all is suspended. Add glassmilk:

   for less than 0.5 µg DNA: 5 µl glassmilk
   for more than 0.5 µg DNA: 5 µl glassmilk + 1 µl per additional 0.5 µg DNA

Shake suspension gently for 5 min (very gently if DNA ≥ 10 kb).

5. Spin for 5-10 s, discard supernatant, and resuspend pellet in 190 µl NEW buffer.
6. Repeat step 5 two more times.
7. After the last centrifugation remove supernatant quantitatively.
8. Add a volume of 0.1xTE that is three-quarters the volume of glassmilk used. Incubate for 2-3 min at 45-55° C. Centrifuge for 30 s; save supernatant.
9. Repeat step 8. Pool supernatants. Silica that is carried over will not interfere in low salt solutions; spin tube in centrifuge before taking out aliquot of DNA.

EtOH Precipitation

1. For a given DNA stock, add 40% of 8.4 mol/l NH_4OAc and two volumes cold, abs. EtOH.
2. Incubate at -20° C for 20 min.
3. Pellet in microfuge (20 min).
4. Resuspend in 80% EtOH.
5. Pellet as before.
6. Dry pellet and resuspend in TE buffer.

Vector Dephosphorylation

Procedure for BAP[12]

1. To 50 µl of restriction digest add: 5 µl 10xBAP buffer
   1 U BAP
2. Incubate at 65° C for at least 1 h.
3. Isolate DNA in the usual manner (2 phenol extractions [!], 1 $CHCl_3$ extraction, EtOH precipitation).

Note that BAP will not be heat-inactivated easily (if at all).

Procedure for CIP[13]

Calf intestinal phosphatase is used as recommended by the manufacturer (Boehringer Mannheim Biochemicals, Inc.).

1. For a 50 µl reaction: 1-20 pmol DNA termini
   5 µl 10x CIP buffer
   0.1 U CIP
2. Incubate at 37° C for 30 min, then heat-inactivate at 68° C for 15-20 min.

[Rule of thumb: 1 µg of a 3 kb piece linear DNA has 1 pmol 5' ends.]
CIP can also be used directly in the restriction digest mix:

A. For 1 pmol of 5' protruding ends, add 0.01 units CIP and incubate at 37° C for 30 min.
B. For 1 pmol of blunt or 3' protruding ends, add 1 unit of CIP and incubate at 50° C for 1 h.
C. At the end, stop reaction by adding 1/10 volume of 500 mmol/l EDTA (pH 8) and incubating at 65° C for 45 min.

Procedure for HK™ Phosphatase

This method is most convenient as the enzyme can be inactivated by heating at 68° C for 30 min (Epicentre Technologies, 1990).

1. Digest DNA in TA buffer.
2. Add $CaCl_2$ to 5 mmol/l and adjust temperature to 30° C.
3. Add one unit for each microgram of plasmid DNA digested with a single cutter.
4. Incubate at 30° C for 1 h.

Reagents

10x BAP buffer:0.5 mol/l Tris (pH 8.0) & 0.5 mol/l NaCl
10x CIP buffer:0.2 mol/l Tris (pH 8.0) & 0.01 mol/l $MgCl_2$ & 0.01 mol/l $ZnCl_2$

Ligations

T4 DNA Ligase activity is inhibited by metal chelators, EtOH, phosphate, NaCl above 25 mmol/l, and Ca^{2+}. Therefore, it is important to dry DNA pellets well and to resuspend DNA in H_2O or 0.1xTE. Precipitation with NH_4OAc is recommended.

Unit definition for T4 DNA Ligase varies between manufacturers. Units given here are as defined by Weiss and adopted by IBI, Inc. (Weiss et al., 1969).

Protocols are optimized for fragments from 1-10 kb. If there is a greater disparity (linkers, etc.), special protocols have to be used (Ferretti & Sgaramella, 1981).

Procedure for Blunt Ends

1. In a microfuge tube, combine vector and insert DNA at a ratio of 1:3 to achieve a total of 200 ng DNA.
2. Add 2 µl 10xLigase Reaction buffer
   2 units T4 DNA Ligase
   H_2O to 20 µl.
3. Incubate for 2 h at 20° C.
4. Transform.

Procedure for Sticky Ends

1. In a microfuge tube, combine vector and insert DNA at a ratio of 1:2-1:3 to give a final concentration of 2.5 µg/ml.
2. Add 2 µl 10x Ligase buffer
   0.01 unit T4 Ligase
   H_2O to 20 µl.
3. Incubate at 4° C for 12-16 h.
4. Transform.

Procedure for Fragments with Sticky and Blunt Ends

1. In a microfuge tube, combine vector and insert DNA at a ratio of 1:1-1:2 to give a final concentration of 30 ng/µl.
2. Add 2 µl 10x Ligase buffer
   1 unit T4 Ligase
   H_2O to 20 µl.
3. Incubate at 12-14° C for 12-16 h.
4. Transform.

Reagents

T4 DNA Ligase 10x reaction buffer:	250 mmol/l	Tris (pH 7.8)
	100 mmol/l	$MgCl_2$
	40 mmol/l	ß-mercaptoethanol
	4 mmol/l	$MgATP^{14}$

Competent Cells

Calcium Chloride Method

Procedure

1. Dilute 2 ml of an overnight culture into 250 ml of YT or LB medium.
2. Grow for 3-4 h at 37° C with mild shaking (mid-log: Klett of 40 with green filter or A_{660nm} of 0.1-0.4).
3. Cool cell culture on ice for 15 min; spin down cells (GSA, 4° C, 5,000 rpm, 10 min).
4. Resuspend cells in 25 ml calcium chloride solution; leave at room temperature for 20 min.
5. Spin down cells (GSA, 4° C, 5,000 rpm, 10 min) and resuspend in 2.5 ml calcium chloride solution.
6. Add 80% glycerol to final concentration of 15-20%. Store at -80° C.

If the cells are left at 4° C, competency increases for about 24 h before decreasing dramatically. Each new batch of competent cells should first be checked for contamination.

Reagents

YT medium:	peptone	8 g/l
	yeast extract	5 g/l
	Adjust pH to 7.5.	
$CaCl_2$ solution:	Tris (pH 8.0 at 4° C)	10 mmol/l
	$CaCl_2$	50 mmol/l

Rubidium Chloride Method

This method works for most strains of *E. coli* and for some of *K. pneumoniae* (Hanahan, 1983).

Procedure

1. Dilute an overnight culture 1:100 into SOB (250 ml final volume).
2. Grow at 37° C with moderate agitation until $A_{550nm} \approx 0.3$.
3. Place culture on ice bath for 15-20 min.
4. Pellet cells (3000 g for 5 min, 4° C [4.5 k in GSA]) and discard supernatant.
5. Resuspend gently in TB1 (80 ml per 250 ml culture) and leave on ice for 15 min.
6. Centrifuge cells as before and resuspend in TB2 (20 ml per 250 ml culture).
7. Flash-freeze in small aliquots in liquid nitrogen or better, in an EtOH/dry ice bath. Store at -80° C.

Reagents

KOAc:	KOAc (pH 7.5)	1.0 mol/l
MOPS[15]:	MOPS (pH 6.8)	0.5 mol/l

TB1:	RbCl	100.0 mmol/l	12.0 g/l
	$MnCl_2$	50.0 mmol/l	9.9 g/l of 4-hydrate
	KOAc	30.0 mmol/l	30 ml 1 mol/l stock (pH7.5)
	$CaCl_2$	10.2 mmol/l	1.5 g/l of the 2-hydrate
	glycerol	1.6 mol/l	150.0 g/l

Adjust pH to 5.8 with 0.2 mol/l AcOH.

Sterilize by filtration through pre-rinsed 0.22 µm filter and keep cold.

TB2:	MOPS	10.0 mmol/l	20 ml 0.5 mol/l stock (pH 6.8)
	RbCl	10.0 mmol/l	1.2 g/l
	$CaCl_2$	75.0 mmol/l	11.0 g/l of the 2-hydrate
	glycerol	1.6 mol/l	150.0 g/l

Sterilize by filtration through pre-rinsed 0.22 µm filter and keep cold.

SOB medium:	tryptone	20.0 g/l
	yeast extract	5.0 g/l
	NaCl	0.6 g/l
	KCl	0.5 g/l

Sterilize by autoclaving; before use add 10 ml of a filter-sterilized, 2 mol/l stock of Mg^{2+} (1 mol/l $MgCl_2$, 1 mol/l $MgSO_4$) to give 20 mM.

Transformations

Standard Method

Procedure

1. Thaw competent cells on ice. Chill one polypropylene tube (Falcon 3033) per transformation.
2. Add 150-200 µl of competent cells to pre-chilled tube.
3. Add up to 20 µl DNA with gentle swirling.
4. Leave on ice for 40 min.
5. Heat shock at 42° C for 45 s. Do not shake!
6. Place on ice for 2 min before adding 1 ml LB. Grow at 37° C for 1 h.

7. Plate 0.1 ml on selective plate with appropriate drug. Centrifuge remainder of culture and resuspend pellet in 50 µl LB. Plate on selective plate with appropriate drug.
8. Invert plates and incubate at 37° C overnight.

For transformations with phage DNA (single-stranded or double-stranded), steps 6-8 are as follows:

6. To each tube with DNA add 3 ml molten H soft top agar (42° C) containing 200 µl log phase *E.coli*, 40 µl X-gal, and 40 µl IPTG. Mix by rolling tube.
7. Pour onto pre-warmed nutrient plate (37° C) and leave to set (15 min).
8. Invert plates and leave at 37° C overnight.

All transformations should be accompanied by suitable controls: at the very least, one positive control with purified DNA and one negative control without DNA (one per drug combination) should be included. For each ng of DNA used in the transformation, more than 100 colonies should be observed.

Reagents

LB medium:	tryptone	10 g/l
	yeast extract	5 g/l
	NaCl	10 g/l
	Adjust pH to 7.5 and sterilize by autoclaving.	
LB plates:	Add 15 g agar to 1 l of medium before autoclaving.	
indicator plates:	• Cool molten agar to 45° C.	
	• Add X-gal to 40 µg/ml.	
	• Add IPTG to 200 µg/ml.	
	• Pour plates and store at 4° C for up to 1 month.	
ampicillin stock:	[500x for *E. coli*, 200x for *Klebsiella pneumoniae*]	
	• Add 5 g ampicillin to 80 ml H_2O.	
	• Add 0.2 n NaOH until solution turns clear.	
	• Bring up to 100 ml with H_2O and filter sterilize.	
	• Store at -20° C or 4° C.	
tetracycline stock:	[1000x for *E. coli*, 1000x for *K. pneumoniae*]	
	• Add 1.5 g/100 ml 50% EtOH.	
	• Filter-sterilize and store wrapped in foil at 4° C.	

X-gal:	2% in DMF
	Prepare just before use.
IPTG:	100 mmol/l in H_2O (23.8 mg/ml)
	Prepare just before use.

Freeze-Thaw Method

This method is most suitable for transformation of *K. pneumoniae* strains (Merrick et al., 1987).

Procedure

1. Add 0.5 ml of an overnight LB culture to 4.5 ml LB and incubate with shaking at 32° C for 2 h.
2. Centrifuge in benchtop centrifuge and resuspend pellet in 0.4 ml of cold $CaCl_2$ buffer by pipetting. Transfer to 15 ml Falcon polypropylene centrifuge tube.
3. Split into 2 aliquots, add 0.5 - 1.0 µg DNA to one aliquot, and flash-freeze both in liquid nitrogen.
4. Thaw cells at 32° C and repeat freeze-thaw cycle once more.
5. Transfer cells to 5 ml LB and grow at 32°-37° C until visibly growing (1-2 h).
6. Plate dilutions and grow at 37° C overnight (e.g., 0.1 ml and 4.9 ml).

Reagents

$CaCl_2$ buffer:	50 mmol/l $CaCl_2$	
	10 mmol/l Tris (pH 7.4)	
LB medium:	tryptone	10 g/l
	yeast extract	5 g/l
	NaCl	10 g/l
	Adjust pH to 7.5 and sterilize by autoclaving.	
LB plates:	Add 15 g agar to 1 l of LB medium before autoclaving.	

For *Klebsiella pneumoniae* use ampicillin at 1 mg/ml on plates and 0.25 mg/ml in liquid culture (Wahl & Orme-Johnson, 1987); tetracycline can be used at the same concentration as for *E. coli*.

DNA Quantitation

Absorption Method

The first procedure is most useful with larger amounts of DNA, e.g., after DNA isolation by CsCl gradient. The second method is much more sensitive and is used for DNA minipreps and DNA isolated from gels.

Procedure

1. Dilute 20 µl of the double-stranded DNA solution with 380 µl TE (dilute single-stranded DNA 1:23).
2. Read A_{260nm} against a TE blank.
3. The absorption equals the DNA concentration in µg/µl.
   [dsDNA in mg/ml] = $A_{260nm}/20$
   [ssDNA in mg/ml] = $A_{260nm}/23$

Reagents

TE: 10 mmol/l Tris (pH 8.0)
1 mmol/l EDTA

Fluorescence Method

The fluorometer used for this experiment is the Hoefer TKO 100 (emission at 365 nm, 100 nm bandwidth; detection at 460 nm, 10 nm bandwidth).
The sample should not contain SDS above 0.01% or ethidium bromide, because they quench the fluorescence of Hoechst dye 33258 (Brunk et al., 1979, and Labarca & Paigen, 1989).

Procedure

1. Allow fluorometer to warm up for 30 min.
   Make up a working dye solution by adding 10 µl of the stock Hoechst dye 33258 to 100 ml of 1xTNE buffer (0.1 µl/ml).
2. Adjust by zeroing against buffer blank (scale at maximum).

3. Adjust scale to 250 for a sample of 2 µl High Standard in 2 ml buffer.
4. Establish calibration curve using blank, Low, Medium, and High Standard. Dilute each standard 1:1000.
5. Dilute sample 1:1000 and take reading. Read concentration from calibration curve. For each standard/sample, re-zero with 2 ml blank, then add 2 µl sample, mix, and take reading. If the reading decreases noticably in a time dependent manner, this is due to photo bleaching and the initial value should be used.

For best results, the DNA used for standardization should be the same kind as assayed, i.e., the standards for single-stranded DNA should be made up with single-stranded DNA, double-stranded DNA should be measured against double-stranded DNA, and so on.

Reagents

standards:	250 µg/ml	High Standard	250 ng/ml final concentration
	100 µg/ml	Medium Standard	100 ng/ml final concentration
	25 µg/ml	Low Standard	25 ng/ml final concentration

dye buffer: 1 µl/ml of Hoechst dye 33258 in 1xTNE

1xTNE:	Tris (pH 7.4)	10 mmol/l
	EDTA	1 mmol/l
	NaCl 100 mmol/l	
	Make up as 10x and filter to remove particulate matter.	
	For crude DNA preparations increase salt concentration.	
stock dye:	Hoechst 33258	10 mg
	H_2O to	10 ml

Isolation of Double-Stranded DNA

Large Scale Prep

This procedure will typically yield 1-5 mg of closed-circular double-stranded DNA, depending on how much culture is used (250 ml - 500 ml).

Procedure

1. Grow culture in 250 - 500 ml LB with appropriate drug overnight at 37° C. Use a shaker flask with baffles that is 4-5 times larger than the culture. Take an aliquot for glycerol stocks (optional).
2. Spin down cells (GSA, 10 min, 4° C, 5,000 rpm) and discard supernatant. Repeat with more culture is desired.
3. Resuspend in 5 ml glucose solution containing lysozyme at 5 mg/ml and RNAse A at 10 µg/ml; incubate at room temperature for 10 min (add fresh lysozyme !).
   Add 16 ml fresh NaOH/SDS and leave on ice for 10 min.
   Add 12 ml cold potassium acetate solution and leave on ice for 10 min.
6. Spin down cell debris (SS34, 20 min, 17,000 rpm, 4° C).
   Transfer supernatant (through cheesecloth) into fresh tube.
7. Add 2.5 volumes cold, abs. EtOH (82.5 ml), centrifuge (GSA, 20 min, 11,000 rpm, 4° C), discard supernatant, and dry centrifuge tube briefly.
8. Resuspend pellet in 9.1 ml TE (heat to 65° C if necessary) and transfer to 15 ml centrifuge tube; add 0.479 ml ethidium bromide and 9.6 g CsCl.
   Incubate for 20-30 min at room temperature. Keep tubes in the dark!
11. Centrifuge in SS34 rotor (25 min, 4° C, 16,000 rpm).
    Transfer to Quickseal tube and balance with a solution of CsCl in H_2O (density of 1.59 mg/ml).
12. Seal tubes and spin (10 h at 55,000 rpm, then 1 h at 45,000 rpm, 19° C).
13. Pull 1-2 ml of bottom band. Extract with 2 volumes of iPrOH[16] until aqueous phase is colorless. Add 2 volumes TE.
14. Add 2 volumes cold EtOH, leave at -20° C for 30 min, and centrifuge (SS34,

10,000 rpm, 4° C, 30 min). Discard supernatant and resuspend pellet in 10 ml cold 80% EtOH. Leave at -20° C for 30 min and centrifuge as before. Discard supernatant and dry pellet in vacuo.

15. Resuspend in 1-2 ml TE depending on size of band in gradient.

Reagents

EtBr solution:	Dissolve solid EtBr in H_2O to a final concentration of 10 mg/ml. Wipe up spills with bleach. Destroy 100 ml of a solution by adding 20 ml 5% hypophosphorous acid and 12 ml 0.5 mol/l sodium nitrite solution; after 20 h neutralize with sodium carbonate and discard (Lunn & Sansone, 1987).	
glucose solution:	Tris (pH 8.0)	25 mmol/l
	glucose	50 mmol/l
	EDTA	10 mmol/l
NaOH/SDS:	(Make fresh each day from 2 n NaOH and 10% SDS.)	
	NaOH	0.2 normal
SDS	1.0 %	
KAcO:	5 mol/l KAcO	600 ml
	AcOH	115 ml
	H_2O	285 ml
cheesecloth:	Fold until four-layered, then autoclave between tin foil.	
iPrOH:	Mix 100 ml iPrOH & 100 ml H_2O; add CsCl untilsaturated.	
terrific broth:	H_2O	900 ml
	tryptone	12 g
	yeast extract	24 g
	glycerol	4 ml
	Sterilize and add 100 ml of 0.17 mol.l KH_2PO_4 0.72 mol/l K_2HPO_4.	

General Miniprep

Procedure

1. Start 3 ml cultures (LB, drugs) from colonies on master plate.
2. Spin 1.5 ml in microfuge for 5 min. Discard supernatant. Repeat with second 1.5 ml aliquot of same culture if desired.
3. Resuspend pellet in 100 µl glucose buffer. Incubate for 5 min at room temperature.
   Add 200 µl NaOH/SDS. Invert to mix. Incubate on ice for 5 min.
   Add 150 µl cold KAcO, mix, and incubate on ice for 5 min.
4. Centrifuge for 10 min and transfer supernatant to fresh tube.
5. Extract with one volume phenol, then back-extract with $CHCl_3$.
6. Fill tube with cold, abs. EtOH, and leave at -20° C for 10 min.
   Centrifuge for 15 min to pellet DNA. Discard supernatant and resuspend in 1 ml 80% EtOH. Repeat precipitation.
7. Dry pellet in vacuo and resuspend in 50 µl TE containing 0.25 µl RNAse. Heat at 65° C for 10 min. Store at 4° C.

Reagents

NaOH/SDS:	make fresh each day from 2 n NaOH and 10% SDS:	
	NaOH	0.2 n
	SDS	1.0 %
glucose solution:	Tris (pH 8.0)	25 mmol/l
	glucose	50 mmol/l
	EDTA	10 mmol/l
KAcO:	5 mol/l KAcO	600 ml
	AcOH	115 ml
	H_2O	285 ml

RNAse: Dissolve 100 mg pancreatic RNase A in 1 ml H_2O. Boil for at least 15 min, allow to cool slowly, and freeze in small aliquots at -20° C.

Miniprep for Rapid Screening

This protocol is most suitably employed when screening a large number of clones. The DNA obtained should not give any problems except with the most sensitive enzymes (Morelle, 1989).

Procedure

1. Grow bacteria overnight in LB medium with drugs.
2. Harvest by centrifugation (9000 rpm, 1.5 min) and resuspend in 0.2 ml lysis buffer by vortexing.
3. Leave at room temperature for 5 min and add 0.4 ml of freshly prepared alkaline solution; mix by inverting tube. The solution should now become clear and viscous. Leave on ice for 5 min.
4. Add 270 µl of NH_4AcO solution and mix by inverting tube.
5. Leave on ice for 15 min to allow high molecular weight fractions to precipitate.
6. Spin for 5 min at top speed.
7. Transfer supernatant to fresh tube, add 450 µl iPrOH, and incubate at room temperature for 10 min.
8. Centrifuge at 12,000 rpm for 15 min, discard supernatant, and wash pellet with 1 ml 70% EtOH.
9. Dry tubes by leaving them inverted on tissue paper for 15 min.
10. Resuspend in 50 µl TE. Add 1 µl RNAse.
11. Heat at 68° C for 5 min and spin for 2-3 min before taking out aliquot. Use 10 µl for one restriction digest in a total volume of 20 µl.

Reagents

lysis buffer:	50 mmol/l glucose
	25 mmol/l Tris (pH 8.0)
	10 mmol/l EDTA
	4 mg/ml lysozyme
alkaline solution:	0.2 n NaOH (from 2 n stock)
	1% (w/v) SDS (from 10% stock)
	Prepare fresh each time.
NH_4OAc solution:	8.4 mol/l (without pH adjustment)

CTAB Miniprep

This procedure is the easiest way to purify double-stranded DNA for DNA sequencing (Biology Department, 1989 and Del Sal et al., 1988).

Procedure

1. Grow overnight culture in LB with appropriate drugs.
2. Collect cells by centrifugation (8,000 rpm, 3-4 min).
3. Resuspend cells in STET buffer.
4. Add 4 µl of lysozyme at 50 mg/ml and incubate at room temperature for 5 min.
5. Boil for 40 sec.
6. Microfuge at top speed for 10 min.
7. Remove pellet with sterile toothpick and add 5 µl of RNAse A at 10 mg/ml. Incubate at 65° C for 15 min.
8. Add 10 µl CTAB solution and microfuge for 5 min.
9. Resuspend pellet in 300 µl of 1.2 mol/l NaCl.
10. Precipitate DNA by adding 750 µl cold absolute EtOH. Spin for 10 min.
11. Aspirate EtOH, vortex pellet with 1 ml 70% EtOH, spin for 10 min, aspirate, and dry pellet.
12. Resuspend in 8 µl H_2O.

This will give enough DNA for one sequencing reaction (template and primer).

Reagents

STET:	sucrose	8.0 %
	Tris (pH 8)	50.0 mmol/l
	EDTA	50.0 mmol/l
	Triton X-100	0.1 % (v/v)

CTAB solution:5% (w/v) of hexadecyltrimethylammonium bromide in H_2O

Isolation of Single-Stranded DNA

Large Scale Prep - M13 Phage

Procedure

1. Pick a single colony of host cells and grow overnight in 10 ml 2xTY medium (shake at 37° C).
   Add one drop to 20 ml fresh medium and shake at 37° C for 3 h.
   Inoculate 1 ml of 2xTY medium with 0.1 ml of three-hour culture.
2. Inoculate this 1 ml culture with fresh plaque using wooden toothpick.
   Grow for 4 h with shaking at 37° C, spin down cells, and store supernatant at 4° C for next day. Set up overnight culture.
3. Inoculate 100 ml 2xTY medium with 1 ml of overnight culture and grow to A_{660nm} of ≈ 0.3. Inoculate with phage stock from step 2.
4. Grow for 4 h at 37° C with vigorous shaking.
5. Spin down cells (GSA, 10,000 rpm, 4° C, 30 min). Transfer supernatant to fresh tube and add 0.2 volumes PEG/NaCl. Leave at 4° C for 1 h.
6. Centrifuge as before and discard supernatant. Re-centrifuge for 5 min and remove all of PEG/NaCl. Resuspend viral pellet in 0.5 ml TE and transfer to microfuge tube.
7. Centrifuge for 15 min to remove remaining cells. Transfer supernatant to fresh tube and add 200 µl PEG/NaCl. Mix, then leave at room temperature for 15 min or at 4° C overnight.
8. Centrifuge for 15 min. Discard supernatant. Centrifuge for 5 min and remove rest of PEG/NaCl. Resuspend in 0.5 ml TE.
9. Extract twice with 1 volume phenol.
10. Extract three times with 1 volume diethylether.
11. Extract twice with 1 volume $CHCl_3$.
12. Split sample and add 625 µl EtOH and 25 µl NaAcO[17]. Invert tube several times.
13. Leave at -20° for 20 min, centrifuge for 20 min, and discard supernatant.
14. Resuspend pellet in 1 ml 80% EtOH and precipitate as before.
15. Dry pellet and resuspend in 50 µl TE. Store at 4° C.

This prep can be scale up to 250 ml by scaling up amounts in the infection steps. In the microfuge tube steps, one additional phenol extraction should be included.

Reagents

2xTY medium:	tryptone	16 g/l
	yeast extract	10 g/l
	NaCl	5 g/l
PEG/NaCl:	Polyethylene glycol 6000	200 g/l
	NaCl	146 g/l
NaOAc:	3 mol/l, pH 6	

Large Scale Prep - Phagemids

Procedure

1. Begin with an overnight culture started from an isolated colony.
2. Add 20 µl of overnight culture to 20 ml 2xTY and shake at 37° C for 3 h (to mid-log).
3. Inoculate 1 ml 2xTY medium with 100 µl of log culture; also inoculate with single plaque and grow for 4 h at 37° C with vigorous shaking.
4. Spin down cells and store supernatant at 4° C. This will be the phage starter. Start overnight cell culture.
5. Dilute 1 ml of overnight culture into 100 ml 2xTY, grow to $A_{550} \approx 0.3$, add phage starter culture, and grow for another 4 h with vigorous aeration.
6. Centrifuge (5,000 g, 30 min, 4° C), discard pellet, and add 0.2 volumes PEG solution. Mix and leave at 4° C for 1 h.
7. Spin as before for 20 min. Discard all of supernatant; respin briefly if necessary.
8. Resuspend in 0.5 ml TE, transfer to microfuge tube, and spin for 10 min to remove last of cells.
9. In a fresh microfuge tube, add 0.2 ml PEG solution and leave for 15 min to overnight at 4° C. Centrifuge for 5 min and again remove all of supernatant.

Resuspend in 0.5 ml TE.

10. Extract twice thoroughly with 0.2 ml phenol, then three times with 0.5 ml diethylether, and twice with 0.5 ml $CHCl_3$.
11. Divide sample into two microfuge tubes, add to each 0.625 ml EtOH and 25 µl 3 mol/l NaAcO, chill, and spin for 15 min. Wash pellet with 1 ml 70% EtOH and dry briefly in vacuum.
12. Resuspend each pellet in 50 µl TE, pool aliquots, and determine concentration.
13. Store at 4° C.

