CHARACTERIZATION OF NITROGEN FIXATION (*nif*) GENES FROM *Paenibacillus polymyxa*

YAM HOK CHAI

UNIVERSITI SAINS MALAYSIA

2007

CHARACTERIZATION OF NITROGEN FIXATION (*nif*) GENES FROM *Paenibacillus polymyxa*

by

YAM HOK CHAI

Thesis submitted in fulfilment of the requirements
for the degree of
Master of Science

March 2007

ACKNOWLEDGEMENTS

Eventually, a book which is less than one inch was culminated. Each page took an average of 10 days. Usually, a thesis does not reflect the real effort that has been invested. Throughout this period, sadness is normally caused by experimental failure; happiness is not only derived from successful experiment but also something that can only be experienced when living in a harmonious community.

The best way to describe the individual connection within this community is ‘symbiotic relationship’. Basically, neither one of us is living alone physically and mentally. The secret is in sharing. When happiness is shared, it becomes greater. In contrast, when sadness is shared, it becomes less burden one.

Professor Nazalan Najimudin who is an expert in molecular genetics was so brave and dared to recruit someone like me (who didn’t even know how to spell ‘science’ correctly) to be involved in a scientific research. It has been an arduous task for him to raise me. Finally, I hope he had turned a naughty boy into a disciplined researcher. My gratitude to him can not be described by mere words.

My heartfelt thanks are also dedicated to Associate Professor Mohd Razip Samian for providing laboratory support. Coincidentally, three of his ‘daughters’ in the lab became very important women in my life.

I am grateful to all the individuals who brought me happiness and made my life complete. They include Aini, Ai Tee, Apai, Boon Choon, Boon Poh, Chee Yong, Emmanuel, Eugene, Hanim, Ivan, Jeremy, Kem, Le Yau, Qis,

Shamanee, Shima, Tham, Yifen and Lab 409 members. I can’t imagine how miserable my life would be without their companionship.

Neo Pei Chin, whatever you have said to me, I had it engraved in my heart!

Contents

List of Tables

List of Figures

Symbols and Abbreviations

%	percentage
ºC	degree Celsius
σ	sigma factor
λ	lambda clone
µg	micro gram
µl	micro litre
µM	micro molar
aa	amino acid
acc. no.	accession number
atm	standard atmospheric pressure
ADP	adenosine diphosphate
ATP	adenosine triphosphate
bp	base pair
DEPC	Diethyl pyrocarbonate
DIG	digoxigenin (non-radioactive DNA labelling)
DNA	deoxyribonucleic acid
dNTP	deoxynucleotide triphosphates
DTT	dithiothereitol
et al.	*et alii* (and others)
EDTA	ethylenediaminetetraacetic
FAD	flavin adenine dinucleotide
g	gram
IPTG	isopropyl β-thiogalactopyranoside
kb	kilo base pair
kDa	kilo Dalton
M	molar
min	minute
mM	millimolar
OD	optical density
ORF	open reading frame

RACE	rapid amplification of cDNA ends
PCR	polymerase chain reaction
RBS	ribosome binding site
RNA	ribonucleic acid
RT	reverse transcription
rpm	revolution per minute
sec	second
SDS	sodium dodecyl sulphate
SSC	standard saline buffer
TAE	Tris acetic acid EDTA
U	unit
w/v	weight per volume
X-gal	5-bromo-4-chloro-3-indolyl-β-galactoside

PENCIRIAN GEN PENGIKAT NITROGEN (*nif*) DARIPADA *Paenibacillus polymyxa*

Abstrak

Paenibacillus polymyxa adalah sejenis bakteria Gram positif yang berupaya menurunkan dinitrogen (N_2) kepada ammonia. Satu fragmen *nifH* separa telah diamplifikasi dengan menggunakan sepasang primer degenerat. Fragmen tersebut digunakan sebagai prob untuk penyaringan perpustakaan genomi lambda dan juga analisis penghibridan Southern terhadap genom *P. polymyxa*. Berasakan keputusan analisis Southern, *P. polymyxa* dipercayai mempunyai salinan tunggal homolog *nifH*. Daripada penyaringan perpustakaan, lima klon lambda positif telah berjaya dipencilkan. Satu fragmen *nifD* separa telah diamplifikasi daripada klon λC2 dengan menggunakan sepasang primer degenerat untuk *nifD*. Prob *nifH* dan *nifD* digunakan untuk penghibridan terhadap DNA lambda λC2 yang telah dicernakan. Satu fragmen *Eco*RI 4 kb dan fragmen *Nde*I 5.5 kb telah berhibridan secara positif masing-masing kepada prob *nifH* dan *nifD*. Kedua fragmen telah disubklon dan dijujuk. Jujukan yang lengkap untuk rangka bacaan terbuka *nifBHD* dan separa untuk *nifK* telah diperolehi. Tapak permulaan transkripsi bagi *nifB* dan *nifK* telah ditentukan melalui kaedah cRACE dan cmRACE. Klon PCR daripada cmRACE mempunyai jujukan 3' *nifK*. Maka jujukan ORF *nifBHDK* yang sepenuhnya telah berjaya diperolehi. Berasaskan tapak permulaan transkripsi, jujukan promoter bagi *nifB* dan *nifH* didapati mempunyai jujukan konsensus yang jelas tetapi berlainan daripada jujukan promoter *nif* yang tradisional. Ini mencadangkan yang ekspresi gen *nif* dalam *P. polymyxa* adalah berkemungkinan menerusi sistem pengawalaturan yang unik.

