Development Of A Novel Prokaryotic Two-Hybrid System For The Detection And Analysis Of Protein-Protein Interactions In Vivo.

Paul Clarke

Ph D. 2000

Development Of A Novel Prokaryotic Two-Hybrid System For The Detection And Analysis Of Protein-Protein Interactions *In Vivo*.

Thesis

Presented for the Degree of

DOCTOR OF PHILOSOPHY

by Paul Clarke, B. Sc.,

under the supervision of Michael O'Connell, B. Sc., Ph.D.

School Of Biotechnology

Dublin City University

August 2000

I hereby certify that this material, which I now submit for assessment on the programme of study leading to the award of Degree of Doctor of Philosophy, is entirely my own work and has not been taken from the work of others save and to the extent that such work has been cited and acknowledged within the text of my work.

Signed: _____ ID No.: 96970928

Date: 22/09/2000

Acknowledgements.

I would like to says thanks to all the gang in DCU, we had some great crack over the last few years, especially during the John Barry days. Looking forward to Friday pints made the week a bit more bearable. I would also like to thank Dr. Michael O'Connell without whose advice and guidance I might not be where I am today. I would also like to thank my family and friends for all their support and for putting up with my endless moaning when things weren't going so well. I would especially like to thank my girlfriend Sonia to whom I dedicate this thesis. Without your love and support I would not have made it through. You saw me through some of the most difficult stages of my thesis and helped me to keep my head when things got tough. I will never forget all you have done.

Table Of Contents.

Index Of Figures.

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Chapter 6

Chapter 7

Chapter 8

Index Of Tables

Chapter 1

Chapter 2

Chapter 6

Chapter 7

Abstract

The purpose of this work was to develop a novel prokaryotic two-hybrid (P2H) system for the detection of protein-protein interactions *in vivo*. The system is based on the transcriptional activation and DNA binding domains of the *Sinorhizobium meliloti 2011* enhancer binding protein NifA. This protein is responsible for the transcriptional activation of several *nif* and *fix* genes, including the *nifHDK* operon, in response to cellular oxygen and/or nitrogen status.

NifA is modular in nature and has three independently folded domains. The central domain of about 220 amino acids is responsible for the activation of transcription. The function of the N-terminal domain is unknown while the C-terminal domain contains a helix-turn-helix motif that is required for binding to a specific upstream activator sequence. PCR was used to amplify the individual domains of the NifA protein and also to amplify the *nifH* promoter. The PCR products obtained were used to construct suitable plasmids and reporter gene constructs for the P2H system.

Three pKK223-3 based Prey plasmids were constructed using the N-terminal-central (NC) transcriptional activation domain of NifA. Cloning of a known gene or a library of DNA fragments into the multiple cloning site (MCS) of any of these plasmids generates hybrid proteins bearing the N terminal-central (NC) activation domain of NifA. The three Prey plasmids differed from each other only in the reading frame of the MCS relative to the upstream NifA activation domain sequence. The purpose of constructing these three plasmids was to facilitate library construction. Another pKK223-3 based Control plasmid expressing the whole NifA protein was also constructed.

Two pACYC184 based Bait plasmids were also constructed encoding the sequence for the C-terminal DNA binding domain of NifA. Cloning into the MCS of the Bait construct on either of the plasmids results in the expression of a hybrid protein bearing the DNA binding domain of NifA. In addition to the Bait construct these plasmids also maintained the reporter gene constructs. One of the Bait plasmids contained a *nifH:lacZα* reporter gene construct while the other contained a *nifH:gent*^{r} (gentamycin resistance) reporter gene construct. These Bait plasmids were compatible with the Prey plasmids and the Control plasmid.

In addition to the construction of the above mentioned plasmids suitable host strains of *E. coli* were developed in which the plasmids could be used. These strains had deletions of the *glnG* gene since the product of this gene, NtrC, had been observed to weakly activate the *nifH* promoter and could therefore contribute to false positives. Two *glnG* mutant strains of *E. coli* called ET6016 and YMC11 were obtained. These lacked the correct *lacZ*α complementation background required for the *nifH:lacZ*α reporter system to function and so the correct genetic background was introduced into the two strains on an F' factor by conjugation to form strains ET6016*rif*F' and YMC11*rif*F'.

Having constructed the appropriate plasmids, developed suitable strains and optimised growth conditions under which the novel NifA based P2H system would be used, the system was evaluated with the model protein-protein interaction between the appropriate domains of the intimin protein and the Tir receptor of enteropathagenic *E. coli* which had been demonstrated in a yeast two-hybrid system by another laboratory. The sequence for the Int280α protein was cloned into the pPC229-A Prey plasmid to form plasmid pPC233 which expressed a hybrid protein consisting of the Int280α protein fused to the NifA transcription activating NC domain. The sequence for the Tir protein was cloned into the pPC187-A Bait plasmid to form plasmid pPC190 which expressed a hybrid protein consisting of the Tir protein fused to the C-terminal DNA binding domain of NifA. The plasmids pPC233 and pPC190 were co-transformed into the *E. coli* strain ET6016*rif*F' which was then plated on appropriate selective media to test for activation of the *nifH:lacZ*α reporter gene. No interaction between Tir and Int280α hybrid proteins was detected and the possible reasons for the failure to detect an interaction are discussed.

Chapter 1

Introduction.

1.1: Protein-Protein Interactions And Their Biological Significance.

Protein-protein interactions are intrinsic to virtually every cellular process. DNA replication, transcription, translation, splicing, secretion, cell cycle control, signal transduction and intermediary metabolism are all processes in which protein complexes and protein-protein interactions have been implicated as essential components *(Phizicky and Fields, 1995)*. In consequence, the analysis of protein-protein interactions is no longer exclusively the domain of the biochemist. Geneticists, cell biologists, molecular biologists and developmental biologists, to name just a few, have all by necessity become involved in the analysis of protein-protein interactions. The identification and characterisation of protein-protein interactions has led to significant insights into many different biological processes and their underlying mechanisms.

Protein-protein interactions are self-evident in multi-subunit proteins. Such proteins are found in many classes of proteins and include classical proteins such as haemoglobin, tryptophan synthetase and core RNA polymerase. Other well known examples of multi-subunit proteins include much more complicated assemblies of polypeptides. These include metabolic enzymes such as pyruvate dehydrogenase and α-ketoglutarate dehydrogenase complexes, the DNA replication complex of *Escherichia coli* and other organisms, the bacterial flagellar apparatus, ribosomes, the nuclear pore complex and the tail assembly of bacteriophage T4. Although some of the subunits of these proteins are not tightly bound, activity is associated with a large structure that in many cases is called a protein machine *(Alberts and Miake-Lye, 1992)*.

There are also a large number of transient protein-protein interactions, which in turn control a large number of cellular processes. All modifications of proteins necessarily involve such transient protein-protein interactions. These include the interactions of protein kinases, protein phosphatases, glycosyl transferases, acyl transferases, proteases etc., with their substrate proteins. Such protein modifying enzymes encompass a large number of protein-protein interactions in the cell and regulate all manner of fundamental processes such as cell growth, cell cycle, metabolic pathways, and signal transduction. Transient protein-protein interactions are also involved in the recruitment and assembly of the transcription complex to specific promoters, the transport of proteins across membranes,

the folding of native proteins (catalysed by chaparones), individual steps in the translation cycle, and the breakdown and reformation of sub-cellular structures during the cell cycle. Such transient complexes are by their very nature difficult to study, especially by biochemical methodologies, because the proteins or conditions responsible for the transient reaction have to be identified first.

Because of their biological significance a vast array of biochemical and genetic methodologies have been developed to facilitate the identification and analysis of protein-protein interactions, and in recent years there has been an explosion in the development of novel innovative technologies for this purpose. The development of such technologies is becoming ever more important with the onset of genomics.

1.2: Genomic Sequencing And Functional Genomics.

Genomic analysis is currently enjoying exponential growth and for many biologists the new millennium marks the beginning of the era of genomics. During the past few years the sequencing of several eukaryotic and dozens of prokaryotic genomes have been completed and many more sequencing projects are under way. A list of completed microbial sequencing projects is given in Table 1.2.1 below while Table 1.2.2 lists microbial sequencing projects that are in progress. A list of current microbial genome sequencing projects can be found at the web site of the institute of genomic research *(www.tigr.org/)*. This genomic sequence information has the potential to revolutionise biological research by providing scientists with a very powerful tool for studying the biology of organisms.

The availability of extensive genomic information makes it possible to carry out comparative genomic analysis, particularly in the context of determining evolutionary relationships between organisms. The comparison of multiple genome sequences can reveal important functional information; characteristics conserved in an evolutionary lineage indicate important functions within that lineage while sudden shifts indicate functional change. We can also begin to compare the genomes of closely related organisms such as the laboratory strain of *E. coli* K12 and the strain θ157 (which was responsible for

several recent outbreaks of food poisoning) with a view to understanding the genetic basis for differences in pathogenicity between the strains. Similar analysis could be performed for other pathogenic species like *Mycobacterium tuberculosis*.

Two major challenges now face us with respect to extensive genomic sequence information. The first of these is concerned with storing, accessing and analysing the vast amount of genetic information. These problems can be addressed with the development of new improved, easily accessible and searchable computer databases. The second challenge is concerned with interpreting the information in biological terms.

Table 1.2.1: Published Microbial Genomes And Chromosomes.

Genome		*Genome*	
Aeropyrum pernix	A	*Leishmania major Chr1*	E
Aquifex aeolicus	B	*Listeria monocytogenes*	B
Archaeoglobus fulgidus	A	*Methanobacterium thermoautotrophicum*	A
Bacillus halodurans	B	*Methanococcus jannaschii*	A
Bacillus subtilis	B	*Mycobacterium tuberculosis*	B
Borrelia burgdorferi	B	*Mycoplasma genitalium*	B
Campylobacter jejuni	B	*Mycoplasma pneumoniae*	B
Chlamydia pneumoniae	B	*Plasmodium falciparum Chr2*	E
Chlamydia pneumoniae	B	*Plasmodium falciparum Chr3*	E
Chlamydia trachomatis	B	*Pyrococcus abyssi*	A
Chlamydia trachomatis	B	*Pyrococcus horikoshii*	A
Deinococcus radiodurans	B	*Rickettsia prowazekii*	B
Escherichia coli	B	*Saccharomyces cerevisiae*	E
Haemophilus influenzae Rd	B	*Synechocystis sp.*	B
Helicobacter pylori	B	*Thermotoga maritima*	B
Helicobacter pylori	B	*Treponema pallidum*	B
Lactococcus lactis	B		

Key: A= Archaebateria B= Eubacteria E= Eukaryotes

The sequencing of whole genomes results in the identification of a large number of novel genes encoding proteins whose functions and interactions with other proteins have to be determined. When the first complete bacterial genome sequence (*Haemophilus influenza*) was published in 1995, 40% of the predicted genes had no known function and half of these were 'new', i.e. they did not match anything in the databases at the time *(Fleischmann et al, 1995)*.

Table1.2.2: Microbial Genomes And Chromosomes Currently Being Sequenced.

Genome		*Genome*	
Actinobacillus actinomycetemcomitans	B	*Neisseria meningitidis (Serogroup C strain FAM18)*	B
Aspergillus nidulans	E	*Neurospora crassa*	E
Bacillus anthracis	B	*Nitrosomonas europaea*	B
Bacillus stearothermophilus	B	*Pasteurella haemolytica*	B
Bordetella bronchiseptica	B	*Photorhabdus luminescens*	B
Bordetella parapertussis	B	*Plasmodium falciparum Chr1,4,5,6,7,8,9,13*	E
Bordetella pertussis	B	*Plasmodium falciparum Chr10,11*	E
Candida albicans (1161)	E	*Plasmodium falciparum Chr14*	E
Candida albicans (SC5314)	E	*Pneumocystis carinii (f. sp. carinii)*	E
Caulobacter crescentus	B	*Pneumocystis carinii (f. sp. hominis)*	E
Chlamydia pneumoniae	B	*Porphyromonas gingivalis*	B
Chlamydia trachomatis	B	*Prochlorococcus marinus*	B
Chlorobium tepidum	B	*Pseudomonas aeruginosa*	B
Clostridium acetobutylicum	B	*Pseudomonas putida*	B
Clostridium difficile	B	*Pyrobaculum aerophilum*	A
Corynebacterium diphtheriae	B	*Pyrococcus furiosus*	A
Corynebacterium glutamicum	B	*Ralstonia solanacearum*	B
Dehalococcoides ethenogenes	B	*Rhodobacter capsulatus*	B
Dictyostelium discoideum Chr 2	E	*Rickettsia conorii*	B
Encephalitozoon cuniculi	E	*Salmonella typhi*	B
Enterococcus faecalis	B	*Salmonella typhimurium (SGSC1412)*	B
Francisella tularensis	B	*Salmonella typhimurium (TR7095)*	B
Geobacter sulfurreducens	B	*Schizosaccharomyces pombe*	E
Giardia lamblia	E	*Shewanella putrefaciens*	B
Haemophilus ducreyi	B	*Shigella flexneri 2a*	B
Halobacterium salinarium	A	*Sinorhizobium meliloti*	B
Halobacterium sp.	A	*Staphylococcus aureus (COL)*	B
Klebsiella pneumoniae	B	*Staphylococcus aureus (8325)*	B
Lactobacillus acidophilus	B	*Staphylococcus aureus (MRSA)*	B
Legionella pneumophila	B	*Staphylococcus aureus (MSSA)*	B
Leishmania major Chr3(Friedlin)	E	*Streptococcus agalactiae*	B
Leishmania major Chr4(Friedlin)	E	*Streptococcus mutans*	B
Leishmania majorChr5,13,14,19,21,23(Friedlin)	E	*Streptococcus pneumoniae*	B
Leishmania major Chr27(Friedlin)	E	*Streptococcus pyogenes*	B
Leishmania major Chr35 (Friedlin)	E	*Streptococcus pyogenes (Manfredo)*	B
Leptospira interrogans serovar icterohaemorrhagiae	B	*Streptomyces coelicolor*	B
Listeria innocua	B	*Sulfolobus solfataricus*	A
Methanococcus maripaludis	A	*Thermoplasma acidophilum*	A
Methanosarcina mazei	A	*Thermoplasma volcanium*	A
Methylobacterium extorquens	B	*Thermus thermophilus*	B
Methylococcus capsulatus	B	*Thiobacillus ferrooxidans*	B
Mycobacterium avium	B	*Treponema denticola*	B
Mycobacterium bovis	B	*Trypanosoma brucei Chr1*	E
Mycobacterium leprae	B	*Trypanosoma brucei*	E
Mycobacterium tuberculosis	B	*Ureaplasma urealyticum*	B
Mycoplasma mycoides subsp. mycoides SC	B	*Ustilago maydis*	E
Mycoplasma pulmonis	B	*Vibrio cholerae*	B
Neisseria gonorrhoeae	B	*Xanthomonas citri*	B
Neisseria meningitidis (MC58)	B	*Xylella fastidiosa*	B
Neisseria meningitidis (serogroup A strain Z2491)	B	*Yersinia pestis*	B

Key: A= Arehaebateria B= Eubacteria E= Eukaryotes

Similarly, 60% of the yeast *Saccharomyces cerevisiae* genes had no assigned function and half of these encoded putative proteins that had no homology with known proteins (Dujon, 1996). The determination of protein functions from genome sequences has emerged as the next major challenge in molecular biology and for this reason much attention is now being focused on functional genomics.

The field of functional genomics seeks to devise and apply novel technologies that take advantage of the growing body of sequence information to analyse the full complement of genes and proteins encoded by an organism *(Fields et al, 1999)*. One of the first steps in determining possible functions of novel proteins is identifying their interactions with other proteins. As already mentioned, there has been an explosion of novel technologies that facilitate the analysis of protein-protein interactions and other molecular interactions of biological significance. The ultimate goal is to apply these technologies for the elucidation of protein-protein interaction networks and the generation of proteomic interaction/protein linkage maps for entire organisms *(Bartel et al, 1996)*. The integration of protein interaction maps with genomic sequence information will greatly advance our understanding of complex biological processes and will help to establish an image that accurately reflects how individual genes and their protein products interact to create the complex world of biological specialisation. In addition to this it will have a dramatic effect on medicine through the identification of genes and mutations involved in disease and by greatly accelerating the discovery of novel therapeutic targets.

1.3: Identification And Analysis Of Protein-Protein Interactions.

Traditionally protein-protein interactions were studied using physical biochemical methods such as affinity chromatography, affinity blotting and co-immunoprecipitation *(methods reviewed in Phizicky and Fields, 1995)*. More recent strategies have taken advantage of genetic methods to isolate genes encoding polypeptides that interact with a protein of interest. Such strategies include the screening of protein expression libraries with labelled protein probes and the use of phage display *(Smith, 1985)*, *E. coli* display *(Wilson and Beveridge, 1993; Lu et al, 1995)* and Yeast display *(Boder and Wiltrup, 1997)* techniques.

The phage display technique was expanded upon with the development of Selective Infective Phage (SIP) technology *(Spada and Plunkthun, 1997)*. All of these methodologies essentially detect protein-protein interactions *in vitro*.

The advent of the yeast two-hybrid system (Y2H) has accelerated the pace of *in vivo* interaction analysis and few would dispute that of all of the methods currently available, the Y2H system is a powerful and efficient method for the identification and analysis of protein-protein interactions. Since its original description by *Fields and Song in 1989* the Y2H system has been used extensively to identify and study protein-protein interactions from a wide range of organisms and a very large number of reviews describing the system have been published *(Fields and Sternglanz, 1994; Allen et al, 1995; Luban and Goff, 1995; Bartel and Fields, 1995; Bai and Elledge, 1996; Brent and Finley, 1997; Warbrick, 1997)*. The system is easy to use and results in the immediate availability of the cloned genes for any proteins identified as interacting with a protein of interest *(Chien et al, 1991)*. These features, along with the fact that interactions are detected *in vivo,* which abrogates the need for purification of proteins, guarantee the continued popularity of the Y2H system as we enter the age of functional genomics.

1.4: The Yeast Two-Hybrid System.

The conceptual basis for the Y2H system was established from the accumulated understanding of the factors involved in the regulation of eukaryotic gene expression.

1.4.1: Gene Expression.

The entire complement of genes in a single cell represents a staggering amount of biological information. Some of this information is needed by the cell all of the time; for example, most cells continuously synthesis ribosomes and thus there is a continuous requirement for transcription of rRNA and ribosomal protein genes. Similarly, genes coding for enzymes such as DNA polymerase or those involved in basal metabolic pathways are active all of the time. These genes are sometimes referred to as housekeeping

genes reflecting the nature of the biological information that they encode. On the other hand, many genes have a more specialised role and expression of their biological information depends on the cell growth rate or environmental conditions.

Control of gene expression is in essence control of the amount of gene product in the cell. This amount is a balance between two factors:

a.) The rate of product synthesis **b.)** The rate of product degradation

The result of this balance is a steady state concentration for each gene product in the cell. If either the rate of synthesis or degradation of a particular gene product changes then the steady state concentration of that gene product will change. In practice critical variations in the steady state concentration of a gene product arise due to changes in the rate of product synthesis while the rate of product degradation seems to be relatively constant.

Synthesis of a particular gene product can be regulated by exerting control over any one or a combination of the various steps involved in the gene expression pathway;

- Transcription.
- mRNA Processing.
- mRNA Turnover.
- Translation.

Control of gene expression is highly complex involving each of the above strategies. However, the best understood mechanisms in both prokaryotes and eukaryotes depend solely on transcriptional regulation. This is not surprising, it makes sense that a cell should not waste energy by making more mRNA than it needs.

1.4.2: Transcriptional Regulation And DNA Binding Proteins.

As well as being the essential step in the conversion of the genetic information in the DNA into protein, the process of transcription is also the major point at which gene expression is regulated. In most cases gene regulation is achieved by activating or repressing the transcription of a particular gene in specific cell types or in response to a specific signal.

Once this has occurred, all of the other stages of gene expression follow and the appropriate protein is produced in a cell type specific or inducible manner.

Processes such as attenuation can influence the rate at which a particular gene is transcribed but specific DNA binding proteins called transcription factors primarily regulate transcription. These proteins bind to specific short DNA sequences that may lie within the promoter or further upstream from the promoter of the gene they are responsible for regulating *(Collado-Vides et al, 1991; Gralla and Collado-Vides)*. Transcriptional activator proteins usually bind to specific sequences that lie upstream of the promoters for genes they regulate. These binding sites are referred to as Upstream Activating Sequences (UAS).

The study of transcription factors is a critical aspect of gene regulation studies. A large number of transcriptional activator proteins have been studied and the majority of these have been found to have distinct functional domains;

- A DNA binding domain (DBD) responsible for binding of the transcription factor to its cognate DNA sequence.
- A transcriptional activation domain (AD) responsible for activation of transcription via interactions with core RNA polymerase.

The important feature of some of these transcription factors such as Gal4 *(Keegan et al, 1986)* and GCN4 *(Hope and Struhl, 1986)* is that these functional domains can be physically separated from each other and still retain their functionality i.e. the domains can function independently of each other. The truly modular nature of transcription factors was demonstrated by an experiment in which a hybrid LexA-Gal4 fusion protein was found to activate transcription in yeast containing a reporter gene under the control of LexA operator sequences *(Brent and Ptashne, 1985)*. In the case of this fusion protein the DNA binding activity was provided by the LexA component and the activation function by the Gal4 component. Not only can heterologous DNA binding and activation domains function as a single fusion protein, but these two domains can also function when non-covalently joined via protein-protein interactions. This was first demonstrated for the naturally occurring protein Oct-1, which possesses a DNA binding domain that locates it to certain promoters. However, it only activates these promoters when it is further complexed with a

transcription activation domain provided by the Herpes Virus protein, VP16 *(McKnight et al, 1987)*. This phenomenon is now called recruitment. At about the same time, *Ma and Ptashne (1988)* demonstrated that a DNA-bound Gal4 derivative, which could not activate transcription, did activate transcription when it recruited a chimeric protein that contained a Gal4-interacting protein (Gal80) and a wild type Gal4 transcriptional activation domain. These observations formed the basis for the Y2H System.

1.4.3: Principle Of The Two-Hybrid System.

It was independent work by Fields and Song *(1989)* that resulted in the first Y2H system which takes advantage of the modular domain structure of eukaryotic transcription factors. The system utilises hybrid genes to detect protein-protein interactions by means of activation of reporter gene expression. Two putative protein partners, X and Y, are genetically fused to a DNA binding domain (DBD) and a transcriptional activation domain (AD), respectively. These hybrid genes are then co-expressed in a suitable yeast strain that contains a reporter gene (usually *lacZ* or a nutritional marker) that is under the transcriptional control of a specially constructed promoter bearing cognate DBD binding sequences. Productive interactions between the two proteins of interest brings the transcriptional AD into proximity with the DNA bound DBD and triggers the transcription of the adjacent reporter gene resulting in a screenable phenotype (see Figure 1.4.1).

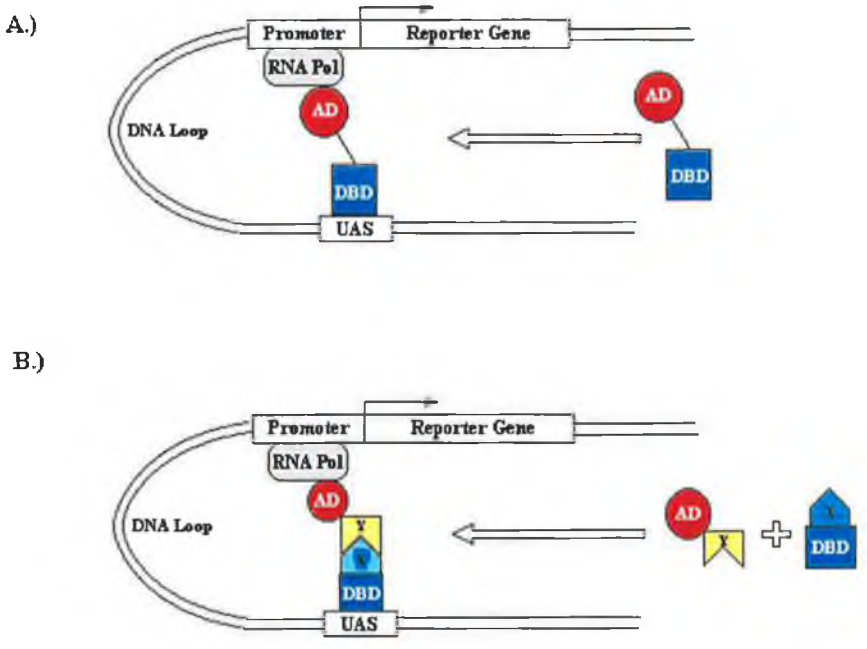
1.4.4: Components Of The Y2H System.

The initial Y2H system devised by Fields and Song was based on the Gal4 transcriptional activator from the yeast *Saccharomyces cerevisiae*. This protein consists of 881 amino acids *(Laughon and Gesteland, 1984)* and is required for the expression of the genes involved in galactose utilisation *(Johnston, 1987)*; gal1, gal2, gal7 and gal10. The Gal4 protein consists of two separable and functionally essential domains *(Keegan et al, 1986; Ma and Ptashne, 1987a)*;

- An N-terminal DNA binding domain (amino acids 1-147) which binds to a specific DNA sequence found upstream of promoters for genes involved in galactose utilisation such as the gal1 gene.
- A C-terminal domain (amino acids 768-881) which is necessary for activation of transcription.

Fields and Song demonstrated the system using two yeast proteins known to interact with each other, SNF1 and SNF4. Using the sequences for the Gal4 domains and the sequences for SNF1 and SNF4 they constructed two plasmids, pEE5 and pNI12.

- **Figure 1.4.1: Schematic Diagram Depicting The Activation Of Transcription In A Typical Eukaryotic System And In The Two Hybrid System.**



(A) Activation of transcription by a typical eukaryotic transcriptional activator with separable DNA binding domain (DBD) and activation domain (AD). **(B)** Activation of transcription by chimeric proteins in a two hybrid system. X represents a given protein fused to a specific DBD. In library screens this protein is termed the 'bait'. Y represents a given protein, or a pool of proteins encoded by a DNA library fused to a transcriptional activation domain. This fusion protein is often termed the 'prey'. If X and Y bind to each other, the AD is brought to the vicinity of the DNA bound DBD and transcription is activated from the adjacent promoter.

Plasmid pEE5 expressed a hybrid protein containing the Gal4 DBD (aa's 1-147) fused N-terminally to SNF1. Hybrid proteins, such as this one, bearing DBD's are now commonly referred to as Bait proteins and the plasmids expressing these proteins are referred to as Bait vectors. The Gal4 DBD contained its own nuclear localisation signal *(Silver et al, 1984)* which meant that the hybrid protein would be localised to the nucleus and the plasmid also had a HIS3 nutritional marker gene to allow independent selection for the plasmid.

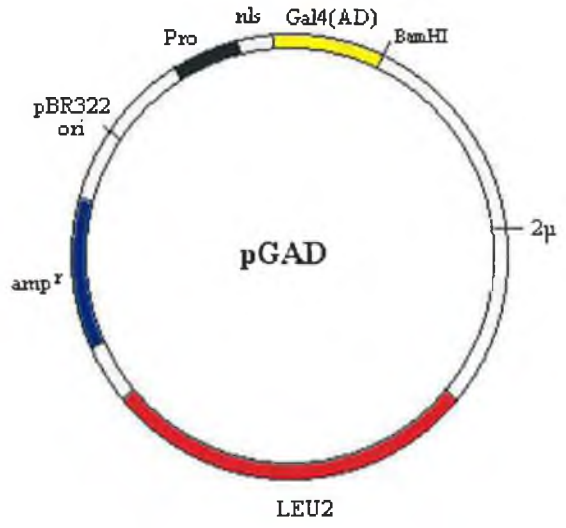
Plasmid pNI12 expressed a hybrid protein containing the SNF4 protein fused N-terminally to the Gal4 AD (aa's 768-881). Hybrid proteins, such as this one, bearing AD's are now often referred to as Prey proteins and the plasmids expressing them are called Prey vectors. This vector contained a LEU2 nutritional marker gene to allow independent selection for the plasmid.

The two plasmids were co-transformed and expressed in the yeast strain GGY1::171 *(Gill and Ptashne, 1987)* which was deleted for both the Gal4 and Gal80 (a negative inhibitor of Gal4). The strain also had a fusion between a *gal1* promoter and a *lacZ* gene integrated into the chromosome and mutations in the HIS3 and LEU2 genes to allow positive independent selection for each of the plasmids. Interaction of the Gal4(DBD)-SNF1 and SNF4-Gal4(AD) was determined on the basis of β-galactosidase activity using X-gal.

The Prey vector used in the initial experiment was later improved upon by *Chien et al (1991)*. In the original paper by *Fields and Song (1989)*, the Gal4 AD was fused C-terminally to SNF4. Although the C-terminal location for the Gal4 activation domain corresponded to its normal position within the Gal4 protein, this location was inconvenient for the construction of libraries of Prey (Activation Domain) hybrid proteins. It was therefore investigated whether the Gal4 AD could function at the N-terminus of hybrid proteins. This required the Gal4 termination sequence to be eliminated by mutagenesis and because the C-terminal DNA sequence for the Gal4 AD lacked its own endogenous initiation codon and nuclear localisation signal sequence, these had to be provided. A heterologous signal sequence consisting of the sequence for the first five codons of the SV40 large T-antigen was used. Three new Prey vectors called pGAD1, pGAD2 and

pGAD3, which differed from each other only in the reading frame of the unique *Bam*HI site, were constructed (see Figure 1.4.2).

- **Figure 1.4.2: pGAD Prey Vector Constructed By *Chien et al 1991.***



Pro = Promoter
nls = Nuclear Localisation sequence

Vector	Reading Frame Of BamHI Site
pGAD1	GGA TCC
pGAD2	XGG ATC CXX
pGAD3	XXG GAT CCX

Using the pGAD3 vector *Chien et al* constructed a hybrid between the Gal4 AD and the yeast SNF4 protein. The hybrid Gal4(AD) protein was tested for its ability to interact with originally constructed Gal4(DBD)-SNF1 hybrid which was expressed from the pEE5 Bait vector. It was found that the Gal4 activation domain could function at the N-terminus of hybrids since significant *gal1-lacZ* activity, comparable to that obtained with the original SNF4 hybrid, was obtained. This demonstrated that the pGAD vectors were suitable for testing defined protein-protein interactions.

Placement of the Gal4 AD at the N-terminus meant that the same vector encoded yeast promoter, initiation codon and nuclear localisation signal sequence could be used for all gene fusions. This facilitated easier construction of libraries of AD hybrid proteins by allowing a single fusion joint between the Gal4 AD and heterologous proteins to be used. In performing library screens a library of AD hybrids would be constructed by cloning a cDNA library into the unique *Bam*HI site of one of the pGAD vectors. However, only a

small proportion of a cloned cDNA population would be in frame with the upstream Gal4 AD sequence. It was for this reason that *Chien et al (1991)* constructed three versions of the pGAD vector which differed only in the reading frame of the unique *Bam*HI site with respect to the upstream Gal4 AD sequence, one vector for each possible reading frame. The same cDNA population would therefore be cloned into each of the three pGAD vectors to ensure that an in frame fusion between every cDNA in a population and the Gal4 AD would be obtained. *Chien et al (1991)* demonstrated the usefulness of the system for performing library screens using the *Sacchromyces cerevisiae* protein Sir4 that plays a role in determining yeast mating type. They transformed a yeast strain that expressed a bait protein, Sir4 fused to the Gal4 DBD, with pGAD based libraries expressing yeast proteins fused to the Gal4 AD and identified transformants that turned blue on the indicator X-gal plates. While some of the identified library plasmids in the blue colonies encoded Gal4, which bound to the Gal4 sites on the reporter, others encoded proteins that interacted with Sir4. These included Sir4 itself (as expected from previous experiments) and a new Sir4-interacting protein, Sfi1.

The work done by *Chien et al (1991)* defined the basic structure for Y2H system vectors and although many improvements have been made to the Y2H system vectors since then, primarily through the expansion of multiple cloning sites (MCS's), the basic structure of Y2H system vectors has remained the same. Several laboratories have constructed Y2H system vectors (see Table 1.4.1) *(Bartel and Fields, 1995)*. *Durfee et al (1993)* constructed vectors using the Gal4 DBD and the Gal4 AD, *Vojtek et al (1993)* constructed a system employing the LexA DBD and the Herpes virus VP16 AD and *Gyuris et al (1993)* constructed a system employing the bacterial LexA DBD and the *E. coli* B42 AD. The system developed by *Gyuris et al (1993)* and further developed by *Golemis et al (1997)* is often referred to as the interaction trap assay and is commercially available from Invitrogen Corporation. Because it employs the LexA DBD, and thus reporter genes controled by LexA binding sites, it has several advantages over the standard two-hybrid systems that employ Gal4 DBD's. First, it is not necessary to use *gal4* deleted strains which are often unhealthy. Second, in the case of library screens using yeast proteins, it is not necessary to remove false positives caused by re-cloning of Gal4.

As with the initial pGAD vectors, many of the newly constructed commercially available vectors have been constructed with the MCS's in all three reading frames. In some instances, such as the pGAD424 Prey and pGBT9 Bait vectors *(Roder et al, 1996)* which are available from Clontech, the versatility of the vectors has been further increased by also making both vectors with inverted MCS's in all three reading frames.

Table 1.4.1:Vectors For The Yeast Two Hybrid System *(Bartel and Fields, 1995)*

Plasmid	Domain	Restriction Sites	Marker
		DNA-Binding Domain Vectors	
pMA424Gal4 bd		*Eco*RI, *Sma*I, *Bam*HI, *Sal*I, *Pst*I	HIS3
pGBT9	Gal4 bd	*Eco*RI, *SmaI*, *Bam*HI, *Sal*I, *Pst*I	TRP1
pAS1	Gal4 bd	*Nde*I, *Nco*I, *Sfi*I, *Sma*I, *Bam*HI	TRP1
pAS2	Gal4 bd	*Nde*I, *Nco*I, *Sfi*I, *Sma*I, *Bam*HI	TRP1, CYH2
pCP62	Gal4 bd	SalI, PstI, SmaI, SpeI, XbaI, NotI, SacII	LEU2
pBTM116	LexA	*Eco*RI, *Sma*I, *Bam*HI, *Sal*I, *Pst*I	TRP1
lex(1-202)PL	LexA	*Eco*RI, *Sma*I, *Bam*HI, *Sal*I, *Pst*I	HIS3
		Activation Domain Vectors	
pGAD.F	Gal4 ad	*Bam*HI	LEU2
pGAD424	Gal4 ad	*Eco*RI, *Sma*I, *Bam*HI, *SalI*, *Pst*I, *Bgl*II	LEU2
pGAD.GH	Gal4 ad	*Eco*RI, *Sma*I, *Bam*HI, *Sal*I, *Spe*I, *Xho*I	LEU2
pACT	Gal4 ad	*Eco*RI, *Bam*HI, *Xho*I, *Bgl*I	LEU2
pGADNOT	Gal4 ad	*Bam*HI, *Not*I, *Sal*I	LEU2
pPC86	Gal4 ad	*Sal*I, *Sma*I, *Eco*RI, *Bgl*II, *Spe*I, *Not*I	TRP1
pJG4-5	B42	*Eco*RI, *Xho*I	TRP1
pVP16	VP16	*Bam*HI, *Not*I	LEU2
pSD.10	VP16	*Eco*RI, *Bst*I, XhoI	URA3

The most significant improvements to the Y2H system since its initial description have been concerned with the development of new reporter gene systems. Reporters derived from the yeast LEU2 *(Gyuris et al, 1993)* and HIS3 *(Vojtek et al, 1993; Durfee et al, 1993)* genes were developed and allow direct selection for yeast cells expressing interacting hybrid proteins. This has greatly facilitated library screens by allowing higher plating densities which increases the probability of detecting low abundance interacting partners in large complex libraries.

Reporter genes can either be integrated into the chromosome or maintained on independently replicating plasmid vectors. For example, the plasmid vectors pSH18-34, pRB1840 and pJK103 *(Golemis et al, 1997),* which are available from Invitrogen Corporation, have *lacZ* reporter constructs on them. These can be transformed into yeast strains bearing integrated auxotrophic reporters, such as strain *EGY48* and *EGY191 (Golemis et al, 1997)* available from Invitrogen (see Table 1.4.2 below), to allow dual reporter selection. The location of the reporter gene has a significant affect on the sensitivity of a system. Phenotypes derived from integrated reporters tend to be weaker than those derived from plasmid borne reporters and thus systems employing integrated reporter genes tend to be less sensitive to weak interactions. However, most systems utilise specific yeast reporter strains with integrated reporter constructs (see Table 1.4.2). The sensitivity of the assay is manipulated by varying the number of UAS's present in the reporter constructs (strains *EGY191* and *EGY48* in Table 1.4.2; *EGY191* has a LEU2 reporter with two LexA operator sites while *EGY48* has a reporter with six LexA operator sites). The greater the number of UAS sites the more sensitive the system but it is important to note that increasing the sensitivity of the assay also has the effect of increasing the number of false positives that are obtained when performing library screens. The reasons for false positives and methods employed to eliminate them are the discussed later.

Many systems now employ dual reporter selection (see strains *L40, Y190* and *Hf7c* in Table 1.4.2). This simply means that they use two distinct reporter genes, usually *lacZ* and HIS3 *(Vojtek et al, 1993; Durfee et al, 1993)* or LEU2 *(Gyuris et al, 1993; Golemis et al, 1997)*, simultaneously. In some yeast reporter strains the two reporter genes employed are driven by unrelated promoters that have had UAS's for a Gal4 or a LexA DBD incorporated upstream of their transcriptional start sites (see strain *Hf7c* in Table 1.4.2: strain utilizes a

gal1 promoter:HIS3 reporter and an unrelated *CYC1* promoter:*lacZ* reporter). *Huang and Schreiber (1997)* produced two such strains by introducing a second URA3 reporter gene under the transcriptional control of a SPO13 promoter with four upstream LexA operator sites into strains *EGY48* and *EGY191*. The use of dual reporter strains like these where individual reporter genes are driven by unrelated promoters can help to significantly reduce the number of false positives detected while performing library screens. This is discussed in greater detail later in this literature review.

Table 1.4.2: Reporter strains For Two-Hybrid Systems (*Bartel and Fields, 1995*)

Strain	*Reporter Genes*	*Markers*
GGY1::171	*gal1-lacZ*	*his3, leu2*
PCY2	*gal1-lacZ*	*his3, leu2, trp1*
Y526	*gal1-lacZ*	*his3, leu2, trp1*
Y527	*CYC1-lacZ*	*his3, leu2, trp1*
Y190	*gal1-lacZ* *gal1*-HIS3	*leu2, trp1*
Hf7c	*CYC1-lacZ* *gal1*-HIS3	*leu2, trp1*
L40	4lexAop-HIS3 8*lexA*op-*lacZ*	*his3, leu2, trp1*
EGY48	6*lexA*op-LEU2	*his3, leu2, trp1*
EGY191	2*lexA*op-LEU2	*leu2, trp1, ura3*

The availability of a wide range of systems employing a wide range of different reporters and exhibiting varying levels of sensitivity is of great benefit to an experimenter, allowing the selection of a system exhibiting the characteristics and the degree of sensitivity best suited to a specific application. The ease of performing library screens has also been significantly increased by the commercial availability of pre-constructed Prey hybrid protein libraries for specific cell types, tissues and organisms from companies such as Invitrogen Corporation.

1.4.5: Why Is The Two-Hybrid System So Sensitive?

The two-hybrid system is an extremely sensitive method for the detection of protein-protein interactions. It is capable of detecting *in vivo* binding that is often beyond the limits of *in vitro* detection methods. For example, the interaction of the mammalian Ras protein with the protein kinase Raf was observed in the Y2H system but had not been detected by co-immunoprecipitation *(Aelst et al, 1993)*. Protein affinity chromatography is the most sensitive of the physical biochemical methods for detecting protein-protein interactions. With appropriate use (high concentrations of immobilised test protein), it can detect interactions with a binding constant as weak as 10^{-5}M *(Formosa et al, 1991)*. On the basis of binding of different proteins to the retinoblastoma protein, *Durfee et al (1993)* estimated that the minimal binding constant required to detect an interaction in their version of the Y2H system was on the order of 1μM (10^{-6}M). The sensitivity of the Y2H system depends on;

- the number of UAS's in the reporter gene construct
- the level of expression of hybrid proteins.
- the amount of reporter protein required for a detectable phenotype.

As mentioned earlier, the greater the number of UAS's placed upstream of the reporter gene construct the more sensitive the system. Most commercially available systems allow selection of the number of UAS's in reporter constructs and thus the selection of a system exhibiting the degree of sensitivity best suited to a specific application. Hybrid proteins are usually over-expressed from strong promoters on high copy number vectors as this favours complex formation and increases the possibility of detecting weak protein-protein interactions. In addition to this, each reporter mRNA transcript induced by productive protein-protein interactions can be translated many times to generate several stable enzyme molecules. This type of amplification leads to a detectable signal even when the initiating interaction is weak.

Physical methods used for the detection of protein-protein interactions, such a co-immunoprecipitation, generally rely on a low rate of dissociation since complexes must survive several washes. This diminishes the signal but is necessary to prove specificity.

This is not the case with Y2H system. In addition to this, the stability of the hybrid protein complex is probably enhanced by the interaction of the activation domain with proteins from the transcription initiation complex that also associates with promoter DNA. This ternary complex formation increases the overall stability of the complex on the DNA and further increases the probability of detection.

Sensitivity of a given system also depends on the strength of the activation domain being used. The Prey vector used by *Gyuris et al (1993)* uses the relatively weak activation domain of *E. coli* B42. While this activation domain is weak, it may increase the spectrum of protein-protein interactions that can be detected by eliminating the toxic effects (called squelching) that strong activators can have in yeast *(Gill and Ptashne, 1988).* The system developed by *Durfee et al (1993)* uses the stronger Gal4 AD while the system employed by *Vojtek et al (1993)* uses the even stronger VP16 AD.

The combination of the above factors generates the remarkable sensitivity of Y2H systems. In fact, the very sensitivity of these systems may present a problem when applying secondary biochemical assays such as co-immunoprecipitation or affinity chromatograph to examine the authenticity of the *in vivo* signal.

1.4.6: Advantages And Applications Of The Yeast Two-Hybrid System.

The Y2H system has many advantages over biochemical and classical genetic methods for the detection and analysis of protein-protein interactions. As already discussed the system is highly sensitive, detecting interactions that are not detected by other methods. As the assay is carried out *in vivo* the need for the detailed and laborious manipulation of conditions necessary for *in vitro* biochemical binding assays is abrogated and detected interactions are more physiologically relevant. In addition to this there is no need for biochemical purification of hybrid proteins.

The system can be used to screen libraries of activation domain hybrids to identify proteins capable of binding to a protein of interest. The use of well characterised genetic backgrounds in yeast hosts allows direct selection for interacting proteins through the use

of appropriate reporter genes like HIS3 *(Vojtek et al, 1993; Durfee et al, 1993)* and LEU2 *(Gyuris et al, 1993; Golemis et al, 1997)* rather than the laborious screening procedures required for biochemical methods. This direct selection allows high plating densities which increases the probability of detecting rare interacting partners in large complex libraries. It also makes Y2H systems simple to use and allows library screens to be performed very rapidly. These screens result in the immediate availability of the cloned gene for any new protein identified *(Chien et al 1991)* and this makes further analysis of identified protein-protein interactions more feasible. Deletion mutants can be made to delineate domains important for interaction. This approach was employed by *Holt et al (1994)* to identify the 192 residue domain of the p85 subunit of phosphatidylinositol-3-kinase required for interaction with the p110 subunit of the enzyme. Once specific interacting domains have been identified point mutations may be made to identify specific residues that are important for interaction. This approach was employed by *Li and Fields (1993)* for the identification of residues important for the interaction of p53 with the SV40 large T antigen. Alternatively, the genes for identified proteins can be cloned into specialised vectors systems that simplify purification of the proteins for analysis by *in vitro* biochemical methods. Several such systems have been described and in each case the protein of interest is genetically fused to a protein or a domain that can be easily purified on an appropriate affinity resin. The most common such fusion contains glutathione S-transferase (GST), which can be purified on glutathione-agarose columns *(Smith and Johnson, 1988)*.

Transcriptional activity can be proportional to binding affinity and so the Y2H system could be used to evaluate different binding affinities of mutant versions of either the bait or prey proteins. For example, with a reporter gene like the HIS3 gene the competitive inhibitor of the HIS3 gene product, 3-aminotriazole, can be used to directly select for constructs which yield increased affinity. The presence of increasing concentrations of 3-aminotriazole requires higher levels of HIS3 transcription *(Durfee et al, 1993)*. In a similar way, 6-azauracil can be used with a URA3 reporter as demonstrated in a version of the Y2H system using the DNA binding domain of the estrogen receptor *(Le Douarin et al, 1995)*. The sensitivity of the LEU2 promoter can also be adjusted by adding 6-fluoroleucine to the medium *(Brent and Finley, 1997)*. The Y2H system could also be used to establish dissociation values. This would first require the generation of a standard curve

based on biochemically determined KD values of existing mutants. The system was used in this manner for the p53-SV40 large T antigen and a good correlation was observed between transcription level and affinity *(Li and Fields, 1993)*. It is important to note however that the concentration of hybrid proteins, the stability of the hybrid proteins, the extent of nuclear transport, accessibility of protein domains to each-other and the accessibility of the activation domain to the transcription machinery can all affect the level of transcriptional activity.

The system is widely applicable and has been used for the analysis of a wide variety of proteins including those that normally reside in the nucleus, cytoplasm, mitocondria, in addition to membrane associated and extracellular proteins. Proteins can be targeted to the nucleus by nuclear localisation signals provided on the vector, such as the nuclear localisation signal of the SV40 large T antigen *(Chien et al 1991)*. However the fact the proteins must be targeted to the nucleus may limit the use of the system for certain extracellular or membrane bound proteins. This is discussed in more detail later.

1.4.7: Application Of The Two Hybrid System To Genomics.

As mentioned earlier attempts are being made to apply the two-hybrid system at a genomic level for the characterisation of protein-protein interaction networks and the construction of proteomic linkage maps. Such protein linkage maps have been successfully completed for small genomes like that of the bacteriphage T7 which encodes 55 proteins *(Bartel et al, 1996)*, and for subsets of eukaryotic proteins like the *Drosophila* cell cycle regulators *(Finley and Brent, 1994)*, the proteins of the pheromone response pathway in *S. cerevisiae (Evangelista et al, 1996)* and the yeast RNA polymerase III subunits *(Flores et al, 1999)*. The current challenge is the systematic application of two-hybrid screens to every protein expressed in eukaryotic cells. Two different strategies are being pursued to apply the two-hybrid screen to the entire complement of proteins ('proteome') of the yeast *S. cerevisiae* whose entire genome sequence is available.

Fromont-Racine et al (1997) have set out on an iterative approach using high frequency mating to screen small sets of related proteins against a high quality genomic library. DBD fusions and the library of AD fusions are expressed in haploid yeast strains of the opposite mating type, mated on filters and plated on selective media to identify diploids positive for two-hybrid interactions. Proteins identified in this way are then used as baits in the next round of screens. A recent addition to this strategy is the attempt to simultaneously identify all two-hybrid positives from a screen by hybridisation of total plasmid DNA to DNA arrays *(Cho et al, 1998).*

Hudson et al (1998) have taken advantage of the availability of the sequences for the approximately 6,000 open reading frames (ORF's) in the yeast genome. Each ORF can be expressed as a DBD fusion in haploid yeast of one mating type and then mated to an ordered array of approximately 6,000 test strains of the opposite mating type that express the complete set of ORF AD fusions. Advantages of the array approach are that the process is easily automated, requires no large-scale sequencing, and has built in controls for specificity of interactions. It is unknown, however, how many interactions may be missed, so called 'false negatives', by using full length ORF fusions because of misfolding, toxicity or steric inhibition.

1.4.8: Problems Associated With The Yeast Two-Hybrid System.

While the two-hybrid system is a very useful genetic technique for the identification and analysis of protein-protein interactions, its application does have some limitations. These limitations include the failure to detect expected interactions, so called 'false negatives', and the detection of large numbers of interactions of no biological significance called 'false positives'.

Failure to detect expected interactions ('false negatives') can occur for any one of a number of reasons. Hybrid proteins generated in the Y2H system have to be targeted to the nucleus where they interact with the promoter of the reporter construct. If the DNA-binding domain or the activation domain of the transcription factor lack a nuclear localisation signal, the

vector usually includes a heterologous nuclear localisation signal sequence such as that of the SV40 large T antigen *(Chien et al 1991)*. The fact that hybrid proteins are targeted to the nucleus means that interactions that are dependent on post-translational modifications that occur within the endoplasmic reticulum (glycosylation and disulfide bond formation) may not be detected. Interactions that require other modifications (phosphorylation or acetylation) by non-yeast proteins may also fail to be detected. For example, proteins with SH2 domains require phosphotyrosine in the proteins to which they bind and this phosphorylation may not occur in yeast. The Y2H system may not therefore be suitable for certain extracellular and plasma membrane proteins *(Fields and Sternglanz, 1994)*.

False negatives can also occur because some hybrid genes are harmful or lethal when expressed in yeast. The introduction of certain genes into yeast may interfere with normal cellular events to the point of non-variability. Essential cellular systems that rely on a balance of positive and negative regulators may be particularly susceptible to interference. One example of this was lethality owing to over-expression of cell cycle regulators cyclin A and cyclin B *(Allen et al, 1995)*.

Proteins must be capable of folding and of stable existence in yeast cells. Western blotting of yeast lysates can usually be used to monitor this. If antibodies are not available for the protein of interest, commercially available antibodies for the transcription factor domains can be used e.g. anti-Gal4 antibodies. Alternatively some vectors are designed with epitope tags such as the V5 epitope tag *(Durfee et al 1993)*, to facilitate monitoring expression of hybrid proteins.

Proteins also have to retain their activity as fusion proteins and the use of hybrid proteins also means that the site of interaction may be occluded by the transcription factor domains *(Fields and Sternglanz, 1994; Beranger et al, 1997)*. Proteins under investigation are usually expressed as C-terminal fusions to AD's or DBD's. This type of construction jeopardises interactions that require a 'free' N terminus for a protein to interact with its molecular partners because the transcription factor domain blocks accessibility. In such cases the orientation of the hybrid could be reversed. This was demonstrated by *Beranger et al (1997)* who, by using N-terminal fusions to the LexA DBD as baits, detected the interaction between the negative Ral (N28) mutant and RalGDS which was previously undetected when C-terminal fusions to the LexA DBD were used. They also document two

cases where using N-terminal fusions exhibited enhanced sensitivity compared to the standard orientation, i.e. the interaction of the Ras negative mutant (N17) with CDC25 and the interaction of the wild type positive Ras (V12) with CDC25. In some cases it might be possible to overcome the occlusion of a site of interaction by swapping transcription factor domains; the protein (X) fused to the DBD could instead be fused to the activation domain of the transcription factor. In assays of defined protein combinations one orientation of the hybrids often activates transcription much more efficiently. This may reflect differences between the expression and/or stability of hybrids containing protein X and those containing the interacting protein Y.

Failure to detect an expected interaction when performing library screens may simply be because the gene encoding an interacting protein may simply not be represented in the library due to a low complexity or a bias in the representation of that clone *(Vidal and Legrain, 1999)*. In many screens cDNA's are derived from random or oligo(dT) primed RNA's. It must be kept in mind that unlike genomic libraries, the relative representation of each cDNA closely reflects the endogenous expression level of the corresponding gene. Thus, interesting proteins might be under-represented if their RNA is expressed at relatively low levels. One solution to this problem is the use of normalised AD libraries. The process of normalising cDNA libraries consists of reducing the representation of highly expressed cDNA's *(Soares et al, 1994)*. In contrast, the complexity of genomic libraries is directly correlated to the number of independent clones that compose the library and to the size of the genome. For organisms that are encoded by compact genomes, i.e. with small intergenic sequences and few introns, screening a genomic library instead of a cDNA library for two-hybrid experiments is advantageous. Screening genomic libraries allows fragments of proteins to be assayed. This is particularly useful since some full-length proteins may be toxic when expressed in yeast.

When performing library screens the bait vector and the AD library are often co-transformed. A problem that is commonly encountered with library screens is achieving optimal transformation efficiency. To overcome this problem some researchers transform the yeast with the bait vector first and then subsequently transform with the AD library plasmids. Other workers transform the bait vector into yeast of one mating type and the

library into yeast of the opposite mating type. The two strains are then mated and the bait and library plasmids are brought together in the diploid cells *(Bartel et al, 1996)*.

Proteins that are involved in transcriptional activation activate transcription of the reporter gene construct when fused to a DBD. However, some proteins that do not normally activate transcription can do so when expressed as DBD hybrid proteins, presumably because these proteins have domains resembling activation domains *(Ma and Ptashne, 1987b)*. Libraries are thus constructed in the activation domain plasmids to avoid detection of random sequences that activate transcription when fused to a DBD *(Chien et al, 1991)*. The target DBD hybrid protein should be tested for its ability to auto-activate reporter gene transcription. If protein is being tested in a defined combination with a known partner and is found to auto-activate reporter gene expression when expressed as a DBD hybrid protein, it can instead be expressed as an AD hybrid. Alternatively, it is often possible to delete a small region of a protein that activates transcription thus eliminating the activation function while retaining other properties of the protein *(Iwabuchi et al, 1993)*. However, if such a deleted protein is being used as a bait, it must be remembered that residues not normally exposed in the intact protein may mediate interactions and thus that positive interactions with other proteins may be artifactual.

In screening libraries it is generally found that in addition to proteins encoded by the library that appear to be able to interact with the target protein a number of other classes of positives are obtained that are artifactual. There are several reasons for these 'false positives' and they can be easily eliminated by additional genetic assays *(Bartel et al 1993)*. These assays simply involve recovering the library plasmids from transformants that score positive in primary library screens and transforming these into *E. coli* for amplification. These plasmids are then re-introduced into the original screening strain in several different combinations.

One class of false positives obtained when performing library screens is capable of activating transcription of the reporter gene system in the absence of any bait hybrid. These library plasmids encode Gal4 or other proteins that apparently bind to sequences in other proteins bound to the promoter used to drive expression of the reporter gene *(Chien et al 1991)*. These false positives are eliminated by re-transforming isolated library plasmids alone and eliminating clones in which the reporter gene is activated. False positives

resulting from the re-cloning of Gal4 are not encountered in the interaction trap assay, a variation of the two-hybrid system developed by *Gyuris et al (1993)* and employing the LexA DBD.

Other false positives may arise if a library plasmid encodes a protein that interacts with the DBD of Gal4. Re-transforming the library plasmid with the Gal4 DBD vector alone and checking for reporter activation eliminates these false positives.

Another class of false positives initiate transcriptional activity of the reporter gene when introduced into a yeast cell with a DBD hybrid consisting of the DBD of Gal4 and the target protein used in the initial screen but fail to do so with the Gal4 DBD alone *(Bartel et al, 1993)*. The inability of these positives to produce a signal with the DBD alone would appear to indicate specificity but when tested with hybrids comprised of the DBD and a totally extrenuous protein they also produce a signal. One explanation for these false positives is that these proteins are themselves transcriptional activators that bind to the promoter of the reporter gene construct (or to proteins already bound there) but require the presence of an inactive hybrid bound to the Gal4 binding sites in order to function. The binding of this inactive hybrid might alter nucleosomes or chromatin structure in some fashion to allow these weak activators to bind or to trigger transcription. This class of false positives can be eliminated by introducing the isolated library plasmid into the original yeast test strain with;

- the Gal4 DBD alone,
- the DBD hybrid consisting of the DBD domain of Gal4 and the target protein used in the initial screen.
- hybrids comprising the DNA-binding domain and a totally extrenuous protein.

Plasmids that are positive with only the target hybrid protein are then introduced with the same set of DBD hybrids into a strain carrying the *lacZ* gene under the control of an unrelated promoter, such as the *CYC1* minimal promoter *(Bartel et al, 1993)*, that has had its own upstream activation sequence deleted and replaced with Gal4 binding sites. Using this reporter strain, false positives will not produce a signal with either the original target

protein, the extrenuous protein or with the Gal4 DBD alone. Although the *gal1* promoter strain appears to be approximately two fold more sensitive than the *CYC1* strain in X-gal filter assays, this class of false positives does not show a transcriptional signal in the *CYC1* strain, even when tested with twice the concentration of X-gal and four fold or longer incubation times. A true positive on the other hand will result in β-galactosidase activity with the original protein target. New reporter strains, like Hf7c (see Table 1.4.2), that contain two or more reporter genes with unrelated reporters help to eliminate many of these false positives in the initial assay. Another possible explanation for proteins capable of interacting with both the target hybrid protein and an extrenuous hybrid protein is that these library encoded proteins might bind to the Gal4 DBD itself. It might be that this DBD alone is not sufficiently stable in yeast to yield a signal and thus the requirement for a DBD hybrid to generate transcription of the reporter gene. However, no such proteins have been detected to date.

The above assays provide a set of rapid genetic criteria for the elimination of false positives detected in Y2H library screens. Library plasmids that pass all of the above criteria are considered true positives and can then be used for DNA sequence analysis.

Even if a given clone passes through all of the control experiments in yeast it is necessary to provide proof of direct interaction between the proteins in a biochemical assay. One common approach is to express the candidate cDNA as a fusion protein with glutathione S-transferase (GST) in bacteria. One then demonstrates that the bait protein is co-precipitated with the GST fusion protein bound to glutathione beads *(Durfee et al, 1993)*. The hybrid fusion proteins may subsequently be co-immunoprecipitated in the original yeast reporter strain but ideally one should show that the two proteins co-precipitate in their native genetic background.

A critical consideration in performing Y2H screens is whether true positives isolated in the system are actually representative of *in vivo* cellular interactions. The system tests every binary combination of protein-protein interaction. Two proteins found to interact very specifically and strongly with one another may never contact each-other in the cell because they are localised in distinct cellular compartments or because they are expressed at different developmental stages or in different cell types.

1.5: Variations Of The Yeast Two Hybrid System.

Since its original conception a large number of variations of the basic Y2H system have been developed (reviewed by *Brent and Finley, 1997; Fredrickson, 1998; Drees, 1999; Vidal and Legrain, 1999)*. Related systems which allow the *in vivo* detection of a wide range of other biologically significant molecular interactions have been developed. These include systems for the detection of DNA-protein interactions (one hybrid system), RNA-protein interactions (RNA based three-hybrid system) and small molecule-protein interactions (ligand based three-hybrid system). The original concept of the two-hybrid system has also been turned upside down with the development of the reverse one-hybrid and reverse two-hybrid systems that allow the identification of mutations, peptides or small molecules that abrogate macromolecular interactions. The original configuration of the system has also been modified to extend the range of possible protein-protein interactions that can be analysed. Three component systems have been developed to allow the *in vivo* detection of trimeric interactions, ligand-receptor interactions and interactions dependent on post-translational modifications. A description of these systems is given in sections 1.5.1 to 1.5.4.

In addition to the above, forward and reverse dual bait systems have also been developed and are very useful for functional dissection of proteins capable of interacting with multiple binding partners *(Serebriiskii et al, 1999)*. Other interaction screens based on reconstituting the activity of proteins other than transcription factors have also been developed which facilitate the identification and analysis of protein-protein interactions that are not conducive to analysis in the standard Y2H system *(Johnson and Varshavsky, 1994; Marsolier et al, 1997; Rossi et al, 1997; Aronheim et al, 1997; Broder et al, 1998)*

Mammalian two-hybrid systems *(Sadowski et al, 1992)* and prokaryotic two-hybrid systems have also been developed which provide alternative host genetic backgrounds to yeast. These permit the analysis of protein-protein interactions in genetic backgrounds more closely related to their native backgrounds. Mammalian systems can be particularly useful for the analysis of protein-protein interactions dependent on post-translational modifications that do not occur in yeast. A description of prokaryotic systems is given in section 1.5.5.

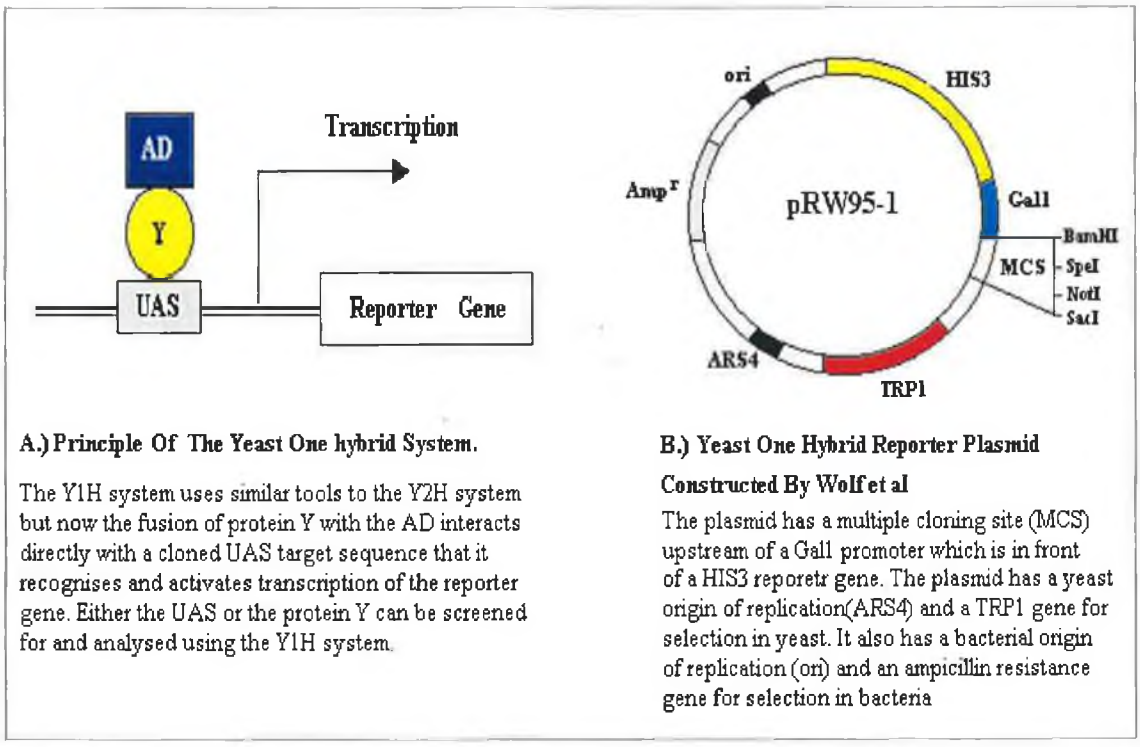
Systems such as those mentioned above greatly facilitate functional genomics by providing scientists with very useful tools for the identification of molecular interactions and for the generation of macromolecular interaction maps. A brief description of some of the systems mentioned above is given in the following sections. The 21st century will no doubt usher in many more novel ways to tap the vast resources of genetic information generated by genomic sequencing projects.

1.5.1: One-Hybrid Systems.

The yeast one-hybrid (Y1H) system is used for the identification and analysis of DNA-protein interactions that play important roles in several biological processes, most notably in the regulation of gene expression as described earlier in this literature review.

The Y1H system is based on principles similar to those of the Y2H system already described and consists of two components. The first of these is a reporter plasmid bearing a specially constructed reporter gene construct. This usually consists of a Gal1 promoter fused to a reporter gene such as *lacZ* or HIS3. The Gal1 promoter is deleted for its cognate upstream Gal4 DNA binding sequences. These are replaced by unique restriction sites into which target DNA binding sequences under investigation are cloned. Cloning target sequences into these restriction sites places reporter gene expression under the transcriptional control of the cloned target sequences. Once the target DNA sequences have been cloned into the reporter plasmid, the plasmid is usually integrated into the genome of a suitable yeast strain. The resulting yeast strain, with the integrated reporter plasmid and cloned target DNA sequences, is then transformed with the second component of the Y1H system, expression libraries identical to those used when performing Y2H library screens. These libraries consist of random proteins fused to a specific transcriptional activation domain like that of the Gal4 protein. Any expressed hybrid proteins capable of binding to the cloned target DNA sequences on the integrated reporter plasmid will activate transcription of the reporter gene resulting in a screenable phenotype (see Figure 1.5.1a).

- **Figure 1.5.1: The One Hybrid System.**



A.) Principle Of The Yeast One hybrid System.

The Y1H system uses similar tools to the Y2H system but now the fusion of protein Y with the AD interacts directly with a cloned UAS target sequence that it recognises and activates transcription of the reporter gene. Either the UAS or the protein Y can be screened for and analysed using the Y1H system.

B.) Yeast One Hybrid Reporter Plasmid Constructed By Wolf et al

The plasmid has a multiple cloning site (MCS) upstream of a Gall promoter which is in front of a HIS3 reporetr gene. The plasmid has a yeast origin of replication(ARS4) and a TRP1 gene for selection in yeast. It also has a bacterial origin of replication (ori) and an ampicillin resistance gene for selection in bacteria

The earliest description of the Y1H system was by *Wilson et al (1991)* who used the system to identify the DNA binding site for the nuclear receptor NGF1-B. Later, *Wang and Reed (1993)* used a Y1H system to clone the olfactory neuronal transcription factor Olf-1 and *Li and Herskowitz (1993)* used the Y1H system to isolate ORC6, a component of the yeast origin recognition complex. *Luo and Vijaychanders (1996)* constructed another Y1H system which they tested using the p53 cognate DNA binding sequences and *Wolf et al (1996)* constructed an improved Y1H system reporter plasmid, pRW95-1, which had an expanded multiple cloning site (MCS) and that could be used with commercially available activation domain libraries (see Figure 1.5.1b).

1.5.2: Reverse One and Two-Hybrid Systems.

Reverse hybrid systems permit the identification of mutations, peptides or small molecules that disrupt protein-protein and protein-DNA interactions. They employ all of the same components as the standard two-hybrid and one-hybrid systems with the exception of the

reporter genes employed. Reverse hybrid systems use reporter genes whose expression causes toxicity and lethality under specific growth conditions (counterselectable markers). Under these growth conditions induced reporter expression resulting from the interaction of two proteins expressed in the context of the Y2H system or the interaction of a DNA binding protein and its binding site in the context of the Y1H system is deleterious to growth. Dissociation of the interactions therefore provides a selectable growth advantage.