Reagents

2xTY medium:	tryptone	16 g/l
	yeast extract	10 g/l
	NaCl	5 g/l
PEG solution:	PEG-6000	20%
	NaCl	2.5 mol/l
alternate PEG solution:	PEG-6000	20%
	NH_4AcO (pH 7.5)	3.5 mol/l

Miniprep for M13 Derivatives

Procedure

1. Pick a well-isolated plaque and grow it in 3 ml of a 1:100 dilution of an overnight culture of JM101 or another suitable host in YT (5-6 h, 37° C with vigorous shaking).
2. Remove cells by centrifuging twice in a microfuge for 10 min; remove supernatant to fresh Eppendorf tube.
3. To 1.2 ml of supernatant add 350 µl of PEG solution, vortex, leave at room temperature for 15 min, and spin for 15 min.
4. Discard supernatant, resuspend pellet in 0.6 ml of TE/PEG solution and incubate at room temperature for 15 min.

5. Spin as before, remove supernatant, respin for 1 min, and remove rest of supernatant quantitatively.
6. Resuspend in 150 µl TE.
7. Extract with an equal volume of phenol twice (vortex for 20 s, leave on bench for 15 min, spin for 5 min, and discard organic layer). It is important to remove all of the interphase! Optional: after the first extraction, remove the aqueous layer to a fresh tube.
8. Back-extract once with an equal volume of $CHCl_3$ (repeat if interphase is observed).
9. Add 10 µl NaAcO solution and 2.5 volumes of cold EtOH, and spin for 20 min at room temperature.
10. Discard supernatant, add 1 ml cold 80% EtOH, vortex, and leave at -20° C for 15 min.
11. Spin as before. Invert tubes briefly on a paper towel and dry under vacuum. Resuspend in 25 µl TE.

Reagents

PEG/TE solution:	5 parts TE	
	1 part PEG solution	
chloroform:	24 parts $CHCl_3$: 1 part isoamylalcohol	
NaOAc solution:	3 mol/l NaAcO (pH 5)	
YT medium:	peptone	8 g/l
	yeast extract	5 g/l
	Adjust pH to 7.5.	
TE:	Tris (pH 8.0)	10.0 mmol/l
	Na_2EDTA	1.0 mmol/l
PEG solution:	PEG-6000	20.0 % w/v
	NaCl	2.5 mol/l

Miniprep for Phagemid Vectors

Procedure

Small scale preps for single-stranded DNA from phagemids don't seem to work well (Stratagene Cloning Systems, 1988). If possible do double-stranded sequencing. In any case, proper titration of the helper phage and thorough aeration are critical to the yield of phage.

1. Grow a small overnight culture of the clone of interest under conditions selecting for episome and phagemid.
2. Add enough of this culture to 3 ml fresh superbroth (SB) to give an OD of 0.1 at 600 nm. Grow with shaking in a 50 ml tube until $A_{600} \approx 0.3$.
3. Add helperphage at a ratio 20:1 phage to cells (ratios of 1:10, 1:1 and 10:1 may also work, depending on clone).
4. Shake vigorously at 37° C for 8 h. At this point, the culture may be stored at 4° C overnight, if it is first heated to 65° C for 15 min and then centrifuged (11,000 g, 2 min).
5. Remove cells by centrifuging twice.
6. Add 300 µl PEG solution to 1.2 ml supernatant, vortex, and leave at room temperature for 15 min. Centrifuge for 20 min.
7. Remove supernatant and resuspend in 4 parts TE and 1 part PEG solution.
8. Collect phage as before and remove all of supernatant.
9. Resuspend in 300 µl TE, extract twice with 200 µl phenol (vigorous vortexing!), and once with $CHCl_3$. Extractions must be repeated until no material is left at interface.
10. Add 200 µl NH_4AcO[18] and 800 µl cold EtOH. Chill for 10 min and centrifuge for 20 min in microfuge. Wash pellet with 1 ml 80% EtOH, dry in vacuo, and resuspend in 20 µl TE. Avoid taking up residue at the side of the tube.
11. Transfer to fresh tube and analyze aliquot by gel electrophoresis.

Yield should be 1-2 µg per 3 ml prep. Helperphage R408 runs approximately at 4 kb when compared to double-stranded standard; pBluescript without insert runs approximately at 1.6 kb.

alternate procedure:

1. Early in the morning, start a culture of bacteria containing the phagemid from a selective plate.
2. Late in the afternoon, add enough of this culture to 3 ml SB to give an OD of 0.05 at 600 nm.
3. Grow culture to an OD of 0.1 (30-60 min). Add helperphage at a low multiplicity of infection (1:10 or 1:1).
4. Grow overnight with shaking at 37° C.
5. Isolate phage as described above.

Reagents

PEG solution:	PEG-6000	20%
	NH_4AcO (pH 7.5)	3.5 mol/l
NH_4AcO:	NH_4AcO (pH 7.5)	7.5 mol/l
SB media:	tryptone	32 g/l
	yeast extract	20 g/l
	NaCl	5 g/l
	1 n NaOH	5 ml

Preparation of Helperphage

Procedure

It is important in this prep to start with a homogeneous phage population.

1. Begin with an overnight culture of DH5αF' in 2xTY.
2. Inoculate 0.75 l SB media in a 2 l flask with 15 ml overnight culture.
3. Let grow with shaking at 37° C until $A_{600} \approx 0.1$.
4. Inoculate with 20 plaques from homogeneous phage plate (excise plaques, vortex with 1 ml medium in Eppendorf tube, and use immediately).
5. Grow for 6-9 h or until lysis is complete.
6. Remove cells by centrifugation and add PEG solution to supernatant.
7. Allow to precipitate in cold room overnight.

8. Collect phage by centrifugation and resuspend in 50 ml storage buffer.
9. Titer phage and freeze in small aliquots.

Reagents

PEG solution:	PEG-6000	20%
	NaCl	2.5 mol/l
storage buffer:	Tris (pH 8)	25 mmol/l
	glycerol	20 %

Site-directed Mutagenesis

Site-directed mutagenesis is performed with the mutagenesis kit from Amersham Corporation (Amersham Corporation, 1986) based on the initial experiments by Eckstein and coworkers (Taylor et al., 1985). All reagents and enzymes are supplied with the kit.

1. Add to microfuge tube:	oligonucleotide	50 pmol
	10x kinasebuffer	3 µl
	H_2O to	30 µl

2. Add 2 units T4 polynucleotide kinase. Incubate at 37° C for 15 minutes, then heat-kill at 65° C for 10 min.
3. Store phosphorylated oligonucleotide at -20° C.

4. Add to microfuge tube:	single-stranded template	10 µg
	phosphorylated oligonucleotide	8 pmol
	buffer 1	7 µl
	H_2O to	12 µl

5. Place at 70° C for 3 min, then at 37° C for 30 min. Put annealed mix on ice.

6. Add to reaction:	$MgCl_2$ solution	10 µl
	nucleotide mix	38 µl
	H_2O	12 µl
	Klenow enzyme	12 units
	T4 DNA ligase	12 units

Mix by pipetting and synthesize mutant strand by placing mix at 16° C overnight.

7. The mix is diluted with 170 µl H_2O and 30 µl 5 mol/l NaCl and filtered through nitrocellulose filters to remove single-stranded DNA. Double-stranded DNA in the filtrate is precipitated with 28 µl 3 mol/l NaAcO and 0.7 ml cold, absolute EtOH. Collect the DNA by centrifugation and resuspend in 50 µl buffer 2.
8. 10 µl of filtered sample are diluted with 65 µl buffer 3 and nicked with 5 units of NciI or other suitable enzyme for 90 min at 37° C.
9. To 65 µl of the nicked mix add 12 µl 0.5 mol/l NaCl, 10 µl buffer 4, and 50 units exonuclease III. Place at 37° C for 30 min. Heat-inactivate enzyme for 15 min at 70° C.
10. Add 13 µl nucleotide mix 2, 5 µl $MgCl_2$ solution, 3 units DNA polymerase I, and 2 units T4 DNA ligase. Perform fill-in reaction/ligation at 16° C for 3 h.
11. Transform 300 µl of competent TG1 with 20 µl of mix and store the remainder at -20° C.

DNA-Sequencing

DNA sequencing is a complex technique that requires different approaches for different problems (Sanger et al., 1977, and Bankier & Barrell, 1983). Single-stranded sequencing is the way to go, especially if the template can be isolated from M13 based phage. However, double-stranded sequencing does not require specialized vectors.
M13mp phage and its ssDNA can be isolated easily in large and in small quantities; the problems here are the instability of large inserts and the inconvenience of having to clone the fragment of interest into a new vector just for the isolation of ssDNA. Phagemid vectors avoid these problems, however, small scale ssDNA isolations are tricky. Plasmids yield dsDNA for double-stranded sequencing; oligonucleotide-directed mutagenesis in this system is not easily achieved.

Reactions for DNA-Sequencing using Klenow

Procedure

1. Plasmid DNA:

   Place 5 µg dsDNA in a microcentrifuge tube and dry down in vacuum centrifuge.
   Dissolve DNA in 40 µl 1xdenaturation buffer and allow to stand at room temperature for 5 min.
   Add 4 µl neutralization buffer and mix.
   Immediately add 100 µl 100% cold EtOH (-20° C), mix, and allow to stand for 10 min at -20° C.
   Centrifuge for 15 min at 4° C, discard supernatant, wash with 1 ml 70% EtOH, respin, and briefly dry precipitated DNA under vacuum (can be stored for days at -70° C).
   [Alternatively, denature DNA by adding concentrated NaOH/EDTA stock solution to regular DNA to achieve desired concentrations.]

   Add to the dried and denatured DNA:
   1 µl primer (2.5 pmol)
   1.5 µl 10xds reaction buffer
   2 µl [^{35}S] dATP (16 µCi)
   11.5 µl H_2O

   Anneal mixture at 37° C for 15 min.

   Pipet 2 µl of each of the four corresponding dN/ddN mixes into four labelled microfuge tubes (A,C,G,T).
   Add 1 µl (2U) of Klenow enzyme to each annealing mixture and mix. Pipet 3 µl of this mix into each of the four microfuge tubes, mix, and place on water bath for 20 min.
   Chase with 1.5 µl chase mix and incubate at 37° C for 15 min.
   Stop by adding 4 µl stop buffer.
   It may be advantageous to dry down the sample before adding stop buffer.
   Store at -20° C.
   Load 2 µl on gel.

2. Single-stranded DNA:

Add to 1.5 ml Eppendorf tube:
1 µl M13 17 bp primer (4 ng)
1 µl 10xss polymerase reaction buffer
6 µl miniprep DNA (1-2 µg)
4.4 µl H_2O

Heat water bath to 95° C , place annealing mixture on it, and let cool down to 30° C for 45 min. Spin down once briefly to collect condensation at bottom of tube after 15 min. The primer/template mixture is stable at -80° C for days. Alternatively, anneal in dry oven at 65° C for 1-2 h.

Using a repeating pipettor, place 2 µl of nucleotide mix on the inside of the wells of a microtiter plate (round bottom); arrange plate so that for each annealing mix there will be a set of four wells with A, C, G, and T mix. Cover and store at -20° C until needed.

Add to the annealing mix:
2 µl [^{35}S] dATP at > 600 Ci/mmol
1.5 µl Klenow at 2 U/µl

Spin down and mix with pipet; add 4 µl primer/template mix to each of the dN/ddN wells on the microtiter plate.

Incubate at 37° C for 20 min, add 1.5 µl chase mix, and incubate for another 10 min at 37° C. Add 5 µl stop buffer.

It may be advantageous to dry down the sample before adding stop buffer.

Store at -20° C.

Load 2 µl on gel.

Reagents

All reagents can be made up individually; however, in order to assure freshness of reagents, it may be advantageous to buy reagents in a kit and discard solutions not to be used in the immediate future. Kits are sold by Boehringer Mannheim Biochemicals (ds sequencing), BRL Life Technologies (ss sequencing), and other companies.

misc. items: round bottom 96-well tissue culture plates, repeating pipettors, vacuum centrifuge with microplate-carrying rotor, aluminum blocks for constant temperature baths ('cold blocks')

denaturation buffer:	NaOH	0.2 mol/l
	EDTA	0.2 mmol/l
neutralization buffer:	NH_4AcO (pH 4.5)	2 mol/l
10xss[19] reaction buffer:	Tris (pH 8)	100 mmol/l
	$MgCl_2$	100 mmol/l
	NaCl	300 mmol/l
1xss dilution buffer:	"K_3PO_4" (pH 7.5)	100 mmol/l
	glycerol	50 %
10xds[20] reaction buffer:	Tris (pH 7.5)	100 mmol/l
	NaCl	500 mmol/l
	EDTA	1 mmol/l
	DTT	50 mmol/l
	Store frozen.	
label:	α-[^{35}S] dATP at > 600 Ci/mmol (Amersham's # SJ304)	
chase mix:	0.5 mmol/l of all four dNTP's	
stop buffer:	deionized formamide	10 ml
	xylene cyanol FF	10 mg
	bromphenol blue	10 mg
	0.5 mol/l EDTA	0.2 ml
	Keep for 2-3 months only.	

dN/ddNTP mixes for single-stranded and double-stranded sequencing:

	A°	C°	G°	T°
0.5 mmol/l dCTP	20	1	20	20
0.5 mmol/l dGTP	20	20	1	20
0.5 mmol/l dTTP	20	20	20	1
10xss buffer	20	20	20	20

For dITP reactions, replace the 0.5 mmol/l dGTP with 2 mmol/l dITP.

To get the dN/ddNTP mixes, mix the N° mixes with ddNTP 1:1 (v/v).

A:	A°	+	0.1 mmol/l ddATP
C:	C°	+	0.3 mmol/l ddCTP
G:	G°	+	0.5 mmol/l ddGTP
T:	T°	+	1.0 mmol/l ddTTP

For dITP reactions, substitute the 1 volume of 0.5 mmol/l ddGTP by 1/8th volume of the same. For deaza-guanosine reactions, use c^7-dGTP instead of dGTP, then proceed as usual.

Keep all mixes at -20° C and avoid thawing.

Reactions for DNA-Sequencing using Sequenase™ 2.0

Sequenase 2.0 is a genetic variant of the T7 DNA polymerase created by in vitro manipulation and is the property of USB Corp. Enzyme and most solutions are provided in a kit (United States Biochemical Corporation, 1989). Therefore, only a brief protocol is listed here.

1. For each set of four single-stranded sequencing reactions, a single annealing reaction is used. Combine the following in a microfuge tube:
   1 µl primer (0.5 pmol)
   2 µl 5x reaction buffer
   6 µl ssDNA
   1 µl H_2O to 10 µl total
2. Warm capped tube at 65° C for 2 min, then allow to cool over period of 30 min. Once temperature is below 30° C, annealing is complete. Place tube on ice and use within 4 h. Thaw label.
3. Dilute labeling mix (dGTP) 5-fold with H_2O (keep for several weeks at -20° C).
4. Dilute Sequenase 1:8 in ice-cold dilution buffer for immediate use. Keep on ice for up to 60 min.

5. Place 2.5 µl of ddNTP termination mix (N = A,C,G,T, I) in microfuge tube labeled N. Cap tubes. Keep on ice, but prewarm at 37° C for 1 min immediately before use.
6. To the annealed primer/template mix add:

   [1.0 µl 0.1 mol/l DTT unless DTT was added in miniprep]

   2.0 µl diluted labeling mix

   0.5 µl α-[^{35}S] dATP at 10 µCi/µl and 10 µmol/l (1,000 Ci/mmol)

   2.0 µl Sequenase enzyme (add last)

   Incubate at room temperature for 2-5 min.

   If compressions are expected, use a 1:5 dilution of dITP labeling mix. In this case, run dGTP reactions alongside the dITP reactions as dITP may cause artifacts.
7. Transfer 3.5 µl of labeling mix to each of the four tubes, returning each of the N tubes to the 37° C bath after addition and mixing.
8. Continue incubations for a total of 3-5 min (up to 30 min is fine except for I reactions). Add 4 µl of stop solution, mix, and leave on ice (samples can be stored at -20° C for days to weeks).
9. Heat samples to 80° C for 2 min before loading 2-3 µl per lane.

It may be advantageous to dry down the sample before adding stop buffer.

For plasmid sequencing, only the following modifications are necessary:

Place 5 µl dsDNA (5 µg) in a microcentrifuge tube and dry down in vacuum centrifuge.

Dissolve DNA in 40 µl 1xdenaturation buffer and allow to stand at room temperature for 5 min.

Add 4 µl neutralization buffer and mix.

[Alternatively, add 1/10th volume of 10x buffer and neutralize later by adding neutralization buffer, e.g.:

8 µl miniprep DNA

1 µl 0.1 mol/l DTT

1 µl 10x denat. buffer]

Immediately add 100 µl 100% cold EtOH (-20° C), mix, and allow to stand for 10 min at -20° C.

Centrifuge for 15 min at 4° C, discard supernatant, wash with 1 ml 70% EtOH, respin, and briefly dry precipitated DNA under vacuum (can be stored for days at -70° C).

Add 7 µl H_2O, 1 µl primer (10-15 ng), and 2 µl Sequenase buffer; anneal at

65° C for 2 min, then let cool over 30 min.

Proceed as for single-stranded sequencing.

Simplified Instructions for Sequenase V2.0

This method works in conjunction with the CTAB miniprep procedure (Biology Department, 1989).

1. Resuspend DNA from 2-3 ml CTAB miniprep to give the following mix:
   8 µl H_2O + DNA
   1 µl primer (10 ng)
   1 µl 1 mol/l NaOH
2. Incubate at 68° C for 10 min.
3. Add 4 µl TDMN buffer.
4. Place at room temperature for 10 min.
5. Add to the template/primer mix:
   2 µl labeling mix (1:10 dilution of stock)
   1 µl α-[^{35}S]-dATP
   2 µl Sequenase enzyme (diluted 1:8 in dilution buffer)
   Place at room temperature for 5 min.
6. Add 3.5 µl of the reaction to 2.5 µl of each of the termination mixes (A,C,G,T) pre-warmed to 37° C.
7. Place at 37° C for 10 min.
8. Add 4 µl of stop solution, denature by heating at 90-100° C for 3 min and place on ice until ready to load gel. Load 4 µl only.

TDMN buffer:	TES	3.21 g
	12 mol/l HCl	0.50 ml
	1 mol/l $MgCl_2$	4.00 ml
	5 mol/l NaCl	2.00 ml
	H_2O to	47.50 ml total volume

Before use, add 2.5 µl 1 mol/l DTT to 47.5 µl TDMN stock.

Gel Electrophoresis

Procedure

The assembly described is that of the IBI Corp. BASE RUNNER.

1. Clean plates, combs, spacers, and buffer reservoirs first with non-abrasive soap, then with destilled H_2O.
   Treat thermoplate with Sigmacote and leave to dry in the hood for several hours.
   Spray plates, combs and spacers with 95% EtOH and wipe dry with lint-free Kimwipes.
2. Use double-sided tape to attach spacers to thermoplate so they will not slide when gel is poured.
   Prepare gel mix as described below. Pour using 'sliding plate technique': place thermoplate on a flat, elevated surface with the edges hanging over so the clamps can be attached later without moving the sandwich; pour activated monomer solution between plates while sliding top plate over bottom plate (see sketch; this will require two operators). Capillary action will suck the solution between the plates; problems like bubbles will probably only occur at the bottom of the gel.
   Slide plate up all the way, top up with monomer solution, place clamps along both sides, and insert combs (sharktooth combs go flat side down 0.5 cm between plates and regular combs go as far as indicated by the depth of the wells). Keep monomer solution on ice in case of leaks. Finally, place clamps on glass plates directly on top of combs to ensure tight fit. Allow to polymerize for 1 h or overnight.
3. Take off all clamps except for one at each corner. Remove combs carefully. Wash away bits and pieces of polyacrylamide. Sharktooth combs are now inverted and placed in such a way that the wells are formed by the flat polyacrylamide gel at the bottom and the plastic material of the comb at both sides; the points of the comb should just touch the gel - just enough to give a good seal, but not enough to either pierce the gel or indent it so much that the bands will smile later because they are loaded on an uneven surface.

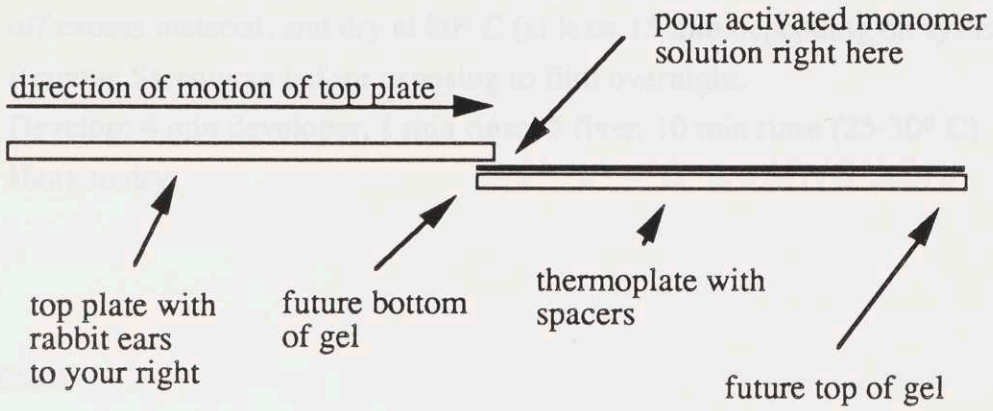


Figure A.1.: DNA Gel Sandwich Assembly

4. The gel sandwich is placed in the plastic rails and the buffer chambers are attached; the chambers are filled with 1xTBE buffer and the wells are rinsed to remove excess urea.
   Pre-electrophorese at 55-60 watts constant power for 30 min.
5. Samples are first heated at 95° C for 5 min and then placed immediately on ice. Before loading samples, the wells have to be rinsed again. After each set of ACGT is loaded, the samples are run into the gel for 1 min. To allow easy orientation of the autoradiogram later, a single lane is loaded as a marker on the far right.
   Suggested loading pattern for difficult sequences:

   dGTP:dITP

   G A T C:T G C A

   This way, all lanes are adjacent to each other and both G and I reactions are run in only 8 lanes.
6. Electrophorese until the desired range of nucleotides is ressolved. Empty buffer chambers and take out gel sandwich. Disassemble the sandwich (thermoplate up!) using a spatula as a lever; the gel should come off the thermoplate and stick to the regular glass plate.
7. Place the glass plate with the gel in a large, solvent-resistant pan with 2 l fixing solution. Shake gently for 20 min (longer if thickest part of the gel is thicker than 0.6 mm). If fixing is not completed properly, the dried gel will 'melt' the X-ray film. Drain the fixing solution and blot off most of the liquid. Place assembly on bench and place large piece of filter paper on top of gel; press

down hard and pull off gel with filter paper. Place Saranwrap on top of gel, cut off excess material, and dry at 80° C (at least 15 min depending on system).

8. Remove Saranwrap before exposing to film overnight.
9. Develop: 4 min developer, 1 min rinse, 5 fixer, 10 min rinse (25-30° C). Hang to dry.

Reagents

gel box:	apparatus for electrophoresis including all accessories like glass plates, assortment of spacers/combs/clamps, surface thermometers, special pipet tips (0.3 mm tips)
misc. items:	solvent-resistant tray with stoppered drain hole and gel drying set-up and exposure holder, all to fit at least 45 cm long gels, film holder for developing film, Whatman #3 filter paper
Sigmacote:	silanizing solution from Sigma
10% APS:	10% w/v ammonium persulfate in H_2O. Store at 4° C for 1 week.
TEMED:	Use neat; store at 4° C.

monomer solutions for 6% gel:

acrylamide stock:	H_2O 40 ml	
	acrylamide	30 g
	bisacrylamide	1.5 g
	Bring up to 75 ml in glass beaker, add 2.5 g Amberlite AG501-X8(D), shake for 20 min, then filter, and store at 4° C.	
Instagel:	urea	230 g
	10xTBE	50 ml
	of the above acrylamide stock solution	75 ml
	Bring up to 500 ml with H_2O, filter, and store in the dark at 4° C. Keep for up to 4 weeks.	
	For an 8% gel, add 100 ml instead of 75 ml to make Instagel.	
	Warm Instagel to room temperature before adding initiators.	

Use 70 µl TEMED and 700 µl 10% APS per 100 ml degassed monomer solution.

Alternatively, buy the pre-made solutions from Boehringer Mannheim Biochemicals.

fixing solution:		AcOH	10%
		MeOH	10%
X-ray film:		Kodak XAR-5 or better Amersham Hyperfilm-βmax	
photographic items:		developer:	Kodak GBX
		fixer:	Kodak GBX
		bath temperature:	25-30° C
10xTBE:	Tris (pH 8.3)		121.0 g/l
	HBO_3		55.0 g/l
	$Na_2EDTAx2H_2O$		9.3 g/l

Data Analysis

Data analysis can be done with local and remote systems. Local systems - like The DNA Inspector IIe running on a Macintosh - have the advantages that they are faster for simple operations and that the user has more control over them. Remote systems - like the University of Wisconsin Genetics Computing Group package (UWGCG) running on a VAX - have the advantage that they allow complex operations like database searches. Moreover, they allow the user to maintain contact with the outside world by bulletin boards and email. An up-to-date listing of commercially available restriction enzymes is especially valuable in this respect.

Some of the databanks like Genbank and PIR maintain sophisticated retrieval programs; by their charter, however, they are prohibited from offering complete analysis software. For a sequencing project in which an unknown sequence is determined, the statistical

analysis programs Codonpreference and Testcode are especially useful after a consensus sequence has been assembled from gels: using various procedures, a statistical probability for the existence and location of open reading frames is calculated for each of the three frames (Doolittle, 1986).

Rules of Thumb (for either of the two left ones) for Klenow:

- the upper band of a double C is always more intense
- the upper band of a double G is often more intense
- the upper band of a double A is often less intense
- the upper band of a double G is often more intense if the sequence is preceded by a T

Oligonucleotide Synthesis and Purification

Procedure

Synthesis procedures are specific to the synthesizer used. Phosphoamidite chemistry is used on most machines; the only requirement is the use of anhydrous technique. Purification of the product is then reduced to the chromatographic separation of the trityl-on oligo from contaminating, shorter oligonucleotides (McBride et al., 1988).

1. After trityl-on synthesis, cleave from support and deprotect as described in instrumentation manual.
2. Flush OPC (Applied Biosystems, 1989) with 5 ml acetonitrile, then with 5 ml triethylamine acetate (rate of 1-2 drops/sec).
3. Add 1 ml H_2O to crude oligonucleotide (trityl-on) in 2-3 ml conc. ammonia.
4. Load oligonucleotide onto column (2-3 passes). This will load 1-5 OD's. Save eluant.
5. Flush with 3 x 5 ml dilute ammonia.
6. Flush with 2 x 5 ml H_2O.
7. Detritylate support-bound oligonucleotide with 2 x 5 ml 2% TFA[21].