CHARACTERIZATION OF NITROGEN FIXATION GENES (*nif*) FROM *Paenibacillus polymyxa*

Abstract

Paenibacilus polymyxa is a Gram positive bacterium capable of converting dinitrogen (N_2) to ammonia. A partial *nifH* fragment was amplified by using a pair of *nifH* degenerate primers. The fragment was used as a probe for screening a genomic library as well as a genomic Southern hybridization analysis. Based on the Southern analysis results, *P. polymyxa* was believed to possess a single copy of *nifH* homologue. From the library screening, five positive lambda clones were successfully isolated. A partial *nifD* fragment was amplified from lambda clone λC2 by using a pair of degenerate primers for *nifD*. The *nifH* and *nifD* probes were then used to hybridize digested lambda λC2 DNA. A 4 kb *Eco*RI and a 5.5 kb *Nde*I fragment hybridized positively to the *nifH* and *nifD* probes, respectively. These fragments were then subcloned and sequenced. An intact sequence of *nifBHD* and a partial sequence of *nifK* open reading frames (ORFs) were obtained. The transcriptional start sites of *nifB* and *nifH* were determined by cRACE and cmRACE methods. The cmRACE PCR clones also carried the 3' end sequence of *nifK.* Therefore, intact ORFs of *nifBHDK* were successfully obtained. Based on their transcriptional start sites, the promoter regions of *nifB* and *nifH* revealed consensus sequences that were distinguishable from the traditional *nif* promoter. This suggested that the expression of *nif* genes in *P. polymyxa* is possibly controlled by a unique regulation system.

Chapter 1 Introduction

The rapid growth of human population and agricultural activity are closely related. The soil nitrogenous source is usually a limiting factor in any increase in agricultural activity. Great endeavor is undertaken in order to solve the nitrogen crisis before it becomes a significant problem.

The largest nitrogen reservoir is atmospheric dinitrogen (N_2). Hence, the seemingly economical way is to develop an industrial scale process to convert N_2 into a form that can be assimilated by plants. The Haber-Bosch process finally realized the conversion of N_2 to ammonia in 1908 under conditions of high temperature (450ºC) and pressure (200 atm) in the presence of an iron-based catalyst (Haber, 1922).

Prior to this, biological nitrogen fixation was the exclusive way to provide nitrogen source into the biosphere (relatively small amount is produced by terrestrial source, lightning and volcanic activity). Today, both biological and chemical processes generate even amount of fixed nitrogen to fulfill the demand of agricultural activity of the world population. However, speedy increase in usage of chemically fixed nitrogen causes a profound environmental problem (Smil, 2001). Thus, researchers of biological nitrogen fixation hope to moderate the usage of chemical fertilizer as well as gain priceless scientific knowledge.

Biological nitrogen fixation is a trademark of prokaryotes. More than 100 species were reported to have nitrogen fixation ability. It is found in most bacterial phylogenetic groups and methanogens in Archaea (Raymond *et al*., 2004). Conversion of N_2 to ammonia is catalyzed by a metalloenzyme called

nitrogenase. The enzyme contains two components that are named according to their metal constituents: Fe-protein and MoFe-protein (Dixon & Kahn, 2004).

The most well studied nitrogen fixers are *Klebsiella pneumoniae* and *Azotobacter vinelandii* and both are Gram negative bacteria. Research activities exploiting Gram positive diazotrophs are still few (Klipp *et al*., 2004). Gram positive diazotrophs have been shown to have distinctly different *nif* genes organization and regulation patterns (Wang *et al*. 1988, Harriott *et al*., 1995). For instance, NifA protein in *K. pneumoniae* is a well known positive regulator that mediates the expression of nitrogenase in respond to various environmental signals such as oxygen and nitrogen. In contrast, *nifA* gene does not even exist in the genome of the Gram positive diazotroph *Frankia alni* (Harriott *et al*., 1995). The intact genome sequence of *Frankia sp*. CcI3 (Acc. no. NC_007777) also does not possess a *nifA*-like gene.

Therefore, we have undertaken investigation on several Gram positive diazotrophs such as *Paenibacillus durus*, *P. macerans* and *P. polymyxa*. In this research, *P. polymyxa* is used to uncover the genetics of nitrogen fixation in a Gram positive bacterium. Using gene isolation methodology and cmRACE approach, the complete DNA sequence of *nifBHDK* was obtained.

The transcriptional start sites of *nifB* and *nifH* were determined using both cRACE and cmRACE approaches. The findings of this research are in line with the notion that Gram positive diazotrophs apparently have a different mode of gene regulation. This speculation is based on dissimilar promoter motif, gene organization and transcriptional pattern.

Chapter 2 Literature Review

2.1 Nitrogen: Essential element for biomolecules

Nitrogen is an essential part of most biological compounds such as amino and nucleic acids and this makes it vital to all life. It is the next most abundant element in the living cell after carbon. For example, a typical bacterial cell possesses 12% nitrogen in terms of dry mass. In nature, nitrogen exists in both organic and inorganic forms. However, the bulk of available nitrogen is in the inorganic form, either as ammonia (NH_3), nitrate (NO_3^-) or N_2. Ammonia and nitrate can be used as nitrogen sources while N_2 (nitrogen gas) requires a reduction process prior to assimilation by plants and other living organisms. Unfortunately, nitrogen, the fifth most abundant element in the universe and makes up about 78% of the earth's atmosphere is inert. Therefore, biological nitrogen fixation is a crucially important process to convert nitrogen gas into ammonia using an environmentally friendly and natural mechanism (Arp, 2000; Finan *et al*., 2002).

2.2 Biological nitrogen fixation

Biological nitrogen fixation, the conversion of atmospheric nitrogen into ammonia by symbiotic, associative and free-living bacteria, is responsible for supplying more than 60% of the world's annual new ammonia source (Schlesinger, 1991). The amount of biologically fixed nitrogen produced is in excess of 2×10^{13} g/year (Falkowski, 1997). The availability of fixed nitrogen is normally the limiting factor for crop productivity. Thus there is a tremendous demand on global agriculture to provide food security which correlates

dramatically to the increase in world's population in the twenty-first century. The increased utilization of chemical fertilizers, which constitutes the largest human interference in the nitrogen cycle, has prompted concerns regarding profound pollution impacts such as increased emissions of nitrogen oxides, soil acidification and water eutrophication. Fortunately, the fixed nitrogen provided by biological nitrogen fixation is less prone to leaching and volatilization as it is utilized *in situ*. Therefore this biological process contributes an important input into agriculture in a sustainable manner (Dixon & Kahn, 2004; Gallon & Chaplin, 1987).