The first reverse one and two-hybrid systems were described by *Vidal et al (1996a)* and utilised URA3 as a reporter gene. In the case of the reverse two-hybrid system the reporter was placed under the transcriptional control of Gal4 DNA binding sequences. Productive protein-protein interactions between hybrid proteins bearing Gal4 AD's and Gal4 DBD's resulted in the expression of the URA3 gene (which is involved in uracil biosynthesis) and could be positively selected for by plating yeast cells on media lacking uracil. However, the URA3 encoded enzyme can also catalyse the transformation of 5-fluoroorotic acid (5-FOA) into a toxic compound and thus cells expressing the URA3 gene (Ura^+) are sensitive to 5-FOA (Foa^S). Thus this reporter gene also provided a means of positively selecting for dissociation of protein-protein interactions by simply plating yeast on media containing 5-FAO since only yeast cells in which the interaction of the hybrid proteins was abrogated would grow on such media (Ura^-, Foa^R).

Another version of the reverse two-hybrid system was described by *Leanna and Hannink (1996)*. Their version of the system employed a CYH2 gene as a reporter gene. This gene is responsible for the sensitivity of yeast to cycloheximide. As with the system devised by *Vidal et al (1996a)* described above, this system allowed for direct positive selection for disruption of a given protein-protein interaction by simply plating on media containing cycloheximide since only yeast cells in which the protein-protein interaction was disrupted would grow on such media.

Shih et al (1996) designed another system called the split-hybrid system. Although the system they designed took a different approach to the other systems described above it is essentially a reverse two-hybrid system. Unlike the other systems the system they designed didn't employ a toxic reporter gene. Their system made use of the *E. coli* Tn*10*-encoded *tet* repressor (TetR) operator system. They engineered LexA binding sites upstream of the TetR gene and placed TetR binding sites (tet operators) upstream from a nutritional

reporter gene, HIS3. Interaction of one protein fused to a LexA DBD with a second protein fused to a VP16 AD resulted in expression of the TetR gene. The TetR gene then bound to its *tet* operator sites preventing expression of the HIS3 gene and thus preventing growth on media lacking histidine.

Huang and Schreiber (1997) constructed a reverse two-hybrid system which utilised the LexA DBD and the B42 AD. Hybrid proteins were expressed from the *gal1* promoter and thus expression of the proteins could be induced by the addition of galactose. The use of a Gal inducible promoter required that the galactose utilisation pathway be intact in the yeast strain used for the system. This ruled out the use of yeast strains developed for Gal4 DBD based forward or reverse systems in which Gal4 and Gal80 are inactivated. A reporter construct consisting of the URA3 gene under the transcriptional control an SPO13 promoter bearing LexA operator sites was integrated into the *ura3* locus of strains *EGY48* and *EGY191 (Golemis et al 1997)*(see Table 1.4.2). The resulting strains therefore had two integrated reporter constructs, one URA3 reporter construct and the original LEU2 reporter which was also under the transcriptional control of LexA operator sites. The presence of two distinct reporter genes, having only the LexA operator sequences in common provided a means of reducing false positives obtained when performing two-hybrid screens as discussed earlier.

Once a protein-protein interaction has been identified by the standard two-hybrid system a great deal of additional work is required to further characterise the functional/biological relevance, structure, and regulation of the observed interaction. The most direct approach to determine the functional relevance of an identified interaction genetically correlates the physical interaction with a biological parameter: the physical interaction is dissociated and the consequences are analysed in a functional assay. Logically one would expect that if a newly identified interaction was critical for a function of interest, dissociation of the interaction would impair that function.

Conceptually, protein-protein interactions can be dissociated by mutations in either one of the interacting partners or by dissociating proteins, peptides or small molecules. Interaction defective alleles, mutant proteins specifically altered in their ability to bind a potential partner, are very useful for probing the structural components of an interaction and they are very useful tools for characterisation of *in vivo* functions of newly identified protein-

protein interactions *(White, 1996; Vidal and Legrain, 1999)*. They can be compared with their wild-type counterparts for their ability to functionally complement knockout mutations in the corresponding gene or for their ability to function in expression assays in relevant cells. They are particularly useful for analysing proteins that have multiple interacting partners. *In vivo* expression of variants that interact with only a subset of their interacting partners can provide information about which interactions are important to mediate specific activities in cells. However, since they usually score as recessive mutations, the use of interaction defective mutants is not always feasible. In some cases the wild-type protein cannot be removed from an assay. This would be the case if no knockout was available for a particular gene of interest or in biochemical assays when the corresponding wild-type protein could not be immunodepleted from the test fraction. In these cases proteins, peptides or small molecules that specifically affect the ability of a particular pair of proteins to interact can be used for determining functional/biological relevance of a protein-protein interaction.

Until recently genetic strategies such as those mentioned above have not been used widely due to the technical difficulties of identifying informative interaction defective alleles or specific dissociating molecules. The main challenge for interaction defective alleles is the creation of subtle mutations that disrupt an interaction without grossly affecting the protein, i.e. mutations that still allow the expression of a full length, stable and correctly folded protein. This problem can be overcome by generating libraries of randomly mutated alleles for each of the interacting partners by error prone PCR. These libraries then have to be screened for alleles exhibiting the desired properties. Similarly, one might expect that complex libraries would need to be screened to find a few specific dissociating proteins, peptides or small molecules. The problems associated with screening such complex libraries have been overcome by the reverse two-hybrid system since dissociation of an interaction provides a selective growth advantage *(Vidal et al, 1996b)*.

Strategies employing reverse two-hybrid systems have been designed which allow for the selection of informative interaction-defective alleles in the context of stable, full length and properly folded proteins, from complex libraries of randomly mutated alleles. In some strategies the systems are used to select for conditional mutations that retain some level of activity. For example, mutations that affect the interaction only weakly can be selected out

from nonsense alleles. *Vidal et al (1996b),* who were studying the interaction of the protein E2F1 with another protein, DP1, employed a two-step strategy in which mutants were first selected for using their URA3 based reverse two-hybrid system. Identified mutants were then subjected to a second round of screening using a standard Gal4 dependent HIS3 reporter gene. They demonstrated that varying the concentration of 3-aminotriazole (3-AT) (a competitive inhibitor of the HIS3 gene product) in growth media in this second round of screening allowed selection of mutations that either weakly or strongly affected the interaction of E2F1 with DP1. For proteins that have multiple interacting partners another approach can be employed. Mutations affecting the ability of the protein to bind to one of its binding partners are selected using a standard reverse two-hybrid screen. Mutants obtained are then subjected to a second round of screening in a standard Y2H system to identify mutants still capable of interacting with another binding partner *(White, 1996). Inouye et al (1997)* designed a differential interaction trap system that allowed such mutant proteins to be identified in one single screening step. They used the system to identify single amino acid substitutions that affected the ability of Ste5, a protein involved in the yeast *S. cerevisiae* mating pheromone response pathway, to interact with only one of its demonstrated partners, Ste7 or Ste11. Ste7 was fused to the Gal4 DBD, Ste11 was fused to the LexA DBD and Ste5 was fused to the Gal4 AD. The three hybrid proteins were simultaneously expressed in the same yeast strain. To provide transcriptional readouts unique to each interaction (the interaction of Ste5/Ste11 and Ste5/Ste7) the yeast strain had two reporter constructs: a URA3 gene under the control of Gal4 UAS's and a *lacZ* gene under the control of LexA operator sequences. Ste5 mutants that were defective for binding to Ste11 but not for Ste7 were selected for by plating on media lacking uracil and containing X-gal (Ura^+/Lac^- mutants). Ste5 mutants that were defective for binding to Ste7 but not for Ste11 were selected for by plating on media containing 5-FAO and containing X-gal (Ura^-/Lac^+mutants).

Many diseases can be attributed to specific protein-protein or protein-DNA interactions and thus specific dissociation of such interactions could be viewed as a potential therapeutic strategy. Target interactions include interactions between proteins of a parasite and its host, unregulated associations between proteins and interactions responsible for the function of the downstream event of a regulatory pathway frequently mutated in a particular disease

(Vidal et al, 1996a). In all cases, peptides or small molecules capable of specifically dissociating abnormal associations could be critical therapeutic reagents. The specificity of protein-protein interactions makes them particularly suitable as drug targets. However, despite their biological importance in human disease protein-protein interactions have not been widely used as drug targets. This has primarily been because of the technical constraints of *in vitro* screening procedures *(Vidal and Endoh, 1999).* However, we may now have the technological tools and strategies that will make protein-protein interactions viable drug targets. Functional genomics projects can rapidly identify large numbers of potential protein-protein interactions for a particular pathway using large-scale Y2H screens. On the other hand, reverse two-hybrid systems facilitate the rapid screening of complex libraries of peptides or small molecules for compounds capable of preventing specific interactions. These hybrid technologies, coupled with advances in combinatorial chemistry such as split pool synthesis *(Huang and Schreiber, 1997)* which allows the synthesis of libraries of compounds of enormous size and complexity, have the potential to make screening for molecules capable of dissociating specific protein-protein or protein-DNA more feasible and to greatly accelerate the development of novel therapeutic agents.

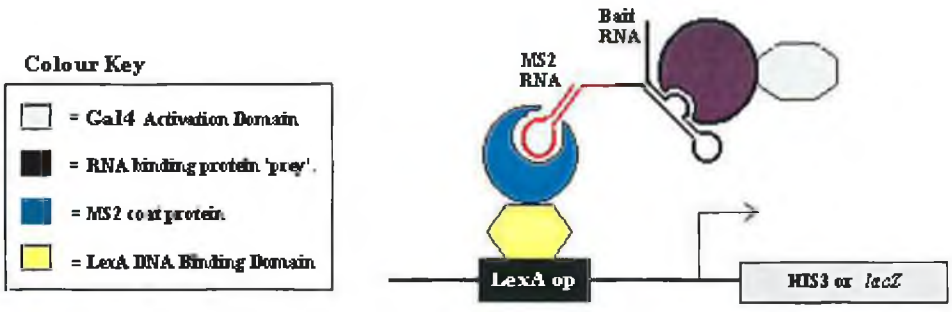
1.5.3: Three Hybrid Systems.

Other two-hybrid system spin-offs include three-hybrid systems for detecting protein-RNA and protein-ligand interactions. A brief description of these systems is given below.

RNA-protein interactions are pivotal in fundamental cellular processes such as transcription, mRNA processing, translation, RNA localisation and infection by RNA viruses. *SenGupta et al (1996)* developed a yeast three-hybrid system for the detection and analysis of protein-RNA interactions. The basic principle of the RNA-protein three- hybrid system is depicted in Figure 1.5.2. In the system they designed one of the hybrids consists of the LexA DBD fused to the well characterised and sequence specific coat protein of bacteriophage MS2. The MS2 coat protein of MS2 recognises a 21nt stem-loop in its genome with high affinity. The second hybrid is a hybrid RNA molecule consisting of the stem-loop structure recognised by the MS2 coat protein, fused to a 'bait' RNA sequence.

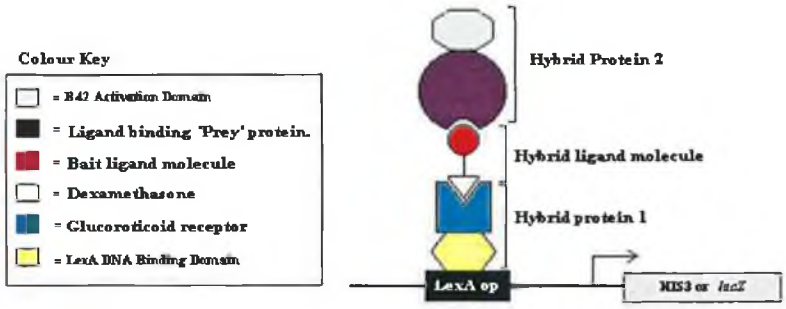
hybrid RNA molecule associates with the LexA-MS2 hybrid molecule bound to sites upstream of a reporter gene. The third hybrid consists of a 'prey' protein fused to the Gal4 transcriptional activation domain as in the standard yeast two-hybrid system. Interaction of the 'prey' protein with the 'bait' RNA sequence results in activation of transcription of a reporter gene resulting in a readily screenable phenotype.

- **Figure 1.5.2: RNA-Protein Three Hybrid System.**



In the initial system designed by *SenGupta et al (1996)*, each of the hybrid molecules was expressed from a separate plasmid. However, in order to make the system easier to use, the gene encoding the LexA-MS2 hybrid protein (which would be constant in all experiments) was integrated into a chromosome in strain L40 (see table1.4) to generate a new strain called L40-coat. This three-hybrid system is now commercially available from Invitrogen with the exception that the commercial system uses the B42 transcriptional activation domain. *Putz et al (1996)* devised a similar system to that described by *SenGupta et al (1996)*. They constructed their system with the RevM10 RNA binding protein from HIV-1 instead of the MS2 coat protein, and the RNA sequence to which RevM10 binds which is called the Rev responsive element (RRE).

- **Figure 1.5.3: Protein-Ligand Three Hybrid System.**



Small ligand-receptor interactions underlie many fundamental processes in biology and form the basis for pharmacological intervention of human diseases. Therefore it is often important to identify the protein targets of small molecules. The development of small molecule hybrid systems began with the use of cell-permeable chemical inducers of dimerisation (CID's). *Spencer et al (1993)* expressed the intracellular domain of the T-lymphocyte antigen receptor as a fusion with the immunophilin FK505-binding protein (FKBP12). They demonstrated that addition of an FK506-FK506 dimer, FK1012, induced aggregation of the T lymphocyte antigen receptor and resulted in activation of a signalling pathway. This demonstrated that small-molecule ligands could be used to activate many cellular processes by forcing oligomerization between proteins fused to ligand binding proteins. *Licitra and Liu (1996)* then developed a novel three-hybrid system for the detection of small ligand-protein interactions. The basic principle of the system is depicted in Figure 1.5.3 below. In the system they designed the LexA DBD is expressed as a fusion with the glucocorticoid receptor (GR). This hybrid protein is expressed in yeast in the presence of a hybrid ligand molecule. The hybrid ligand molecule consists of dexamethasone, which is a ligand for GR, and a 'bait' ligand molecule. Interaction of a 'prey' protein, expressed as a fusion with a transcriptional activation, results in expression of a reporter gene generating a screenable penotype. Systems such as that described above can be used either to screen libraries of 'prey' AD proteins to identify targets for a particular small ligand of interest or for screening libraries of ligands to identify novel ligands for a given 'prey' protein. They could be particularly useful in screening for novel therapeutic

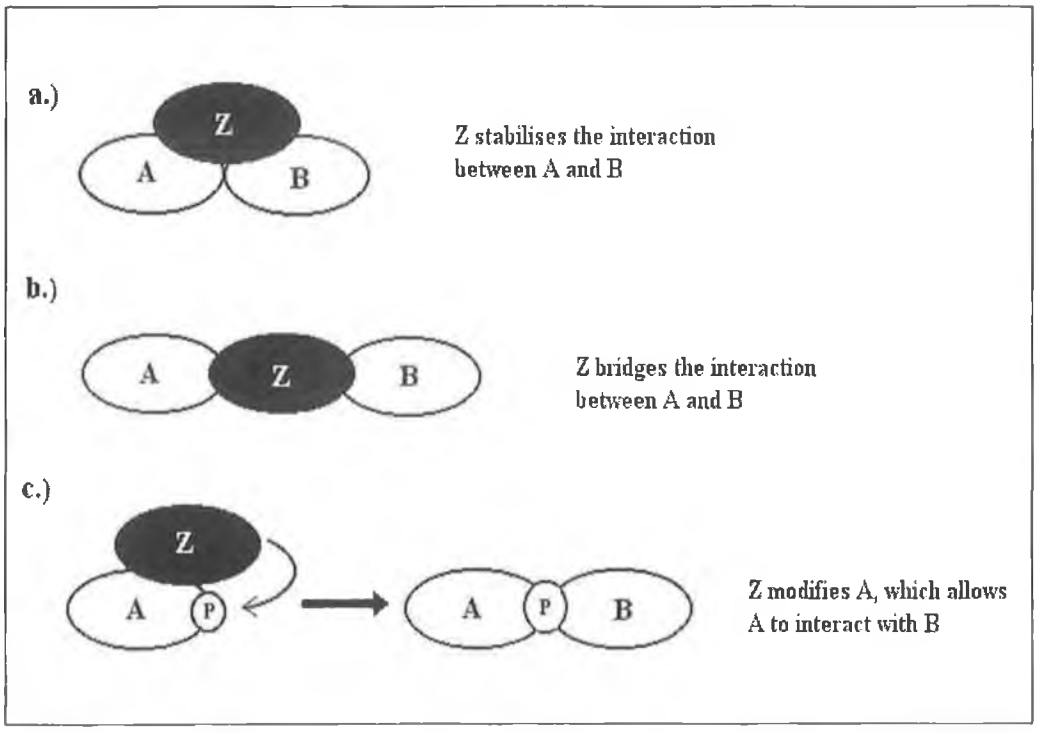
ligand molecules since any ligand molecules identified in a screen have already been screened for cell permeability and for their stability in a cellular environment.

1.5.4: Three Component Systems.

Some protein-protein interactions cannot be detected with standard two-hybrid methods because they require a third molecule not normally available in yeast. For example, the affinity of two proteins that contact each other directly may be enhanced by the expression of a third protein that contacts both (see figure 1.5.4a).

Expression of a third protein can also be used for detection of interactions between proteins that make no direct contact but interact solely via a third, bridging protein (see Figure 1.5.4b).

- **Figure 1.5.4: Different Kinds Of Ternary Protein Complexes.**



Finally, expression of a third protein can sometimes facilitate a protein-protein interaction without forming a lasting part of a stable complex (see figure 1.5.4c). This situation arises for interactions that depend on post-translational modifications that are relatively uncommon in yeast. *Osborne et al (1995)* co-expressed a tyrosine kinase, Lck, capable of phosphorylating a DBD-bait hybrid on a target residue to facilitate its interaction with an SH-domain containing AD-hybrid. Using intracellular portions of receptors as baits they were able to isolate proteins with immunoreceptor based tyrosine activating motifs (ITAMs) that interacted with tyrosine phosphorylated forms of receptors.

Zhang and Lauter (1996) demonstrated the use of their three component system for the detection of three component complexes using the epidermal growth factor (EGF) receptor, Grb2 and Sos proteins. They then screened a library with Gal4-DBD-EGF receptor and Gal4 AD-Sos hybrid proteins and isolated a plasmid encoding Grb2.*Tirode et al (1997)* also developed a three component system and used it to demonstrate that the cdk7-MAT1 interaction was stabilized by the presence of a protein called cyclin H.

1.5.5: Prokaryotic Two-hybrid Systems.

In recent years several prokaryotic two-hybrid systems have been developed. These systems have several advantages over the Y2H system. Firstly, they facilitate the analysis of prokaryotic protein-protein interactions in a prokaryotic genetic background. The ability to perform such assays in an *E. coli* background is particularly attractive considering the fact that there is a vast resource of *E. coli* strains with specific mutations which can be chosen for specific applications. P2H systems are also useful for the analysis of eukaryotic proteins in situations where homologous yeast proteins interfere with the interactions of the proteins under investigation. In addition to this manipulation of bacterial DNA is generally easier than that of yeast and much higher transformation efficiencies are attainable in *E. coli*. This facilitates the rapid and simple screening of large complex libraries of protein to identify putative binding partners for a given protein of interest. This potential makes

prokaryotic systems particularly attractive for the purpose for the construction of proteomic linkage maps.

It may be argued that prokaryotic based systems would be of limited use for the analysis of eukaryotic proteins on the basis that such proteins may not be stable or be capable of folding correctly in a prokaryotic background. Also some eukaryotic proteins require post-translational modifications such as phosphorylation or glycosylation in order to fold and function correctly and thus would not function in *E. coli*. However, it is important to note that limitations such as these are also encountered in the conventional Y2H system where hybrid proteins are targeted to and interact in the nucleus and thus are not subject to post-translational modifications such as glycosylation and disulfide bond formation that occur in the endoplasmic reticulum. Proteins that require phosphorylation or acetylation by non-yeast proteins also fail to interact in the conventional Y2H system.

As already mentioned above several P2H systems have been developed. *Cairns et al (1997)* developed a system that permitted the detection and analysis of homodimers. The system was based on the cI repressor of λ phage. This repressor binds to specific operator sites on the λ genome as a dimer and prevents transcription from the major leftward (P_L) and rightward (P_R) promoters. The cI monomer can be divided into two functional domains, an amino-terminal DBD (repB) and a C-terminal dimerization domain (repD). In the system devised by *Cairnes et al (1996)* the repD domain is removed and replaced with a heterologous protein. If the protein is capable of dimerization the cI repression is recovered and detected on the basis of repression of a λP_L *promoter:lacZ* reporter. This system facilitated the study of the dimerization properties of proteins in *E. coli.*

Dimitrova et al (1998) devised a system based on the LexA repressor. This protein functions as a dimer and binds to specific operator sequences in *E. coli.* Like the cI repressor the C-terminal dimerization domain of this protein can be removed and replaced with a heterologous protein. If the fused protein is capable of dimerization LexA repression is recovered. Systems such as this and then one devised by *Cairns et al (1997)* are of limited use for studying the interaction of heterodimers in cases where one of the proteins is also capable of forming homodimers (eg. Jun/Fos). *Dimitrova et al (1998)* overcame this problem by using two LexA DBD fusion proteins with different DNA binding specificities.

In their system one protein (A) is fused to a wild type LexA DBD while the another protein (B) is fused to a variant LexA DBD with a altered operator sequence requirement. They constructed a hybrid promoter with the wild type LexA operator sequence and the operator sequence recognized by the variant LexA DBD. This promoter was fused to a *lacZ* gene and thus only interaction between proteins (A) and (B) results in the formation of a heterodimer capable of binding to the hybrid promoter and repressing *lacZ* expression.

Karimova et al (1998) developed a system based on the reconstitution of a signal transduction pathway. In their system proteins of interest are fused to two complementary fragments of the catalytic domain of *Bordetella pertussis* adenylate cyclase. Interaction between the two proteins results in functional complementation between the two adenylate cyclase fragments leading to cAMP systhesis which in turn can trigger the expression of several resident genes via interaction with the CAP protein.

Pelletier et al (1999) devised a similar system in which proteins are fused to two designed fragments of the murine enzyme dihydrofolate reductase (mDHFR). Interaction of the proteins results in reconstitution of mDHFR activity. Activity was detected by an *E. coli* based selection assay, in which the bacterial DHFR was specifically inhibited with trimethoprim, preventing the biosynthesis of purines, thymidylate, methionine, and pantothenate and therefore cell division. The reconstituted mDHFR, which was insensitive to the low trimethoprim concentration present in selection, restored the biosynthetic reactions required for bacterial propagation. As a result, the interaction between proteins was directly linked with cell survival and detected by colony formation.

1.6: Construction Of A Novel Prokaryotic Two-Hybrid System.

The purpose of the work undertaken for this thesis was to develop a novel P2H system based on the prokaryotic transcriptional activating protein NifA from *Sinorhizobium meliloti* 2011. This protein belongs to a unique class of bacterial transcriptional regulatory proteins called Enhancer Binding Proteins (EBP's). These proteins activate transcription

from a special class of bacterial promoters that require the alternative bacterial sigma factor called σ^{54} bound to core RNA polymerase (σ^{54}-holoenzyme or $E\sigma^{54}$). The EBP's bind to specific DNA sequences found upstream of promoters that they are responsible for regulating and activate transcription by a mechanism that more closely resembles that observed for eukaryotic transcription factors. Bacterial EBP's also resemble eukaryotic transcription factor modularity in that they have distinct functionally independent DNA binding and transcriptional activation domains. It was therefore proposed that these proteins could be utilized for the construction of a novel P2H system homologous to the Y2H system. Construction of a novel NifA based P2H system requires knowledge of both the mechanism by which EBP's activate transcription and an understanding of the specific characteristics of the *S. meliloti* 2011 NifA protein itself. The following sections provide an overview of σ^{54} dependent transcription and of the EBP's.

1.7: The Transcription Cycle And the Role Of Bacterial Sigma (σ) Factors.

Transcription is a complex biological process that involves many different protein factors. In the eubacteria the core RNA polymerase contains four subunits with the stoichiometry $\beta\beta'\alpha 2$. This core enzyme interacts with individual specificity proteins, known as sigma (σ) factors, to form a holoenzyme with the minimal composition $\beta\beta'\alpha 2\sigma$. The σ factors are required for recognition and binding of specific promoter sequences and so only the holoenzyme can accurately initiate transcription at promoter sequences. Following transcription initiation, the σ factor is released and the core enzyme elongates an RNA chain. Transcription termination is subsequently triggered by interaction of the elongating RNA polymerase with an RNA termination structure or specific termination factors. This complex series of reactions has been termed the transcription cycle.

1.7.1: Alternative Bacterial Sigma Factors.

The majority of cellular transcription requires the predominant, or primary, σ factor. In *E. coli* the primary σ factor is referred to as σ^{70} while in *B. subtilis* the primary σ factor is referred to as σ^{43}. In addition to the most abundant σ factor, both gram-positive and gram-negative eubacteria employ alternative σ factors (reviewed by *Helmann and Chamberlin, 1988)* that confer different promoter specificity's on the core RNA polymerase. These alternative σ factors are required for the transcription of co-ordinately regulated sets of genes from promoter sequences that are quite distinct from those recognised by the primary bacterial σ factors (see Table 1.7.1).

Table 1.7.1: Bacterial σ Factors And Their Consensus Promoter Sequences *(Helmann and Chamberlin, 1988).*

Factor	Gene	Function	Consensus sequence	
Bacillus subtilis			−35	−10
σ^{43}	rpoD, sigA	housekeeping functions	TTGACA	TATAAT
σ^{28}	sigD	flagellar synthesis/chemotaxis	CTAAA	CCGATAT
σ^{29}	spoIIGB, sigE	sporulation genes	TT-AAA	CATATT
σ^{30}	spoOH, sigH	sporulation genes	unknown	
σ^{32}	sigC	unknown	AAATC	TA-TG-TT-TA
σ^{37}	sigB	unknown	AGG-TT	GG-ATTG-T
$\sigma^{spoIIAC}$	spoIIAC	sporulation genes	unknown	
?	spoIIIC	sporulation genes	unknown	
σ^{gp28}	SPO1 28	phage middle genes	T-AGGAGA—A	TTT-TTT
$\sigma^{gp33/34}$	SPO1 33,34	phage late genes	CGTTAGA	GATATT
Escherichia coli and Related Bacteria				
σ^{70}	rpoD	housekeeping functions	TTGACA	TATAAT
σ^{54}	glnF, ntrA, rpoN	nitrogen-regulated genes (+ C_4-dicarboxylate transport in *R. meliloti*)	CTGGCAC N_5 TTGCA	
σ^{32}	htpR, rpoH	heat-shock genes	CTTGAA	CCCCAT-TA
?	flbB+flaI	flagellar synthesis/chemotaxis	TAAA	GCCGATAA
σ^{gp55}	T4 gene 55	phage late genes	none	TATAAATA
?	unknown	flagellar genes *(C. crescentus)*	TGGC-C N_5 TTGC	

All of the alternative bacterial σ factors interact with the common core RNA polymerase and facilitate recognition of distinct promoter sequences by making sequence-specific contacts with promoters. Most of the alternative bacterial σ factors resemble the primary *E. coli* σ^{70} factor in that they interact with blocks of conserved sequences in the -10 and -35 regions of promoters (see Table 1.7.1) and that they do not bind tightly to DNA in the absence of core RNA polymerase.

One model for the binding RNA polymerase holoenzyme to promoters which helps to account for this suggests that the holoenzyme first binds to the -35 region via specific interactions involving the σ factor, perhaps with core RNA polymerase contributing some contacts at the -10 region. This complex may then isomerize to a conformation allowing specific contacts between the σ factor and the -10 region *(Helmann and Chamberlin, 1988).* The σ factor therefore requires core RNA polymerase in order to bind tightly to promoters. It is now known that the α subunit of core RNA polymerase also plays a role in the recognition of promoters via binding to A/T rich sequences immediately upstream of core -10 and -35 core promoter sequences *(Ross et al, 1993).*

In addition to their role in the recognition and binding of specific promoter sequences bacterial σ factors can also display other activities such as DNA melting and inhibition of non-specific transcription by core RNA polymerase *(Helmann and Chamberlin, 1988).* Core RNA polymerase binds tightly and non-specifically to DNA but this non-specific binding is suppressed by σ^{70}. Isolated β' subunit of core RNA polymerase is able to bind to DNA in a non-specific manner. Upon addition of σ^{70}, β' dissociates from DNA and complexes with the σ^{70} subunit. This is consistent with the fact that non-specific binding of core RNA polymerase is suppressed by σ^{70} and that the main contact site for σ^{70} on core RNA polymerase is located on β', although the C-terminal region of β is also involved in σ^{70} binding *(Ishihama, 1992).*

Individual alternative bacterial σ factors allow the transcription of sets of genes whose products have a common physiological role. For example, σ^{32} of enteric bacteria is required for the transcription of genes whose products are required for protection from heat shock and certain other stresses while σ^{28} of *B. subtilis* is required for the transcription of genes involved in motility and chemotaxis. The alternative σ factor, σ^{54}, differs from all

other bacterial σ factors in several respects, one of which being that it allows the transcription of genes whose products have diverse physiological roles *(reviewed by Kustu et al, 1989).*

σ^{54} (encoded by *ntrA* [*glnF*, *rpoN*]) was initially identified as a positive regulatory factor needed for the expression of the gene encoding glutamine synthetase, *glnA*, in enteric bacteria *(Kustu et al, 1989)*. It was later found to be required for the expression of other genes whose products function in the assimilation of nitrogen. For example, σ^{54} is required for transcription of genes encoding amino acid transport components (such as the *glnHPQ* operon of *E. coli* encoding components for the glutamine transport system; *Claverie-Martin and Magasanik, 1991*), degradative enzymes (such as the *hutUH* operon encoding genes for the degradation of histidine) and for transcription of the nitrogen fixation (*nif* and *fix*) genes from a number of bacteria, including *Klebsiella pneumoniae*, *Rhodobacter capsulatus*, and members of the genera *Azotobacter, Azospirillum, Sinorhizobium*, *Azorhizobium* and *Bradyrhizobium*.

It is now known that σ^{54}-holoenzyme ($\mathbf{E\sigma^{54}}$) transcribes genes whose products have diverse physiological functions. Examples of such genes are;

- The *dctA* gene of *Rhizobia* which encodes a transport component for dicarboxylic acids.
- Genes on the TOL (toluene) plasmid of *Pseudomonas putida* that encode proteins required for the catabolism of toluene and xylenes.
- Genes encoding two of the components of a formate-degradative pathway in *E. coli.*
- Genes encoding hydrogenases responsible for oxidation of molecular hydrogen in *Alcaligenes eutrophus* and *Pseudomonas facilis*.
- Genes encoding the hook and filament proteins of *Caulobacter flagella.*
- Genes encoding pilins in *Pseudomonas aeruginosa* and *Neisseria gonorrhoeae* that allow these organisms to adhere to human epithelial cells.

In addition to the above, σ^{54} also differs structurally and functionally from all other bacterial σ factors and activates transcription by a unique mechanism which more closely resembles that observed for the transcriptional activation of eukaryotic RNA polymerase II. These characteristics of σ^{54} are described in the following sections.

1.7.2: The Alternative Bacterial Sigma Factor σ^{54}.

σ^{54} differs functionally from all other σ factors in that all known σ^{54}-dependent promoters are subject to positive control rather than repression, and transcription initiation by RNA polymerase containing σ^{54} (σ^{54}-holoenzyme or $E\sigma^{54}$) is totally dependent on an activator protein. These activator proteins bind to palindromic or nearly palindromic sequences upstream of σ^{54} promoters (~100bp to 200bp) that have properties of eukaryotic transcriptional enhancers in that they can be moved more than a kilobase upstream and still mediate transcriptional activation *(Reitzer and Magasanik, 1986; Buck et al, 1986; Ninfa et al, 1987)*. These sequences are called Enhancer-Like Elements (ELE) or Upstream Acitivation Sequences (UAS) and the bacterial activator proteins that bind to them are collectively called Enhancer Binding Proteins (EBP's). These activator proteins are described in detail in section 1.8.

Promoters recognized by σ^{54} lack the conserved -10 and -35 regions found in typical promoters and instead are characterized by a consensus sequence TGGCAC - N5 - TTGCA between -26 and -12 *(Buck et al, 1985)*. Over 60 such promoters have been identified in a variety of genera. The GG doublet at -25/-24 and the GC doublet at -13/12 are the most invariant nucleotides amongst the σ^{54}-dependent promoters. Mutational analysis (*Buck et al, 1985*) and $E\sigma^{54}$-dependent methylation protection in DMS footprinting assays indicates that these nucleotides are important contact points for $E\sigma^{54}$ *(Morett and Buck, 1989)*.

Activity of the alternative sigma factor σ^{54} has been studied most intensively at the *glnA* and *nif* promoters. $E\sigma^{54}$ has been demonstrated to bind to promoter sequences independently of activator proteins but interactions between $E\sigma^{54}$ and promoter sequences (closed complexes) are nonproductive transcriptionally because the DNA remains double

stranded (*Ninfa et al, 1987; Popham et al, 1989; Moret and Buck, 1989; Buck and Cannon, 1992a*). In the case of the *glnA* promoter open complex formation and initiation of transcription requires an activator called NtrC (encoded by *ntrC* [*glnG*]). NtrC binds to sites located upstream of the promoter and catalyzes the isomerization of closed complexes between $E\sigma^{54}$ and the *glnA* promoter to transcriptionally productive open complexes in which the DNA strands are locally denatured in the region of the transcription start site (*Popham et al, 1989*). The isomerization reaction requires ATP hydrolysis by NtrC activator protein (*Popham et al, 1989; Weiss et al, 1991; Austin et al, 1991; Austin and Dixon, 1992*). As is true for *glnA*, transcription of other σ^{54}-dependent promoters also requires an activator and hydrolysis of ATP. Transcription of *nif* genes in a variety of bacteria requires the activator NifA and transcription of the *dctA* gene is dependent on DctD. The absolute requirement for an activator protein and hydrolysis of ATP to catalyze closed $E\sigma^{54}$ promoter complex to open complexes is in stark contrast to σ^{70}-dependent promoters. Closed $E\sigma^{70}$ complexes can generally make the transition to the open complex without the aid of an activator and the process does not require the hydrolysis of ATP *(Kustu et al, 1991).*

In addition to the absolute requirement for an activator protein for initiation of transcription, σ^{54} also differs from other bacterial σ factors in the way it interacts with core RNA polymerase and promoter sequences. The ability of $E\sigma^{54}$ to bind to promoter sequences and form detectable closed complexes was found to be dependent not only on the conserved GG and GC doublets found at positions -24 and -12 respectively but also on a tract of thymine residues located at -15 to -17 in most σ^{54}-dependent promoters. $E\sigma^{54}$ was found to have a very low affinity for promoters lacking this stretch of thymine residues such as the *nifH*, *nifB* and *nifU* promoters of *Klebsiella pneumoniae*. This made it more difficult to detect closed complexes between $E\sigma^{54}$ and these promoters by DMS footprinting. In *K. pneumoniae* the wild type *nifH* promoter has a tract of cytosine residues in place of the thymine residues. *Morett and Buck (1989)* demonstrated the importance of the tract of thymine residues for $E\sigma^{54}$ binding by mutating the cytosine bases in the -15 to -17 region of the *K. pneumoniae nifH* promoter to thymines (see Figure 1.7.1). The formation of closed complexes between the resulting mutant *K. pneumoniae nifH* promoter sequence (*Kp nifH*49) and $E\sigma^{54}$ was readily detected by DMS protection footprinting

indicating an increase in the binding affinity of $E\sigma^{54}$. It was initially thought that σ^{54}, like other σ factors, interacted first with core RNA polymerase to form $E\sigma^{54}$ which then interacted with promoter sequences to form closed complexes. However, *Buck and Cannon (1992b)* demonstrated that σ^{54}, unlike other bacterial σ factors, was capable of specifically binding to promoter DNA in the absence of core RNA polymerase. Again binding of σ^{54} was not detected for the wild type *K. pneumoniae nifH* promoter sequence and thus the -15 to -17 tract of thymine residues found in most σ^{54}-dependent promoters was implicated as being important in determining the strength of binding of σ^{54} to promoter sequences. The detection of $E\sigma^{54}$ binding when σ^{54} binding was not evident (i.e. for the *K. pneumoniae* wild type *nifH* promoter) indicated that core RNA polymerase in some way stabilized σ^{54} binding to promoter sequences.

- **Figure 1.7.1: σ^{54} Recognition Sequences In Various σ^{54} Dependent Promoters.**

	-24												-12		
Consensus	C	T	G	G	C	A	C	G	G	C	C	T	T	T	G	C	A
Rm nifH	C	T	G	G	C	A	C	G	A	C	T	T	T	T	G	C	A
Kp nifH	C	T	G	G	T	A	T	G	T	T	C	C	C	T	G	C	A
Kp nifH49	C	T	G	G	T	A	T	G	T	T	T	T	T	T	G	C	A
Kp glnAp2	T	T	G	G	C	A	C	A	G	A	T	T	T	C	G	C	T

N.B: Region shown in yellow indicates base changes between the wild type and mutant *K. pneumoniae nifH* promoters
Bases shown in red are those that differ from the consensus sequence

Using DNaseI footprinting and gel mobility assays *Cannon et al (1993)* demonstrated that core RNA polymerase afforded a significant binding advantage to σ^{54}. They provided evidence that this was due to induced conformational changes in σ^{54} that improved its direct interactions with promoter sequences rather than to non-specific interactions between the core RNA polymerase with upstream DNA sequences. This was consistent

with the observation that the DNaseI footprint for $E\sigma^{54}$ is moved upstream towards the -12 promoter element when compared with that of σ^{54} (*Buck and Cannon, 1992b*). *Casaz and Buck (1997)* later provided further evidence that core RNA polymerase induced conformational changes in σ^{54} by demonstrating that core RNA polymerase caused changes in σ^{54} protease sensitivity. *Cannon et al (1993)* also demonstrated that the thymine tract at -15 to -17 favored σ^{54} binding and that the non-specific DNA binding activity of core-RNA polymerase was inhibited by σ^{54}.

Binding of σ^{54} to promoter DNA may provide an advantage in the assembly of closed promoter complexes by favoring $E\sigma^{54}$ formation at promoter sequences, to perhaps compensate for the apparently weak (compared to σ^{70}) association of σ^{54} with core RNA polymerase in solution (*Cannon et al, 1993*). It is now known that σ^{54} remains bound to the bottom template strand at the promoter after core RNA polymerase begins elongation and DNA at the start site remains open. This indicates that strong contacts between σ^{54} and promoter sequences formed as a result of conformational changes following transcriptional activation are not easily broken (*Tintut et al, 1995*). After polymerase and σ^{54} separate the system has the potential to re-initiate but new core RNA polymerase that comes in for re-initiation cannot engage using a proper conformation, despite the fact that the start site is already open. *Tintut et al (1995)* suggested that σ^{54} either needs to be released or undergo a conformational change prior to re-initiation. This novel transcription cycle involving σ^{54} raises interesting possibilities for transcriptional regulation, especially with regard to control at the level of re-initiation.

The above mentioned differences between σ^{54} and all other bacterial σ factors are reflected in the fact that amino acid sequence of σ^{54} shows no homology to that of other bacterial σ factors. It instead possess a number of conserved motifs more commonly associated with eukaryotic transcription factors (*Sasse-Dwight and Gralla, 1990*). These domains probably explain why σ^{54}-dependent promoters can respond to remote activation by EBP's as do eukaryotic promoters (*Reitzer and Magasanik, 1986*).

1.8: The Bacterial Enhancer Binding Proteins (EBP's).

As already described above the EBP's activate transcription from σ^{54}-dependent promoters by binding to specific palindromic or nearly palindromic DNA sequences normally found between 100bp to 200bp upstream of σ^{54} promoters. These binding sites resemble eukaryotic enhancers in that they can be moved thousands of nucleotides upstream *(Reitzer and Magasanik, 1986; Buck et al, 1986)* or downstream *(Reitzer et al, 1989)* from the transcriptional start and still activate transcription. These binding sites have thus been termed Enhancer-Like Elements (ELE) or Upstream Activating Sequences (UAS). The EBP's bind as dimers (*North et al, 1993*) and once bound they cataylse the isomerization of transcriptionally inactive closed $E\sigma^{54}$-promoter complexes to transcriptionally active open complexes in a reaction that requires the hydrolysis of ATP *(Popham et al, 1989*).

It has been observed that certain strong promoters can be partially activated even if they lack an UAS. For example, NtrC was observed to weakly activate transcription form the *nifH* promoter, which is regulated by NifA, despite the fact that the *nifH* promoter lacked an NtrC UAS *(Sundaresan et al 1983; Buck et al, 1985). Buck et al (1985)* found that deleting the upstream NifA UAS from the *nifH* promoter diminished activation by NifA but that NifA could still weakly activate transcription. This activation occurs from solution within the cell and is now known to be related to the strength with which certain promoters independently bind σ^{54}. This fact poses the question as to what the role of the UAS is? Most likely, the UAS serves to increase the local concentration of the EBP at the promoter and to tether it in the right position for activation. Since each EBP recognizes a specific UAS binding to these elements confers specificity to the initiation of transcription.

Like eukaryotic enhancer-binding proteins, prokaryotic EBP's have several domains. Most of the prokaryotic EBP' s have three domains, an amino-terminal (N-terminal) domain, a central domain and a carboxy-terminal (C-terminal) domain. Those proteins whose sequences are known (NtrC, NifA, DctD, and XylR) show a high degree of sequence similarity at the amino acid level *(Buikema et al 1985; Drummond et al, 1986).* This indicates both evolutionary relatedness *(reviewed by Morett and Segovia, 1993*) and that these proteins activate transcription by a common mechanism.

The N-terminal domain of the EBP's is regulatory and is a true domain in that it is joined to the remainder of the protein by a flexible Q-linker *(Drummond et al, 1986; Wootton and Drummond, 1989)*. While there is low homology between the N-terminal domains of the *K. pneumoniae* and *S. meliloti* NifA proteins, this is not shared with other EBP's. This reflects the fact that different EBP's are regulated by different mechanisms involving different proteins *(Drummond et al, 1986)*. The sequences for the central and C-terminal domains exhibit a high degree of homology, especially in the central domain. The central domain possess an ATP binding motif and appears to be directly responsible for nucleotide hydrolysis and transcriptional activation (*Drummond et al, 1986; Huala and Ausubel, 1989; Weiss et al 1991; Austin et al 1991; Austin and Dixon, 1992; Cannon and Buck, 1992)* and the C-terminal domain contains a helix-turn-helix DNA binding motif and is responsible for binding of the protein to specific UAS's *(Drummond et al, 1986; Morett et al, 1988; Morett and Buck, 1988, Lee et al, 1993)*.

Prokaryotic EBP's reflect eukaryotic modularity. It was demonstrated for the NifA and DctD proteins that the central domain is an independent functional unit and that transcriptional activation and DNA binding are truly separable. In each case the C-terminal domain of the protein was removed without eliminating transcriptional activation, which due to the inability of the proteins to bind DNA, occurred from solution (*Morett et al, 1988*; *Huala and Ausubel, 1989; Huala et al, 1992*). Conversely in the case of NifA the isolated C-terminal domain retained the ability to bind specifically *(Lee et al, 1993)*. In both cases the N-terminal domain has also been removed, and the isolated central domain retains the ability to activate transcription (*Huala and Ausubel, 1989; Huala et al, 1992*). Moreover, a functional chimera has been constructed between the N-terminal and central domains of DctD and the C-terminal domain of NifA (*North et al, 1993*). This chimera seems to have the DNA binding specificity of NifA and can activate transcription.

Attempts to separate the DNA binding and transcriptional activation functions of NtrC have been unsuccessful. Although the isolated C-terminal domain of NtrC retains the ability to bind to DNA the remainder of the protein does not retain the ability to activate transcription. There are several lines of evidence to suggest that the reason for this is that the dimerization determinants for the NtrC protein lie in the C-terminal domain region while those for the other activators do not (*North et al, 1993*). The isolated C-terminal

domain of *Salmonella typhimurium* NtrC recognizes its DNA binding site as a dimer and is a dimer in solution while the isolated C-terminal domain of *K. pneumoniae* NifA, which is also capable of binding to its DNA binding site at high concentrations *(Lee et al, 1993),* is largely monomeric in solution. Finally, although it was possible to produce an active chimera between the C-terminal domain of NifA with the upstream region of DctD, it was not possible to do so with upstream regions of NtrC (*Drummond et al, 1990*). Failure to obtain active chimeras between NifA and NtrC might be explained by the fact that both portions of the chimera were monomeric. *North and Kustu (1997)* demonstrated the ability of NtrC to activate transcription from solution by mutating the second helix in the C-terminal DNA-binding domain which was known to be required for recognition of specific DNA sequences. This abolished the ability of the mutant NtrC protein to bind DNA but it was still capable of dimerization and of activating transcription from solution.

Control of transcription at σ^{54}-dependent promoters is accomplished by modulation of the activity state and abundance of the Enhancer Binding Proteins (EBP's). The amount of σ^{54} does not vary much, at least under different conditions of nitrogen availability.

Each EBP allows σ^{54} to initiate transcription in response to a distinct physiological signal such as;

- Limitation of combined nitrogen (NtrC)
- Low oxygen tension (NifA)
- Availability of dicarboxylic acids (DCTD).

The functions of the activators themselves are controlled by a variety of mechanisms. The best studied EBP's are NtrC and NifA that stimulate the expression of genes required for nitrogen assimilation and nitrogen fixation, respectively, in a number of organisms. These EBP's and the mechanisms by which they regulate transcription are described in detail in the following sections.

1.8.1: The General Nitrogen Regulatory Protein NtrC.

The related species *K. aerogenes, E. coli, S. typhimurium and K. pnuemoniae* regulate the expression of operons and genes whose products are involved in the assimilation of alternative nitrogen sources when ammonia, the preferred source, is lacking in the growth medium. Among the nitrogen-regulated operons are *glnALG,* which contains *glnA,* the structural gene for glutamine synthetase, the primary ammonia assimilating enzyme; *argT* and *hisJQHP*, which code for the components of the uptake systems for arginine and histidine; *hutUH*, which codes for enzymes required for the degradation of histidine; and *nifLA* which contains the *nifA* gene whose product activates the expression of the structural genes for the proteins required in nitrogen fixation.

Activation of the Ntr (Nitrogen regulated) genes requires the product of the *ntrA* gene (encoding σ^{54}), NtrC (product of *ntrC* [*glnG*]) *(Pahel et al 1982)* and on NtrB (product of the *ntrB* [*glnL*]) *(Chen et al, 1982).* The NtrC protein is often also referred to as NRI (Nitrogen regulator I) and NtrB is often referred to as NRII (Nitrogen regulator II) and these terms are used in the following sections. The two regulator proteins are encoded by the *glnALG* (*glnA ntrBC*) operon which also contains the structural gene for gluamine synthetase (*glnA*).

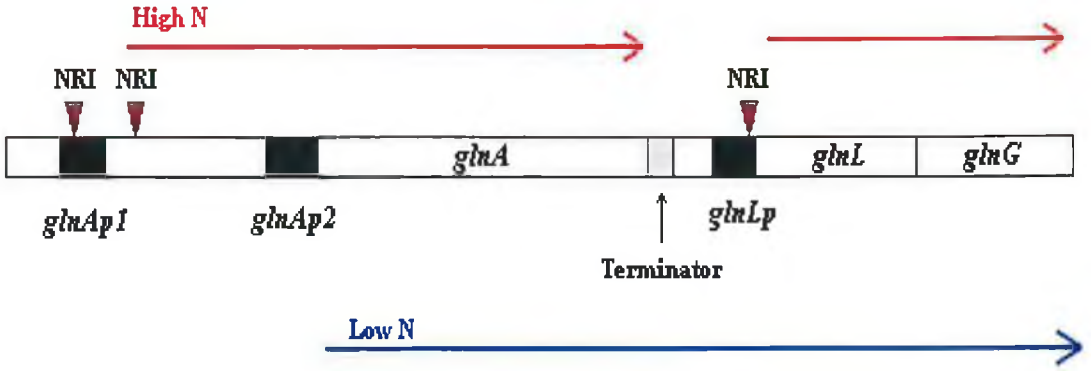
NRI and NRII are part of a two component regulatory system. NRI is initially synthesized as an inactive protein that is incapable of activating transcription. It is inter-converted between an active phosphorylated and inactive dephosphorylated forms through the activity of the NRII protein *(Kustu et al, 1989)* which in turn receives an accurate assessment of the availability of cellular nitrogen from the products of *glnB* (product of which is called P_{II}) and *glnD* (product of which is called uridylyltransferase [UTase]). The NRI protein has an ATP binding motif in its central transcriptional activating domain which is responsible for the hydrolysis of ATP. This ATPase activity is required to catalyze open complex formation *(Weiss et al, 1991; Austin et al, 1991))* and was shown to be activated by phosphorylation of the N-terminal domain of NRI by NRII *(Austin and Dixon, 1992).* When nitrogen is present NRII associates with the small protein P_{II} to become a phosphatase which dephosphorylates NRI and thus inactivates it. When ammonia levels in the media drop off the intracellular level of glutamine drops. This enables the enzyme

UTase to use UTP to uridylate P_{II}. The resulting P_{II}-UMP is released from its association with NRII enabling it to phosphorylate and thus activate NRI. NRI then activates transcription from the *glnALG* operon encoding glutamine synthetase and NRI. The increased NRI concentration then induces the expression of several other operons encoding proteins facilitating the utilization of numerous alternative nitrogen sources *(Magasanik, 1996)*. In the nitrogen fixing bacteria *K. pneumoniae* NRI is also responsible for activating the transcription of the *nifLA* operon which encodes the NifA EBP. The NifA protein then activates the expression of all other nitrogen fixation (*nif*) genes.

As mentioned above the NRI and NRII proteins are encoded by the *glnALG* (*glnA ntrBC*) operon. This operon is autogenously regulated by NRI at three promoters; *glnAp1, glnAp2* and *glnLp*. Initiation of transcription at the *glnApI* promoter is stimulated by the catabolite activator protein charged with cAMP. Transcription at this promoter serves to maintain a low intracellular level of glutamine synthetase in cells growing in conditions of carbon deficiency and nitrogen excess (see Figure 1.8.1). Initiation of transcription at *glnLp* promoter serves to maintain a low intracellular level of NRI and NRII (see Figure 1.8.1). The *glnAp2* promoter is a σ^{54} dependent promoter and is activated by phosphorylated NRI. Initiation of transcription at this promoter results in high intracellular concentrations of glutamine synthetase and NRI (see Figure 1.8.1). NRI binds to the consensus sequence TGCACCA---TGGTGCA which is a perfect inverted repeat of 7 bp separated by 3 bp. Binding sites for NRI overlap the start site and the TATAAT sequence of the *glnLp* promoter of *E. coli* and inhibits initiation of transcription from this promoter by σ^{70}-RNA polymerase holoenzyme *(Reitzer and Magasanik, 1986)*. Two binding sites lie 110bp and 140bp upstream of the *glnAp2* promoter. The first of these, designated site 1, overlaps the portion of the *glnAp1* promoter homologous to the canonical TTGACA sequence, and site 2 overlaps the transcriptional start site. Binding of NRI to these sites thus inhibits the initiation of transcription from the *glnAp1* promoter and results in the activation of transcription from the *glnAp2* promoter *(Reitzer and Magasanik, 1986)*. These sites still activate transcription from the *glnAp2* promoter when they are moved much further upstream of the promoter *(Reitzer and Magasanik, 1986)* and can activate transcription from either side of the promoter (*Reitzer et al, 1989)*.The increase in the intracellular level of NRI caused by NRI activated transcription from the *glnAp2* promoter because not all

transcripts initiated at *glnAp2* are terminated at the end of *glnA* and because the presence of NRI at the sites overlapping *glnLp* does not interfere significantly with the elongation of the transcripts (*Magasanik, 1996*). It can thus be seen that the NRI protein is capable of acting both as a repressor (of σ^{70} dependent promoters *glnAp1* and *glnLp*) and as an activator (of the σ^{54}-dependent *glnAp2* and other σ^{54}-dependent promoters).

- **Figure 1.8.1: Transcription Of The *glnALG* Operon Under Excess Nitrogen (High N) Or Limiting Nitrogen (Low N).**



Early experiments devised to address how EBP's bound far from the promoter could activate transcription showed that the relative position of the UAS and the promoter was critical for activation. It was observed that activation of the *K. pneumoniae nifH* promoter by NifA (see section 1.8.2 below) only occurred efficiently when its UAS was located over a certain face of the DNA helix *(Buck et al, 1987)*. It was therefore proposed that EBP functions might be mediated by direct contact with bound $E\sigma^{54}$ and that EBP's might simultaneously contact upstream UAS's and the $E\sigma^{54}$ holoenzyme bound at the promoter causing looping of DNA the intervening DNA. *Reitzer et al (1989)* reported the same findings for NRI-mediated activation of the *glnAp2* promoter and again proposed that activation of transcription by NRI might involve loop formation. This hypothesis was elegantly demonstrated by direct observation of such loop structures by electron microscopy with the *glnAp2* promoter (*Su et al, 1990*). *Claverie-Martin and Magasanik (1991)* later demonstrated that activation of the *glnHPQ* operon also involved loop

formation but that in this case loop formation was assisted by a protein called integration host factor (IHF). This protein had been demonstrated to play an important role in NifA-mediated activation of transcription from the *K. pneumoniae nifH* and *nifU* promoters *(Hoover et al 1990)* (see section 1.8.2 below). It was found to bind DNA between the promoter and the upstream NifA UAS. It then induced a bend in the DNA thus facilitating direct productive interactions between the upstream bound NifA and the promoter bound $E\sigma^{54}$. IHF was found to perform a similar function in the case of the *glnHp2* promoter thus promoting expression of the *glnHPQ* operon. *Claverie-Martin and Magasanik (1992)* demonstrated that the ability of IHF to enhance transcription of the *glnHp2* promoter required that all three protein (NRI, IHF and $E\sigma^{54}$) were bound on the same face of the DNA helix which supported the fact that IHF exerts its effects through bending of the DNA.

1.8.2: The Nitrogen Fixation Regulatory Protein NifA

The free living nitrogen fixing bacterium *K. pneumoniae* utilizes the enzyme nitrogenase to reduce dinitrogen (N_2) to NH_4^+ under conditions of nitrogen starvation, low oxygen tension and an ambient temperature below 37 ^oC *(Fischer and Hennecke, 1987)*. Nitrogenase is composed of polypeptides encoded by the genes *nifH, nifD* and *nifK* which are situated within an operon transcribed in the direction *nifH* to *nifK*. The *nifHDK* operon is itself located within a larger cluster of 20 contiguous *nif* genes which are organized into 7 or 8 operons. One *nif* operon, the *nifLA* operon, codes for the regulatory proteins NifL and NifA. NifA is an EBP responsible for the activation of all other *nif* operons and, unlike NRI, NifA is initially expressed as a fully active protein (i.e. it is capable of activating transcription). The NifL protein is involved in modulating the activity of the NifA protein in response to certain physiological conditions (i.e. oxygen tension and nitrogen availability). The mechanism by which it does this is unknown *(Sundaresan et al, 1983)* but appears to involve the N-terminal domain of the NifA protein.

The *nif* genes of the enteric bacterium *K. pneumoniae* are indirectly under the control of the *ntr* system. Under conditions of nitrogen limitation the NRI protein activates the transcription of the *nifLA* operon *(Sundaresan et al, 1983)*. The *nifA* gene product then activates the transcription of all other nitrogen fixation (*nif* and *fix*) genes and also autogeneously activates the *nifLA* operon *(Buck et al, 1985)*. The *nifL* gene product prevents NifA-mediated activation of other *nif* operons in response to high oxygen levels and intermediate concentrations of nitrogen. Expression of the *nifLA* operon is only partially repressed by oxygen but fully repressed by high concentrations of fixed nitrogen *(Fischer and Hennecke, 1987)*. The fact that *nifLA* expression itself is somewhat repressed upon aeration is due to the NifL protein preventing NifA-mediated autoactivation of the *nifLA* operon *(Fischer and Hennecke, 1987)*.

The NifA protein binds to the consensus sequence TGT-N_4-T-N_5-ACA located more than 100bp upstream of *nif* promoters *(Buck et al, 1986; Morett and Buck, 1988)*. *Buck et al (1987)* found that NifA-mediated activation of transcription required that the NifA protein be bound upstream and to the correct face of the DNA helix in order to interact with downstream transcription factors (i.e. $E\sigma^{54}$ bound at the promoter). It was proposed that this interaction was brought about by loop formation. *Hoover et al* (1990) discovered that a protein called the integration host factor, IHF, bound to DNA between the *nifH* promoter and the upstream NifA UAS. They discovered that IHF greatly stimulated NifA-mediated activation of transcription from the *nifH* promoter and that it caused DNA adjacent to the *nifH* promoter to bend. They deduced that the IHF induced bend facilitated direct productive contacts between upstream bound NifA and $E\sigma^{54}$ bound at the *nifH* promoter thus resulting in activation of transcription. They also demonstrated that IHF could bind to *nif* promoter regulatory regions in a variety of nitrogen fixing bacteria and that in each case it bound between the promoter and the upstream site(s) for NifA. *Santero et al (1992)* provided further evidence to support the role of IHF in NifA-mediated activation of transcription. They found that IHF greatly stimulated NifA-mediated transcription of the *nifH* promoter but that it did not enhance the level of activation of the *nifH* promoter caused by NRI (which had been observed to activate transcription from the *nifH* promoter). They proposed that this was because NRI did not bind to the *nifH* promoter and that it activated the *nifH* promoter from solution. They removed the upstream NifA UAS from the

nifH promoter and replaced it with the UAS for NRI and found that IHF then enhanced NRI activation of the *nifH* promoter.

Genetic control of nitrogen fixation by nitrogen status is much less rigorous in the legume root and stem nodule bacteria belonging to the genera *Sinorhizobium*, *Bradyrhizobium* and *Azorhizobium* even though these bacteria possess a nitrogen control circuitry for utilization of amino acids and nitrate. In these organisms *nif* gene expression is primarily regulated in response to oxygen tension *(Fischer and Hennecke, 1987)*. Unlike *K. pneumoniae, S. meliloti* and *B japonicum* only fix nitrogen in symbiotic nodules where microaerobic conditions prevail, and in free living cells under low oxygen tension *(Fischer and Hennecke, 1987)*. The NifA gene has been identified in several *Rhizobia* species and is required for *nif* expression in all of them. It was initially thought that the activity of NifA might be modulated in response to oxygen in all of these species by a similar mechanism to that observed in *K. pneumoniae*. This was supported by the observation that *K pneumoniae NifA* could activate transcription of the *S. meliloti nifHDK* operon *(Sundaresan et al, 1983)*. However, no *nifL*-like gene was identified in any *Rhizobial* species.

Fischer and Hennecke (1987) reported that the NifA protein from *B. japonicum* ($NifA_{Bj}$) was sensitive to oxygen in both a *B. japonicum* and an *E. coli* background while the *K. pneumoniae* NifA protein was not. Based on this observation they suggested that regulation of $NifA_{Bj}$ activity in relation to oxygen tension might not be mediated indirectly through the action of another protein and that the $NifA_{Bj}$ was itself inherently sensitive to oxygen. Comparison of the amino acid sequences of NifA from *B. japonicum, S. meliloti, K. pneumoniae* and *Rhizobium leguminosrum bioxar viciae* revealed that while these proteins exhibited a high degree of homology in their central and C-terminal domains that there was little homology between their N-terminal domians $NifA_{Bj}$ *(Fischer et al 1988)*(see Figure 1.8.2). This supported the view that the activity of NifA in the *Rhizobial* species was regulated in a different manner to that observed in *K. pneumoniae. Fischer et al (1988)* also found that when the N-terminal domain of the $NifA_{Bj}$ was deleted that the resulting deleted protein was still active and that it retained the oxygen sensitivity characteristic of the wild type protein. This supported the theory that the $NifA_{Bj}$ was itself inherently oxygen sensitive and that this sensitivity did not require the N-terminal domain of the protein. Like $NifA_{Bj}$, the *S. meliloti* NifA ($NifA_{Sm}$) was also reported to be oxygen sensitive. Comparison

of the amino acid sequences of NifA from *B. japonicum, S. meliloti, K. pneumoniae* and *Rhizobium leguminosrum bioxar viciae* also resulted in the identification of an interdomain linker region between the central and C-terminal domains of NifA that was conserved among all of the *Rhizobial* NifA proteins but absent from the *K. pneumoniae* NifA protein (see Figure 1.8.2). Within this interdomain region where two highly conserved cysteine residues *(Fischer et al, 1988)*. Since cysteine residues are very often involved in redox-reactive processes *Fischer et al (1988)* decided to alter the two conserved cysteine residues in the $NifA_{Bj}$ to determine the functional importance of the conserved interdomain region and in particular the highly conserved cysteine residues. They also mutated two other conserved cysteine residues in the central domain (see Figure 1.8.2). They found that alteration of any of the cysteine residues resulted in an inactive protein indicating that the interdomain region and in particular the four conserved cysteine residues were indispensible to $NifA_{Bj}$ function. *Fischer et al (1988)* also found that $NifA_{Bj}$ was sensitive to chelating agents and that this inhibition could be overcome by the addition of divalent metal ions. In light of the above finding *Fischer et al (1988)* proposed a model in which the four conserved cysteines coordinated a metal cofactor and that the redox state of the NifA-metal complex was responsible for regulating $NifA_{Bj}$ activity. *Morett et al (1991)* found that oxygen inhibited the ability of $NifA_{Bj}$ to both bind to its UAS and its ability to activate transcription. They proposed that the redox-state of a NifA-metal complex might determine its conformation and thus its ability to activate transcription.

The inherent oxygen sensitivity of the $NifA_{Bj}$ protein plays a role in regulating the transcription of the *nifA* gene itself in response to oxygen. In *B. japonicum* expression of the *nifA* gene is autoregulated by $NifA_{Bj}$. Under conditions of high oxygen tension *nifA* is only expressed at very low levels while under conditions of low oxygen tension the $NifA_{Bj}$ is active and activates expression of the *nifA* gene resulting in maximal expression of the NifA protein *(Morett et al 1988)*. Like $NifA_{Bj}$ the *S. meliloti* NifA ($NifA_{Sm}$) is also inherently sensitive to oxygen and thus its activity in activating the transcription of *nif* genes is directly controlled by oxygen.. Transcription of the *nifA* gene is also regulated but in a different manner to that observed in *B. japonicum*. In *S. meliloti nifA* transcription in regulated by a two component regulatory system comprising of the proteins FixL and FixJ. FixL is a membrane bound protein which senses cellular oxygen status and under low

oxygen tension it phosphorylates the FixJ protein. The phosphorylated FixJ protein then activates transcription of the *nifA* gene *(Agron et al, 1993).*

- **Figure 1.8.2 :Comparison Of Deduced NifA Amino Acid Sequences**

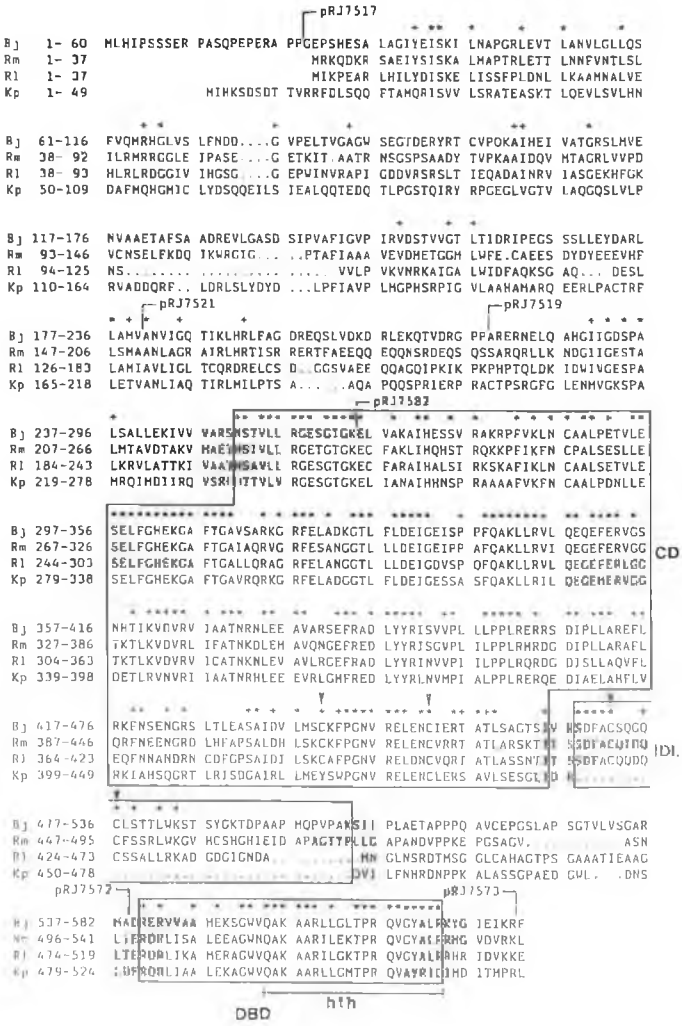


Figure 1. Comparison of the deduced amino acid sequences of the B.japonicum (Bj), R.meliloti (Rm), R.leguminosarum biovar viciae (Rl) and K.pneumoniae (Kp) NifA proteins. For references see text. Identical amino acids in all four proteins are marked by an asterisk (*) and identical amino acids occurring only in the rhizobial NifA proteins by the symbol (+). Functional domains originally proposed by Drummond et al. (7) were modified to boxes termed central domain (CD), interdomain linker (IDL) and DNA binding domain (DBD). The conserved helix-turn-helix (hth) motif common to many DNA binding proteins is denoted by a horizontal bar within DBD. The endpoints of the truncated NifA proteins encoded by the plasmids shown in Fig. 2 are indicated by the respective plasmid numbers. In the Bj NifA protein sequence the cysteine residues that were converted to serine residues by site-directed mutagenesis (see also Table 2) are marked with vertical arrowheads.