8. Flush with 2 x 5 ml H_2O.
9. Elute purified oligonucleotide with 3 x 1 ml 20% acetonitrile and dry fractions in Speedvac.
10. Resuspend in H_2O to determine yield.

Reagents

Oligonucleotide Purification Cartridge:	OPC, from Applied Biosystems, Inc.
acetonitrile:	HPLC grade
20% acetonitrile:	20% in H_2O
triethylamine acetate:	2 mol/l (titrate in the cold or buy from Applied Biosystems, Inc.)
dilute ammonia:	1:10 dilution of concentrated ammonia in HPLC H_2O

Protein Techniques

Reagent grade chemicals are from Sigma Chemical Company, electrophoresis reagents are from BIORAD Laboratories, ultrafiltration membranes and Western blotting membranes are from Millipore Corporation, and peptide synthesis reagents are from Aldrich Chemical Company; Richelieu Biotechnologies, Inc.; Sigma Chemical Company; Fluka Chemical Corporation; and BACHEM Bioscience, Inc.

Measurement of Protein Concentrations

Lowry Assay

Interference in this procedure from Tris requires a blank (Lowry et al., 1951).

Procedure

1. Make reagent C (alkaline copper) fresh by mixing 50 ml reagent A and 1 ml reagent B.
2. Prepare protein standard solution: Dissolve bovine serum albumin (BSA) in H_2O to a final concentration of 5 mg/ml. Determine actual concentration using $\varepsilon_{279nm} = 4.467 \times 10^4$ l/(molxcm) and $m_w = 66{,}000$ or $\varepsilon_{1\%aqueous} = 6.6$ at 280 nm. Set up standard curve (total protein 0-150 µg) by combining suitable amounts of BSA stock solution and H_2O to give 100 µl final volume. Includea blank.
3. Dilute samples so they contain less than 150 µg protein in 100 µl. Include a suitable blank as many substances interfere with assay (Tris, dithionite, etc.). Vortex to oxidize dithionite, if present.
4. Add 4 ml alkaline copper reagent to each sample.
5. Add 0.4 ml phenol reagent while vortexing.
6. Take reading after 30 min. Color is stable for at least 3 h.

7. Read absorbance at 660 nm. Establish standard curve and read unknown protein quantity from plot.

Reagents

Reagent A:	Na_2CO_3	20.0 g/l
	NaOH	4.0 g/l
	sodium tartrate	0.2 g/l
Reagent B:	$CuSO_4$	5.0 g/l
Phenol reagent:	Folin-Ciocalteau reagent diluted 1:1 with H_2O (1 n)	

Lowry Assay in Presence of Interfering Substances

The presence of DTT requires a modification of the above method (Bensadoun & Weinstein, 1976).

Procedure

1. Prepare samples and standards as in regular assay (1000 µl volumes in microfuge tubes).
2. Add 70 µl deoxycholate, mix, and leave at room temperature for 10 min.
3. Add 250 µl TCA and mix.
4. Centrifuge for 15 min and remove supernatant (re-centrifuge for 5 sec).
5. Transfer with two 1 ml washes of alkaline copper reagent to test tube.
6. Add 2 ml alkaline copper reagent.
7. Add 0.4 ml phenol reagent while vortexing.
8. Read A_{660nm} after 30 min.

Reagents

0.5% (w/v)	sodium deoxycholate (make up fresh if precipitate is observed)
30% (w/v)	trichloroacetic acid

SDS-Polyacrylamide Electrophoresis

Procedure

1. Assemble gel sandwich and check for leaks with H_2O.
2. Prepare running gel mix fresh and degas using H_2O aspirator for 15 min. Stir gently while degassing.
   Add 500 µl APS and 50 µl TEMED per 100 ml monomer solution. Stir and pour gel into frame until it reaches desired level (usually 2-3 cm lower than lower edge of comb).
3. Overlay carefully with 2-3 ml H_2O and allow to polymerize for 30-45 min. Blot off H_2O.
4. Add 120 µl APS and 12 µl TEMED to 20 ml degassed stacking gel and pour into frame until it overflows. Place comb, taking care not to catch bubbles. Allow to polymerize for 45-60 min.
5. Remove comb by pulling it straight up and flush wells with running buffer. Clamp sandwich into frame and fill buffer chambers with running buffer.
6. Add 25% 4x loading buffer to samples and heat at 95° C for 5 min.
   Load samples and electrophorese at constant current.
7. Stain for 0.5 h, destain for 1-3 h, layer gel between thick filter paper and Saranwrap or between layers of cellophane, and dry at 80° C.

Reagents

To separate the *nif*D and *nif*K peptides, add 0.01% n-dodecanol to the running buffer and to the gel mixes.

acrylamide/Bis stock solution (30%T stock solution, 37.5:1 crosslinker)

$$\%T = 100 \times \frac{\text{g acrylamide} + \text{g Bis}}{\text{total volume}}$$

acrylamide	2.05 mol/l	292.0 g/l
N'N'-Bis(methylene)acrylamide	0.052 mol/l	8.0 g/l

Store at 4° C in the dark for 1 month.

4x sample buffer: (Laemmli, U.K., 1970)

Tris (pH 6.8)	62.5 mmol/l	1 ml 0.5 mol/l Tris (pH 6.8)
SDS	2.0%	1.6 ml 10% (w/v) SDS
glycerol	1.1 mmol/l	0.8 ml
β-mercaptoethanol	0.3 mmol/l	0.4 ml
bromophenol blue	0.0013% (w/v)	0.2 ml 0.05% stock
H_2O		4 ml

5x running buffer:

Tris (pH 8.3)	15 g/l	24.8 mmol/l
glycine	72 g/l	19.2 mmol/l
SDS	5 g/l	3.5 mmol/l

Dilute 300 ml into 1.5 l before use. Store at 4° C.

running gel mix: (12%) for proteins with m_w from 10,000-100,000

Tris (pH 8.8)	375.0 mmol/l	25 ml 1.5 mol/l Tris (pH 8.8)
acrylamide/bis	12.0%	40 ml 30% stock solution
SDS	0.1% (w/v)	1 ml of 10% (w/v) stock
H_2O		33.5 ml

sufficient for two gels (20 cm x 20 cm x 1.5 mm)

stacking gel mix: (4%)

Tris (pH 6.8)	125.0 mmol/l	5 ml 0.5 mol/l Tris (pH 6.8)
SDS	0.1% (w/v)	0.2 ml 10% (w/v) SDS
acrylamide/bis	4.0%	2.6 ml stock solution
H_2O		12.2 ml

sufficient for two stacking gels (1.5 mm thick)

stain: 0.1% (w/v) Coomassie Blue R-250 in 40% MeOH, 10% AcOH
destain: 40% MeOH, 10% AcOH

running conditions: At these settings, run time is 5-6 h.

gel thickness	stacking gel (mA/gel)	running gel (mA/gel)
0.5	8	12
0.75	13	18
1.0	16	24
1.5	25	35
3.0	50	70

Ultrafiltration

1. Select pore size:
   - at least twice the nominal molecular weight to ensure passage of the protein of interest
   - at the most half the nominal molecular weight to ensure retention of the protein of interest
2. Pre-equilibrate the membrane by floating it on H_2O according to manufacturer's instructions (for Amicon membranes float it shiny side down).
3. Install membrane in ultrafiltration cell (for Amicon membranes dull side down).
4. Test for leaks at 65 psi.
5. Concentrate sample with stirring.

Note: The faster the flow, the stronger the polarization of the membrane.

Anaerobic Native Gel Electrophoresis

All procedures, as far as possible, are done outside the box. Pouring, loading, and running the gel are performed inside the box.

Procedure

1. Assemble gel sandwich and test for leaks with H_2O. Degas electrode buffer and add 0.75 ml 1 mol/l dithionite for 1.5 l.
2. Mix 12 ml acrylamide stock and 33 ml stock 1 and degas. Add 7.5 µl TEMED (degassed), 45 µl 10% APS (degassed) and 4.5 µl 1 mol/l dithionite. Mix and pour running gel immediately: overlay is not necessary but helpful in getting a straight upper edge to the running gel. Allow to polymerize for 30 min.
3. Mix 3.3 ml acrylamide stock and 20 ml stock 2 and degas. Add 7.5 µl TEMED (degassed), 38 µl 10% APS (degassed) and 1.5 µl 1 mol/l dithionite. Mix and pour stacking gel immediately; insert comb. Allow to polymerize for 45-60 min.
4. Degas micro-assay vials containing 30 mg sucrose and dye (guaiazulene 3-sulfate; McKenna et al., 1977). Add 200 µl anaerobic sample. Leave on ice until ready to load. Make up standards.
6. Load samples and run gel overnight at constant current (e.g. 12 mA for 1.5 mm gel/13 h) or until dye runs off.

Reagents

acrylamide/bis stock is same as for SDS-PAGE

electrode buffer:	Tris (pH 8.5)	32.5 mmol/l	5.91 g for 1.5 l
	glycine	191.5 mmol/l	21.56 g for 1.5 l
	glycerol	10% (v/v)	150 ml for 1.5 l
	H_2O to 1.5 l		
stock 1:	glycerol		67.5 ml
	1.5 mol/l Tris (pH 8.8)		168.8 ml
	H_2O		93.8 ml
stock 2:	glycerol		33.8 ml
	0.5 mol/l Tris (pH 6.8)		49.5 ml
	H_2O		116.25 ml

Western Blotting

Procedure

1. Electrophorese protein sample to completion.
2. Equilibrate the polyacrylamide gel, two sheets of filter paper, and one piece of membrane per gel in blotting buffer for at least 15 min before assembling blotting sandwich. Immobilon membrane must be pre-wet in MeOH for 2-5 s before it can be wetted by aqueous buffer. Nitrocellulose can be pre-wet in H_2O.
3. Assemble the gel sandwich as shown; blotting is from the cathode to the anode. Contact, especially between the gel and the membrane, has to be tight; bubbles must be excluded anywhere between the two sheets of filterpaper.

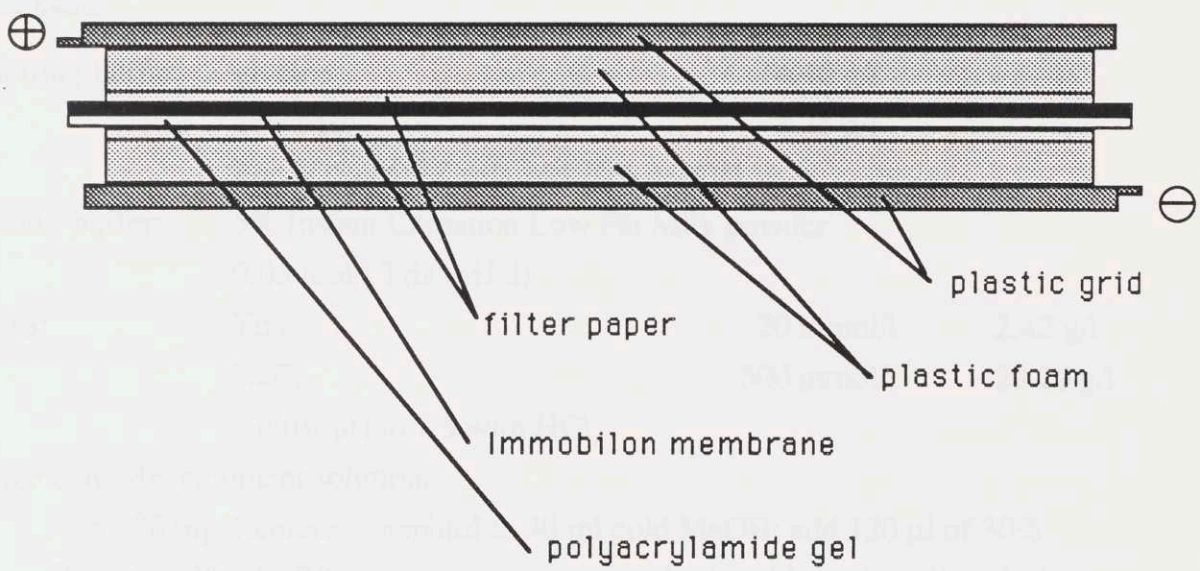


Figure A.2.: Blotting Sandwich

4. Insert cassette into holder, fill to the line with blotting buffer, and blot for 5 h at 200 mA (for *nif* proteins anywhere between 2 h at 400 mA and overnight at 200 mA seems to be fine). Blotting should be at 4° C.
5. At the end of the run disassemble cassette and shake membrane slowly in 200 ml Blotto buffer for 45 min - 1 h. Add the appropriate dilution of antiserum (1:2000 - 1:4000 of antiserum) in 200 ml Blotto buffer and incubate 2 h -

overnight with gentle shaking. Typically, a 1:3000 dilution of 4-534 is used to detect *nif*D, *nif*K, and *nif*H peptide in 10 µg of cell extracts.

anti-Kp1: 4-534 and 4-535 (written around the perimeter of the tube)

anti-Kp2: 4-536 and 4-537 (written along the long side of the tube)

6. Rinse thoroughly with 2x200 ml Blotto buffer and add secondary antibody (goat anti-rabbit antibody, BIORAD #170-6516, 33 µl per 100 ml Blotto buffer). Shake gently for 1 h.
7. Rinse membrane with 2x200ml TBS. Add color development solution and allow color to develop at room temperature for up to 15 min. Stop color development by rinsing generously with H_2O. Let dry and keep in the dark.

Reagents

blotting buffer:	glycine	14.40 g/l	
	Tris base	3.25 g/l	
	For SDS-PAGE gels add 20% MeOH.		
Blotto buffer:	5% Instant Carnation Low Fat Milk powder		
	0.05 mol/l Tris (pH 8)		
TBS:	Tris	20 mmol/l	2.42 g/l
	NaCl	500 mmol/l	29.20 g/l
	Adjust pH to 7.5 with HCl.		

HRP color development solution:

Add 120 mg 4-chloro-1-naphtol to 40 ml cold MeOH; add 120 µl of 30% H_2O_2 to 200 ml TBS at room temperature and mix with methanolic solution; use immediately.

membranes: Immobilon PVDF from Millipore, Inc.
nitrocellulose membrane (0.2 µm pore size) from S&S, Inc.

N-terminal Peptide Sequencing

N-terminal protein sequencing can be performed on purified proteins or even on proteins present in crude cell extracts, if they are present in sufficient quantities. Rule of thumb: if you can see it well by Coomassie staining, you can sequence it.
Typically, a protein sample is resolved by SDS-PAGE, then blotted onto Immobilon membrane and subsequently sent into a service facility for the actual gas phase sequencing and HPLC (Matsudaira, 1987).
The sensitivity of this method is appr. 10 pmol; therefore, reagents of high purity must be used. Also, the less reagent is used, the less contamination is introduced; therefore, only the smallest amounts of reagents possible should be used.

Procedure

1. Run SDS-PAGE gel to give maximum resolution around the protein of interest.
2. Blot the proteins onto Immobilon membrane (15 V for 1-2 h).
3. Stain membrane briefly with Coomassie.
4. Destain as long as possible (of course, the protein of interest must still be visible!).
5. Wash exhaustively with ultrapure H_2O (shaker, 15 min).
6. Cut out protein of interest and dry in Speedvac. Slices of membrane may be pooled; however, they all have to fit into a circle of 1 cm diameter (cartridge size of sequencer).
7. Store sample in freezer.

Reagents

blotting buffer:	CAPS (pH 11)	10 mmol/l
	MeOH	10%

Solid-Phase Peptide Synthesis

Using the method described, peptides up to ten amino acids can be easily if laboriously synthesized. This method uses the FMOC group (Carpino & Han, 1972) to protect the alpha-amino group on the nascent chain and HOBt (1-hydroxy-benzotriazole), BOP (Castro reagent, Hudson et al., 1988), and NMM (N-methylmorpholine) to activate the amino acid to be added to the chain (Stewart & Young, 1986). Because the N-terminal protecting group FMOC is base-labile, the protecting groups chosen for the tertiary functions of the amino acid precursors must be acid-labile (so is the linkage to the polystyrene resin - Wang resin). The peptide is synthesized by deprotecting the last amino acid with piperidine, adding activated amino acid (Bolin et al., 1989), capping non-reacted product with acetic anhydride, and so on and so forth (one cycle takes 3-4 h). Deprotection, coupling, and capping should be immediately followed by the Kaiser test (ninhydrin-based colorimetric assay) to assess the extent of the reaction.

- Preactivate coupler for 10 min before adding to resin/DMF[22].
- Store overnight in 10% MeOH, 90% DCM[23]; for long term storage filter and dry out of same solvent. Keep dry resin at -20° C.
- Deprotection: Shake 2x30 min with 14 ml/g resin of 2% H_2O, 2% p-cresol, 2% thioanisole, 94% TFA (add 3% DTT for met, 1.5% N-methyl indole for trp). Elute cleaved product from resin after each step, then rinse with 10 ml TFA and pool fractions. Leave at room temperature for 2 h (3 h at 50° C for Mtr!). Evaporate to near dryness, wash 5x with ether, 1x with ethylacetate:ether (2:1), collecting solid after each step by centrifugation in polypropylene tube. Rotorvap all solvent, resuspend in appr. 15 ml H_2O, rotorvap briefly, and lyophilize.
- HPLC: C18 reverse phase, 214 nm detector. Solvent A: 0.1% TFA in H_2O, Solvent B: acetonitrile. Run 10% B to 50% B in 30-50 min at 4 ml/min on 3.9x30 cm column (Rivier et al., 1984).

Step	Reagent	volume/ml	time/min
1	DCM wash (3x) [omit after first cycle]	15 ml/g resin	3
2	DMF wash (2x) [3x for n≥2]	15 ml/g resin	3
3	deprotect: DMF:piperidine @ 1:1 (1x)	15 ml/g resin	20
4	DMF wash (2x)	15 ml/g resin	3
5	wash with dioxane:H_2O @ 2:1 (1x)	25 ml/g resin	10
6	DMF wash (6x)	15 ml/g resin	5
7	couple:add to DMF 3 equ. of coupler (FMOCAA:BOP:NMM:HOBt @1:1:1.5:1 in DMF [0.25M NMM in DMF]; preactivate for 10 min) [recouple if necessary]	15 ml/g resin	120
8	DMF wash (5x)	15 ml/g resin	8
9	iPrOH wash (2x)	25 ml/g resin	8
10	DCM wash (5x) [good step for storage]	15 ml/g resin	8

Table A.1.: General Scheme for Peptide Synthesis

Reagents

wrist action shaker:
- at least 45° action radius

dark room timer:
- timer with provisions to turn off power to shaker when preset time for shaking has elapsed

dichloromethane:	ACS grade directly from bottle
dimethylformamide:	HPLC grade left over act. molecular sieves/4Å for 1 week
iso-propanol:	ACS grade directly from bottle
dioxane:	ACS grade from still
TFA:	protein synthesis grade directly from bottle
diethylether:	ACS grade from still
ethylacetate:	ACS grade directly from bottle
thioanisole:	ACS grade directly from bottle
p-cresol:	ACS grade directly from bottle
H_2O:	SuperQ Millipore H_2O
piperidine:	highest grade available from Aldrich

solid phase reactor:
- custom-made from glass
- solvent-resistant screw top, teflon stopcocks side arm for gas pressure
- 20-30 ml volume for each g of resin used
- 10 µm glass frit (test after annealing!)

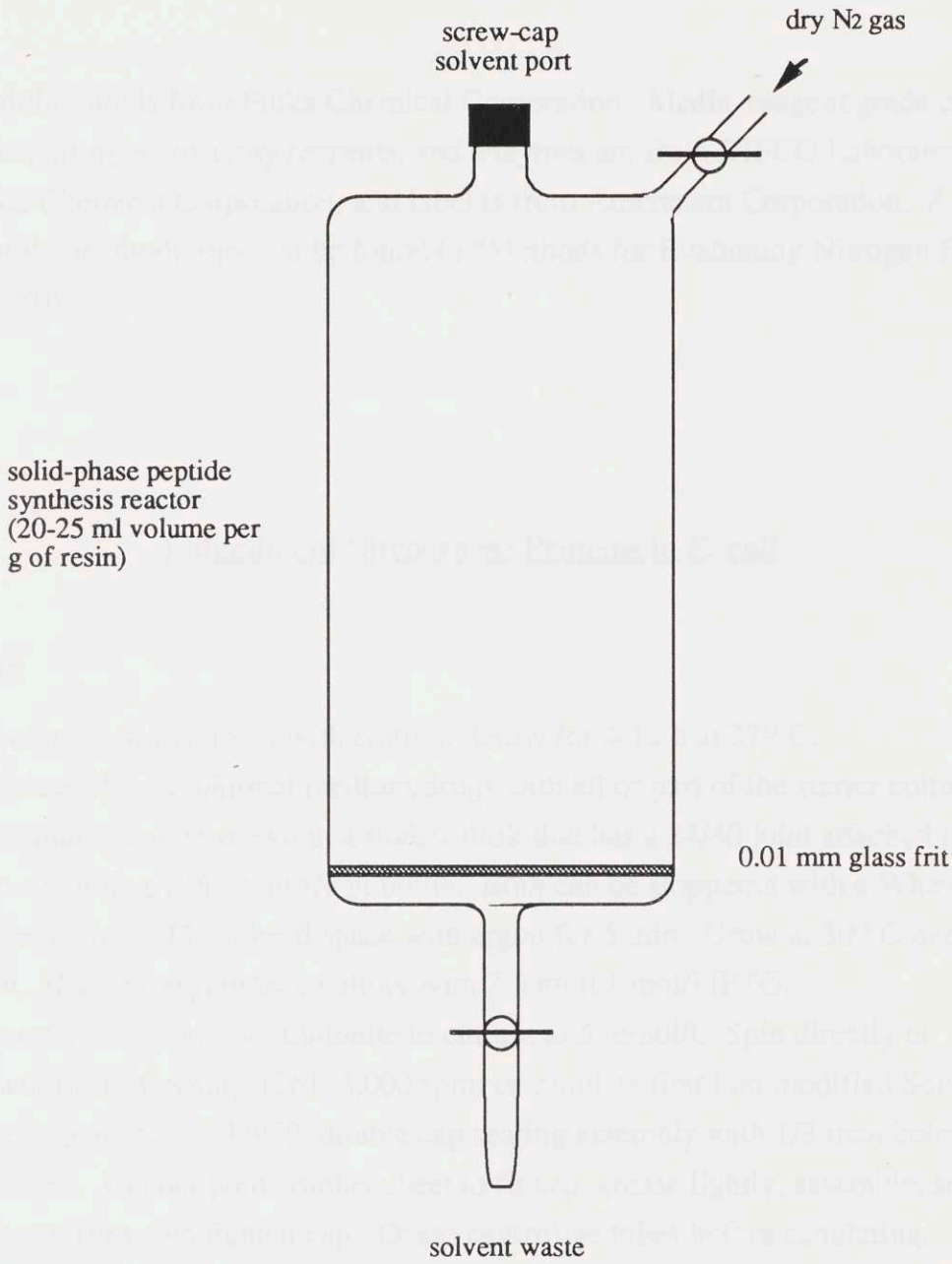


Figure A.3.: Solid Phase Peptide Synthesis Reactor

Nitrogenase Techniques

Sodium dithionite is from Fluka Chemical Corporation. Media, reagent grade chemicals, antibiotics, nitrogenase assay reagents, and enzymes are from DIFCO Laboratories, Inc. and Sigma Chemical Corporation, and label is from Amersham Corporation. A general review of the methodology can be found in "Methods for Evaluating Nitrogen Fixation" (Eady, 1980).

Induction of Nitrogenase Proteins in *E. coli*

Procedure

1. Inoculate 5 ml LB/drugs with culture. Grow for 4-12 h at 37° C.
2. Inoculate 250 ml minimal medium/drugs with all or part of the starter culture. This culture can be grown in a shaker flask that has a 24/40 joint attached to it or in a Corning 1261 centrifuge bottle. Both can be stoppered with a Wheaton rubber septum. Flush head space with argon for 5 min. Grow at 30° C overnight. If necessary, induce culture with 2.5 ml 0.1 mol/l IPTG.
3. Harvest cells by adding dithionite to culture to 5 mmol/l. Spin directly in growth flask (Corning 1261, 3,000 rpm) or canulate first into modified Sorvall centrifuge bottles (#03939, double cap sealing assembly with 1/3 inch holes cut in center). Cut neoprene rubber sheet to fit cap, grease lightly, assemble, secure with parafilm, and tighten cap. Degas centrifuge tubes before canulating.
4. Canulate off supernatant and resuspend in cracking buffer (1% volume of culture).
5. Crack anaerobically in French Press at 20,000 psi using rapid fill kit. Rinse before and in between samples with anaerobic buffer. Collect samples in degassed serum vials and freeze in liquid nitrogen.

Reagents

media: (modified M9 medium, Neidhardt et al., 1974)

50.0 mmol/l Na_2HPO_4	6.0 g/l
22.0 mmol/l KH_2PO_4	3.0 g/l
8.6 mmol/l NaCl	0.5 g/l
18.7 mmol/l NH_4Cl	1.0 g/l
[0.1% Casamino acids	1.0 g/l]

Adjust pH to 7.4, autoclave, and add to 1 l:

2.0 mmol/l $MgSO_4$	2.0 ml of 1 mol/l stock
0.2% glucose	10.0 ml of 20% stock
0.1 mmol/l $CaCl_2$	0.1 ml of 1 mol/l stock
3.0 mg/l iron citrate	0.1 ml of 30 mg/ml stock

Add 1 ml of micronutrients to get final concentrations of:

3 x 10-6 mmol/l $(NH_4)_6(MoO_7)_{24}$

4 x 10-4 mmol/l H_3BO_3

3 x 10-5 mmol/l $CoCl_2$

1 x 10-5 mmol/l $CuSO_4$

8 x 10-5 mmol/l $MnCl_2$

1 x 10-5 mmol/l $ZnSO_4$

Add antibiotics as necessary.

Modified MM9: Make as M9, but use 50% more phosphate salt.

cracking and rinse buffer:

Tris or HEPES (pH 7.4)	100 mmol/l
dithionite	5 mmol/l

anaerobic centrifuge bottles: Assemble as described below and cycle until anaerobic; monitor mercury manometer while isolating gas train from vacuum source to check for leaks.