2.3 Diazotrophic microorganisms

Biological nitrogen fixation is exclusively an ability of prokaryotes. These organisms are called diazotrophs and are widely distributed among the prokarya and archaea with more than 100 species reported (Raymond *et al*., 2004). The ability to fix nitrogen is found in most bacterial phylogenetic groups, including green sulphur bacteria, Firmibacteria, actinomycetes, cyanobacteria and all subdivisions of the Proteobacteria. In Archaea, nitrogen fixation is mainly restricted to the methanogens. The ability to fix nitrogen is compatible with a wide range of physiological state such as aerobic (for example, *Azotobacter*), facultatively anaerobic (for example, *Klebsiella*) or anaerobic (for example, *Clostridium*) heterotrophs; anoxygenic (for example, *Rhodobacter*) or oxygenic (for example, *Anabaena*) phototrophs; and chemolithotrophs (for example, *Leptospirillum ferrooxidans*). Diazotrophs are also found in a wide variety of habitats including soil and water. They are also found in various states of associations; symbioses with grasses, termite guts,

woody plants, and root–nodule formation in legumes (Raymond *et al*., 2004; Madigan, 2000).

2.4 *Paenibacillus polymyxa*

Paenibacillus polymyxa was firstly isolated by Prazmowski (Buchanan dan Gibbons, 1974). *Paenibacillus durus* (formerly *P. azotofixans*), *P. polymyxa* and *P. macerans* are nitrogen fixers which form a monophyletic cluster in the genus *Paenibacillus*. These are formerly labeled as *Bacillus* until they were reclassified based on rDNA sequence analysis (Ash *et al*., 1993).

P. polymyxa is a bacterial species that is often found in soil and in the rhizosphere and rhizoplane of grasses such as wheat, maize, sorghum and sugarcane (Line and Loutit, 1971; Holl *et al*., 1988; Mavingui *et al*., 1992; von der Weid *et al*., 2000). Many *P. polymyxa* strains are free-living nitrogen fixers (Grau and Wilson 1962; Seldin *et al*., 1983) and there was evidence on the secretion of plant growth-enhancing substances by root-associated *P. polymyxa* (Holl *et al*., 1988). Besides their ecological importance as plant-growth-promoting rhizobacteria, *P. polymyxa* strains are also important for the pharmaceutical and food industries (Debabov, 1982; Priest, 1993). *P. polymyxa* is known to produce two types of peptide antibiotics (Beatty & Jensen, 2002). One group is active against bacteria and includes the polymyxin-colistin-circulin family, polypeptins, jolipeptin, gavaserin, and saltavalin, which contain a 2,4-diaminobutyric acid. The other is active against fungi and actinomycetes and includes gatavalin and fusaricidins. This species also synthesizes plant hormones auxin (Lebuhn *et al*., 1997) and cytokinin (Timmusk *et al*., 1999).

2.5 Key enzymes involved in biological nitrogen fixation

The conversion of dinitrogen(N_2) to ammonia (NH_3) is catalysed by the nitrogenase enzyme, a complex of metalloproteins with conserved structural and mechanistic features. Nitrogenase is a combination of two soluble proteins, known as MoFe protein and Fe protein, according to their metal property. Fe protein functions as an ATP-dependent electron donor to the larger component, the MoFe protein, which contains the enzyme catalytic site (Rees *et al*., 2005).

2.5.1 MoFe protein

MoFe protein has a size of 200 to 250-kDa (also called dinitrogenase) and is composed of four subunits ($\alpha_2\beta_2$ tetramer). Each subunit has a size of approximately 60-kDa; the α subunit is encoded by the *nifD* gene while the β subunit is encoded by the *nifK*. The MoFe protein contains two types of metal-sulfur clusters: the P-cluster (8Fe-7S cluster) and the FeMo cofactor (Mo-7Fe-9S cluster). The latter is usually designated as FeMo-co and is believed to contain the substrate binding site and reduction process (Rees *et al*., 2005). The P-cluster is coordinated by six Cys ligands, bridges the α and β subunits and intermediates the electron transport pathway between Fe-protein and FeMo-co. FeMo-co cluster is connected to the protein by only two ligands (Cys and His) located within the α subunit (Kim and Rees, 1994; Eady, 1995; Igarasshi and Seefeldt, 2003).

2.5.2 Fe-protein

Fe-protein (also called dinitrogenase reductase) is a dimer composed of two copies of a single α subunit and is encoded by *nifH*. It has a size of 55 to 65-kDa. The Fe-protein contains a metal cluster ([4Fe–4S] cluster) that covalently connects the two α subunits. The [4Fe–4S] cluster is the redox-active site involving in electron transfer to FeMo-protein which cycles between reduced state and oxidized state (Eady, 1995; Igarashi and Seefeldt, 2003).

2.5.3 Stoichiometry

The overall stoichiometry of dinitrogen reduction under normal condition is as follows:

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

The enzyme mechanism starts with a reduction of the Fe protein by electron donors such as ferredoxin and flavodoxin (Dixon and Kahn, 2004, Igarashi and Seefeldt, 2003). This is followed by a transfer of a single electron from the Fe protein to the MoFe protein (which is dependent on MgATP hydrolysis). Finally, an internal electron transfer takes place in the MoFe protein by the P cluster to the FeMo-co cluster at the substrate-binding site. Each electron-transfer step requires a continuous cycling step of association and dissociation of the Fe and MoFe proteins as shown in Figure 2.1. Nitrogenase is a relatively slow enzyme with a turnover time of about $5s^{-1}$, and thus the dissociation of the complex is the rate-limiting step. Complex formation has a crucial role in the enzyme mechanism as it is required for the coupling of ATP hydrolysis to electron transfer (Thornely & Lowe, 1985).