1.9: Project Outline.

As was already described in section 1.6 the aim of the work undertaken in this thesis was to construct a novel P2H system based on the NifA protein of *S. meliloti* 2011. The planned system would be homologous in structure to the Y2H system.

The first step in the project would involve the construction of plasmid vectors for the P2H system. Three Prey plasmids would be constructed using the sequence for the transcriptional activating N-terminal domain-central domain (NC) portion of NifA. In these plasmids the NC portion of the NifA protein would be expressed from a strong *ptac* promoter. C-terminally to the sequence for the NC portion of NifA would be a multiple cloning site (MCS) into which a library of DNA fragments could be cloned to generate a library of hybrid proteins bearing the activation domain of NifA. The NC portion of the NifA protein was used rather than just the central activating domain of the NifA protein by itself because the N-terminal domain of the protein had been shown to stabilize the protein when expressed in *E. coli (Huala and Ausubel, 1989)*. The three Prey plasmids would differ from each other only in the reading frame of the multiple cloning site relative to the upstream sequence for the NC portion of NifA. Another Control plasmid expressing the whole NifA protein from the *ptac* promoter would also be constructed. This could be used as a positive control when performing P2H assays

In addition to the Prey plasmids two Bait plasmids would be constructed. The bait construct present on these plasmids would be constructed using the C-terminal DNA binding domain of NifA. It would consist of a *ptac* promoter followed by a MCS into which the sequence for a protein of interest could be cloned. C-terminally to the MCS would be the sequence for the DNA binding domain of NifA. Cloning into the MCS would result in the expression of a hybrid protein bearing the DNA binding domain of NifA. In addition to the bait construct these plasmids would also contain reporter gene constructs. One of the Bait plasmids would contain a *nifH:lacZα* reporter gene construct while the other would contain a *nifH:gent*^{r} (gentamycin resistance) reporter gene. These Bait plasmids would be compatible with the constructed Prey and Control Plasmids

In addition to the construction of these above mentioned plasmids a suitable strain of *E. coli* would also be developed in which the plasmids could be used. This strain would have

to have the *glnG* gene deleted since the product of this gene, NtrC, had been observed to activate the *nifH* promoter *(Buck et al, 1985)* and would thus could give rise false positives.

Having constructed the plasmids and developed suitable strains for the P2H system, The system would be tested for it ability to detect protein-protein interactions *in vivo* using a pair of proteins known to interact as a model. The details of how all of the above plasmids would be made can be found in the following chapters of this thesis.

Chapter 2

Methods And Materials.

2.1: Bacterial Strains, Primer Sequences And Plasmids.

The bacterial strains, primer sequences and plasmids used in this study are described in tables 2.1, 2.2, and 2.3 respectively.

Table 2.1: Bacterial Strains.

Strain	Phenotype/Genotype	Source / Reference
Sinorhizobium meliloti 2011		*Meade et al, 1982.*
Escherichia coli strains		
DH5α	F⁻, *recA*1, *hsdR*17(r_K^-, m_K^+), *supE*44, ϕ80d*lacZ*ΔM15 Δ(*lacZYA-argF*)U169.	Bethesda research Laboratories
XL1-Blue	*recA*1, *hsdR*17(r_K^-, m_K^+), *supE*44, *lac*, [F', *proAB*$^+$, *lacI*^{q}, *lacZ*ΔM15::Tn10(Tet^r)]	Stratagene
INVαF'	F', *recA*1, *hsdR*17(r_K^-, m_K^+), *supE*44, ϕ80d*lacZ*ΔM15, Δ(*lacZYA-argF*)U169.	Invitrogen
YMC11	*supE*44, *hsdR*17(r_K^-, m_K^+), Δ(*glnA-glnG*)2000, Δ*lac*U169.	*Chen et al, 1982.*
YMC11*rif*	Rifampicin mutant of YMC11	*This work*
YMC11*rif*F'	Same as YMC11*rif*, [F', *proAB*$^+$, *lacI*^{q}, *lacZ*ΔM15::Tn10(Tet^r)].	*This work*

Table2.1 continued

ET6016	Δ(*glnG-glnL*?)228, *strR*, Δ*lac*U169,	*E. coli* genetic stock center.
ET6016*rif*	Rifampicin mutant of ET6016	*This work*
ET6016*rif*F'	Same as 6016*rif*, [F ', *proAB*$^{+}$, *lacI*^{q}, *lacZ*ΔM15::Tn10(Tet^{r})].	*This work*

Table 2.2: Primers Sequences.

Primer Name	Primer Sequence 5' 3'	Conc. (nmol/μl)	Tm (°C)
***gent*-F**	**(GAATTC)** -A- TTA CGC AGC AGC AAC GAT GTT	0.60	62
***gent*-R**	**(AAGCTT)** - TTA GGT GGC GGT ACT TGG GT	0.60	62
***nifH:gent'*-F**	**(AAGCTT)** - GCA TGC CTG CAG CGC CCA	0.60	62
***ptac*-F**	**(TCTAGA)** - GAG CTT ATC GAC TGC ACG GT	0.60	64
***ptac*-R**	**(CCCGGG)** - TTT CCT GTG TGA AAT TGT TAT CCG	0.60	62
NC-F1	TCCC **(CCCGGG)** - ATG GCC CCC ACT CGT CTT GAG	0.87	68
NC-R	CCG **(GGATCC)** - GGG CTC TTT CGG CGG	0.91	72
NC-F2	**(CCCGGG)** - ATG GCC CCC ACT CGT CTT GAG ACC ACG C	0.60	92
C-term-F	**(GGATCC)** - GCA GGA GTG GCA TCC	0.63	70
C-term-R	**(AAGCTT)** - GAC GGA GAA AAG AGG CGA CGC	0.65	68
Bate-F	**(CGGCCG)** - ATC CGG AGC TTA TCG ACT	0.60	54
Bate-R	ACG **(GTCGAC)** - GCG GAT ACA TAT TTG AAT GTA	0.60	56
***nifH*-F1**	AAAA **(CTGCAG)** - CGC CCA TAC GAC ACT GTC CGT	0.86	68
***nifH*-F2**	**(GTCGAC)** - GCA TGC CTG CAG CGC CCA	0.60	62
***nifH*-R**	CCA **(GAATTC)** - AA - CAT CTT GCT TCC TTT GTT GTT	0.82	58
HZ -F	CC **(AAGCTT)** - GCA TGC CTG CAG	0.69	64
HZ -R	**(AAGCTT)** - CGG CAT CAG AGC AGA TTG TAC	0.69	64
P1H-F	**(GGATCC)** - GTT GTT CGC TCA ACC ATC TGG	0.60	60
***tir*- F**	**(GGATCC)** - CAC AGG AAA GAA TTC ATG CCT	0.60	80
***tir*-R**	**(GGATCC)** - AAC GAA ACG TAC TGG	0.60	66
***280α*-F**	**(GGATCC)** - ATT ACT GAG ATT AAG GCT	0.0095	48
***280α*-R**	**(AAGCTT)** - TTA TTT TAC ACA AGT GGC	0.0182	48
***280C/Aα*-R**	**(AAGCTT)** - TTA TTT TAC AGC AGT GGC	0.60	50

Note: Underlined bases are bases included for maintenance of reading frames.

Added restriction sites are in bold type and are bracketed.

Table 2.3: Plasmids.

Plasmid	Description	Source/Reference.
pCR2.1TA	Cloning Kit Vector: Amp^R, Km^R, *lacZα*	Invitrogen
pKK223-3	*ptac* promoter expression vector pBR322 derivative, Amp^R.	Pharmacia
pACYC184	p15a origin of replication, Cm^R, Tc^R.	New England Biolabs
pUC19	Amp^R, *lacZα*	Pharmacia
pJQ200ks	Gm^R, *sacB.*	*Quandt (1993)*
pSUP5011	pBR322 derivative, Tn5-mob, Cm^R, Tc^R, Km^R.	*Simon (1984)*
pICC10	pGBT9 expressing *tir*	Gift From Gad Frankel
pICC19	pGAD424 expressing *int280α*	Gift From Gad Frankel
pICC20	pGAD424 expressing *int280C/Aα*	Gift From Gad Frankel
	pCR2.1 Derived Vectors	
pPC1A	*nifH* promoter PCR product in pCR2.1 (amplified as a *Pst*I-*Eco*RI fragment)	*This work*
pPC1B	*nifH* promoter PCR product in pCR2.1 (amplified as a *Sal*I-*Eco*RI fragment)	*This work*
pPC2	*nifH:lacZα* PCR product in pCR2.1	*This Work*
pPC3	NifA C-terminal domain PCR product in pCR2.1	*This work*
pPC4A.	Bait construct PCR product in pCR2.1 (amplified as a *Sal*I fragment)	*This work*
pPC4B.	Bait construct PCR product in pCR2.1 (amplified as a *Eag*I-*Sal*I fragment)	*This work*

Table 2.3 continued

pPC5.	*nifA* PCR product in pCR2.1	*This work*
pPC6A, B&C.	NifA N-terminal-central domain PCR product in pCR2.1	*This work*
pPC7	*gent*^{r} gene PCR product in pCR2.1	*This work*
pPC8	*nifH:gent*^{r} PCR product in pCR2.1	*This work*
pPC9	*ptac* promoter PCR product in pCR2.1	*This work*
pPC10	*nifH:lacZα* PCR product in pCR2.1 (one-hybrid system reporter construct)	*This work*
pPC11	*int280α* gene PCR product in pCR2.1 (amplified as a *Bam*HI-*Hind*IIII fragment)	*This work*
pPC12	*int280* C/Aα gene PCR product in pCR2.1 (amplified as a *Bam*HI-*Hind*IIII fragment)	*This work*
pPC13	*tir* gene PCR product in pCR2.1 (amplified as a *Bam*HI fragment)	*This work*

pKK223-3Derived Vectors.

pPC223	pKK223-3 with *Bam*HI site outside the MCS destroyed and replaced by a unique *Xba*I site.	*This work*
pPC224	NifA C-terminal domain from pPC3 in pPC223.	*This work*
pPC225	*Bam*HI Fragment from pSUP5011 cloned in *Bam*HI site of pPC223.	*This work*
pPC226	*gent*^{r} gene from pPC7 in pPC223.	*This work*
pPC227	*nifH* promoter from pPC1B in pPC226.	*This work*
pPC228	**Final NifA Control Vector.** *nifA* from pPC5 in pPC223.	*This work*

pPC229A, B&C	**Final Prey Vectors.** NifA N-terminal-central domain from pPC6A, B&C in pPC223.	*This work*
pPC230	**Alternative Control Vector.** *ptac* promoter from pPC9 in pPC228.	*This work*
pPC231A, B&C	**Alternative Prey Vectors.** *ptac* promoter from pPC9 in pPC229A,B&C.	*This work*
pPC232	*tir* gene from pICC10 in pPC223.	*This work*
pPC233	*int280α* from pPC11 in pPC229A.	*This work*
pPC234	*int280C/Aα* from pPC12 in pPC229A.	*This work*

pACYC184 Derived Vectors.

pPC184	pACYC184 with its unique *Bam*HI site destroyed. Plasmid is no longer Tet^R	*This work*
pPC185A&B	*nifH:lacZα* from pPC2 in pPC184	*This work*
pPC186A&B	*nifH:gent*^{r} from pPC8 in pPC184	*This work*
pPC187A&B	**Final *nifH:lacZα* Bait Vectors**. Bait construct from pPC4-B in pPC185A&B respectively	*This work*

Table 2.3 continued

pPC188A&B	**Final *nifH:gent* Bait Vectors.** Bait construct from pPC4B in pPC186A&B respectively.	*This work*
pPC189	**One-Hybrid System Bait Vector.** *nifH:lacZα* frompPC10 in pACYC184	*This work*
	pUC 19 Derived Vectors.	
pPC19	*nifH* promoter from pPC1A in pUC19	*This work*

2.2: Microbiological Media.

Solid complex media contained 15g/L Oxoid #3 agar. Solid minimal media contained 15g/L Oxoid#1 purified technical agar. Tryptone and Yeast extract were also from Oxoid. Other chemicals were from Sigma Chemicals Co. and BDH Chemicals LTD. All chemicals were analar grade.

Distilled water was used to prepare all media and sterilisation was achieved by autoclaving at 15lb/in^2 for 20min.

- **TY Medium(Beringer, 1974).**

Used for the routine culturing of fast growing *Sinorhizobium* strains.

Tryptone	5 g
Yeast Extract	3 g
$CaCl_2 \cdot 2H_2O$	0.7 g

Adjusted to pH 7.0 with NaOH and volume brought to 1L with dH_20. The solution was then sterilised by autoclaving.

- **Luria Bertani Broth (LB).**

Used for routine culturing of *E. coli.*

Tryptone	10 g
Yeast Extract	1 g
NaCl	10 g

Adjusted to pH 7.0 with NaOH and volume brought to 1L with dH_20. The solution was then sterilised by autoclaving.

- **SOB Medium.**

Tryptone	20 g
Yeast Extract	5 g
NaCl	0.5 g
KCl	2.5 mM
dH_2O	1 L
pH	7.0

After autoclaving the solution was allowed to cool to 55 °C and sterile solutions of $MgCl_2$ and Mg_2SO_4 were added to final concentrations of 10mM.

- **SOC Medium.**

  After making SOB as above 7.2ml of 50% sterile glucose was added to give a final concentration of 20mM.

2.3: Solutions And Buffers.

- **TE Buffer.**

Tris-HCl	10 mM
Na_2-EDTA	1 mM
pH	8.0

- **TES Buffer.**

Tris-HCL	10 mM
Na_2-EDTA	1 mM
NaCl	50 mM
pH	8.0

- **STET Buffer (100 ml)** ***(Holmes and Quigley, 1981).***

Tris-HCl	50 mM (5ml of 1M solution)
Na_2-EDTA	50 mM (10ml of 0.5M solution)
Trition X 100	5 % (5ml)
Sucrose	8 % (8g)
pH	8.0

- **Solutions For The 1,2,3 Method *(Birnboim and Doly. 1979).***

Solution 1

Glucose (0.5M)	1 ml
Na_2-EDTA (0.1M)	1 ml
Tris-HCl (1M)	0.25 ml
dH_2O	7.75 ml

Solution 2

NaOH (1N)	2 ml
SDS (10%)	1 ml
dH_2O	7 ml

made up fresh every month and stored at room temp.

Solution 3

Potassium acetate	3 M
pH	4.8

To 60ml of 5M Potassium acetate add 11.5ml of glacial acetic acid and 28.5ml of dH_2O. The resulting solution is 3M with respect to Potassium and 5M with respect to acetate.

Kirby Mix

Phenol	100 ml
Chloroform	100 ml
Isoamyl alcohol	4 ml
8-Hydroxy Quinolone	0.8 g

Stored under 100mM Tris-HCl, pH 7.5 at 4 ^{o}C.

- **50 x Tris Acetate EDTA (TAE) Buffer.**

EDTA (0.5M)	100 ml
Glacial Acetic Acid	57.1 ml
Tris	242 g
pH	8.0

diluted to 1X with dH_2O before use.

- **Gel Loading Dye (6X).**

Bromophenol Blue	0.25 %
Xylene cyanol	0.25 %
Sucrose (w/v)	40 %

Made in dH_2O and stored at 4 °C.

- **Solutions For Gene Clean Procedure (*Mather et al*, 1993).**

Sodium Iodide

NaI	90.8 g
dH_2O to	100 ml
Na_2SO_3	1 g

The solution was stirred until as much of the sodium iodide as possible had dissolved and was then filtered through a 0.45µm filter. Then 0.5g of Na_2SO_3 was added (did not dissolve) and the solution stored at 4 °C in the dark.

New Wash Solution

Ethanol	50 %
Tris-HCl, pH 7.5	10 mM
Na_2-EDTA	1 mM
NaCl	100 mM

the solution was stored at -20 °C

- **TB Buffer For Competent Cells *(Inoue et al, 1990).***

Pipes	10 mM
$CaCl_2$	15 mM
KCL	250 mM
dH_2O	1 L
pH with KOH	6.7

Once the pH had been adjusted, $MnCl_2$ was added to a final concentration of 55mM. The solution was then filter sterilised through a 0.45µm sterile filter and stored at 4 °C.

- **Z Buffer For β-Galactosidase Activity Assay *(Miller, 1972).***

Na_2HPO_4	16.1 g
NaH_2PO_4	5.5 g
KCL	0.75 g
$MgSO_4$	0.246 g
Mercaptoethanol	2.7 ml
dH_2O	1 L

Do not autoclave this solution

- **8% L-Glutamine Stock Solution.**

  The solution had to be heated to 60 °C in order to get the glutamine to dissolve. The solution was then filter sterilised through a 0.28μm sterile filter. Upon cooling the glutamine precipitated out of solution. The solution therefore had to be heated to dissolve the glutamine prior to its addition to media.

2.4: Antibiotics.

Antibiotics used were from Sigma. Antibiotic stocks were prepared to a concentration of 100mg/ml and stored at -20 °C

Ampicillin was prepared in dH_2O and was used at a final concentration of 100μg/ml in solid media and 25μg/ml in liquid broth.

Chloramphenicol was prepared in ethanol and used at a concentration of 25μg/ml in both solid and liquid media.

Tetracycline was prepared in 50% ethanol and stored in the dark as it is light sensitive. It was used at a concentration of 10μg/ml in both solid and liquid media.

Kanamycin was prepared in dH_2O and was used at a final concentration of 50μg/ml in both solid media and liquid media.

Rifampicin was prepared in methanol and was stored in the dark as it is light sensitive. It was used at a concentration of 100μg/ml in both solid and liquid media.

2.5: Storing And Culturing Bacteria.

Strains were stored as glycerol stocks. A 0.5ml aliquot of a late log phase culture was added to an equal volume of sterile glycerol in microfuge tube which was then mixed and stored at -20 °C. A duplicate set of long term stocks were stored at -80 °C. Where

hosts were harbouring plasmids, the appropriate antibiotic was added to the stock medium. Working stocks were stored on plates at 4 °C.

2.6: Phenol Preparation.

A 500g container of phenol was placed in a 68 °C water bath and heated until the phenol had melted. 8-hydroxyquinolone was added to a final concentration of 0.1%. This yellow compound is an antioxidant, a partial inhibitor of RNase and a weak chelator of metal ions. In addition the yellow colour provides a convenient way to identify the phenol phase during separations.

The melted phenol was then extracted several times with an equal volume of Tris buffer. The first extraction was carried out with 1M Tris pH 8.0. Subsequent extraction steps were carried out using 0.1M Tris pH 8.0 to which 0.2% β-mercaptoethanol had been added. This was continued until the pH of the aqueous phase was >7.6. The phenol solution was then stored at 4 °C under 100mM Tris-HCl pH 8.0.

2.7: Plasmid Preparation By The 1,2,3 Method.

This method was described by *Birnboim and Doly (1979).* A 1.5ml aliquot of a bacterial culture grown in selective media was pelleted at 4000rpm in a microfuge and the supernatant removed. The pellet was resuspended by vortexing in 200µl of solution 1 and was then left for 5min at room temperature. Then 200µl of solution 2 was added, the tube was mixed by inversion and placed on ice for 5min. Then 200µl of solution 3 was added, the tube was mixed by inversion and placed on ice for 10min. A clot of chromosomal DNA formed and was pelleted by centrifugation at 13,000rpm in a microfuge for 10min. The supernatant (600µl) was placed into a fresh tube and 400µl of Kirby mix was added and mixed by vortexing. After centrifugation

at 13,000rpm for 5min the aqueous layer was removed to a fresh tube and an equal volume of isopropanol was added. After 10min at room temperature the tube was centrifuged for 10min at 13,000rpm to pellet plasmid DNA. The pellet was washed twice with 70% ethanol, dried briefly in a vacuum dryer and resuspended in 50µl of TE buffer. Plasmid preps were stored at 4 °C.

2.8: Plasmid Preparation By Rapid Boiling Method.

This method was described by *Holmes and Quigley (1981)* and used instead of the 1,2,3, procedure outlined above for the screening of large numbers of transformants. A 3ml aliquot of an overnight culture was spun at 4000rpm in a microfuge for 5min and the supernatant removed. The pellet was resuspended in 350µl of STET buffer. A 20µl aliquot of a 10mg/ml lysozyme solution (prepared fresh in STET buffer) was added and the microfuge tube incubated at 30 °C for 10min. The tube was then placed in a boiling water bath for 60sec and then spun at 10,000rpm for 10min. The supernatant was removed to a fresh microfuge tube and an equal volume of isopropanol was added. The tube was left at room temperature for 10min and then the plasmid DNA was pelleted by centrifugation at 13,000rpm for 10min. The pellet was washed twice with 70% ethanol, dried briefly in a vacuum dryer and then dissolved in 50µl of TE buffer. Plasmid DNA was then stored at 4 °C.

2.9: Preparation Of Total Genomic DNA From *Sinorhizobium.*

Ten ml of an early stationary phase culture of *Sinorhizobium meliloti* 2011 was sedimented in a Bechman J2-21 centrifuge at 10,000rpm for 5min. The cells were washed with an equal volume of TES buffer and then resuspended in 5ml of TE buffer. A 0.5ml aliquot of freshly prepared lysozyme solution (2mg/ml in TE) was added and the suspension incubated at 30 °C for 20min. A 0.5ml aliquot of a

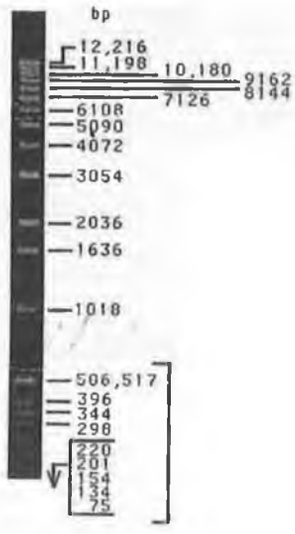
sarkosyl/proteinase solution (10% sarkosyl in TE containing 5mg/ml proteinase) was then added and the suspension was incubated for 1 hour. Lysis was evident by an increase in the viscosity of the suspension. A 0.7ml aliquot of 3M sodium acetate was added and the suspension mixed gently. Then 2.5ml of Kirby mix was added and the suspension was mixed slowly by inversion for 15min. After centrifugation at 10,000rpm for 10min the aqueous phase was removed to a glass corex tube and centrifuged at 10,000rpm for another 15min. The supernatant was removed to a clean plastic centrifuge tube and 2.5ml of Kirby mix was added. The suspension was mixed gently for 15min and before extraction with chloroform:isoamylalcohol(24:1). The DNA was precipitated by the addition of an equal volume of isopropanol and was evident as a coiled thread. The DNA was sedimented by centrifugation at 10,000rpm for 15min, washed twice with 70% ethanol, dried under vacuum and resuspended in 400µl of TE.

2.10: Agarose Gel Electrophoresis For DNA Characterisation.

DNA was analysed by running on agarose gels in a horizontal gel apparatus. Gels were prepared by dissolving agarose in 1xTAE buffer to the required concentration (typically 0.7-1.2%) and boiling until the solution went clear. The 1xTAE buffer was also used as the running buffer. A tracker dye was incorporated into DNA samples (2µl to 5-7µl of sample) to facilitate loading of samples. Mini-gels were frequently run at 100Volts for 1hour or until the tracker dye reached the base of the gel while maxi gels were frequently run at 40Volts overnight. Gels were stained by immersing in a bath of ethidium bromide for 30min and then destaining in a water bath for 10min. Gels were then visualised on a UV transilluminator and photographed using a UV image analyser.

All agarose gels shown in the results chapters of this thesis were 0.7% agarose unless otherwise stated. A standard 1Kb ladder was run on all gels to facilitate estimation of DNA fragmnent size. Figure 2.10.1 below shows a photograph of this ladder run on a 1% agarose gel and indicates the fragment sizes within the ladder.

- **Figure 2.10.1: Fragment Sizes Within The 1Kb ladder.**



2.11: Preparation Of Ethidium Bromide.

One 10mg tablet of ethidium bromide was added to 1ml of water. The container was wrapped in tin foil and stored at 4 °C. A 100µl aliquot of this stock solution was added to 1L of dH_2O for staining agarose gels. Gloves were worn at all time when handling solutions containing ethidium bromide. Ethidium bromide waste was collected and treated with activated charcoal before being filtered through 3mm Whatman filter paper. The clear liquid was disposed of normally and the solids contained on the filter paper were incinerated.

2.12: Preparation Of Silica 325 Mesh Glass Beads For Gene Clean Procedure.

Approximately 250ml of silica powder was suspended in distilled water and the volume made up to 500ml in a 600ml beaker. The suspension was stirred for 1 hour and then allowed to settle for 1 hour. The supernatant was then removed to two centrifuge tubes and spun at 5000rpm in a Bechman J2-21 centrifuge for 5min. The pellet was resuspended in 100ml of distilled water and transferred to a 600ml beaker

and the volume adjusted to approximately 140ml. Then 350ml of 70% HNO_3 was added to give a final concentration of 50% HNO_3. The suspension was heated to near boiling on a hot plate with continuous stirring. The suspension was then transferred to two centrifuge tubes and spun at 5000rpm for 5min. Each pellet was then resuspended in 100ml of distilled water and spun down at 5000rpm for 5min. This was repeated three times. The volume of the final pellet was then estimated, resuspended in an equal volume of distilled water and the suspension stored at 4 °C.

2.13: Gene Clean Procedure For Isolation Of DNA From Agarose Gels.

This method was described by *Mather et al* (1993). The band to be isolated from an agarose gel was excised with a clean sharp scalpel and placed in a microfuge tube. The excised piece of agarose was weighed and then 2-3 volumes of saturated NaI was added to the agarose. The agarose was then incubated at 55 °C for 15 min or until the agarose had dissolved. Approximately 2µl of silica 325 mesh glass beads were added and mixed by vortexing. This suspension was then left at -20 °C for approximately 15min and then spun for 5sec at 13,000rpm in a microfuge. The supernatant was then removed and discarded. The pellet was washed three times with 300µl of ice cold New Wash solution to remove residual agarose and salt contaminants. The pellet was resuspended in a desired volume of TE buffer and incubated at 55 °C for 15min to elute the DNA. The suspension was spun at 13,000rpm for 1min in a microfuge to pellet the Silica beads. The supernatant was removed to a fresh microfuge tube and stored at 4 °C.

2.14: Preparation Of Competent Cells By $CaCl_2$ Treatment.

One ml of a 5ml overnight culture of an appropriate strain of *E. coli* was added to 100ml of LB broth and grown with shaking at 37 °C until an OD_{600} of 0.3-0.4 was

reached. The culture was then chilled on ice for 30min. A 10ml aliquot was then pelleted by centifugation in a Bechman J2-21 centrifuge at 4000rpm and 4 °C for 5min. The pellet was then resuspended in 5ml of ice-cold 100mM $MgCl_2$ in order to wash the cells. The cells were pelleted as before and then resuspended in 5ml of ice-cold 50mM $CaCl_2$ and the suspension was placed on ice for 30min. The cells were pelleted again, resuspended gently in 1ml of ice-cold 50mM $CaCl_2$ and placed on ice overnight. The cells were then competent.

For storage the cells were mixed with an equal volume of sterile glycerol, split into 200µl aliquots in microfuge tubes, flash frozen to -80 °C and then stored at -80 °C.

2.15: Transformation Of Competent Cells Prepared By $CaCl_2$ Treatment.

A 3-5µl aliquot of a ligation mixture was added to 200µl of competent cells that had been allowed to thaw on ice. The cells were left on ice for 30min and were then heat shocked at 42 °C for exactly 2min before being returned to ice for a further 2min. A 0.8ml aliquot of fresh SOC medium was added to the cells and the cells were incubated at 37 °C for 1 hour. A 100-200µl aliquot of the resulting transformation mixture was then plated on appropriate selective media and plates were incubated at 37 °C for 24 hours.

2.16: Preparation Of High Efficiency Competent Cells.

This method was described by *Inoue et al (1990).* A frozen stock the appropriate *E. coli* strain was thawed, streaked on LB agar and incubated at 37 °C overnight. Approximately 10 to 12 large colonies were removed with an inoculating loop and inoculated into 250ml of SOB medium in a 2L baffled flask. The culture was grown at 18 °C with vigorous shaking (200-250rpm) until an OD_{600} of 0.6 was reached. The

flask was then placed on ice for 10min. The culture was transferred to a 500ml centrifuge bottle and spun in a Bechman J2-21 centrifuge at 4,000rpm and 4 °C for 5min. The pellet was resuspended in 80ml of ice-cold TB buffer, placed on ice for 10min and then spun down as before. The cell pellet was gently resuspended in 20ml of ice cold TB buffer and DMSO was added slowly with gentle swirling to a final concentration of 7%. After incubation in an ice bath for 10min the cell suspension was dispensed in 1ml aliquots into microfuge tubes. The cells were then flash frozen in liquid nitrogen and stored at -80 °C. Cells prepared in this manner frequently gave transformation efficiencies of the order of 10^8-10^9 transformants/μg DNA which is comparable with those attainable by electroporation.

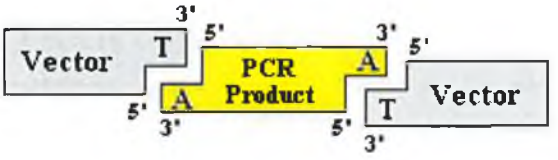
2.17: Transformation Of High Efficiency Competent Cells.

A 1-5μl aliquot of a plasmid preparation was added to 200μl of competent cells prepared according to the procedure outlined in section 2.16. Polypropylene tubes were used as glass tubes decreased (10 fold) the competence of cells. The cells were incubated in an ice bath for 30min. The cells heat shocked at 42 °C for 30sec and then transferred back to ice for 2min. After the addition of 0.8ml of fresh SOC medium the cells were incubated at 37 °C with vigorous shaking for 1 hour. A 100μl aliquot of the resulting transformation mixture was plated on an appropriate selective media and the plates were incubated at 37 °C overnight.

2.18: TA Cloning Of PCR Products.

PCR products were routinely cloned using Invitrogens Original TA Cloning Kit vector pCR2.1. The diagram below shows the concept behind the TA Cloning method.

- **Figure 2.18.1: Principle Of TA Cloning**



The method is dependent on the fact that thermostable polymerases like *Taq* DNA polymerase, which lacks 3'-5' exonuclease activity, leave 3' A-overhangs. PCR products generated with *Taq* DNA polymerase have a high efficiency of cloning in the TA Cloning system. Other thermostable polymerases like *Vent* and *Pfu,* which have 3'-5' exonuclease activity, do not leave these 3' A-overhangs.

PCR products were amplified using a standard PCR reaction mixture (see section 2.21) and using *Taq* DNA polymerase from Promega. They were subsequently ligated with the TA pCR2.1 vector. The ligation was set up as follows;

Fresh PCR Product	2 µl
PCR2.1 Vector (25ng/µl)	2 µl
10X Ligation Buffer	1 µl
Sterile dH_2O	4 µl
T4 DNA Ligase (4.0 Weiss units)	1 µl
Total Volume	10 µl

The reaction was then incubated at room temperature overnight. Following incubation, 2-5µl of the ligation was used to transform either DH5α cells prepared by the high efficiency method (section 2.17) or INVαF' one shot competent cells that were supplied with the TA Cloning Kit.

To transform INVαF' cells, the cells were first thawed on ice. Then 2µl of β-mercaptoethanol (0.5M) was added and mixed gently with the pipette tip. Between 2-5µl of the ligation reaction mixture was added to the cells and mixed in gently with the pipette tip. The cells were incubated on ice for 20min and then heat shocked at

42 °C for 30sec. The cells were placed back on ice for 2min and then 250µl of SOC media was added. The cells were incubated at 37 °C for 1 hour. A 50µl aliquot of the transformation mixture was plated on LB agar containing ampicillin (100µg/ml) and Xgal to select for transformants and to test for α-complementation of β-galactosidase. In addition to an ampicillin resistance gene the TA pCR2.1 vector also contained a kanamycin resistance gene. Kanamycin (50µg/ml) was thus used to select for transformants instead of ampicillin when PCR products amplified from ampicillin resistant plasmids were being cloned.

2.19: Testing For α-Complementation Of β-Galactosidase.

Cloning in the TA vector works on the same basis as cloning in pUC based vectors. Transformants bearing plasmids with inserts, or PCR products in the case of TA cloning, were distinguished from transformants bearing recircularised plasmids without inserts by plating on media containing Xgal to test for α-complementaion of β-galactosidase. Transformants harbouring plasmids with inserts did not express a functional *lacZα* gene product, which is the α peptide of β-galactosidase, and therefore appeared white on X-gal plates because α-complementation could not occur. Transformants harbouring recircularised plasmid DNA without any inserts expressed a functional *lacZα* gene product and thus appeared blue on Xgal plates because α-complementation of β-galactosidase occurred resulting in the cleaving of X-gal. White colonies were thus picked from plates and screened for the presence of inserts.

In order to test for α-complementation transformations were plated out on LB agar containing an appropriate antibiotic, usually ampicillin in the case of TA cloning, and Xgal as mentioned above. The Xgal was not incorporated into the media. Instead 40µl of a stock solution of Xgal (40mg/ml in dimethylformamide) was spread on the surface of premade LB agar plates containing ampicillin and the plates were allowed to dry. The stock solution of Xgal was not filter sterilised. It was stored at -20 °C and in the dark in order to prevent damage by light. Transformations were plated out on

the LBampicillin/Xgal plates and the plates were then incubated at 37 °C for 12-16 hours. The plates were then removed from the incubator and stored at 4 °C for several hours in order to allow colour development. White colonies were then picked off plates and screened for the presence of plasmids bearing inserts.

IPTG was not required in the media when transforming strains DH5α or INVαF' because these strains did not contain the *lacI*^{q} allele. The *lacI* gene encodes the Lac repressor which represses expression of the *lacZ* gene from the *lac* promoter. IPTG can be used to relieve this repression and to induce expression of the *lacZα* gene from the *lac* promoter. However, plasmids like pUC18 and pCR2.1 are such high copy number plasmids that the amount of the LacI repressor produced in cells is insufficient to repress expression of all copies of the *lacZα* gene present on the vectors and thus IPTG is not required to induce expression of the *lacZα* gene. On the other hand the *lacI*^{q} allele expresses 10 times as much LacI repressor than the normal *lacI* gene. In strains bearing this allele IPTG has to be used to induce expression and therefore has to be added to the media when testing for α-complementation. A stock solution of IPTG (100mM) is prepared by dissolving IPTG in water. The solution has to be filter sterilised and can then be stored at -20 °C. A 4µl aliquot of this stock solution can be spread on the surface of pre-made LB agar plates along with Xgal when testing for α-complementation.

2.20: Miller Assay For β-Galactosidase Activity.

The method used was as described by *Miller (1972).* A 1ml aliquot of the culture to be tested was centrifuged at 5000rpm in a microfuge for 5min and the cell pellet was resuspended in 0.75ml of Z buffer. The cells were permeablised by the addition of 100µl of chloroform and 50µl of 0.1% SDS. The tube was vortexed for 10sec and then equilibrated at 30 °C for 5min. The reaction was started by the addition of 0.2ml of ONPG (4mg/ml in Z buffer, prepared fresh) and the tube was vortexed again for a 10sec. The reaction was timed for 10-20min and then stopped by the addition of 0.375ml of 1M Na_2CO_3. The cells were then pelleted by centrifugation at 13000rpm

in a microfuge and the OD420 of the supernatant recorded. β-galactosidase activity was calculated using the equation below.

$$\text{Miller units} = \frac{OD420 * 1000}{V * T * OD600}$$

OD420	=	the absorbance of the supernatant at 420nm
OD600	=	the absorbance of the culture used at 600nm
V	=	the volume of the culture used in ml
T	=	the time of the reaction in minutes

2.21: Enzymatic Reactions.

- **Klenow Reaction.**

DNA	18 µl
dNTPs (0.5mM)	1 µl of each
Klenow Buffer (10X)	3 µl
Sterile dH_20	4 µl
Klenow (0.5u/µl)	1 µl

Incubated at room temperature for 1 hour and then the enzyme was heat inactivated at 65 °C for 30min.

- **Calf Intestinal Phosphatase (CIP) Reaction.**

DNA	13 µl
CIP Buffer (10X)	1.5 µl
CIP	0.5 µl

Incubated at 37 °C for 30min and then the enzyme was heat inactivated at 65 °C for 30min.

- **Standard PCR Reaction Mixture.**

$MgCl_2$ (25mM)	3 µl
dNTPs (10mM)	1 µl of each
Thermo Buffer (10X, Mg^{2+} free)	5 µl
Primers (0.6nm/µl)	1 µl of each
Template DNA	1 µl
Sterile dH_2O	34.5 µl
Taq Polymerase (5u/ul)	0.5 µl

The reaction was then overlayed with 50µl of sterile mineral oil. Total genomic DNA from *S. meliloti* 2011 was used as a source of template DNA for PCR reactions unless otherwise stated.

- **Standard PCR Program Cycle.**

Stage 1: **Step 1:** 95 °C for 10 min.

Stage 2: **Step 1:** 95 °C for 1 min.
Step 2: Annealing Temperature for 30sec.
Step 3: 72 °C for 1min for every Kb to be synthesised.
(Stage 2 was usually repeated for 30 cycles)

Stage 3: **Step 1:** 72 °C for 10min.

Chapter 3

Construction Of The Prey And The Control Plasmids For The Prokaryotic Two-Hybrid System.

3.1: Introduction.

As described in chapter one the proposed NifA based Prokaryotic Two-Hybrid **(P2H)** system would consist of the following components;

- Three Prey plasmids into which genes of interest could be cloned in the appropriate reading frame to generate hybrid proteins containing the transcriptional activating N-terminal-central **(NC)** domain of *Sinorhizobium meliloti* 2011 NifA.
- A Bait plasmid carrying a reporter gene construct consisting of the *nifH* promoter fused to the *lacZα* gene. In addition to this it would also carry a Bait construct into which genes of interest could be cloned to generate hybrid proteins containing the C-terminal DNA binding domain of NifA.
- A Control plasmid expressing the whole *S. meliloti* 2011 NifA protein.
- Suitable *E. coli* strains bearing *glnG* (*ntrC*) mutations in which the newly constructed plasmids could be used.

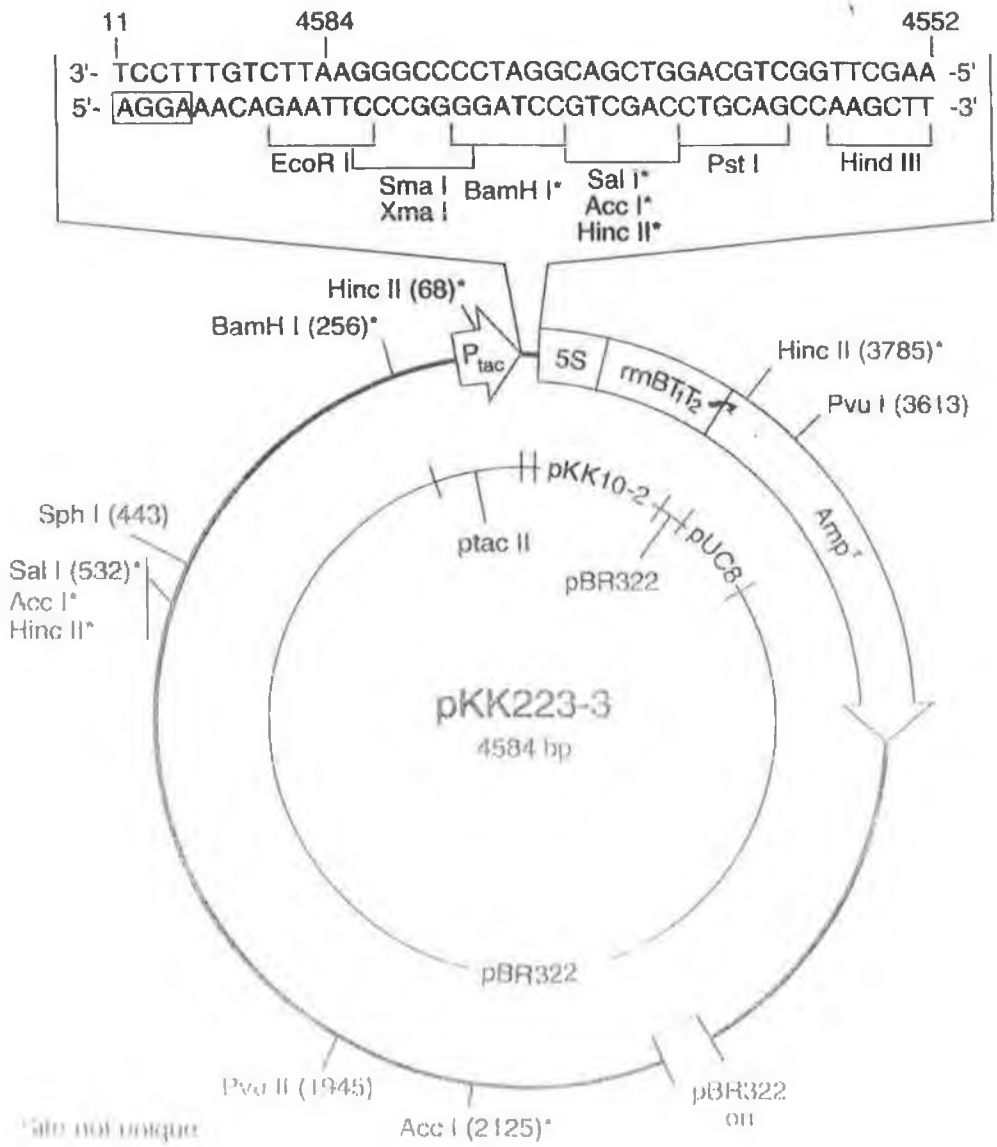
The construction of each of these components is described in the following chapters. This chapter describes the strategy employed for the construction of the Prey plasmids and the Control plasmid as all of these plasmids are related and were constructed in a similar manner.

3.2: Selection Of A Suitable Starting Plasmid For The Construction Of The Prey Plasmids.

The first step in the construction of the Prey plasmids involved the selection of a suitable starting plasmid upon which the Prey plasmids would be based. The expression plasmid vector pKK223-3, which was commercially available from Pharmacia, was selected as the starting plasmid. This plasmid has the following features (see Figure 3.2.1);

- A strong *ptac* promoter that would permit IPTG inducible expression of proteins.

- **Figure 3.2.1: The pKK223-3 Protein Expression Plasmid.**



- A pUC8 multiple cloning site **(MCS)** adjacent to the *ptac* promoter that could be used for the cloning of the transcriptional activation domains of the NifA protein.
- A ribosome binding site **(RBS)** for expression of genes cloned in the MCS.
- An ampicillin resistance gene to allow independent selection for the plasmid.
- A strong *rrn*BT_1T_2 transcriptional termination sequence to prevent read through transcription induced from the *ptac* promoter into the ampicillin resistance gene.

If a gene is cloned into the MCS of the pKK223-3 plasmid so that its start codon is within 13 base pairs of the *Eco*RI in the MCS, the plasmid's RBS could be used for expression of the protein.

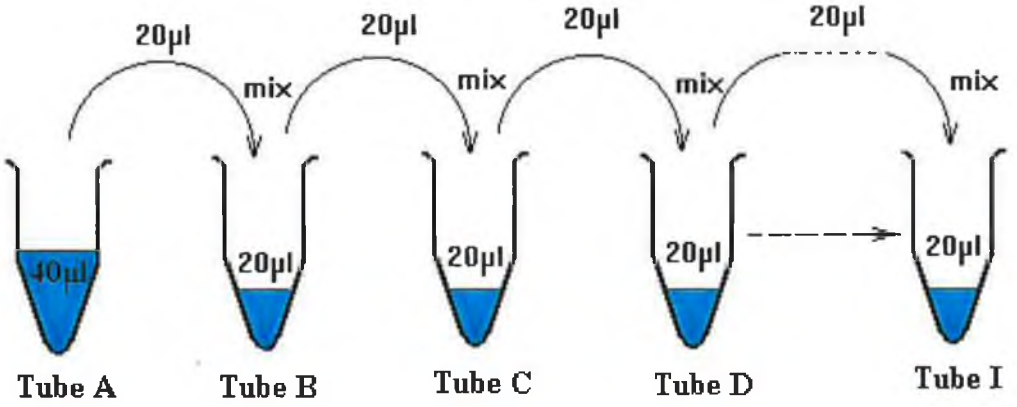
The above features of pKK223-3 made it a desirable starting plasmid for the construction of the Prey and the Control plasmids.

3.2.1: Preparation Of The pKK223-3 Plasmid For Use In The Construction Of The Prev Plasmids.

Inspection of Figure 3.2.1 shows that the pKK223-3 plasmid has two *Bam*HI sites, one within the MCS of the plasmid and one outside of the MCS. A preliminary step in the construction of the Prey plasmids was the elimination of the second *Bam*HI site outside the MCS. This made the *Bam*HI site within the MCS a unique restriction site which was then used for the cloning of the NC domain of the NifA protein to form the Prey plasmid.

The elimination of the *Bam*HI site involved the use of partial digests to generate pKK223-3 plasmid DNA cleaved at only one of its two *Bam*HI sites. A set of serial dilutions of *Bam*HI was set up to determine which concentration of *Bam*HI would yield the greatest proportion of partially digested plasmid DNA. Figure 3.2.2 shows how the set serial dilutions was set up.

- **Figure 3.2.2: Preparation Of BamHI Partial Digests By Serial Dilution.**



An initial *Bam*HI digest was set up as follows;

Tube A Digest.

pKK223-3 plasmid DNA	34 µl
React Buffer 2 (10X)	4 µl
*Bam*HI enzyme (10u/µl)	2 µl

Eight other tubes designated tubes (B) to (I) were set up as follows;

Tube B to I Reaction Mixtures.

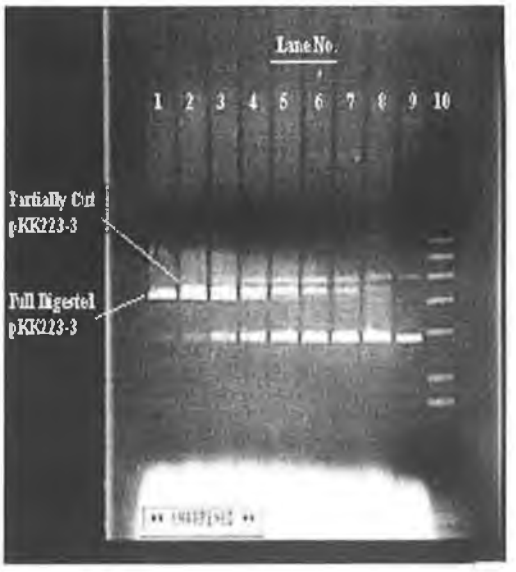
pKK223-3 plasmid DNA	18 µl
React 2 Buffer (10X)	2 µl

As illustrated in Figure 3.2.2, 20µl was taken from the Tube A digest and added to the Tube B reaction mixture. The concentration of *Bam*HI in the resulting Tube B digest was half that of the Tube A digest. Then 20µl was taken from the Tube B digest and added to the Tube C reaction mixture. The concentration of *Bam*HI in the resulting Tube C digest was half that in the preceding Tube B digest. This procedure was continued until a complete set of *Bam*HI digests, A to I, in which each digest had half the concentration of *Bam*HI of the preceding digest, was obtained. These

digests were then incubated at 37 °C for 30min and then separated on a 0.7% agarose gel. A photo of the gel obtained is shown in Figure 3.2.3 below.

- **Figure 3.2.3: *Bam*HI Partial Digests Of pKK223-3.**

Lane No.	Sample	BamHI Conc.
1	Tube A Digest	0.5u/µl
2	Tube B Digest	0.25u/µl
3	Tube C Digest	0.125u/µl
4	Tube D Digest	0.062u/µl
5	Tube E Digest	0.031u/µl
6	Tube F Digest	0.015u/µl
7	Tube G Digest	0.007u/µl
8	Tube H Digest	0.003u/µl
9	Tube I Digest	0.001u/µl
10	1kb Ladder	



As can be seen from the photo of the gel in Figure 3.2.3, the Tube B digest gave two bands that ran very close together. The lower band corresponded to fully digested pKK223-3 DNA that had been cleaved at both of its *Bam*HI sites. The upper band corresponded to partially digested pKK223-3 DNA that had been cleaved at only one of its two *Bam*HI sites. As the concentration of *Bam*HI became more dilute the proportion of partially digested DNA present in the digests increased. It was estimated that the Tube F digest in lane 6 of the gel seen in Figure 3.2.3 gave the greatest proportion of partially digested pKK223-3 plasmid DNA. This digest was therefore set up on a larger scale to give a total volume of 80µl and was incubated at 37 °C for 30mins. The digest was then cleaned by treatment with kirby mix to remove protein followed by precipitation with isopropanol. The plasmid DNA was then resuspended in TE buffer and a Klenow reaction was set up in order to fill in the 5' cohesive ends generated by cutting with *Bam*HI. The DNA was then cleaned again and a ligation with an *Xba*I linker was set up. The DNA was again cleaned and then restricted with *Xba*I. The pKK223-3 plasmid itself did not have any restriction sites for the *Xba*I enzyme and thus digesting with this enzyme only resulted in the cleaving of the linker molecules. This generated *Xba*I cohesive ends on the ends of

the linearised pKK223-3 plasmid DNA. Finally the DNA was cleaned, ligated and then transformed into the *E. coli* strain DH5α. The transformation was plated on LB ampicillin plates to select for transformants containing the recircularised pKK223-3 plasmid DNA.

The above process resulted in the elimination of one of the two *Bam*HI sites present on the original pKK223-3 plasmid, replacing the destroyed site with a unique *Xba*I site. As partial digests were used the partially digested pKK223-3 plasmid DNA would have been cut at either one of the *Bam*HI. Transformants could therefore contain plasmid DNA with either one of the two *Bam*HI sites destroyed. Transformants therefore had to be screened to identify clones containing plasmid DNA with the desired *Bam*HI site destroyed.

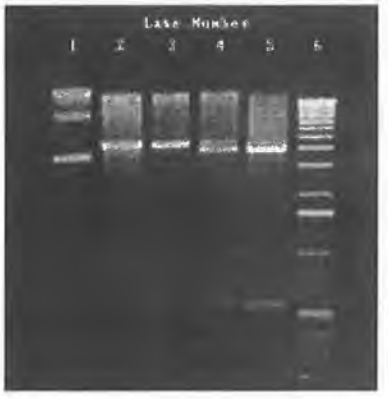
In order to identify transformants containing plasmid DNA with the desired *Bam*HI site destroyed, plasmid DNA was prepared from several transformants. The DNA was initially digested with *Xba*I to verify the presence of linker. *Bam*HI digests were also performed on the plasmid DNA from each of the transformants to verify that in each case only one of the *Bam*HI sites had been destroyed. Having verified this, *Sal*I-*Bam*HI double digests were performed in order to determine which of the two possible *Bam*HI sites had been destroyed in each case. Examination of Figure 3.2.1 shows that performing such a digest on plasmid DNA with the desired *Bam*HI site destroyed (the one lying outside of the MCS) was expected to yield two bands on an agarose gel, one band at approximately 4036bp and one at approximately 543bp. Performing such a *Sal*I-*Bam*HI double digest on plasmid DNA in which the other *Bam*HI site was destroyed (the one in the MCS) was expected to yield 3 bands, one at 4036bp and a doublet at around 280bp. Screening transformants in this manner thus resulted in the identification of the desired plasmid which was called pPC223 (see Figure 3.3.2) and the results of the above mentioned digests on this plasmid can be seen in Figure 3.2.4 below.

As will be described in section 3.3 of this chapter it was intended that the NC domain of the NifA protein was to be amplified by PCR as an *Xma*I-*Bam*HI fragment and then cloned into the pPC223 plasmid as a *Xma*I-*Bam*HI fragment to generate the Prey plasmid. However, examination of Figure 3.2.1 shows that the *Xma*I and *Bam*HI recognition sites in the MCS of the plasmid partially overlap. Cutting the plasmid with either one of the enzymes therefore prevented the other enzyme from

cutting at its site. It was therefore impossible to generate restricted pPC223 plasmid DNA with the correct cohesive ends to facilitate cloning of the NifA NC domain as an *Xma*I-*Bam*HI fragment. In order to overcome this problem, a *Bam*HI fragment approximately 1.8kb in size was cut from plasmid pSUP5011 and cloned into the *Bam*HI site of the pPC223 plasmid to generate plasmid pPC225 (see Figure 3.3.2). Figure 3.2.5 shows the verifying restriction analysis of the pPC225 plasmid.

- **Figure 3.2.4: Restriction Analysis On Plasmid pPC223.**

Lane No.	Sample.
1	pPC223 uncut
2	pPC223 cut with *Xba*I
3	pPC223 cut with *Bam*HI
4	pPC223 cut with *Sal*I
5	pPC223 cut with *Sal*I-*Bam*HI
6	1Kb Ladder

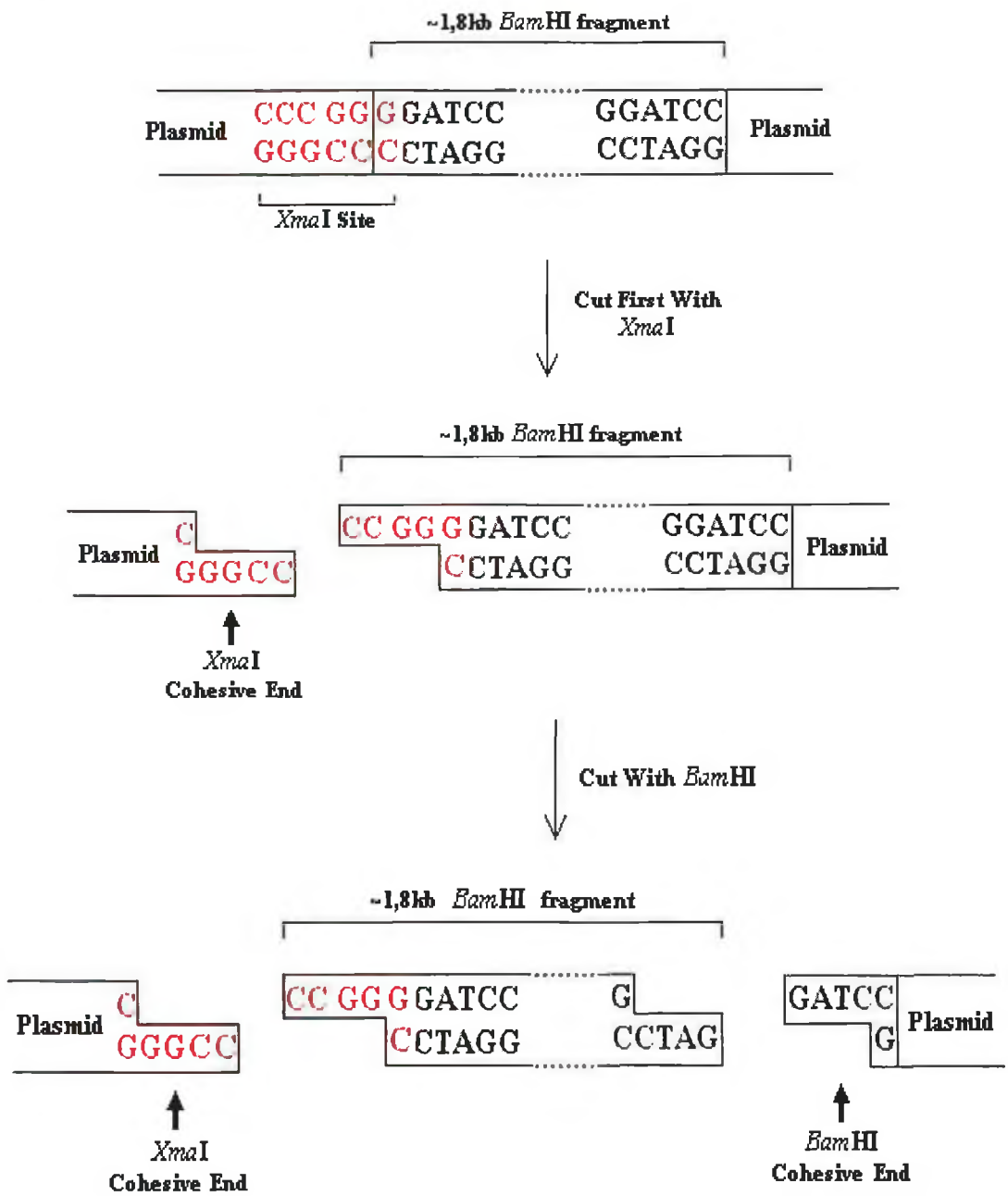


- **Figure 3.2.5: Restriction Analysis Of Plasmid pPC225.**

Lane No.	Sample.
1	1Kb Ladder
2	pPC225 uncut
3	pPC223 uncut
4	pPC225 cut with *Bam*HI
5	pSUP5011 cut with *Bam*HI
6	pPC223 cut *Bam*HI
7	1Kb Ladder

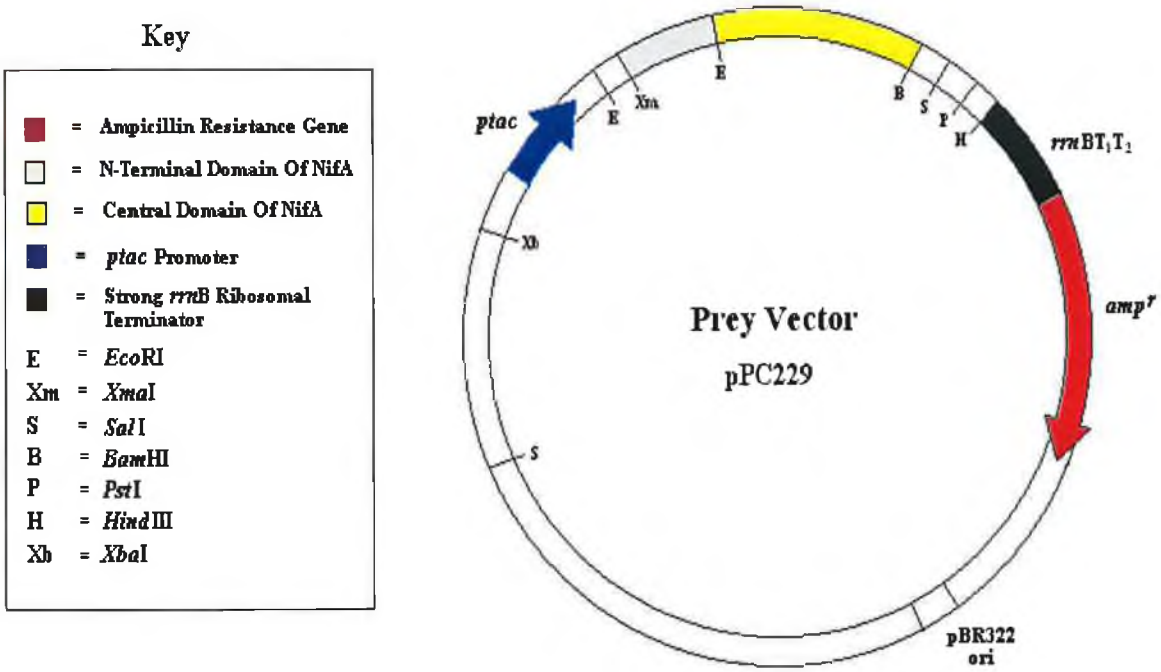
As can be seen from Figure 3.2.6 below, cutting the 1.8Kb insert back out of plasmid pPC225 by sequential digestion first with *Xma*I and then with *Bam*HI would generated pPC223 plasmid DNA with the appropriate cohesive ends to facilitate cloning of the NifA NC domain as an *Xma*I-*Bam*HI fragment. Plasmid pPC225 was thus used as the starting plasmid for the construction of the Prey plasmids

- **Figure 3.2.6: Sequential Digestion Of Plasmid pPC225 With *Xma*I And *Bam*HI.**

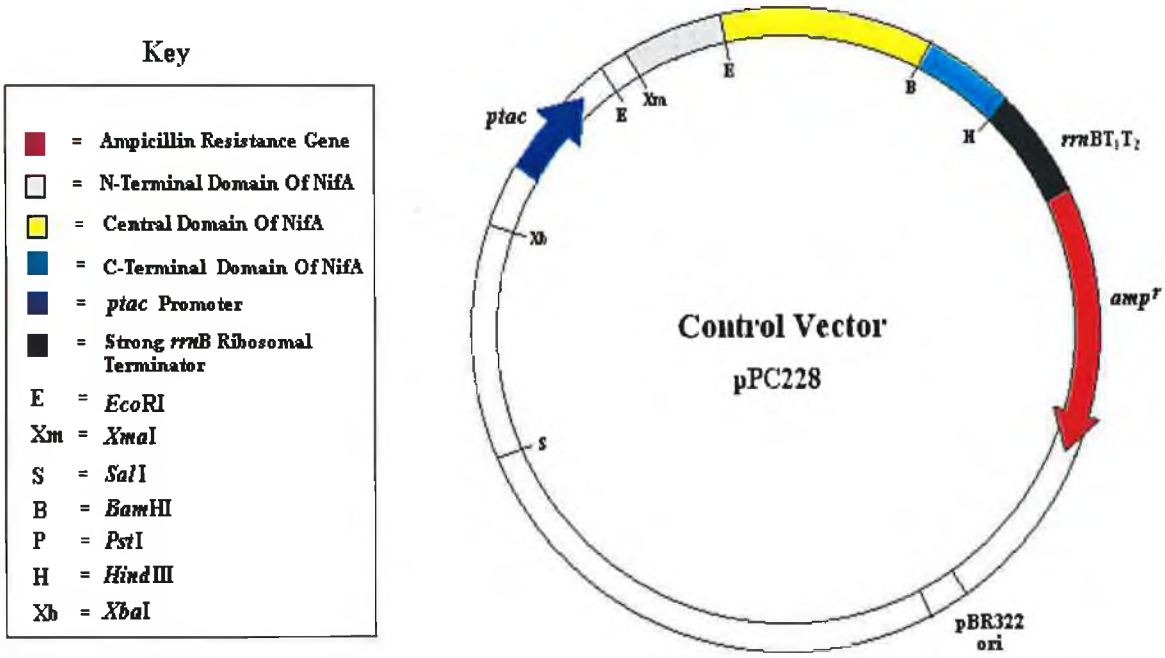


- **Figure 3.3.1: Basic Structure Of The Planned Prey And Control Plasmids.**

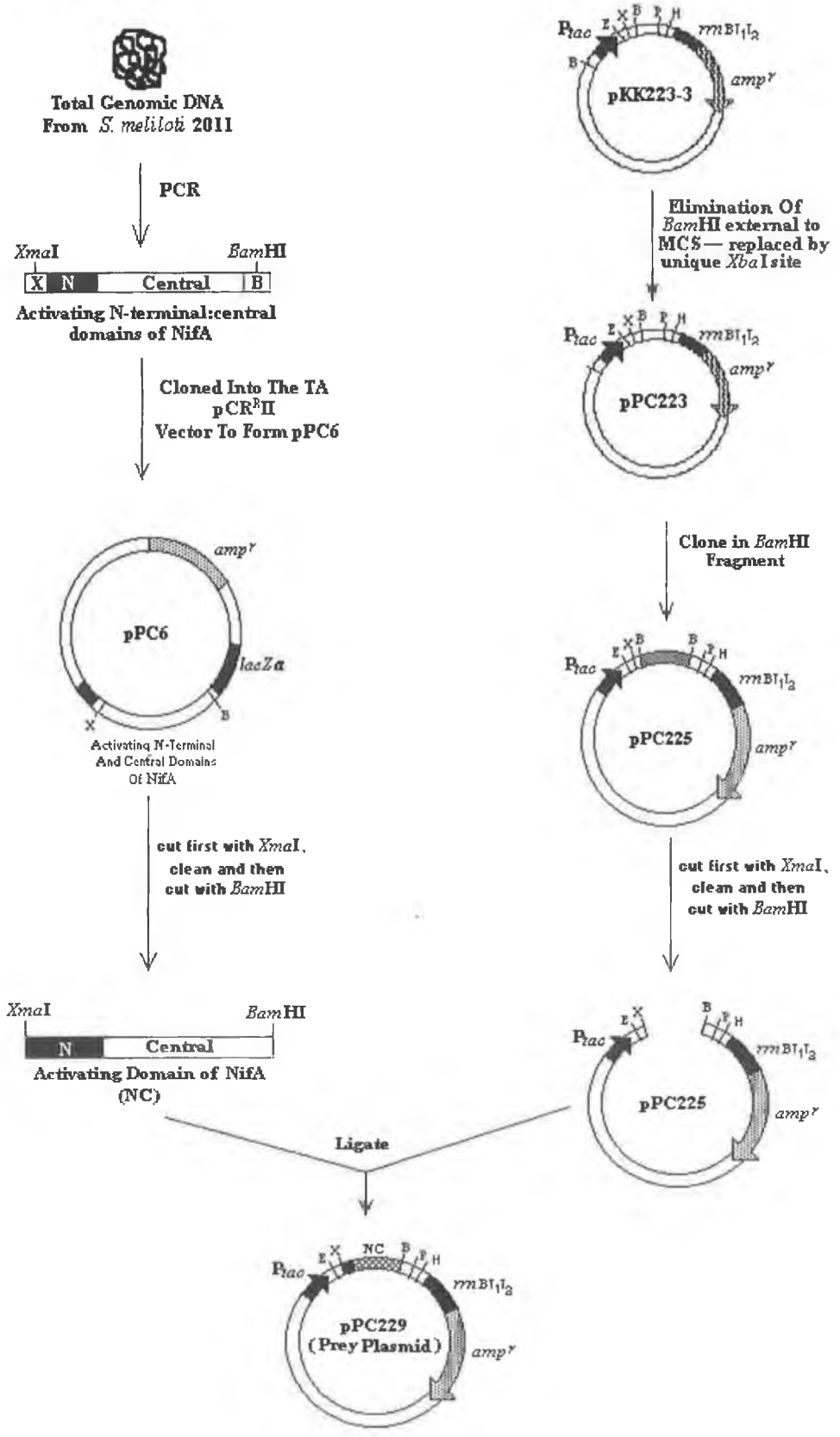
a.) Prey Plasmid Expressing The Activating NC Domain Of NifA



b.) Control Plasmid Expressing The Whole NifA Protein



- **Figure 3.3.2: Strategy For The Construction Of The Prey Plasmid pPC229-A.**



- **Figure 3.3.3: Strategy For The Construction Of The Control Plasmid pPC228.**

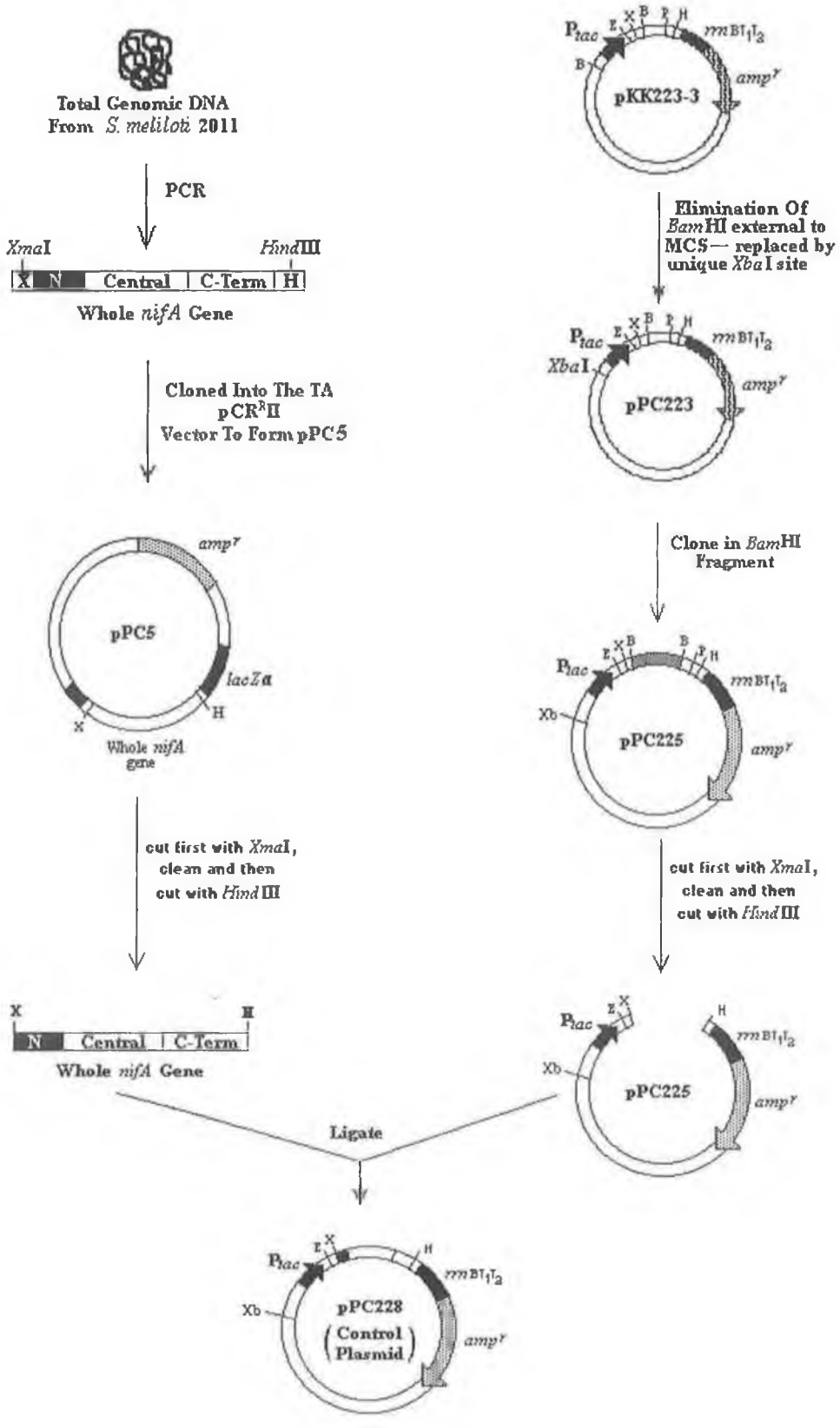


Figure 3.3.1 depicts the structure of the planned Prey and Control plasmids. Schematic representations of the cloning strategies employed for the construction of the Prey and Control plasmids can be seen in Figure 3.3.2 and Figure 3.3.3 respectively. These strategies are discussed in detail in the following sections of this chapter.