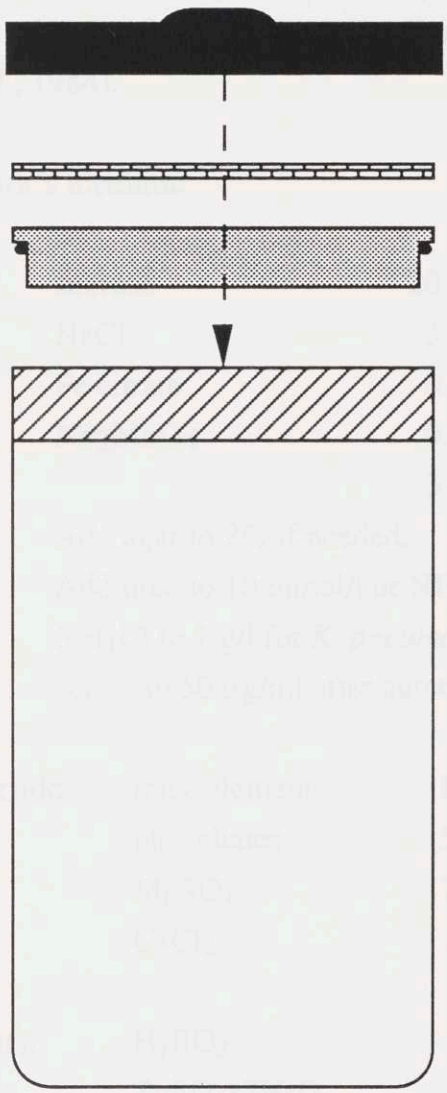


Figure A.4.: Anaerobic Centrifuge Bottle

Growth of *Azotobacter vinelandii* and *Klebsiella pneumoniae*

Azotobacter vinelandii is grown aerobically in the absence of fixed nitrogen and in the presence of molybdenum to derepress its molybdenum-*nif* system. Growth in modified Burk's medium yields the standard morphology: two pear-shaped cells touch each other at the pointed ends (like an 'eight'). Growth for *K. pneumoniae* (Mahl et al., 1965) is anaerobic or microaerobic in the same medium (Hill, 1976 and Hill et al., 1984) at 30° C (Brooks et al., 1984).

Modified Burk's medium:

For 1 l mix:	sucrose	20 g solid sucrose
	NaCl	3.4 ml of 1 mol/l stock solution
	Fe citrate	0.4 ml of 30 mg/ml stock solution
	Na_2MoO_4	0.1 ml of $Na_2MoO_4x2H_2O$ @ 25 g/l in 3 mmol/l NaOH

Add agar to 2% if needed.

Add urea to 10 mmol/l or NH_4AcO to 25 mmol/l for *A. vinelandii* and NH_4Cl to 1 g/l for *K. pneumoniae*, if nitrogen source is desired. Add L-serine to 50 µg/ml after autoclaving if derepressing *K. pneumoniae*.

Autoclave & add:	trace elements	1 ml
	phosphates	5 ml of Burk's salts
	$MgSO_4$	2 ml of $MgSO_4x7H_2O$ at 100 g/l
	$CaCl_2$	1 ml of $CaCl_2x2H_2O$ at 44 g/l

trace elements:	H_3BO_3	0.88 g/l
	$ZnSO_4x7H_2O$	0.44 g/l
	$CoCl_2x6H_2O$	0.20 g/l
	$MnSO_4xH_2O$	17.0 mg/l
	$CuSO_4x5H_2O$	10.0 mg/l

Burk's salts (200x):	K_2HPO_4	160.2 g/l
	KH_2PO_4	40.8 g/l
	pH should be 7.1	

In vivo Assay of *K. pneumoniae* and *A. vinelandii*

Procedure

1. Cap 10 ml serum vial with stopper and transfer to gas train. Cycle between vacuum and argon three times.
2. Evacuate vial and backfill with 10% acetylene in argon.
3. Transfer to 30° C waterbath and relieve overpressure.
4. Measure A_{660nm} of culture. Inject 1 ml of culture using anaerobic technique (syringe flushed with 2 mmol/l dithionite in Tris buffer).
5. Incubate at 30° C for 20 min with gentle shaking.
6. Quench with 0.1 ml 30% TCA.
7. Inject 100 µl of headspace onto GC and compare to standard.
8. Calculate specific activity as:

$$\text{specific activity} = \frac{\text{nmol } C_2H_4 \text{ produced}}{\text{min x } A_{660nm}}$$

For *Azotobacter vinelandii* the following modification is necessary:

Before injecting the aerobically growing culture, inject air to 6,8,10,12 % final concentration. Vary concentrations to match growth of culture.

In vitro Nitrogenase Assay

Procedure

1. In a "12.56" ml serum vial combine:
   H_2O to 1 ml
   250 µl RM
   50 µl CK
   Cycle between vacuum and argon three times. Evacuate vial and backfill with 10% acetylene in argon. Store on ice before use.
2. Add 25 µl dithionite and incubate with shaking at 30° C for 5 min. Relieve

overpressure.

3. Add sample to be assayed and purified proteins (if necessary). Total volume should be 1 ml. Stop reaction by addition of 0.1 ml 30% TCA.
   Typically 0.1 ml of cell extracts (3-10 mg/ml protein) are assayed ± purified Kp1, Kp2 at 30° C for 20 min.
4. Inject 0.1 ml of head space into GC. Compare to standard.

Calculations

number of nmol C_2H_4 injected in injection of standard:

$$\frac{\text{nmol } C_2H_4}{\text{injection of standard}} = \frac{0.05}{532} \times \frac{0.1}{22.4 \times 10^3} \times 10^9 = 0.4196$$

a correction for temperature and pressure yields:

$$\frac{\text{nmol } C_2H_4 \text{ (corr)}}{\text{injection of standard}} = \frac{\text{nmol } C_2H_4 \text{ (uncorr)}}{\text{injection of standard}} \times \frac{273.15}{\text{temperature/K}} \times \frac{\text{pressure/torr}}{760}$$

for a sample injection this means:

$$\text{nmol } C_2H_4 \text{ injected} = \frac{\text{size of sample peak}}{\text{size of standard peak}} \times \text{nmol } C_2H_4 \text{ in standard injection}$$

and

$$\text{nmol } C_2H_4 \text{ per assay} = \text{nmol } C_2H_4 \text{ injected} \times \frac{12.56}{0.1}$$

Assay for Component 1 and Component 2

A given amount of cell extract to be assayed for activity of component A is titrated with increasing amounts of purified component B until no increase in activity is observed upon further addition of component B. Under these conditions, the amount of A present is limiting and a specific activity can be calculated for component A.

Typically, 20 µl of pGH1/pVL15 extract with an estimated specific activity of 100 nmol/(minxmg) for Kp2 will be saturated by addition of 8 µl of purified Kp1 at 30 mg/ml and 985 nmol/(minxmg) specific activity.

Assay for apoMoFe Protein

apoMoFe protein assays are essentially the same as for Kp1 with one exception: extracts are incubated with isolated FeMoCo before the actual assay to reconstitute holoprotein.

In degassed assay vials mix: 100 µl extract
300 µl anaerobic buffer
increasing amounts of cofactor (1,2,4,6,8 µl)

Incubate on ice for 15 min before injecting 100 µl into an assay vial with RM, CK, dithionite, H_2O, argon/acetylene, Kp2.

For each new batch of cofactor, determine how many microliters are needed to saturate an extract assumed to be very active (use a constant amount of Kp2). Then determine how much Kp2 is needed for maximum activity. Typically, 100 µl of extract with an approximate specific activity of 35 nmol/(minxmg) for apoprotein is saturated by 18 µl of cofactor in DMF (110 nmol/(minxnmol molybdenum), [Mo] = 47 µmol/l). To saturate 100 µl of activated extract, 60 µl of Kp2 is needed at 13.3 mg/ml and 1202 nmol/(minxmg) specific activity.

Reagents

dithionite: Degas 0.2 g $Na_2S_2O_4$ per ml 1 mol/l dithionite desired. Degas 0.25 mol/l KOH separately.
Add KOH to degassed dithionite to make 1 mol/l dithionite.

serum vial: Use 10 ml serum vial; cap and fill with H_2O to determine exact head space gravimetrically. Wheaton vials measure 12.56 ml head space.

standard: Degas 25 ml serum vial as far as possible (e.g., 10 mtorr).
Fill with pure ethylene from lecture gas bottle set to 3-5 psi.
Vent to ambient pressure with needle topped up with H_2O.
Degas 1 l round-bottom flask (actually 532 ml) and fill with argon. Inject 0.05 ml pure ethylene. Mix and inject 0.1 ml onto GC (Poropak N column, nitrogen carrier gas, FID).

creatine phosphokinase: 20 mg in 10 ml 25 mmol/l HEPES (pH 7.5)

80 mmol/l ATP: Dissolve 4.8 g in 100 ml H_2O, adjust pH to 7.0,
and check concentration using $\varepsilon_{259nm} = 1.54 \times 10^4$ l/(molxcm).

RM:

4x concentration	final concentration	add per 50 ml of 4x
42.0 mmol/l	10.5 mmol/l	687 mg creatine phosphate
100.0 mmol/l	25.0 mmol/l	25 ml 200 mmol/l HEPES (pH 7.5)
25.0 mmol/l	6.25 mmol/l	12.5 ml 0.1 mol/l $MgCl_2$
20.0 mmol/l	5.0 mmol/l	12.5 ml 80 mmol/l ATP

Store in small aliquots at -20° C.

Pulse Labeling

Procedure

1. Inoculate 3 ml LB with cultures to be labelled and grow overnight at 37° C.
2. In the morning, inoculate 4 ml minimal medium with 4 µl of overnight culture and grow at 37° C for 2 h or until sparsely grown.
3. Split each culture into 2 x 2 ml. Add 20 µl 0.1 mol/l IPTG to one tube. Grow for 1 h at 30° C (for NifA-induced systems).
4. Dilute ^{35}S-methionine/^{35}S-cysteine to 5-10 µCi/10 µl.
5. Add 10 µl of label to each culture tube. Incubate at 30° C for 10 min. Chase cold with 10 µl of chase mix.
6. Quench reaction by putting tubes on ice. Collect cells by spinning for 3 min in a microfuge. Resuspend cells in 0.1 ml Laemmli sample buffer.
7. Analyze 5-10 µl by SDS-PAGE.
8. Stain with Coomassie Blue. Destain, dry, and expose to X-ray film overnight. Develop.

Reagents

chase mix: Add each amino acid to final concentration of 25 mg/l.
film: Kodak X-omat AR film or SP5

Manual for the Faircrest Glove Box

Introduction and General Operating Precautions

The Faircrest glove box is designed to allow manipulations in an inert atmosphere containing between one and two ppm oxygen. These values can be obtained only if extreme care is taken in the daily operation and maintenance of the glove box. Because the manufacturer is now out of business, spare parts must be either machined or special ordered; breakage of any part will lead to considerable downtime.
Discipline in operation and maintenance of these glove boxes is essential: 'minor' mistakes in operating valves or handling gloves will lead to loss of anaerobicity or (much) worse.
Pressure in the inert gas feed line should be limited to 5 psi, at the most to 10 psi during gas intensive operations. Should the pressure in the box rise out of control, immediate shut-off of the feed gas valve (green) at the main tank is necessary: this is easier than checking the dozen or so valves that might be the problem in the few seconds that remain before the gloves blow out. By the same token, it is advisable to leave at least one glove unplugged if a problem with the pressure regulation system is suspected (this will avoid blow-out of the front window).
On the low pressure side, every time one of the ante-chambers is evacuated, the pressure in the box must be monitored to ascertain that a leak on the box side of the ante-chamber does not lead to evacuation and destruction of the box.

Operating Procedures

The Faircrest glove box maintains an inert atmosphere by combining minimization of direct air leakage with catalytic removal of oxygen entering through diffusion through gloves, windows, etc.; needed materials are brought into the box through an ante-chamber that is made anaerobic through repeated evacuation and back-filling with helium gas. The system consists of four components: main chamber, ante-chamber, recirculation

system, and purification system.

Two ante-chambers, one large and one small, work in conjunction with a two-stage, direct-drive vacuum pump. They can be used one at a time or in parallel: for each, at least three cycles - each to equilibrium at the pressure gauge - are required. Apparatus cycled into the box must be able to withstand the pressure changes. Paper, cardboard, and other items with large surfaces and oxygen capacities must be evacuated overnight. If an ante-chamber is loaded, this should be indicated on the outside to prevent inadvertant opening of the chamber.

Space in a glove box is at a premium; this applies all the more to storage areas inside a glove box, e.g., a fridge. Samples must be labelled clearly to prevent accumulation of junk (and the accidental disposal of important samples by other users). Waste should be removed immediately, containers with liquids should be capped as much as possible to prevent contamination of the atmosphere and subsequent impairment of the catalyst, and apparatus should be returned to their designated places.

The t-butyl rubber gloves are the weak point in this system; diffusion through them is the major source of oxygen during box use. Users should not wear watches or jewelry (no exceptions!). Use of cotton gloves and talcum will facilitate entering the gloves and minimize stress on the material. When a port is not in use, it should be plugged: install the plug with the metal pin at 12 o'clock, pull until the o-ring around the plug seals, then turn the plug clockwise by pushing on the pin with your thumb until the pin points to 1 o'clock. Tighten the plug by tightening the large knob. Before opening the port, cycle the glove 3-6 times depending on when the port was used last. Do not just let the plug fall after turning the metal pin to 12 o'clock, but store it with the knob pointing into the box by jamming it between front window and the aluminum pipe just in front of it.

The pressure inside the main chamber is maintained by the lute (a mineral oil bubbler) and a top-up valve set to maintain a pressure of about 2 inches H_2O on its low pressure side, i.e., inside the box. A foot-operated valve can be used to increase the pressure manually.

The nitrogen hook-up for the ultrafiltration apparatus is set to 66 psi; in order to prevent accidents, always turn off the nitrogen gas at the valve just outside the box when done. An emergency shut-off switch can be used to shut down the system electrically.

The purification system consists of two columns which can be used one at a time, together sequentially, or together in parallel. Removal of oxygen is by means of a copper-based catalyst working at elevated temperatures, and regeneration is achieved by flowing forming gas (95% N_2, 5% H_2) over the column at a higher temperature. It

is convenient to use one column at a time while keeping the second column as a back-up. Turnover of the box atmosphere is by means of a set of two maintenance-free diaphragm pumps. Gas is transported from the box to the oxygen analyzer by a standard aquarium pump hooked up to copper tubing. For detailed procedures see below.

Maintenance

0. Write log book.
1. Change oil of the pump evacuating the ante-chamber every month.
2. Top up the lute with oil to the 2 inch mark when needed.
3. Check for leaks: gloves, pipes, ante-chamber, etc.
4. Regenerate spent catalyst immediately after switching to new column.
   Regeneration becomes necessary when the oxygen value of the box in its resting state goes above 1 ppm.
   a. Set both switches in the main switch box to "OFF" (the diaphragm pumps will go off).
   b. Consult the diagram and the table below on valve positions to set the system to the new column that will be used in purification.
   c. Turn the two switches so that the switch for the spent column points to "RE-FORM" and the switch for the new column points to "RUN" (the diaphragm pumps will go on).
   d. Open the drain valve for the column to be regenerated.
   e. Open the forming gas valve on the column to be regenerated until a slow but steady flow of bubbles is observed at the bubbler attached to the drain valve.
   f. Regenerate for 12-20 h, longer if the catalyst is new (the column being regenerated should become significantly hotter than the column used in purification).
   g. Close the forming gas valve. Open the helium valve until a fast flow is observed and turn switch from "REFORM" to "PUMP" (the diaphragm pumps will go off).
   h. Cool and dry catalyst under helium for 1/2 h (after 10-15 min the columns should have cooled enough for the pumps to come on again).
   i. Close drain valve and then quickly close helium valve so that only a slight overpressure remains in the column.

The switch box with one knob is now obsolete with the use of the new oxygen analyzer: the knob should be left at "BEFORE" and all fittings in the back should be closed tightly.

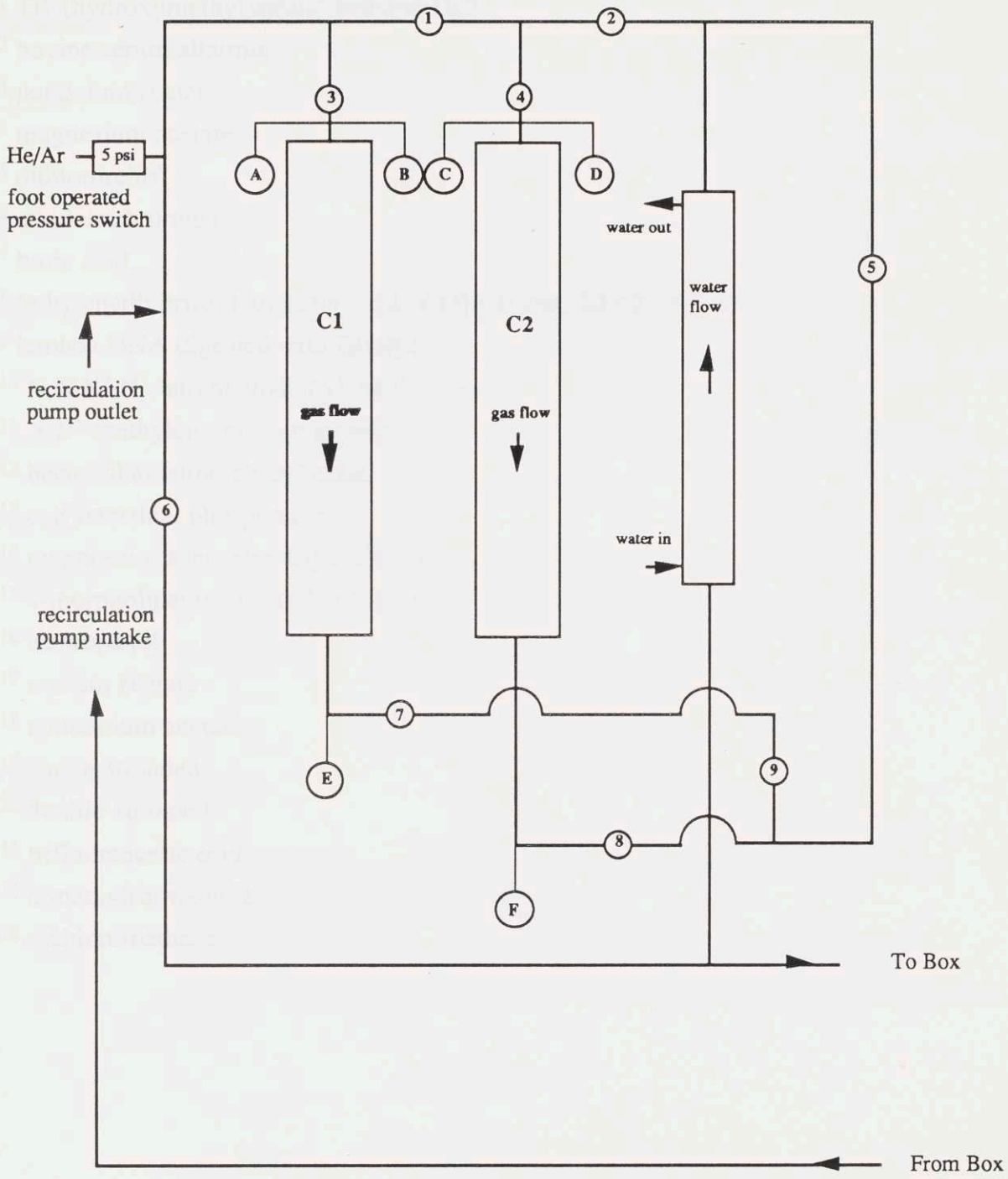
Valve Positions

	1	2	3	4	5	6	7	8	9
C1 & C2	O	O							
C1 only			O		O		O		O
C2 only	O			O	O			O	
C1 & C2 in parallel	O		O	O	O		O	O	O
C1 & C2 in series		O	O	O			O	O	O

O = open valve

Table A.2.: Valve Positions on Glove Box

Recirculation System



#'s 1-9 are 1/2 inch valves, A-F are 1/4 inch valves. A and C are helium valves, B and D are forming gas valves, E and F are H_2O drain valves.

Figure A.5.: Recirculation Scheme for Glove Box

Footnotes to Appendix A

[1] Tris(hydroxymethyl)aminomethane-HCl

[2] bovine serum albumin

[3] potassium acetate

[3] magnesium acetate

[5] dithiothreitol

[6] ethidium bromide

[7] boric acid

[8] (ethylenedinitrilo) tetraacetic acid disodium salt, dihydrate

[9] lambda DNA digested with HindIII

[10] N,N,N',N'-tetramethylethylenediamine

[11] N,N'-methylene-bis-acrylamide

[12] bacterial alkaline phosphatase

[13] calf intestinal phosphatase

[14] magnesium adenosine triphosphate

[15] 4-morpholinepropanesulfonic acid

[16] isopropanol

[17] sodium acetate

[18] ammonium acetate

[19] single-stranded

[20] double-stranded

[21] trifluoroacetic acid

[22] dimethylformamide

[23] dichloromethane

Appendix B: Bacterial Strains

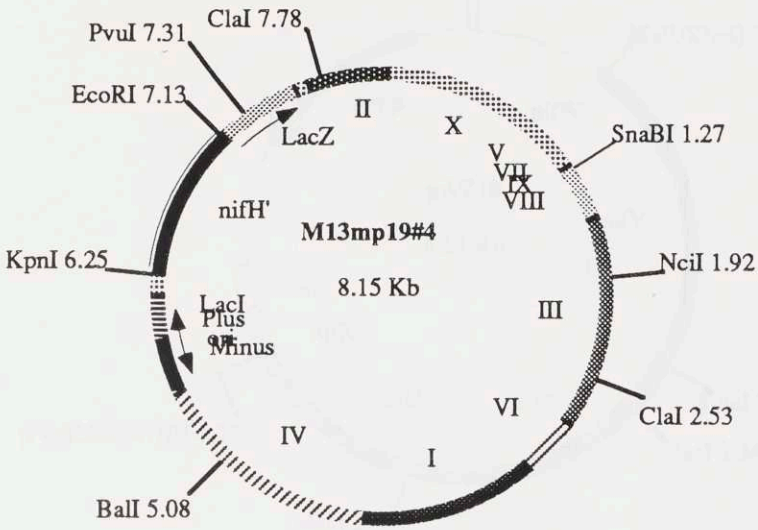
Escherichia coli K12

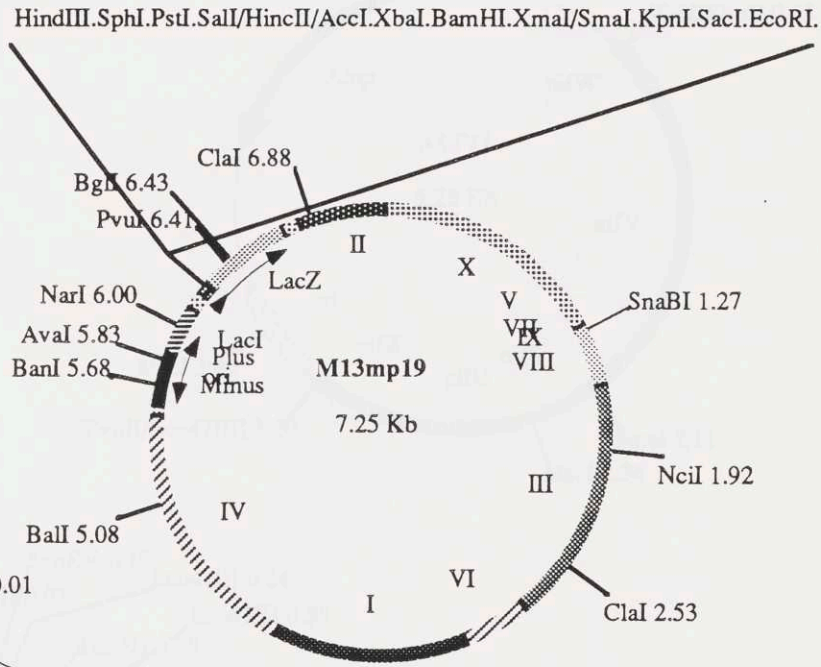
DH5αF'	F', φ80d*lac*ZΔM15, Δ(*lac*ZYA-*arg*F) U169, *rec*A1, *end*A1, *hsd*R17 ($r_K^-m_K^+$), *sup*E44, λ^-, *thi*-1, *gyr*A, *rel*A1 [obtained from BRL Life Technologies, Inc.]
JM101	*sup*E, *his*, Δ(*lac-pro*AB), ($r_M^+r_K^+$) *mcr*A$^+$, {F', *tra*D36, *pro*AB, *lac*I^{q}ZΔM15} [obtained from BRL Life Technologies, Inc.]
JM105	*thi*, *rps*L, *end*A, *sbc*B15, *hsd*R4, Δ(*lac-pro*AB) {F' *tra*D36, *pro*AB, *lac*I^{q}ΔM15}[obtained from Pharmacia LKB Biotechnology, Inc.] N.B.: Sm resistant
TG1	Δ(*lac-pro*), *sup*E, *his*, *hsd*D5, {F', traD36, *pro*AB, *lac*I^q, *lac*ZΔM15} [obtained from Amersham Corp.]
W3110	*lac*I^q, *lac*L8, F'$^-$
XL1-Blue	*rec*A1, *end*A1, *gyr*A96, *his*, *hsd*R17 ($r_K^-m_K^+$), *sup*E44, *rel*A1, λ^-, *lac*, {F', *pro*AB, *lac*I^{q}ZΔM15, Tn10 (tet)} [obtained from Stratagene Cloning Systems, Inc.] N.B.: Tet resistant

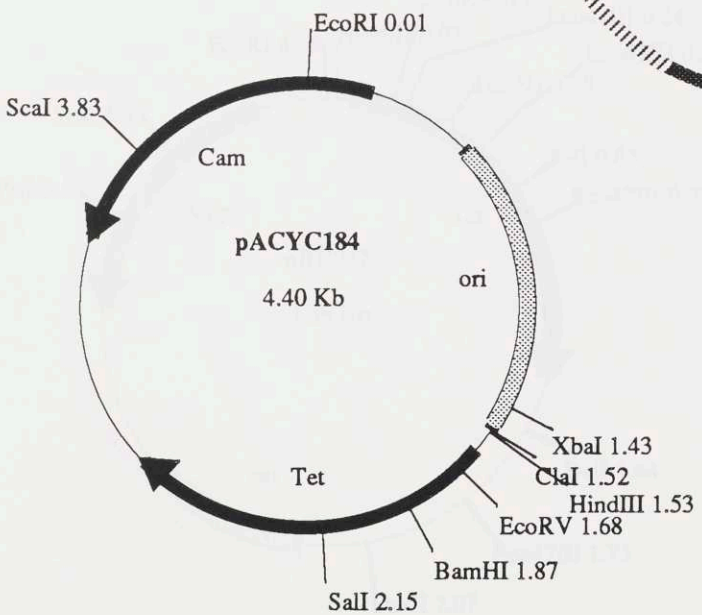
Klebsiella pneumoniae

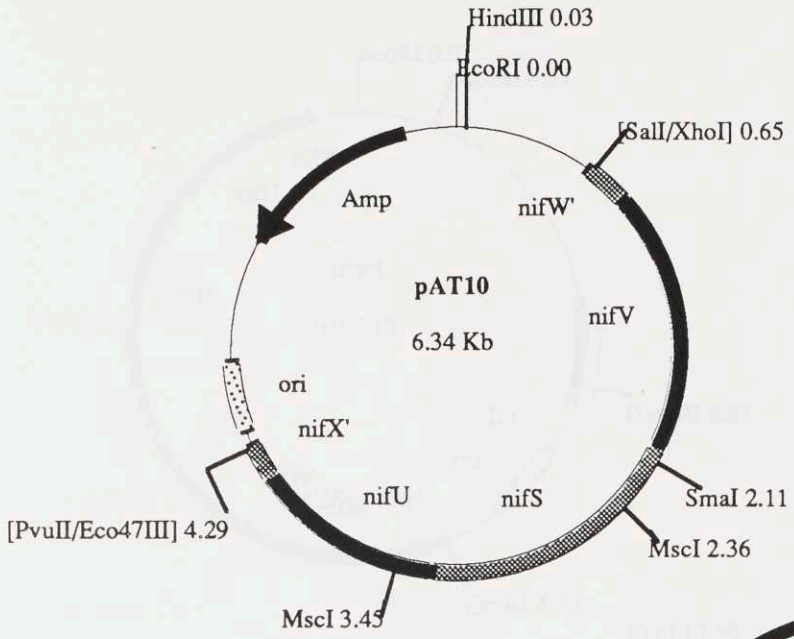
UN	*Klebsiella pneumoniae* isolate M5a1 [obtained from W.J. Brill at the University of Wisconsin, Madison, 1984]
UN1990	point mutant of UN in *nif*V(allele number 4944, MacNeil et al., 1978b)[obtained from W.J. Brill at the University of Wisconsin, Madison, 1984]