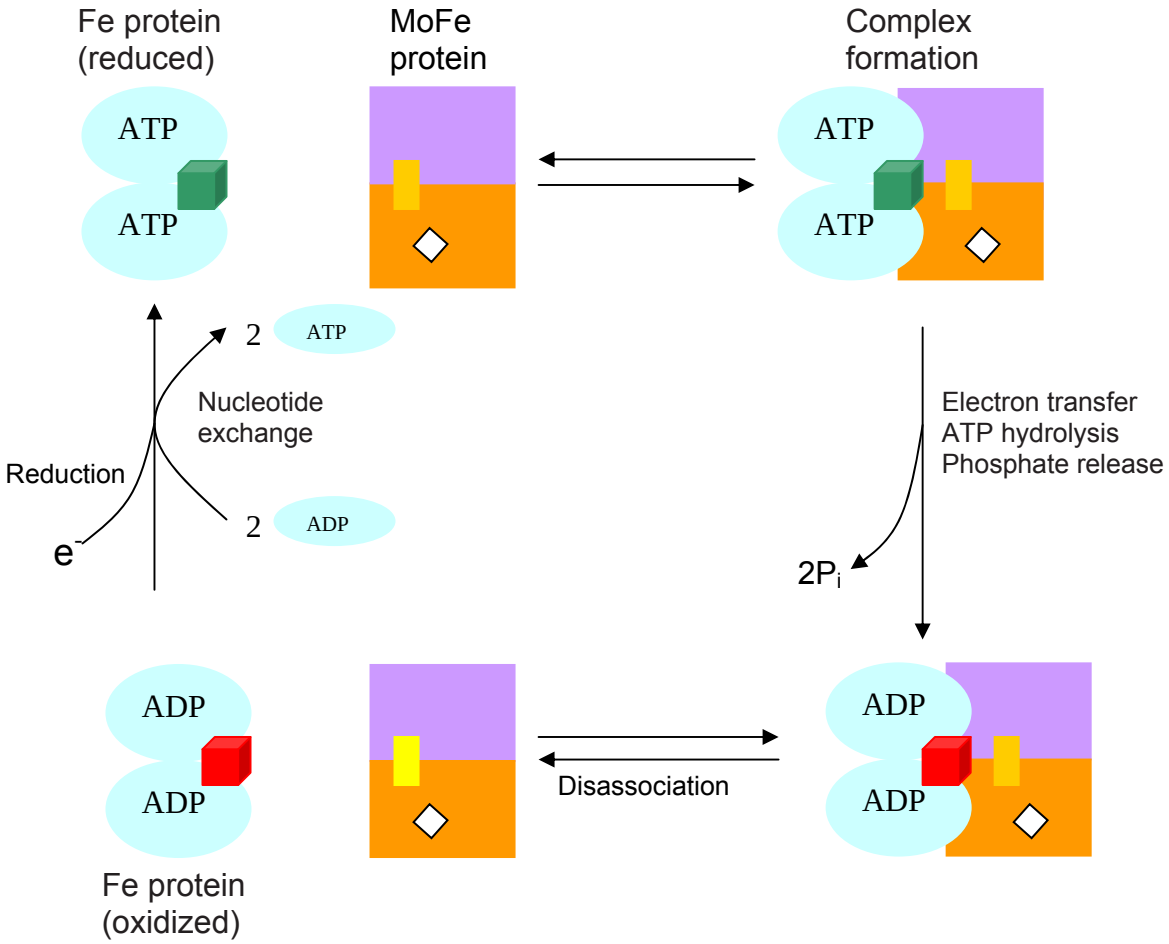

Figure 2.1 : Schematic representation of the nitrogenase Fe protein cycle. The Fe protein dimer is shown in light blue with the cube representing the [4Fe–4S] cluster coloured green to indicate the reduced form and red to represent the oxidized form. The α and β subunits of the MoFe protein are depicted as orange and purple, respectively, the yellow squares represent the P cluster and the white diamond represent the FeMo cofactor (adapted from Dixon & Kahn, 2004).

There is a conformational change in the Fe protein upon ATP turnover coupled to repositioning of the [4Fe–4S] cluster. This brings the cluster in closer proximity to the MoFe protein, thereby facilitating inter-protein electron transfer from the Fe protein to the MoFe protein (Schindelin *et al.*, 1997).

2.5.4 Alternative nitrogenase

Some microorganisms contain alternative nitrogenase whereby Mo is replaced by either Fe or V in condition lacking Mo source. Molybdenum-lacking nitrogenase was firstly reported in *Azotobacter vinelandii* (Bishop *et al.*, 1980). These so-called alternative nitrogenases are found only in limited diazotrophs and their presence is secondary to the MoFe protein. An alternative nitrogenase from *Azotobacter chroococcum* was first purified and characterized by Robson *et al.* (1986), who demonstrated that the enzyme contained vanadium rather than molybdenum. The vanadium nitrogenase is encoded by the *vnf* genes (Robson *et al*., 1989). After the description of the vanadium nitrogenase, Chisnell *et al.* (1988) isolated a second alternative nitrogenase apparently lacking any metal other than iron in its dinitrogenase protein. The gene encoding this second alternative nitrogenase was designated as *anf* genes (Joerger *et al*., 1989). Vanadium containing nitrogenase is expressed preferentially if Fe and V are both present in Mo lacking situation and also if the microorganism has all three types of nitrogenases. The MoFe-co containing nitrogenase has highest affinity and efficiency in nitrogen fixation compared to the other two nitrogenases in the order of Mo>V > Fe (Joerger and Bishop, 1988).

Obviously, the structural genes of *nif*, *vnf*, and *anf* nitrogenases show significant sequence similarity. The general properties of the component proteins are also quite similar. All nitrogenase systems contain two components; a dinitrogenase protein and a dinitrogenase reductase protein. The usual *nif*-encoded dinitrogenase protein is $\alpha_2\beta_2$ tetramer of the *nifDK* gene products whereas the alternative dinitrogenases encoded by the *vnfDK* and *anfDK* genes consist a third subunit known as VnfG and AnfG respectively (Robson *et al*., 1989; Premakumar *et al*., 1989). In contrast to the reduction of acetylene to ethylene by the *nif*-encoded nitrogenase, *anf*-encoded nitrogenase and *vnf*-encoded nitrogenase are able to reduce acetylene to a mixture of ethylene and ethane (Dilworth *et al*., 1988; Scott *et al*., 1990).