3.3: Construction Of The Prey Plasmids - Overview Of Strategy.

The transcriptional activating N-terminal-central **(NC)** domain portion of the *S. meliloti* 2011 NifA protein was used in the construction of the Prey plasmids. The NC domain was amplified by PCR as an *Xma*I-*Bam*HI fragment using the primers NC-F1 and NC-R1 shown in Figure 3.3.4 below. This fragment was first cloned into the TA pCR2.1 vector and then sub-cloned into the pPC225 plasmid as an *Xma*I-*Bam*HI fragment to generate the Prey plasmid (see Figure 3.3.1a and Figure 3.3.2)). The *Bam*HI site and the downstream *Pst*I and *Hind*III sites of the Prey plasmid could then be used for cloning genes C-terminally to the NifA NC domain sequence to generate hybrid proteins.

The forward primer NC-F1 was designed to incorporate an *Xma*I site into the NC domain PCR product and to amplify the NC domain sequence from a second internal start codon rather than from the natural start codon for the NifA protein. The reason for this was that it had been previously shown by *Huala and Ausubel (1989)* that the *S. meliloti* 2011 NifA protein was more stable in *E. coli* when expressed from the second internal start codon. The sequence of the NC-F1 forward primer is given in Figure 3.3.4 and the region of the *nifA* sequence for which this primer was designed is shown in Figure 3.3.5.

The *Eco*RI site in the MCS of the pKK223-3 plasmid could not be used for the cloning of the NifA NC domain as there was an *Eco*RI site within the *nifA* sequence (see Figure 3.3.5). The *Xma*I site therefore had to be used for cloning in order to place the start codon of the NifA NC domain PCR product close enough to the RBS on the pPC225 plasmid (essentially the same MCS as pKK223-3) to allow expression.

The reverse primer, NC-R1, was designed for the inter-domain region separating the central and C-terminal domains of the NifA protein. This region possessed a natural *Bam*HI site so the primer did not have to be designed to incorporate a *Bam*HI site. The sequence of the reverse primer NC-R1 is shown in Figure 3.3.4 and the region of the *nifA* sequence for which this primer was designed is shown in Figure 3.3.5.

Three Prey vectors were constructed which differed only in the reading frame of the *Bam*HI site relative to the upstream NifA NC domain sequence, one plasmid for each possible reading frame. The availability of these three plasmids would facilitate easy construction of libraries of Prey hybrid proteins containing the transcriptional activating NifA NC domain. The three plasmids were generated by amplifying the NC domain sequence with three different reverse primers, NC-R1, NC-R2 and NC-R3, to generate three separate PCR products. The reverse primers NC-R2 and NC-R3 were designed to incorporate additional bases to shift the reading frame of the *Bam*HI site relative to the upstream NC domain sequence (see Figure 3.3.4). Each of the resulting PCR products was cloned into the TA pCR2.1 vector and then sub-cloned into pPC225 plasmid to generate the three proposed Prey plasmids.

- **Figure 3.3.4: Primers Used For The Amplification Of The NifA NC Domain Sequence.**

NC-F1: **CCC GGG** - ATG GCC CCC ACT CGT CTT GAG

NC-F2: **CCC GGG** - ATG GCC CCC ACT CGT CTT GAG ACC ACG C

***Xma*I**

NC-R1: **GGA TCC** - GGG CTC TTT CGG CGG

NC-R2: **GGA TCC** - T - GGG CTC TTT CGG CGG

NC-R3: **GGA TCC** - TT - GGG CTC TTT CGG CGG

***Bam*HI**

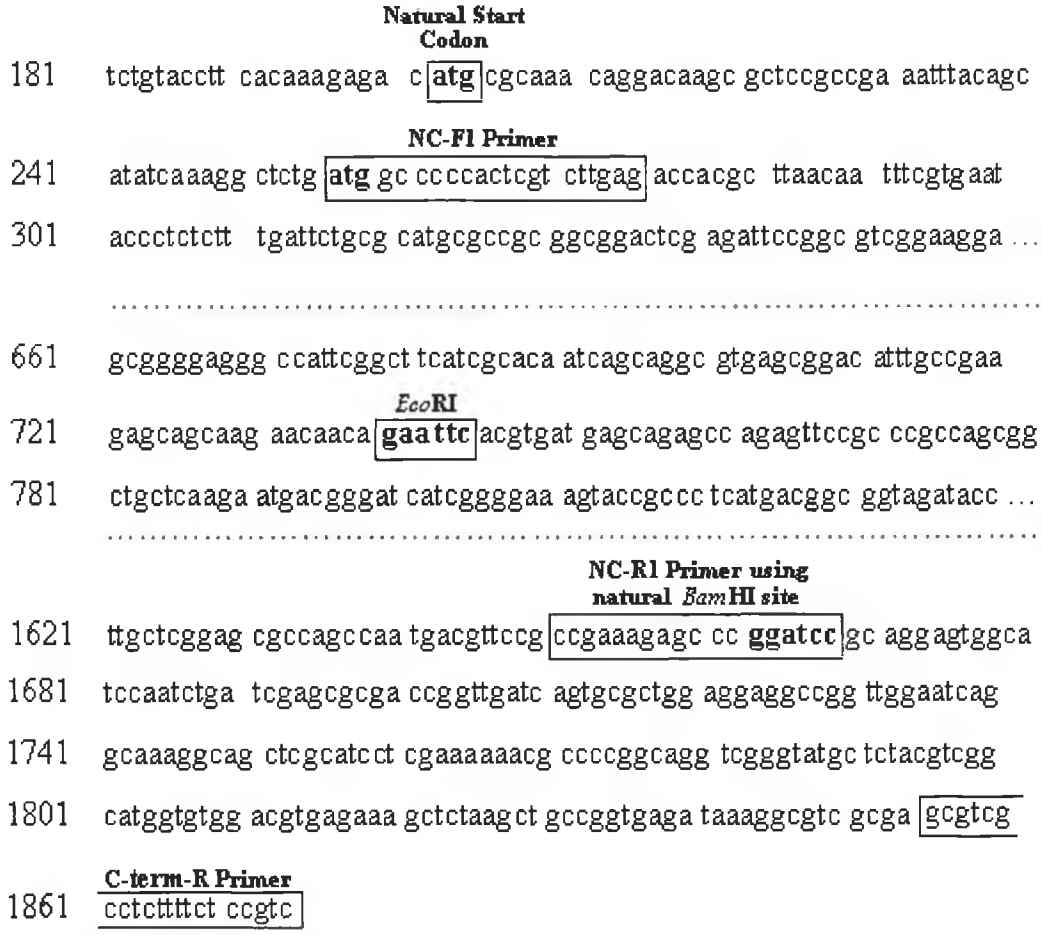
C-term-R: **AAG CTT** - GAC GGA GAA AAG AGG CGA CGC

***Hind*III**

Underlined bases in red are bases incorporated to shift the reading frame of the *Bam*HI site relative to the upstream NifA NC domain sequence.

The Control plasmid was constructed in a similar manner to the Prey plasmids and expresses the whole *S. meliloti* 2011 NifA protein. The entire *nifA* gene was amplified as an *Xma*I-*Hind*III fragment, cloned first into the TA pCR2.1 vector and then sub-cloned as an *Xma*I-*Hind*III fragment into plasmid pPC225 to form the Control plasmid (see Figure 3.3.1b and Figure 3.3.3). The *nifA* gene was amplified using the same NC-F1 forward primer that was used for the amplification of the NifA NC domain. The sequence of the reverse primer used, C-term-R, is given in Figure 3.3.4 and the region of the *nifA* sequence for which this primer was designed is shown in Figure 3.3.5. The C-term-R reverse primer was designed to incorporate a *Hind*III site into the *nif*A PCR product.

- **Figure 3.3.5: Primed Regions Of The *S. meliloti* 2011 *nifA* Sequence.**



3.3.1: Amplification Of The NifA NC Domain By PCR.

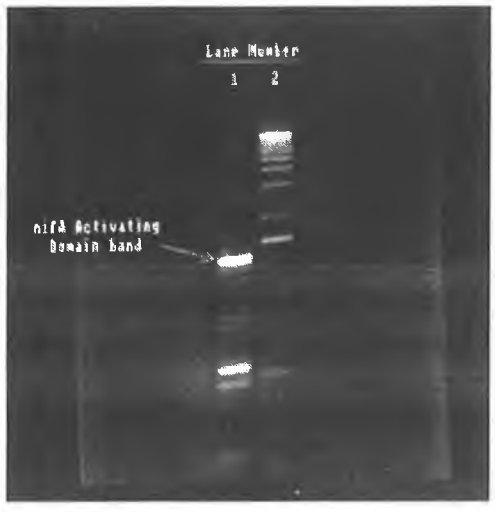
Having manipulated the initial starting plasmid pKK223-3 to give plasmid pPC225 that would permit the cloning of the NifA NC domain as an *Xma*I-*Bam*HI fragment, the next step in the construction of the proposed Prey plasmids was the amplification of the NifA NC domain sequence by PCR. The NC domain sequence was to be amplified using the forward NC-F1 and reverse NC-R1 primers described earlier. Total genomic DNA was prepared from *S. meliloti* 2011 and was used as the template DNA in the PCR (see Figure 3.3.2). Figure 3.3.6 below shows the results of the PCR and the optimized annealing and extension times that were used for the reaction.

- **Figure 3.3.6: PCR Amplification Of The NifA NC Domain With The NC-F1 Forward Primer.**

Lane No.	Sample
1	NifA NC Domain PCR
2	1Kb Ladder

PCR Conditions.

Annealing temperature	61 °C
Annealing time	1 min
Extension time	2 min



As can be seen from Figure 3.3.6 the PCR reaction yielded several smaller non-specific bands in addition to the desired 1418bp PCR product. These non-specific products could not be eliminated through manipulation of PCR conditions. As total genomic DNA was used as template in the initial PCR reactions, it was thought that the non-specific bands could be eliminated by purifying back the desired NifA NC domain PCR product from an agarose gel by the gene clean procedure and using the purified NifA NC domain PCR product as template DNA in a second PCR reaction. However, this did not eliminate the non-specific PCR products completely indicating

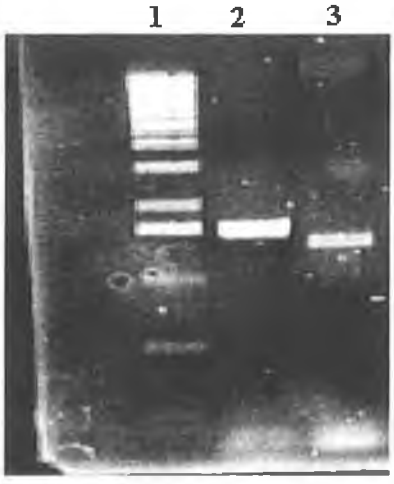
that they were caused by non-specific binding of the primers within the NifA sequence itself. The same NC-F1 forward primer was also used in conjunction with the C-term-R primer for the amplification of the whole *nifA* gene sequence and as can be seen from Figure 3.3.10 a similar pattern of non-specific PCR products was obtained. As with the NifA NC domain PCR reaction the non-specific bands could not be eliminated by using purified NifA PCR product as template DNA in a second PCR reaction. It was therefore determined that the specificity problem was due to non-specific binding of the forward NC-F1 primer within the *nifA* gene sequence itself as this was the common primer to both the NifA NC domain PCR reaction and the whole *nifA* PCR reaction. The non-specific bands were eventually eliminated by designing a new forward primer, NC-F2, that was extended a further 7 bases at the 3' end. The sequence of this primer is shown in Figure 3.3.4. The NifA NC domain PCR was repeated using the NC-F2 forward primer and using the same conditions described in Figure 3.3.6 above. The results of this PCR reaction can be seen in Figure 3.3.7. As can be seen from Figure 3.3.7, performing the PCR with the longer NC-F2 primer eliminated the non-specific PCR products. Using the longer NC-F2 primer for the amplification of the whole *nifA* had the same effect of eliminating previously obtained non-specific PCR products.

- **Figure 3.3.7: PCR Amplification Of The NifA NC Domain With The NC-F2 Forward Primer.**

Lane No.	Sample
1	1Kb Ladder
2	Whole *nifA* PCR
3	NifA NC Domain PCR

PCR Conditions.

Annealing temperature	61 °C
Annealing time	1 min
Extension time	2 min



Having eliminated the non-specific PCR bands from the NifA NC domain PCR reaction, two more PCR reactions were performed using the NC-F2 primer with each of the NC-R2 and NC-R3 reverse primers. In this way three NifA NC domain PCR products were obtained, one amplified using the NC-R1 reverse primer (this PCR product was called NC-A), one amplified with the NC-R2 primer (this PCR product was called NC-B) and one amplified using the NC-R3 primer (this PCR product was called NC-C).

3.3.2: Cloning Of The NifA NC Domain PCR Products Into The TA Vector To Form Plasmids pPC6-A, pPC6-B and pPC6-C.

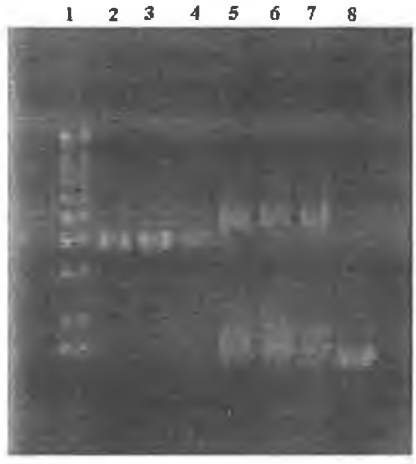
Having optimized the NifA NC domain PCRs and obtained the three desired PCR products NC-A, NC-B and NC-C the next step was to clone each of the PCR products into the TA pCR2.1 vector (see Figure 3.3.2). Ligations were set up according to the TA protocol (see section 2.18 of Chapter 2) using the PCR reaction mixtures directly without any purification. The ligations were used to transform INVαF' cells that came with the TA cloning kit and the transformations were plated out on LBamp/Xgal plates. White colonies were picked off plates and screened to identify clones containing the TA pCR2.1 vector with inserted NifA NC domain PCR products. However, the desired clones could not be identified despite extensive screening. All of the clones screened, although white in color, appeared to contain recircularised pCR2.1 vector DNA.

The TA vector only recircularises to give white colonies at a very low frequency yet a very large number of white colonies were obtained from transformations. It was therefore deduced that the NifA NC domain PCR reaction, although it appeared to be specific and to only contain the desired 1418bp NifA NC domain PCR product, must have contained a very small PCR product that was causing recircularisation of the TA pCR2.1 vector to yield white colonies. It was decided that the best way to overcome this problem was to purify the desired NC domain PCR products from an agarose gel and to then ligate the purified products with the TA pCR2.1 vector. Attempts to ligate PCR products purified from 0.7% agarose gels by the gene clean procedure failed to yield any transformants. It was felt that the silica beads used to

absorb DNA in gene clean procedure may have been too abrasive and that the 3'A overhangs on the ends of the PCR products required for TA cloning may have been sheared off during the purification procedure. The procedure was therefore modified in that rather than using the silica beads to absorb the DNA, a spin column with a DNA absorbing filter was used. It was felt that this would be a more gentle procedure and this adjustment to the protocol overcame the problems. Large numbers of white transformants were obtained when the resulting ligations were used to transform INVαF' cells. Screening of the colonies obtained by performing *Xma*I-*Bam*HI double digests on plasmid DNA prepared from each clone resulted in the identification of the large numbers of clones containing the desired plasmids. The NC-A PCR product was cloned into the pCR2.1 vector to give plasmid pPC6-A, and each of the NC-B and NC-C PCR products were also cloned into the TA vector to yield plasmids pPC6-B and pPC6-C respectively. Figure 3.3.8 below shows the results of the verifying restriction analysis of these plasmids.

- **Figure 3.3.8: Restriction Analysis Of Plasmids pPC6-A, pPC6-B And pPC6-C.**

Lane No.	Sample
1	1Kb Ladder
2	pPC6-A uncut
3	pPC6-B uncut
4	pPC6-C uncut
5	pPC6-A cut *Xma*I-*Bam*HI
6	pPC6-B cut *Xma*I-*Bam*HI
7	pPC6-C cut *Xma*I-*Bam*HI
8	NifA NC domain PCR
9	1Kb Ladder



3.3.3: Sub-cloning Of The NifA NC Domains Into pPC225 To Form The Prey Plasmids pPC229-A, pPC229-B And pPC229-C.

Having successfully cloned the 3 PCR amplified NifA NC domains into the TA pCR2.1 vector to form plasmids pPC6-A, pPC6-B and pPC6-C, the next step in the

construction of the Prey plasmids was to sub-clone each of the NC domain fragments into plasmid pPC225 (see Figure 3.3.2). Plasmid pPC6-A was digested first with *Xma*I and then with *Bam*HI and the digest was separated on a 0.7% agarose gel. The 1.4Kb band corresponding to the excised NC-A fragment was then purified from the gel using the gene clean procedure. The pPC225 plasmid was digested first with *Xma*I and then with *Bam*HI as described earlier in section 3.2.1 and illustrated in Figure 3.2.6. The digested pPC225 DNA was then separated on a 0.7% agarose gel and the 4.5Kb band corresponding to the pPC223 portion of pPC225 was cleaned back from the gel by the gene clean procedure. This cleaned plasmid DNA was ligated with the purified NC-A fragment and the ligated DNA was used to transform DH5α. Clones were screened by performing *Xma*I-*Bam*HI double digests on the plasmid DNA prepared from each clone. In this manner the desired Prey plasmid, pPC229-A was obtained. The NC-B and NC-C products were sub-cloned from plasmids pPC6-B and pPC6-C into pPC225 in the same manner to give Prey plasmids pPC229-B and pPC229-C respectively. Figure 3.13 below shows the verifying restriction analysis of plasmids pPC229-A, pPC229-B and pPC229-C and detailed restriction maps of the plasmids can be seen in Figure 3.5.1.

- **Figure 3.3.9: Restriction Analysis Of The Prey Plasmids pPC229-A, pPC229-B And pPC229-C.**

Lane No.	Sample
1	1Kb Ladder
2	pPC223 uncut
3	pPC229-A uncut
4	pPC229-B uncut
5	pPC229-C uncut
6	pPC229-A cut *Xma*I-*Bam*HI
7	pPC229-B cut *Xma*I-*Bam*HI
8	pPC229-C cut *Xma*I-*Bam*HI
9	NifA NC domain PCR
10	1Kb Ladder

3.3.4: Construction Of The Control Plasmid pPC228.

In addition to the Prey plasmids discussed up to this point, a Control plasmid also had to be constructed. The proposed Control plasmid (see Figure 3.3.1b) would express the whole *S. meliloti* 2011 NifA protein. It could be used to test the planned P2H system reporter constructs and would also be used as a positive control in any two-hybrid tests.

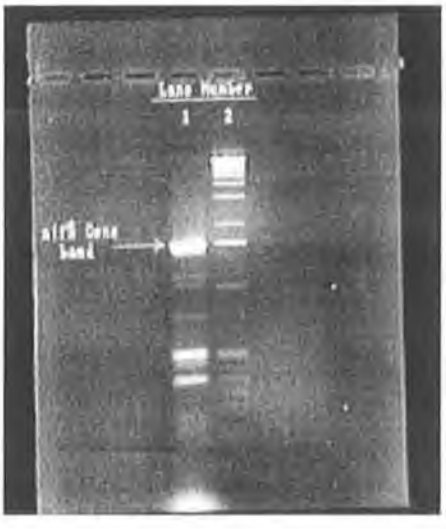
The Control plasmid was constructed in a very similar manner to the Prey plasmids (see figure 3.3.3). The entire *nifA* gene was amplified by PCR as an *Xma*I-*Hind*III fragment as described earlier in section 3.3. Initially the PCR was carried out using the NC-F1 forward primer and the C-Term-R reverse primer and the same PCR conditions as those used for the amplification of the NifA NC domain. The results of this PCR can be seen in Figure 3.3.10 below.

- **Figure 3.3.10: PCR Amplification Of The Whole *nifA* Gene Using The NC-F1 Primer.**

Lane No.	Sample
1	Whole *nifA* PCR
2	1 Kb Ladder

PCR Conditions.

Annealing temperature	61 °C
Annealing time	1 min
Extension time	2 min

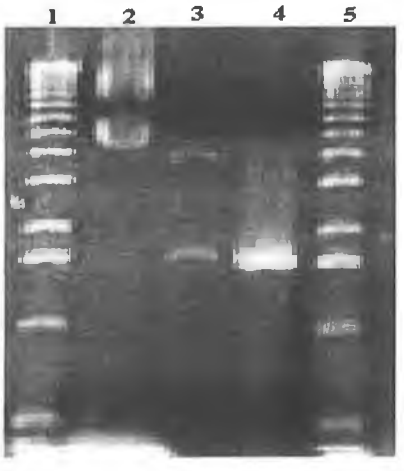


As with the PCR for the NifA NC domain PCR reaction, several non-specific PCR products were obtained in addition to the desired 1631bp whole *nifA* PCR product. As with the PCR reaction for the NifA NC domain PCR reaction these non-specific PCR products were eliminated by the use of the NC-F2 forward primer. The results of the PCR for the entire *nifA* gene using this forward primer can be seen in Figure 3.3.7.

Having optimized the *nifA* PCR, the next step was to clone the *nifA* PCR product into the TA pCR2.1 vector. The same problems that were encountered when trying to clone the NifA NC domain PCR products into the TA vector were encountered when trying to clone the *nifA* PCR product. However, these problems were also overcome in exactly the same way. The *nifA* PCR product was separated on a 0.7% agarose gel, the desired 1.6Kb *nifA* band was cleaned back using the modified gene clean procedure described earlier in section 3.3.2. The purified *nifA* PCR product was ligated with the TA pCR2.1 vector and transformed into INVαF' cells. The transformation was plated on LBamp/Xgal plates and white colonies were picked off and screened. Clones were screened by performing *Xma*I-*Hind*III double digests on plasmid DNA prepared from each clone. In this way the desired clone, pPC5, was obtained. Figure 3.3.11 below shows the verifying restriction analysis of plasmid pPC5.

- **Figure 3.3.11: Restriction Analysis Of Plasmid pPC5.**

Lane No.	Sample
1	1Kb Ladder
2	pPC5 uncut
3	pPC5 cut *Xma*I
4	pPC5 cut *Xma*I-*Hind*III
5	PCR for whole *nifA*
6	1Kb ladder

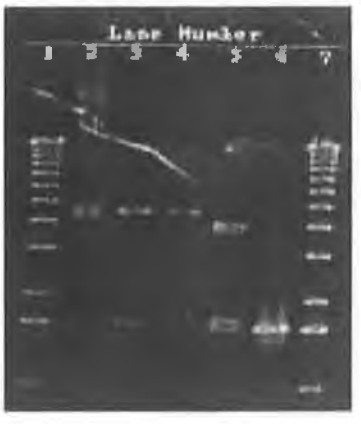


The last step in the construction of the Control plasmid pPC228 was the sub-cloning of the *nifA* gene from pPC5 into pPC225. Plasmid pPC5 was digested first with *Xma*I and then with *Hind*III. The digested DNA was separated on a 0.7% agarose gel and the desired 1.6KB *nifA* fragment was purified using the gene clean procedure. Plasmid pPC225 was digested with *Xma*I and then with *Hind*III and the digested DNA was separated on a 0.7% agarose gel. The 4.5kb band corresponding to the pPC223 portion of plasmid pPC225 was purified by the gene clean procedure and

ligated with the purified *nifA* fragment. The ligated DNA was used to transform DH5α and the transformation was plated on LB ampicillin plates. Clones were screened by performing *Xma*I-*Hind*III double digests on plasmid DNA isolated from each clone. In this manner the desired Control plasmid, pPC228, was identified. Figure 3.3.12 below shows the results of the verifying restriction analysis of plasmid pPC228.

- **Figure 3.3.12: Restriction Analysis Of The Control Plasmid pPC228.**

Lane No.	Sample.
1	1Kb Ladder
2	pPC228 Uncut
3	pPC228 cut *Xma*I-*Hind*III
4	pPC225 cut *Xma*I-*Hind*III
5	pPC5 cut *Xma*I-*Hind*III
6	*nifA* PCR Product
7	1Kb Ladder



3.4: Testing Expression Of NifA Protein From The Control Plasmid pPC228.

Having constructed the Control plasmid pPC228, expression of the NifA protein from the plasmid had to be verified. This was done by testing for NifA induced transcriptional activation of a *nifH:lacZα* reporter gene. The construction of this reporter gene is discussed in chapter 4. The reporter gene is maintained on a pPCYC184 based plasmid called pPC185-A. This plasmid is compatible with the pKK223-3 plasmid upon which the pPC228 Control plasmid is based. The pPC185-A plasmid also had a chloramphenicol resistance gene that allows it to be selected for independently of the pPC228 plasmid which has an ampicillin resistance gene.

Plasmids pPC228 and pPC185-A were co-transformed into DH5α. The transformation was plated on LB media to which ampicillin and chloramphenicol had been added to select for the presence of the two plasmids. IPTG was also added to the media to a concentration of 0.5mM to induce expression of the NifA protein from the *ptac* promoter on pPC228. Xgal was also incorporated into the media to detect NifA induced expression of the *nifH:lacZ*α reporter. Plates were incubated at 30 °C. After 48 hours at 30 °C colonies began to develop a blue color but plates had to be incubated for a total of 72 hours at 30 °C for significant color development.

It was felt that the slow color development was due to poor expression of the *nifA* gene from the *ptac* promoter on plasmid pPC228. It was thought that this may have been due to the start codon on the *Xma*I-*Hind*III cloned *nifA* fragment being too far away from the plasmid borne RBS. The *ptac* promoter region of plasmid was sequenced to verify the distance of the start codon from the RBS and sequence of the N-terminal region of the cloned *nifA* gene. The sequence obtained can be seen in Figure 3.4.1 below.

- **Figure 3.4.1: Sequence Of The *ptac* Promoter And The N-Terminal Of NifA From pPC228 .**

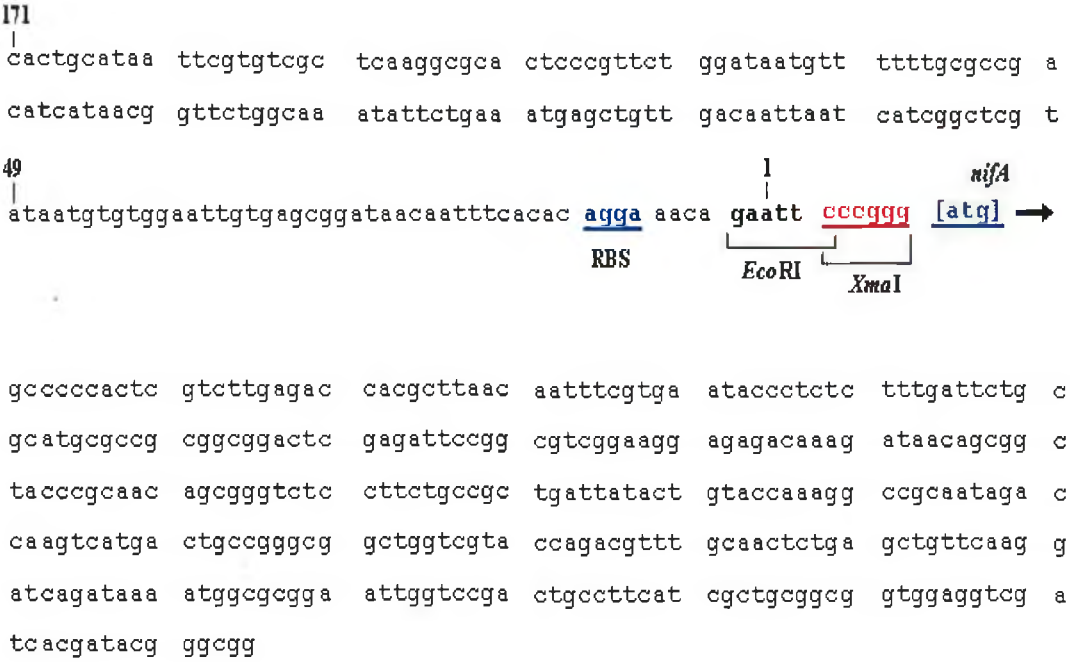
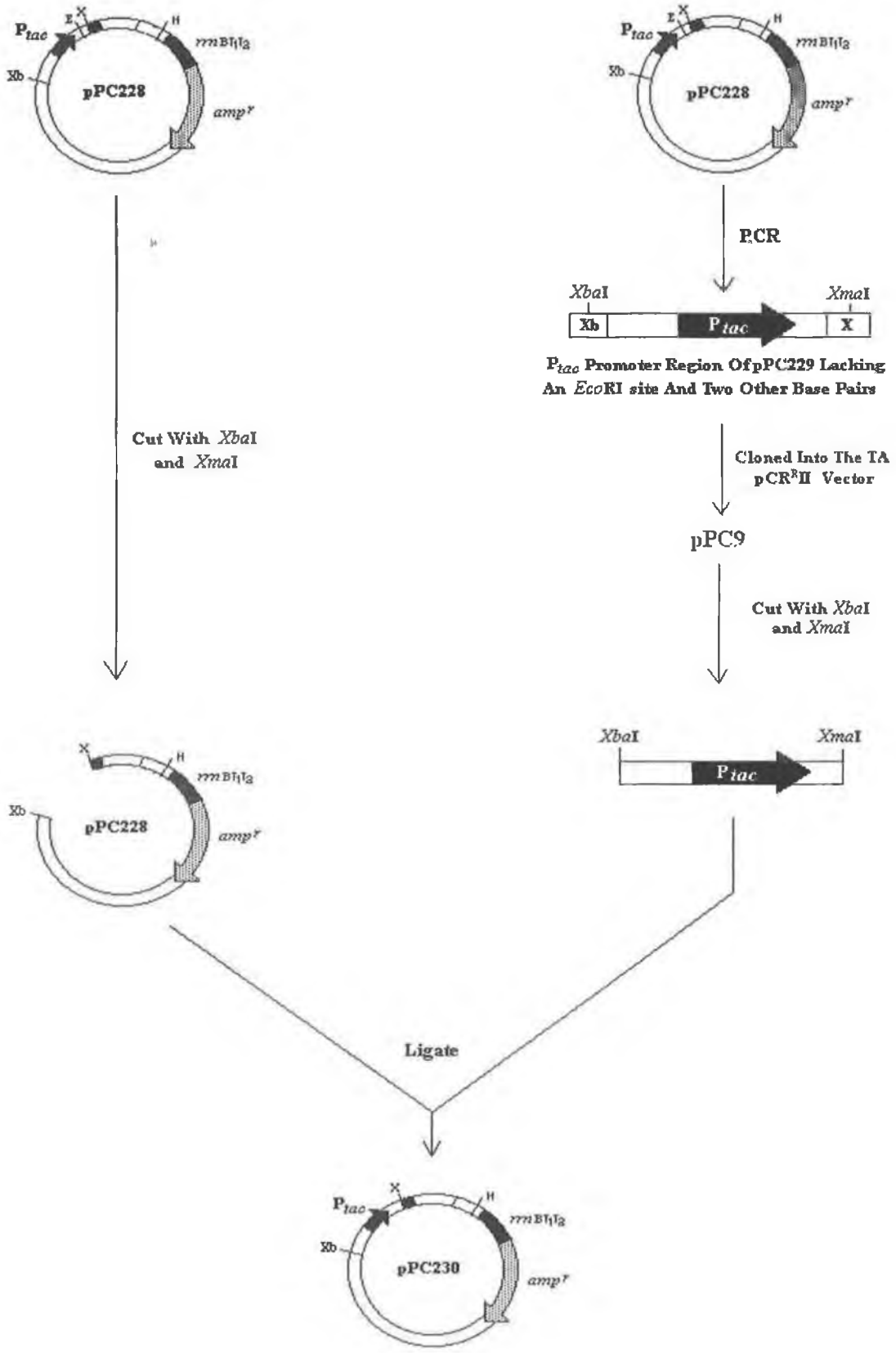


Figure 3.4.2: Strategy For The Manipulation Of The *ptac* Promoter Of The Control Plasmid pPC228 To Form pPC230.



The sequence of the *ptac* promoter and the *nifA* 5' end, encoding the N-terminal domain of NifA, obtained for plasmid pPC228 was as expected. The RBS, *Eco*RI site, *Xma*I site and the start codon are all highlighted in Figure 3.4.2 and it can be seen that there are 15 bases between the start codon and the RBS. For most known promoters the start codon is within 13 base pairs of the RBS (*Gold, 1988*) with spacings of about 9 bases being optimal. It was therefore felt that the *nifA* start codon on plasmid pPC228 was fractionally too far away from the RBS to allow efficient translation of the NifA protein. It was thus decided that the *ptac* promoter would be manipulated to reduce the distance between the *nifA* start codon and the RBS.

A schematic representation of the strategy employed for the manipulation of the *ptac* promoter region of plasmid pPC228 can be seen in Figure 3.4.2. The strategy made use of the unique *Xba*I site incorporated into pPC223 and thus the pPC225 starting plasmid when the second *Bam*HI site outside the MCS on plasmid pKK223-3 was destroyed. The *ptac* promoter was amplified by PCR as an *Xba*I-*Xma*I fragment using the primers shown in Figure 3.4.3.

- **Figure 3.4.3: Primers Used For The Amplification Of The *ptac* Promoter.**

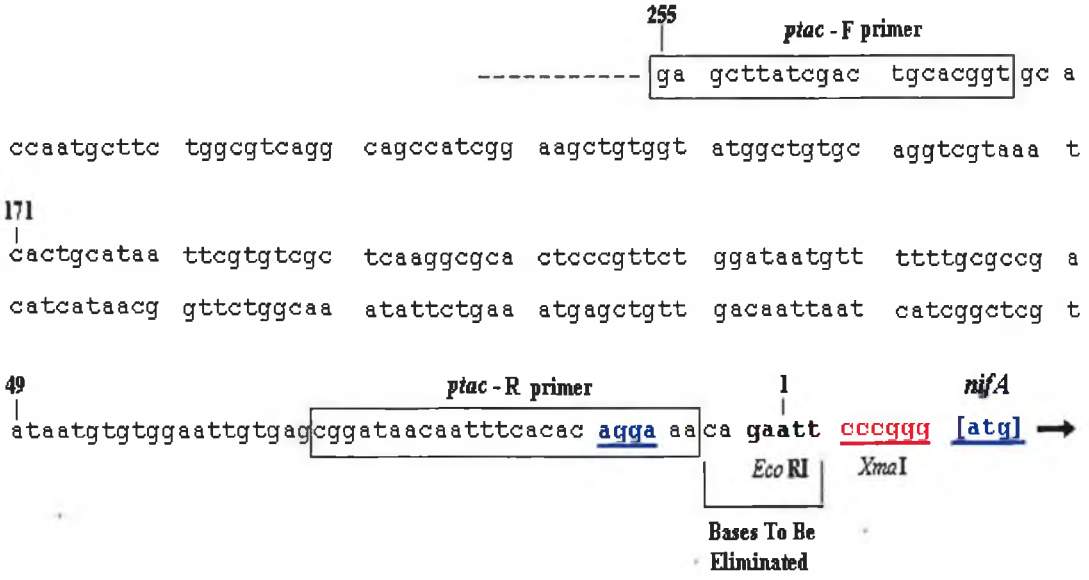
ptac - F: **(TCT AGA)** GAG CTT ATC GAC TGC ACG GT
XbaI site

ptac - R: **(CCC GGG)** TTT CCT GTG TGA AAT TGT TAT CCG
XmaI site

The *ptac*-R reverse primer was designed to eliminate 7 of the bases between the *nifA* start codon and the RBS. The bases that were eliminated are shown in Figure 3.4.4 below as are the regions of the pPC228 sequence for which the *ptac*-F and *ptac*-R primers were designed. Elimination of the 7 bases indicated in Figure 3.3.4 resulted in the elimination of the *Eco*RI site in the *ptac* promoter region on the resulting *ptac* PCR product. This PCR product was first cloned into the TA pCR2.1 vector and then sub-cloned into the pPC228 plasmid as an *Xba*I-*Xma*I fragment thus replacing the original *ptac* promoter region on the pPC228 Control plasmid with the manipulated

PCR amplified *ptac* promoter with it's deleted bases. The resulting alternative Control plasmid thus formed was called pPC230. Similarly the *ptac* PCR product was cloned into each of the pPC229-A, pPC229-B and pPC229-C Prey plasmids to generate three new alternative Prey plasmids called pPC231-A, pPC231-B and pPC231-C respectively.

- **Figure 3.4.4: Primed Regions Of The pPC228 Sequence Used For The Amplification Of The *ptac* Promoter.**



3.4.1: PCR Amplification And Cloning Of The *ptac* Promoter.

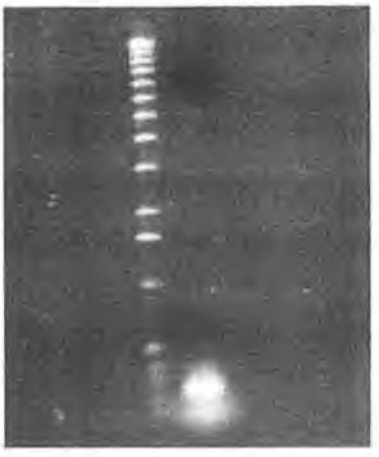
The *ptac* promoter was amplified from pPC228 using the primers *ptac-F* and *ptac-R* described above. The results of this PCR and the PCR conditions used are shown in Figure 3.4.5 below. As can be seen from Figure 3.4.5, a single band of approximately 262bp corresponding to the desired *ptac* promoter PCR product was obtained.

- **Figure 3.4.5 : PCR Amplification Of The *ptac* Promoter From pPC228.**

Lane No.	**Sample.**
1	1Kb Ladder
2	*ptac* Promoter PCR

PCR Conditions.

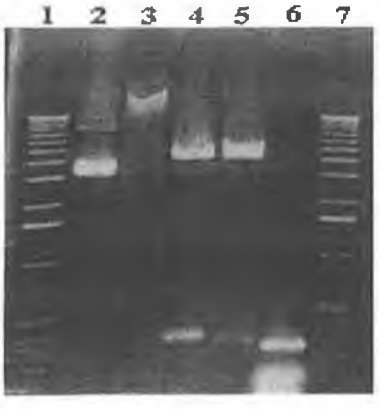
Annealing Temperature	58 °C
Annealing Time	30 sec
Extension Time	1 min



Having optimized the *ptac* PCR and obtained the desired PCR product the next step was to clone the PCR product into the TA pCR2.1 vector. A ligation was set up according to the TA protocol given in section 2.18 and the ligated DNA was used to transform INVαF' cells. The transformation was plated on LBamp/Xgal plates and incubated at 37 °C overnight. White colonies were picked off and screened by performing *Eco*RI digests on plasmid DNA isolated from each clone. *EcoRI* sites flank the insertion site for the PCR product in the TA vector and thus *Eco*RI digestion was expected to excise cloned PCR products as there were no *Eco*RI sites within the desired *ptac* PCR product. Screening in this manner resulted in the identification of the desired plasmid called pPC9. The results of the verifying restriction analysis of this plasmid can be seen in Figure 3.4.6 below.

- **Figure 3.4.6: Restriction Analysis Of Plasmid pPC9**

Lane No.	**Sample.**
1	1Kb Ladder
2	pPC9 Uncut
3	pPC9 cut *Xma*I
4	pPC9 cut *Eco*RI
5	pPC9 cut *Xma*I-*Xba*I
6	*ptac* promoter PCR product
7	1Kb ladder



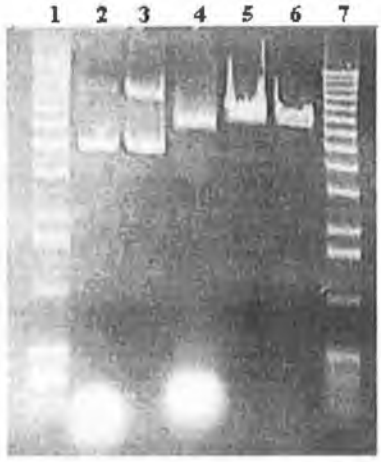
3.4.2: Sub-cloning Of The *ptac* Promoter Into pPC228 To Form The Alternative Control Plasmid pPC230.

Plasmid pPC9 was digested with *Xma*I and *Xba*I and the digested plasmid DNA was separated on a 1% agarose gel. The desired 262bp band corresponding to the *ptac* promoter PCR product was purified from the gel using the gene clean procedure. Plasmid pPC228 was also digested with *Xma*I and *Xba*I to excise its existing 272bp *ptac* promoter region. The digested pPC228 DNA was separated on a 1% agarose gel to verify complete digestion. The 5931bp band corresponding to the remaining portion of the pPC228 plasmid was purified from the agarose gel and ligated with the purified *ptac* PCR product. The ligated DNA was used to transform DH5α cells and the transformation was plated on LBamp agar. Colonies obtained were screened by performing *Eco*RI digests on plasmid DNA isolated from each clone. *Eco*RI digestion of plasmid pPC228 yields two bands on an agarose gel, one of 493bp and one of approximately 5710bp. This is because it possesses two *Eco*RI sites, one within the *nifA* sequence itself and one in the upstream *ptac* promoter region *Eco*RI digestion thus cleaves off the 493bp NifA N-terminal domain sequence. *Eco*RI digestion of the desired clone with the manipulated PCR amplified *ptac* promoter region was expected to generate a single band corresponding to linearised plasmid DNA as the *Eco*RI site in the original *ptac* promoter region would not be present in the PCR amplified *ptac* promoter region as described earlier. Screening clones using *Eco*RI digestions resulted in the identification of the desired plasmid called pPC230. Figure 3.4.7 below shows the results of the verifying restriction analysis of plasmid pPC230 and Figure 3.5.1(A) shows a detailed restriction map of the plasmid.

When the newly constructed control plasmid pPC230 was introduced into DH5α with plasmid pPC185-A (bearing the *nifH:lacZα* reporter gene) and plated on selective media containing IPTG and Xgal much stronger colour development of colonies was observed than that observed for the initial control plasmid pPC228. This indicated that the manipulation of the *ptac* promoter region of plasmid pPC228 to reduce the distance between the plasmid borne RBS and the ATG start codon of the cloned *nifA* gene (thus generating the control plasmid pPC230) had the desired effect of improving the efficiency with which the NifA protein was translated.

- **Figure 3.4.7: Restriction Analysis Of Plasmid pPC230.**

Lane No.	Sample.
1	1Kb Ladder
2	pPC228 Uncut
3	pPC230 uncut
4	pPC228 cut *Eco*RI
5	pPC230 cut *Eco*RI
6	pPC230 cut *Xba*I
7	1Kb Ladder.



3.4.3: Construction Of The Alternative Prey Plasmids pPC231-A, pPC231-B And pPC231-C.

In each of the previously constructed Prey vectors pPC229-A, pPC229-B and pPC229-C the start codon of the cloned NifA NC domain was the same distance from the RBS as it was for the *nifA* gene in plasmid pPC228. Therefore, the *ptac* promoter region of each of these plasmids also had to be replaced with the modified *ptac* promoter PCR product and this was done in the same way as for plasmid pPC228.

Plasmid pPC229-A was digested with *Xma*I and *Xba*I to excise its existing 272bp *ptac* promoter region and the digested plasmid DNA was separated on a 1% gel to verify complete digestion. The 5738bp band corresponding to the remaining portion of the pPC229-A plasmid was purified back from the agarose gel using the gene clean procedure and ligated with the purified *Xma*I-*Xba*I *ptac* promoter fragment cut from pPC9. The ligated DNA was used to transform DH5α cells and clones obtained were screened by *Eco*RI digestion as with the manipulation of the *ptac* region of pPC228. In this manner the desired plasmid pPC231-A was obtained. The modified *ptac* promoter PCR product was also sub-cloned from pPC9 into plasmids pPC229-B and pPC229-C in the same way to yield plasmids pPC231-B and pPC231-C respectively. The results of the verification restriction analysis of these plasmids can

be seen in Figures 3.4.8 and Figure 3.5.1(B-D) shows detailed restriction maps of the alternative Prey plasmids.

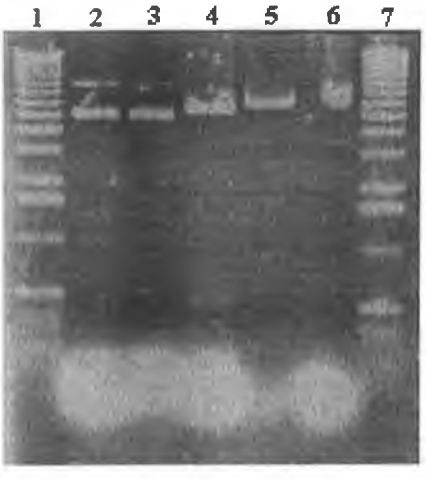
- **Figure 3.4.8: Restriction Analysis Of Plasmids pPC231-A, pPC231-B And pPC231-C.**

A.) Restriction Of pPC231-A.

Lane No.	Sample.
1	1Kb Ladder
2	pPC229-A uncut
3	pPC231-A uncut
4	pPC229-A cut *Eco*RI
5	pPC231-A cut *Eco*RI
6	pPC231-A cut *Xba*I
7	1Kb Ladder.

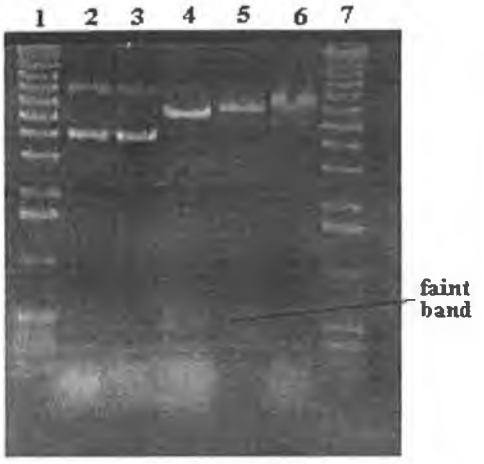
B.) Restriction Of pPC231-B.

Lane No.	Sample.
1	1Kb Ladder
2	pPC229-B uncut
3	pPC231-B uncut
4	pPC229-B cut *Eco*RI
5	pPC231-B cut *Eco*RI
6	pPC231-B cut *Xba*I
7	1Kb Ladder.



C.) Restriction Of pPC231-C.

Lane No.	Sample.
1	1Kb Ladder
2	pPC229-C uncut
3	pPC231-C uncut
4	pPC229-C cut *Eco*RI
5	pPC231-C cut *Eco*RI
6	pPC231-C cut *Xba*I
7	1Kb Ladder.

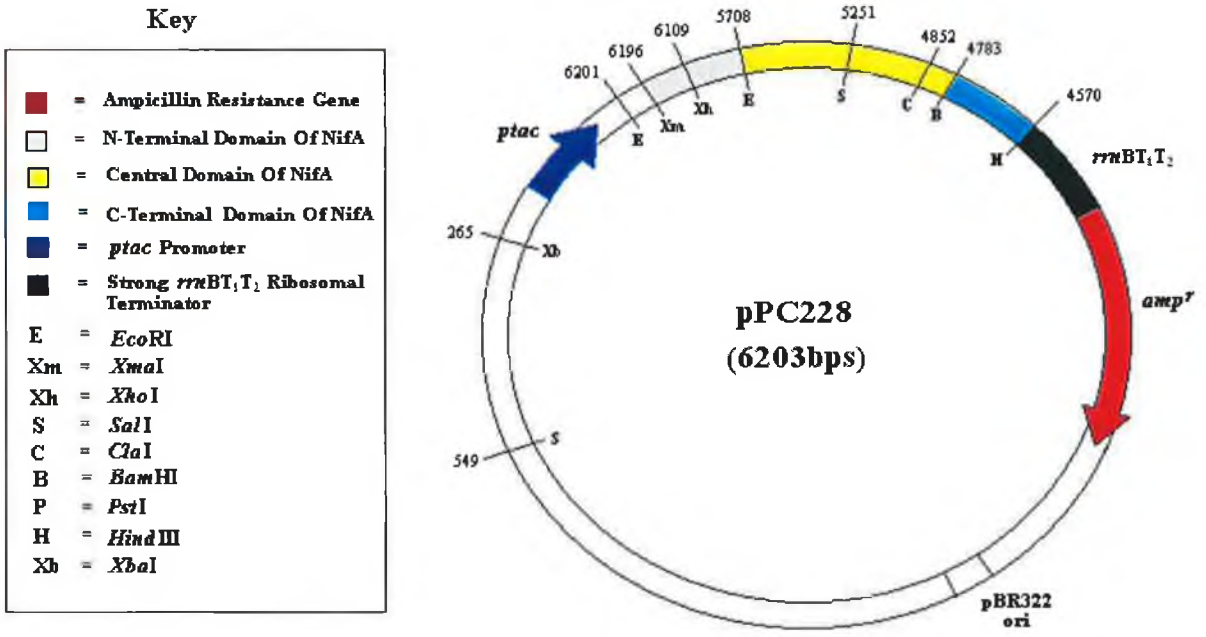


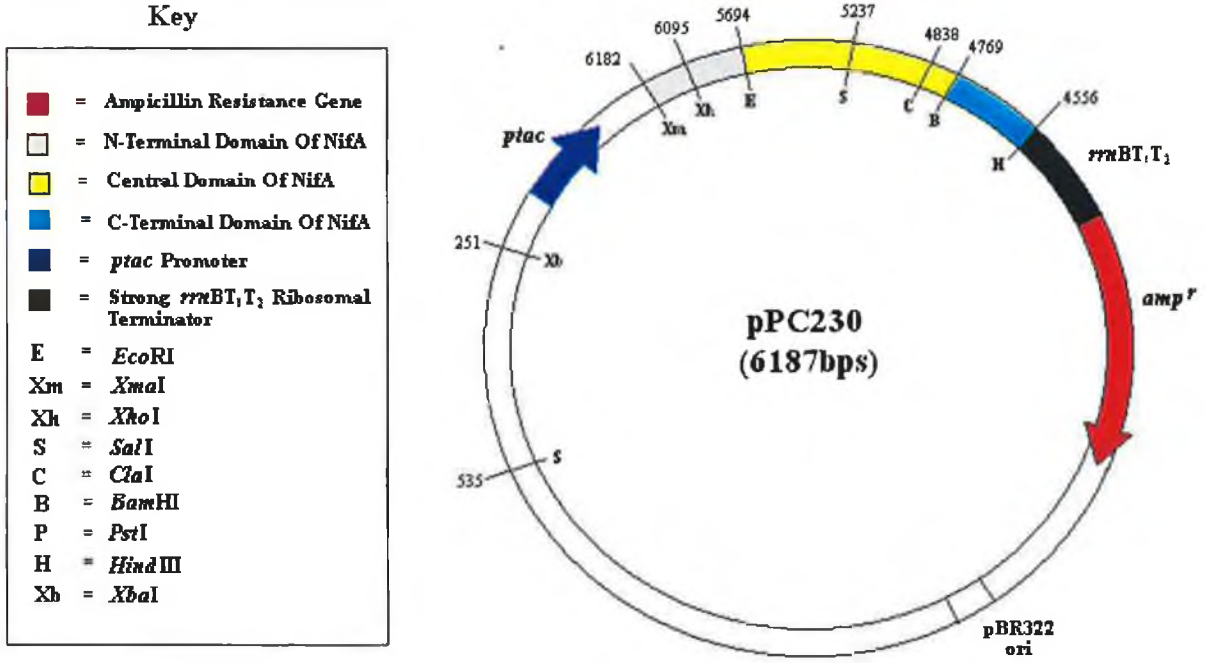
3.5: Summary Of Chapter 3.

The end result of all of the above strategies was the construction of the initial Control plasmid pPC228 and the alternative Control plasmid pPC230 with its manipulated *ptac* promoter region. Also three initial Prey plasmids called pPC229-A, pPC229-B and pPC229C were constructed. The *ptac* promoter region of each of these was manipulated to form the alternative Prey plasmids pPC231-A, pPC231-B and pPC231-C respectively. Detailed restriction maps of all of these plasmids can be seen in Figure 3.5.1 below.

- **Figure 3.5.1: Restriction Maps Of The Constructed Control And Prev Plasmids For The NofA Based P2H System.**

A.) Control Plasmids pPC228 And pPC230



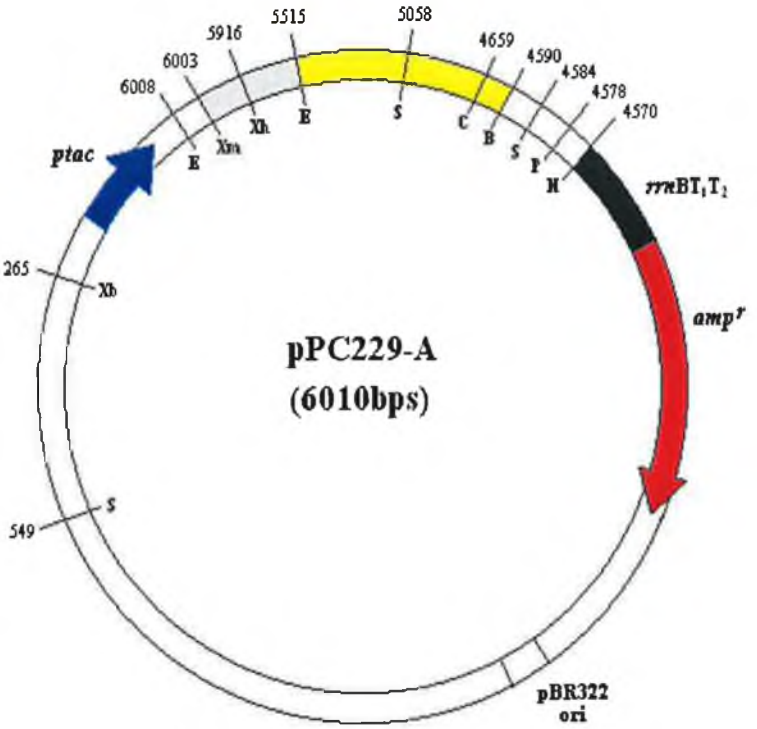


B.) Prev Plasmids pPC229-A And pPC231-A.

Key

= Ampicillin Resistance Gene
= N-Terminal Domain Of NifA
= Central Domain Of NifA
= *ptac* Promoter
= Strong *rrn*BT$_1$T$_2$ Ribosomal Terminator

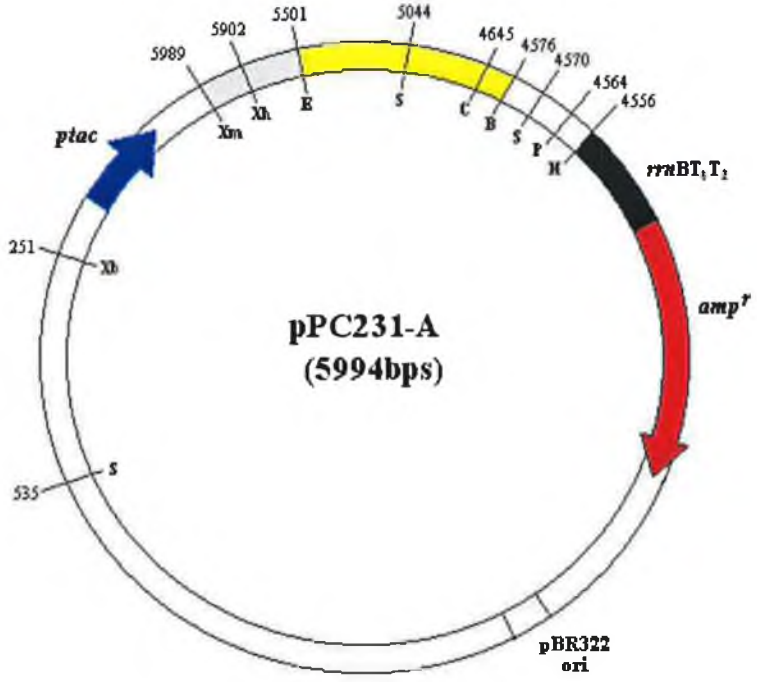
E = *Eco*RI
Xm = *Xma*I
Xh = *Xho*I
S = *Sal*I
C = *Cla*I
B = *Bam*HI
P = *Pst*I
H = *Hind*III
Xb = *Xba*I



Key

= Ampicillin Resistance Gene
= N-Terminal Domain Of NifA
= Central Domain Of NifA
= *ptac* Promoter
= Strong *rrn*BT$_1$T$_2$ Ribosomal Terminator

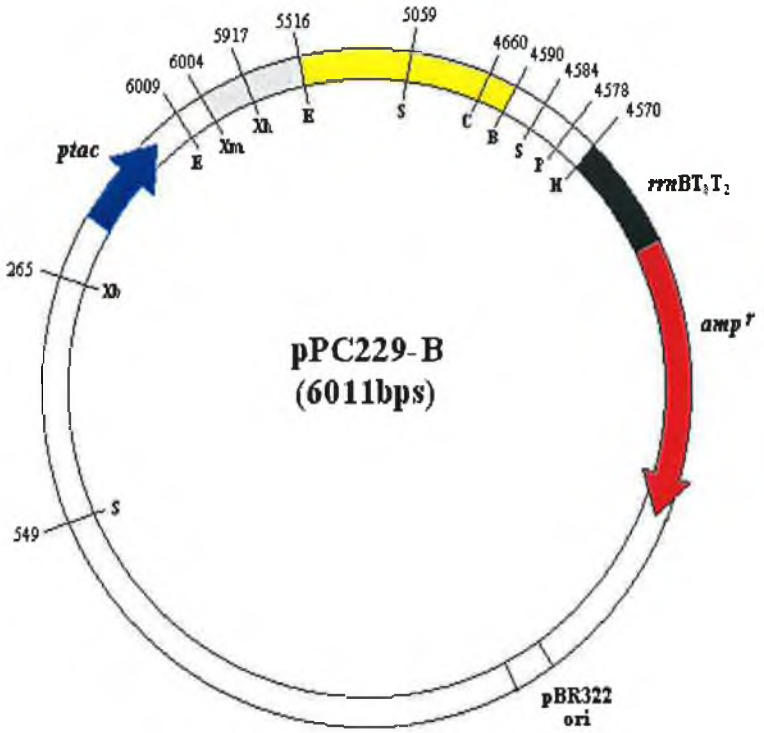
E = *Eco*RI
Xm = *Xma*I
Xh = *Xho*I
S = *Sal*I
C = *Cla*I
B = *Bam*HI
P = *Pst*I
H = *Hind*III
Xb = *Xba*I



C.) Prey Plasmids pPC229-B And pPC231-B.

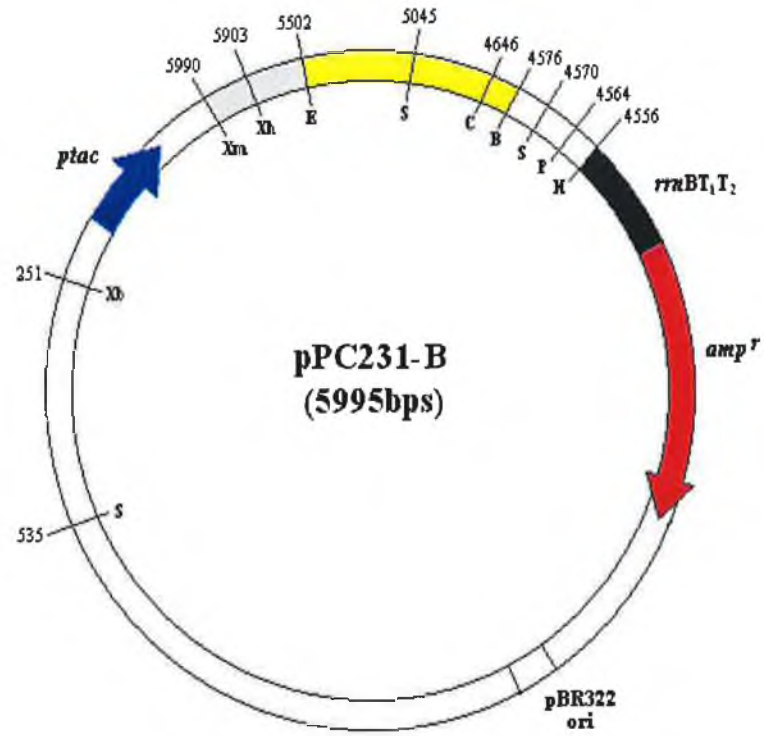
Key

= Ampicillin Resistance Gene
= N-Terminal Domain Of NifA
= Central Domain Of NifA
= *ptac* Promoter
= Strong *rrn*BT$_1$T$_2$ Ribosomal Terminator
E = *Eco*RI
Xm = *Xma*I
Xh = *Xho*I
S = *Sal*I
C = *Cla*I
B = *Bam*HI
P = *Pst*I
H = *Hind*III
Xb = *Xba*I



Key

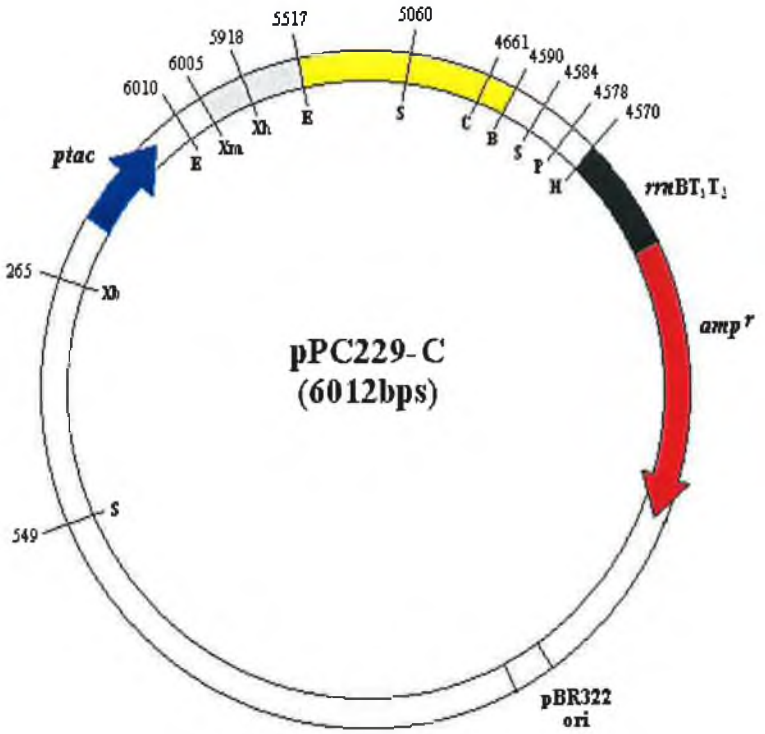
= Ampicillin Resistance Gene
= N-Terminal Domain Of NifA
= Central Domain Of NifA
= *ptac* Promoter
= Strong *rrn*BT$_1$T$_2$ Ribosomal Terminator
E = *Eco*RI
Xm = *Xma*I
Xh = *Xho*I
S = *Sal*I
C = *Cla*I
B = *Bam*HI
P = *Pst*I
H = *Hind*III
Xb = *Xba*I



D.) Prev Plasmids pPC229-C And pPC231-C.