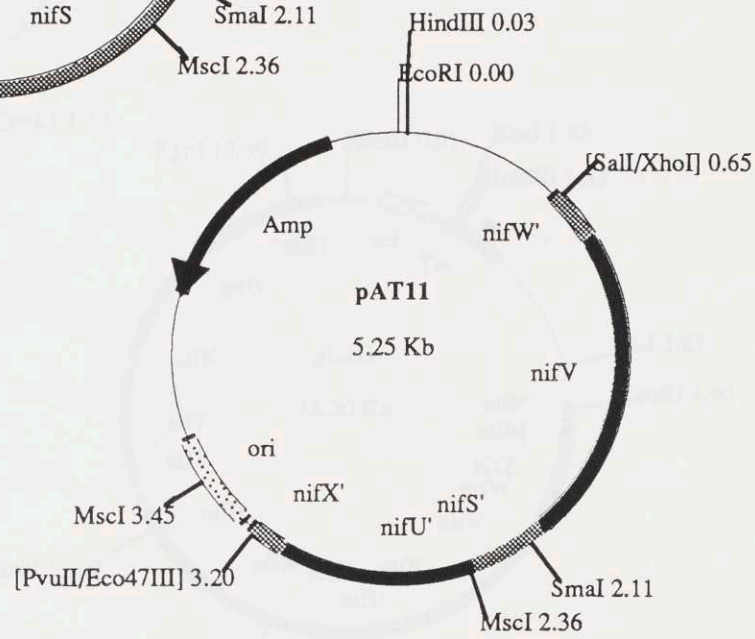
Appendix C: Plasmid Maps

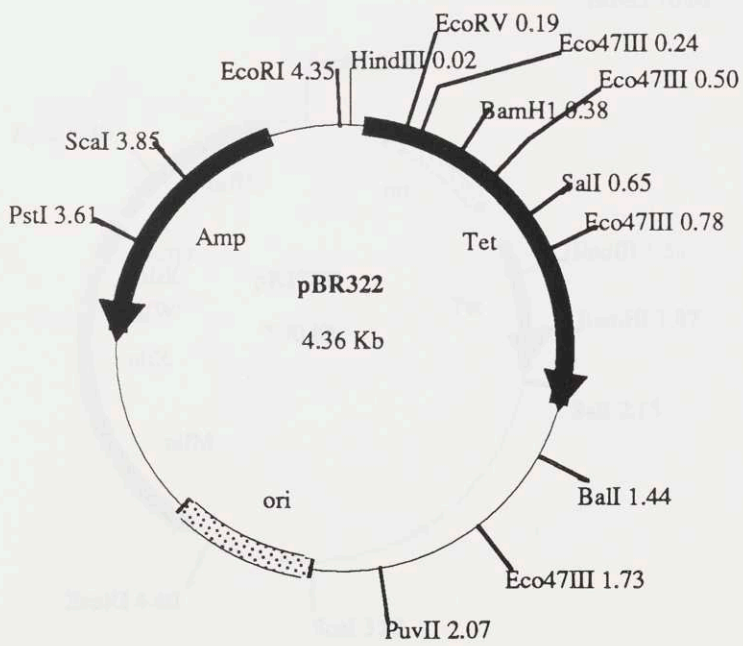


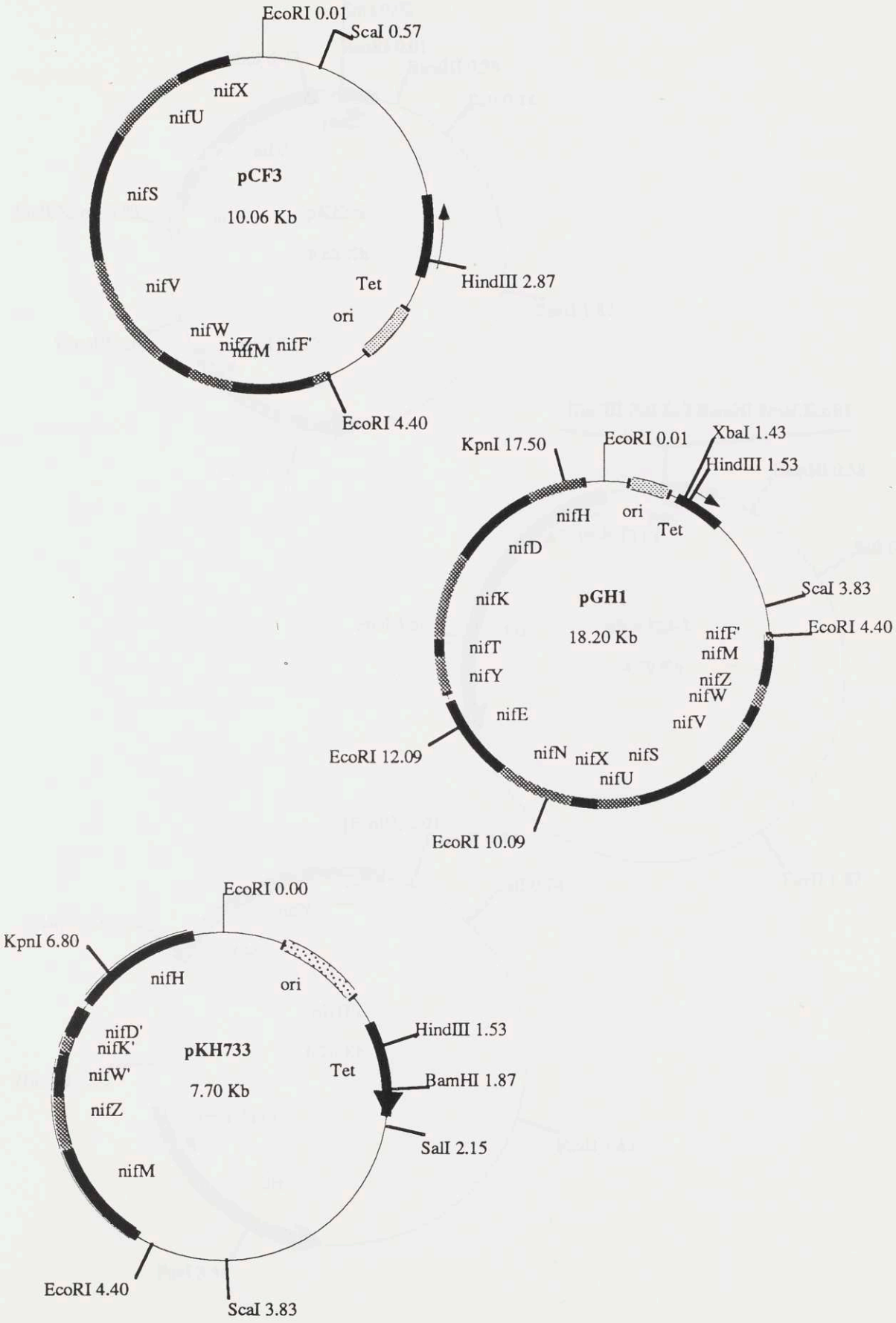


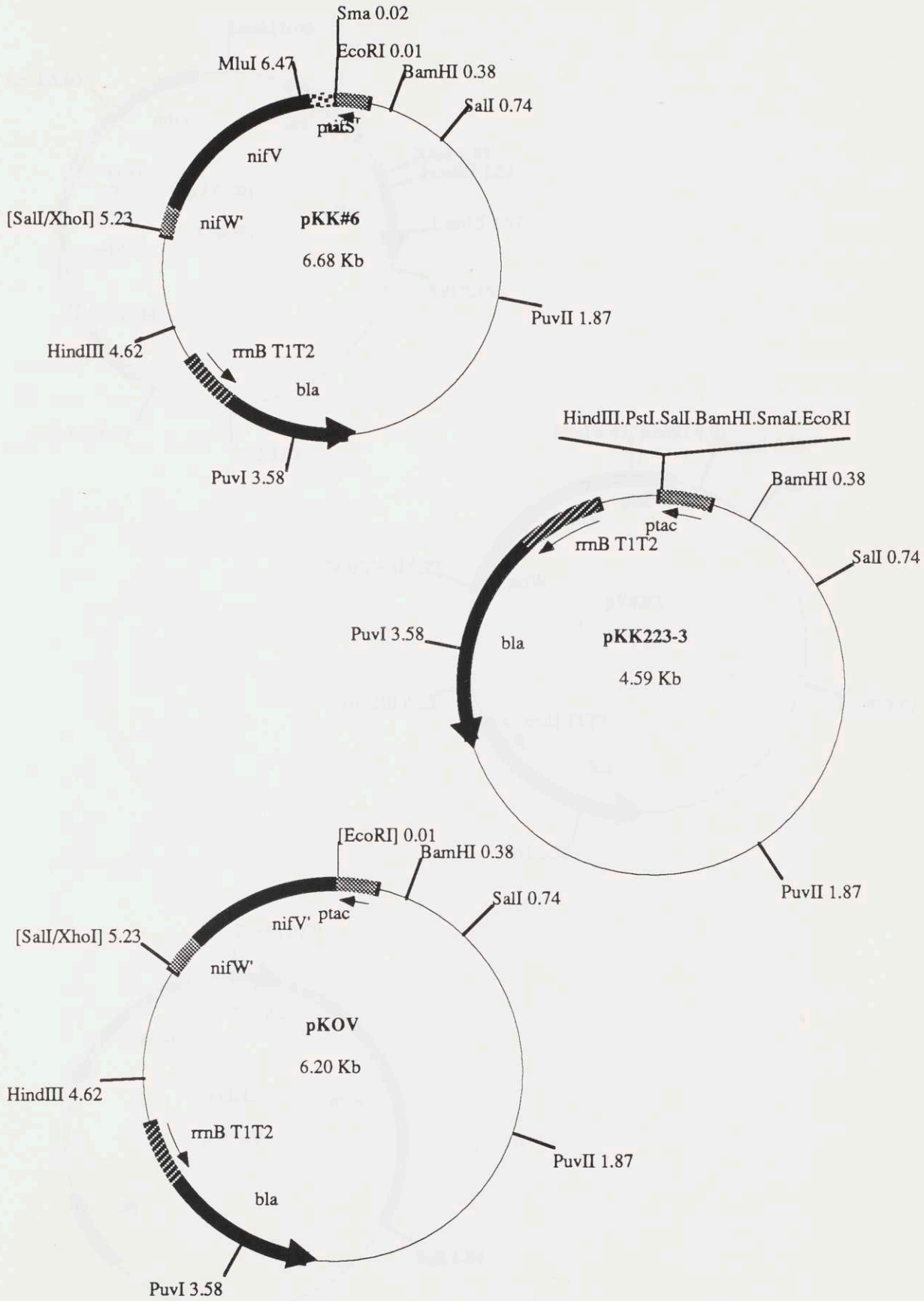


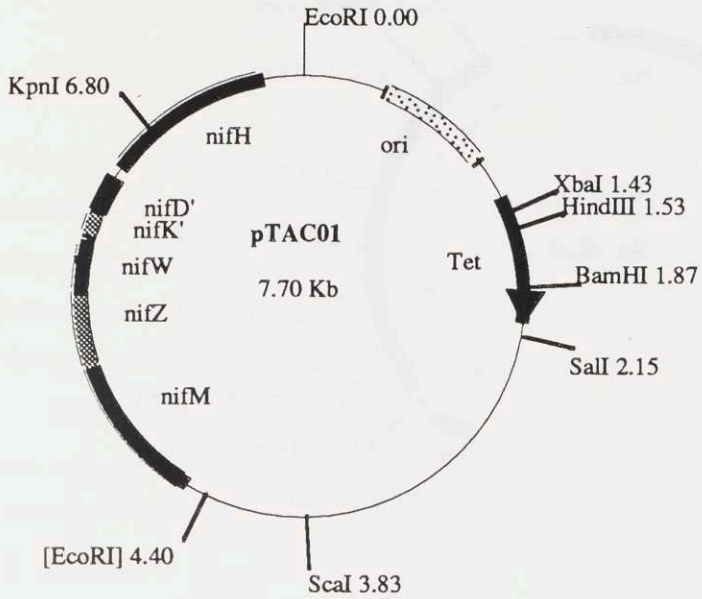

HindIII 0.03
EcoRI 0.00
[SalI/XhoI] 0.65
Amp
nifW'
pAT10
6.34 Kb
nifV
ori
nifX'
nifU
nifS
SmaI 2.11
[PvuII/Eco47III] 4.29
MscI 2.36
MscI 3.45

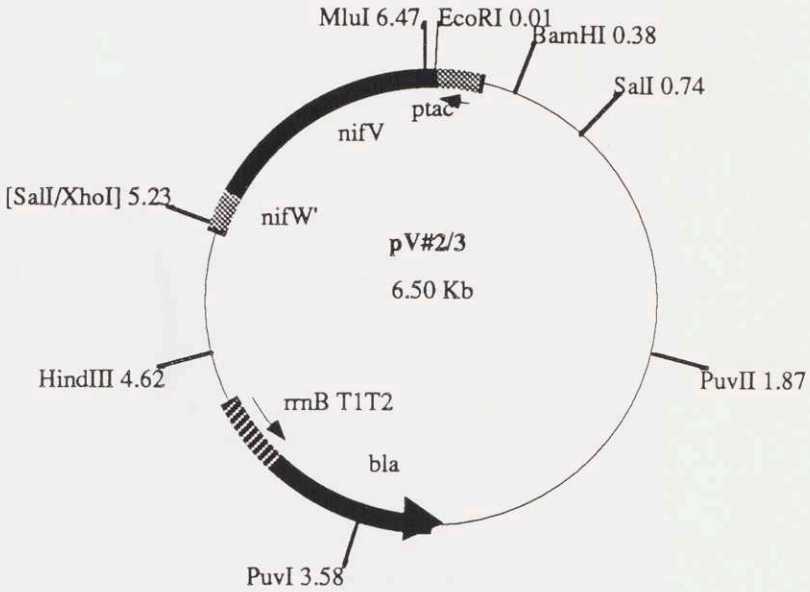

HindIII 0.03
EcoRI 0.00
[SalI/XhoI] 0.65
Amp
nifW'
pAT11
5.25 Kb
nifV
ori
nifX'
nifS'
nifU'
MscI 3.45
[PvuII/Eco47III] 3.20
SmaI 2.11
MscI 2.36

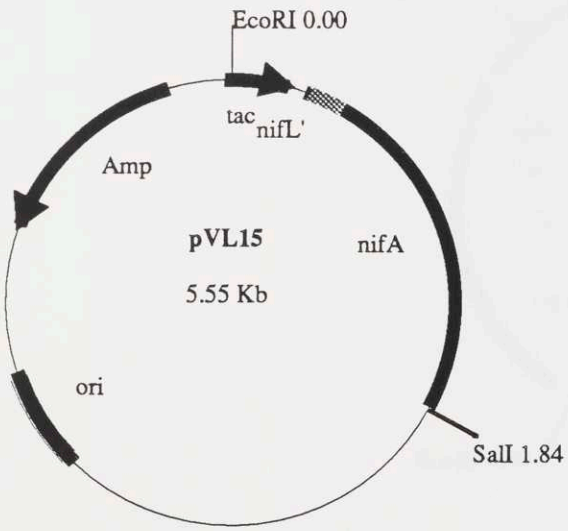

EcoRV 0.19
Eco47III 0.24
EcoRI 4.35
HindIII 0.02
Eco47III 0.50
BamH1 0.38
ScaI 3.85
SalI 0.65
PstI 3.61
Eco47III 0.78
Amp
Tet
pBR322
4.36 Kb
ori
BalI 1.44
Eco47III 1.73
PuvII 2.07

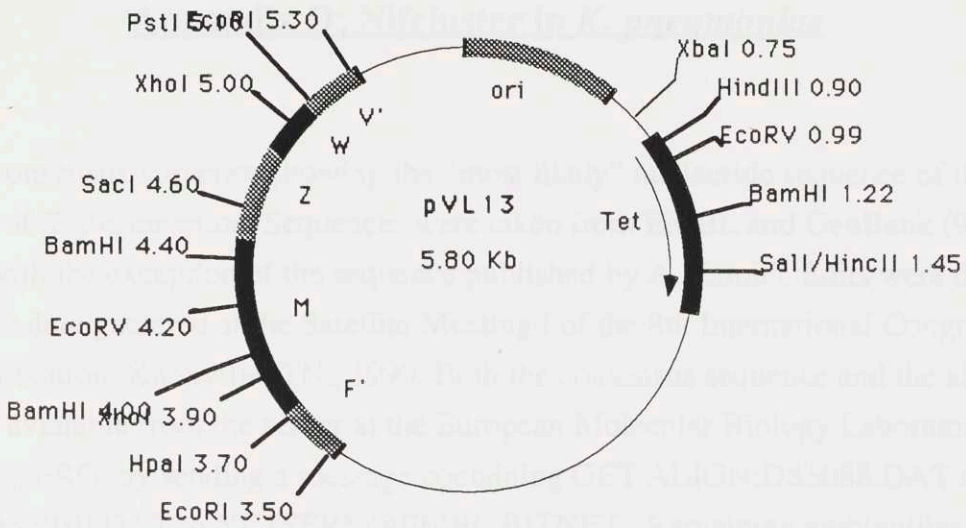

EcoRI 0.01
ScaI 0.57
nifX
nifU
pCF3
nifS
10.06 Kb
nifV
Tet
HindIII 2.87
ori
nifW
nifZ
nifM
nifF'
EcoRI 4.40
KpnI 17.50
EcoRI 0.01
XbaI 1.43
HindIII 1.53
nifH
ori
Tet
nifD
ScaI 3.83
nifK
pGH1
EcoRI 4.40
18.20 Kb
nifF'
nifT
nifM
nifY
nifZ
nifW
nifE
nifV
EcoRI 12.09
nifN
nifX
nifS
nifU
EcoRI 10.09
EcoRI 0.00
KpnI 6.80
nifH
ori
nifD'
HindIII 1.53
nifK'
pKH733
Tet
nifW'
BamHI 1.87
7.70 Kb
nifZ
SalI 2.15
nifM
EcoRI 4.40
ScaI 3.83

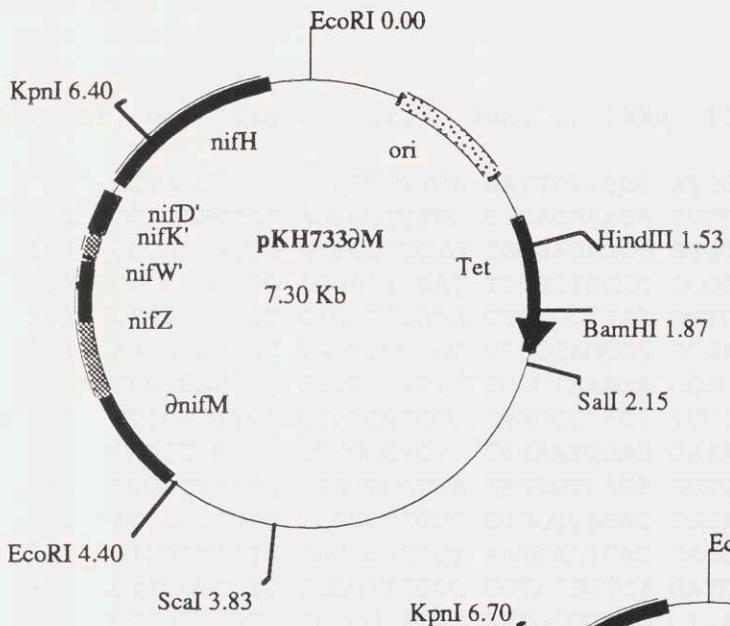

Sma 0.02
EcoRI 0.01
MluI 6.47
BamHI 0.38
SalI 0.74
ptaS
nifV
[SalI/XhoI] 5.23
nifW'
pKK#6
6.68 Kb
PuvII 1.87
HindIII 4.62
rrnB T1T2
bla
PuvI 3.58
HindIII.PstI.SalI.BamHI.SmaI.EcoRI
BamHI 0.38
ptac
rrnB T1T2
SalI 0.74
PuvI 3.58
bla
pKK223-3
4.59 Kb
PuvII 1.87
[EcoRI] 0.01
BamHI 0.38
SalI 0.74
nifV'
ptac
[SalI/XhoI] 5.23
nifW'
pKOV
6.20 Kb
HindIII 4.62
rrnB T1T2
PuvII 1.87
bla
PuvI 3.58

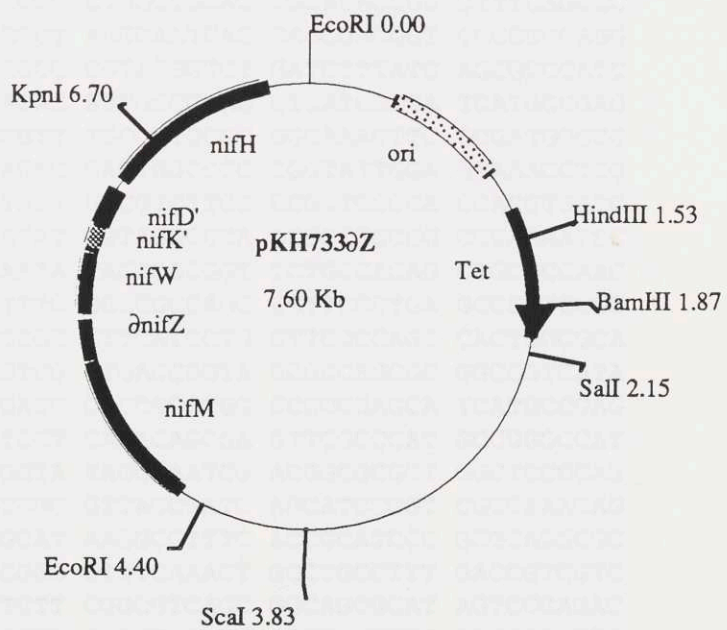

EcoRI 0.00
KpnI 6.80
nifH
ori
XbaI 1.43
HindIII 1.53
nifD'
nifK'
pTAC01
nifW
Tet
BamHI 1.87
7.70 Kb
nifZ
SalI 2.15
nifM
[EcoRI] 4.40
ScaI 3.83


MluI 6.47
EcoRI 0.01
BamHI 0.38
SalI 0.74
ptac
nifV
[SalI/XhoI] 5.23
nifW'
pV#2/3
6.50 Kb
HindIII 4.62
PuvII 1.87
rrnB T1T2
bla
PuvI 3.58


EcoRI 0.00
tac
nifL'
Amp
pVL15
nifA
5.55 Kb
ori
SalI 1.84


PstI
EcoRI 5.30
XhoI 5.00
V'
W
Z
M
F'
SacI 4.60
BamHI 4.40
EcoRV 4.20
BamHI 4.00
XhoI 3.90
HpaI 3.70
EcoRI 3.50
ori
XbaI 0.75
HindIII 0.90
EcoRV 0.99
BamHI 1.22
SalI/HincII 1.45
Tet
pVL13
5.80 Kb


EcoRI 0.00
KpnI 6.40
nifH
ori
nifD'
nifK'
nifW'
nifZ
∂nifM
pKH733∂M
7.30 Kb
Tet
HindIII 1.53
BamHI 1.87
SalI 2.15
EcoRI 4.40
ScaI 3.83


EcoRI 0.00
KpnI 6.70
nifH
ori
nifD'
nifK'
nifW
∂nifZ
nifM
pKH733∂Z
7.60 Kb
Tet
HindIII 1.53
BamHI 1.87
SalI 2.15
EcoRI 4.40
ScaI 3.83