The fourth family of nitrogenases is represented by enzyme from the carboxydotrophic bacterium *Streptomyces thermoautotrophicus* (Ribbe *et al.*, 1997). This nitrogenase contains two component proteins. One component, a CO-dehydrogenase, oxidizes CO to CO_2 and reduces O_2 to the superoxide anion radical (O_2^-). The second component is a manganese-dependent oxidoreductase that oxidizes O_2^-, providing electrons to the N_2 and reducing the MoFeS active site (Ribbe *et al.*, 1997). Among the most striking properties of the *S. thermoautotrophicus* nitrogenase system are the dependence on O_2 and O_2^-, the complete insensitivity to neither O_2 nor H_2O_2, the inability to reduce acetylene or ethylene, and a low MgATP requirement. In addition, the subunit structure of the *S. thermoautotrophicus* nitrogenase components and the polypeptides involved distinctly dissimilar from the other three nitrogenase families (Ribbe *et al.*, 1997).

2.6 Genes involved in biological nitrogen fixation

Klebsiella pneumoniae and *Azotobacter vinelandii* are the best-studied diazotrophs and most of the biological nitrogen fixation related genes were discovered in these bacteria. About 20 genes were found to be responsible in the complex nitrogen fixation mechanism for both nitrogenase protein synthesis and genetic regulation of the process (Jacobson *et al*., 1989; Arnold *et al.*, 1988). Knowledge obtained from this two species became essential reference for the studies of other diazotroph. Table 2.1 shows a summary of *nif* genes and functions and their corresponding proteins.

2.7 Regulation of nitrogenase expression

2.7.1 Oxygen control

Nitrogenase is irreversibly inactivated by oxygen and hence environmental oxygen tension is one of the major regulatory factors. In the case of strict anaerobes, oxygen regulation has less importance since the anaerobes can only survive in oxygen free environment. For the facultative aerobes, oxygen regulation becomes very significant in order to avoid nitrogenase denaturation by oxygen (Merrick, 1992).

In the case of the strict aerobe *Azotobacter*, a very high respiratory activity rapidly reduces the oxygen, thus decreasing the internal oxygen concentration and thereby protecting the nitrogenase (Philips &Johnson, 1961). Most aerobic diazotroph produce gummy colonies on agar medium and this may also play a role in protecting nitrogenase from oxygen (Hill *et al*., 1972).

Table 2.1 Nitrogen fixation (*nif*) genes products and functions.

Gene(*nif*)	Function
Q	Incorporation of molybdenum into nitrogenase (Imperial *et al*., 1984)
B	FeMo-cofactor synthesis (Curatti *et al*., 2006)
A	Positive regulation (Lee *et al*., 1993; Eydmann *et al*., 1995; Schmitz *et al*., 2002)
L	Negative regulation (Lee et al., 1993; Eydmann *et al*., 1995; Schmitz *et al*., 2002)
F	Electron carrier : flavodoxin (Arnold *et al*., 1988; Taylor *et al*., 1990)
M	Nitrogenase reductase processing (Roberts *et al*., 1978; Arnold *et al*., 1988)
Z	Maturation and activation of FeMo-protein (Paul and Merrick, 1989)
W	Maturation and activation; oxygen protection of FeMo-protein (Paul and Merrick, 1989; Lee *et al*., 2000)
V	FeMo-cofactor synthesis: homocitrate-synthase (Zheng *et al*., 1997; Allen *et al*., 1994)
S	Homodimeric cysteine desulfurase, S activation in metallocluster synthesis (Zheng *et al.*,1993; Zheng and Dean, 1994)
U	FeMo-protein processing (Harris *et al*., 1990; Fu *et al*., 1994)
X	FeMo cofactor synthesis. Negative regulation (Lee *et al.*,2000)
E	Synthesis and insertion of FeMo-co into dinitrogenase protein (Orme-Johnson, 1985; Arnold *et al.*, 1988)

Table 2.1 (continued)

Y	Processing of MoFe-protein (Homer *et al*., 1993)
T	Nitrogenase maturation (Simon *et al*., 1996)
K	FeMo-protein β subunit (Arnold *et al*,. 1988; Kim and Rees, 1994)
D	FeMo-protein α subunit (Arnold *et al.*, 1988; Kim and Rees, 1994)
H	Fe-protein subunit (Arnold *et al*,. 1988; Kim and Rees, 1994)
J	Electron transfer: pyruvate-flavodoxin oxidoreductase (Schmitz *et al*., 2001)

In heterocystous cyanobacteria, the enzyme nitrogenase seems to only exist in the heterocysts-thick-walled cells that lack oxygen due to photosynthetic process. Some heterocyst defective mutants are unable to fix nitrogen. Therefore, heterocysts play an essential role in nitrogen fixation (Fleming & Haselkorn, 1973).

In *Klebsiella pneumoniae* and *Azotobacter vinelandii* the nitrogen regulatory proteins NifL and NifA tightly control the synthesis of nitrogen fixation genes in response to oxygen (and nitrogen). The transcriptional activator NifA is required for transcription of other nitrogen fixation (*nif*) genes. In this regulation, the negative regulator NifL inhibits NifA activity. Immunological studies, chromatography and complex formation analyses using the yeast two-hybrid system demonstrates that NifA interacts directly with NifL by protein-protein interaction (Henderson *et al*., 1989; Money *et al*., 1999 and 2001; Lei *et al*., 1999). This indicates that the oxygen molecule finally results in a complex formation between NifL and NifA which inhibits NifA activity and thus prevents the transcription of other *nif* genes (Figure 2.2). The inhibitor NifL is a flavoprotein which regulates NifA activity depending on the reduction status of its N-terminally bound FAD-cofactor. It allows NifA to function only under anaerobic conditions. Thus, the redox-sensitive FAD-cofactor appears to be involved in oxygen signal-transduction (Hill *et al*., 1996; Klopprogge *et al.*, 2002; Grabbe & Schmitz, 2003).