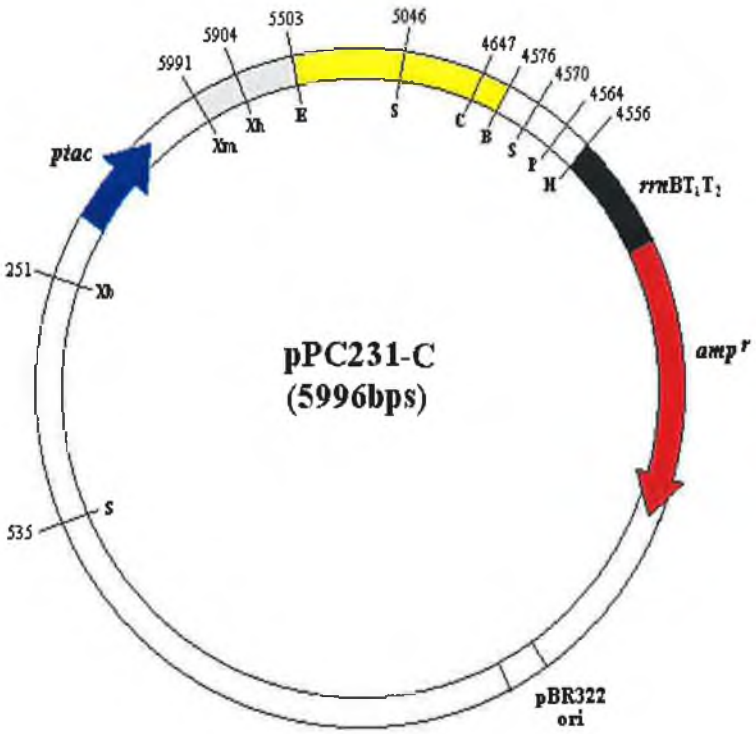
Key

= Ampicillin Resistance Gene
= N-Terminal Domain Of NifA
= Central Domain Of NifA
= *ptac* Promoter
= Strong *rrn*BT$_1$T$_2$ Ribosomal Terminator
E = *Eco*RI
Xm = *Xma*I
Xh = *Xho*I
S = *Sal*I
C = *Cla*I
B = *Bam*HI
P = *Pst*I
H = *Hind*III
Xb = *Xba*I



Key

= Ampicillin Resistance Gene
= N-Terminal Domain Of NifA
= Central Domain Of NifA
= *ptac* Promoter
= Strong *rrn*BT$_1$T$_2$ Ribosomal Terminator
E = *Eco*RI
Xm = *Xma*I
Xh = *Xho*I
S = *Sal*I
C = *Cla*I
B = *Bam*HI
P = *Pst*I
H = *Hind*III
Xb = *Xba*I



Chapter 4

Construction Of The Reporter Plasmids pPC185 And pPC186 For The Prokaryotic Two-Hybrid System.

4.1: Introduction.

This chapter describes the construction of the reporter genes for the proposed P2H system. These were to be constructed using the *S. meliloti* 2011 *nifH* promoter, a promoter naturally regulated by the *S. meliloti* 2011 NifA protein.

Two reporter genes were to be constructed, a *nifH:lacZα* reporter gene and a *nifH:gent*^{r} (gentamycin resistance) reporter gene. The *nifH:lacZα* reporter gene was to be constructed first as it would allow the levels of *lacZα* expression from the *nifH* promoter to be monitored both on the basis of color intensities of colonies on Xgal plates and quantitatively by the Miller assay. The NifA N-terminal-central domains were known to activate transcription from *nif* promoters like the *nifH* promoter despite lacking the C-terminal DNA binding domain, although at a lower level than the entire NifA protein. In addition to this *E. coli* NtrC was known to be capable of inducing low level expression from the *nifH* promoter. The *S. meliloti* 2011 NifA protein was also known to be both temperature sensitive and oxygen sensitive. Considering all of these properties of the NifA protein the *nifH:lacZα* reporter gene would be very useful for;

- monitoring background levels of expression from the *nifH* promoter.
- testing NtrC mutant *E. coli* strains developed for use with the proposed P2H system
- studying the levels of transcriptional activation induced by the entire NifA protein and the NifA NC domain expressed from the Control and Prey plasmids discussed in chapter 3.
- optimization of the growth conditions to be used in the P2H system.

Having optimized the P2H system using the *nifH:lacZα* reporter gene the *nifH:gent*^{r} reporter could be used perform library screens, permitting direct selection for clones expressing interacting hybrid proteins on the basis of gentamycin resistance.

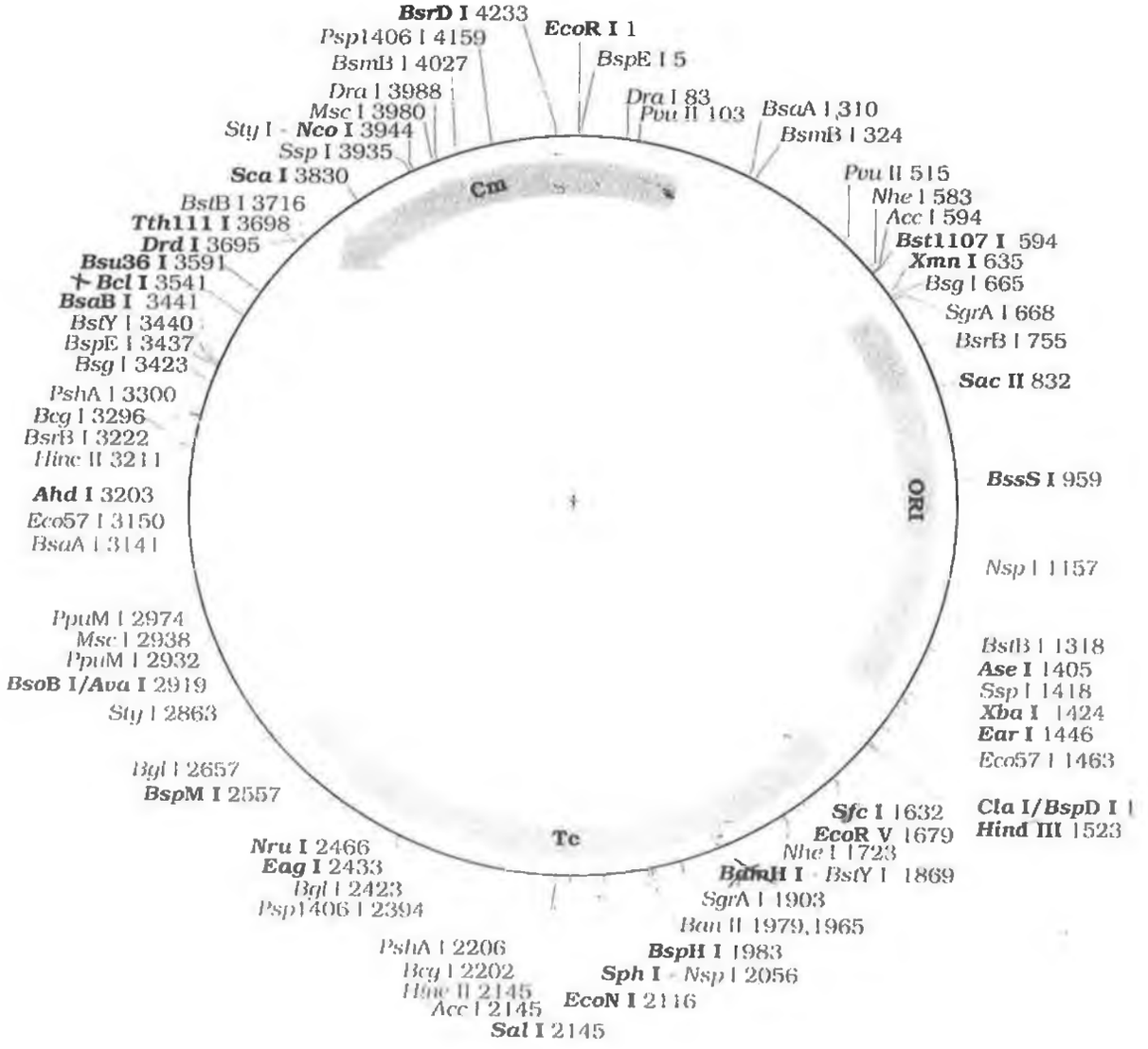
4.2: Selection Of A Suitable Plasmid For Maintaining Reporter Genes - The Formation Of Plasmid pPC224.

The constructed reporter genes were to be maintained on a multi-copy plasmid rather than being integrated into the genome of an *E. coli* strain. It was expected that the resulting increased copy number of the reporter genes would increase the sensitivity of the final P2H system and that it would add to the versatility of the system as specific reporter strains would not be required with the *nifH:gent*^{r} reporter gene. It would thus be possible to perform the assay in a wide range of genetic backgrounds. A suitable plasmid on which the reporter genes could be maintained therefore had to be selected. In addition to the reporter genes, the selected plasmid would also maintain a Bait construct for expressing hybrid proteins containing the NifA C-terminal DNA binding domain. This fact also had to be taken into consideration when selecting a suitable plasmid.

Any plasmid selected for maintaining the reporter genes and the Bait construct had to be compatible with the pKK223-3 plasmid upon which the Control and Prey plasmids discussed in chapter 3 were based and so plasmid pACYC184 was chosen. The pACYC184 plasmid had a p15A origin of replication known to be compatible with the pMB1 origin of replication present on plasmid pKK223-3. The plasmid also had a tetracycline resistance gene with unique restriction sites into which *nifH:lacZ*α or *nifH:gent*^{r} reporter genes, and later the Bait construct, could be cloned (see Figure 4.2.1). The plasmid also had a chloramphenicol resistance gene that could be used to select for the plasmid independently of the pKK223-3 based Control and Prey Vectors described in chapter 3.

As mentioned above, in addition to maintaining the reporter genes the pACYC184 plasmid would also maintain a Bait construct for expressing hybrid proteins containing the NifA C-terminal DNA binding domain. The planned Bait construct, which is discussed in chapter 5, would contain a *Bam*HI site for cloning genes of interest and making fusions with the NifA C-terminal domain sequence. In order for the *Bam*HI site within the Bait construct to be a unique restriction site and to thus allow it to be used for cloning purposes the existing *Bam*HI site in the tetracycline resistance gene of the pACYC184 plasmid had to be destroyed.

- **Figure 4.2.1: Map Of The Cloning Vector pACYC184.**



In order to destroy the *Bam*HI site of the pACYC184 plasmid the plasmid was digested with *Bam*HI. The digested plasmid DNA was then separated on a 0.7% agarose gel to verify complete digestion and the band corresponding to the linearised plasmid was purified by the gene clean procedure. The 5' cohesive *Bam*HI ends were filled by performing a Klenow reaction, the DNA was cleaned and then blunt end ligated. The ligation was used to transform DH5α and the transformation was plated out on LB media to which chloramphenicol had been added to select for the presence of the plasmid. Elimination of the *Bam*HI site would disrupt the tetracycline resistance gene on pACYC184 leading to a tetracycline sensitive phenotype so transformants were replica plated onto LB chloramphenicol and LB tetracycline plates. In this way chloramphenicol resistant, tetracycline sensitive clones were identified. Plasmid DNA was isolated from each of the identified clones and digested with *Bam*HI to verify elimination of the *Bam*HI site. In this way the desired plasmid, pPC184, was obtained. Figure 4.2.2 below shows the results of the verifying *Bam*HI digestion on plasmid pPC184.

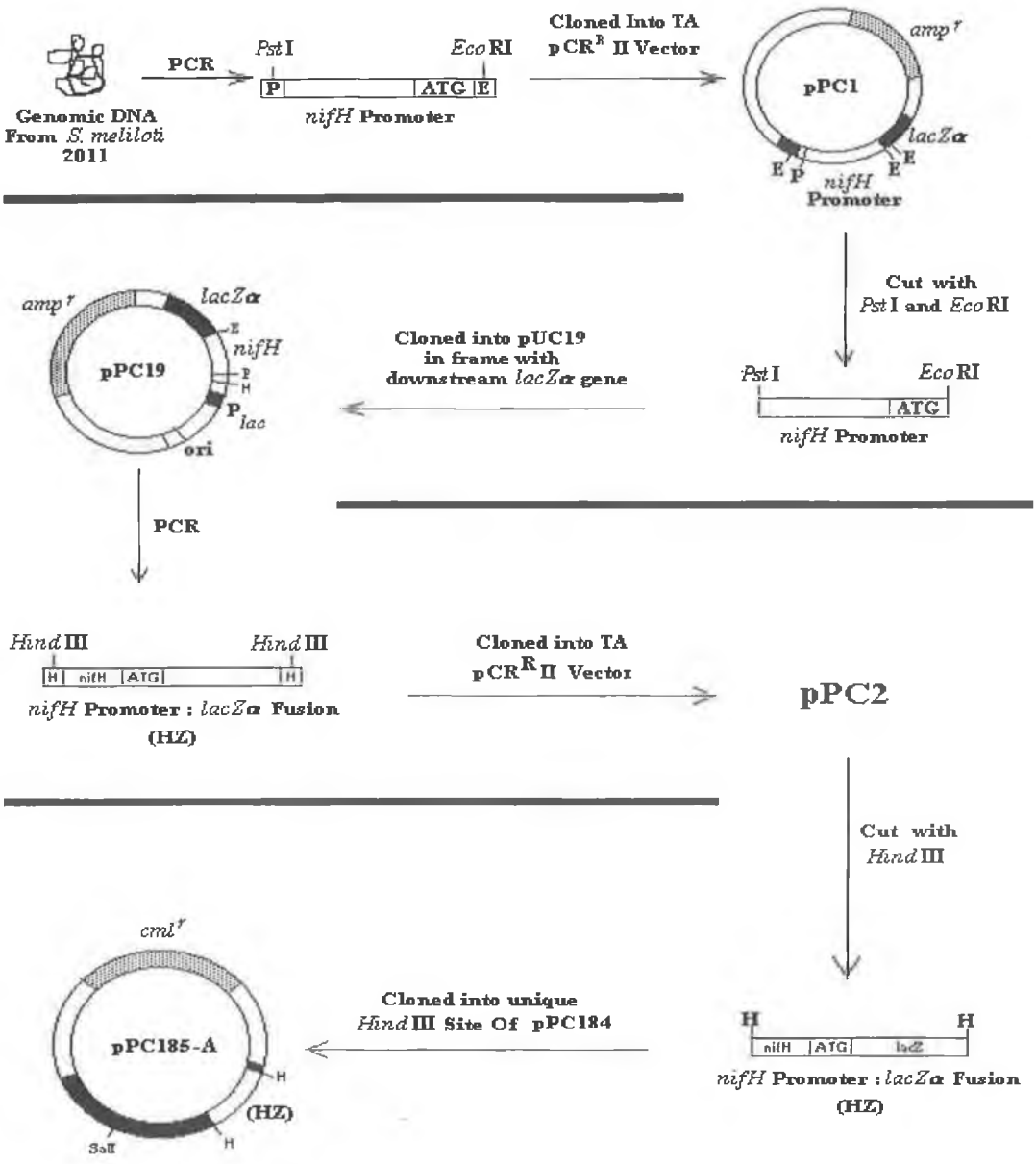
Once constructed the *nifH:lacZα* reporter gene would be cloned onto the pPC184 plasmid rather than directly into the pACYC184 plasmid to form a reporter plasmid called pPC185. The *nifH:gent*^{r} reporter would also be cloned into pPC184 to form a second reporter plasmid called pPC186. Later the Bait construct would be cloned into both the pPC185 and pPC186 reporter plasmids to form the final Bait plasmids (see chapter 5).

- **Figure 4.2.2: Restriction Analysis Of Plasmid pPC184.**

Lane No.	Sample.
1	pPC184 Uncut
2	pACYC184 Uncut
3	pPC184 cut *Bam*HI
4	pACYC184 cut *Bam*HI
5	pPC184 cut *Hind*III
6	1Kb Ladder

Figure 4.3.1 below shows a schematic representation of the strategy employed for the construction of the *nifH:lacZα* reporter gene. This strategy is described in detail in section 4.3 below. The strategy employed for the construction of the *nifH:gent*^{r} is discussed in section 4.4 and a schematic representation of the strategy employed can be seen in Figure 4.4.2.

- **Figure 4.3.1: Strategy For The Construction Of The *nifH:lacZα* Reporter Gene.**



4.3: Construction Of The *nifH:lacZα* Reporter Gene - Overview Of Strategy.

In order to construct the *nifH:lacZα* reporter gene the *S. meliloti* 2011 *nifH* promoter was amplified from total *S. meliloti* 2011 genomic DNA as an *Pst*I-*Eco*RI fragment by PCR (see Figure 4.3.1) using the primers *nifH*-F1 and *nifH*-R shown in Figure 4.3.2 below. The primed regions of the *S. meliloti* 2011 *nifH* sequence for which these primers were designed are shown in Figure 4.3.3.

- **Figure 4.3.2: Primers Used For The Amplification Of The *nifH* Promoter. From *S. meliloti* 2011 Total Genomic DNA.**

nifH-F1: AAAA - **(CTG CAG)** - CGC CCA TAC GAC ACT GTC CGT

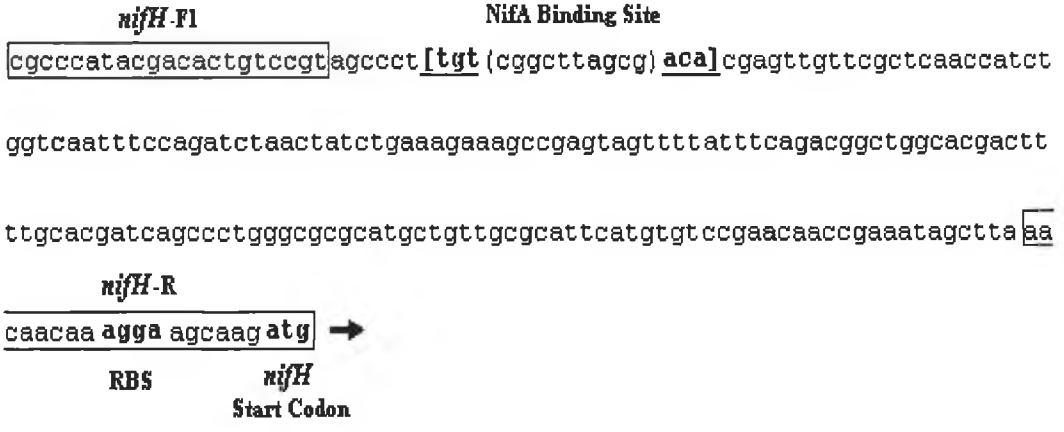
***Pst*I Site**

nifH-R: CCG - **(GAA TCC)** - AA -CAT CTT GCT TCC TTT GTT GTT

***Eco*RI site**

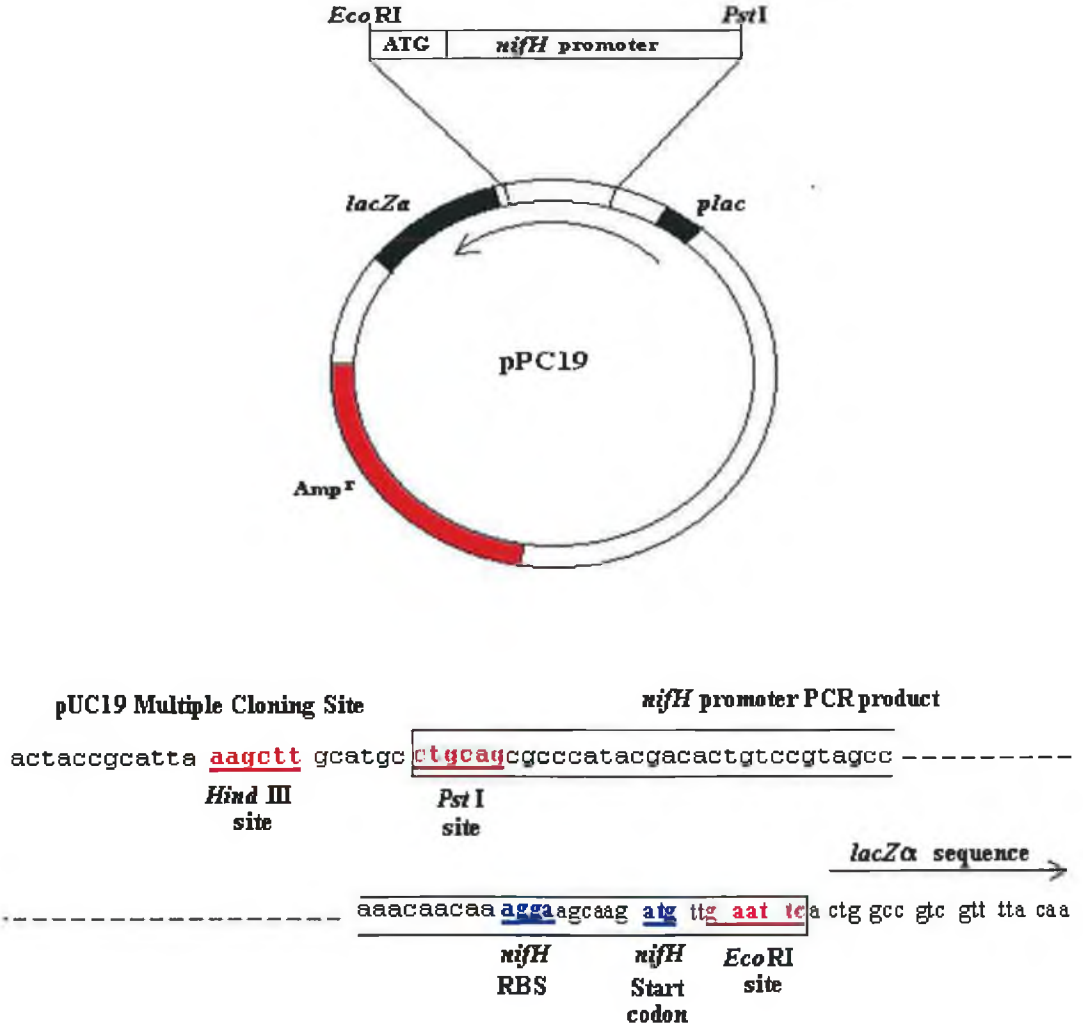
NB: Underlined Bases In Red Are Included For Maintenance Of The Reading Frame Between The *nifH* Start Codon And The *lacZα* Gene Subsequently Fused Downstream.

- **Figure 4.3.3: Primed Regions Of The *S. meliloti* 2011 *nifH* Sequence.**



The *nifH* promoter was amplified with its own start codon and the resulting PCR was cloned first into the TA vector and then sub-cloned as an *Pst*I-*Eco*RI fragment into the multiple cloning site of pUC19 to form plasmid pPC19. The *nifH*-R was designed to incorporate two additional bases into the *nifH* PCR product between the added *Eco*RI site and the *nifH* start codon. These bases were incorporated so that when the *nifH* promoter PCR product was cloned as a *Pst*I-*Eco*RI, the start codon of the *nifH* promoter PCR product would be in the correct reading frame with the downstream *lacZ*α sequence of pUC19 as shown in Figure 4.3.4.

- **Figure 4.3.4: Cloning Of The *nifH* Promoter PCR Product Into The MCS Of pUC19.**



Having made a the fusion between the *nifH* promoter and the *lacZα* gene the entire *nifH:lacZα* (HZ) reporter gene was to be amplified form pPC19 as a *Hind*III fragment using primers HZ-F and HZ-R and the resulting PCR product cloned into the TA vector. The sequences of the HZ-F and HZ-R primers used for the amplification of the *nifH:lacZα* reporter are shown in Figure 4.3.5 below and the primed regions of the pPC19 plasmid sequence are shown in Figure 4.3.6.

- **Figure 4.3.5: Primers Used For Amplification Of The *nifH:lacZα* Reporter Gene From pPC19.**

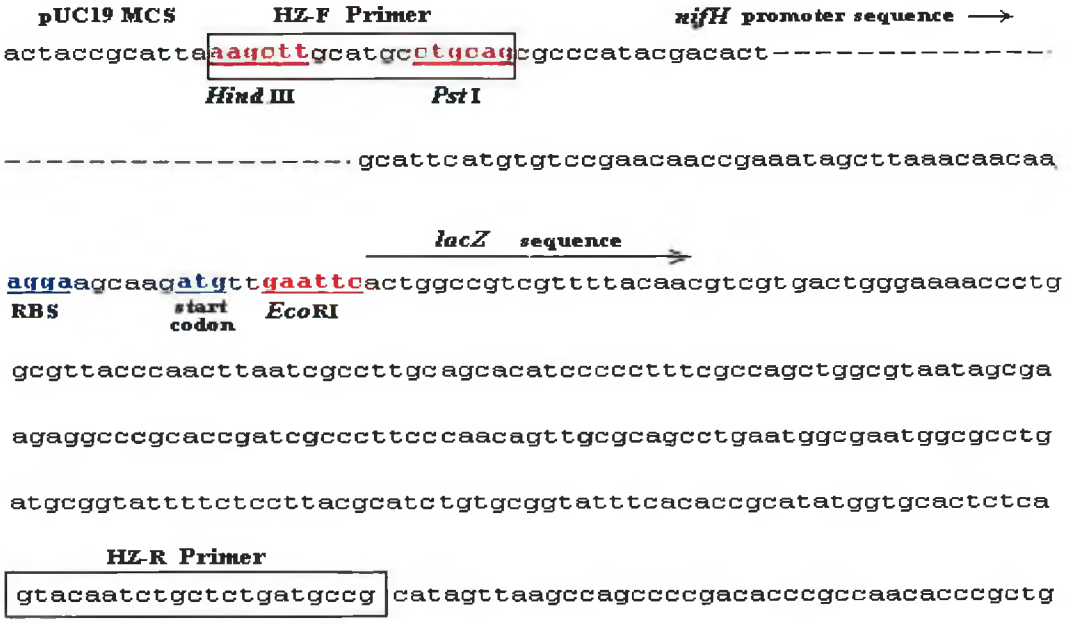
HZ-F: CC - **(AAGCTT)** - GCA TGC CTG CAG

***Hind*III site**

HZ-R: **(AAG CTT)** - CGG CAT CAG AGC AGA TTG TAC

***Hind*III site**

- **Figure 4.3.6: Primed Regions Of The pPC19 Sequence Used For The Amplification Of The *nifH:laZα* Reporter Gene.**



The *nifH:lacZα* reporter gene was then cut from the TA vector and cloned into the unique *Hind*III site in the tetracycline resistance gene of plasmid pPC184 to form the reporter plasmid pPC185. The results of the above mentioned cloning steps employed for the construction of the *nifH:lacZα* reporter gene and the reporter plasmid pPC185 are given below.

4.3.1: Amplification Of The *nifH* Promoter.

The *nifH* promoter was amplified from *S. meliloti* 2011 total genomic DNA as a *Pst*I-*Eco*RI fragment using the *nifH*-F1 and *nifH*-R primers described earlier. The PCR was separated on a 1% agarose gel. A photo of the gel obtained and the conditions used for the PCR are shown in Figure 4.9 below. As can be seen from Figure 4.3.7 a single 243bp PCR product corresponding to the desired *nifH* promoter was obtained. The next step in the construction of the *nifH:lacZα* reporter was to clone this PCR product into the TA pCR2.1 vector.

- **Figure 4.3.7: PCR Amplification Of The *S. meliloti* 2011 *nifH* Promoter.**

Lane No.	**Sample.**
1	*nifH* PCR Product
2	1Kb Ladder

PCR Conditions.

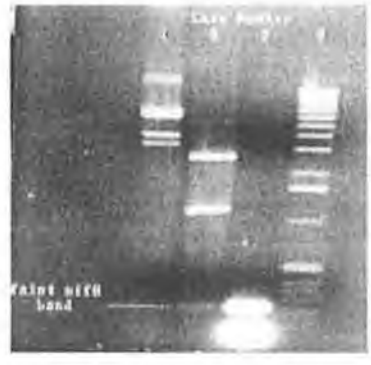
Annealing temperature	60 °C
Annealing time	30 sec
Extension time	1 min

4.3.2: Cloning Of The *nifH* Promoter PCR Product Into The TA Vector To Form Plasmid pPC1-A.

The *nifH* promoter PCR product was ligated with the TA pCR2.1 vector according to the TA protocol. The resulting ligation was used to transform INVαF' cells and the transformation was plated on LB ampicillin/Xgal plates. White colonies were isolated and screened by performing *Eco*RI digests on plasmid DNA isolated from each clone. In this way the desired plasmid, pPC1-A, was obtained. Figure 4.3.8 shows the results of the verifying restriction analysis on the pPC1-A plasmid.

- **Figure 4.3.8: Restriction Analysis Of Plasmid pPC1-A.**

Lane No.	Sample.
1	pPC1-A uncut
2	pPC1-A cut *Pst*I-*Eco*RI
3	*nifH* PCR Product
4	1Kb Ladder



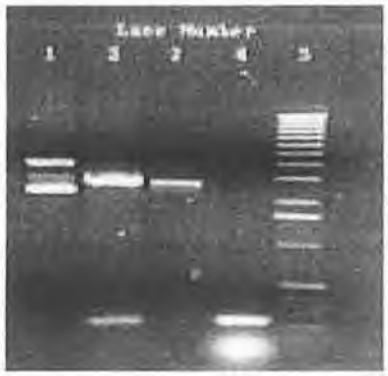
4.3.3: Sub-cloning Of The *nifH* PCR Product From pPC1-A Into pUC19 To Form Plasmid pPC19.

Plasmid pPC1-A was digested with *Pst*I-*Eco*RI and the digested DNA was separated on a 1% agarose gel. The 243bp band corresponding to the *nifH* promoter fragment was purified from the gel using the gene clean procedure. pUC19 was also digested with *Pst*I-*Eco*RI, the digested DNA was cleaned and then ligated with the purified *nifH* promoter fragment. The ligation was used to transform DH5α cells and the transformation was plated out on LB ampicillin Xgal plates. All of the colonies obtained on the plates were dark blue in color. A number of the blue colonies were picked and screened by performing *Pst*I-*Eco*RI double digests on plasmid DNA

isolated from each clone. In this way the desired plasmid, pPC19, was obtained. Figure 4.3.9 shows the verifying restriction analysis on the isolated pPC19 plasmid.

- **Figure 4.3.9: Restriction Analysis Of Plasmid pPC19.**

Lane No.	Sample
1	pPC19 uncut
2	pPC19 cut *Pst*I-*Eco*RI
3	pUC19 cut *Pst*I-*Eco*RI
4	*nifH* PCR product
5	1Kb Ladder



As mentioned above, all of the transformants obtained were blue in color. Cloning into the MCS of pUC19 usually results in white clonies due to disruption of the reading frame of the *lacα* peptide sequence with respect to the upstream *plac* promoter. However, in this case the *nifH* promoter was being cloned into the MCS in frame with the downstream *lacZα* sequence. *E. coli* NtrC was known to activate transcription of NifA promoters like the *nifH* promoter and thus this was thought to be the cause of the blue color development. It was thought that the *lacZα* gene was being expressed from the cloned *nifH* promoter (rather from the upstream *plac* promoter on pUC19) due to activation of the *nifH* promoter by *E. coli* NtrC. Although the degree of activation of the *nifH* promoter by NtrC was known to be low, the extremely high copy number of pUC19 compensated for the low level of activation thus resulting in dark blue colonies.

4.3.4: Amplification Of The *nifH:lacZα* Fusion From pPC19.

Having isolated plasmid pPC19, the *nifH:lacZα* was amplified as a *Hind*III fragment using the HZ-F and HZ-R primers described earlier. The PCR was separated on a 1% agarose gel. A photo of the gel and the conditions used for the PCR are shown in

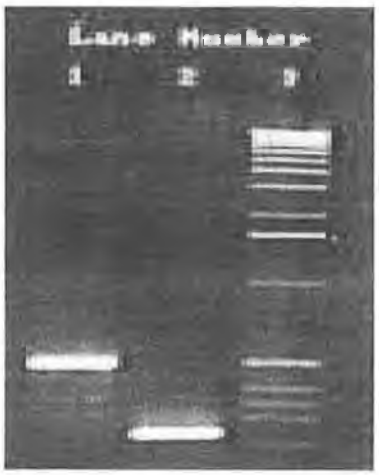
Figure 4.3.10 below. As can be seen from Figure 4.3.10 a single band corresponding to the desired *nifH:lacZα* sequence was obtained. This PCR product was then cloned into the TA pCR2.1 vector.

- **Figure 4.3.10: PCR Amplification Of The *nifH:lacZα* Reporter Gene From pPC19.**

Lane No.	Sample
1	*nifH:lacZα* PCR product
2	*nifH* promoter PCR product
3	1Kb Ladder

PCR Conditions

Annealing temperature	57 °C
Annealing time	30 sec
Extension time	1 min



4.3.5: Cloning Of The *nifH:lacZα* Reporter Gene Into The TA Vector To Form Plasmid pPC2.

The *nifH:lacZα* reporter gene PCR product was ligated with the TA pCR2.1 vector according to the TA protocol. The ligation was used to transform INVαF' cells and the transformation was plated out on LB ampicillin Xgal plates. Again all of the transformants obtained were dark blue in color. As before, it was thought that this was due to NtrC activation of the *nifH:lacZα* fusion present on the high copy number TA pCR2.1 vector. Some of the blue colonies were picked and screened by performing *Hind*III digests on plasmid DNA isolated from each of the clones. In this way the desired plasmid, pPC2, was obtained. Figure 4.3.11 shows the results of the verifying restriction analysis on plasmid pPC2.

- **Figure 4.3.11: Restriction Analysis Of The pPC2 Plasmid.**

Lane No.	Sample
1	pPC2 uncut
2	pPC2 cut *Hind*III
3	*nifH:lacZα* PCR product
4	1Kb Ladder

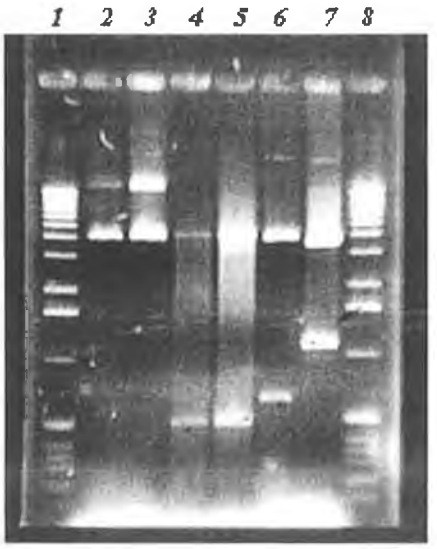
4.3.6: Sub-cloning Of The *nifH:lacZα* Reporter Gene PCR Product From pPC2 Into pPC184 To Form The Reporter Plasmid pPC185.

Plasmid pPC2 was digested with *Hind*III and the digested DNA was separated on a 1% agarose gel. The band corresponding to the *nifH:lacZα* reporter gene fragment was purified from the gel by the gene clean procedure. pPC184 was also digested with *Hind*III and the linearised plasmid DNA was then treated with Calf Intestinal Phosphatase (CIP) to remove the 5' phosphates. The dephosphorylated pPC184 DNA was then ligated with the purified *nifH:lacZα* reporter gene fragment. The ligation was used to transform DH5α and the transformation was plated out on LB chloramphenicol Xgal media. The reason for the incorporation of the Xgal into the media was to make use of *E. coli* NtrC activation of the *nifH:lacZα* reporter gene for identification of the desired plasmids containing the *nifH:lacZα* reporter gene. Two types of colonies were obtained, one type were very light blue in color while the other were a darker blue color. Colonies of each type were screened by performing *Hind*III digests on plasmid DNA isolated from each clone. Having identified several clones containing the desired *nifH:lacZα* *Hind*III fragment the orientation of the *Hind*III fragment was determined for each clone by performing *Sal*I-*Pst*I double digests. Two plasmids were to be isolated, pPC185-A and pPC185-B, differing only in the

orientation of the *nifH:lacZα* reporter gene. Restriction maps of these plasmids can be seen in Figure 4.3.13. As can be seen from the restriction maps in Figure 4.3.13 a *Sal*I-*Pst*I double digest of pPC185-A was expected to yield two bands, one of 638bps and one of 4104bps. A *Sal*I-*Pst*I double digest of pPC185-B was also expected to yield two bands, one of 1107bps and one of 3635bps. Thus by performing *Sal*I-*Pst*I double digests the two plasmids, pPC185-A and pPC185-B, were obtained. The results of the verifying restriction analysis on both of these plasmids can be seen in Figure 4.3.12.

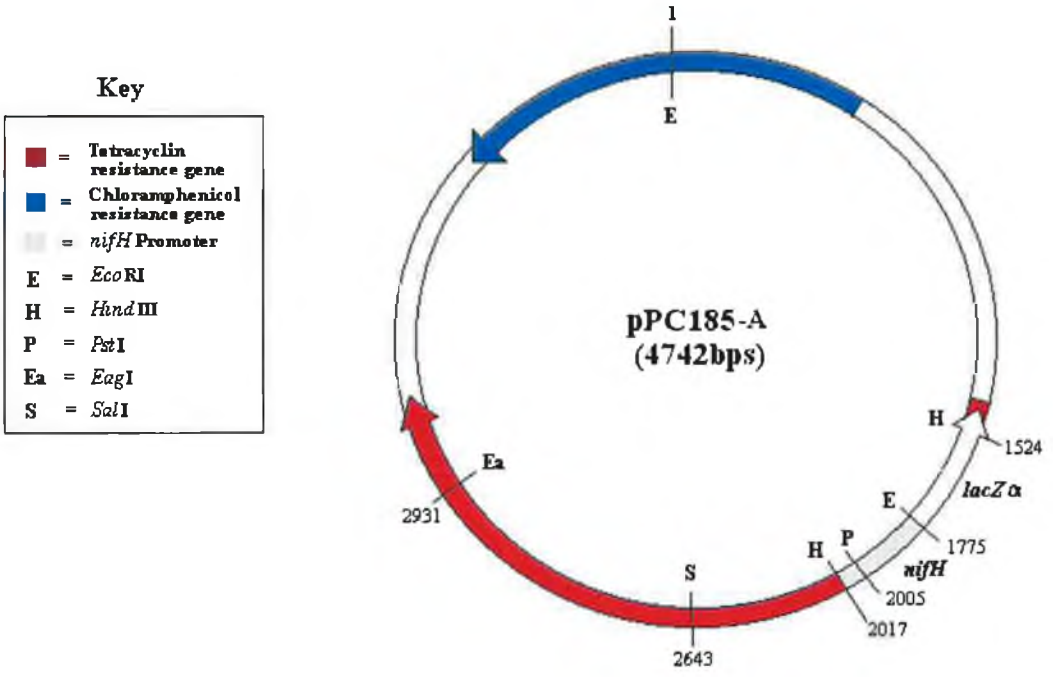
- **Figure 4.3.12: Restriction Analysis Of Plasmid pPC185-A And pPC185-B.**

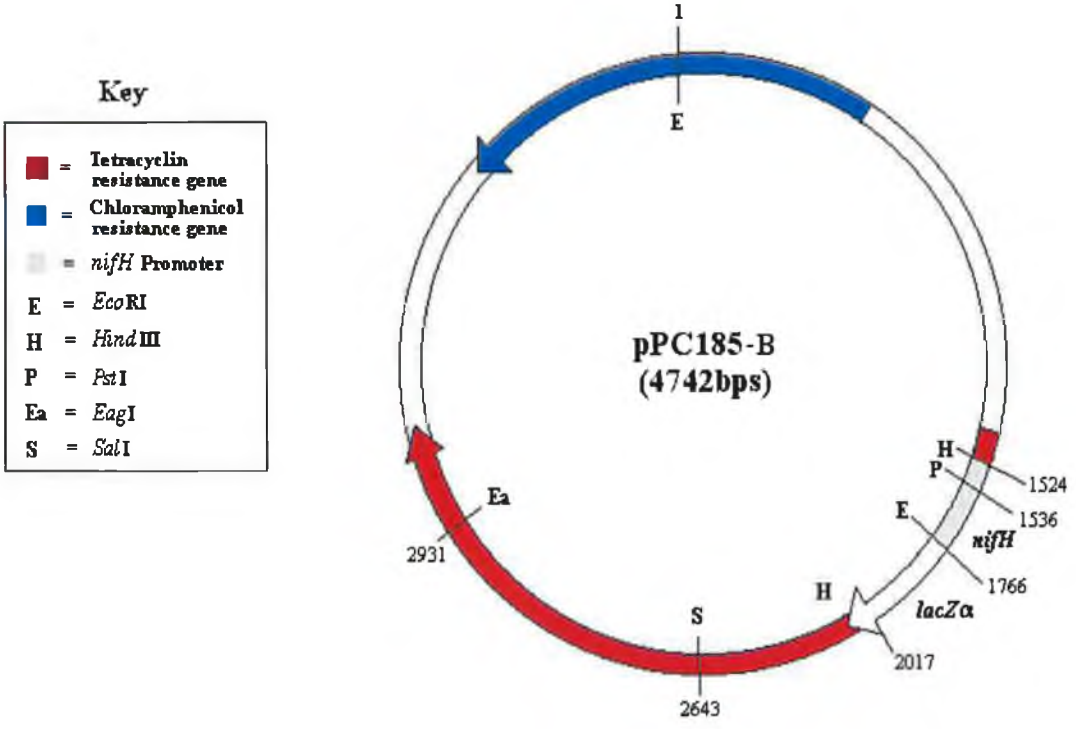
Lane No.	Sample.
1	1Kb Ladder
2	pPC185-A uncut
3	pPC185-B uncut
4	pPC185-A cut *Hind*III
5	pPC185-B cut *Hind*III
6	pPC185-A *Sal*I-*Pst*I
7	pPC185-B cut *Sal*I-*Pst*I
8	1Kb Ladder



Plasmid pPC185-A was found to generate the very light blue colored colonies while pPC185-B was found to produce the slightly darker blue colonies. As explained earlier the light blue color caused by the pPC185-A plasmid was thought to be due to weak activation of the *nifH:lacZα* reporter gene by *E. coli* NtrC. The color development was not as intense as that observed when the *nifH* promoter was cloned into pUC19 or when the *nifH:lacZα* reporter gene was cloned into the TA vector to form pPC2 because the pACYC184 based pPC185-A plasmid was a much lower copy number than pUC19 or TA vector. The slightly darker blue color development of colonies with plasmid pPC185-B was thought to be due to additional expression of the *nifH:lacZα* reporter gene caused by mRNA transcripts initiated from the upstream tetracycline resistance gene promoter present on plasmid pPC184 (pACYC184).

- **Figure 4.3.13: Restriction Maps Of The LacZα Reporter Plasmids pPC185-A And pPC185-B.**

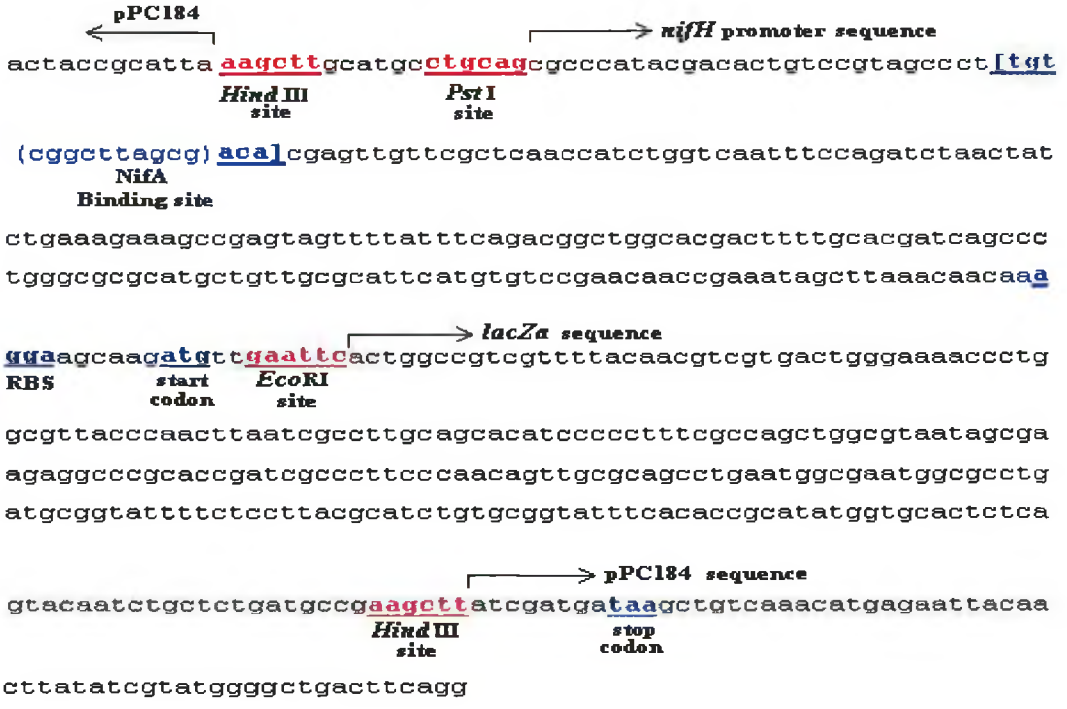




As can be seen from Figure 4.3.13, on the pPC185-B plasmid the tetracycline resistance gene and the *nifH:lacZα* reporter gene are both transcribed in the same direction and thus from the same strand. The *nifH:lacZα* was cloned into the *Hind*III site which lies within the promoter for the tetracycline resistance gene on plasmid pPC184. Cloning into this site does not always result in the inactivation of the promoter and so transcripts initiated from this promoter would read through the *nifH:lacZα* reporter gene generating mRNA transcripts containing the coding sequence for LacZα. This additional expression of the *nifH:lacZα* reporter would not occur on the pPC185-A plasmid because the tetracycline resistance gene would be transcribed in the opposite direction to the *nifH:lacZα* reporter gene and thus from the opposite strand. Transcripts initiated from the tetracycline resistance gene would read through the *nifH:lacZα* reporter gene but the mRNA generated would be the non-sense strand of the *nifH:lacZα* reporter gene and thus would not be translated to give functional LacZα.

To verify that the sequence of the *nifH:lacZα* reporter was correct the *nifH:lacZα* reporter present on plasmid pPC185-A was sequenced and the sequence obtained is shown in Figure 4.3.14 below.

- **Figure 4.3.14: Sequence Of The *nifH:lacZα* Reporter On pPC185-A.**



4.4: Construction Of The *nifH:gent*^{r} Reporter Gene - Overview Of Strategy.

Figure 4.4.2 shows a schematic representation of the strategy employed for the construction of the *nifH:gent*^{r} reporter gene and the strategy is discussed in detail below.

In order to construct the *nifH:gent*^{r} reporter gene the gentamycin resistance gene was amplified without its own start codon and as an *EcoR*I-*Hind*III fragment from the plasmid pJQ200ks using the primers *gent*-F and *gent*-R shown in Figure 4.4.1. The primed regions of the gentamycin resistance gene are shown in Figure 4.4.3. The resulting PCR product was first cloned into the TA pCR2.1 vector and then sub-cloned into pPC225 (described in Chapter 3) to form plasmid pPC226. The pPC225 plasmid was used in this case simply because it possessed suitable restriction sites in the correct order to allow a fusion to be made between the *S. meliloti* 2011 *nifH* promoter and the gentamycin resistance gene.

- **Figure 4.4.1: Primers Used For The Amplification Of The Gentamycin Resistance Gene From pJQ200ks.**

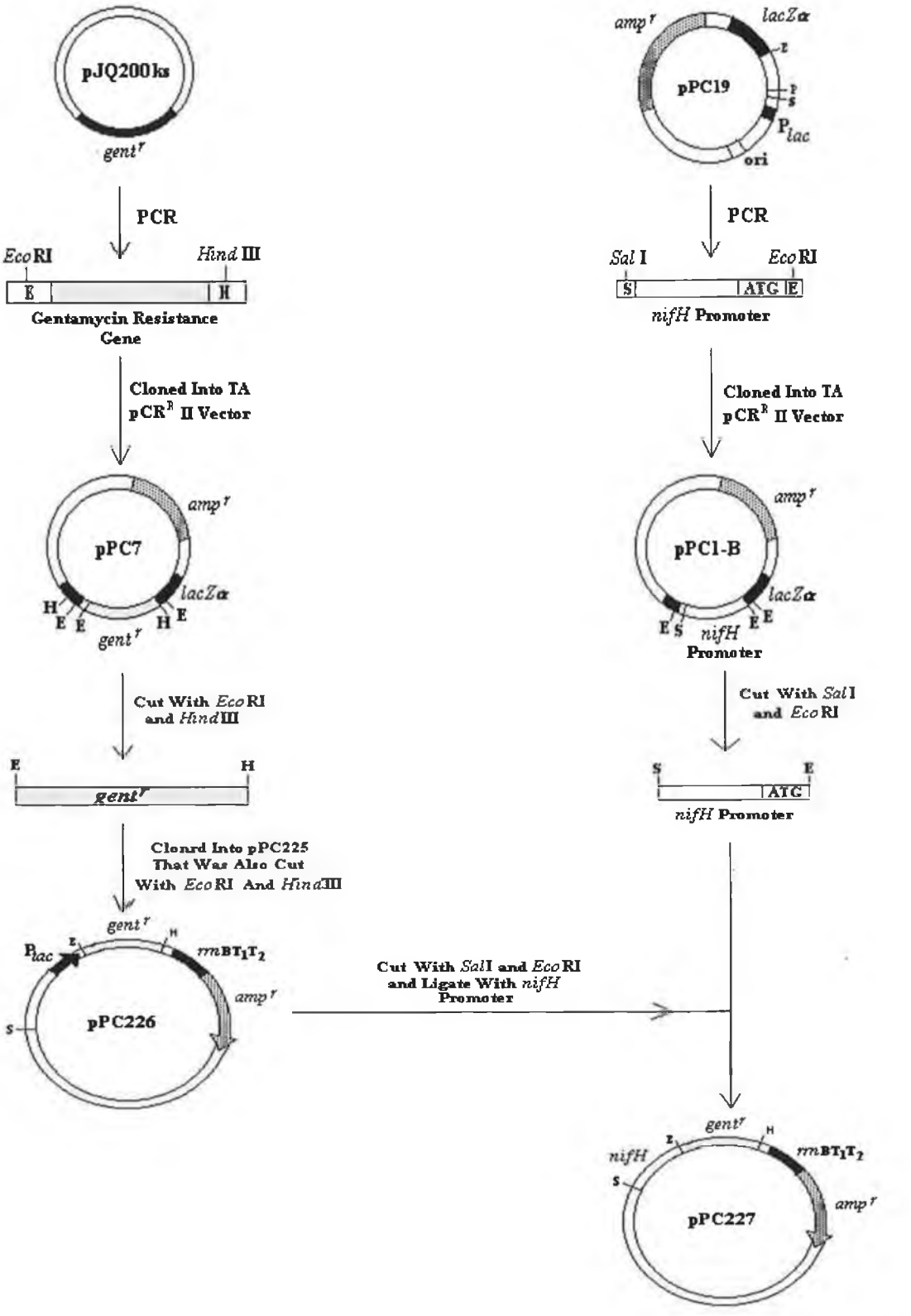
gent-F: **(GAA TTC)** - A - TTA CGC AGC AGC AAC GAT GTT
***Eco*RI site**

gent-R: **(AAG CTT)** - TTA GGT GGC GGT ACT TGG GT
***Hind*III site**

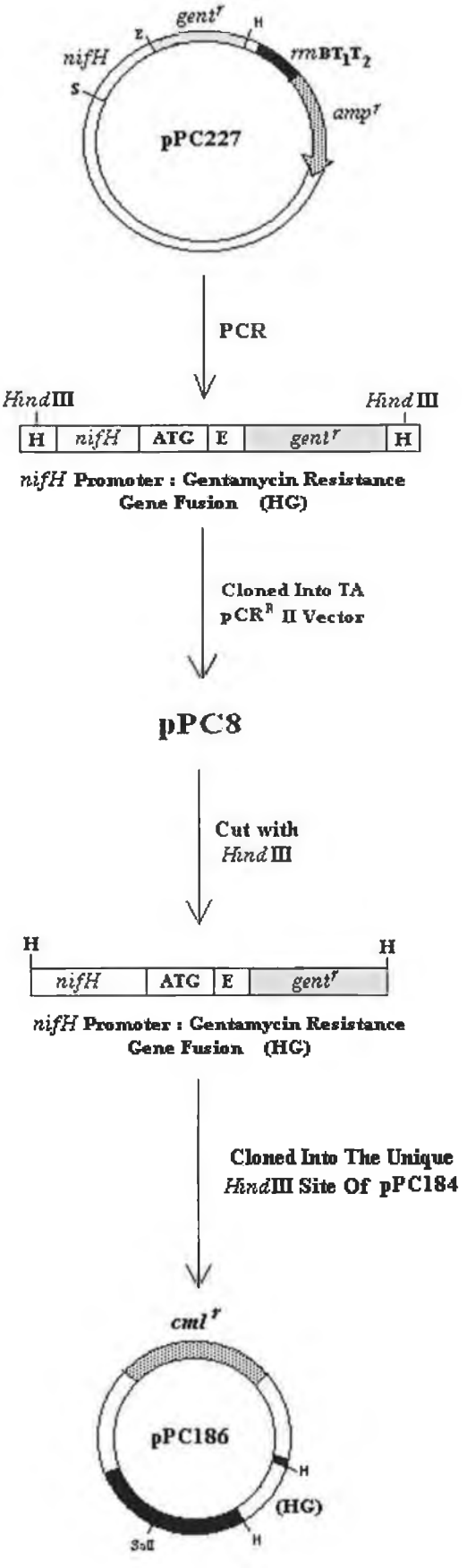
NB: Underlined Base In Red Was Included For Maintenance Of The Reading Frame Between The Gentamycin Resistance Gene And The Start Codon Of The *nifH* Promoter With Which It Would Later Be Fused

The *nifH* promoter was amplified from pPC19 as an *Sal*I-*Eco*RI fragment using the primer *nifH*-F2 shown in Figure 4.4.4 and the *nifH*-R primer used in its initial amplification from *S. meliloti* 2011 genomic DNA. The primed regions of the *nifH* promoter sequence present on plasmid pPC19 for which these primers were designed

- **Figure 4.4.2: Strategy For The Construction Of The *nifH:gent*^{r} Reporter Gene And The pPC186 Reporter Plasmid.**

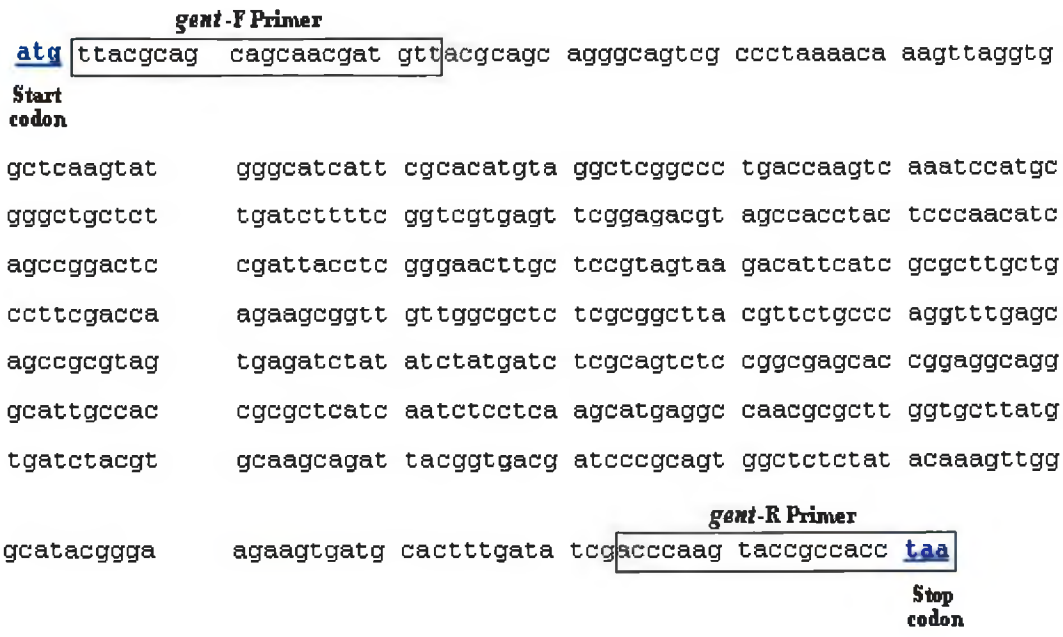


Continued overleaf →


gent^r
H
E
nifH
S
rrnBT1T2
amp^r
pPC227
PCR
HindIII
HindIII
H
nifH
ATG
E
gent^r
H
nifH Promoter : Gentamycin Resistance Gene Fusion (HG)
Cloned Into TA pCR^R II Vector
pPC8
Cut with HindIII
H
H
nifH
ATG
E
gent^r
nifH Promoter : Gentamycin Resistance Gene Fusion (HG)
Cloned Into The Unique HindIII Site Of pPC184
cml^r
pPC186
H
(HG)
H
SalI

are shown in Figure 4.4.5. Once amplified the resulting *nifH* promoter PCR product was cloned into the TA vector.

- **Figure 4.4.3: Primed Regions Of The Gentamycin Resistance Gene From pJQ200ks.**



- **Figure 4.4.4: Primers Used For The Amplification Of The *nifH* Promoter From pPC19.**

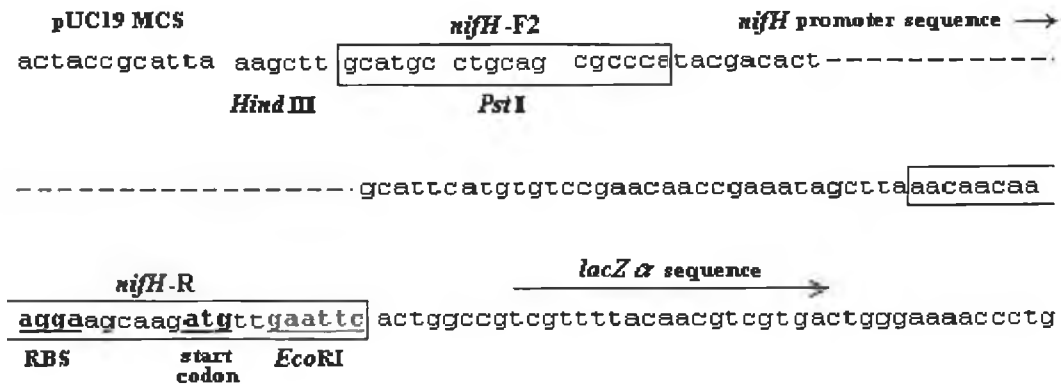
nifH-F2: **(GTC GAC)** - GCA TGC CTG CAG CGC CCA

***Sal*I site**

nifH-R: CCG - **(GAA TCC)** - AA - CAT CTT GCT TCC TTT GTT GTT

***Eco*RI site**

- **Figure 4.4.5: Primed Regions Of The *nifH* Promoter On pPC19.**



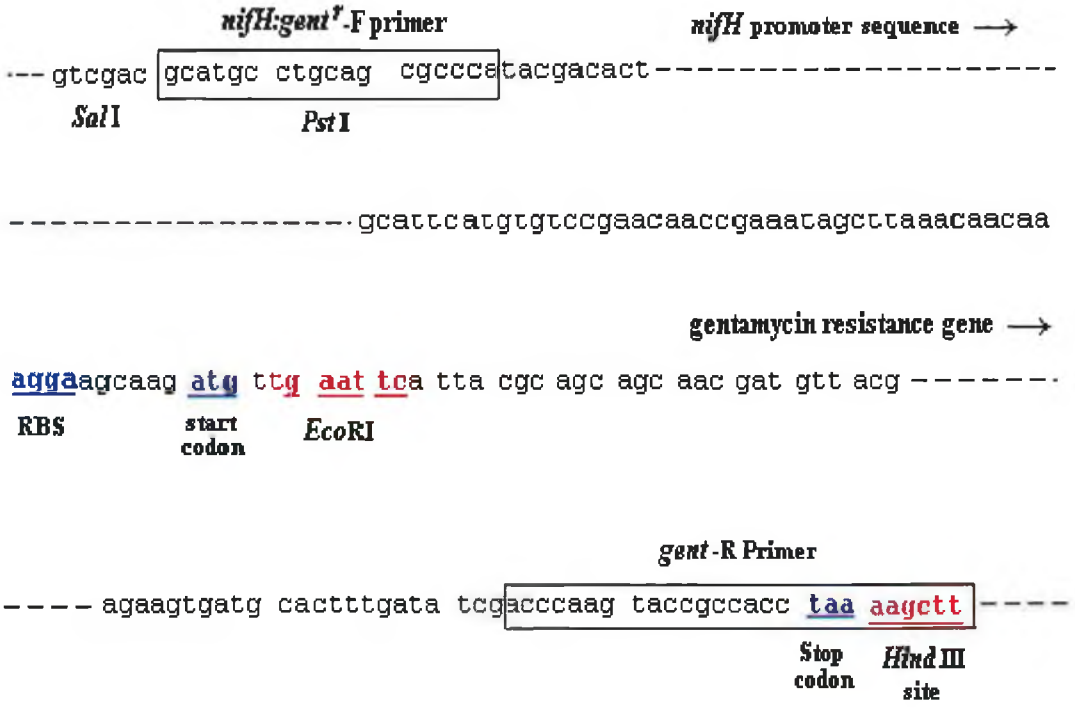
The amplified *nifH* promoter *Sal*I-*Eco*RI fragment was cut from the TA vector and sub-cloned into pPC226 to make an in frame fusion with the gentamycin resistance gene . The resulting plasmid with the *nifH:gent*^{r} reporter gene was called pPC227. Having made a fusion between the *nifH* promoter and the gentamycin resistance gene the entire *nifH:gent*^{r} reporter gene was amplified from pPC227 as a *Hind*III fragment using a forward primer called *nifH:gent*^{r}-F shown in Figure 4.4.6 and the *gent*-R reverse primer used earlier for the amplification of the gentamycin gene from pJQ200ks. Figure 4.4.7 shows the primed regions of the *nifH:gent*^{r} sequence.

- **Figure 4.4.6: Primers Used For The Amplification Of The *nifH:gent*^{r} Reporter Gene From pPC227.**

nifH:gent^r-F: **(AAG CTT)** - GCA TGC CTG CAG CGC CCA
***Hind*III site**

gent-R: **(AAG CTT)** - TTA GGT GGC GGT ACT TGG GT
***Hind*III site**

- **Figure 4.4.7: Primed Regions Of The *nifH:gent*^{r} Sequence On pPC227.**



The resulting *nifH:gent*^{r} PCR product was first cloned into the TA pCR2.1 vector and then sub-cloned into the unique *Hind*III site of plasmid pPC184 to form the reporter.plasmid pPC186. The results of the above mentioned cloning steps employed for the construction of the *nifH:gent*^{r} reporter gene and the pPC186 reporter plasmid are given below.

4.4.1: Amplification Of The Gentamycin Resistance Gene From pJQ200ks.

The gentamycin resistance gene (*gent*^{r}) was amplified from pJQ200ks as an *EcoRI-Hind*III fragment using the *gent*-F and *gent*-R primers described earlier. The PCR product was separated on a 1% agarose gel. A photo of the gel obtained and the

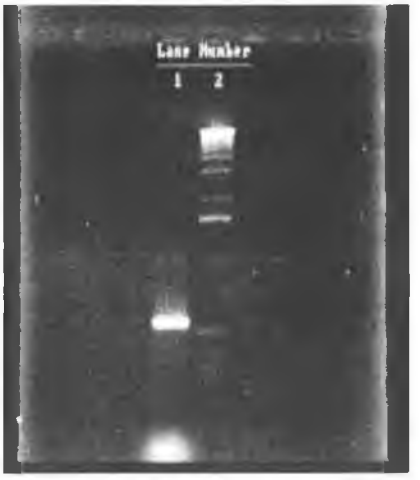
conditions used for the PCR are shown in Figure 4.4.8 below. As can be seen from Figure 4.4.8, a band of 544bps corresponding to the desired gentamycin resistance gene was obtained.

- **Figure 4.4.8: PCR Amplification Of The Gentamycin Resistance Gene (*gent*^{r}) From pJQ200ks.**

Lane No.	Sample.
1	*gent*^{r} PCR product
2	1Kb Ladder

PCR Conditions

Annealing temperature	57 ^oC
Annealing time	30 sec
Extension time	1 min

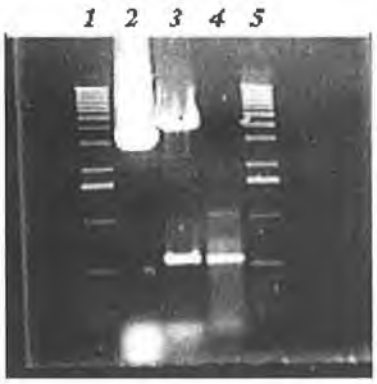


4.4.2: Cloning Of The *gent*^{r} PCR Product Into The TA Vector To Form Plasmid pPC7.