Appendix D: Nifcluster in *K. pneumoniae*

This is a consensus sequence showing the "most likely" nucleotide sequence of the nifcluster of *K. pneumoniae*. Sequences were taken from EMBL and GenBank (9/89) and aligned (with the exception of the sequence published by A. Zamir). Edits were done to reflect new data gathered at the Satellite Meeting I of the 8th International Congress on Nitrogen Fixation, Knoxville, TN., 1990. Both the consensus sequence and the alignments are available from the server at the European Molecular Biology Laboratory in Heidelberg, FRG, by sending a message containing GET ALIGN:DS5088.DAT and GET ALIGN:DS5089.DAT to NETSERV@EMBL.BITNET. Remaining ambiguities are listed below:

```
ckpcluster.seq  Length: 24213  June 8, 1990  13:21  Check: 2397  ..

      1  GGTAACCCGC TACGGCTTGA GATTATCCGC ATCCTTGCCG ACGGCAGCGA
     51  GCAGAGCTGT AACGCCCTGC GTCACGAAGA TGTGGCGAAG TCGACCATGA
    101  CCCACCACTG GCGCGTCCTG CGCGACAGCG GTGTGATCTG GCAGCGCCCA
    151  CAGGGGCGGG AGAACTTGAT TTCGCTGCGC CGGGAAGATT TAGACGCGCG
    201  CTTTCCCGGC CTGCTGGATA CGCTGCTTAA GGTCATGCAG CAGGAGAACT
    251  AAAGGCCCGC TACTCCTCGC CGGCCAGCCG CCGATACTGG GCAAAGCGGG
    301  CCCGCGCGTC CTCCTCGGTT CGGCTAAAGA GCGCATCCGC CAGATGCGGC
    351  GTCGTTTTGT GCAGCGAGGC GTAGCGCACT TCGCCAAGCA AAAAGTCGCG
    401  GAAGCTCTCC TCCGGCTCTT CGGAATCGAG CATAAACGGC GTCTTACCTT
    451  CCGCTTCCCG CTGCGGATGA TAGCGCCACA GGTGCCAGTA TCCCGCCTCA
    501  ACCGCCCGTT TCGCCTCGCG CTGGCTGCAG CGCATACCGG CTTTCAGCCC
    551  GTGGTTAATG CAGGCGGCGT AGGCAATCAC CAGCGACGGT CCCGGCCAGG
    601  CTTCGGCCTC GGCGATCGCC CGTAGGGTCT GATCTTTATC AGCGCCCATC
    651  GCGACCTGGG CCACGTACAC ATTGCCGTAG CTCATCGCCA TCATGCCGAG
    701  ATCTTTTTTC CGCGTGCGTT TGCCCTGCGC GGCAAACTTC GCGATGGCCG
    751  CCACCGGGGT CGATTTAGAC GACTGGCCGC CGGTATTGGA GTAAACCTCG
    801  GTGTCAAACA CCAGAATATT GACGTCTTCC CCGCTCGCCA GCACGTGATC
    851  GAGACCGCCG AAGCCGATAT CGTAGGCCCA GCCGTCGCCG CCGAAAATCC
    901  ACTGCGAACG ACGAACAAAA TAGTCGCGGT TCTGCCACAG CTGCTCCAAC
    951  AGCGGCACGC CCTCTTTTTC CGCCGCCAGC CGTTCGCTGA GCCGGTCCGC
   1001  GCGCTCGCGG GTGCCCTCGC CTTCATCCTG CTTCGCCAGC CACTGGCGCA
   1051  TTGCGTCGCT AAGTTCGTCG CTGACCGGTA GCGCCAGCGC GGCGGTCATA
   1101  TCATCGGCGA TTTGTTGACG CACCGCCTGG CCGCCGAGCA TCATGCCGAG
   1151  GCCAAACTCC GCATTATCCT CAAACAGCGA GTTCGCCCAT GCCGGGCCAT
   1201  GGCCGCGGTG GTTGGTGGTA TAGGGAATCG ACGGCGCGCT GGCTCCCCAG
   1251  ATAGAAGAGC AGCCGGTGGC GTTAGCGATC AGCATCCGGT CGCCAAACAG
   1301  CTGGGTTATC AGGCGGGCAT AAGGCGTTTC ACCGCATCCC GCGCAGGCGC
   1351  CGGAAAACTC CAGCAGCGGG GTTTCAAACT GGCTGCCTTT GACCGTCGTC
   1401  TTACGAAACG GATTGCTCTT CGGCGTCAGC GCCAGCGCAT AGTCCCAGAC
   1451  CGGCGCCATC TGACGCTGGC TATCGAGAGA CTGCATTTTT AACGCCTTGC
```

```
1501  CGCGCGCGGG ACAGATATCC ACGCAGTTGC CGCAGCCGGA ACAATCCAGC
1551  GGCGAGATAG CCAGATGGTA GTGATACTCC TTCGCTCCCT GCGCGGGTTT
1601  GCTCAGCAGC CCAACCGGCG CGGCGTCATG CTCTTCGCCG TTGAGCAGCG
1651  CCGGGCGGAT CGCCGCATGC GGGCAGATAA AGGCGCACTG GTTACACTGC
1701  GTGCAGCCCT CCGGCTGCCA GACCGGCACT TCCAGCGCGA TCCCGCGTTT
1751  CTCCCACGCG GCGGTGCCCG AAGGAAAGGT CCCGTCCTCC ATACCGACGA
1801  ACGCGCTCAC CGGCAGCTGG TCGCCGCACT GGCGGTTCAT CGGCTGCAGA
1851  ATATCGCGGA TGAAATCCGG CATCATGGCT GATGCTTGCG CCGCGGGTTC
1901  ATCCAGCGTC GCCCAGTGCG CCGGAATCGT CACCTGATGC AGCGAGGCCA
1951  TGCCCAGCTC GATCGCCCGC TGGTTCATCT CAATCACCGC CGCCCCTTTG
2001  CTGCCGTAGC TTTTTTCAAC CGCCTGCTTG AGGTAATCCG CCGCGGTCTG
2051  CGGGTCGATA ATCGCCGCCA GCTTAAAGAA CGCCGCCTGC ATCAGCATAT
2101  TAAAGCGCCC GCCCAGCCCG AGCTCGCGGG CGATATCCAC GGCGTTCAGG
2151  GTATAAAAAT GGATATTTTC CCGCGCCAGA TAGCGTTTAA AGCCGACCGG
2201  CAGATGCTGC TCCAGCTCCG CATCGGACCA GCTGCAGTTG AGTAAAAAGG
2251  TCCCGCCCGG CTTTAATCCG TCCAGCAGAT CGTAGCGCTC AACGTAGGAC
2301  TGCTGCGAAC AGGAGATAAA ATCGGCCCGA TGGATCAGGT AGGGCGAATT
2351  GATCGGCCGG TCGCCGAAGC GTAAATGTGA AACGGTAATG CCGCCGGATT
2401  TTTTCGAGTC ATAAGAAAAG TAGGCCTGCG CGTAGAGCGG CGTTTTATCG
2451  CCGATAATTT TGATCGCGCT TTTATTGGCC CCGACGGTGC CGTCCGAGCC
2501  CATGCCCCAA AATTTACAGG CGGTGATGCC GTCATGCGAG ACCGCCAGCG
2551  TCTGCTGGGC CGGCGGTAAC GAAGTAAAGG TTACATCATC GACAATCCCG
2601  AGGGTAAACC CGTCCATCGG CAGCGGTTTA TTGAGGTTAT CAAAGACGGC
2651  CGCGATATCG TTGGGCAGAA CATCCTTCCC GCCAAGCGCA TAGCGGCCGC
2701  CGACGATTAG CGGCGCATCG TCGTGGTGGT AGAAGGCGTT TTTCACATCC
2751  AGGCACAGCG GTTCAGCCTG AGCGCCGGGC TCTTTGGTAC GGTCAAGGAC
2801  GGCAATCCGC TGCACGGTTT TCGGCAGCTG GGCGAAGAAG TGGGCCAGCG
2851  AAAAAGGGCG AAACAGATGC ACGCTGAGCA GCCCGACCTT CTCTCCCGCC
2901  GCGTTCAGCG TATCCACCAC TTCCTGAACG GTATCGCAGA CCGATCCCAT
2951  TGCGATAATC ACCCGTTCGG CATCCGCCGC GCCGGTATAG TTAAACAGAT
3001  GATACTCCCG GCCGGTGAGC GCGCTGATTT GCGTCATATA GCTTTCGACA
3051  ATGTCGGGCA GCGCCTGATA AAAACGGTTG CCCGCCTCCC GCTCCTGGAA
3101  GTAGATATCC GGGTTCTGCG CCGTTCCGCG GATGACCGGA TGATCCGGAT
3151  GCAGCGCGTT ACGGCGGAAG CTGTCGAGCG CGGGCCGGTC CAGCAGCGTC
3201  GCCAGCTGCT CATATTCCAA CACCTCGATT TTTTGAATTT CGTGCGAGGT
3251  GCGAAAACCG TCGAAGAAGT TAACAAACGG GATGCGTCCC TTAATCGCCG
3301  CCAGATGCGC CACCGCCGAC AAATCCATCA CCTGCTGCAC GTTGTTCTCC
3351  GCCAGCATCG CGCAGCCGCT CTGGCGGACC GCCATCACAT CCTGGTGATC
3401  GCCAAAAATA TTCAGCGAAT TGGTCGCCAG CGCCCGGGCG CTGACGTGAA
3451  AGACGCCCGG CAGCAGTTCA CCGGCGATTT TGTACATGTT GGGGATCATC
3501  AGCAGCAGCC CCTGGGAGGC CGTATAGGTG GTGGTGAGCG CCCCGGCCTG
3551  CAGCGCGCCG TGGACCGCGC CTGCCGCGCC GGCCTCCGAC TGCATCTCCA
3601  TTAAGCGCAC CGGCTGGCCA AAAAGGTTCT TTTTCCCCTG CGCCGCCCAC
3651  TCGTCGACGT TTTCCGCCAT CGGCGTGGAG GGGGTTATGG GGTAAATCGC
3701  CGCGACCTCG GTAAAGGCAT AAGAGATCCA GGCCGCCGCG GCGTTGCCAT
3751  CCATTGTTTT CATTTTTCCG GACATTGTTC AATCCTCGAA GGTGAGAGGC
3801  ATCTTCGCCG CCTCAAATAA GCGGCAAACC CAGTTGTTGC CTCAAGCACA
3851  GCCTGTGCCA GCTCGCGGAT GACAGAAGAG TTAGCGCGAA TTCAACGCGT
3901  TATGAAGAGA GTCGCCGCGC AGCGCGCCAA GAGATTGCGT GGAATAAGAC
3951  ACAGGGGGCG ACAAGCTGTT GAACAGGCGA CAAAGCGCCA CCATGGCCCC
4001  GGCAGGCGCA ATTGTTCTGT TTCCCACATT TGGTCGCCTT ATTGTGCCGT
4051  TTTGTTTTAC GTCCTGCGCG GCGACAAATA ACTAACTTCA TAAAAATCAT
4101  AAGAATACAT AAACAGGCAC GGCTGGTATG TTCCCTGCAC TTCTCTGCTG
4151  GCAAACACTC AACAACAGGA GAAGTCACCA TGACCATGCG TCAATGCGCT
```

```
4201  ATTTACGGTA AAGGCGGTAT CGGTAAATCC ACCACCACGC AGAACCTCGT
4251  CGCCGCGCTG GCGGAGATGG GTAAGAAAGT GATGATCGTC GGCTGCGATC
4301  CGAAGGCGGA CTCCACCCGT CTGATTCTGC ACGCCAAAGC ACAGAACACC
4351  ATTATGGAGA TGGCCGCGGA AGTCGGCTCG GTCGAGGACC TCGAACTCGA
4401  AGACGTGCTG CAAATTGGCT ACGGCGATGT GCGCTGCGCG GAATCCGGCG
4451  GCCCGGAGCC AGGCGTCGGC TGCGCGGGAC GCGGCGTGAT CACGGCGATC
4501  AACTTTCTTG AAGAAGAAGG CGCCTACGAG GACGATCTCG ATTTCGTGTT
4551  CTATGACGTG CTCGGCGACG TGGTCTGCGG CGGCTTCGCC ATGCCGATCC
4601  GCGAAAACAA AGCCCAGGAG ATCTACATCG TCTGCTCCGG CGAAATGATG
4651  GCGATGTACG CGGCCAACAA TATCTCCAAA GGGATCGTTA AATACGCCAA
4701  ATCCGGCAAG GTGCGCCTCG GCGGCCTGAT CTGTAACTCA CGTCAGACCG
4751  ACCGTGAAGA CGAACTGATT ATTGCCCTGG CGGAAAAGCT CGGTACCCAG
4801  ATGATCCACT TTGTGCCCCG CGACAACATC GTGCAGCGCG CGGAGATCCG
4851  CCGCATGACG GTTATCGAGT ACGACCCCGC CTGTAAACAG GCCAACGAAT
4901  ACCGCACCCT GGCGCAGAAG ATCGTCAACA ACACCATGAA AGTGGTGCCG
4951  ACGCCCTGCA CCATGGATGA GCTGGAATCG CTGCTGATGG AGTTCGGCAT
5001  CATGGAAGAG GAAGACACCA GCATCATTGG CAAAACCGCC GCCGAAGAAA
5051  ACGCGGCCTG AGCACAGGAC AATTATGATG ACCAACGCAA CGGGCGAACG
5101  TAATCTGGCG CTGATCCAGG AAGTCCTGGA GGTGTTCCCG GAAACCGCGC
5151  GAAAAGAGCG CAGAAAGCAC ATGATGGTCA GCGATCCGAA AATGAAGAGC
5201  GTCGGCAAGT GCATTATCTC TAACCGCAAA TCACAACCCG GCGTAATGAC
5251  CGTACGCGGC TGCGCCTACG CCGGTTCCAA AGGGGTGGTA TTTGGGCCGA
5301  TTAAGGATAT GGCCCATATT TCGCACGGAC CGGCTGGCTG CGGCCAGTAT
5351  TCCCGCGCCG AACGACGCAA CTACTACACC GGAGTCAGCG GCGTCGATAG
5401  CTTCGGCACG CTGAACTTCA CCTCTGATTT TCAGGAGCGC GACATCGTCT
5451  TCGGCGGCGA TAAAAAGCTC AGCAAGCTGA TTGAAGAGAT GGAGTTGCTG
5501  TTCCCGCTCA CCAAAGGGAT CACCATTCAG TCGGAATGCC CGGTGGGGCT
5551  GATCGGTGAT GATATCAGCG CGGTGGCCAA CGCCAGCAGC AAGGCGCTGG
5601  ATAAACCGGT GATCCCGGTA CGCTGCGAAG GCTTTCGCGG CGTGTCGCAG
5651  TCTCTGGGGC ACCATATCGC CAACGACGTG GTGCGCGACT GGATCCTGAA
5701  CAATCGCGAA GGACAGCCGT TTGAAACCAC CCCTTACGAT GTGGCGATCA
5751  TCGGCGACTA CAACATCGGC GGCGACGCCT GGGCCTCGCG CATTCTGCTG
5801  GAAGAGATGG GGCTACGGGT AGTCGCGCAG TGGTCCGGCG ACGGCACGCT
5851  GGTGGAGATG GAGAATACCC CATTCGTCAA GCTGAACCTG GTTCACTGCT
5901  ACCGTTCGAT GAACTATATC GCCCGCCATA TGGAGGAGAA ACATCAGATT
5951  CCGTGGATGG AGTACAACTT CTTCGGGCCG ACCAAAATCG CCGAATCGCT
6001  GCGCAAAATC GCCGACCAGT TCGACGATAC CATTCGCGCG AACGCCGAAG
6051  CGGTGATCGC CCGGTATGAG GGGCAGATGG CGGCGATTAT CGCCAAATAT
6101  CGCCCGCGCC TGGAGGGGCG TAAGGTGCTG CTCTATATGG GCGGCCTGCG
6151  GCCGCGCCAC GTTATTGGCG CCTATGAGGA TCTCGGGATG GAGATCATCG
6201  CCGCCGGCTA CGAGTTTGCC CATAACGATG ATTACGACCG CACCCTGCCG
6251  GATCTGAAAG AGGGCACGCT GCTGTTCGAT GACGCCAGCA GCTACGAGCT
6301  GGAAGCGTTC GTCAAGGCGC TGAAGCCCGA CCTTATCGGC TCCGGCATCA
6351  AGGAAAAATA TATCTTCCAG AAAATGGGCG TGCCGTTCCG CCAGATGCAC
6401  TCGTGGGACT ATTCCGGCCC GTACCACGGC TACGATGGTT TCGCCATTTT
6451  CGCCCGCGAT ATGGATATGA CCCTGAACAA CCCGGCGTGG AACGAACTGA
6501  CCGCTCCGTG GCTGAAGTCT GCGTGATTGC CCACTCACTG TCCCGTCTGT
6551  TCACCGATTT GTGGCGCGGG AGGAGAACAC CATGAGCCAA ACGATTGATA
6601  AAATTAATAG CTGTTATCCG CTATTCGAAC AGGATGAATA CCAGGAGCTG
6651  TTCCGCAATA AGCGGCAGCT GGAAGAGGCG CACGATGCGC AGCGCGTGCA
6701  GGAGGTCTTT GCCTGGACCA CCACCGCCGA GTATGAAGCG CTGAATTTCC
6751  GACGCGAGGC GCTGACCGTT GACCCGGCGA AAGCCTGCCA GCCGCTTGGC
6801  GCGGTGCTTT GCTCGCTGGG ATTTGCCAAC ACCCTGCCGT ATGTGCACGG
6851  CTCTCAGGGG TGCGTGGCCT ACTTTCGCAC CTATTTTAAC CGCCATTTCA
```

```
6901  AAGAGCCGAT CGCCTGCGTC TCCGACTCGA TGACCGAAGA CGCGGCGGTC
6951  TTCGGCGGCA ACAACAATAT GAACCTGGGC CTGCAGAACG CCAGCGCGCT
7001  GTACAAACCG GAGATCATTG CGGTGTCCAC CACCTGCATG GCGGAAGTTA
7051  TCGGCGATGA CCTGCAGGCG TTTATCGCCA ACGCTAAAAA AGATGGCTTC
7101  GTCGACAGCA GCATCGCCGT GCCCCACGCC CATACGCCAA GCTTTATCGG
7151  CAGCCACGTC ACCGGCTGGG ATAACATGTT TGAAGGCTTC GCCAAAACCT
7201  TCACTGCGGA CTACCAGGGG CAGCCGGGCA AATTGCCGAA GCTCAATCTG
7251  GTGACCGGCT TTGAAACCTA TCTCGGCAAC TTCCGCGTAT TAAAGCGGAT
7301  GATGGAACAG ATGGCGGTGC CGTGCAGCCT GCTCTCCGAT CCGTCGGAAG
7351  TTCTCGACAC GCCCGCCGAC GGTCACTATC GGATGTATTC CGGCGGCACC
7401  ACGCAGCAGG AGATGAAAGA GGCCCCTGAC GCCATCGATA CGCTGCTCCT
7451  GCAGCCGTGG CAGCTGCTGA AGAGCAAAAA AGTGGTGCAG GAGATGTGGA
7501  ACCAGCCCGC CACCGAGGTC GCCATTCCGC TGGGGCTGGC CGCCACCGAT
7551  GAACTGCTGA TGACCGTCAG CCAGCTTAGC GGCAAGCCGA TTGCCGACGC
7601  CCTCACCCTT GAGCGCGGCC GGCTGGTTGA CATGATGCTC GACTCCCACA
7651  CCTGGCTGCA CGGCAAGAAG TTTGGCCTGT ACGGCGATCC GGACTTCGTG
7701  ATGGGCCTCA CCCGCTTCCT GCTGGAGCTG GGCTGCGAGC CAACGGTGAT
7751  CCTGAGCCAT AACGCCAACA AACGCTGGCA AAAAGCGATG AACAAAATGC
7801  TCGATGCCTC GCCGTACGGG CGCGATAGCG AAGTGTTTAT CAACTGCGAT
7851  TTGTGGCACT TCCGTTCGCT GATGTTCACC CGTCAGCCGG ACTTTATGAT
7901  CGGCAACTCC TACGGCAAGT TTATCCAGCG CGATACCCTG GCGAAGGGTA
7951  AAGCCTTTGA AGTGCCGCTT ATCCGCCTCG GCTTTCCGCT GTTCGACCGC
8001  CACCATCTGC ACCGCCAGAC AACCTGGGGT TATGAAGGGG CGATGAACAT
8051  TGTGACGACG CTGGTGAACG CCGTGCTGGA GAAACTGGAT AGCGATACCA
8101  GCCAGCTGGG CAAAACCGAT TACAGCTTCG ATCTCGTCCG TTAACCATCA
8151  GGTGCCCCGC GTCATGCGGC GGCAGGAGGG AGTATGCCCA TCGTGATTTT
8201  CCGTGAGCGC GGCGCGGACC TGTACGCCTA TATCGCGAAA CAGGATCTGG
8251  AAGCGCGAGT GATCCAGATT GAGCATAACG ACGCTGAACG CTGGGGCGGC
8301  GCGATTTCGC TGGAGGGGGG ACGCCGCTAC TACGTGCATC CGCAGCCGGG
8351  GCGTCCCGTC TTTCCGATAA GCCTGCGCGC GACGCGCAAT ACCTTGATAT
8401  AAGGAGCTAG TGATGTCCGA CAACGATACC CTATTCTGGC GTATGCTGGC
8451  GCTGTTTCAG TCTCTGCCGG ACCTACAGCC GGCGCAAATC GTCGACTGGC
8501  TGGCGCAGGA GAGCGGCGAG ACGCTGACGC CAGAGCGTCT GGCGACCCTG
8551  ACCCAGCCGC AGCTGGCCGC CAGCTTTCCC TCCGCGACGG CGGTGATGTC
8601  CCCCGCTCGC TGGTCGCGGG TGATGGCGAG CCTGCAGGGC GCGCTGCCCG
8651  CCCATTTACG CATCGTTCGC CCTGCCCAGC GCACGCCGCA GCTGCTGGCG
8701  GCATTTTGCT CCCAGGATGG GCTGGTGATT AACGGCCATT TCGGCCAGGG
8751  ACGACTGTTT TTTATCTACG CGTTCGATGA ACAAGGCGGC TGGTTGTACG
8801  ATCTGCGCCG CTATCCCTCC GCCCCCCACC AGCAGGAGGC CAACGAAGTG
8851  CGCGCCCGGC TTATTGAGGA CTGTCAGCTG CTGTTTTGCC AGGAGATAGG
8901  CGGGCCCGCC GCCGCGCGGC TGATCCGCCA TCGCATCCAC CCGATGAAAG
8951  CGCAGCCCGG GACGACGATT CAGGCACAGT GCGAGGCGAT CAATACGCTG
9001  CTGGCCGGCC GTTTGCCGCC GTGGCTGGCG AAGCGGCTTA ACAGGGATAA
9051  CCCTCTGGAA GAACGCGTTT TTTAATCCCT GTTTTGTGCT TGTTGCCCGC
9101  TGACCCCGCG GGCTTTTTTT CGCGTATGGA CGCTCTTCCC CACGTTACGC
9151  TCAGGGGAAT ATTCCGTTCA CGGTTGTTCC GGGCTTCTTG ATGCGCCTAA
9201  CCCCCTCGCT GCCAGCCTTT CATCAACAAA TAGCCATCCC AGCGCGATAG
9251  GTCATAAAGC ATCACATGCC GCCATCCCTT GTCCGATTGT TGGCTTTGTC
9301  GCAAAGCCAA CAACCTCTTT TCTTTAAAAA TCAAGGCTCC GCTTCTGGAG
9351  CGCGAATTGC ATCTTCCCCC TCATCCCCCA CCGTCAACGA GGTCACTATG
9401  AAGGGAAATG AAATTCTGGC GCTGCTGGAT GAACCGGCCT GTGAACACAA
9451  CCATAAACAA AAATCCGGCT GCAGCGCGCC CAAACCCGGC GCCACCGCCG
9501  CGGGCTGCGC GTTCGACGGC GCGCAGATAA CCCTGCTGCC CATCGCCGAC
9551  GTGGCGCATC TGGTCCACGG CCCCATCGGC TGCGCCGGAA GCTCATGGGA
```

```
 9601  TAACCGCGGC AGCGCCAGCT CCGGCCCCAC CCTTAATCGG CTCGGGTTCA
 9651  CCACCGATCT CAACGAACAG GACGTGATTA TGGGCCGCGG CGAACGCCGA
 9701  CTGTTTCACG CCGTGCGCCA TATCGTCACC CGCTATCATC CGGCGGCGGT
 9751  CTTTATCTAC AACACCTGCG TACCGGCCAT GGAGGGCGAT GACCTGGAAG
 9801  CGGTATGCCA GGCCGCGCAG ACCGCCACCG GCGTACCGGT TATCGCTATT
 9851  GACGCCGCCG GTTTCTACGG CAGTAAAAAT CTCGGTAACC GGCTGGCGGG
 9901  CGACGTCATG GTCAAACGGG TCATCGGCCA GCGCGAGCCC GCCCCCTGGC
 9951  CGGAGAGCAC GCTCTTTGCC CCGGAGCAGC GTCACGATAT TGGCCTGATT
10001  GGCGAATTCA ATATTGCCGG CGAGTTCTGG CATATTCAGC CGCTGCTCGA
10051  CGAACTGGGG ATCCGCGTGC TCGGCAGCCT CTCCGGTGAT GGCCGCTTCG
10101  CCGAGATCCA GACCATGCAC CGGGCGCAGG CCAATATGCT GGTCTGCTCG
10151  CGGGCGTTAA TTAACGTCGC CAGAGCCCTG GAGCAGCGCT ACGGCACGCC
10201  GTGGTTCGAA GGCAGCTTTT ACGGGATCCG CGCCACCTCT GACGCCCTGC
10251  GCCAGCTGGC GGCGCTGCTG GGCGACGACG ACCTTCGCCA GCGCACCGAA
10301  GCGCTGATTG CGCGGGAGGA ACAGGCGGCG GAACTGGCGC TACAGCCGTG
10351  GCGCGAACAG CTGCGCGGCC GCAAAGCGCT GCTCTATACC GGCGGGGTGA
10401  AATCCTGGTC GGTGGTATCG GCGCTGCAGG ATTTGGGCAT GACCGTGGTG
10451  GCAACCGGCA CGCGTAAATC CACCGAAGAG GATAAACAGC GGATCCGCGA
10501  GCTGATGGGC GAAGAGGCGG TAATGCTGGA AGAGGGCAAC GCCCGCACGC
10551  TGCTGGATGT GGTCTATCGC TATCAGGCCG ACCTGATGAT TGCCGGCGGA
10601  CGCAATATGT ACACCGCCTA TAAAGCCAGG CTGCCGTTTC TCGATATCAA
10651  TCAGGAGCGC GAACACGCCT TCGCTGGCTA TCAGGGGATC GTCACCCTCG
10701  CCCGCCAGCT GTGTCAGACC ATCAACAGCC CCATCTGGCC GCAAACCCAT
10751  TCTCGCGCCC CGTGGCGCTA AGGAGCTCAC CATGGCAGAC ATTTTCCGCA
10801  CCGATAAGCC GCTGGCGGTC AGCCCCATCA AAACCGGCCA GCCGCTCGGC
10851  GCAATCCTCG CCAGCCTCGG GATCGAACAC AGCATCCCTC TGGTCCACGG
10901  CGCGCAGGGG TGCAGCGCCT TCGCCAAAGT CTTTTTTATT CAACATTTCC
10951  ACGACCCGGT TCCCCTGCAG TCGACGGCGA TGGACCCCAC GTCGACGATT
11001  ATGGGCGCGG ACGGCAATAT TTTTACCGCC CTGGATACCC TCTGCCAGCG
11051  CAACAATCCG CAGGCTATCG TACTGCTCAG CACCGGGCTG TCGGAGGCCC
11101  AGGGCAGCGA TATTTCCCGC GTGGTTCGCC AGTTTCGCGA AGAGTATCCC
11151  CGGCATAAGG GGGTGGCGAT ATTGACGGTT AACACGCCGG ATTTTTATGG
11201  CTCCATGGAG AACGGCTTCA GCGCGGTGTT AGAGAGCGTC ATTGAGCAGT
11251  GGGTGCCGCC GGCGCCGCGC CCGGCTCAGC GCAATCGCCG GGTCAATCTG
11301  CTGGTCAGCC ATCTCTGTTC GCCGGGCGAT ATCGAGTGGC TGCGCCGATG
11351  CGTCGAAGCC TTTGGTCTGC AGCCGATAAT CCTGCCGGAC CTGGCGCAAT
11401  CGATGGACGG CCACCTGGCG CAGGGCGATT TCTCGCCGCT GACCCAGGGC
11451  GGGACGCCGC TGCGCCAGAT AGAGCAGATG GGGCAAAGCC TGTGCAGCTT
11501  CGCCATTGGC GTCTCCCTTC ATCGCGCCTC ATCGCTGCTG GCCCCGCGCT
11551  GCCGCGGCGA GGTTATCGCC CTGCCGCACC TGATGACCCT CGAACGCTGC
11601  GACGCCTTTA TTCATCAACT GGCGAAAATT TCCGGACGCG CCGTTCCCGA
11651  GTGGCTGGAA CGCCAGCGCG GCCAGCTACA GGATGCGATG ATCGACTGCC
11701  ATATGTGGCT CCAGGGCCAG CGCATGGCGA TAGCGGCGGA AGGCGATTTG
11751  CTGGCGGCGT GGTGTGATTT CGCCAACAGC CAGGGGATGC AGCCCGGCCC
11801  GCTGGTGGCC CCTACCGGTC ATCCCAGCCT GCGCCAGCTG CCGGTGGAAC
11851  GGGTGGTGCC GGGGGATCTG GAGGATCTGC AAACCCTGCT GTGCGCGCAT
11901  CCCGCCGACC TGCTGGTGGC GAACTCGCAC GCCCGCGACC TGGCGGAGCA
11951  GTTTGCGCTG CCGCTGGTGC GCGCGGGTTT TCCGCTCTTT GACAAGCTCG
12001  GCGAATTCCG CCGGGTGCGA CAGGGGTATA GCGGGATGCG CGATACGCTG
12051  TTTGAGCTGG CAAACCTGAT ACGCGAGCGT CACCACCACC TCGCCCACTA
12101  CCGATCGCCG CTGCGCCAGA ACCCCGAATC GTCACTCTCC ACAGGAGGCG
12151  CTTATGCCGC CGATTAACCG TCAGTTTGAT ATGGTCCACT CCGATGAGTG
12201  GTCTATGAAG GTCGCCTTCG CCAGCTCCGA CTATCGTCAC GTCGATCAGC
12251  ACTTCGGCGC TACCCCGCGG CTGGTGGTGT ACGGCGTCAA GGCGGATCGG
```

```
12301  GTCACTCTCA TCCGGGTGGT TGATTTCTCG GTCGAGAACG GCCACCAGAC
12351  GGAGAAGATC GCCAGGCGGA TCCACGCCCT GGAGGATTGC GTCACGCTGT
12401  TCTGCGTGGC GATTGGCGAC GCGGTTTTTC GCCAGCTGTT GCAGGTGGGC
12451  GTGCGTGCCG AACGCGTTCC CGCCGACACC ACCATCGTCG GCTTACTGCA
12501  GGAGATTCAG CTCTACTGGT ACGACAAAGG GCAGCGCAAA AATCAGCGCC
12551  AGCGCGACCC GGAGCGCTTT ACCCGTCTGC TGCAGGAGCA GGAGTGGCAT
12601  GGGGATCCGG ACCCGCGCCG CTAGCCGTGT CGTTTCTGTG ACAAAGCCCA
12651  CAAAACATCG CGACACTGTA GGACGAACCT TGTCAGGACT AATACACAAC
12701  CATTTGAAAA ATATTAATTT TATTCTCTGG TATCGCAATT GCTAGTTCGT
12751  TATCGCCACC GCGCTTCCGC GGTGAACCGC GCCCCGGCGT TTTCCGTCAA
12801  CATCCCTGGA GCTGACAGCA TGTGGAATTA CTCCGAGAAA GTGAAAGACC
12851  ATTTTTTTAA CCCCCGCAAT GCGCGCGTGG TGGACAACGC CAACGCGGTA
12901  GGCGACGTCG GTTCGTTAAG CTGCGGCGAC GCCCTGCGCC TGATGCTGCG
12951  CGTCGACCCG CAAAGCGAAA TCATTGAGGA GGCGGGCTTC CAGACCTTCG
13001  GCTGCGGCAG CGCCATCGCC TCCTCCTCCG CGCTGACGGA GCTGATTATC
13051  GGCCATACCC TCGCCGAAGC CGGGCAGATA ACCAATCAGC AGATTGCCGA
13101  TTATCTCGAC GGACTGCCGC CGGAGAAAAT GCACTGCTCG GTGATGGGCC
13151  AGGAGGCCCT GCGCGCGGCC ATCGCCAACT TTCGCGGCGA AAGCCTTGAA
13201  GAGGAGCACG ACGAGGGCAA GCTGATCTGC AAATGCTTCG GCGTCGATGA
13251  AGGGCATATT CGCCGCGCGG TACAGAACAA CGGGCTGACC ACCCTTGCCG
13301  AGGTGATCAA CTACACCAAA GCGGGCGGCG GCTGCACCTC TTGCCACGAA
13351  AAAATCGAGC TGGCCCTGGC GGAGATCCTC GCCCAGCAGC CGCAGACGAC
13401  GCCAGCCGTG GCCAGCGGCA AAGATCCGCA CTGGCAGAGC GTCGTCGATA
13451  CCATCGCAGA ACTGCGGCCG CATATTCAGG CCGACGGCGG CGATATGGCG
13501  CTACTCAGCG TCACCAACCA CCAGGTGACC GTCAGCCTCT CCGGCAGCTG
13551  TAGCGGCTGC ATGATGACCG ATATGACCCT GGCCTGGCTG CAGCAAAAAC
13601  TGATGGAACG TACCGGCTGT TATATGGAAG TGGTGGCGGC CTGAGCCGCG
13651  GTTAACTGAC CCAAGGGGGA CAAGATGAAA CAGGTTTATC TCGATAACAA
13701  CGCCACCACC CGTCTGGACC CGATGGTCCT GGAAGCGATG ATGCCCTTTT
13751  TGACCGATTT TTACGGCAAC CCCTCGTCGA TACACGATTT TGGCATTCCG
13801  GCCCAGGCGG CTCTGGAACG CGCGCATCAG CAGGCTGCGG CGCTGCTGGG
13851  CGCGGAGTAT CCCAGCGAGA TCATCTTTAC CTCCTGCGCC ACCGAAGCCA
13901  CCGCCACCGC CATCGCCTCG GCGATCGCCC TGCTGCCTGA GCGTCGCGAA
13951  ATCATCACCA GCGTGGTCGA ACATCCGGCG ACGCTGGCGG CCTGCGAGCA
14001  CCTGGAGCGC CAGGGCTACC GGATTCATCG CATCGCGGTG GATAGCGAGG
14051  GGGCGCTGGA CATGGCGCAG TTCCGCGCGG CGCTCAGCCC GCGCGTCGCG
14101  TTGGTCAGCG TGATGTGGGC GAATAACGAA ACCGGGGTGC TTTTCCCGAT
14151  CGGCGAAATG GCGGAGCTGG CCCATGAACA AGGGGCGCTG TTTCACTGCG
14201  ATGCGGTGCA GGTGGTCGGG AAAATACCGA TCGCCGTGGG CCAGACCCGC
14251  ATCGATATGC TCTCCTGCTC GGCGCATAAG TTCCACGGGC CAAAAGGCGT
14301  AGGCTGTCTT TATCTGCGGC GGGGAACGCG CTTTCGCCCG CTGCTGCGCG
14351  GCGGTCACCA GGAGTACGGT CGGCGAGCCG GGACAGAAAA TATCTGCGGA
14401  ATCGTCGGCA TGGGCGCGGC CTGCGAGCTG GCGAATATTC ATCTGCCGGG
14451  AATGACGCAT ATCGGCCAAT TGCGCAACAG GCTGGAGCAT CGCCTGCTGG
14501  CCAGCGTGCC GTCGGTCATG GTGATGGGCG GCGGCCAGCC GCGGGTGCCC
14551  GGCACGGTGA ATCTGGCCTT TGAGTTTATT GAAGGTGAAG CCATTCTGCT
14601  GCTGTTAAAC CAGGCCGGGA TCGCCGCCTC CAGCGGCAGC GCCTGCACCT
14651  CAGGCTCGCT GGAACCCTCC CACGTGATGC GGGCGATGAA TATCCCCTAC
14701  ACCGCCGCCC ACGGCACCAT CCGCTTTTCT CTCTCGCGCT ACACCCGGGA
14751  GAAAGAGATC GATTACGTCG TCGCCACGCT GCCGCCGATT ATCGACCGGC
14801  TGCGCGCGCT GTCGCCCTAC TGGCAGAACG GCAAGCCGCG CCCGGCGGAC
14851  GCCGTATTCA CGCCGGTTTA CGGCTAAGGC GGAGGTGGCT GATGGAACGC
14901  GTGCTGATTA ACGATACCAC CCTGCGCGAC GGCGAGCAGA GCCCCGGCGT
14951  CGCCTTTCGC ACCAGCGAAA AGGTCGCCAT TGCCGAGGCG CTTTACGCCG
```