2.7.2 Nitrogen control

Nitrogen itself is a major regulatory factor in free living diazotrophs. These microorganisms can only fix nitrogen when a nitrogenous source is

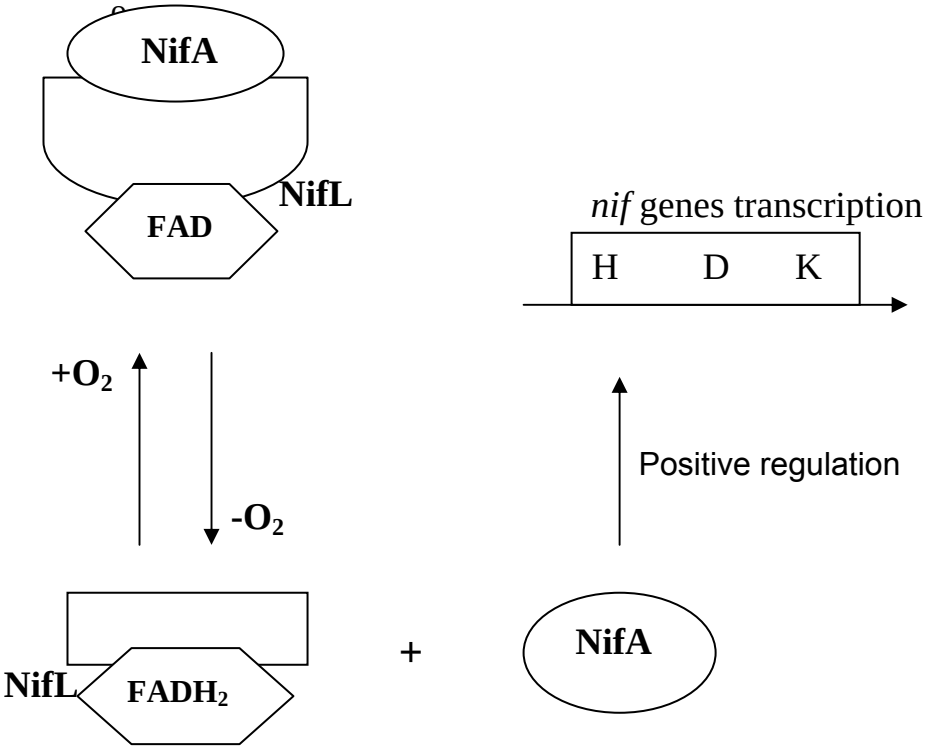

Figure 2.2: Model for oxygen control in *K. pneumoniae* (Adapted from Dixon and Kahn, 2004). In low oxygen condition, NifL dissociates from NifA, thus allowing the latter to drive the expression of *nif* genes.

limited. A general nitrogen regulation (ntr) system controls the expression of many genes including the *nif* system in the diazotroph. Four key proteins are involved in nitrogen control system; a uridylyltransferase(UTase) encoded by *glnD*, a small tetrameric effector protein (PII) encoded by *glnB*, NtrB encoded by *ntrB* and a transcriptional activator NtrC encoded by *ntrC* (Merrick, 1995).

UTase participates as the primary sensor of the cellular nitrogen status. When cells are lacking nitrogen, Utase mediates the uridylation of PII by transferring a uridylyl group onto a tyrosine residue on each of the four PII subunits (Figure 2.3). The uridylylated PII (PII-UMP) indirectly causes the phosphorylation of NtrC. It actually mediates autophosphorylation of NtrB followed by the transfer of the phosphate to NtrC. Under nitrogen rich condition, UTase acts as a uridylyl-removing enzyme, converting PII-UMP to its original form, and NtrB now promotes the dephosphorilation of NtrC. In the second level of regulation, phosphorylated NtrC activates the expression of *nifA* and hence control the nitrogenase expression response to the low nitrogen level in the cell (Merrick 1994; Merrick 2004).

2.7.3 *nif* promoter is regulated by DNA supercoiling

In prokaryotes, negatively supercoiled genomic DNA promotes DNA recombination, replication and transcription (Snoep *et al*., 2002). In addition, DNA supercoiling plays an important role in the cellular perception of environmental signals. Changes in oxygen content directly affect DNA superhelical density via the action of gyrase which increases negative supercoiling. In terms of oxygen content, anaerobic conditions activate bacterial gyrase, whereas aerobic conditions activate topoisomerase I. Indeed,