The *gent*^{r} PCR product was ligated with the TA pCR2.1 vector according to the TA protocol. The ligation was used to transform INVαF' cells and the transformation was plated onto LB ampicillin Xgal plates. White colonies were picked and screened by performing *Eco*RI digests on plasmid DNA isolated from each clone. In this way the desired plasmid pPC7 was obtained. Figure 4.4.9 shows the results of the verifying restriction analysis on plasmid pPC7.

- **Figure 4.4.9: Restriction Analysis Of Plasmid pPC7.**

Lane No.	Sample
1	1Kb Ladder
2	pPC7 uncut
3	pPC7 cut *Eco*RI-*Hind*III
4	*gent*^{r} PCR product
5	1Kb Ladder



4.4.3: Sub-cloning Of The *gent*^{r} PCR Product From pPC7 Into pPC225 To Form Plasmid pPC226.

Plasmid pPC7 was digested with *Eco*RI-*Hind*III and the digested DNA was separated on a 1% agarose gel. The 544bp band corresponding to the *gent*^{r} fragment was purified from the gel using the gene clean procedure. Plasmid pPC225 was also digested with *Eco*RI-*Hind*III and the digested DNA was separated on a 0.7% agarose gel. The 4.5Kb band corresponding to the pPC223 portion of the pPC225 plasmid was purified from the gel using the gene clean procedure and ligated with the purified *gent*^{r} *Eco*RI-*Hind*III fragment. The transformation was used to transform DH5α cells and the transformation was plated out on LB ampicillin. Colonies were screened by performing *Eco*RI-*Hind*III double digests on plasmid DNA isolated from each clone. In this way the desired plasmid pPC226 was obtained. Figure 4.4.10 shows the results of the verifying restriction analysis on plasmid pPC226.

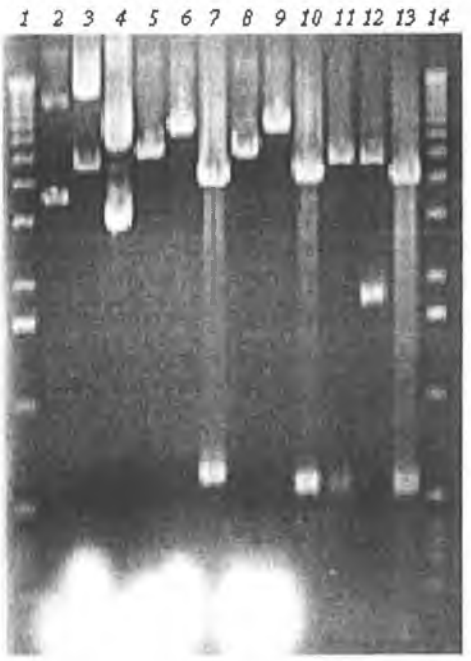
4.4.4: Amplification Of The *nifH* Promoter From Plasmid pPC19.

The *nifH* promoter was amplified from plasmid pPC19 as a *Sal*I-*Eco*RI fragment using the *nifH*-F2 and *nifH*-R primers described earlier. The PCR was separated on a 1%

agarose gel. A photo of the gel obtained and the PCR conditions used are shown in Figure 4.4.11. As can be seen from Figure 4.4.11 a single band of 251bps corresponding to the desired *nifH* promoter was obtained.

- **Figure 4.4.10: Restriction Analysis Of Plasmid pPC226.**

Lane No.	Sample
1	1Kb Ladder
2	pPC226 uncut
3	pPC225 uncut
4	pPC7 uncut
5	pPC226 cut *Hind*III
6	pPC225 cut *Hind*III
7	pPC7 cut *Hind*III
8	PC226 cut *Eco*RI
9	pPC225 cut *Eco*RI
10	pPC7 cut *Eco*RI
11	pPC226 cut *Eco*RI-*Hind*III
12	pPC225 cut *Eco*RI-*Hind*III
13	pPC7 cut *Eco*RI-*Hind*III
14	1Kb Ladder

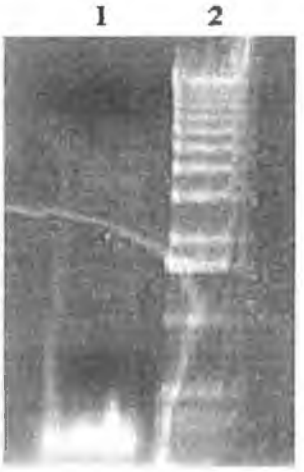


- **Figure 4.4.11: PCR Amplification Of The *nifH* Promoter From pPC19.**

Lane No.	Sample
1	*nifH* promoter PCR product
2	1Kb Ladder

PCR Conditions

Annealing temperature	60 °C
Annealing time	30 sec
Extension time	1 min

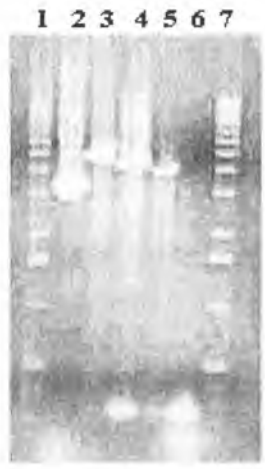


4.4.5: Cloning Of The *nifH* Promoter *Sal*I-*Eco*RI PCR Product Into The TA vector To Form Plasmid pPC1-B.

The *nifH* promoter PCR product amplified from plasmid pPC19 was ligated with the TA pCR2.1 vector and the ligation was used to transform INVαF' cells. The transformation was plated on LB ampicillin Xgal and white colonies where picked off for screening. Clones were screened by performing *Eco*RI digests on plasmid DNA isolated from each clone. In this way the desired plasmid pPC1-B was obtained and Figure 4.4.12 shows the results of the verifying restriction analysis on this plasmid.

- **Figure 4.4.12: Restriction Analysis Of Plasmid pPC1-B.**

Lane No.	Sample
1	1Kb Ladder
2	pPC1-B uncut
3	pPC1-B cut *Sal*I
4	pPC1-B cut *Eco*RI
5	pPC1-B *Sal*I-*EcoRI*
6	*nifH* promoter PCR product
7	1Kb Ladder



4.4.6: Sub-cloning Of The *nifH* Promoter PCR Product From pPC1-B Into pPC226 To Form Plasmid pPC227.

Plasmid pPC1-B was digested with *Sal*I-*Eco*RI and the digested plasmid DNA was separated on a 1% agarose gel. The 251bp *Sal*I-*Eco*RI fragment corresponding to the *nifH* promoter was purified from the gel by the gene clean procedure. Plasmid pPC226 was also digested with *Sal*I-*Eco*RI to excise a 532bp fragment and the digested DNA was separated on a 0.7% agarose gel. The 4.5kb band corresponding to the remainder of the plasmid was purified from the gel by the gene clean procedure and ligated with the purified *nifH* promoter *Sal*I-*Eco*RI fragment. The ligation was used to transform

DH5α cells. Transformants obtained were first screened by performing *Pst*I digests on plasmid DNA isolated from each clone. The reason for this was that the pPC226 plasmid did not possess any *Pst*I sites (the plasmid is basically pPC223 with the *gent*^{r} gene on an *Eco*RI-*Hind*III fragment). The *Sal*I-*Eco*RI *nifH* promoter fragment amplified by PCR contained a *Pst*I site (see Figure 4.4.5). Thus, only the desired plasmid, pPC227, containing the *nifH* promoter fragment would be linearised by *Pst*I digestion. In this way the plasmid pPC227 was obtained and Figure 4.4.13 below shows the results of the verifying restriction analysis on plasmid pPC227.

- **Figure 4.4.13: Restriction Analysis Of Plasmid pPC227.**

Lane No.	Sample
1	1Kb Ladder
2	pPC226 uncut
3	pPC227 uncut
4	pPC226 cut *Pst*I
5	pPC227 cut *Pst*I
6	pPC226 cut *Sal*I
7	pPC227 cut *Sal*I
8	pPC226 cut *Eco*RI
9	pPC227 cut *Eco*RI
10	pPC226 cut *Sal*I-*Eco*RI
11	pPC227 cut *Sal*I-*Eco*RI
12	pPC1-B cut *Sal*I-*Eco*RI
13	1Kb Ladder

4.4.7: Amplification Of The *nifH:gent*^{r} Reporter Gene From Plasmid pPC227.

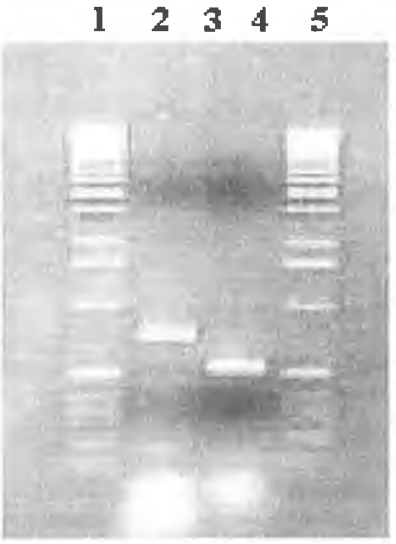
Having made the fusion between the *nifH* promoter and the *gent*^{r} gene the entire *nifH:gent*^{r} reporter gene was amplified from plasmid pPC227 as a *Hind*III fragment using the *nifH:gent*^{r}-F and *gent*-R primers described earlier. The PCR product was separated on a 1% agarose gel. A photo of the gel obtained and the conditions used for the PCR are shown in Figure 4.4.14. As can be seen from Figure 4.4.14 a single band corresponding to the desired *nifH:gent*^{r} reporter gene was obtained.

- **Figure 4.4.14: PCR Amplification Of The *nifH:gent*^{r} Reporter Gene From pPC227.**

Lane No.	Sample
1	1Kb Ladder
2	*nifH:gent*^{r} PCR product
3	*gent*^{r} PCR product
4	1Kb Ladder

PCR Conditions

Annealing temperature	58 °C
Annealing time	30 sec
Extension time	1 min

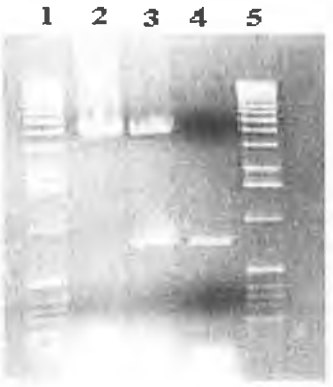


4.4.8: Cloning Of The *nifH:gent*^{r} PCR Product Into The TA Vector To Form Plasmid pPC8.

The *nifH:gent*^{r} reporter gene PCR product amplified from plasmid pPC227 was ligated with the TA pCR2.1 vector according to the TA protocol. The ligation was used to transform INVαF' cells and the transformation was plated on LB ampicillin Xgal plates. White colonies were screened by performing *Hind*III digests on plasmid DNA isolated from each clone. In this way the desired clone pPC8 was obtained and Figure 4.4.15 below shows the results of the verifying restriction analysis on this plasmid.

- **Figure 4.4.15: Restriction Analysis Of Plasmid pPC8.**

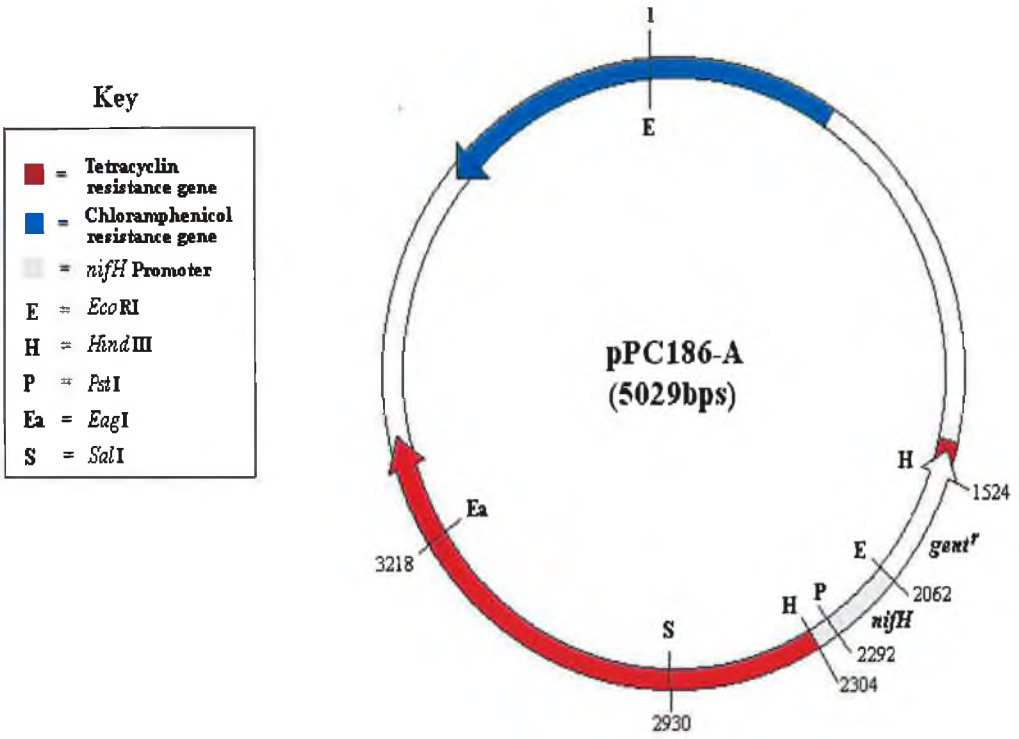
Lane No.	Sample
1	1Kb Ladder
2	pPC8 uncut
3	pPC8 cut *Hind*III
4	*nifH:gent*^{r} PCR product
5	1Kb Ladder

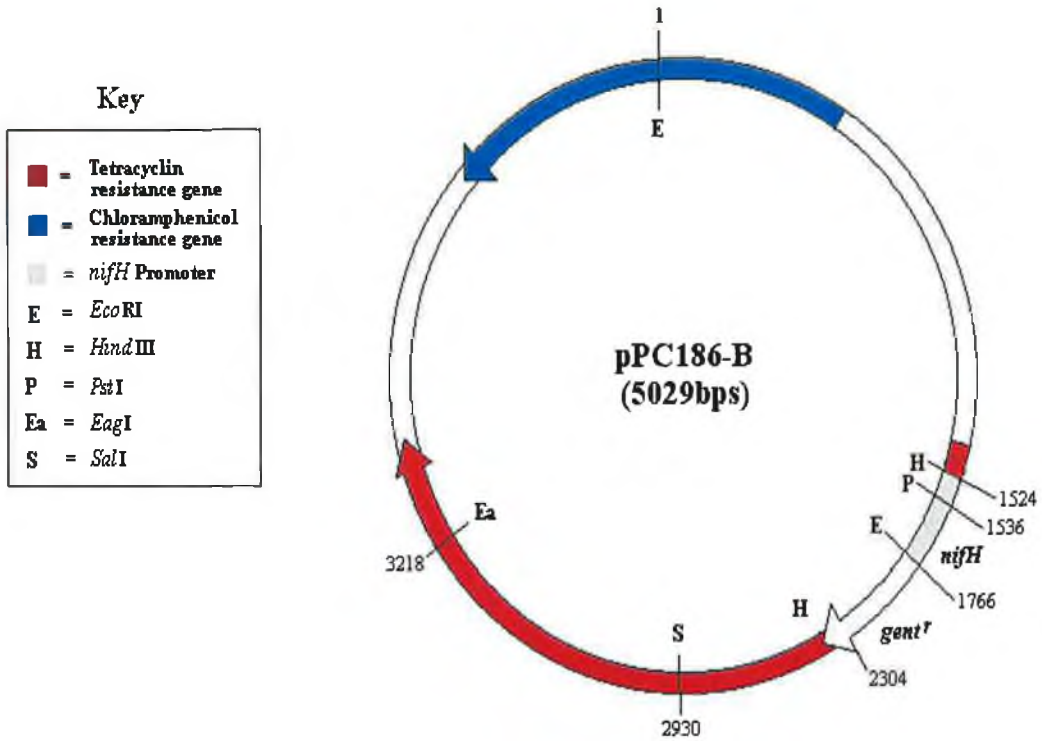


4.4.9: Sub-Cloning Of The *nifH:gent*^{r} Reporter Gene From pPC8 Onto Plasmid pPC184 To Form The Reporter Plasmid pPC186.

Plasmid pPC8 was digested with *Hind*III and the digested DNA was separated on a 0.7% agarose gel. The 780bp fragment corresponding to the *nifH:gent*^{r} reporter gene was purified from the gel using the gene clean procedure. pPC184 was also digested with *Hind*III and the linearised plasmid DNA was dephosphorylated with CIP to prevent recircularisation of the plasmid. The *Hind*III restricted dephosphorylated pPC184 DNA was ligated with the purified *nifH:gent*^{r} *Hind*III fragment and the transformation was plated out on LB chloramphenicol. Transformants were initially screened by performing *Hind*III digests on plasmid DNA isolated from each clone. Having identified several clones with the desired pPC186 plasmid, plasmid DNA from each of the clones was subjected to *Sal*I-*Pst*I double digestion to determine the orientation of the *nifH:gent*^{r} reporter gene in each case. Two plasmids were to be isolated, pPC186-A and pPC186-B, differing only in the orientation of the *nifH:gent*^{r} reporter gene. Restriction maps of the two desired plasmids can be seen in Figure 4.4.16. From the restriction maps it can be seen that a *Sal*I-*Pst*I digestion of pPC186-A would be expected to yield two bands, one of 638bps and one of 4391bps. *Sal*I-*Pst*I digestion of pPC186-B would be expected to also yield two bands, one of 1394bps and one of 3635bps. Thus by performing *Sal*I-*Pst*I double digests the two desired plasmids pPC186-A and pPC186-B were obtained. Figure 4.4.17 shows the results of restriction analysis on the two plasmids.

- **Figure 4.4.16: Restriction Maps Of The *nifH:gent*^r Reporter Plasmids pPC186-A And pPC186-B.**





- **Figure 4.4.17: Restriction Analysis Of The *nifH:gent*^{r} Reporter Plasmid pPC186-A And pPC186-B.**

Lane No.	Sample
1	1Kb Ladder
2	pPC186-A uncut
3	pPC186-B uncut
4	pPC186-A cut *Hind*III
5	pPC186-B cut *Hind*III
6	*nifH:gent*^{r} PCR product

4.5: Summary Of Chapter 4.

The end product of all of the cloning strategies discussed in this chapter was the construction of four reporter plasmids. Two *nifH:lacZα* plasmids, pPC185-A and pPC185-B (see Figure 4.3.13), differing only in the orientation of the *nifH:lacZα* reporter gene were constructed. Two *nifH:gent*^{r} plasmids, pPC186-A and pPC186-B (see Figure 4.4.16), again differing only in the orientation of the *nifH:gent*^{r} reporter gene, were also constructed. Each of these reporter plasmids would later have a Bait construct cloned into them in order to form the final Bait plasmids for the proposed P2H system. The construction of the Bait construct and the final Bait plasmids is discussed in Chapter 5.

Chapter 5

Construction Of The Bait Plasmids pPC187 And pPC188 For The Prokaryotic Two-Hybrid System.

5.1: Introduction.

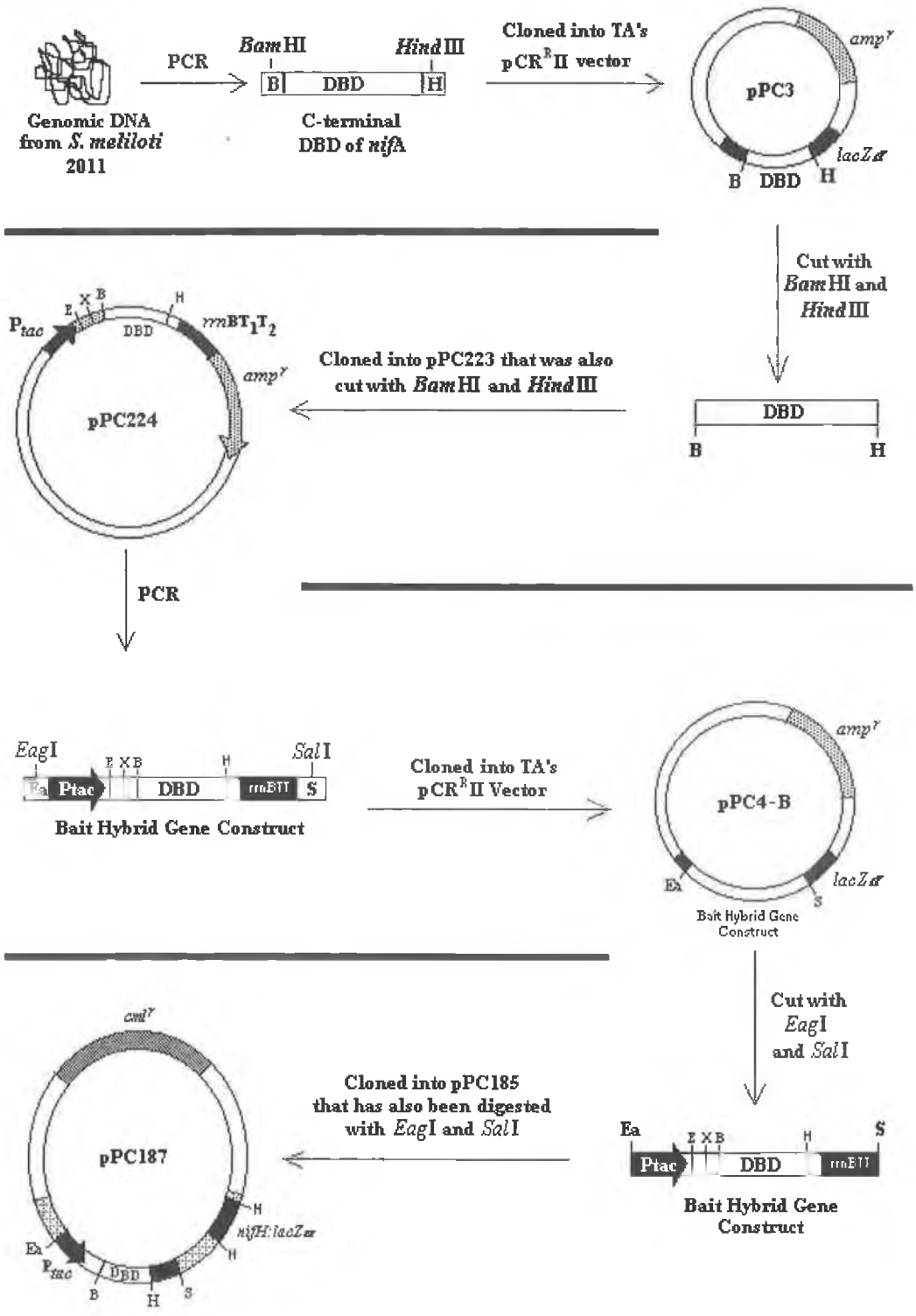
This chapter describes the construction of the Bait plasmids for the P2H system. Two Bait plasmids were to be constructed, pPC187 and pPC188. These plasmids would maintain the reporter genes for the P2H system and a Bait construct for the expression of hybrid proteins containing the C-terminal DNA binding domain (DBD) of *S. meliloti* 2011 NifA.

A Bait construct was to be cloned into the previously described pPC185 reporter plasmid maintaining the *nifH:lacZα* reporter gene (see chapter 4) in order to form the final Bait plasmid pPC187. Similarly the pPC186 reporter plasmid maintaining the *nifH:gent*^{r} reporter gene was to have a Bait construct cloned into it to form the second Bait plasmid pPC188. The strategy employed for the construction of the Bait construct and the results of the steps involved in the construction of the Bait plasmids is discussed below. Schematic representations of the strategies for the construction of the pPC187 and pPC188 Bait plasmids can be seen in Figure 5.2.1 and 5.2.2 respectively.

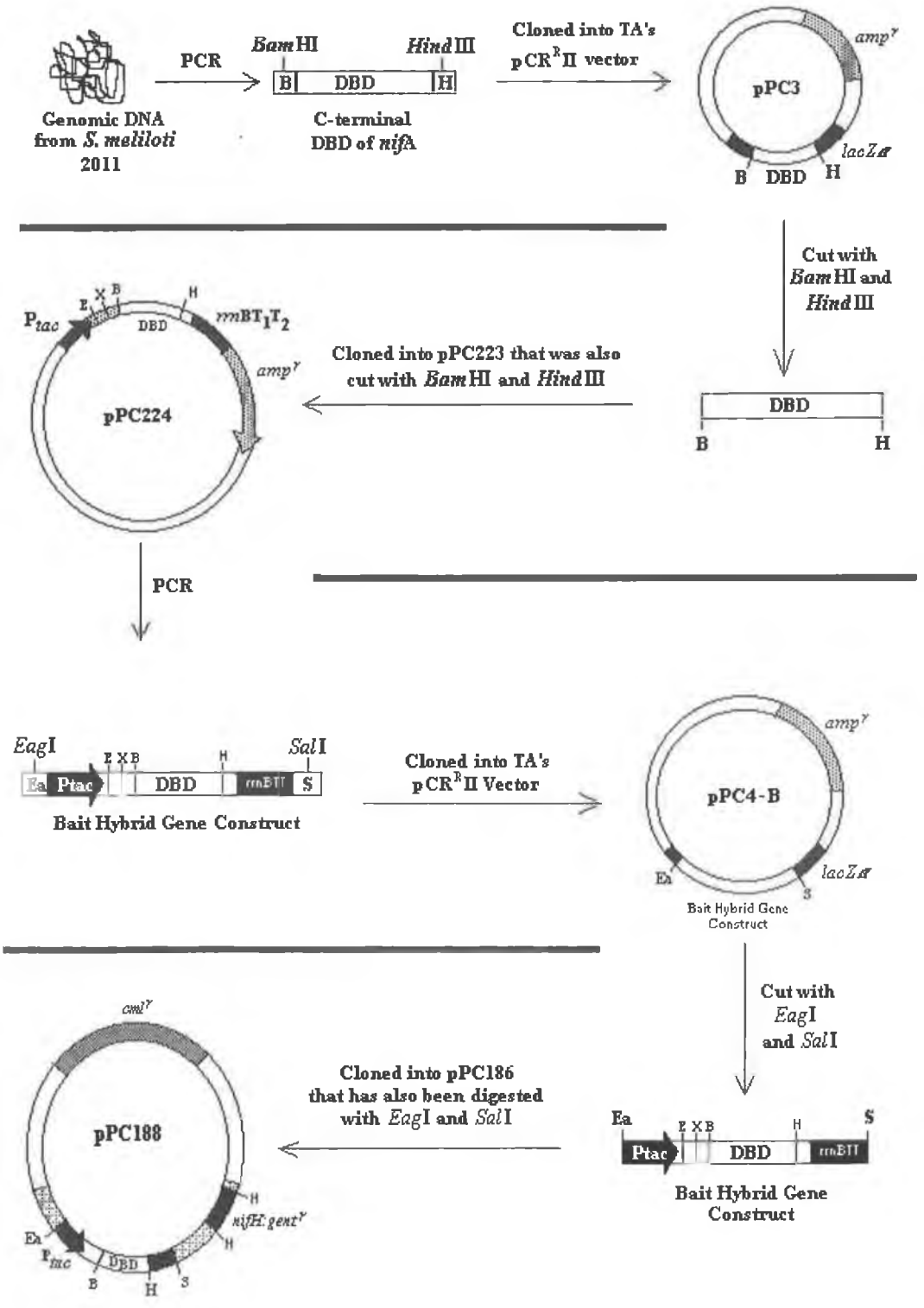
5.2: Construction Of The Bait Construct And The Final Bait Plasmids pPC187 And pPC188 - Overview Of Strategy.

In order to make the Bait construct for the expression of hybrid proteins containing the C-terminal DNA binding domain (DBD) of *S. meliloti* 2011 NifA, the sequence for the C-terminal DBD of the NifA protein was amplified from *S. meliloti* 2011 genomic DNA as a *Bam*HI-*Hind*III fragment using the C-term-F and C-term-R primers shown in Figure 5.2.3 below. The C-term-F forward primer was designed for an inter-domain region of the NifA protein separating the central and C-terminal DBD domains. The DNA sequence for this region possessed a natural *Bam*HI site and so the C-term-F primer did not have to be designed to incorporate a *Bam*HI site. The C-term-R primer was designed to incorporate a *Hind*III site into the NifA DBD PCR product. The regions of the *S. meliloti* 2011 *nifA* sequence for which these primers were designed are shown in Figure 5.2.4.

- **Figure 5.2.1: Strategy For The Construction Of The Bait Construct And Final Bait Plasmid pPC187 With The nifH:lacZα Reporter Gene.**



- **Figure 5.2.2: Strategy For The Construction Of The Bait Construct And The Final Bait Plasmid pPC188 With *nifH:gent*^{r} Reporter System.**

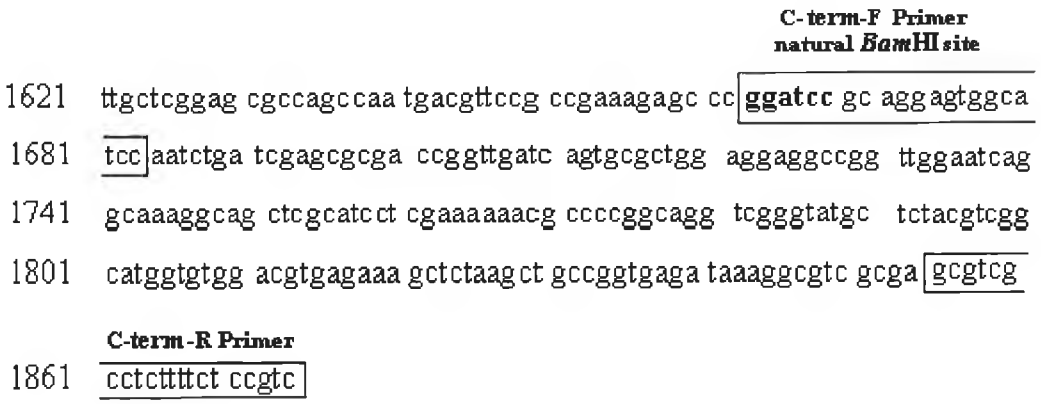


- **Figure 5.2.3: Primers Used For The Amplification Of the C-Terminal DBD Of NifA From *S. meliloti* 2011 Genomic DNA.**

C-term-F: **(GGA TCC)** - GCA GGA GTG GCA TCC
***Bam*HI site**

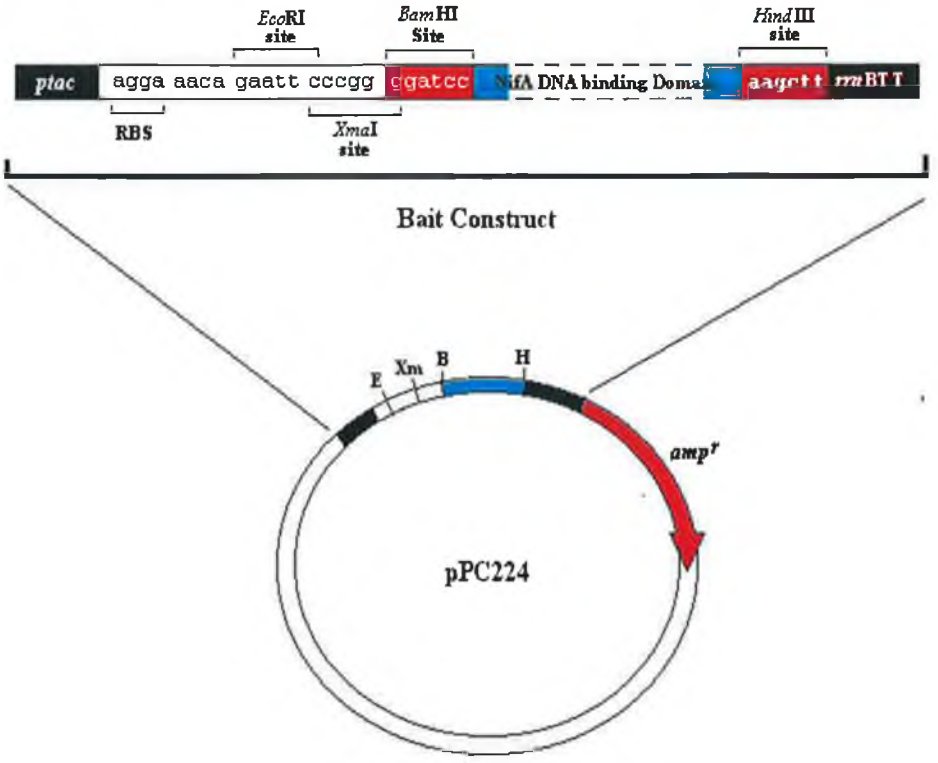
C-term-R: **(AAG CTT)** - GAC GGA GAA AAG AGG CGA CGC
***Hind*III site**

- **Figure 5.2.4: Primed Regions Of The *S. meliloti* 2011 *nifA* Sequence Used For The Amplification Of The C-Terminal DBD Of NifA.**



Once amplified the NifA DBD domain sequence was first cloned into the TA vector and then sub-cloned as a *Bam*HI-*Hind*III fragment into the MCS of plasmid pPC223 to form a plasmid called pPC224. Cloning the C-terminal domain fragment in this fashion essentially formed the Bait construct on plasmid pPC224. As depicted in Figure 5.2.5 below the Bait construct consisted of the strong *ptac* promoter followed C terminally by a MCS with unique *Xma*I and *Bam*HI sites that could be used for cloning genes of interest. C-terminally to the MCS was the sequence for the NifA DBD and then the *rrn*BT_1T_2 transcriptional termination sequence.

- **Figure 5.2.5: Schematic Of The Proposed Bait Construct On Plasmid pPC224**



Cloning genes of interest into the unique restriction sites of the MCS of this construct would fuse them N-terminally to the NifA DBD. The NifA DBD sequence would thus be located in a position mirroring its natural location within the *nifA* gene sequence. Transcription of the resulting hybrid genes would be initiated from the strong upstream *ptac* promoter and could be regulated using IPTG. Genes of interest would have to be cloned with their own ATG start codons to facilitate translation from the Bait construct RBS. They would also have to be cloned in the correct reading frame with the downstream NifA DBD sequence, and without their own stop codons to facilitate read-through translation into the NifA DBD sequence. Depending on the distance of the ATG start codon of the final cloned gene of interest from the Bait construct RBS, genes of interest might also have to be cloned with their own RBS in order to facilitate efficient translation of the hybrid proteins. It was not felt

that these stipulations would be a major hindrance to investigators using the final P2H system since the DNA sequence of the protein to be used as a Bait would most likely be known. The gene could therefore be amplified from a suitable source by PCR to meet the above mentioned requirements for cloning into the Bait construct and for efficient expression of hybrid proteins from the Bait construct.

Having made the proposed Bait construct the next step in the construction of the final Bait plasmids was to amplify the Bait construct from plasmid pPC224 as an *Eag*I-*Sal*I fragment using the Bait-F and Bait-R primers shown in Figure 5.2.6 below. The primed regions of the pPC224 sequence for which these primers were designed are shown in Figure 5.2.7 below.

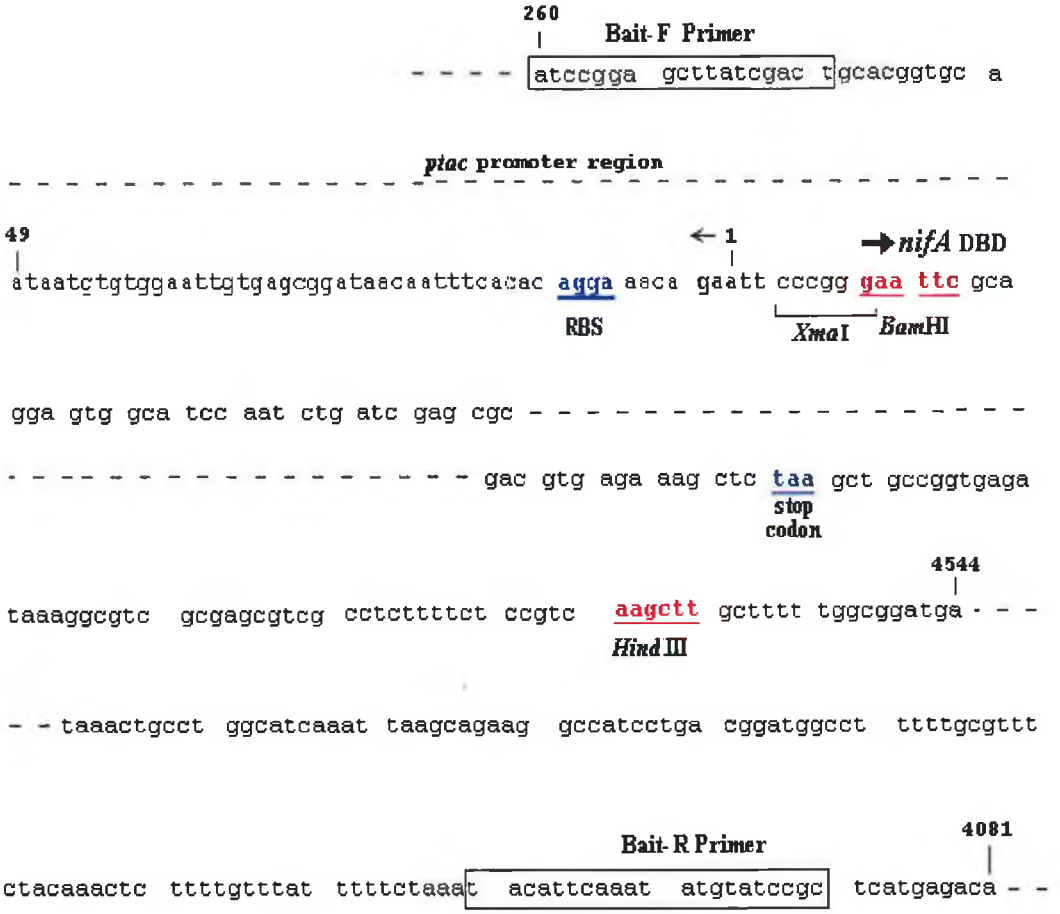
- **Figure 5.2.6: Primers Used For The Amplification Of The Bait Construct From Plasmid pPC224.**

Bait-F: **(CGG CCG)** - ATC CGG AGC TTA TCG ACT
***Eag*I site**

Bait-R: ACG - **(GTC GAC)** - GCG GAT ACA TAT TTG AAT GTA
***SalI* site**

Once amplified the Bait construct was first cloned into the TA vector. It was then sub-cloned into the tetracycline resistance gene of the pPC185-A and pPC185-B *nifH:lacZα* reporter plasmids described in chapter 4 as an *Eag*I-*Sal*I fragment to form the final Bait plasmids pPC187-A and pPC187-B respectively. Similarly the Bait construct was sub-cloned into the pPC186-A and pPC186-B *nifH:gent*^{r} reporter plasmids to form the final Bait plasmids pPC188-A and pPC188-B respectively. The orientation in which the Bait construct was cloned into each of the reporter plasmids placed the *rrn*BT_1T_2 transcriptional terminator sequence between the Bait construct and the reporter genes already on the reporter plasmids. This would prevent any read-through transcription into the reporter genes from the Bait construct and *vise versa*.

- **Figure 5.2.7: Primed Regions Of Plasmid pPC224 Used For The Amplification Of The Bait Construct.**



5.2.1: Amplification Of The NifA C-Terminal DNA Binding Domain Sequence From *S. meliloti* 2011.

The NifA C-terminal DBD sequence was amplified from *S. meliloti* 2011 genomic DNA as a *Bam*HI-*Hind*III fragment using the C-term-F and C-term-R primers described earlier. The PCR was separated on a 1% agarose gel. A photo of the gel obtained and the conditions used for the PCR can be seen in Figure 5.2.8 below. As can be seen from Figure 5.2.8 a single band of 219bp corresponding to the desired NifA DBD sequence was obtained.

- **Figure 5.2.8: PCR Amplification Of The NifA DBD sequence From *S. meliloti* 2011 Genomic DNA.**

Lane No.	Sample
1	NifA DBD PCR product
2	1Kb Ladder

PCR Conditions

Annealing temperature	64 °C
Annealing time	30 sec
Extension time	1 min

5.2.2: Cloning Of The NifA DBD PCR Product In The TA Vector To Form Plasmid pPC3.

The NifA DBD PCR product was ligated with the TA pCR2.1 vector according to the TA protocol. The resulting ligation was used to transform INVαF' cells and the transformation was plated out on LB ampicillin Xgal plates. White colonies were selected and screened by performing *Bam*HI-*Hind*III double digests on plasmid DNA isolated from each clone. In this way the desired plasmid, pPC3, was obtained and Figure 5.2.9 below shows the results of the verifying restriction analysis on this plasmid.

- **Figure 5.2.9: Restriction Analysis Of Plasmid pPC3.**

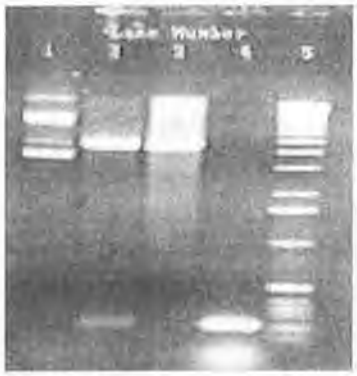
Lane No.	Sample.
1	pPC3 uncut
2	pPC3 cut *Bam*HI-*Hind*III
3	NifA DBD PCR product
4	1Kb Ladder

5.2.3: Sub-cloning Of The NifA DBD PCR Product From pPC3 Into pPC223 To Form Plasmid pPC224.

Plasmid pPC3 was digested with *Bam*HI-*Hind*III and the digested DNA was separated on a 1% agarose gel. The 219bp band corresponding to the NifA DBD *Bam*HI-*Hind*III fragment was purified from the gel using the gene clean procedure. Plasmid pPC223 was also digested with *Bam*HI-*Hind*III and the restricted DNA was run on a 0.7% agarose gel. The 4.5Kb band corresponding to the linearised pPC223 plasmid was purified from the gel using the gene clean procedure and ligated with the purified NifA DBD *Bam*HI-*Hind*III fragment. The ligation was used to transfrom DH5α and the transformation was plated on LB ampicillin plates. Colonies obtained were screened by performing *Bam*HI-*Hind*III double digests on plasmid DNA isolated from each of the clones. In this way the desired plasmid, pPC224, was obtained. Figure 5.2.10 below shows the results of the verifying restriction analysis on this plasmid.

- **Figure 5.2.10: Restriction Analysis Of Plasmid pPC224.**

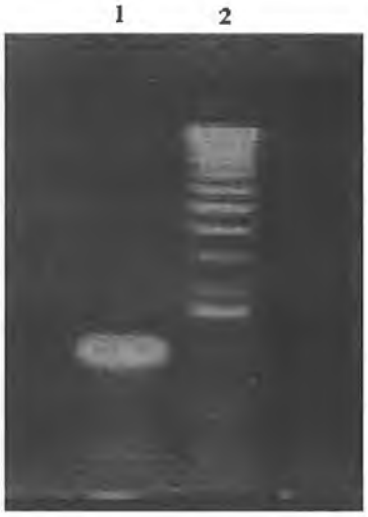
Lane No.	Sample.
1	pPC224 uncut
2	pPC224 cut *Bam*HI-*Hind*III
3	pPC223 cut *Bam*HI-*Hind*III
4	NifA DBD PCR product
5	1Kb Ladder



5.2.4: Amplifiction Of The Bait Construct From Plasmid pPC224.

The Bait construct was amplified from plasmid pPC224 as an *Eag*I-*Sal*I fragment using the Bait-F and Bait-R primers described earlier. The PCR was separated on a 0.7% agarose gel. A photo of the gel obtained and the conditions used for the PCR are shown in Figure 5.2.11 below. As can be seen from Figure 5.2.11 below a single 971bp band corresponding to the desired Bait construct was obtained.

- **Figure 5.2.11: PCR Amplification Of The Bait Construct From Plasmid pPC224.**



Lane No.	Sample
1	Bait PCR product
2	1Kb Ladder

PCR Conditions	
Annealing temperature	48 °C
Annealing time	30 sec
Extension time	2 mins

5.2.5: Cloning Of The Bait Construct PCR Product Into The TA Vector To Form Plasmid pPC4-B.

The Bait construct PCR product was ligated with the TA pCR2.1 vector according to the TA protocol and the ligation was then used to transform INVαF' cells. The transformation was plated out LB ampicillin Xgal plates. White colonies were picked and screened by performing *Eag*I-*Sal*I double digests on plasmid DNA isolated from each clone. In this way the desired plasmid pPC4-B was obtained. Figure 5.2.12 below shows the results of the verifying restriction analysis on this plasmid.

- **Figure 5.2.12: Restriction Analysis On Plasmid pPC4-B.**

Lane No.	Sample
1	1Kb Ladder
2	pPC4-B uncut
3	pPC4-B cut *Sal*I
4	pPC4-B cut *Eag*I-*Sal*I
5	1Kb Ladder

The final step in the construction of the final Bait plasmids involved the sub-cloning of the Bait construct from pPC4-B onto each of the reporter plasmids.

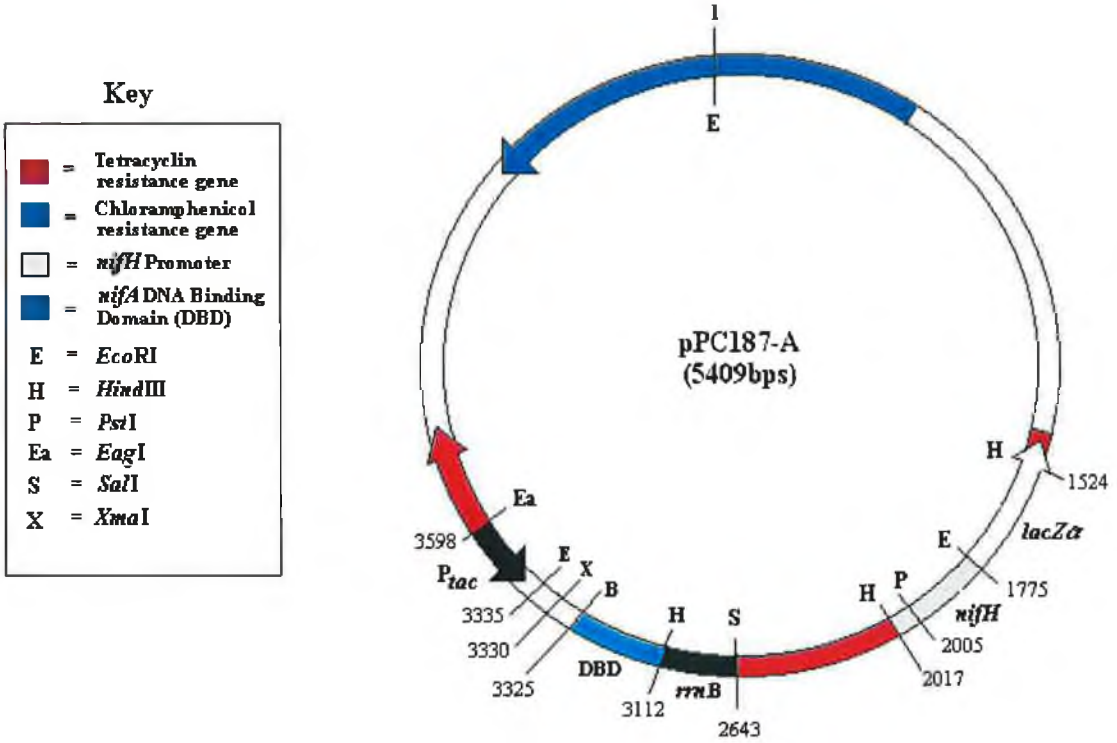
5.2.6: Sub-cloning Of The Bait Construct From Plasmid pPC4-B Into The pPC185 Plasmids To Form The Final *nifH:lacZα* pPC187 Bait Plasmids.

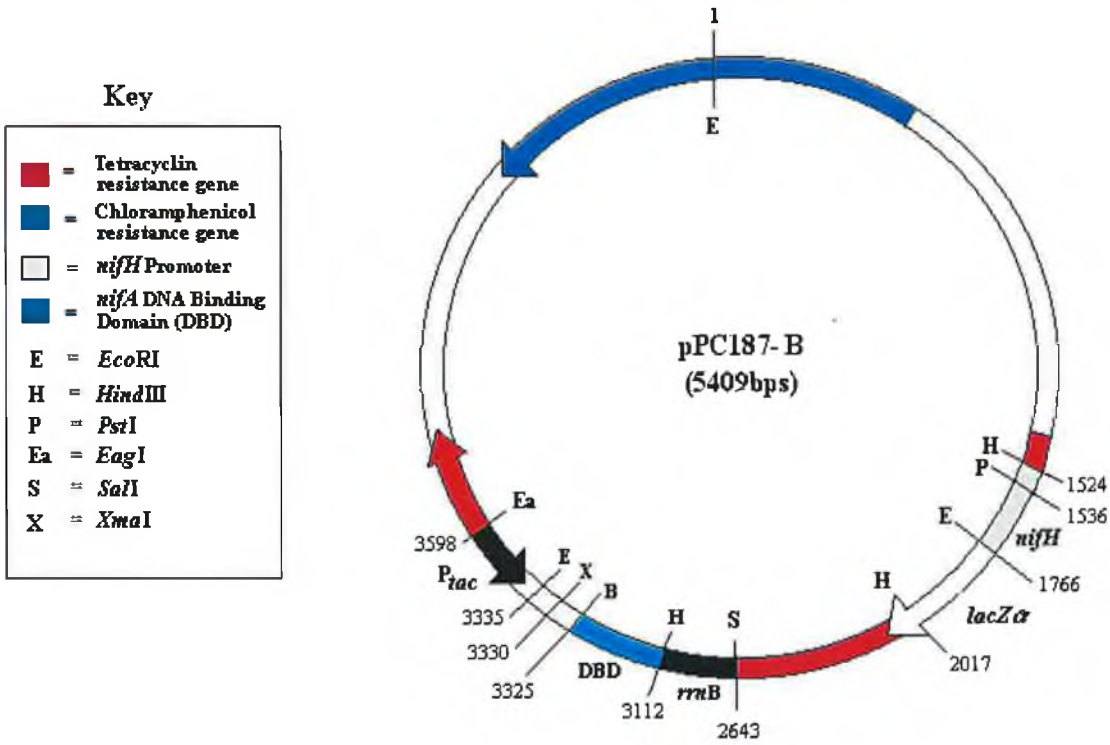
The final step in the construction of the Bait plasmids was to sub-clone the Bait construct from pPC4-B as an *Eag*I-*Sal*I fragment into the tetracycline resistance gene of each of the *nifH:lacZα* reporter plasmids pPC185-A and pPC185-B to form the Bait plasmids pPC187-A and pPC187-B respectively.

Plasmid pPC4-B was digested with *Eag*I-*Sal*I and the digested DNA separated on a 0.7% agarose gel. The 971bp band corresponding to the Bait construct *Eag*I-*Sal*I fragment was purified from the gel using the gene clean procedure. Plasmid pPC185-A was also digested with *Eag*I-*Sal*I and the digested DNA separated on a 0.7% agarose gel. The 4454bp band corresponding to the linearised plasmid was purified from the gel by the gene clean procedure and ligated with the purified *Eag*I-*Sal*I Bait construct fragment. The ligation was used to transform DH5α and the transformation was plated on LB chloramphenicol plates. Colonies were picked and screened initially by performing *Bam*HI digests. Only plasmid DNA containing the Bait construct with its unique *Bam*HI restriction site would be cleaved and thus by screening in this way the desired Bait plasmid pPC187-A was obtained. The above procedure was repeated for the pPC185-B reporter plasmid and resulted in the isolation of the final Bait plasmid pPC187-B. Detailed restriction maps of these two Bait plasmids can be seen in Figure 5.2.13 and the results of the verifying restriction analysis on plasmids pPC187-A and pPC187-B can be seen in Figures 5.2.14 and 5.2.15 respectively.

In order to verify the orientation of the *nifH:lacZα* reporter genes on these plasmids relative to the Bait construct *Sal*I-*Pst*I digests were performed and the results of these digests can be seen in Figure 5.2.16 below.

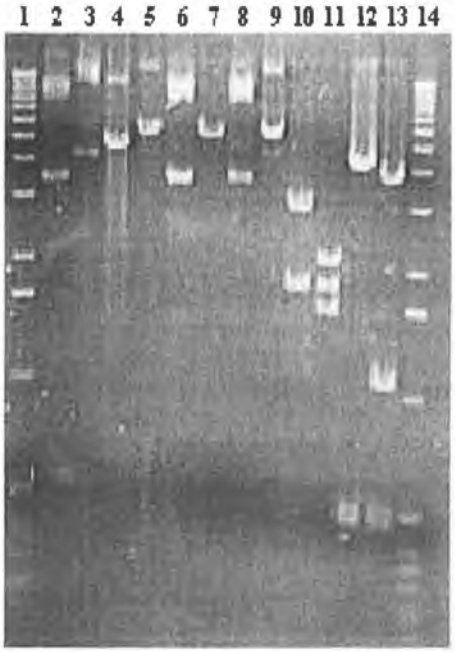
- **Figure 5.2.13: Restriction Maps Of The *nifH:lacZα* Bait Plasmids pPC187-A And pPC187-B.**





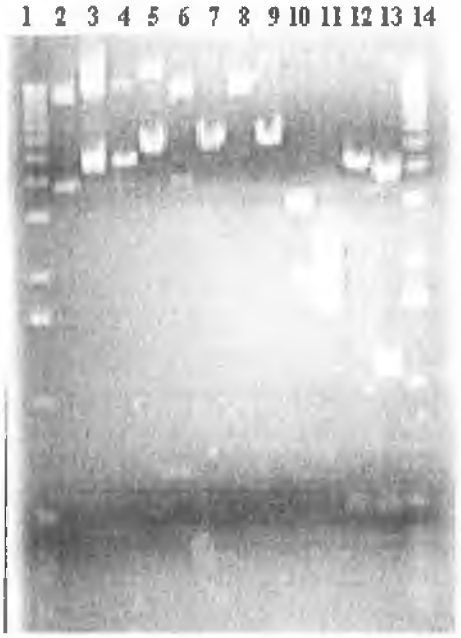
- **Figure 5.2.14: Restriction Analysis Of Bait Plasmid pPC187-A.**

Lane No.	Sample
1	1Kb Ladder
2	pPC185-A uncut
3	pPC187-A uncut
4	pPC185-A cut *Sal*I
5	pPC187-A cut *Sal*I
6	pPC185-A cut *Bam*HI
7	pPC187-A cut *Bam*HI
8	pPC185-A cut *Xma*I
9	pPC187-A cut *Xma*I
10	pPC185-A cut *Eco*RI
11	pPC187-A cut *Eco*RI
12	pPC185-A cut *Hind*III
13	pPC187-A cut *Hind*III
14	1Kb Ladder



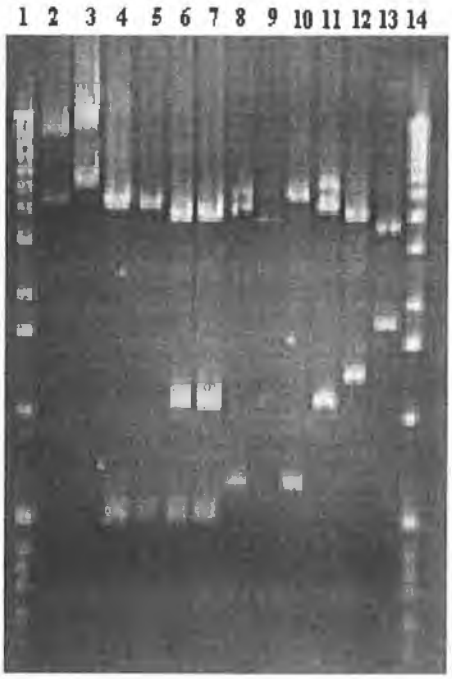
- **Figure 5.2.15: Restriction Analysis Of Bait Plasmid pPC187-B.**

Lane No.	Sample
1	1Kb Ladder
2	pPC185-B uncut
3	pPC187-B uncut
4	pPC185-B cut *Sal*I
5	pPC187-B cut *Sal*I
6	pPC185-B cut *Bam*HI
7	pPC187-B cut *Bam*HI
8	pPC185-B cut *Xma*I
9	pPC187-B cut *Xma*I
10	pPC185-B cut *Eco*RI
11	pPC187-B cut *Eco*RI
12	pPC185-B cut *Hind*III
13	pPC187-B cut *Hind*III
14	1Kb Ladder



- **Figure 5.2.16: Verification Of The Orientation Of The *nifH:lacZα* Reporter Gene On Plasmids pPC187-A And pPC187-B.**

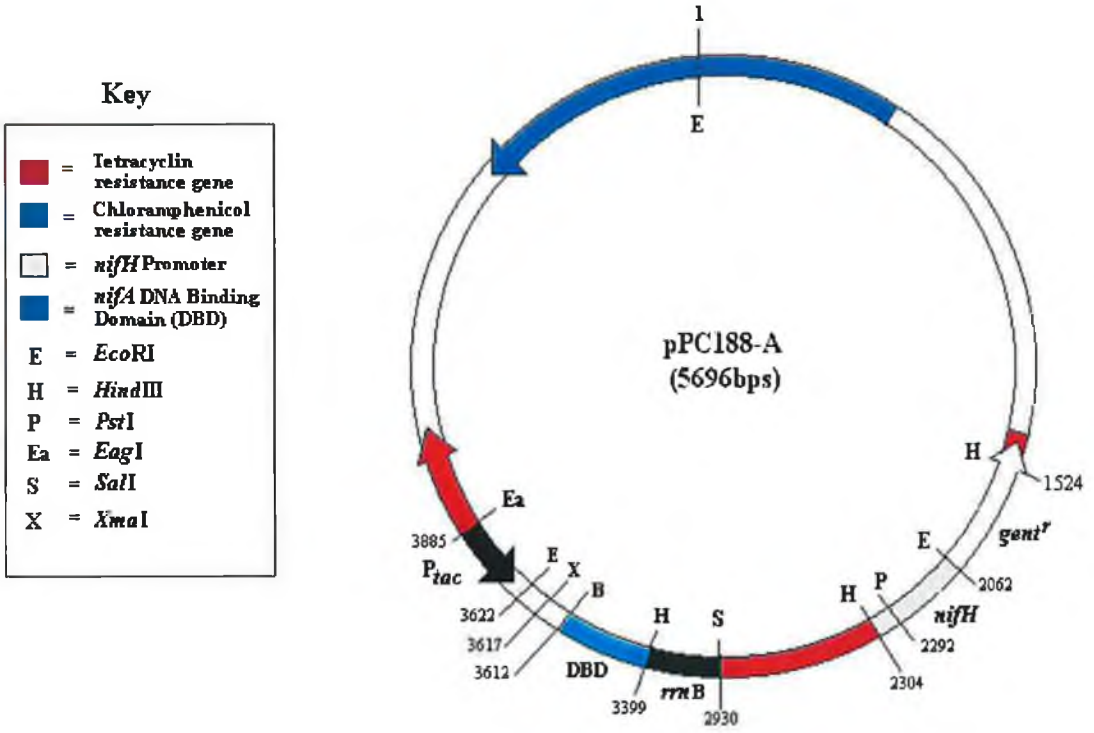
Lane No.	Sample
1	pPC185-A uncut
2	pPC187-A uncut
3	pPC185-A cut *Hind*III
4	pPC185-B cut *Hind*III
5	pPC187-A cut *Hind*III
6	pPC187-B cut *Hind*III
7	pPC185-A cut *Sal*I-*Pst*I
8	pPC185-B cut *Sal*I-*Pst*I
9	pPC187-A cut *Sal*I-*Pst*I
10	pPC187-B cut *Sal*I-*Pst*I
11	pPC187-A cut *Bam*HI-*Pst*I
12	pPC187-B cut *Bam*HI-*Pst*I

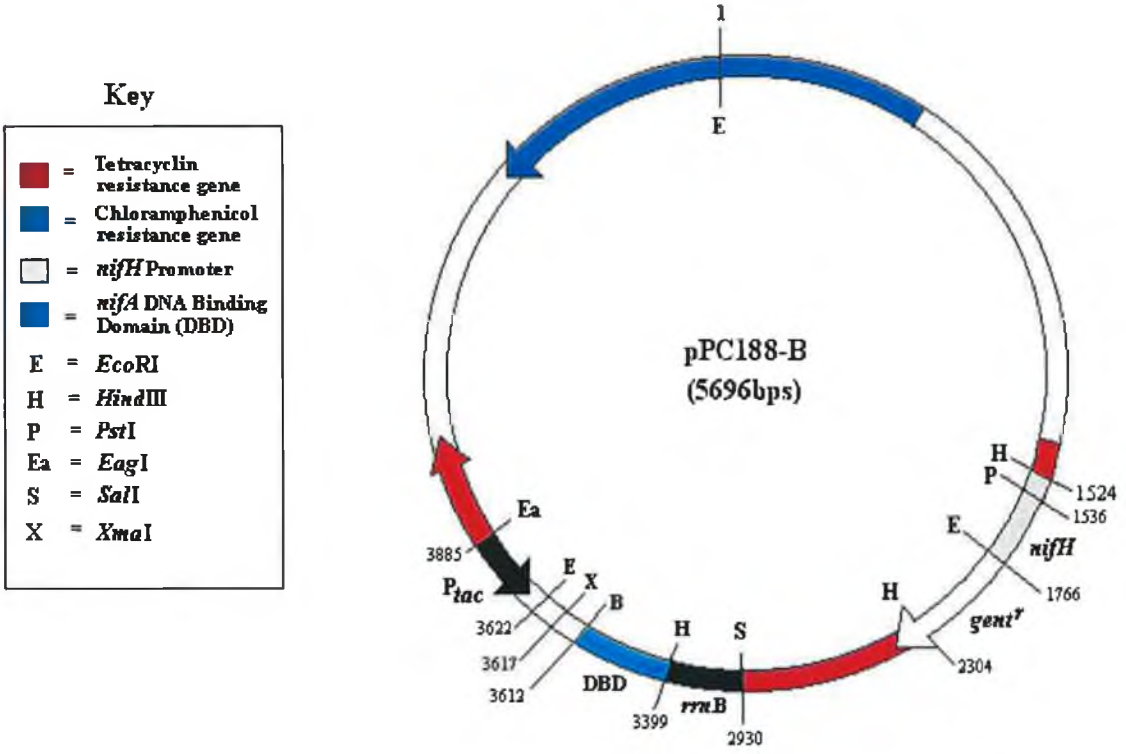


5.2.7: Sub-cloning Of The Bait Construct From Plasmid pPC4-B Into The pPC186 Plasmids To Form The Final *nifH:gent*' pPC188 Bait Plasmids.

The Bait construct was sub-cloned from plasmid pPC4-B onto each of the *nifH:gent*' reporter plasmids pPC186-A and pPC186-B in exactly the same way as that described above for the pPC185 *nifH:lacZα* reporter plasmids. This generated the other desired Bait plasmids pPC188-A and pPC188-B. Detailed restriction maps of each of these Bait plasmids can be seen in Figure 5.2.17. The results of verifying restriction analysis on plasmids pPC188-A and pPC188-B can be seen in Figures 5.2.18 and 5.2.19 respectively.

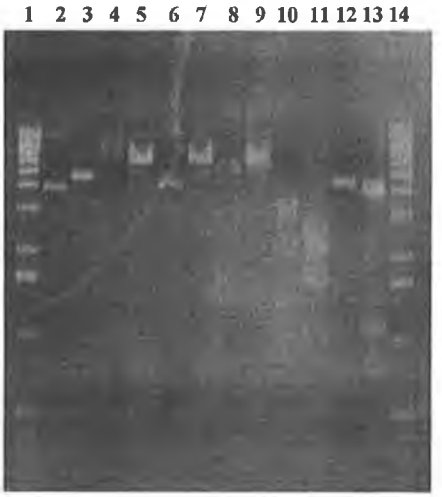
- **Figure 5.2.17: Restriction Maps Of The *nifH:gent*^r Bait Plasmids pPC188-A And pPC188-B.**





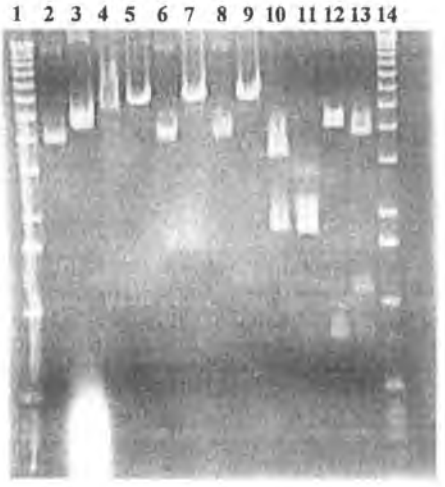
- **Figure 5.2.18: Restriction Analysis Of Bait Plasmid pPC188-A.**

Lane No.	Sample
1	1Kb Ladder
2	pPC186-A uncut
3	pPC188-A uncut
4	pPC186-A cut *Sal*I
5	pPC188-A cut *Sal*I
6	pPC186-A cut *Bam*HI
7	pPC188-A cut *Bam*HI
8	pPC186-A cut *Xma*I
9	pPC188-A cut *Xma*I
10	pPC186-A cut *Eco*RI
11	pPC188-A cut *Eco*RI
12	pPC186-A cut *Hind*III
13	pPC188-A cut *Hind*III
14	1Kb Ladder



- **Figure 5.2.19: Restriction Analysis Of Bait Plasmid pPC188-B.**

Lane No.	Sample
1	1Kb Ladder
2	pPC186-B uncut
3	pPC188-B uncut
4	pPC186-B cut *Sal*I
5	pPC188-B cut *Sal*I
6	pPC186-B cut *Bam*HI
7	pPC188-B cut *Bam*HI
8	pPC186-B cut *Xma*I
9	pPC188-B cut *Xma*I
10	pPC186-B cut *Eco*RI
11	pPC188-B cut *Eco*RI
12	pPC186-B cut *Hind*III
13	pPC188-B cut *Hind*III
14	1Kb Ladder



5.3: Summary Of Chapter 5.