15001 CAGGAATAAC GGCGATGGAG GTCGGCACCC CGGCGATGGG CGACGAGGAG
15051 ATCGCGCGGA TCCAGCTGGT GCGTCGCCAG CTGCCCGACG CGACCCTGAT
15101 GACCTGGTGT CGGATGAACG CGCTGGAGAT CCGCCAGAGC GCCGATCTGG
15151 GCATCGACTG GGTGGATATC TCGATTCCGG CTTCGGATAA GCTGCGGCAG
15201 TACAAACTGC GCGAGCCGCT GGCGGTGCTG CTGGAGCGGC TGGCGATGTT
15251 TATCCATCTT GCGCATACCC TCGGCCTGAA GGTATGCATC GGCTGCGAGG
15301 ACGCCTCGCG GGCCAGCGGC CAGACCCTGC GCGCTATCGC CGAGGTCGCG
15351 CAGCAATGCG CCGCCGCCCG CCTGCGCTAT GCCGATACGG TCGGCCTGCT
15401 CGACCCTTTT ACCACCGCGG CGCAAATCTC GGCCCTGCGC GACGTCTGGT
15451 CCGGCGAAAT CGAAATGCAT GCCCATAACG ATCTGGGTAT GGCGACCGCC
15501 AATACGCTGG CGGCGGTAAG CGCCGGGGCC ACCAGCGTGA ATACGACGGT
15551 CCTCGGTCTC GGCGAGCGGG CGGGCAACGC GGCCGCATGG AAACCGTCCG
15601 CGCTGGGCCT TGAACGCTGC CTGGGCGTGG AGACCGGCGT GCATTTTTCG
15651 GCGCTGCCCG CGCTCTGTCA GAGGGTCGCG GAAGCCGCGC AGCGCGCCAT
15701 CGACCCGCAG CAGCCGCTGG TCGGCGAGCT GGTGTTTACC CATGAGTCAG
15751 GTGTCCACGT GGCGGCGCTG CTGCGCGACA GCGAGAGCTA CCAGTCCATC
15801 GCCCCTTCCC TGATGGGCCG CAGCTACCGG CTGGTGCTGG GCAAACACTC
15851 CGGGCGTCAG GCGGTCAACG GCGTTTTTGA CCAGATGGGC TATCACCTCA
15901 ACGCCGCGCA GATTAACCAG CTGCTGCCCG CCATCCGCCG CTTCGCCGAG
15951 AACTGGAAGC GCAGCCCGAA AGATTACGAG CTGGTGGCTA TCTACGACGA
16001 GCTGTGCGGT GAATCCGCTC TGCGGGCGAG GGGGTAATGA TGGAGTGGTT
16051 TTATCAAATT CCCGGCGTGG ACGAACTTCG CTCCGCCGAA TCTTTTTTTC
16101 AGTTTTTCGC CGTCCCCTAT CAGCCCGAGC TGCTTGGCCG CTGCAGCCTG
16151 CCGGTGCTGG CAACGTTTCA TCGCAAACTC CGCGCGGAGG TGCCGCTGCA
16201 AAACCGGCTC GAGGATAACG ACCGCGCGCC CTGGCTGCTG GCGCGAAGAC
16251 TGCTCGCGGA GAGCTATCAG CAACAGTTTC AGGAGAGCGG AACATGAGAC
16301 CGAAATTCAC CTTTAGCGAA GAGGTCCGCG TCGTACGCGC GATTCGTAAC
16351 GACGGCACCG TGGCGGGCTT CGCGCCCGGC GCGCTGCTGG TCAGGCGCGG
16401 CAGCACCGGC TTTGTGCGCG ACTGGGGCGT TTTTTTGCAA GATCAGATTA
16451 TCTACCAGAT CCACTTTCCG GAAACCGATC GGATCATCGG CTGCCGCGAG
16501 CAGGAGCTGA TCCCCATCAC CCAGCCGTGG CTGGCCGGAA ATTTGCAATA
16551 CAGGGATAGC GTGACCTGCC AGATGGCGCT CGCGGTCAAC GGCGATGTGG
16601 TCGTGAGCGC CGGCCAGCGG GGACGCGTTG AGGCTACCGA TCGGGGAGAG
16651 CTCGGCGACA GCTACACCGT CGACTTTAGC GGCCGCTGGT TCAGGGTCCC
16701 GGTGCAGGCC ATCGCCCTTA TAGAGGAAAG AGAAGAATGA ACCCATGGCA
16751 ACGTTTTGCC CGGCAGCGGC TGGCGCGCAG CCGCTGGAAT CGCGATCCGG
16801 CGGCCCTGGA TCCGGCCGAT ACGCCGGCTT TTGAACAGGC CTGGCAACGC
16851 CAGTGCCATA TGGAGCAGAC GATCGTCGCG CGGGTCCCTG AAGGCGATAT
16901 TCCGGCGGCG TTGCTGGAGA ATATCGCTGC CTCCCTTGCC ATCTGGCTCG
16951 ACGAGGGGGA TTTTGCGCCG CCCGAGCGCG CTGCCATCGT GCGCCATCAC
17001 GCCCGGCTGG AACTCGCCTT CGCCGATATC GCCCGCCAGG CGCCGCAGCC
17051 GGATCTCTCC ACGGTACAGG CATGGTATCT GCGCCACCAG ACGCAGTTTA
17101 TGCGCCCGGA ACAGCGTCTG ACCCGCCATT TACTGCTGAC GGTCGATAAC
17151 GACCGCGAAG CCGTGCACCA GCGGATCCTC GGCCTGTATC GGCAAATCAA
17201 CGCCTCGCGG GACGCTTTCG CGCCGCTGGC CCAGCGCCAT TCCCACTGCC
17251 CGAGCGCGCT GGAAGAGGGT CGTTTAGGCT GGATTAGCCG TGGCCTGCTC
17301 TATCCGCAGC TCGAGACCGC GCTGTTTTCA CTGGCGGAAA ACGCGCTAAG
17351 CCTTCCCATC GCCAGCGAAC TGGGCTGGCA TCTTTTATGG TGCGAAGCGA
17401 TTCGCCCCGC CGCGCCCATG GAGCCGCAGC AGGCGCTGGA GAGCGCGCGC
17451 GATTATCTTT GGCAGCAGAG CCAGCAGCGC CATCAGCGCC AGTGGCTGGA
17501 ACAGATGATT TCCCGTCAGC CGGGACTGTG CGGGTAGCCT CGGCGGCTAC
17551 CCGTTAACGC CTACAGCACG GTGCGTTTAA TCTCCTCAAG CCAGCTCGCC
17601 AGACGCGCTT CGGTCTGGTC GAACTGGTTA TCCTGATCCA GCACCAGCCC
17651 AACAAAGCGG TCGCCTTCCA GCGCCGAGGA CGCGCTGAAT TCATAACCCT

```
17701  CATTTGGCCA GCTGCCAATC ATCTGCGCGC CGCGCGCGCT CAGGGCGTCG
17751  AACAGCGGGC GCATCCCGCT GACGAAGTTG TCCGGATAGC CTCTCTGATC
17801  GCCGAGGCCG AACAGCGCCA CGGTTTTCCC TTTCAGGCTG GCGTCGTCGA
17851  GGCCGCTGAT AAATTCGCTC CATGACTCGC TTTCGCATCC GGCCTCCAGC
17901  CCCGGCAGCT GGCCGTCGCC GAGCGTCGGC GTGCCCAGCA GCAGCACCGG
17951  ATAGGCCATA AAGTCGTCCA GCGTCGTGCG GTTAATGTTG ACCGGGGCAT
18001  CCGCCAGCTC GCCCAGTTGC TTATGGATCA TTTTCGCGAT TTTGCGGGTT
18051  TTACCGGTAT CGGTGCCAAA GAAAATACCA ATGTTCGCCA TGTTGCGCTC
18101  CTGTCGGAAA AGGGGGTTGA AAATACGCGT TCTCGCAGGG GTATTGCGAA
18151  GGCTGTGCCA GGTTGCTTTG CACTACCGCG GCCCATCCCT GCCCCAAAAC
18201  GATCGCTTCA GCCCTCTCCC GCCGCGCGCG GCGGGGCTGG CGGGGCGCTT
18251  AAAATGCAAA AAGCGCCTGC TTTTCCCCTA CCGGATCAAT GTTTCTGCAC
18301  ATCACGCCGA TAAGGGCGCA CGGTTTGCAT GGTTATCACC GTTCGGAAAA
18351  CACCGCGGCG TCCCTGTCAC GGTGTCGGAC AAATTGTCAT AACTGCGACA
18401  CAGGAGTTTG CGATGACCCT GAATATGATG CTCGATAACG CCGTACCCGA
18451  GGCGATTGCC GGTGCGCTGA CTCAACAACA TCCGGGGCTG TTTTTTACAA
18501  TGGTCGAACA GGCATCGGTA GCGATTTCCC TCACCGATGC CCGGGCGAAT
18551  ATTATCTACG CCAACCCGGC GTTTTGCCGC CAGACTGGAT ACTCGCTGGC
18601  GCAATTGCTC AATCAAAACC CGCGCCTGCT GGCCAGCAGC CAGACGCCGC
18651  GCGAGATCTA CCAGGAGATG TGGCAAACCC TGCTCCAGCG CCAGCCGTGG
18701  CGCGGTCAGC TAATTAATCA GCGCCGCGAC GGCGGCCTGT ATCTGGTAGA
18751  TATCGATATC ACGCCGGTGC TGAATCCGCA GGGCGAGCTG GAGCATTATC
18801  TGGCGATGCA GCGGGATATC AGCGTCAGCT ATACCCTGGA ACAGCGGCTG
18851  CGCAATCATA TGACGCTAAT GGAAGCGGTG CTCAATAACA TCCCCGCCGC
18901  CGTGGTCGTG GTCGATGAGC AGGATCGGGT GGTGATGGAT AATCTCGCCT
18951  ACAAAACGTT CTGCGCGGAC TGCGGCGGGA AAGAGCTGCT GGTCGAGCTC
19001  CAGGTTTCCC CGCGCAAAAT GGGGCCCGGC GCGGAGCAAA TCCTGCCGGT
19051  GGTGGTTCGC GGCGCGGTCC GCTGGCTGTC GGTAACCTGC TGGGCGCTGC
19101  CCGGCGTGAG TGAAGAAGCC AGCCGCTACT TCGTCGACAG CGCCCCGGCG
19151  CGCACGCTGA TGGTGATCGC CGACTGTACC CAGCAGCGCC AGCAGCAGGA
19201  GCAGGGCCGG CTCGACCGTC TGAAACAGCA AATGACCGCC GGTAAGCTGC
19251  TGGCCGCGAT TCGCGAGTCG CTGGACGCGG CGCTGATTCA GCTTAATTGC
19301  CCAATCAATA TGCTGGCGGC GGCCCGCCGG CTGAACGGCG AAGGCAGCGG
19351  CAACGTGGCG CTGGACGCGG CGTGGCGCGA AGGTGAAGAG GCCATGGCGC
19401  GCCTGCAGCG CTGCCGCCCT TCTCTTGAGC TGGAAAGCAA TGCCGTCTGG
19451  CCGCTTCAGC CCTTTTTTGA CGACCTGTAC GCCCTCTACC GCACCCGCTT
19501  TGACGATCGC GCGCGGCTGC AGGTGGACAT GGCATCGCCG CATCTGGTCG
19551  GCTTCGGCCA GCGTACCCAG CTGCTGGCCT GCTTGAGTTT ATGGCTCGAC
19601  CGGACGCTGG CCCTCGCCGC CGAGCTGCCC TCCGTACCGC TGGAGATCGA
19651  GCTTTACGCC GAAGAGGACG AGGGCTGGCT CTCTTTGTAT CTCAACGACA
19701  ATGTCCCGCT GCTGCAGGTG CGCTACGCCC ACTCCCCCGA TGCCCTAAAC
19751  TCTCCCGGCA AAGGGATGGA GCTGCGGCTG ATCCAAACGC TGGTCGCCTA
19801  CCACCGCGGC GCGATTGAAC TGGCTTCGCG ACCGCAGGGA GGCACCAGCC
19851  TGGTTCTGCG TTTCCCGCTC TTTAATACCC TGACCGGAGG TGAGCAATGA
19901  TCCATAAATC CGATTCGGAC ACCACCGTCA GACGTTTCGA TCTCTCCCAG
19951  CAGTTTACCG CCATGCAGCG GATAAGCGTG GTCCTGAGTC GCGCCACCGA
20001  AGCGAGCAAA ACCCTGCAGG AGGTTCTGAG CGTGCTACAT AACGATGCCT
20051  TTATGCAGCA CGGGATGATT TGCCTGTACG ACAGCCAGCA GGAGATCCTG
20101  AGCATCGAAG CGCTGCAGCA AACGGAAGAT CAGACGCTGC CCGGCAGTAC
20151  GCAAATTCGC TACCGGCCGG GGGAAGGATT AGTCGGTACC GTGCTGGCGC
20201  AGGGCCAGTC GCTGGTGCTG CCGCGCGTCG CCGACGACCA GCGTTTTCTC
20251  GATCGTCTGA GCCTGTACGA CTATGACCTG CCGTTTATCG CCGTTCCGCT
20301  GATGGGCCCC CACTCCCGGC CCATCGGCGT ACTGGCGGCG CACGCGATGG
20351  CGCGTCAGGA AGAGCGGCTG CCCGCCTGCA CGCGCTTTCT CGAAACCGTC
```

```
20401  GCCAATCTGA TCGCCCAGAC GATTCGCCTG ATGATCCTGC CAACCTCCGC
20451  CGCGCAGGCG CCGCAGCAGA GCCCCAGAAT AGAGCGCCCG CGCGCCTGTA
20501  CCCCTTCGCG CGGTTTCGGC CTGGAAAATA TGGTCGGTAA AAGCCCGGCG
20551  ATGCGGCAGA TTATGGATAT TATTCGTCAG GTTTCCCGCT GGGATACCAC
20601  GGTGCTGGTA CGCGGCGAGA GCGGCACCGG GAAAGAGCTC ATCGCCAACG
20651  CCATCCACCA TAATTCTCCG CGCGCCGCCG CGGCGTTCGT CAAATTTAAC
20701  TGCGCGGCGC TGCCGGACAA CCTGCTGGAG AGCGAGCTGT TTGGTCATGA
20751  GAAAGGCGCG TTTACCGGCG CGGTGCGCCA GCGGAAAGGC CGCTTTGAGC
20801  TGGCGGACGG CGGCACCTTA TTCCTCGATG AGATCGGCGA AAGCAGCGCC
20851  TCGTTTCAGG CTAAGCTACT GCGTATTCTG CAAGAGGGGG AGATGGAGCG
20901  CGTCGGCGGC GACGAAACCC TGCGGGTCAA CGTGCGCATT ATCGCGGCGA
20951  CCAACCGCCA TCTGGAAGAG GAGGTGCGGC TGGGTCATTT CCGCGAGGAT
21001  CTATACTACC GCCTGAACGT AATGCCTATC GCGCTGCCGC CGCTGCGCGA
21051  GCGCCAGGAG GATATCGCCG AGCTGGCGCA CTTTCTGGTG CGAAAAATCG
21101  CCCACAGCCA GGGGCGAACG CTGCGCATCA GCGATGGGGC GATTCGCCTG
21151  CTGATGGAGT ACAGCTGGCC GGGAAACGTG CGCGAACTGG AAAACTGTCT
21201  CGAACGTTCG GCGGTGCTGT CGGAAAGCGG CCTGATAGAC CGGGACGTGA
21251  TTCTGTTCAA CCATCGCGAT AACCCGCCGA AAGCGCTCGC CAGCAGCGGC
21301  CCGGCGGAGG ACGGCTGGCT CGATAACAGC CTCGACGAGC GCCAGCGGCT
21351  GATCGCCGCC CTGGAAAAAG CGGGCTGGGT GCAGGCCAAA GCGGCGCGGC
21401  TGCTCGGCAT GACCCCGCGC CAGGTGGCGT ATCGCATTCA GATTATGGAT
21451  ATCACCATGC CGCGACTGTG AAGCCTTATG TGAGATTCAG GACATTGTCG
21501  CCAGCGCGGC GGAATTGCGA CAATTCAGGG ACGCGGGTTG CCGGTTAAAA
21551  AGTCTACTTT TCATGCGGTT GCGAAATTAA CCTCTGGTAC AGCATTTGCA
21601  GCAGGAAGGT ATCGCCCAAC CACGAAGGTA CGACCATGAC TTCCTGCTCC
21651  TCTTTTTCTG GCGGCAAAGC CTGCCGCCCG GCGGATGACA GCGCATTGAC
21701  GCCGCTTGTG GCCGATAAAG CTGCCGCGCA CCCCTGCTAC TCTCGCCATG
21751  GGCATCACCG TTTCGCGCGG ATGCATCTGC CCGTCGCGCC CGCCTGCAAT
21801  TTGCAGTGCA ACTACTGTAA TCGCAAATTC GATTGCAGCA ACGAGTCCCG
21851  CCCCGGGGTA TCGTCAACGC TGCTGACGCC TGAACAGGCG GTCGTGAAAG
21901  TGCGTCAGGT CGCGCAGGCG ATCCCGCAGC TTTCGGTGGT GGGCATCGCC
21951  GGGCCCGGCG ATCCGCTCGC CAATATCGCC CGCACCTTTC GCACCCTGGA
22001  GCTGATCCGC GAACAGCTGC CGGACCTGAA ATTATGCCTG TCGACCAACG
22051  GACTGGTGCT GCCTGACGCG GTGGACCGCC TGCTGGATGT CGGCGTTGAC
22101  CACGTCACGG TCACCATTAA CACCCTCGAC GCGGAGATTG CCGCGCAAAT
22151  CTACGCCTGG CTATGGCTGG ACGGCGAACG CTACAGCGGG CGCGAAGCGG
22201  GAGAGATCCT GATTGCCCGT CAGCTTGAGG GCGTACGCAG GCTGACCGCC
22251  AAAGGCGTGC TGGTGAAAAT AAATTCGGTG CTGATCCCCG GTATCAACGA
22301  TAGCGGCATG GCCGGCGTGA GCCGCGCGCT GCGGGCCAGC GGCGCGTTTA
22351  TCCATAATAT TATGCCGCTG ATCGCCAGGC CGGAGCACGG CACGGTGTTT
22401  GGCCTCAACG GCCAGCCGGA GCCGGACGCC GAGACGCTCG CCGCCACCCG
22451  CAGCCGGTGC GGCGAAGTGA TGCCGCAGAT GACCCACTGC CACCAGTGTC
22501  GCGCCGACGC CATTGGGATG CTCGGCGAAG ACCGCAGCCA GCAGTTTACC
22551  CAGCTTCCGG CGCCAGAGAG TCTCCCGGCC TGGCTGCCGA TCCTCCACCA
22601  GCGCGCGCAG CTGCACGCCA GCATTGCGAC CCGCGGCGAA TCTGAAGCCG
22651  ATGACGCCTG CCTGGTCGCC GTGGCGTCAA GCCGCGGGGA CGTCATTGAT
22701  TGTCACTTTG GTCACGCCGA CCGGTTCTAC ATTTACAGCC TCTCGGCCGC
22751  CGGTATGGTG CTGGTCAACG AGCGCTTTAC GCCCAAATAT TGTCAGGGGC
22801  GCGATGACTG CGAGCCGCAG GATAACGCAG CCCGGTTTGC GGCGATCCTC
22851  GAACTGCTGG CGGACGTTAA AGCCGTATTC TGCGTGCGTA TCGGCCATAC
22901  GCCGTGGCAA CAGCTGGAAC AGGAAGGCAT TGAACCCTGC GTTGACGGCG
22951  CGTGGCGGCC GGTCTCCGAA GTGCTGCCCG CGTGGTGGCA ACAGCGTCGG
23001  GGGAGCTGGC CTGCCGCGTT GCCGCATAAG GGGGTCGCCT GATGCCGCCG
23051  CTCGACTGGT TGCGGCGCTT ATGGCTGCTG TACCACGCGG GGAAAGGCAG
```

```
23101  CTTTCCGCTG CGCATGGGGC TTAGCCCGCG CGATTGGCAG GCGCTGCGGC
23151  GGCGCCTGGG CGAGGTGGAA ACGCCGCTCG ACGGCGAGAC GCTCACCCGT
23201  CGCCGCCTGA TGGCGGAGCT CAACGCCACC CGCGAAGAGG AGCGCCAGCA
23251  GCTGGGCGCC TGGCTGGCGG GCTGGATGCA GCAGGATGCC GGGCCGATGG
23301  CGCAGATTAT CGCCGAGGTT TCGCTGGCGT TTAACCATCT CTGGCAGGAT
23351  CTTGGTCTGG CATCGCGCGC CGAATTGCGC CTGCTGATGA GCGACTGCTT
23401  TCCACAGCTG GTGGTGATGA ACGAACACAA TATGCGCTGG AAAAAGTTCT
23451  TTTATCGTCA GCGCTGTTTG CTGCAACAGG GGGAAGTTAT CTGCCGTTCG
23501  CCAAGCTGCG ACGAGTGCTG GGAACGCAGC GCCTGTTTTG AGTAGCCGTT
23551  TCCCGAAGGG GGCGCTGCAA ACAAAAAAGC CGGAGGTTTC CCTCCGGCTT
23601  TTCACATCAT CAAATGTGAT TATGCGACGT CTTCGTACTG CGGCACCGGG
23651  TTGCGGAAGC TTTTGGTCAC GCAGGCCTCC GTAGACCAGA CCAATACCGC
23701  CCCAGATCAG GCCGAGAACC ATGGAGCTCT CTTCGAGGTT AATCCACAGT
23751  GCGCCGACGG TCAGCGCGCC GCAGACCGGC AGAATCAGAT AGTTGAAGTG
23801  GTCTTTCAGC GTTTTGTTGC GCTTTTCACG GATCCAGAAC TGGGAGATCA
23851  CCGACAGGTT AACGAAGGTG AACGCCACCA GCGCGCCGAG GTTAATCGGC
23901  GCCGTCGCCG TGACGAGGTC GAGTTTAATC GCCAGCAGCG CGATCGCGCA
23951  ACCAGCAGCA CGTTCCATGC CGGAGTACGC CGTTTCGGAT GCACGTAGCC
24001  GAAGAAACGC GTCGGGAACA CGCCGTCGCG GCCCATCACG TACATCAGAC
24051  GGGAAACGCC CGCGTGCGCG GCCGTGCCGG ATGCCAGTAC GGTAACGCTG
24101  GAGAAAATCA GCACGCCCCA CTGGAAGGTT TTGCCCGCCA CGTACAGCAT
24151  GATTTCAGGC TGCGAGGCGT CCGGATCTTT GAAGCGCGAG ATGTCCGGGA
24201  AGTACAGCTG CAG
```