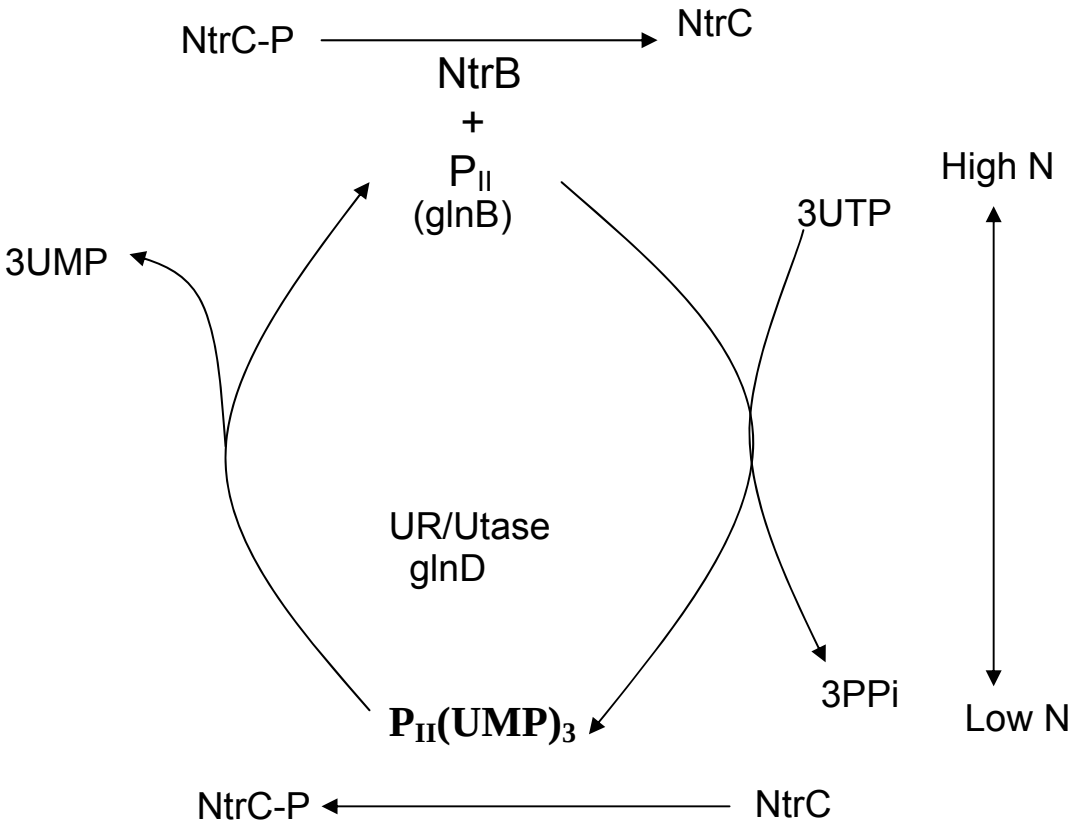

Figure 2.3: The nitrogen regulation Ntr system of enteric bacteria. The activity of the response regulator NtrC is regulated in response to the intracellular nitrogen status. UTase (*glnD* product) catalyses the uridylylation and deuridylylation of P_{II} (*glnB* product). P_{II} in turn regulates the activity of the sensor histidine kinase NtrB which catalyses the phosphorylation and dephosphorylation of NtrC (Adapted from Merrick, 2004).

it appears that the latter activates expression of aerobically-induced genes, whereas gyrase favors the expression of anaerobically-induced genes such as the hydrogenase gene (Friedman *et al*., 1995).

In *Klebsiella pneumoniae*, the transcription of *nifLA* operon requires DNA gyrase activity while the transcription of other nitrogen fixation (*nif*) genes does not appear to be dependent on DNA gyrase activity (Dimri & Das, 1988). *K. pneumoniae nifLA* promoter activity requires a certain level of DNA negative supercoiling which provides a connection between the regulation of *nifLA* expression and the aerobic/anaerobic conditions known to alter supercoiling (Dixon *et al.*, 1988). Hu *et al.* (2000) also reported that DNA gyrase activity is essential for *Enterobacter cloacae nifLA* promoter expression. Liu *et al.* (2005) found that both the oxygen-insensitive *nifH* promoter and the anaerobically induced *nifLA* promoter required DNA supercoiling for optimal expression of the *nifH* and *nifLA* genes in *Sinorhizobium meliloti*. The finding suggests that promoter activities may require a certain level of DNA supercoiling, regardless of direct oxygen control.

2.8 Promoter motive of *nif* genes

The NifA protein is a specific activator of the expression of *nif* genes promoters by interacting with SigN-containing RNA polymerase (SigN is a Sigma factor in *Escherichia coli*, also known as σ^N, σ^{54} and *ntrA*). This RNA polymerase holoenzyme recognizes the consensus dinucleotides sequence GG and GC at the positions -12 and -24 with respect to the transcriptional start site, respectively. This consensus sequence exists in the most of the *nif*

promoters in *K. pneumoniea* including *nifH, nifE, nifU, nifM, nifF, nifL* and *nifB* (Barrios *et al*., 1999; Jack *et al*., 1999).

Purified SigN can bind to certain promoters in the absence of core RNA polymerase, indicating that the DNA binding determinants of SigN can function in the isolated protein (Buck and Cannon, 1992). This is in contrast to the major sigma factor of enteric bacteria SigA (also known as σ^{70}), in which the DNA binding determinant is masked until the protein binds to the core RNA polymerase (Dombroski *et al*., 1993).

Based on promoter study of *K. pneumoniae*, NifA (a member of enhancer-binding proteins) binds to a specific upstream activating sequence, UAS (also known as enhancer sequence), and interacts with the SigN-containing RNA polymerase bound to the -12,-24 portion. Mutational analysis of the *nifH* UAS has supported the suggestion that the TGT-N_{10}-ACA motif (where N is any nucleotide), which characterizes the UAS of *nif* promoters, is a NifA binding site (Buck *et al*., 1987; Morett *et al*., 1988). This UAS sequence is normally found at the 100 to 200bp region upstream of *nif* genes. NifA protein participates in the activation of transcription by binding to the UAS and contacting the downstream SigN-containing RNA polymerase complex by forming a loop forming in the DNA between the UAS and the -12,-24 (GG-N_{10}-GC) conserved sequence promoter element (Buck *et al*., 1987; Morett & Buck, 1989; Jack *et al*., 1999).

2.9 Gram-positive, free living diazotrophs

Most of nitrogen fixation related research focussed on Gram-negative bacteria such as *K. pneumoniae* while Gram-positive diazotrophs received

very little attention. Interestingly, Gram-positive diazotrophs show exclusively contrasting features either in their *nif* genes promoter motive or genes arrangement compared to Gram-negative diazotroph (Wang *et al*., 1988; Harriott *et al*., 1995). Thus, research on Gram-positive diazotrophs is much needed, especially to look at new features of *nif* genes regulation. Below are some examples of research of Gram-positive diazotrophs.

2.9.1 *Clostridium pasteurianum*

Clostridium pasteurianum is a Gram-positive anaerobic bacterium with a low G+C content of 26 to 28% (Cummins & Johnson, 1971) which distinguishes it from the rest of well-studied nitrogen-fixing microorganisms. The primary structure of *C. pasteurianum* nitrogenase components is significantly less related to that of nitrogenases from other microorganisms (Chen *et al*., 1973). *C. pasteurianum* nitrogenase has high activity, but its components are distinctly ineffective in forming active heterologous complexes (Emerich & Burris, 1978, Smith *et al*., 1976) and are less sensitive to H_2 as an inhibitor (Guth & Burris, 1983).

Wang *et al*. (1988) reported that *C. pasteurianum* has five *nifH*-like sequences in addition to *nifH* (for a total of 6 copies). The nucleotide sequence similarities among these *nifH* and *nifH*-like sequences range from 68 to nearly 100%, and most of them are transcribed under N2-fixing conditions. The gene *nifH* is adjacent to *nifH2* and the rest of *nifH* copies (*nifH3*, *nifH4*, *nifH5 and nifH6*) are located separately (Wang *et al*., 1988).