The end product of the cloning strategies described in this chapter was the construction of four final Bait Plasmids for the proposed P2H system. The pPC187-A and pPC187-B Bait plasmids (see Figure 5.2.13) contain the *nifH:lacZα* reporter gene while the pPC188-A and pPC188-B Bait plasmids (see Figure 5.2.17) contained the *nifH:gent*^{r} reporter gene. The usefulness of each of these plasmids would be assessed in *E. coli* strains developed for the proposed P2H system described in Chapter 6.

Chapter 6

Development Of Suitable E. coli Strains For The Prokaryotic Two-Hybrid System And Optimization Of The P2H Assay.

6.1: Introduction.

In addition to the construction of the plasmid vectors for the NifA based P2H system it was also necessary to develop a suitable host strain of *E. coli* in which the plasmids would be used. The reason for this was that NtrC, another well studied member of the Enhancer Binding Protein (EBP) family, had been observed to activate the same promoters as NifA. Although this activation was very weak it would be necessary to eliminate it as it could contribute to false positives in the proposed P2H system. For this reason two strains of *E.* coli in which the gene for the NtrC protein (*glnG*) was deleted were obtained and developed for use with the proposed NifA based P2H system plasmids.

Like NifA, NtrC is part of the EBP family of proteins. It is a transcriptional activator responsible for regulating the expression of genes involved in nitrogen assimilation. It has three functional domains: an N-terminal domain which is involved in the regulation of the protein's activity by NtrB, a central transcriptional activation domain and a C-terminal DNA binding domain responsible for binding of the protein to UAS's found upstream of promoters regulated by the NtrC. Homologous NtrC proteins have been identified in a large number of bacterial strains. In *E.coli* the gene for NtrC, *glnG*, is part of the complex *glnALG* operon and expression of this operon has been extensively studied. The *glnL* gene product is NtrB which regulates the activity of NtrC through phosphorylation. NtrC is initially expressed as an inactive protein and requires phosphorylation of its N-terminal domain by NtrB in order to be activated. The *glnA* gene of the *glnALG* operon codes for glutamine synthase.

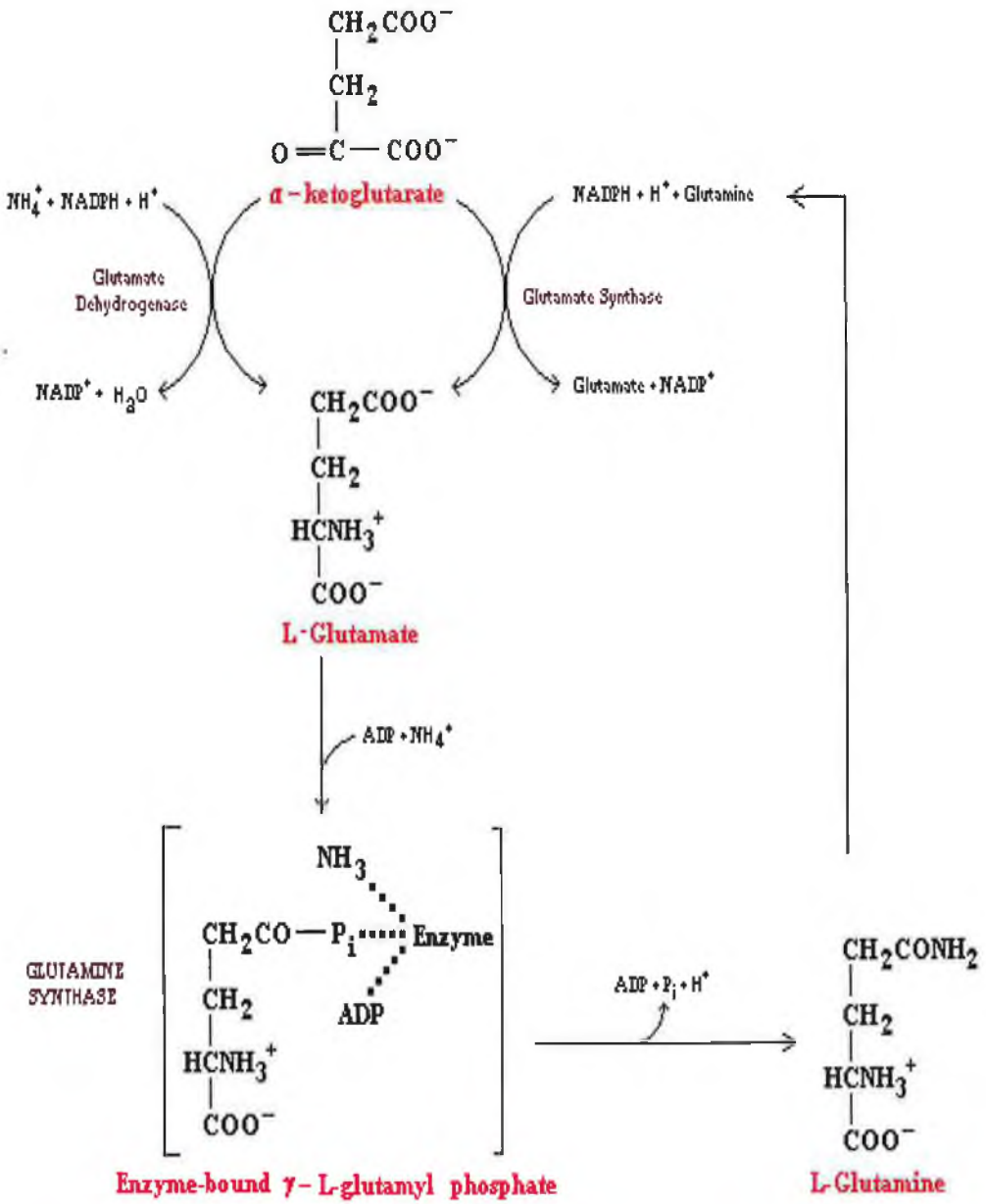
As already mentioned above two *glnG* mutant *E. coli* strains were obtained and developed for use with the proposed NifA based prokaryotic system. One of these strains, *YMC11*, has the entire *gln* operon deleted (*Δgln2000*). Because of the *glnA* deletion, growth of this *E.coli* strain is severely retarded unless media is supplemented with 0.2% L-glutamine. This is because glutamine synthase plays a key role in amino acid biosynthesis in the cell.

A common element in all amino acids is the α-amino group. All α-amino groups are derived from ammonia by way of the amino groups of L-glutamate. The simplest route to glutamate and therefore to all amino group formation is the direct reductive amination of α-ketoglutarate with ammonia, a reaction that is catalysed by glutamate dehydrogenase (see Figure 6.1.1). This means of making glutamate is only exhibited by bacteria when grown in a medium containing an ammonium salt as the sole nitrogen source. The prevelent reaction by which glutamate is formed is the amination of α-ketoglutarate by glutamate synthase, a reductive reaction in which the source of the amino group is not ammonia itself but an "activated" form, the amide group of L-glutamine. This reaction generates two molecules of glutamate, one of which must be amidated again to form L-glutamine in order to sustain the continued synthesis of amino groups. Glutamine synthase is responsible for the amidation of glutamate to form L-glutamine (see Figure 6.1.1) and is therefore the key enzyme in the conversion of NH_4^+ to organic nitrogen. Although this mode of amino group formation requires a high energy phosphate there is considerable advantage since amide group formation can proceed at a much lower concentration of NH_4^+ than can amination by glutamate dehydrogenase. The K_m for NH_4^+ of the *E. coli* glutamate dehydrogenase is 1.1mM, whereas that of *E. coli* glutamine synthase is about 0.2mM. NH_4^+ is not often present in high concentrations in the soil due to microbial oxidation of NH_4^+ and regulatory mechanisms that control both NO_3^- reduction and nitrogen fixation. The formation of amino groups in plants and microorganisms usually therefore occurs in a low NH_4^+ environment and is achieved by amination of α-ketoglutarate through the activities of glutamate synthase and glutamine synthase.

From the above description of the role of glutamine synthase in L-glutamate formation and thus in the biosynthesis of all amino acids, it is clear that deletion of the gene for this enzyme, *glnA*, would seriously retard the growth of a bacterial cell. This was evident from the extremely slow growth rate exhibited by the *YMC11* strain of *E. coli* on LB agar plates unless media was supplemented with L-glutamine. Supplementing media with L-glutamine effectively bypasses the need for the

reaction catalyzed by glutamine synthase in the formation of L-glutamate and thus overcomes the growth retardation.

- **Figure 6.1.1: The Conversion Of Ammonia Into The α-Amino Group Of Glutamate And Into The Amide Group Of Glutamine**



In addition to the *YMC11* strain described above, another strain of *E. coli* called *ET6016* was obtained. This strain had an undefined deletion of the *glnL* and/or *glnG* gene. Although the deletion was undefined, a deletion of either the *glnL* or *glnG* genes would ensure that active NtrC was absent. Because the *glnA* gene was still present growth of this strain did not require the addition of L-glutamine to media.

Both of the above strains were selected for development as host strains for the NifA based P2H system by virtue of their deletions in the *glnG* gene. Unfortunately, these strains also had the entire *lac* operon deleted (ΔU169) and therefore did not posses the correct *lac* genetic background for α-complementation of β-galactosidase. This genetic background was required for the P2H system's *nifH:lacZα* reporter system (described in chapter 4) to work. Most strains designed for α-complementation of β-galactosidase, like *DH5α,* initially have the entire *lac* operon deleted and the correct complementing genetic background, *lacZ*ΔM15, is re-introduced. In the case of DH5α this complementing background was re-introduced using the transducing phage ϕ80 while in the case of JM103 the correct complementing background was re-introduced on an F'-factor by conjugation. Once the complementing genetic background is re-introduced into strains the strains are stabilized by the introduction of mutations in the *recA* gene. The RecA protein plays a role in homologous recombination and in the SOS DNA repair system. It is also required for induction of lambdoid prophage, like ϕ80, by ultraviolet light. In cases were the complementing background was introduced on an F'-factor, mutations are often introduced in the *traD* gene of the F' factor. This mutation significantly reduces the frequency of conjugation of the F' factor.

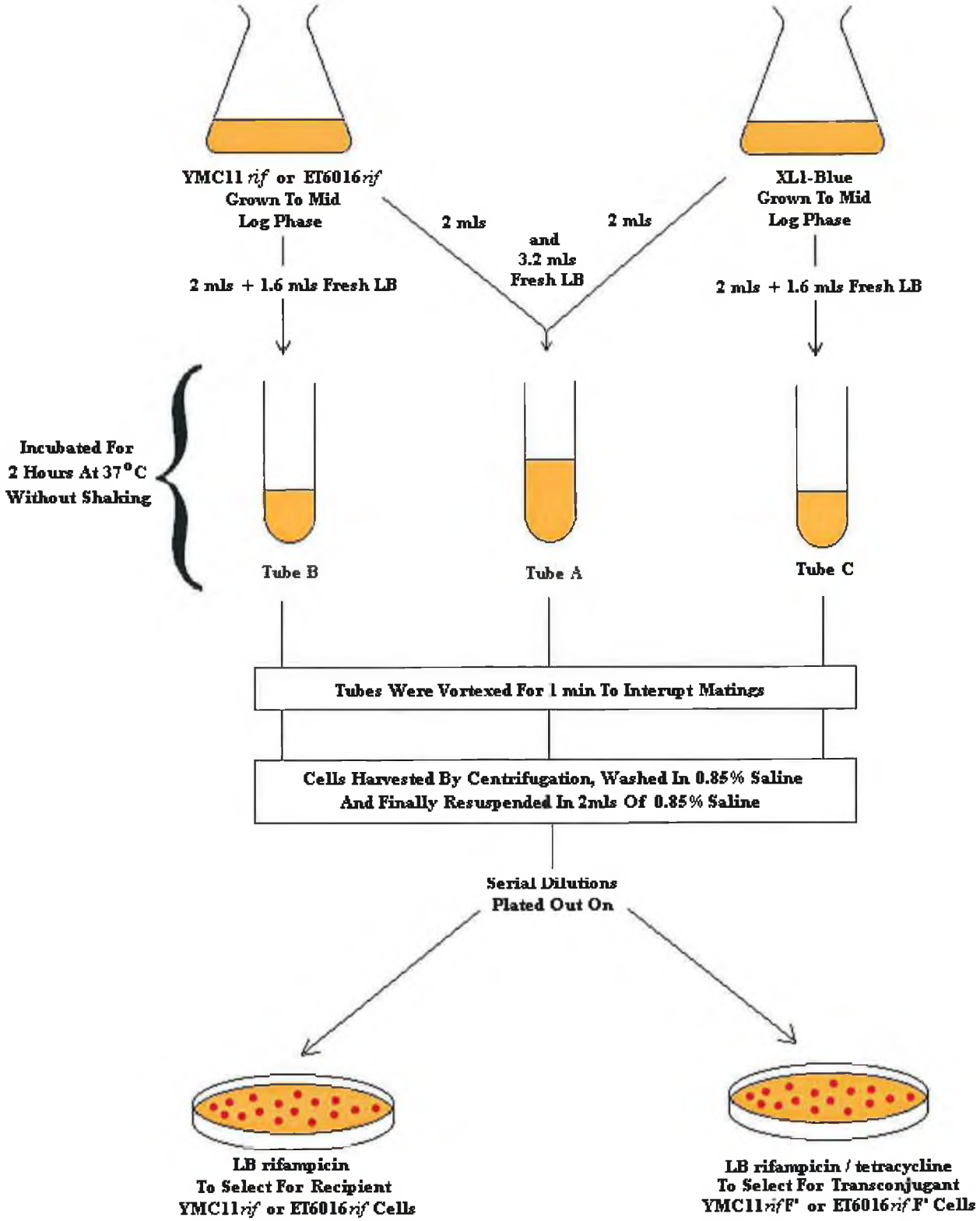
Due to the lack of availability of a suitable transducing phage bearing the required *lac* complementing genetic background it was decided that the required *lac* genetic background would be introduced into the two *glnG* mutant *E.coli* strains, YMC11 and ET6016, on an F' factor by conjugation.

6.2:Introduction Of The *lacZα* Complementing Genetic Background Into The *E. coli* Strains YMC11 And ET6016.

The required *lacZα* complementing genetic background was to be introduced into the two selected *glnG* mutant *E. coli* strains YMC11 and ET6016 on an F' Factor by conjugation with the *E. coli* strain XL1-Blue. This strain of *E. coli* possesses an F' Factor with the required *lacZα* complementing genetic background, *lacZ*ΔM15. It also possesses the $lacI^q$ allele that expresses 10 times as much *lacI* repressor as the wild type allele. This was also desirable since both the Prey and the Bait plasmids constructed for the NifA based P2H system expressed hybrid proteins from a *ptac* promoter. The LacI repressor represses expression from this promoter and this repression can be relieved through the use of IPTG. The presence of the $lacI^q$ allele would ensure that LacI levels in the cell would be sufficiently high to repress expression from all copies of the *ptac* promoter present on the Prey and Bait plasmids and would thus allow expression of hybrid proteins to be regulated using IPTG. In addition to this the F' Factor did not possess a *traD* mutation and thus efficient conjugal transfer of the F' Factor into the two *glnG* mutant *E. coli* strains would be expected to occur. The F' Factor also possessed a Tn10 transposon bearing a tetracycline resistance gene which would provide a useful means of selecting for transconjugants when performing conjugations (see Figure 6.2.1).

Rifampicin resistant mutant strains of each of the *glnG E. coli* mutant strains, YMC11 and ET6016, were generated in order to facilitate selection of transconjugants in conjugations performed with XL1-Blue (see Figure 6.2.1). These rifampicin resistant mutants were generated by plating 200µl of overnight cultures of each of the strains on LB agar to which rifampicin had been added (LBrif) to a concentration of 100µg/ml. Plates were incubated at 37 °C for 24 hours. Approximately 15 to 20 colonies were obtained on each plate. Colonies were picked from each plate and purified by streaking on fresh LBrif agar. The rifampicin resistant YMC11 and ET6016 strains obtained were called YMC11*rif* and ET6016*rif* respectively. These rifampicin resistant mutants were used in conjugations with XL1-Blue.

- **Figure 6.2.1: Schematic Representation Of YMC11*rif* or ET6016*rif* Conjugation With XL1-Blue.**



6.2.1: Conjugation Of YMC11*rif* With XL1-Blue.

Figure 6.1.1 shows a schematic representation of the manner in which the conjugation between YMC11*rif* and XL1-Blue was performed. Culture of YMC11*rif* and XL1-Blue were grown to an OD_{600nm} of approximately 0.4 corresponding to mid log phase. Three tubes, A, B and C were then set up. Tube A contained 2mls of the YMC11*rif* culture, 2mls of the XL1-Blue culture and 3.2mls of fresh LB media. This was the conjugation mixture. Tube B contained 2mls of the YMC11*rif* culture and 1.6mls of fresh LB media while Tube C contained 2mls of the XL1-Blue culture and 1.6mls of fresh LB media. Tubes B and C were to serve as control tubes. The tubes were incubated at 37 ^{o}C for 2 hours without shaking to allow mating to occur and were then vortexed to interupt mating. Cells were harvested by centrifugation at 3000rpm for 5 mins, washed twice with 5mls of 0.85% sterile saline solution and each of the final pellets was then resuspended in 2mls of 0.85% sterile saline solution.

Serial dilutions of the Tube A conjugation were prepared to a dilution factor of 10^{-6}. A 200µl sample from each dilution was plated in duplicate onto LB agar to which rifampicin and tetracycline had been added (LBrif/tet). Only YMC11*rif* cells that had received an F' Factor with its tetracycline resistance gene from XL1-Blue, YMC11*rif*F' transconjugants, could grow on this media. The number of transconjugants on each of the plates was counted and these counts were used to calculate the total number of transconjugants in the original mating mixture. The counts obtained and the calculated transconjugant total count can be seen in Table 6.2.1 below.

A 200µl sample from each dilution was also plated out in duplicate on LBrif media. Only recipient YMC11*rif* cells and YMC11*rif*F' transconjugants could grow on this media. The number of colonies on each of the plates was counted and these counts were used to calculate a recipient count for the mating mixture. The counts obtained and the calculated total recipient count can be seen in Table 6.2.2 below.

Table 6.2.1:YMC11*rif*F' Transconjugant Counts Obtained On LBrif/tet Media.

Dilution Factor	Colony Counts		Average Count
200µl Neat	TNTC	TNTC	TNTC
200µl 10^{-1}	TNTC	TNTC	TNTC
200µl 10^{-2}	401	370	386
200µl 10^{-3}	35	43	39
200µl 10^{-4}	1	2	2
200µl 10^{-5}	-	-	-
200µl 10^{-6}	-	-	-

Total Number Of YMC11*rif*F'

Transconjugants In 200µl Neat Sample $= 3.88 \times 10^4$

Total Number Of Transconjugants

In Total 2mls Of The Mating Mixture $= (3.88 \times 10^4) \times 10$

$= 3.88 \times 10^5$ **$\sim 3.88 \times 10^5$**

Table 6.2.2: YMC11*rif* Recipient Counts Obtained On LBrif Media.

Dilution Factor	Colony Counts		Average Count
200µl Neat	TNTC	TNTC	TNTC
200µl 10^{-1}	TNTC	TNTC	TNTC
200µl 10^{-2}	TNTC	TNTC	TNTC
200µl 10^{-3}	TNTC	TNTC	TNTC
200µl 10^{-4}	TNTC	TNTC	TNTC
200µl 10^{-5}	492	522	507
200µl 10^{-6}	47	59	53

Total Number Of YMC11*rif*

Recipients In 200µl Neat Sample $= 5.18 \times 10^7$

Total Number Of Recipients In

Total 2mls Of The Mating Mixture $= (5.18 \times 10^7) \times 10$ **$= 5.18 \times 10^8$**

Using the calculated recipient count and transconjugant count the transfer efficiency (Tf) was calculated as:

$$Tf = \frac{\text{number of transconjugants}}{\text{number of recipients}}$$

$$Tf = \frac{3.88 \times 10^{5}}{5.18 \times 10^{8}}$$

$$Tf = \mathbf{7.5 \times 10^{-4}}$$

Serial dilutions of the Tube C mixture containing only XL1-Blue were prepared to a dilution factor of 10^{-3}. A 200µl neat sample of the XL1-Blue cells and a 200µl sample of each of the prepared dilutions of the XL1-Blue cells was plated on LBrif media and LBrif/tet media. These plates were to serve as negative controls. As expected, a small number of colonies, 6 on the LBrif and 8 on the LBrif/tet media, were obtained when the neat sample of XL1-Blue cells from the Tube C mixture was plated. These colonies were XL1-Blue rifampicin resistant mutants generated in the same way the YMC11*rif* strain had been generated earlier. No colonies were obtained when the other dilutions were plated on the LBrif or LBrif/tet media. It could therefore be deduced that all of the colonies obtained when 200µl of the 10^{-2}, 10^{-3} and 10^{-4} dilutions of the Tube A conjugation mixture were plated on the LBrif/tet media were transconjugant YMC11*rif*F' cells and not spontaneously generated XL1-Blue rifampicin resistant mutants. A 200µl sample of the Tube B mixture containing only YMC11*rif* cells was also plated on LBrif/tet media as a negative control and no growth was obtained.

One of the YMC11*rif*F' transconjugant colonies obtained when the 10^{-4} dilution of the Tube A conjugation mixture was plated on LBrif/tet media was picked off and purified by streaking on fresh LBrif/tet media. In order to verify that the desired

lacZα complementing genetic background had been introduced, competent YMC11*rif*F' cells were prepared and transformed with pUC19. The transformation was plated out on LBrif/tet Xgal media. Blue colonies were obtained thus verifying that the correct *lacZα* complementing genetic background had been successfully introduced.

6.2.2: Conjugation Of ET6016*rif* With XL1-Blue.

The conjugation between ET6016*rif* and XL1-Blue was carried out in exactly the same way as that described for the conjugation between YMC11*rif* and XL1-Blue. Cultures of ET6016*rif* and XL1-Blue were grown to an OD_{600nm} of approximately 0.4 corresponding to mid log phase. Three tubes, A, B and C were then set up. Tube A contained 2mls of the ET6016*rif* culture, 2mls of the XL1-Blue culture and 3.2mls of fresh LB media. This was the conjugation mixture. Tube B contained 2mls of the ET6016*rif* culture and 1.6mls of fresh LB media while Tube C contained 2mls of the XL1-Blue culture and 1.6mls of fresh LB media. As before Tubes B and C were control tubes. The tubes were incubated at 37 ^{o}C for 2 hours without shaking to allow mating to occur and then vortexed to interupt mating. Cells were harvested by centrifugation at 3000rpm for 5 mins, washed twice with 5mls of 0.85% sterile saline solution and each of the final pellets was then resuspended in 2mls of 0.85% sterile saline solution.

Serial dilutions of the Tube A conjugation were prepared to a dilution factor of 10^{-6}. A 200µl sample from each dilution was plated in duplicate onto LBrif/tet media. Only ET6016*rif* cells that had received an F' Factor with its tetracycline resistance gene from XL1-Blue, ET6016*rif*F' transconjugants, could grow on this media. The number of transconjugants on each of the plates was counted and these counts were used to calculate the total number of transconjugants in the original mating mixture. The counts obtained and the calculated transconjugant total count can be seen in Table 6.2.3 below.

Table 6.2.3: ET6016*rif*F' Transconjugant Counts Obtained On LBrif/tet Media.

Dilution Factor	Colony Counts		Average Count
200µl Neat	TNTC	TNTC	TNTC
200µl 10^{-1}	TNTC	TNTC	TNTC
200µl 10^{-2}	TNTC	TNTC	TNTC
200µl 10^{-3}	TNTC	TNTC	TNTC
200µl 10^{-4}	352	381	367
200µl 10^{-5}	36	32	34
200µl 10^{-6}	3	2	3

Total Number Of 6016*rif*F'

Transconjugants In 200µl Neat Sample = **3.36×10^6**

Total Number Of Transconjugants

In Total 2mls Of The Mating Mixture = $(3.36 \times 10^6) \times 10$

= **3.36×10^7**

Table 6.2.4: ET6016*rif* Recipient Counts Obtained On LBrif Media.

Dilution Factor	Colony Counts		Average Count
200µl Neat	TNTC	TNTC	TNTC
200µl 10^{-1}	TNTC	TNTC	TNTC
200µl 10^{-2}	TNTC	TNTC	TNTC
200µl 10^{-3}	TNTC	TNTC	TNTC
200µl 10^{-4}	TNTC	TNTC	TNTC
200µl 10^{-5}	TNTC	TNTC	TNTC
200µl 10^{-6}	446	471	459

Total Number Of ET6016*rif*

Recipients In 200µl Neat Sample = **4.59×10^8**

Total Number Of Recipients In
Total 2mls Of The Mating Mixture $= (4.59 \times 10^8) \times 10$

$$= \mathbf{4.59 \times 10^9}$$

A 200µl sample from each dilution was also plated out in duplicate on LBrif media. Only recipient ET6016*rif* cells and ET6016*rif*F' transconjugants could grow on this media. The number of colonies on each of the plates was counted and these counts were used to calculate a recipient count for the Tube A mating mixture. The counts obtained and the calculated total recipient count can be seen in Table 6.2.4 below. Using the calculated total recipient count and transconjugant count the transfer efficiency (Tf) was calculated as:

$$Tf = \frac{\text{number of transconjugants}}{\text{number of recipients}}$$

$$Tf = \frac{3.36 \times 10^7}{4.59 \times 10^9}$$

$$Tf = \mathbf{7.3 \times 10^{-3}}$$

Serial dilutions of the Tube C mixture containing only XL1-Blue were prepared to a dilution factor of 10^{-3}. A 200µl neat sample of the XL1-Blue cells and a 200µl sample of each of the prepared dilutions of the XL1-Blue cells was plated on LBrif media and LBrif/tet media. As for the YMC11*rif* conjugation, these plates were to serve as negative controls. Again a small number of XL1-Blue rifampicin resistant mutant colonies, 6 on the LBrif and 7 on the LBrif/tet media, were obtained when the neat sample of XL1-Blue cells from the Tube C mixture was plated. No colonies were obtained when the other dilutions were plated on the LBrif or LBrif/tet media.

It could therefore be deduced that all of the colonies obtained when 200µl of the 10^{-5} and 10^{-6} dilutions of the Tube A conjugation mixture were plated on the LBrif/tet media were transconjugant ET6016*rif*F' cells and not spontaneously generated XL1-Blue rifampicin resistant mutants. A 200µl sample of the Tube B mixture containing only ET6016*rif* cells was also plated on LBrif/tet media as a negative control and no growth was obtained.

One of the ET6016*rif*F' transconjugant colonies obtained when the 10^{-5} dilution of the Tube A conjugation mixture was plated on LBrif/tet media was picked off and purified by streaking on fresh LBrif/tet media. As before the presence of the desired *lacZα* complementing genetic background was confirmed by preparing competent ET6016*rif*F' cells and transforming them with pUC19. The transformation was plated out on LBrif/tet Xgal media. Blue colonies were obtained thus verifying that the correct *lacZα* complementing genetic background had been successfully introduced.

6.3: Optimization Of The P2H System.

Having constructed two *E. coli* host strains, ET6016*rif*F' and YMC11*rif*F', in which the constructed plasmids for the NifA based P2H system could be used, the next step was to assess the usefulness of all of the constructed plasmids in each of the host strains and to optimize the growth conditions under which the P2H assay would be carried out in each of the two strains.

As previously mentioned the NC domain of the NifA protein was known to be capable of activating transcription from the *nif*H promoter despite lacking the C-terminal DBD of the protein. The level of expression from the *nif*H promoter induced by the NifA NC domain was known to be lower than that observed for the whole NifA protein. It was hoped that by using the Control plasmid pPC228 (expressing the whole NifA protein) as a positive control and the Prey plasmid pPC229-A (expressing the NC domain of NifA) as negative control with each of the *nifH::lacZα* Bait plasmids, pPC187-A and pPC187-B, that the growth conditions could be

manipulated to optimize the contrast between the colony colour intensities given by the positive pPC228 and pPC229-A negative controls on Xgal plates. Growth conditions were to be optimized in the same way for the alternative Control and Prey Plasmids, pPC230 (expressing the whole NifA protein from manipulated *ptac* promoter) and pPC231-A (expressing the NifA NC domain), with each of the Bait plasmids pPC187-A and pPC187-B and their usefulness compared with that of the pPC228/pPC229-A combination of Control/Prey plasmids. In this way the most useful Control/Prey plasmid combination and the most useful Bait plasmid would be determined along with the growth conditions under which they gave the best results.

Several factors had to be taken into consideration when optimizing growth conditions. First of all, it was known that the NifA protein is oxygen sensitive. Oxygen was known to inhibit the ability of the NifA protein to bind to its cognate DNA sequence and also its ability to activate transcription. However, it was known that even in the presence of oxygen some NifA induced expression from the *nifH* promoter could still occur. It was hoped that the oxygen sensitivity of the NifA protein could be used to advantage by helping to reduce the low level of *nifH:lacZα* expression induced by the NifA NC domain (expressed from the Prey plasmids pPC229-A and pPC231-A) to an undetectable level while still permitting the detection of more strongly induced *nifH:lacZα* expression by the whole NifA protein (expressed from the Control plasmids pPC228 and pPC230) or by positive protein-protein interactions in the context of an actual P2H assay. In this way clear positive and negative results could be distinguished.

Another factor that had to be considered was the fact that the NifA protein is temperature sensitive, working optimally at a temperature of 28 ^{o}C. Also the optimal IPTG level to be used for inducing the expression of the whole NifA protein from the Control plasmids and the NifA NC domain from the Prey plasmids would have to be determined. It was expected that lower levels of induction would produce the best contrast between positive pPC228/pPC230 controls and their corresponding pPC229-A/pPC231-A negative controls. The theory behind this was that the NifA NC domain activated transcription from the *nifH* promoter from solution as it did not possess a C-terminal DBD. The level to which NifA NC domain induced expression from the

nifH promoter was therefore expected to be dependent on its concentration in the cell and thus the lower its concentration in the cell the lower the level of induced expression from the *nifH* promoter. On the other hand, the level of expression induced from the *nifH* promoter by the whole NifA protein was not expected to be as dependent on its concentration in the cell due to its ability to bind at the *nifH* promoter. This binding activity would be expected to increase the local concentration of the whole NifA protein in the vicinity of the *nifH* promoter.

The results of the optimization process for both strains are given below and a discussion of the results can be found in Chapter 8.

6.3.1: Optimization Of The P2H System In The *E. coli* Host Strain ET6016*rif*F'.

Initial attempts to optimize the P2H system in the *E. coli* host strain ET6016*rif*F' focused on determining which combination of Control and Prey plasmids (pPC228/pPC229-A or pPC230/pPC231-A) gave the best positive control (whole NifA protein) to negative control (NifA NC domain protein) contrast with each of the two *nifH:lacZα* Bait plasmids (pPC187-A and pPC187-B) on Xgal plates. In order to do this the constructed Control, Prey and *nifH:lacZα* Bait plasmids were transformed into ET6016*rif*F' in the combinations shown in Figure 6.3.1 and plated on appropriate selective media (see Figure 6.3.1) .

Each combination of plasmids was plated on three different levels of IPTG; 0.5mM, 1.0mM and 1.5mM. The purpose of this was to determine which level of IPTG would give the best results for each combination of plasmids. The resulting set of plates was incubated at 30 °C for 72 hours. Some colour development of colonies was visible after 48 hours but a further 24 hours incubation was required for strong color development. A duplicate set of plates was prepared and incubated first at 37 °C for 24 hours and then switched to 30 °C for another 24 hours after which

strong colour development of colonies was obtained. The reason for the 37 °C incubation is discussed in Chapter 8.

- **Figure 6.3.1: Combinations Of Control, Prey And Bait Plasmids Transformed Into ET6016*rif*F'**

Plasmid Combinations		Plated On
pPC185-A	pC185-B	LB tetracycline / chloramphenicol Xgal (50µg/ml) IPTG
pPC187-A	pPC187-B	
pPC187-A / pPC228	pPC187-B / pPC228	LB tetracycline / chloramphenicol ampicillin Xgal (50µg/ml) IPTG
pPC187-A / pPC229-A	pPC187-B / pPC229-A	
pPC187-A / pPC230	pPC187-B / pPC230	
pPC187-A / pPC231-A	pPC187-B / pPC231-A	

The colour intensity of colonies on all of the plates was compared and the colony colour intensity on each plate was graded on a visual basis. In the final P2H system positive protein-protein interactions would be identified by visual inspection of plates and so it was decided that the system should be initially optimized on a visual basis. The results obtained are given in Table 6.3.1 and Table 6.3.2 below

Table 6.3.1:Results For ET6016*rif* F' Plates Incubated At 30 °C For 72 hours.

Plasmids	*0.5 mM*	*1.0 mM*	*1.5 mM*
pPC185A	+	+	-
pPC187A	+	+	-
pPC187A-pPC228	++++	++++	++++
pPC187A-pPC229	+++	+++	++
pPC187A-pPC230	++++	++++	++++
pPC187A-pPC231	++++	++++	+++
pPC185B	+++	+++	++
pPC187B	+++	+++	++
pPC187B-pPC228	++++	++++	++++
pPC187B-pPC229	++++	++++	++++
pPC187B-pPC230	++++	++++	++++
pPC187B-pPC231	++++	++++	++++

Table 6.3.2: Results For ET6016*rif*F' Plates Incubated At 37 °C For 24 hours And Then 30 °C For 24 hours.

Plasmids	*0.5 mM*	*1.0 mM*	*1.5 mM*
pPC185A	+	+	-
pPC187A	+	+	-
pPC187A-pPC228	++++	++++	++++
pPC187A-pPC229	+++	+++	++
pPC187A-pPC230	++++	+++	++++
pPC187A-pPC231	++++	+++	+++
pPC185B	+++	+++	++
pPC187B	+++	+++	++
pPC187B-pPC228	++++	++++	++++
pPC187B-pPC229	++++	++++	++++
pPC187B-pPC230	++++	++++	++++
pPC187B-pPC231	++++	++++	++++

The results obtained for the plates incubated at 37 °C for 24 hours followed by a 30 °C incubation for another 24 hours mirrored those obtained for the plates incubated at 30 °C for 72 hours.

As can be seen from Table 6.3.1 and Table 6.3.2 the Bait plasmid pPC187-B was of very little use with any of the constructed Control or Prey plasmids under any of the IPTG concentrations or incubation conditions used. This Bait plasmid exhibited high levels of *lacZα* expression from its *nifH:lacZα* reporter gene in the absence of any expressed NifA or NifA NC domain proteins. Because of the high level of background expression from the pPC187-B Bait plasmid the colour intensity of the pPC228 positive control (expressing the whole NifA protein) could not be clearly distinguished from its corresponding pPC229-A negative control (expressing the NifA NC domain). Similarly the pPC230 positive control could not be clearly distinguished from its corresponding pPC231-A negative control. The pPC187-B Bait plasmid was thus eliminated from further optimization experiments.

The pPC187-A Bait plasmid, which has the *nifH:lacZα* reporter in the opposite orientation, also exhibited some background expression from its *nifH:lacZα* reporter gene in the absence of any expressed whole NifA or NifA NC domain protein but the level of this background *lacZα* expression was much lower than that observed for the pPC187-B Bait plasmid. Because the level of background expression from the pPC187-A plasmid was low the colony colour intensity given by each of the positive controls (pPC228 and pPC230) could be more clearly distinguished from that given by their corresponding negative controls (pPC229-A and pPC231-A respectively). The possible reasons for the background expression from the two pPC187 *nifH:lacZα* Bait plasmids and the reason for the difference in the level of background expression exhibited by each of the plasmids are discussed in chapter 8.

Another interesting and unexpected trend was also observed for both of the pPC187 Bait plasmids although it was more obvious for the pPC187-A plasmid because of the lower levels of background *nifH:lacZα* expression it exhibited. It was observed that increasing the IPTG concentration decreased the level of background expression from the *nifH:lacZα* reporter gene from these plasmids and the possible reasons for this are also discussed in chapter 8.

Of the two Control plasmid/Prey plasmid combinations, pPC228/pPC229-A and pPC230/pPC231-A, the pPC228/pPC229-A combination with the pPC187-A Bait plasmid was found to give the best positive control to negative control contrast. Strong activation of the *nifH:lacZα* reporter on the pPC187-A plasmid by the whole NifA protein expressed from the pPC228 Control plasmid was observed for all levels of IPTG and incubation temperatures. Activation of the *nifH:lacZα* reporter by the NifA NC domain expressed from pPC229-A Prey plasmid was also evident but was lower than that observed for the pPC228 plasmid. The strength of *nifH:lacZα* activation induced by the NifA NC domain protein seemed to decrease as the level of IPTG increased and the contrast between the positive pPC228 and negative pPC229-A controls was thus enhanced by increasing the IPTG level in the media. It was thought that the level of activation caused by the NifA NC domain expressed from the pPC229-A plasmid was actually very low. The improved contrast between the positive pPC228 and negative pPC229-A controls caused by increasing the IPTG

concentration was attributed to the decreased background *nifH:lacZα* expression from the pPC187-A Bait plasmid caused by increasing the IPTG concentration rather than to an actual decrease in the level of activation of the *nifH:lacZα* reporter by the NifA NC domain.

Strong activation of the *nifH:lacZα* reporter gene on pPC187-A by NifA expressed from the pPC230 Control plasmid was also observed and was comparable to that observed for the pPC228 Control plasmid. However, the level of *nifH:lacZα* expression induced by the NifA NC domain protein expressed from the pPC231-A Prey plasmid was stronger than that observed when it was expressed from the pPC229-A Prey plasmid. As a result of this the contrast between the pPC230 positive control and the pPC231-A negative control was not as distinct as that observed for the pPC228/pPC229-A combination of plasmids. Increasing the IPTG concentration did enhance the contrast between the pPC230 positive control and the pPC231-A negative control but the contrast obtained for the pPC228/pPC229-A combination of plasmids was still more distinct. The pPC230/pPC231-A combination of Control/Prey plasmids was therefore eliminated from further optimization experiments.

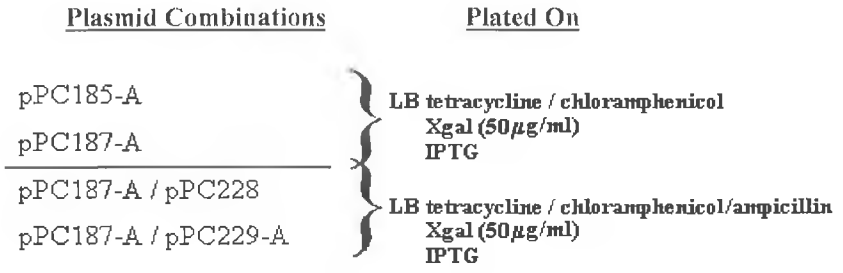
Overall the best results were obtained when the pPC228 Control plasmid and the pPC229-A Prey plasmid were used with the pPC187-A Bait plasmid and when a high concentration of IPTG (1.5mM) was used to induce expression of the NifA and NifA NC domains. Results obtained when a 30 ^{o}C incubation for 72 hours was used were slightly better than those obtained when plates were first incubated at 37 ^{o}C for 24 hours followed by incubation at 30 ^{o}C for another 24 hours. The pPC228, pPC229-A and pPC187-A plasmids were selected for a second round of optimization experiments to try and further enhance the contrast between the positive pPC228 and negative pPC229-A controls.

In the second round of optimization experiments the level of IPTG used to induce expression of the NifA and NifA NC domain proteins from the pPC228 and pPC229-A plasmids was extended to 2mM to see if this would further enhance the contrast between the positive pPC228 and negative pPC229-A controls. Plasmids where

transformed into ET6016*rif*F' in the combinations shown in Figure 6.3.2 and plated on appropriate selective media.

Each combination of plasmids was plated on 4 different levels of IPTG; 0.5mM, 1.0mM, 1.5mM and 2mM. The resulting set of plates was incubated at 30 °C for 72 hours. These plates were designated the (A) set of plates.

- **Figure 6.3.2: Combinations Of Control, Prey And Bait Plasmids Transformed Into ET6016*rif*F'.**



A duplicate set of plates was also prepared, sealed with parafilm and incubated at 30 °C for 72 hours. This set of plates was designated the (B) set of plates. The purpose of wrapping these plates in parafilm was to try and provide micro-aerobic conditions under which NifA worked optimally. Initially bacteria growing on the plates would have sufficient oxygen to grow and thus would grow more quickly to form visible colonies than if they were placed in an anaerobic jar. As the bacteria grew the oxygen within the plates would gradually be depleted and micro-aerobic conditions would ensue. It was hoped that under these conditions the expressed NifA protein from the pPC228 Control plasmid would activate the *nifH:lacZα* reporter gene more efficiently and generate more rapid and more intense colour development. It was also hoped that this would enhance the contrast between the colour intensity of the positive pPC228 control and the negative pPC229-A.

The colour intensity of colonies on both the (A) and (B) sets of plates was compared and plates were graded on a visual basis as before. The results obtained are given in

Table 6.3.3 below. The experiment was repeated again with the exception that plates were incubated first at 37 °C for 24 hours and then at 30 °C for another 24 hours. The colour intensity of colonies on all of these plates was compared and plates were graded on a visual basis. The results obtained are given in Table 6.3.4 below.

Table 6.3.3: Results For Second Round Of Optimization Experiments In ET6016*rif*F' - Plates Incubated At 30 °C For 72 hours.

Plasmids	0.5mM IPTG		1.0mM IPTG		1.5mM IPTG		2.0mM IPTG	
	A	B	A	B	A	B	A	B
pPC228-pPC187A	+++	++++	+++	++++	+++	++++	+++	++++
pPC229A-pPC187A	++	++++	++	++++	+	++++	+	++++
pPC185A	++	+++	++	+++	+	++	-	++
pPC187A	++	+++	++	+++	+	++	-	++

A = unwrapped plates B = plates wrapped in parafilm

Table 6.3.4: Results For Second Round Of Optimization Experiments In ET6016*rif*F' - Plates Incubated At 37 °C For 24 hours And 30 °C For 24 hours.

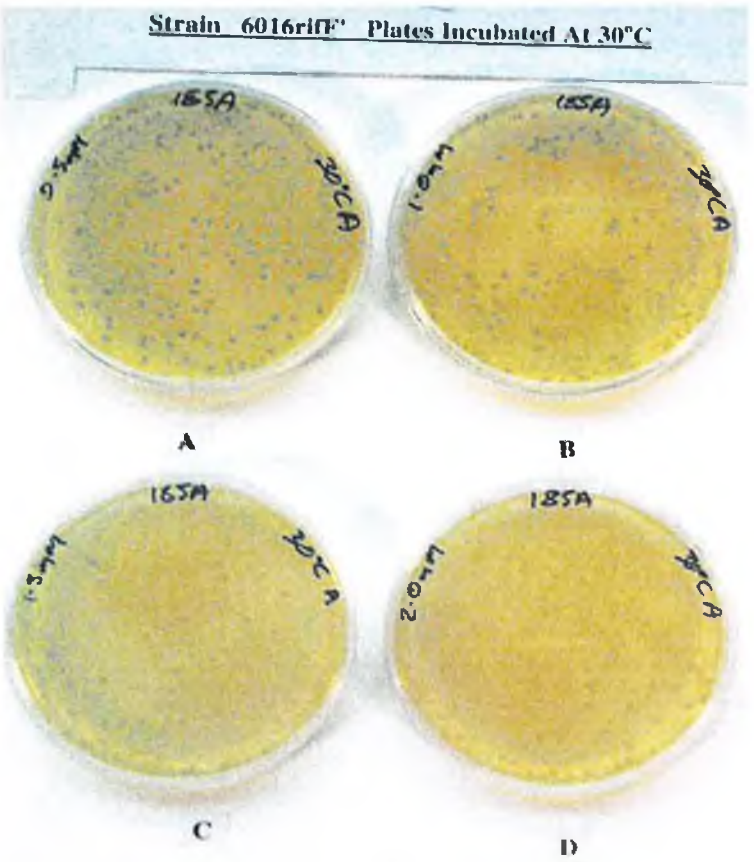
Plasmids	0.5mM IPTG		1.0mM IPTG		1.5mM IPTG		2.0mM IPTG	
	A	B	A	B	A	B	A	B
pPC228-pPC187A	+++	++++	+++	++++	+++	++++	+++	++++
pPC229A-pPC187A	++	+++	++	+++	+	++	+	+
pPC185A	++	++	+	++	+	+	-	+
pPC187A	++	++	+	++	+	+	-	+

A = unwrapped plates B = plates wrapped in parafilm

The results obtained for plates that were not wrapped in parafilm (the A set of plates) mirrored those obtained in the initial optimization experiments. Increasing IPTG concentration caused a reduction in the amount of background expression from the *nifH:lacZα* reporter gene on the pPC187-A Bait plasmid. At the highest concentration of IPTG (2mM) background expression from the *nifH:lacZα* reporter gene was not detected on Xgal plates (see Figure 6.3.3). The greatest positive pPC228 control to negative pPC229-A control contrast was also obtained when 2mM

IPTG was used (see Figure 6.3.4). Activation of the *nifH:lacZα* reporter by the pPC229-A expressed NifA NC domain protein was barely detectable on Xgal plates when 2mM IPTG was used to induce its expression (see Figure 6.3.5). As in the initial optimization experiments the results obtained for plates incubated at 30 °C for 72 hours mirrored those obtained for the set of plates that were incubated first at 37 °C for 24 hours and then switched to 30 °C for another 24 hours. However the results obtained for the plates that were incubated at 30 °C for 72 hours were slightly better.

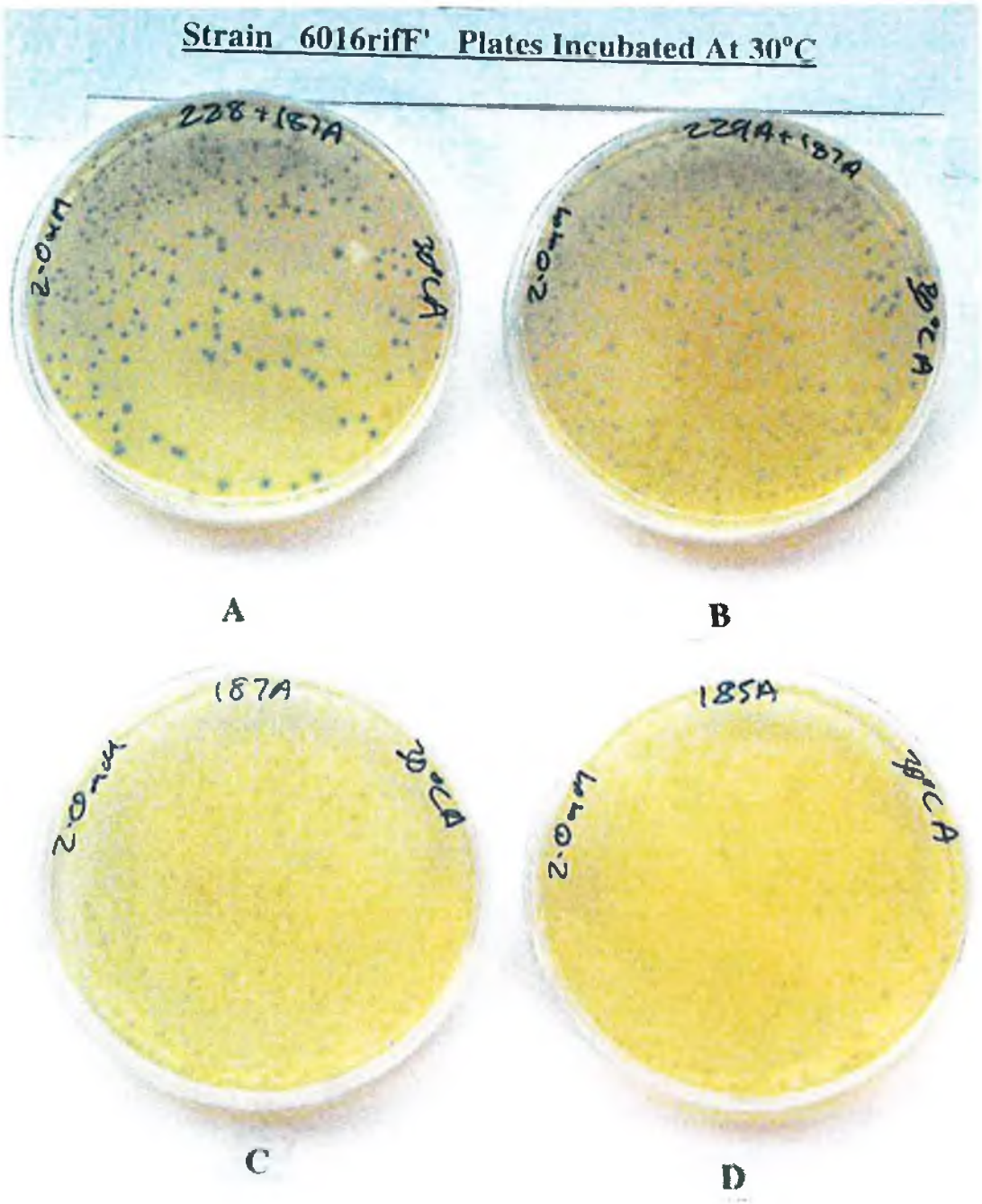
- **Figure 6.3.3: Plates Showing IPTG Effect On Background Expression From pPC185-A.**



A = pPC185-A/ 0.5mM IPTG **B = pPC185-A/ 1.0mM IPTG**

C = pPC185-A/ 1.5mM IPTG **D = pPC185-A/ 2.0mM IPTG**

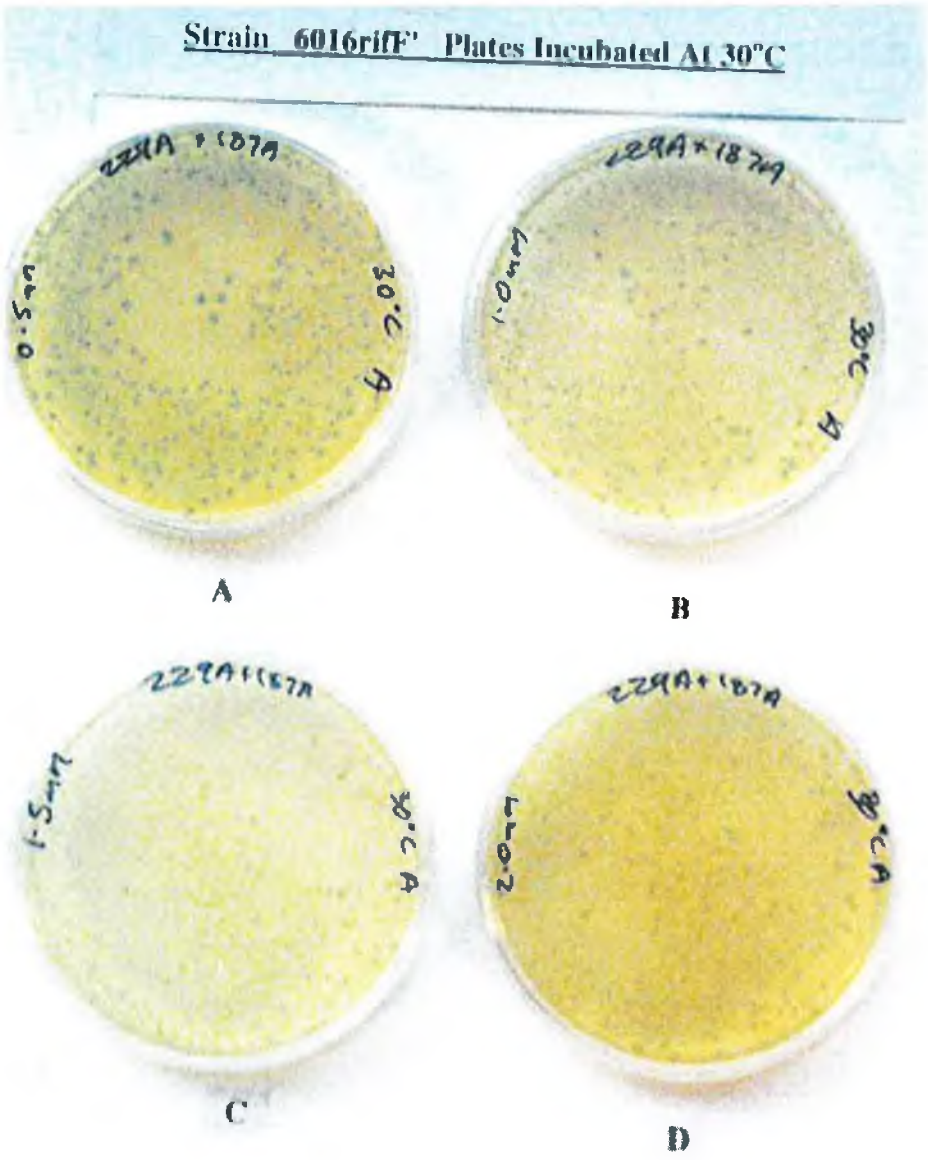
- **Figure 6.3.4: Results Obtain For pPC228/pPC229-A Combination Of Control/Prey Plasmids Using 2mM IPTG**



A = pPC228/pPC187-A **B = pPC229-A/pPC187-A**

C= pPC187-A **D = pPC185-A**

- **Figure 6.3.5: Plates Showing IPTG Effect On pPC229-A Induced Reporter Expression**



A = pPC229-A / 0.5mM IPTG **B = pPC229-A / 1.0mM IPTG**

C = pPC229-A / 1.5mM IPTG **D = pPC229-A / 2.0mM IPTG**

Miller assays were performed in order to quantify the levels of *nifH:lacZα* expression observed for each of the combinations of plasmids on Xgal plates containing different concentrations of IPTG. ET6016*rif*F' transformants containing each combination of plasmids were inoculated into 30ml universals containing 20mls of selective LB broth to which IPTG had been added to a final concentration of 0.5mM, 1.0mM, 1.5mM and 2.0mM IPTG. The lids of the tubes were closed tightly and the tubes were incubated at 30 °C with low shaking for 48 hours. The tubes were then assayed in duplicate for β-galactosidase activity. The results obtained are shown in Table 6.3.5 below and mirror the results that were obtained on Xgal plates. Similar levels of β-galactosidase activity were obtained when assays were performed after 48 hours incubation

Table 6.3.5: β-Galctosidase Activity For Each Combination Of Plasmids In ET6016*rif*F' Under Different Concentrations Of IPTG.

Plasmids	β-Galactosidase Activity (Miller Units)			
	0.5mM IPTG	1.0mM IPTG	1.5mM IPTG	2.0mM IPTG
pPC185A	4.99	4.87	4.75	4.50
pPC187A	4.20	4.38	4.90	4.05
pPC228-pPC187A	6.99	7.27	11.29	12.62
pPC229A-pPC187A	5.81	5.73	5.60	5.52

Wrapping plates in parafilm had the unexpected effect of increasing background expression from the *nifH:lacZα* reporter on the pPC187-A Bait plasmid, especially for the plates that were incubated at 30 °C for 72 hours (see Table 6.3.3 and Table 6.3.4). This made it difficult to distinguish the color intensity of colonies obtained for the pPC228 positive control from those obtained for the pPC229-A negative control. Although increasing the IPTG concentration had the effect of decreasing this background expression from the *nifH:lacZα* reporter and enhancing the contrast between the positive pPC228 and negative pPC229-A controls, the best results were still obtained for the unwrapped set of plates. Again the results obtained for the

wrapped set of plates incubated at 30 °C for 72 hours were slightly better than those obtained for those that were incubated first at 37 °C for 24 hours and then switched to 30 °C for another 24 hours.

From all of the above optimization experiments it was determined that the best combination of Control/Prey plasmids was the pPC228/pPC229-A combination and that these gave the best results when used with the pPC187-A Bait plasmid. The optimal growth conditions for these plasmids in the *E. coli* strain ET6016*rif*F' was incubation on media containing 2mM IPTG at 30 °C for 72 hours without wrapping the plates in parafilm (see plate D in Figure 6.3.3).

6.3.2: Optimization Of The P2H System In The *E. coli* Host Strain YMC11*rif*F'.

Having determined that the pPC228/pPC229-A combination of Control/Prey plasmids gave the best results with the pPC187-A Bait plasmid in the ET6016*rif*F' strain of *E. coli* it was decided to test these plasmids in the YMC11*rif*F' strain of *E. coli* to see if the contrast between the positive pPC228 control the negative pPC229 control would be enhanced. This strain has a gross deletion of the *glnALG* operon an grows very slowly in the absence of glutamine.

Plasmids where transformed into YMC11*rif*F' in the same combinations as those shown in Figure 6.3.2 and plated on appropriate selective media (see Figure 6.3.2). Each combination of plasmids was plated on 4 different levels of IPTG; 0.5mM, 1.0mM, 1.5mM and 2mM. These plates were designated the (A) set of plates and where incubated at 30 °C. A duplicate set of plates was prepared, sealed with parafilm and also incubated at 30 °C. This set of plates was designated the (B) set of plates. Colour development of colonies was very slow and, contrary to what was observed in the ET6016*rif*F' strain of *E. coli*, only developed on the plates that were wrapped in parafilm (the B set of plates). Although some colour development was evident on these plates after 72 hours they required another 24 hours incubation for distinct colour development to occur. The colour intensity of colonies on these plates

was graded on a visual basis as before and the results obtained are given in Table 6.3.6 below. The experiment was repeated again with the exception that plates were initially incubated first at 37 °C for 24 hours. These plates then required incubated at 30 °C for another 48 hours before strong color development was obtained and again color development was only obtained for the plates wrapped in parafilm (the B set of plates). The color intensity of colonies on these plates was graded on a visual basis and the results obtained are given in Table 6.3.7 below.

Table 6.3.6: Results Optimization Experiments In YMC11*rif*F' - Plates Incubated At 30 °C For 96 hours.

Plasmids	0.5mM IPTG		1.0mM IPTG		1.5mM IPTG		2.0mM IPTG	
	A	B	A	B	A	B	A	B
pPC228-pPC187A	-	+++	-	+++	-	+++	-	+++
pPC229A-pPC187A	-	++	-	++	-	+	-	+
pPC185A	-	++	-	++	-	+	-	-
pPC187A	-	++	-	++	-	+	-	-

A = unwrapped plates B = plates wrapped in parafilm

Table 6.3.7: Results For Optimization Experiments In YMC11*rif*F' - Plates Incubated At 37 °C For 24 hours And 30 °C For 48 hours.

Plasmids	0.5mM IPTG		1.0mM IPTG		1.5mM IPTG		2.0mM IPTG	
	A	B	A	B	A	B	A	B
pPC228-pPC187A	-	+++	-	+++	-	+++	-	+++
pPC229A-pPC187A	-	++	-	++	-	+	-	+
pPC185A	-	++	-	++	-	+	-	+
pPC187A	-	++	-	++	-	+	-	+

A = unwrapped plates B = plates wrapped in parafilm

While no colour development was observed for unwrapped plates the results obtained for the wrapped plates mirrored those obtained for the ET6016*rif*F' unwrapped plates. The reason for no color development on the unwrapped YMC11*rif*F' plates was attributed to the slow growth rate of the YMC11*rif*F' strain. It was felt that because of its slow growth rate this strain only expressed very low

levels of the NifA protein from plasmid pPC228. This fact coupled with the oxygen sensitivity of NifA protein resulted in a failure to detect NifA induced *nifH:lacZα* expression on Xgal plates. Wrapping plates in parafilm had the same effect of enhancing *nifH:lacZα* background expression from the pPC187-A Bait plasmid as it did in ET6016*rif*F' but in the YMC11*rif*F' strain this increase in *nifH:lacZα* background expression was required in order to boost NifA induced *nifH:lacZα* expression to a level that could be detected on Xgal plates. Increasing IPTG concentration caused a decrease in the level of background expression from the *nifH:lacZα* reporter gene on pPC187-A and enhanced the contrast between the positive pPC228 and negative pPC229-A controls. Slightly better results were obtained when 2mM IPTG was used and when plates were incubated at 30 °C for 96 hours rather than at 37 °C for 24 hours followed by 48 hours at 30 °C.

Miller assays were performed in order to quantify the levels of *nifH:lacZα* expression observed for each of the combinations of plasmids on Xgal plates containing different concentrations of IPTG. YMC11*rif*F' transformants containing each combination of plasmids were inoculated into 30ml universals containing 20mls of selective LB broth to which IPTG had been added to a final concentration of 0.5mM, 1.0mM, 1.5mM and 2.0mM IPTG. The lids of the tubes were closed tightly and the tubes were incubated at 30° C with low shaking for 24hours. The tubes were then assayed in duplicate for β-galactosidase activity. No β-galactosidase activity could be detected even when cultures were incubated for 48hours or 72hours without shaking.

From the above optimization experiments it was determined that the best contrast between the positive pPC228 and negative pPC229-A controls was obtained when media containing 2mM IPTG was used and when plates were sealed with parafilm and incubated at 30 °C for 96 hours.

6.4: Summary Of Chapter 6.

The work described in this chapter resulted in the identification of the most useful set of Control, Prey and Bait plasmids out of all of the plasmids that were constructed. It also resulted in the determination of the optimal growth conditions under which these plasmids should be used when performing P2H assays. The optimal set of plasmids comprised the pPC228 Control plasmid, the pPC229-A Prey plasmid and the pPC187-A Bait plasmid. The optimal growth conditions where determined to be 30 °C for 72 hours on media containing 2mM IPTG and under aerobic conditions for the strain ET6016*rif*F'. The same conditions were optimal for the YMC11*rif*F' strain with the exception that plates had to be sealed with parafilm in order to provide microaerobic conditions.

Chapter 7

Testing Of The NifA Based Prokaryotic Two-Hybrid System.

7.1: Introduction.

Having identified the most useful set of constructed Control, Prey and Bait plasmids (pPC228, pPC229-A and pPC187-A respectively) and having optimized the growth conditions under which the P2H assay would be carried out, the ability of the constructed NifA based P2H system to detect protein-protein interactions *in vivo* was tested. In order to do this the genes for two proteins known to interact were to be cloned into the pPC229-A Prey and pPC187-A Bait plasmids to see if the interaction between the proteins could be detected in the NifA based P2H system. The proteins selected for this purpose were the Intimin and Tir proteins of Enteropathogenic *E. coli* (EPEC) which had been shown to interact in the yeast two-hybrid (Y2H) system by *Hartland et al (1999)*. The structural and biological characteristics of these two proteins are described below and in the paper by *Hartland et al (1999)*.

7.1.1: The Tir And Intimin Proteins Of EPEC.

EPEC and Enterohaemorrhagic *E. coli* (EHEC) constitute a significant risk to human health worldwide. EPEC and EHEC adhere to host cells and induce characteristic attaching and effacing (A/E) lesions on epithelial cells. A/E lesions are characterised by localised destruction (effacement) of brush border microvilli and reorganisation of the host cell cytoskeleton beneath the intimately attached bacteria. This event is mediated, in part, by binding of the bacterial outer membrane protein, intimin, to a second EPEC protein called Tir (Translocated Intimin Receptor) which is translocated to the host cell where it becomes integrated into the host cell plasma membrane. The genes coding for the intimin protein (*eae* gene) and the Tir protein (*tir*) are located in a pathagenicity island known as the locus of enterocyte effacement or the 'LEE' region

The product of the *eae* gene in EPEC is intimin α and its cell binding activity is localised to the C-terminal 280 amino acids (Int280α). Int280α comprises three separate domains,

two immunoglobulin-like domains and a C-type lectin-like module. Int280α can bind to host cells even in the absence of Tir and this binding requires a correctly folded C-type lectin-like domain. Correct folding of the C-type lectin-like module of Int280α relies on a disulphide bridge between Cys-858 and Cys-937. Int280C/Aα and Int280C/Sα are two biologically inactive forms of Int280α in which Cys-937 has been substituted by Serine or Alanine. Neither of the mutant forms of Int280α binds to host cells in the absence of Tir. This data provides strong evidence that intimin interacts not only with Tir but also with a host cell intimin receptor.

Tir consists of at least three functional regions, an extracellular domain(s) that interact with intimin, a transmembrane domain(s) and a cytoplasmic domain(s) that can focus polymerised actin and other cytoskeletal proteins beneath intimately attached bacteria, leading to the formation of characteristic pedestal-like structures. *Kenny et al (1997)* highlighted the presence of two putative transmembrane domains along the Tir polypeptide. This implies that Tir is inserted into the host cell membrane in a 'hairpin loop' topology in which the loop may be facing either side of the host cell membrane. *Hartland et al (1999)* utilised the yeast two-hybrid (Y2H) system to localise the intimin-binding region of Tir to the central 107-amino-acid region, designated Tir-M, located between the two putative transmembrane domains.

7.1.2: Analysis Of Tir-Int280α Binding Using The Y2H System.