#	Comments
2559	gc, or cg
3369	c, or g
3394	g, or a
3982	aaagcgccaccatggcccc (our data)
4979	c (our data)
5107	g, or c
5189	a, or g
5195	a, or g
5334	ct, or tc
5361	a, or g
6139	g, or c
6142	c, or a
6751	ga, or ag
7373	t, or c
7830	g, or a
7832	a, or g
8138	c, or t
8170	c, or g
8172	g, or c
8921	t, or c
9034	c may not be there
9344	t is confirmed
9502	g, or c
9700	ac, or ct

9894	t is correct
10395	g is correct
12544	ca, or ac
14044	a is correct
15354	c is correct (our data & A. Pühler's)
13649	cg, or gc
13662	c, or g
13665	g, or c
13887	c may not be there
13892	c may not be there
13895	a may not be there
13901	c may not be there
13907	c may not be there
13914	c may not be there
13916	c may not be there
13919	c may not be there
13920	g may not be there
14002	c, or a
14011	c, or g
14040	g, or a
14254	g, or c
1428	gg, or cc
14523	g, or c
14542	cg, or gc
15196	g, or c
15363-4	most likely no g in between
15584	may not be there
15587	may not be there
15599	may not be there
15663	ct is correct
15776	cg, or gc
18159	c may not be there
18216,7,8,9	ctcc may not be there
18223	c may not be there
18227	c may not be there
18232,3,4	may not be there
18555	t is correct
18567	may not be there
18583	g, or a
18722	c is correct
21617-18	there may be a g inserted
22056	g, or a
22198	cg, or gc
22315	g, or a
22401	g, or c
22406	c, or g
23349-50	there may be an a inserted
23367	g may not be there
23578	a is correct

gene	from	to
nifJ	3775	260
nifH	4180	5061
nifD	5075	6526
nifK	6582	8144
nifT	8184	8402
nifY	8413	9075
nifE	9398	10771
nifN	10782	12167
nifX	12154	12624
nifU	12820	13644
nifS	13675	14877
nifV	14892	16037
nifW	16037	16297
nifZ	16294	16740
nifM	16737	17537
nifF	18091	17561
nifL	18413	19900
nifA	19897	21471
nifB	21636	23042
nifQ	23042	23545

genes without problems: H, T, Y, N, W, Z, M, F, A
genes with minor problems: J, D, E, X, U, V, L, B, Q
genes with major problems: K, S

References

Amann, E., Brosius, J., & Ptashne, M. *Gene* 1983, **25**, 167-178.

Amersham Corporation, *Manual for Oligonucleotide-directed in vitro Mutagenesis System*, 1986.

Anderson, D.H. & Rodwell, V.W. *J. Bacteriol.* 1989, **171**, 6468-6472.

Anderson, G.L. & Howard, J.B. *Biochem.* 1984, **23**, 2118-2122.

Applied Biosystems, Tech sheet #400771, 1989.

Arnold, W., Rump, A., Klipp, W., Priefer, U.B., & Pühler, A. *J. Mol. Biol.* 1988, **203**, 715-738.

Ausubel, F.M. *Cell* 1983, **37**, 5-6.

Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A., & Struhl, K. *Current Protocols in Molecular Biology,* 1987, Greene Publishing Associates & Wiley-Interscience, New York, New York.

Bachhuber, M., Brill, W.J., & Howe, M.M. *J. Bacteriol.* 1976, **128**, 749-753.

Bankier, A.T. & Barrell, B.G. in *Techniques in Nucleic Acid Biochemstry: Shotgun DNA Sequencing* 1983, R.A. Flavell Ed., Elsevier Scientific Publishing Co., Inc., New York, New York.

Beach, M.J. & Rodwell, V.W. *J. Bacteriol.* 1989, **171**, 2994-3001.

Bensadoun, A. & Weinstein, D. *Anal. Biochem.* 1976, **70**, 241-250.

Berman, J., Gershoni, J.M., & Zamir, A. *J. Biol. Chem.* 1985a, **260**, 5240-5243.

Berman, J., Zilberstein, A., Salomon, D., & Zamir, *A. Gene* 1985b, **35**, 1-9.

Beynon, J., Cannon, M., Buchanan-Wollaston, V., & Cannon, F. *Cell* 1983, **34**, 665-671.

BIO101 Inc., *Manual for Glassmilk*™, 1986.

Biology Department at MIT, Cambridge, MA, Anonymous Protocol, 1989.

Bishop, P.E., Jarlenski, M.L., & Hetherington, *D.R. Proc. Natl. Acac. Sci.* 1980, **77**, 7342-7346.

Boehringer Mannheim Inc. Tech. Sheet 8711132/1M, Mannheim, FRG, 1988.

Bolin, D.R., Sytwo, I.-I., Humiec, F., & Meienhoffer, J. *Int. J. Pept. Protein Res.* 1989, **33**, 353-359.

Bolin, J. Presentation at the 8th International Congress on Nitrogen Fixation, May 1990, Knoxville, TN.

Brooks, S.J., Collins, J.J., & Brill, W.J. *J. Bact.* 1984, **157**, 460-464.

Brunk, C.F., Jones, K.C., & James, T.W. *Anal. Biochem.* 1979, **92**, 497-500.

Buck, M, Miller, S., Drummond, M., & Dixon, R. *Nature* 1986, **320**, 374-378.

Carpino, L.A. & Han, G.Y. *J. Org. Chem.* 1972, **37**, 3404-3409.

C&E News, 1990, April 9 Issue, 11-14.

Chisnell, J.R., Premakumar, R., & Bishop, P.E. *J. Bacteriol.* 1988, **170**, 27-33.

Cole, S.T. & Raibaud, O. *Gene* 1986, **42**, 201-208.

Coucouvanis, D. & Kanatzidis, M.G. *J. Am. Chem. Soc.* 1985, **107**, 5005-5006.

Davis, L.C., Henzl, M.T., Burris, R.H., & Orme-Johnson, W.H. *Biochemistry* 1979, **18**, 4860-4869.

Del Sal, G., Manfioletti, G., & Schneider, C. *Nucl. Ac. Res.* 1988, **16**, 9878.

Dixon, R.O.D. & Wheeler, C.T. *Nitrogen Fixation in Plants*, 1986, Chapman & Hall Publ. New York.

Doolittle, R.F. *Of URF's & ORF's* 1986, University Science Books.

Eady, R.R. in *Methods for Evaluating Biological Nitrogen Fixation*, F.H. Bergersen Ed. 1980. pp. 213-264. J. Wiley & Sons, Ltd.

Earl, C. Ph.D. Thesis, Harvard University, 1985.

Elmerich, C., Houmard, J., Sibold, L., Manheimer, I., & Charpin, N. *Molec. Gen. Gen.* 1978, **165**, 181-189.

Epicentre Technologies, *Manual for HK™ Phosphatase* 1990, Madison, WI 53713.

Erbes, D., & Orme-Johnson, W.H. Unpublished Observations, 1975.

Ferretti, L. & Sgaramella, V. *Nucl. Ac. Res.* 1981, **9**, 85-93.

Filler, W.A., Kemp, R.M., NG, J.C., Hawkes, T.R., Dixon, R.A., & Smith, B.E. *Eur. J. Biochem.* 1986, **160**, 371-377.

Fournier, A., Wang, C. & Felix, A. *Int. J. Peptide Res.* 1988, **31**, 86-97.

Georgiadis, M.M., Chakrabarti, P, & Rees, D.C. in *Nitrogen Fixation: Achievements and Objectives*, Gresshoff, P.M., Roth, E., Stacy, G., & Newton, W.G., Eds., Chapman & Hall, New York, in press 1990.

Gosink, M.M., Franklin, N.M., Roberts, G.P. *J. Bact.* 1990, **172**, 1441-1447.

Gray, G.S. & Bhattacharjee, J.K. *Can. J. Microbiol.* 1976, **22**, 1664-1667.

Greenwood, N.N. & Earnshaw, A. *Chemistry of the Elements*, 1984, pp. 466-473, Pergamon Press, New York.

Gribskov, M., Devereux, J., & Burgess, R.R. *Nuc. Ac. Res.* 1984, **12**, 539-549.

Gussin, G.N., Ronson, C.W., & Ausubel, F.M. *Ann. Rev. Gen.* 1986, **20**, 567-592.

Hageman, R.V., Orme-Johnson, W.H., & Burris, R.H. *Biochem.* 1980, **19**, 2333-2342.

Hamaguchi, K. & Geiduschek, E.P. *J. Am. Chem. Soc.* 1962, **84**, 1329-1338.

Hanahan, D. *J. Mol. Biol.* 1983, **166**, 557-580.

Harris, G.S., White, T.C., Flory, J.E., & Orme-Johnson, W.H., in press *J. Biological Chemistry*, 1990.

Hausinger, R.P., & Howard, J.B. *J. Biol. Chem.* 1983, **258**, 13486-13492.

Hawkes, T.R., McLean, P.A., & Smith, B.E. *Biochem. J.* 1984, **217**, 317-321.

Hill, S. *J. Gen. Microbiol.* 1976, **93**, 335-345.

Hill, S., Turner, G.L., & Bergersen, F.J. *J. Gen. Microbiol.* 1984, **130**, 1061-1067.

Holland, D., Zilberstein, A., Govezensky, D., Salomon, S., & Zamir, A. *J. Biol. Chem.* 1987, **262**, 8814-8820.

Holleman, A.F. & Wiberg, E. *Lehrbuch der Anorganischen Chemie*, 1976, pp. 363-375, Walter de Gruyter Publ., New York.

Hoover, T.R., Imperial, J., Ludden, P.W., & Shah, V.K. *J. Bacteriol.* 1988a, **170**, 1978-1989.

Hoover, T.R., Imperial, J., Liang, J., Luddern, P.W., & Shah, V.K. *Biochemistry*, 1988b, **27**, 3647-3652.

Hoover, T.R., Imperial, J., Ludden, P.W., & Shah, V.K. *Biochem.* 1989, **28**, 2768-2771.

Hoover, T.R., Robertson, A.D., Cerny, R.L., Hayes, R.N., Imperial, J., Shah, V.K., & Ludden, P.W. *Nature* 1987, **329**, 855-857.

Howard, J.B., Davis, R., Moldenhauer, B., Cash, V.L., & Dean, D. *J. Biol. Chem.* 1989, **264**, 11270-11274.

Howard, K.S., McLean, P.A., Hansen, F.B., Lemley, P.V., Koblan, K.S., & Orme-Johnson, W.H. *J. Biol. Chem.* 1986, **261**, 772-778.

Hudson, D. *J. Org. Chem.* 1988, **53**, 617-624.

Imperial, J., Hoover, T.R., Madden, M.S., Ludden, P.L., & Shah, V.K. *Biochem.* 1989, **28**, 7796-7799.

Jacobson, M.R., Brigle, K.E., Bennett, L.T., Setterquist, R.A., Wilson, M.S., Cash, V.L., Beynon, J., Newton, W.E., Dean, D.R. *J. Bacteriol.* 1989a, **171**, 1017-1027.

Jacobson, M.R., Cash, V.L., Weiss, M.C., Laird, N.F., Newton, W.E., & Dean, D.R. *Mol. Gen. Genet.* 1989b, **219**, 49-57.

Jameson, B.A. & Wolf, H. *CABIOS* 1988, **4**, 181-186.

Kohlhaw, G., Leary, T.R., & Umbarger, H.E. *J. Biol. Chem.* 1969, **244**, 2218-2225.

Kong, Q., Wu, Q,, Ma, Z., & Shen, S. *J. Bact.* 1986, **166**, 353-356.

Kyte, J. & Doolittle, R.F. *J. Mol. Biol.* 1982, **157**, 105-132.

Labarca, C. & Paigen, K *Anal. Biochem.* 1989, **102**, 344-352.

Laemmli, U.K. Nature 1970, 227, 680-685.

Lindahl, P.A., Day, E.P., Kent, T.A., Orme-Johnson, W.H., & Münck, E. *J. Biol. Chem.* 1985, **260**, 11160-11173.

Lindahl, P.A., Gorelick, N.J., Münck, E., & Orme-Johnson, W.H. *J. Biol. Chem.* 1987a, **262**, 14945-14953.

Lindahl, P.A., Teo, B.-K., & Orme-Johnson, W.H. *Inorg. Chem.* 1987b, **26**, 3912-3916.

Lowry, O.H., Rosegrough, N.J., Farr, A.L., & Randall, R.J. *J. Biol. Chem.* 1951, **193**, 265-275.

Lunn, G. & Sansone, E.B. *Anal. Biochem.* 1987, **162**, 453-458.

MacNeil, T., Brill, W.J., & Howe, M.M. *J. Bacteriol.* 1978a, **134**, 821-829.

MacNeil, T., MacNeil, D., Roberts, G.P., Supiano, M.A., and Brill, W.J. *J. Bacteriol.* 1978b, **136**, 253-266.

Mahl, M.C., Wilson, P.W., Fife, M.A., & Ewing, W.H. *J. Bact.* 1965, **89**, 1482-1487.

Maniatis, T., Fritsch, E.F., & Sambrook, J. *Molecular Cloning: A Laboratory Manual,* 1982, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

Martin, A.E., Burgess, B.K., Stout, C.D., Cash, V.L., Dean, D.R., Jensen, G.M., & Stephens, P.J. *Proc. Nat. Acac. Sc.* 1990, **87**, 598-602.

Matsudaira, P. *J. Biol. Chem.* 1987, **262**, 10035-10038.

McBride, L.J., McCollum, C., Davidson, C., Efcavitch, J.W., Andrus, A., & Lombardi, S.J. *BIOTECHNIQUES* 1988, **6**, 362-367.

McKenna, C.E., Nguyen, H.T., Huang, C.W., & Smith, S.A. *Anal. Biochem.* 1977, **83**, 337-345.

McLean, P.A., & Dixon, R.A. *Nature* 1982, **292**, 655-656.

McLean, P.A., Smith, B.E., & Dixon, R.A. *Biochem. J.* 1983, **211**, 589-597.

McLean, P.A., Wink, D.A., Chapman, S.K., Hickman, A.B., McKillop, D.M., & Orme-Johnson, W.H. *Biochem.* 1989, **28**, 9402-9406.

Merrick, M.J., Gibbins, J.R., & Postgate, J.R. *J. Gen. Microbiol.* 1987, **133**, 2053-2057.

Messing, J., Gronenborn, B., Mueller-Hill, B., & Hofschneider, P.H. *Proc. Nat. Acad. Sc.* 1977, **74**, 3642-3646.

Mizusawa, S., Nishimura, S., & Seela, F. *Nuc. Ac. Res.* 1986, **14**, 1319-1324.

Morelle, G. *FOCUS* 1989, **11**, 7.

Neidhardt, F.C., Bloch, P.L., & Smith, D.F. *J. Bacteriol.* 1974, **119**, 736-747.

Nelson, M.J., Levy, M.A., & Orme-Johnson, W.H. *Proc. Natl. Acad. Sci.* 1983, **80**, 147-150.

Nieva-Gomez, D., Roberts, G.P., Klevickis, S, & Brill, W. J. *Proc. Natl. Acad. Sci.* 1980, **77**, 2555-2558.

O'Farrell, P.H., Kutter, E., & Nakanishi, M. *Mol. Gen. Genet.* 1980, **179**, 421-435.

Orme-Johnson, W.H. *Ann. Rev. Biophys. Biophys. Chem.* 1985, **14**, 419-459.

Orme-Johnson, W.H., Hamilton, W.D., Jones, T.L., Tso, M.-Y. W., Burris, R.H., Shah, V.K., & Brill, W.J. *Proc. Natl. Acad. Sci.* 1972, **69**, 3142-3145.

Pagani, S., Eldridge, M, & Eady, R.R. *Biochem. J.* 1987, **244**, 485-488.

Paul, W.& Merrick, M. *Eur. J. Biochem.* 1987, **170**, 259-265.

Paul, W. & Merrick, M. *Eur. J. Biochem.* 1989, **178**, 675-682.

Paustian, T.D., Shah, V.K., & Roberts, G.P. *Biochem.* 1990, **29**, 3515-3522.

Pharmacia/LKB Biotechnology, Inc., *Manual for NAP-10*, Piscataway, NJ, 1989.

Promega Corp. *Manual for ERASE-A-BASE Kit*, Madison, WI, 1987.

Pühler, A. & Klipp, W. in *Biology of Inorganic Nitrogen and Sulfur* 1982, pp. 276-286, Bothe, H. & Trebst, A. Eds., Springer Verlag, Berlin, FRG.

Punita, Jafri, S., Reddy, M.A., & Das, H.K. *J. Bacteriol.* 1989, **171**, 3133-3138.

RIBI ImmunoChem Research, Inc., Tech Sheet "Reconstitution Instruction", Hamilton, Montana, 1989.

Ricca, E. & Calvo, J.M. *Nucl. Ac. Res.* 1990, **18**, 1290.

Rivier, J., McLintock, R., Galyean, R., & Anderson, H. *J. Chromatography* 1984, **288**, 303-328.

Roberts, G.P. & Brill, W.J. *J. Bact.* 1980, **144**, 210-216.

Roberts, G.P., MacNeil, T., MacNeil, D., & Brill, W.J. *J. Bacteriol.* 1978, **136**, 267-279.

Robinson, A.C., Burgess, B.K., & Dean, D.R. *J. Bacteriol.* 1986, **166**, 180-186.

Robinson, A.C., Chun, T.W., Li, C.-G., & Burgess, B.K. *J. Biol. Chem.* 1989, **264**, 10088-10095.

Robinson, A.C., Dean, D.R., & Burgess, B.K. *J. Biol. Chem.* 1987, **262**, 14327-14332.

Robson, R.L., Chesshyre, J.A., Wheeler, C., Jones, R., Woodley, P.R., & Postgate, J.R. *J. Gen. Microbiol.* 1984, **130**, 1603-1612.

Sanger, F., Nicklen, S., & Coulson, A.R. *Proc. Nat. Acad. Sci.* 1977, **74**, 5463-5476.

Schwarz, E., Oesterhelt, D, Reinke, H., Beyreuther, K., & Dimroth, P. *J. Biol. Chem.* 1988, **263**, 9640-9645.

Scott, K.F., Rolfe, B.G., & Shine, J. *J. Mol. Appl. Gen.* 1981, **1**, 71-81.

Shah, V.K., Imperial, J., Ugalde, R.A., Ludden, P.W., & Brill, W.J. *Proc. Natl. Acac. Sci.* 1986, **83**, 1636-1640.

Smith, M. *Ann. Rev. Genet.* 1985, **19**, 423-462.

Staros, J.V., Wright, R.W., & Swingle, D.M. *Anal. Biochem.* 1986, **156**, 220-222.

Stewart, J.M. & Young, J.D. *Solid Phase Peptide Synthesis* 1986, 2nd. Ed. Pierce Chemical Company, Il.

St. John, R.T., Johnston, H.M., Seidman, C., Garfinkel, D., Gordon, J.K., Shah, V.K., & Brill, W.J. *J. Bacteriol.* 1975, **121**, 759-765.

Strandberg, G.W. & Wilson, P.W. *Can. J. Microbiol.* 1968, **14**, 25-31.

Stratagene Cloning Systems, pBluescript manual, 1988.

Tainer, J.A., Getzoff, E.D., Alexander, H., Houghten, R.A., Olson, A.J., Lerner, R.A., & Hendrickson, W.A. *Nature* 1984, **312**, 127-134.

Taylor, J.W., Ott, J., & Eckstein, F. *Nucl. Ac. Res.* 1985, **13**, 8749-8764, 8764-8785.

Thorneley, R.N.F. & Lowe, D.J. *Biochem. J.* 1983, **215**, 393-403, & 1984, **224**, pp. 877-886, 887-894, 903-909 (I-IV).

Thorneley, R.N.F., Ashby, G., Howarth, J.V., Millar, N.C., & Gutfreund, H., *Biochem. J.* 1989, **264**, 657-661.

United States Biochemical Corporation, Manual for Sequenase V. 2.0, 1989.

Van Regenmortel, M.H.V., Briand, J.P., Muller, S., & Plaué, S. in *Synthetic Polypeptides as Antigens* 1988, pp. 101-102, Elsevier, New York.

Wahl, R.C. & Orme-Johnson, W.H. *J. Biol. Chem.* 1987, **262**, 10489-10496.

Walker, G.A. & Mortenson, L.E. *Biochem.* 1974, **13**, 2382-2388.

Weiss, B, Jacquemin-Sablon, A., Live, T.R., Fareed, G.C., & Richardson, C.C. *J. Biol. Chem.* 1969, **243**, 4543-4555.

White, T., Ph.D. Thesis, in preparation, 1990.

Wink, D.A., McLean, P.A., Hickman, A.B., & Orme-Johnson, W.H. *Biochem.* 1989, **28**, 9407-9412.

Wolf, H., Modrow, S., Motz, M., Jameson, B.A., Hermann, G., & Förtsch, B. *CABIOS* 1988, **4**, 187-191.

Yoch, D.C. & Pengra, R.M. *J. Bact.* 1966, **92**, 618-622.

Zoller, M.J. & Smith, M. *Meth. Enzymol.* 1983, **100**, 468-500.

Biographical Note

THOMAS A. COLLET

BORN	December 13, 1963 in Beirut, Lebanon
EDUCATION	MASSACHUSETTS INSTITUTE OF TECHNOLOGY, CAMBRIDGE, MASSACHUSETTS Ph.D., Chemistry, September 1990 RWTH AACHEN, AACHEN, FRG Vordiplom, Chemistry, 1985
EXPERIENCE	MASSACHUSETTS INSTITUTE OF TECHNOLOGY, CAMBRIDGE, MASSACHUSETTS
Jan. 1987 - present	Research Assistant. Studied maturation processes and conserved cysteine residues in nitrogenase proteins.
Sept. 1985 - Dec. 1986	Teaching Assistant. Taught undergraduate chemistry lab with emphasis on organic chemistry and instrumentation.
PROFESSIONAL ACTIVITIES	Co-organizer of the Satellite Meeting I of the 8th International Congress on Nitrogen Fixation in Knoxville, Tennessee (May 1990)
HONORS AND ASSOCIATIONS	Fellow of the Studienstiftung Des Deutschen Volkes, 1985-1989 Member of the American Chemical Society, 1986 - present