Both *nifH1DK* and *nifH5* genes have sequences of TATTG at the -35 and -15 regions and TATAT or TATACT at the -10 region which are

homologous to those of other Gram-positive bacteria. The genes *nifH2* and *nifH6* match perfectly with the *E. coli* SigA promoter sequences of TTGACA at -35 region and TATAAT at -10 region, these are in contrast to neither *nif*A box nor SigN consensus sequence as detailed (Wang *et al*., 1988).

Interestingly, in the -100 regions, the sequence ATCAATAT-N_{6-10} ATGGATTC is homologous among all the six transcription units. This sequence consists of three segments, with N6-10 appearing to serve as a bridge which separates the first segment, ATCAATAT, from the third segment, ATGGATTC (Wang *et al*., 1988).

2.9.2 Paenibacillus massiliensis T7

Paenibacillus massiliensis is a Gram-positive, free living, spore forming diazotroph, isolated from the country side of Beijing, China. The isolated *nif* cluster of *Paenibacillus massiliensis* consists of seven *nif* genes in the arrangement *nifBHDKENX* (*nifX* partial) (Zhao *et al*., 2005).

The -180, -166 region (upstream from *nifB*) has a consensus UAS sequence (presumably a NifA binding site): TGT-N8-ACA. The -155, -140 region has SigN recognition motive which is TGGCA-N6-ATGA. The ribosome binding site (AGAA) is located 8 bp before the *nifB* start codon.

A *nifB* promoter activity analysis showed that high oxygen tension suppresse the expression of *nifB* promoter, while ammonium (nitrogen source) had less effect on *nifB* promoter activity in anaerobic condition (Zhao e*t al*., 2005).

2.9.3 Paenibacillus durus

Paenibacillus durus (formerly *P. azotofixans*) is a Gram-positive, facultatively anaerobic diazotroph, classified into a broad cluster of nitrogen fixers in rRNA group 3. This cluster also includes *Paenibacillus macera*ns and *Paenibacillus polymyxa* (Ash *et al*., 1991). It fixes atmospheric nitrogen with high efficiency and this ability is not affected by the presence of nitrate which is in contrast to the majority nitrogen fixers (Rosado *et al*., 1998).

Choo *et al*. (2003) reported the presence of three *nifH* homolog in the *P. durus* and these are designated as *nifH1*, *nifH2* and *nifH3*. Adjacent to *nifH1* are *nifB1, nifD1* and *nifK1* in the following arrangement *nifB1H1D1K1*. It was firstly found that the gene *nifB* is located upstream to *nifH*. NifH1 and NifH2 protein comparison yielded 97% identity. In contrast, comparing either *nifH1* or *nifH2* to *nifH3* yielded a relatively low 43% identity (Choo *et al*., 2003).

An analysis of NifH phylogeny demonstrated clustering of *P. durus* NifH1 and NifH2 within the *Cyanobacteriaceae* grouping while NifH3 clustered with the NifH proteins of members of the *Archaea* domain. Interestingly, none of the NifH proteins from *P. durus* clustered with the NifH of other gram-positive diazotroph such as *Frankia* sp. (a high-G+C firmicute) and *Clostridium pasteurianum* (a low-G+C firmicute). Unusual placement of NifH3 among the highly divergent members of the Archaea suggests the occurrence of horizontal gene transmission (Choo *et al*., 2003).

2.10 Transcriptional Start Site (TSS) determination methods

Determination of the first nucleotide (5'-end) of mRNA is a crucial step to identify and analyse a gene's promoter. Several methods have been

documented to map mRNA 5’ end including RNase protection, S1 mapping and primer extension (Ausubel *et al.*, 1988). These methods require relatively large amounts of targeted mRNA and often fail to identify the 5' end of rare mRNAs. An alternative procedure known as the rapid amplification of cDNA ends (RACE) has been suggested (Frohman *et al.*, 1988). The mRNA is reverse-transcribed with a gene-specific primer and the resultant cDNA is modified at the 5' end with a homo-oligonucleotide synthesized by terminal deoxynucleotidyl transferase. The tailed cDNA is subjected to PCR amplification with a complimentary oligonucleotide to the tail and a gene-specific primer. Using this technique, it is however often difficult to detect the 5' end of rare mRNAs due to inefficiency of the tailing reaction and non-specific priming by a homopolymeric primer. To overcome this problem, anchored RACE has been introduced (Liu and Gorovsky, 1993). The mRNA is ligated with an anchor (short chain of RNA oligonucleotide) followed by cDNA synthesis using a gene specific primer. This is followed by a PCR amplification using primers corresponding to the sequences of anchor and targeted gene, cloning and sequencing. The method requires multi-step chemical and enzymatic reactions. Further more, anchor ligation is non selective in which the anchor may ligate to non-target mRNA (as well as rRNA and tRNA) in the RNA pool.

In this research, a method named cRACE (circular or concatemeric first-strand cDNA-mediated RACE) was used to determine transcriptional start site of *nif* genes (Maruyama *et al*., 1995). This method simplifies the multi enzymatic steps and all the primers involved are gene-specific. Thus, non-

specific PCR products should be less likely to be amplified. The complete protocol is detailed in Chapter 3.

RT-PCR analysis of circularized mRNA is the only option to concurrently determine the 5' as well as 3' mRNA sequences. This approach requires target mRNA to be religated by using RNA ligase, followed by cDNA synthesis and PCR amplification targeting the joined regions. The reliability of this method was firstly ensured by Mandl *et al.* (1991) on identification 5' and 3'-terminal regions of linear RNA genome from several tick-borne flaviviruses. This method was also included in this research to refine the results obtained from cRACE. An exhaustive protocol is detailed in Chapter 3.

2.11 Research objectives

The objectives of this research are summarized as below:

a. To obtain complete sequence of the *nifH* gene of *P. polymyxa* as well as *nif*-related genes adjoining to *nifH*.
b. To analysis phylogenetic of *nifH* isolated from *P. polymyxa*.
c. To determine the transcriptional start sites of *nif* genes.
d. To postulate probable promoter and regulatory regions.