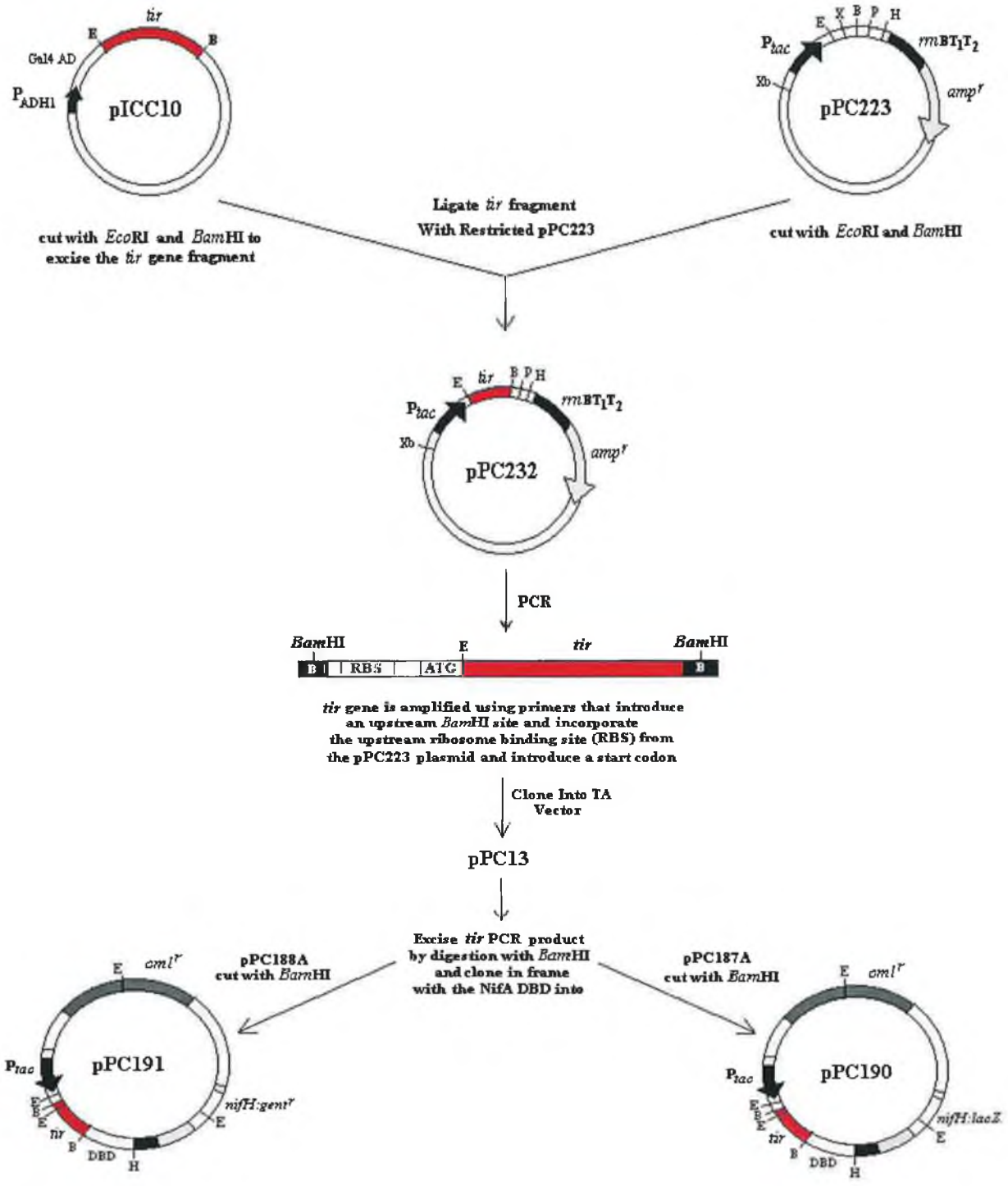
In order to analyse the interaction between Tir and Int280α *Hartland et al (1999)* amplified *tir* and the DNA sequences for the Int280α, Int280C/Aα and Int280C/Sα proteins by PCR and cloned the resulting PCR products into the Y2H system plasmids pGBT9 and pGAD424. The *tir* gene was cloned into pGBT9 as an *Eco*RI-*Bam*HI fragment to form the plasmid pICC10 that expresses Tir as C terminal fusion with the Gal4-DBD. The DNA sequences for the Int280α (*int280α*), Int280C/Aα (*int280C/Aα*) and Int280C/Sα (*int280C/Sα)* proteins were cloned as *Eco*RI-*Bam*HI fragments into the pGAD424 plasmid to form plasmids pPC19, pPC20 and pPC21 respectively. These plasmids express the

Int280α, Int280C/Aα and Int280C/Sα proteins as C-terminal fusions with the Gal4 AD. Each of these plasmids was co-transformed with plasmid pICC10 into the reporter yeast strain PJ69-4A (*trp, leu, ura3, his3*, *gal4, gal80*, Gal1-HIS3, Gal2-ADE, Gal7-*lacZ*). Int280α, Int280C/Aα and Int280C/Sα were all found to interact with Tir indicating that the C-type lectin like domain was not required for the binding of Tir to Int280α.

In order to localise the intimin-binding domain of Tir, three fragments of *tir* corresponding to the amino-terminal region (Tir-N), the carboxy-terminal region (Tir-C) and the middle region located between the two putative transmembrane domains (Tir-M) were amplified by PCR and cloned into the pGBT9 plasmid to form plasmids pICC13, pICC14 and pICC15 respectively. Each of these plasmids expresses its intimin protein as a C-terminal fusion with the Gal4 AD. Each of the plasmids was co-transformed with pICC19 (expressing Int280α as a fusion with the Gal4 DBD) into the yeast reporter strain PJ69-4A and interaction was only detected between Tir-M (pICC15) and Int280α (pICC19). This localised the intimin-binding domain to the Tir-M region of *tir*.

Because the interaction between the two EPEC proteins Tir and Int280α proteins had been detected in the Y2H system is was decided that these two prokaryotic proteins would be used to test the newly constructed NifA based P2H system. Plasmids pICC10, pICC19 and pICC20 were obtained as a gift from Gad Frankel at the Imperial College Of Science, Technology And Medicine In London. These plasmids were used as sources from which the DNA sequences for Tir (*tir*), Int280α (*int280α*) and Int280C/Aα (*int280C/Aα*) were amplified by PCR. The resulting PCR products were cloned into the newly constructed Prey and Bait plasmids for the NifA based P2H system. The *tir* gene was cloned into the pPC187-A Bait plasmid while the *int280α* and *int280C/Aα* sequences were cloned into the pPC229-A Prey plasmid. Figure 7.2.1 and Figure 7.3.1 show schematic representations of the strategies employed for the cloning of *tir* and *int280α* respectively. These cloning strategies are described in detail in the following sections of this chapter.

- **Figure 7.2.1: Cloning Of *tir* Into The Bait Plasmid pPC187-A To Generate A N-Terminal Fusion With The Downstream NifA DNA Binding Domain Sequence.**



7.2: Cloning Of *tir* Into The pPC187-A Bait Plasmid To Form Plasmid pPC190 - Overview Of Strategy.

The *tir* gene was to be amplified from plasmid pICC10 by PCR and then cloned into the pPC187-A Bait plasmid N-terminally to the NifA DBD sequence. The *Eco*RI site within the Bait construct of the *nifH:lacZα* Bait plasmid pPC187-A could not be used for cloning *tir* as there are several *Eco*RI restriction sites within the pPC187-A Bait plasmid (see Figure 5.2.13). Only the unique *Xma*I and *Bam*HI restriction sites within the Bait construct on pPC187-A can be used for cloning genes N-terminally to the NifA-DBD. Although there are no *Xma*I or *Bam*HI sites within the *tir* gene it could not be directionally cloned into the pPC187-A Bait plasmid as an *Xma*I-*Bam*HI fragment. The reason for this was that the *Xma*I and *Bam*HI restriction sites within the Bait construct of pPC187-A partially overlap so that cutting of the plasmid by either one of the two enzymes prevents the other from cutting (this problem was encountered when cloning the NifA NC domain sequence into pPC223 - see section 3.2.1). It was therefore planned that the *tir* gene would be amplified from plasmid pICC10 as a *Bam*HI fragment which would then be cloned into pPC187-A to form a plasmid called pPC190. This plasmid would express a *tir*-NifA DBD hybrid protein from its *ptac* promoter.

Cloning of the above *tir* gene PCR product into the Bait plasmid pPC187-A as a *Bam*HI fragment would present a problem in that it would place the start codon of the *tir* PCR product too far from the pPC187-A plasmid borne RBS to facilitate efficient translation. The *tir* gene therefore had to be cloned with its own RBS. In order to add a RBS onto the *tir* gene prior to its cloning into the pPC187-A Bait plasmid the *tir* gene was first cut from plasmid pICC10 as an *Eco*RI-*Bam*HI fragment and cloned into pPC223 to form plasmid pPC232 (see Figure 7.2.1). The *tir* gene was then amplified from pPC232 as a *Bam*HI fragment using the primers *tir*-F and *tir*-R. The sequence of these primers is shown in Figure 7.2.2 below and the primed regions of the pPC232 sequence for which these primers were designed are shown in Figure 7.2.3.

- **Figure 7.2.2: Primers Used For The Amplification Of The *tir* Gene From pPC232.**

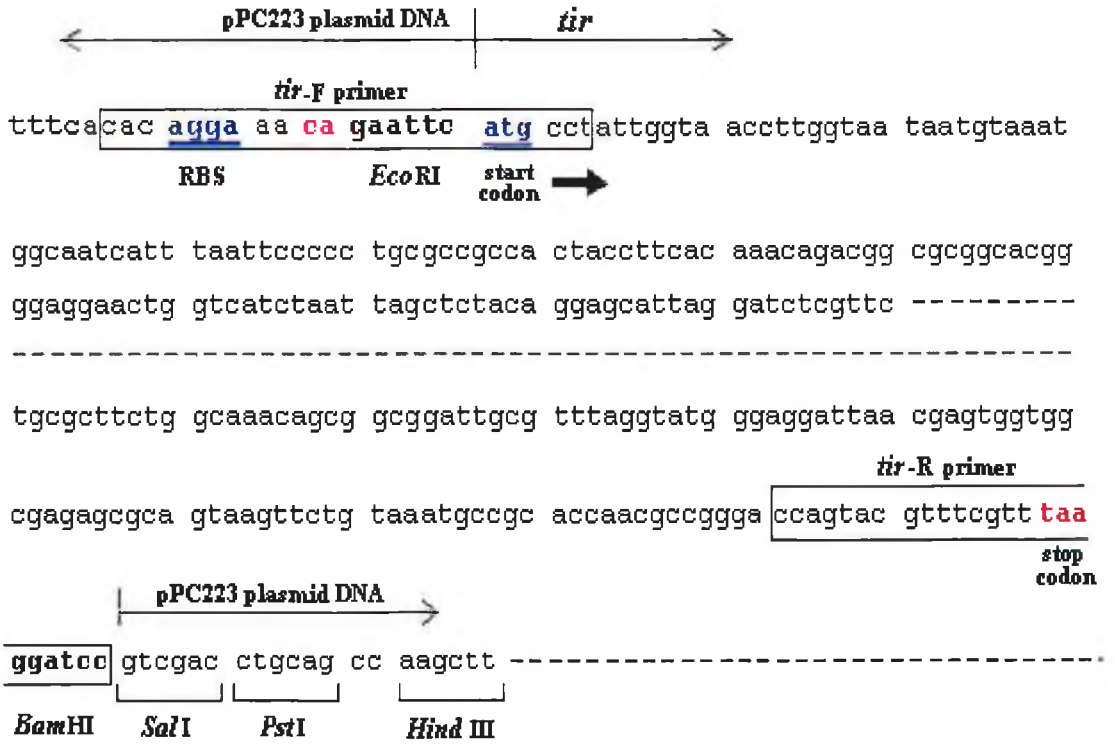
tir-F: (GGA TCC) - CAC AGG AAA GAA TTC ATG CCT

***Bam*HI**

tir-R: (GGA TCC) - AAC GAA ACG TAC TGG

***Bam*HI**

- **Figure 7.2.3: Primed Regions Of The *tir* Gene On Plasmid pPC232.**



NB: Bases in red were eliminated by the primer sequences

As can be seen from Figure 7.2.3 the *tir*-F primer was designed to incorporate the RBS of pPC223 into the *tir* PCR product. This primer also incorporated a *Bam*HI site onto the end of the *tir* PCR product. The *tir*-R primer was designed to eliminate the *tir* stop codon so that when the *tir* PCR product was cloned into the pPC187-A Bait plasmid translation could proceed through the *tir* sequence into the downstream NifA DBD sequence to generate a Tir-NifA DBD hybrid protein.

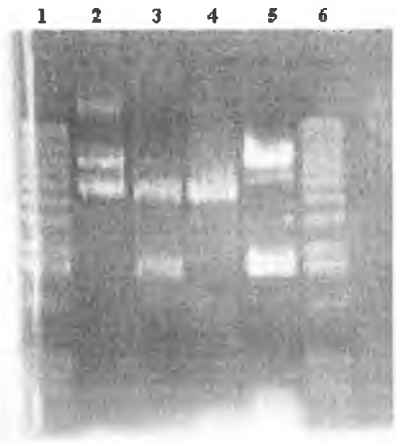
Once amplified the *tir* PCR product with its added RBS was cloned into the TA pCR2.1 vector. It was then sub-cloned into pPC187-A as a *Bam*HI fragment N-terminally and in the correct reading frame with the downstream NifA DBD sequence to form plasmid pPC190 (see Figure 7.2.1). The *tir* PCR product was also cloned into the pPC188-A *nifH:gent*^{r} Bait plasmid in the same manner to form the plasmid pPC191 (see Figure 7.2.1). Both the pPC190 and pPC191 plasmids express a Tir-NifA DBD hybrid protein.

7.2.1: Sub-cloning Of *tir* From pICC10 Into pPC223 To Form Plasmid pPC232.

The plasmid pICC10 was digested with *Eco*RI-*Bam*HI and the digested plasmid DNA was separated on a 0.7% agarose gel. The 1661kb band corresponding to the excised *tir* gene was purified from the gel using the gene clean procedure. Plasmid pPC223 was also digested with *Eco*RI-*Bam*HI and the restricted plasmid DNA was run on a 0.7% agarose gel. The 4.5kb band corresponding to the linearised plasmid DNA was purified from the gel using the gene clean procedure. The linearised pPC223 plasmid DNA was ligated with the purified *tir* *Eco*RI-*Bam*HI fragment. The ligated DNA was used to transform DH5α and the transformation was plated on LBamp media. Clones were screened by performing *Eco*RI-*Bam*HI double digests on plasmid DNA prepared from each clone. Screening in this manner resulted in the identification of the desired plasmid called pPC232. Figure 7.2.4 shows the results of the verifying restriction analysis of pPC232.

- **Figure 7.2.4: Restriction Analysis Of Plasmid pPC232.**

Lane No.	Sample
1	1Kb Ladder
2	pPC232 uncut
3	pPC232 cut *Eco*RI-*Bam*HI
4	pPC223 cut *Eco*RI-*Bam*HI
5	pICC10 cut *Eco*RI-*Bam*HI
6	1Kb Ladder



7.2.2: PCR Amplification Of *tir* From pPC232.

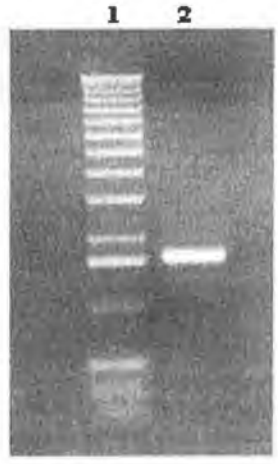
The *tir* gene was amplified from the plasmid pPC232 as a *Bam*HI fragment using the *tir*-F and *tir*-R primers described earlier. The PCR product was separated on a 0.7% agarose gel. A photo of the gel obtained and the conditions used for the PCR are shown in Figure 7.2.5 below. As can be seen from Figure 7.2.5 a single 1676bp band corresponding to the desired *tir* gene PCR product was obtained.

- **Figure 7.2.5: PCR Amplification Of *tir* From pPC232.**

Lane No.	Sample
1	1Kb Ladder
2	*tir* PCR Product

PCR Conditions

Annealing temperature	60 °C
Annealing time	30 sec
Extension time	2 mins

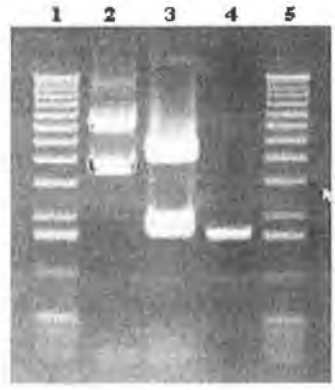


7.2.3: Cloning Of The *tir* PCR Product Into The TA Vector To Form Plasmid pPC13.

The *tir* PCR product was ligated with the TA pCR2.1 vector according to the TA protocol. The ligated DNA was used to transform INVαF' cells and the transformation was plated on LB ampicillin/Xgal plates. White colonies were isolated and screened by performing *Bam*HI digests on plasmid DNA isolated from each clone. Screening in this way resulted in the identification of the desired plasmid called pPC13. Figure 7.2.6 below shows the results of the verifying restriction analysis of plasmid pPC13.

- **Figure 7.2.6: Restriction Analysis Of Plasmid pPC13.**

Lane No.	Sample
1	1Kb Ladder
2	pPC13 uncut
3	pPC13 *Bam*HI
4	*tir* PCR Product
5	1Kb Ladder



7.2.4: Sub-cloning Of The *tir* PCR Product From pPC13 Into The pPC187-A Bait Plasmid To Form Plasmid pPC190.

Plasmid pPC13 was digested with *Bam*HI and the restricted plasmid DNA was separated on a 0.7% agarose gel. The 1676bp band corresponding to the excised *tir Bam*HI fragment was purified from the gel by the gene clean procedure. The *nifH:lacZα* Bait plasmid pPC187-A was also digested with *Bam*HI and the digested plasmid DNA was separated on a 0.7% agarose gel to verify complete digestion. The 5.4kb band corresponding to the linearised plasmid DNA was purified from the gel by the gene clean procedure and ligated

with the purified *tir* *Bam*HI fragment. The ligated DNA was used to transform DH5α and the transformation was plated on LBamp media. Clones obtained were screened by performing *Bam*HI digests on plasmid DNA isolated from each clone. Several clones containing a pPC187-A plasmid with an inserted *tir* *Bam*HI fragment were identified. Plasmid DNA from each of these clones was then digested with *Eco*RI to determine the orientation of the inserted *tir* *Bam*HI fragment as the *tir* PCR product has an *Eco*RI site at it N-terminal (see Figure 7.2.1). *Eco*RI digestion of the pPC187-A Bait plasmid generates 3 fragments of 1560bp, 1775bp, and 2160bp. *Eco*RI digestion of the desired pPC190 plasmid was expected to yield three fragments of 1775bp, 2160bp and 3236bp while a pPC187-A plasmid with the *tir* gene cloned in the opposite orientation was expected to yield four fragments of 1560bp, 1676bp, 1775bp and 2160bp. Screening by performing *Eco*RI resulted in the identification of a clone containing a pPC187-A plasmid with a *tir* *Bam*HI fragment inserted in the desired orientation and the resulting plasmid was called pPC190. The results of the verifying restriction analysis of the pPC190 plasmid can be seen in Figure 7.2.7 below.

- **Figure 7.2.7:Restriction Analysis Of Plasmid pPC190.**

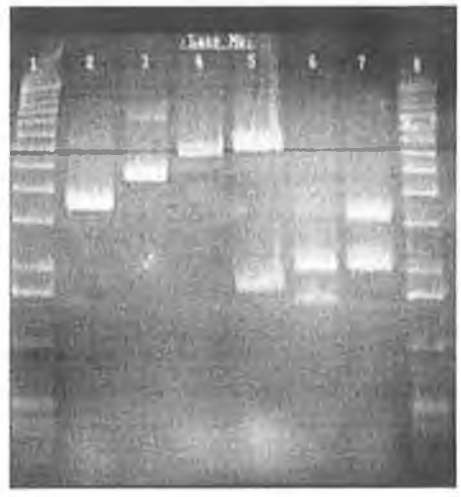
Lane No.	Sample
1	1Kb Ladder
2	pPC187-A uncut
3	pPC190 uncut
4	pPC187-A cut *Bam*HI
5	pPC190 cut *Bam*HI
6	pPC187-A cut *Eco*RI
7	pPC190 cut *Eco*RI
8	1Kb Ladder

7.2.5: Sub-cloning Of The *tir* PCR Product From pPC13 Into The pPC188-A Bait Plasmid To Form Plasmid pPC191.

The *tir* PCR product was sub-cloned from pPC13 into the *nifH:gent*^{r} Bait plasmid pPC188 to form plasmid pPC191 in the same way it was sub-cloned into the pPC187-A Bait plasmid. Plasmid pPC13 was digested with *Bam*HI and the restricted plasmid DNA was separated on a 0.7% agarose gel. The 1676bp band corresponding to the excised *tir Bam*HI fragment was purified from the gel by the gene clean procedure. The *nifH:gent*^{r} Bait plasmid pPC188-A was digested with *Bam*HI and the digested plasmid DNA was separated on a 0.7% agarose gel to verify complete digestion. The 5.6kb band corresponding to the linearised pPC188-A plasmid DNA was purified from the gel by the gene clean procedure and ligated with the purified *tir Bam*HI fragment. The ligated DNA was used to transform DH5α and the transformation was plated on LBamp media. Clones obtained were screened by performing *Bam*HI digests on plasmid DNA isolated from each clone. Several clones containing a pPC187-A plasmid with an inserted *tir Bam*HI fragment were identified. Plasmid DNA from each of these clones was then digested with *Eco*RI to determine the orientation of the inserted *tir Bam*HI fragment. *Eco*RI digestion of the pPC188-A Bait plasmid generates 3 fragments of 1560bp, 2062bp, and 2074bp.

- **Figure 7.2.8: Restriction Analysis Of Plasmid pPC191.**

Lane No.	Sample
1	1Kb Ladder
2	pPC188-A uncut
3	pPC191uncut
4	pPC188-A cut *Bam*HI
5	pPC191 cut *Bam*HI
6	pPC188-A cut *Eco*RI
7	pPC190 cut *Eco*RI
8	1Kb Ladder



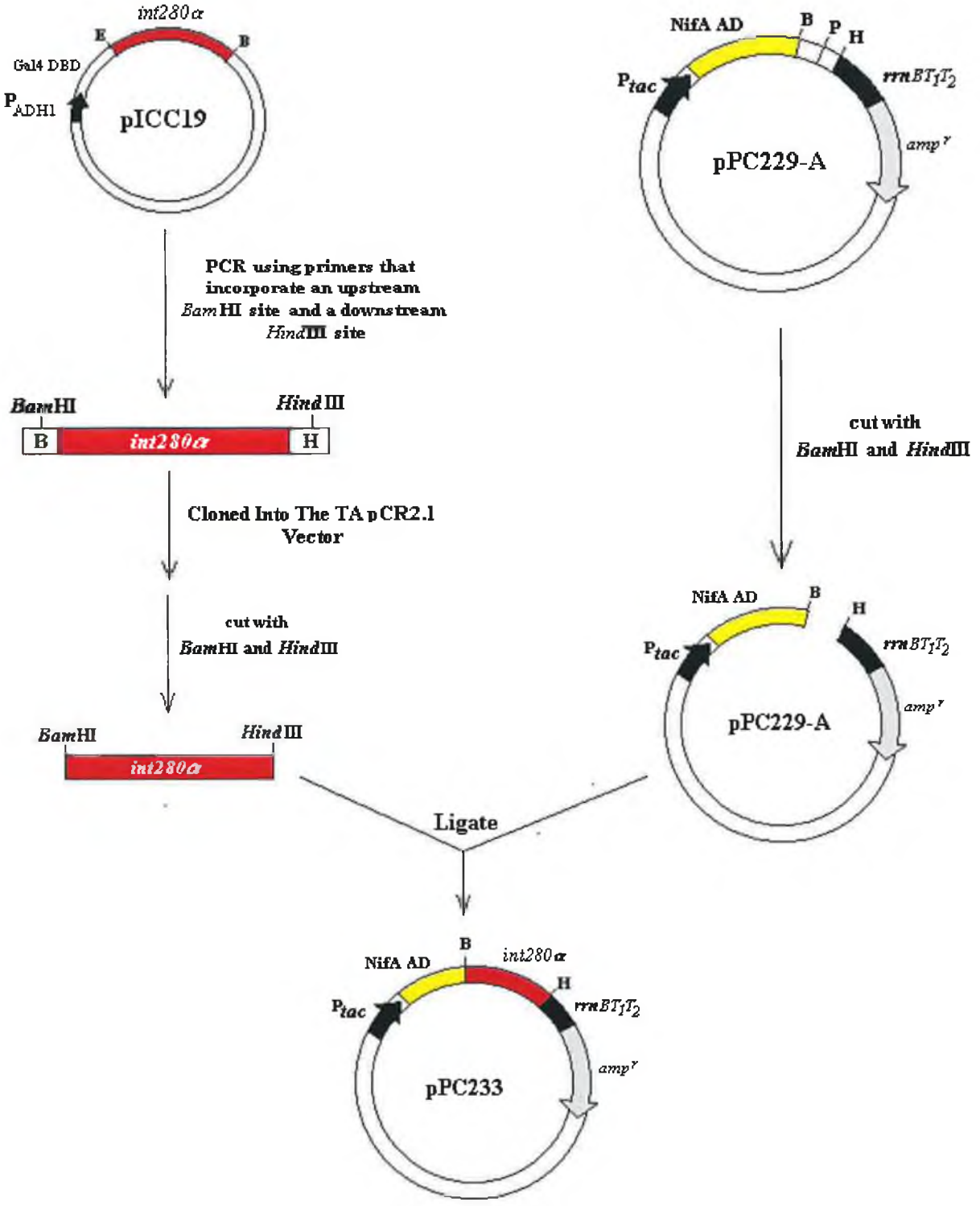
*Eco*RI digestion of the desired pPC191 plasmid was expected to yield three fragments of 2062bp, 2074bp and 3236bp while a pPC188-A plasmid with the *tir* gene cloned in the opposite orientation was expected to yield four fragments of 1560bp, 1676bp, 2062bp, and 2074bp. Screening by performing *Eco*RI digests thus resulted in the identification of a clone containing a pPC188-A plasmid with a *tir Bam*HI fragment inserted in the desired orientation and the resulting plasmid was called pPC191. The results of the verifying restriction analysis of the pPC191 plasmid can be seen in Figure 7.2.8.

7.3: Cloning Of *int280α* And *int280C/Aα* Into The pPC229-A Prey Plasmid - Overview Of Strategy.

Figure 7.3.1 shows a schematic representation of the strategy employed for the cloning of *int280α* into the pPC229-A Prey plasmid. The *int280α* sequence was amplified by PCR from plasmid pICC19 as a *Bam*HI-*Hind*III fragment using the primers *280α*-F and *280α*-R and cloned. The sequences of these primers are given in Figure 7.3.2 and the primed regions of the pICC19 are shown in Figure 7.3.3. Once amplified the *int280α* PCR product was cloned into the TA pCR2.1 vector and then sub-cloned into the pPC229-A Prey plasmid as a *Bam*HI-*Hind*III fragment C-terminally and in the correct reading frame with the upstream NifA-AD sequence. The resulting plasmid was called pPC233 and expresses a NifA AD-Int280α hybrid protein from its *ptac* promoter.

The *int280C/Aα* sequence differs from the *int280α* sequence by only one codon. This codon exchange results in the substitution of the Cys-937 residue of Int280α with an Alanine residue. The sequences of plasmids pICC19 and pICC20 are thus identical with the exception of this single codon exchange. The *int280C/Aα* sequence was amplified from pICC20 using the *280α*-F forward primer and the reverse primer *280C/Aα*-R. The sequence of the *280C/Aα*-R primer is shown in Figure 7.3.2. The *280C/Aα*-R primer encompasses the single codon exchange in the *int280C/Aα* sequence and this codon is indicated in red in Figure 7.3.2. As with *int280α*, *int280C/Aα* was amplified as a *Bam*HI-*Hind*III fragment.

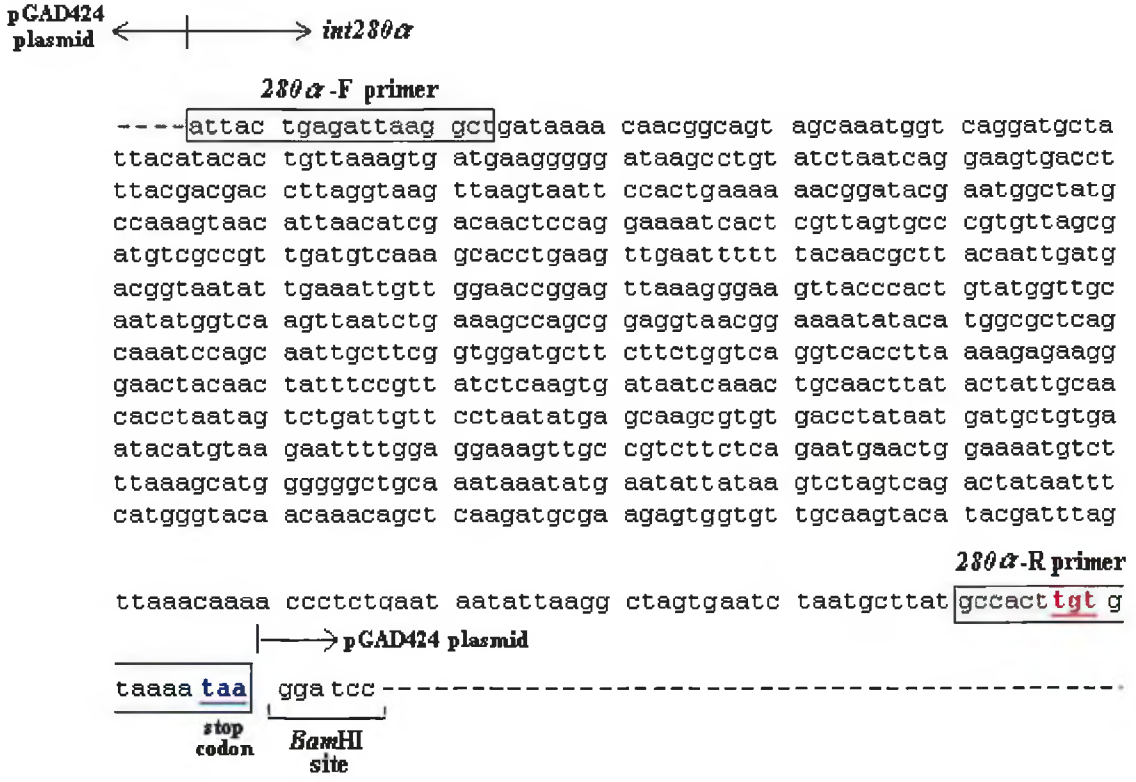
- **Figure 7.3.1: Cloning Of *int280α* Into The Prey Plasmid pPC229-A To Generate A C-Terminal Fusion With The Upstream NifA AD Sequence.**



- **Figure 7.3.2: Primers Used For The Amplification Of *int280α* From Plasmid pICC19.**

280α -F:	**(GGA TCC)** - ATT ACT GAG ATT AAG GCT ***Bam*HI**
280α-R:	**(AAG CTT)** - TTA TTT TAC ACA AGT GGC
280C/Aα-R	**(AAG CTT)** - TTA TTT TAC **AGC** AGT GGC ***Hind*III**

- **Figure 7.3.3: Primed Regions Of The *int280α* Sequence On Plasmid pICC19.**



NB: **sequence of pICC20 is identical to that of pICC19 shown above with the exception of the three bases in red which are exchanged for gct**

The *int280C/Aα* PCR product was first cloned into the TA pCR2.1 vector and then sub-cloned as a *Bam*HI-*Hind*III fragment into the pPC229-A Prey plasmid C-terminally and in the correct reading frame with the upstream NifA NC domain sequence. The resulting plasmid was called pPC234 and expresses a NifA AD-Int280C/Aα hybrid protein from its *ptac* promoter.

7.3.1: PCR Amplification Of *int280α* And *int280C/Aα* From Plasmids pICC19 And pICC20 Respectivelv.

The *int280α* sequence was amplified from the plasmid pICC19 as a *Bam*HI-*Hind*III fragment using the *280α*-F and *280α*-R primers described earlier. The PCR product was separated on a 0.7% agarose gel. A photo of the gel obtained and the conditions used for the PCR are shown in Figure 7.3.4 below. As can be seen from Figure 7.3.4 a single 855bp band corresponding to the desired *int280α* PCR product was obtained.

- **Figure 7.3.4: PCR Amplification Of *int280α* From pICC19.**

Lane No.	Sample
1	1Kb Ladder
2	*int280α* PCR Product

PCR Conditions

Annealing temperature	43 °C
Annealing time	30 sec
Extension time	1 min

The *int280C/Aα* sequence was amplified from the plasmid pICC20 as a *Bam*HI-*Hind*III fragment using the *280α*-F and *280C/Aα*-R primers described earlier. The PCR product was separated on a 0.7% agarose gel. A photo of the gel obtained and the conditions used for the PCR are shown in Figure 7.3.5 below. As can be seen from Figure 7.3.5 a single 855bp band corresponding to the desired *int280C/Aα* PCR product was obtained.

- **Figure 7.3.5: PCR Amplification Of *int280C/Aα* From pICC20.**

Lane No.	**Sample**
1	1Kb Ladder
2	*int280C/Aα* PCR Product

PCR Conditions

Annealing temperature	43 °C
Annealing time	30 sec
Extension time	1 min

7.3.2: Cloning Of The *int280α and int280C/Aα* PCR Products Into The TA Vector To Form Plasmids pPC11 And pPC12 Respectively.

The *int280α* PCR product was ligated with the TA pCR2.1 vector according to the TA protocol. The ligated DNA was used to transform INVαF' cells and the transformation was plated on LB ampicillin/Xgal plates. White colonies were isolated and screened by performing *Eco*RI digests (*Eco*RI sites flank the PCR product insertion point on the TA pCR2.1 vector) DNA isolated from each transformant. Clones identified as possessing an pCR2.1 vector with an inserted *int280α* were then screened further by performing *Bam*HI-*Hind*III double digests on plasmid DNA prepared from each clone. Screening in this way

resulted in the identification of the desired plasmid called pPC11. Figure 7.3.6 below shows the results of the verifying restriction analysis of plasmid pPC11.

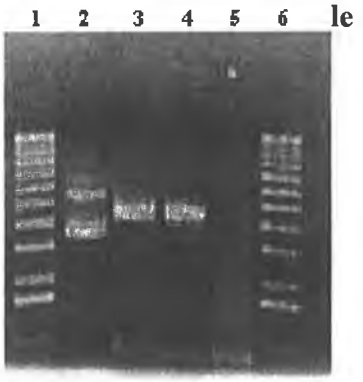
The *int280C/Aα* PCR product was cloned into the TA pCR2.1 vector in exactly the same way and the desired plasmid, called pPC12, was identified by first performing *Eco*RI and then *Bam*HI-*Hind*III double digests on plasmid DNA isolated from individual transformants. Figure 7.3.7 below shows the results of the verifying restriction analysis of plasmid pPC12.

- **Figure 7.3.6: Restriction Analysis Of Plasmid pPC11.**

Lane No.	Sample
1	1Kb Ladder
2	pPC11 uncut
3	pPC11 cut *Eco*RI
4	pPC11 cut *Bam*HI-*Hind*III
5	*int280α* PCR Product
6	1Kb Ladder

- **Figure 7.3.7: Restriction Analysis Of Plasmid pPC12.**

Lane No.	
1	1Kb Ladder
2	pPC12 uncut
3	pPC12 cut *Eco*RI
4	pPC12 cut *Bam*HI-*Hind*III
5	*int280C/Aα* PCR Product
6	1Kb Ladder



7.3.3: Sub-cloning Of The *int280α* PCR Product From Plasmid pPC11 Into The pPC229-A Prey Plasmid To Form Plasmid pPC233.

The *int280α* PCR product was sub-cloned from pPC11 into the pPC229-A Prey Plasmid to form plasmid pPC233. Plasmid pPC11 was digested with *Bam*HI-*Hind*III and the restricted plasmid DNA was separated on a 0.7% agarose gel. The 855bp band corresponding to the excised *int280α* *Bam*HI-*Hind*III fragment was purified from the gel by the gene clean procedure. The pPC229-A Prey plasmid was also digested with *Bam*HI-*Hind*III and the digested plasmid DNA was separated on a 0.7% agarose gel to verify complete digestion. The 6kb band corresponding to the linearised plasmid DNA was purified from the gel by the gene clean procedure and ligated with the purified *int280α* *Bam*HI-*Hind*III fragment. The ligated DNA was used to transform DH5α and the transformation was plated on LBamp media. Clones obtained were screened by performing *Bam*HI-*Hind*III double digests on plasmid DNA isolated from each clone. Screening in this manner resulted in the identification of the desired plasmid pPC233 and Figure 7.3.8 shows the results of the verifying restriction analysis of plasmid pPC233.

- **Figure 7.3.8: Restriction Analysis Of Plasmid pPC233.**

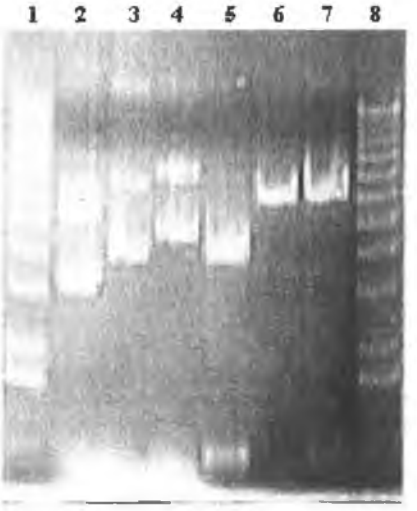
Lane No.	Sample
1	1Kb Ladder
2	pPC11 uncut
3	pPC229-A uncut
4	pPC233 uncut
5	pPC11 cut *Bam*HI-*Hind*III
6	pPC229-A cut *Bam*HI-*Hind*III
7	pPC233 cut *Bam*HI-*Hind*III
8	1Kb Ladder

7.3.4: Sub-cloning Of The *int280C/Aα* PCR Product From Plasmid pPC12 Into The pPC229-A Prey Plasmid To Form Plasmid pPC234.

The *int280C/Aα* PCR product was sub-cloned from pPC12 into the pPC229-A Prey Plasmid to form plasmid pPC234. Plasmid pPC12 was digested with *Bam*HI-*Hind*III and the restricted plasmid DNA was separated on a 0.7% agarose gel. The 855bp band corresponding to the excised *int280C/Aα Bam*HI-*Hind*III fragment was purified from the gel by the gene clean procedure. The pPC229-A Prey plasmid was also digested with *Bam*HI-*Hind*III and the digested plasmid DNA was separated on a 0.7% agarose gel to verify complete digestion. The 6kb band corresponding to the linearised plasmid DNA was purified from the gel by the gene clean procedure and ligated with the purified *int280C/Aα Bam*HI-*Hind*III fragment. The ligated DNA was used to transform DH5α and the transformation was plated on LBamp media. Clones obtained were screened by performing *Bam*HI-*Hind*III double digests on plasmid DNA isolated from each clone. Screening in this manner resulted in the identification of the desired plasmid pPC234 and Figure 7.3.9 shows the results of the verifying restriction analysis of plasmid pPC234.

- **Figure 7.3.9: Restriction Analysis Of Plasmid pPC234.**

Lane No.	Sample
1	1Kb Ladder
2	pPC12 uncut
3	pPC229-A uncut
4	pPC234 uncut
5	pPC12 cut *Bam*HI-*Hind*III
6	pPC229-A cut *Bam*HI-*Hind*III
7	pPC234 cut *Bam*HI-*Hind*III
8	1Kb Ladder



7.4: Performing The P2H Assay For The *In vivo* Detection Of Tir Interactions With Int280α And Int280C/Aα.

Having constructed the plasmids pPC190 (expressing a Tir-NifA DBD hybrid protein), pPC233 (expressing a NifA AD-Int280α hybrid protein) and pPC234 (expressing a NifA AD-Int280C/Aα hybrid protein) they were used to perform a P2H system assay in order to determine if the interactions between Tir with Int280α or Tir with Int280C/Aα could be detected in the newly constructed NifA based system. As already described earlier these interactions had been detected in the Y2H system by *Hartland et al (1999)* and it was therefore expected that these interactions would be detected in the NifA based P2H system. In order to perform the NifA based P2H assay the combinations of plasmids shown in Figure 7.4.1 were transformed into the *E. coli* strain ET6016*rif*F' and plated on appropriate selective media. IPTG was added to the media to a final concentration of 2mM to induce expression of hybrid proteins as this level of IPTG had been determined to be the optimal concentration during the optimization experiments described in chapter 6.

- **Figure 7.4.1: Combinations Of Plasmids Transformed Into ET6016*rif*F' To Perform The P2H Assay.**

pPC185-A	pPC187-A	pPC190	**LB chloramphenicol / tetracycline Xgal, IPTG**
pPC185-A / pPC228	pPC187-A / pPC228	pPC190 / pPC228	**LB chloramphenicol / tetracycline ampicillin Xgal, IPTG**
pPC185-A / pPC229-A	pPC187-A / pPC229-A	pPC190 / pPC229-A	
pPC185-A / pPC233	pPC187-A / pPC233	pPC190 / pPC233	
pPC185-A / pPC234	pPC187-A / pPC234	pPC190 / pPC234	

pPC185-A = *nifH:lacZα* reporter plasmid
pPC187-A = nifH:lacZα Bait plasmid
pPC228 = Control plasmid
pPC229 = Prey plasmid
pPC233 expresses NifA AD-Int280α
pPC234 expresses NifA AD-Int280C/Aα

Plates were incubated at 30 °C for 72 hours after which the colour intensity of colonies on each of the plates was compared and then graded on a visual basis. The results obtained are given in Table 7.4.1 below.

- **Table 7.4.1: Colony Colour Intensities Obtained For P2H Assay.**

Plasmids	*Colony Colour Intensity*	*Expected Results*
pPC185-A	-	-
pPC185-A / pPC228	++++	++++
pPC185-A / pPC229-A	+	+
pPC185-A / pPC233	+	+ (same as pPC229-A)
pPC185-A / pPC234	++++	+ (same as pPC229-A)
pPC187-A	-	-
pPC187-A / pPC228	++++	++++
pPC187-A / pPC229-A	+	+
pPC187-A / pPC233	+	+ (same as pPC229-A)
pPC187-A / pPC234	++++	+ (same as pPC229-A)
pPC190	-	-
pPC190 / pPC228	++++	++++
pPC190 / pPC229-A	+	+
pPC190 / pPC233	+	++++
pPC190 / pPC234	++++	++++

As can be seen from Table 7.4.1 the results obtained were not as expected. No interaction was detected between the NifA AD-Int280α hybrid protein (expressed from pPC233) and Tir-NifA DBD hybrid protein (expressed from pPC190). In addition to this the NifA AD-

Int280C/Aα hybrid protein (expressed from pPC234) generated a strong positive result even in the absence of the Tir-NifA DBD hybrid protein. The possible reasons for these results are discussed in Chapter 8.

Chapter 8

Discussion.

8.1: Discussion.

The purpose of this work was to construct a novel prokaryotic two-hybrid (P2H) system for the detection and analysis of protein-protein interactions *in vivo*. The system would function in *E. coli* and was to be constructed using the DNA binding and transcriptional activation domains of the *S. meliloti* 2011 NifA protein. Such a system would be a valuable addition to the ever growing range of methods for the detection of molecular interactions *in vivo* of which the yeast two-hybrid system (Y2H) is currently most powerful.

8.1.1: Benefits Of A Prokaryotic Two-Hybrid System.

A P2H system would facilitate the analysis of prokaryotic protein-protein interactions in a prokaryotic genetic background. Such a system could also be of use for the analysis of eukaryotic proteins in situations where homologous yeast proteins interfere with the interactions of the proteins under investigation. In addition to this manipulation of bacterial DNA is generally easier than that of yeast and much higher transformation efficiencies are attainable in *E. coli*. This facilitates the rapid and simple screening of large complex libraries of protein to identify putative binding partners for a given protein of interest. This potential makes prokaryotic systems particularly attractive for the purpose for the construction of proteomic linkage maps.

It may be argued that a prokaryotic based system would be of limited use for the analysis of eukaryotic proteins on the basis that such proteins may not be stable or be capable of folding correctly in a prokaryotic background. Also some eukaryotic proteins require post-translational modifications such as phosphorylation or glycosylation in order to fold and function correctly and thus would not function in *E. coli*. However, it is important to note that limitations such as these are also encountered in the conventional Y2H system where hybrid proteins are targeted to and interact in the nucleus and thus are not subject to post-translational modifications such as glycosylation and disulfide bond formation that occur in the endoplasmic reticulum. Proteins that require phosphorylation or acetylation by non-yeast proteins also fail to interact in the conventional Y2H system.

8.1.2: Overview Of The S. meliloti 2011 *NifA Protein - Reasons For Its Selection For The Construction Of A Novel P2H System.*

The novel P2H system constructed in this work is based on the transcriptional activating protein NifA from *S. meliloti* 2011. NifA activates the expression of several *nif* and *fix* genes in response to cellular oxygen and/or nitrogen status. It belongs to a large family of bacterial Enhancer Binding Proteins (EBP's) that activate transcription from promoters requiring the alternate sigma factor σ^{54} (NtrA or RpoN).

The RNA polymerase containing σ^{54} (σ^{54}-holoenzyme or $E\sigma^{54}$) is capable of binding to distinct promoter sequences to form transcriptionally closed complexes but is unable to initiate transcription unless an EBP, such as NifA, is present. NifA activates transcription by binding to sites characterized by the palindromic motif TGT-N_{10}- ACA that are located 100 to 200bp upstream of *nif* promoters and contacts $E\sigma^{54}$ bound to such promoters by means of DNA loops or kinks. In many instances the formation of the DNA loop is assisted by the DNA bending protein Integration Host Factor (IHF) which binds at a site located between the NifA and $E\sigma^{54}$. NifA activates transcription by catalyzing the isomerization of transcriptionally inactive closed complexes between $E\sigma^{54}$ and promoter sequences to transcriptionally active open complexes in a reaction that depends upon hydrolysis of a nucleoside triphosphate by NifA. Other EBP's, such as NtrC and DctD, activate transcription of their associated promoters by a similar mechanism.

NifA, as well as the other EBP's, is a three domain protein. Like many eukaryotic transcription factors, many EBP's are modular in nature. This simply means that the domains of the proteins are functionally independent of one another and can be physically separated from each other while still retaining their functionality. In the case of NifA, the central domain of about 220 amino acids is responsible for the activation of transcription. This domain is highly conserved among all of the EBP's. The function of the N-terminal domain of NifA is unknown whereas the C-terminal domain contains a helix-turn-helix motif that is required for binding to the UAS.

The modular nature of the EBP's and the functional separation of DNA binding and transcriptional activation activities, makes them potential candidates for the development of P2H systems. The *S. meliloti* 2011 NifA protein was chosen for this purpose because it is one of the best studied EBP's and because unlike the other EBP's it

is synthesized as a completely functional protein. Other EBP's are synthesized in an inactive form and require activation through the activity of another protein in response to certain environmental stimuli. For example, NtrC requires phosphorylation of its N-terminal domain by NtrB before it is capable of activating the expression of σ^{54} dependent genes concerned with nitrogen assimilation. This phosphorylation only occurs under nitrogen limiting conditions. The requirement for specific environmental conditions (such as nitrogen starvation in the case of NtrC) for activation of most of the EBP's would complicate the application of any P2H system constructed from them.

The *S. meliloti* 2011 NifA protein was also chosen because unlike NtrC it is completely heterologous to other proteins found in *E. coli* (i.e. there is no equivalent NifA protein in *E. coli* as *E. coli* does not carry out nitrogen fixation). There is a homologous NtrC protein in *E. coli* encoded by the *glnG* gene in the *glnALG* operon. NifA could therefore be utilised in *E. coli* without substantially interfering with normal *E. coli* cellular processes.

Thus NifA was selected as the most suitable of all of the studied EBP's for the construction of a novel P2H system that would function in *E. coli*. The DNA sequences for the transcriptional activating N-terminal-central (NC) domain and the DNA binding domain (DBD) of the NifA protein were amplified from *S. meliloti* 2011 genomic DNA and used for the construction of the P2H system Prey and Bait plasmids respectively. The sequence for the NifA dependent *nifH* promoter (the promoter for the *nifHDK* operon in *S. meliloti* 2011) was also amplified from *S. meliloti* 2011 genomic DNA by PCR and used for the construction of two reporter genes for the P2H system, one *nifH:lacZα* reporter gene and one *nifH:gent*^{r} reporter gene. In addition to the construction of the Prey plasmids, Bait plasmids and reporter genes for the P2H system, suitable *E. coli* strains had to be developed in which the P2H assay would be performed. The reason for this was that the product of the *glnG* gene in *E. coli*, NtrC, had been observed to weakly activate the same promoters as NifA. Although this activation was weak it would contribute to false positives when performing P2H assays and so strains of *E. coli* in which the *glnG* gene had been deleted had to be obtained and developed for use as host strains for the P2H system.

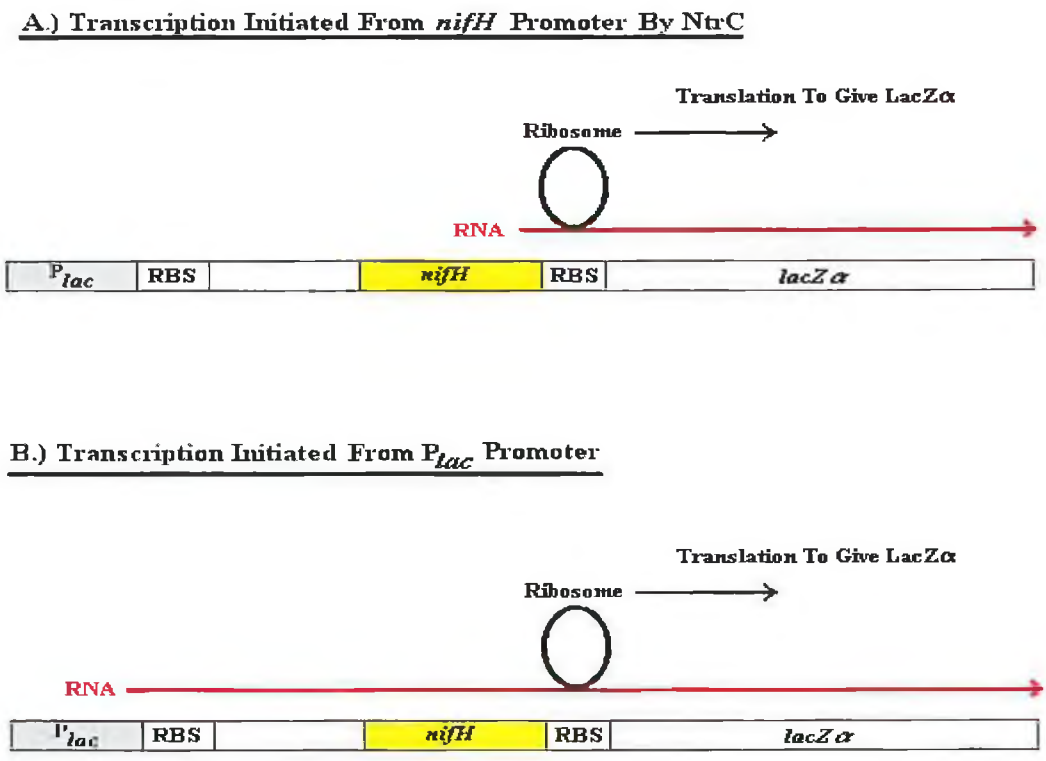
8.1.3: Construction Of The Reporter Genes For The NifA Based P2H System.

Two reporter genes were constructed for the novel NifA based P2H system, one *nifH:lacZα* reporter gene and one *nifH:gent*^{r} reporter gene. As described in chapter 4 the *nifH:lacZα* reporter gene was constructed first as it would be of greater use for the optimization of the P2H system.

Several interesting observations were made during the construction of the *nifH:lacZα* reporter gene. When the amplified *nifH* promoter was cloned into the multiple cloning site of pUC19 to make an in frame fusion between it and the downstream *lacZα* sequence (see Figure 4.3.4) DH5α colonies containing the desired plasmid pPC19 were found to produce a very intense dark blue colour on Xgal plates. There where two possible reasons for the colour development. Since DH5α did not have a deletion of the *glnG* gene the reason for the intense blue colour development may have been due to *E. coli* NtrC activation of the *nifH:lacZα* reporter fusion present on pPC19 (see Figure 8.1a). Although the level of activation of *nifH* by NtrC was reported to be low it was amplified by the high copy number of pUC19 to a level easily detectable on Xgal plates. Another possible reason for the intense colour development was that transcription of the lacZα sequence was being initiated from the upstream *plac* promoter. Transcripts initiated at this point would first read through the *nifH* promoter sequence and then into downstream *lacZα* sequence. A functional LacZα peptide would not be translated from the resulting RNA transcript using the ribosome binding site (RBS) and start codon present in the *plac* promoter because of disruption of the reading frame by the intervening cloned *nifH* promoter sequence. However a functional LacZα peptide could be translated from RNA transcripts using the RBS and start codon of the cloned *nifH* promoter since it was cloned into the MCS of pUC19 so that its start codon would be in frame with the downstream *lacZα* sequence (see Figure 8.1b).

When the *nifH:lacZα* reporter gene was amplified by PCR and cloned into the high copy number TA pCR2.1 vector all of the colonies obtained when the transformation was plated out on selective Xgal were dark blue in colour. This observation supported the theory that the intense blue colour development obtained for colonies containing plasmid pPC19 was caused by NtrC activation of transcription from the *nifH* promoter of the *nifH:lacZα* fusion on the plasmid rather than as a result of transcription initiated

- **Figure 8.1: Two Possible Mechanisms For The Expression Of Functional LacZα From pPC19.**



from the upstream *plac* promoter. If the latter had been the case an equal number of white and blue colonies would have been obtained upon cloning of the *nifH:lacZ*α PCR product into the TA pCR2.1 vector since the *nifH:lacZ*α PCR product could insert into the pCR2.1 plasmid in either orientation with respect to the *plac* promoter on the pCR2.1 plasmid. In one orientation transcripts initiated at the *plac* promoter would contain the coding sequence of the *nifH:lacZ*α reporter gene and functional LacZα peptide would be translated to from the *nifH* derived RBS and start codon to generate intense blue colour development on Xgal plates in the same way as that described for pPC19. In the other orientation transcripts initiated at the *plac* promoter would contain the nonsense strand of the *nifH:lacZ*α reporter and thus no functional LacZα would be produced and white colonies would be the result. NtrC activation of the *nifH:lacZ*α reporter gene would be independent of its orientation and thus all colonies would appear blue. While this appeared to be the case it could not be ruled out that both transcription initiated from the *nifH* promoter as a result of NtrC activation and transcription initiated

from the *plac* promoter on pPC19 and pCR2.1 contributed to the expression of functional LacZα and the intense blue colour development observed on Xgal plates. However the extreme high copy number of the TA pCR2.1 vector made it impossible to differentiate between colonies in which both processes contributed to functional LacZα expression (when the *nifH:lacZα* PCR product was cloned in the correct orientation with respect to the *plac* promoter on the pCR2.1 vector) and colonies in which only NtrC activated transcription from the *nifH:lacZα* reporter contributed to function LacZα expression (when the *nifH:lacZα* PCR product was cloned in the opposite orientation). This fact became clear when the *nifH:lacZα* reporter gene was sub-cloned from the high copy number TA pCR2.1 vector into the much lower copy number pPC184 (pACYC184 with the unique *Bam*HI site in the tetracycline resistance gene eliminated) plasmid to form the reporter plasmids pPC185-A and pPC185-B (see Figure 4.3.13).

The *nifH:lacZα* reporter was cut from plasmid pPC2 (pCR2.1 vector with the *nifH:lacZα* PCR product cloned into it) as a *Hind*III fragment and ligated with *Hind*III restricted and dephosphorylated pPC184 plasmid DNA. The ligation was used to transform DH5α and the transformation was plated on selective media to which Xgal had been added. The reason for the addition of the Xgal was to screen for NtrC activation of the *nifH:lacZα* reporter gene and thus to help detect colonies containing plasmids that had the *Hind*III cloned *nifH:lacZα* reporter gene fragment. Interestingly two classes of colonies were obtained on plates, some being very light blue in colour while others were slightly darker. Because the *nifH:lacZα* reporter was not directionally cloned into pPC184 it could insert into the plasmid in either of two orientations and upon identification of the desired reporter plasmids pPC185-A and pPC185-B it was discovered that pPC185-A generated the very light blue colonies while plasmid pPC185-B generated the slightly darker blue colonies. This finding supported the earlier theory that the blue colour development observed for plasmid pPC19 and when the *nifH:lacZα* reporter was cloned into the TA pCR2.1 vector was probably due to a combination of transcription initiated from both the *plac* promoter and as a result of NtrC activation from the *nifH* promoter. It was determined that plasmid pPC185-A produced very light blue colonies on Xgal plates as a result of weak NtrC activation of the *nifH:lacZα* reporter. The pPC185-B plasmid produced a slightly darker blue colour intensity because in addition to the weak NtrC induced activation of the *nifH:lacZα*

there was also some transcription initiated from the upstream tetracycline resistance gene promoter (see Figure 4.3.13). The *Hind*III site into which the *nifH:lacZα* reporter gene was cloned lies within the promoter for the tetracycline resistance gene and cloning into this site does not always result in inactivation of the tetracycline promoter. As described earlier, functional LacZα was translated from the *nifH* derived RBS within the resulting transcripts initiated from the tetracycline promoter (see Figure 8.1b).

In addition to the *nifH:lacZα* reporter gene a *nifH:gent*^{r} reporter gene was also constructed and cloned as a *Hind*III fragment into the *Hind*III site of plasmid pPC184 to form the reporter plasmids pPC186-A and pPC186-B. In order to generate the final Bait plasmids for the P2H system a Bait construct was constructed using the C-terminal DBD of NifA (described in Chapter 5). It consists of a *ptac* promoter followed by a MCS into which the sequence for a protein of interest can be cloned. C-terminally to the MCS is the sequence for the DNA binding domain of NifA. Cloning into the MCS results in the expression of a hybrid protein bearing the DNA binding domain of NifA. This construct was directionally cloned as a *Eag*I-*Sal*I into each of the reporter plasmids pPC185-A, pPC185-B, pPC186-A and pPC186-B to form the final Bait plasmids pPC187-A, pPC187-B, pPC188-A and pPC188-B respectively (see Figure 5.2.13 for the *nifH:lacZα* Bait plasmids pPC187-A and pPC187-B and see Figure 5.2.17 for the *nifH:gent*^{r} Bait plasmids pPC188-A and pPC188-B).

8.1.4: Construction Of The Prey Plasmids For The NifA Based P2H System.

Three Prey plasmids, pPC229-A, pPC229-B and pPC229-C, were constructed (see Figure 3.5.1) using the sequence for the N-terminal domain-central domain (NC) portion of NifA. In these plasmids the NC portion of the NifA protein is expressed from a strong *ptac* promoter. C-terminally to the sequence for the NC portion of NifA is a multiple cloning site (MCS) into which a library of DNA fragments can be cloned to generate a library of hybrid proteins bearing the activation domain of NifA. The NC portion of the NifA protein was used rather than just the central activating domain of the NifA protein by itself because the N-terminal domain of the protein had been shown to stabilize the protein when expressed in *E. coli*. The NC domain of NifA was also

amplified from a second internal start codon rather than from its natural start codon because proteins expressed from this start codon had been demonstrated to be more stable in *E. coli*. The three Prey plasmids differ from each other only in the reading frame of the multiple cloning site relative to the upstream sequence for the NC portion of NifA. The reason for constructing three Prey plasmids was to facilitate the easy construction of libraries of hybrid proteins. Another related Control plasmid expressing the whole NifA protein from the *ptac* promoter, pPC228, was also constructed (see Figure 3.5.1). This plasmid was constructed to be used as a positive control when performing P2H assays.

When expression of the NifA protein from plasmid pPC228 was tested using the *nifH:lacZα* reporter plasmid pPC185-A very poor colour development was observed. As described earlier in Chapter 3, it was felt that this was due the ATG start codon on the cloned *nifA* gene sequence being too far from the plasmid borne RBS to facilitate efficient translation of NifA from RNA transcripts initiated at the *ptac* promoter. It was therefore decided that the *ptac* promoter region of plasmid pPC228 would be manipulated to reduce the distance between the *nifA* start codon and the plasmid borne RBS in the hope that this would result in improved translation of the NifA protein. The manner in which this manipulation was carried out is described in detail in Chapter 5 and resulted in the formation of the alternative Control plasmid pPC230. When this Control plasmid was introduced into *E. coli* with the *nifH:lacZα* reporter plasmid pPC185-A it was found to produce more intense colour development on Xgal plates. As the distance between the start codon and the plasmid borne RBS on each of the constructed Prey plasmids pPC229-A, pPC229-B and pPC229-C was the same as it was on pPC228 the *ptac* promoter region of each of these plasmids was also manipulated to generate three alternative Prey plasmids, pPC231-A, pPC231-B and pPC231-C respectively (see Figure 3.5.1).

8.1.5: Development Of Suitable Host Strains And Optimization Of Growth Conditions For The NifA Based P2H System.

In addition to the construction of the plasmids for the P2H system a suitable strain of *E. coli* had to be developed in which the plasmids could be used. This strain had to have

the *glnG* gene deleted since the product of this gene, NtrC, had been observed to activate the *nifH* promoter and would thus give rise false positive. Such activation by *E. coli* NtrC was observed during the construction of the *nifH:lacZα* reporter construct. Two *glnG* mutant strains of *E. coli* were obtained, YMC11 and ET6016, but unfortunately both of these lacked the correct genetic background for *lacZα* complementation which the *nifH:lacZα* reporter system requires. The correct genetic background was introduced into the *E. coli* strains on an F' factor by conjugation with XL1-Blue as described in Chapter 6 and two suitable *E. coli* host strains, YMC11*rif*F' and ET6016*rif*F', were thus obtained for the NifA based P2H system.

Having constructed two *E. coli* host strains, ET6016*rif*F' and YMC11*rif*F', in which the constructed plasmids for the NifA based P2H system could be used, the next step was to assess the usefulness of all of the constructed plasmids in each of the host strains and to optimize the growth conditions under which the P2H assay would be carried out in each of the two strains. The manner in which these optimization experiments were carried out and the results obtained are described in Chapter 6 (see section 6.3). These experiments yielded some strange and unexpected results which are discussed below in relation to the strain ET6016*rif*F'. Results obtained for the alternative YMC11*rif*F' strain mirrored those obtained for the ET6016*rif*F' strain.

The first unexpected result was obtained when the *nifH:lacZα* reporter plasmids pPC185-A and pPC185-B and the final *nifH:lacZα* Bait plasmids pPC187-A and pPC187-B were introduced into ET6016rifF' alone. This *E. coli* strain was known to have a deletion of its *glnG* gene and thus should not express any functional NtrC. It was therefore expected that when the reporter plasmids pPC185-A and pPC185-B, were introduced into ET6016*rif*F' that pPC185-A would not produce any colour development on Xgal plates and that pPC185-B would produce a significantly reduced level of colour development compared to that observed in DH5α (which has an intact *glnG* gene) if any colour development at all. The final Bait plasmids pPC187-A and pPC187-B were expected to yield the same results since they differed from their related reporter plasmids only in that they possessed the additional Bait construct. However, similar levels of colour development to those observed in DH5α were obtained for plasmids pPC185-A and pPC185-B indicating that the observed colour development in DH5α was not the result of NtrC induced transcription from the *nifH* promoter of the *nifH:lacZα* reporter but rather the result of leaky baseline transcription from the *nifH*

promoter. Such leaky transcription from σ^{54} promoters has been reported to be caused by transient melting of DNA and is enhanced by conditions that favour DNA melting, such as DNA supercoiling, elevated temperatures, and lower ionic strengths (*Wang et al, 1997*). The final Bait plasmids pPC187-A and pPC187-B generated colony colour intensities equivalent to their corresponding related reporter plasmids (pPC185-A and pPC185-B respectively) as expected. The colour intensity of colonies generated by pPC187-B was stronger than that observed for pPC187-A and as described earlier in section 8.1.3 (see Figure 8.1b) this was due to initiation of transcription from the tetracycline promoter reading through the *nifH:lacZα* reporter (see Table 6.3.1). Due to the high level of background expression exhibited by plasmids pPC185-B and pPC187-B these plasmids were quickly eliminated from further optimization experiments.

The NC domain of the NifA protein was known to be capable of activating transcription from the *nif*H promoter despite lacking the C-terminal DBD of the protein. This transcriptional activation by the NifA NC domain occurs from solution (rather than from non-specific interaction with DNA sequences) and the level of expression from the *nif*H promoter induced by the NifA NC domain was known to be lower than that observed for the whole NifA protein. It was hoped that by using the Control plasmid pPC228 (expressing the whole NifA protein) as a positive control and the Prey plasmid pPC229-A (expressing the NC domain of NifA) as negative control that growth conditions could be manipulated to optimize the contrast between the colony color intensities given by the positive pPC228 and negative pPC229-A controls on Xgal plates. Growth conditions were to be optimized in the same way for the alternative Control and Prey Plasmids (i.e. those that had manipulated *ptac* promoter regions to reduce the distance between the RBS and the start codon of the cloned *nifA* and NifA NC domain sequences), pPC230 (expressing the whole NifA protein) and pPC231-A (expressing the NifA NC domain), with each of the Bait plasmids pPC187-A and pPC187-B and their usefulness compared with that of the pPC228/pPC229-A combination of Control/Prey plasmids. In this way the most useful Control/Prey plasmid combination and the most useful Bait plasmid would be determined along with the growth conditions under which they gave the best results.

Several approaches were taken to try and optimize the contrast between positive pPC228/pPC230 and negative pPC229-A/pPC231-A controls. Firstly it was known that the *S. meliloti* 2011 NifA protein was intrinsically sensitive to oxygen. Oxygen inhibits

the ability of NifA to both bind to its cognate DNA binding site and to initiate transcription. This is in contrast to the NifA protein from *Klebsiella pneumoniae* which is not sensitive to oxygen and inactivation of this NifA protein under aerobic conditions is entirely dependent on the activity of the *K. pneumoniae* NifL protein. However, the inhibition of *S. meliloti* 2011 NifA induced transcriptional activation under aerobic conditions is not complete. The reason why the *S. meliloti* 2011 NifA was chosen for use in the construction of the P2H system rather than the *K. pneumoniae* NifA protein was that it was hoped that the oxygen sensitivity of the protein could be used to advantage. It was hoped that performing the P2H assay under aerobic conditions would minimize the strength of induced *nifH:lacZα* activation from solution by the NifA NC domain.

A second approach involved initial incubation of plates at 37 °C for 24 hours followed by incubation at 30 °C for a further 24hours. The theory behind this approach is discussed below in relation to the initially constructed Control and Prey plasmids pPC228 and pPC229-A. NifA is temperature sensitive working optimally at a temperature of 28 °C and is not active at 37 °C. IHF also plays an important role in the activation of the *nifH* promoter by NifA by mediating DNA bending to bring NifA bound to its cognate DNA binding sequence upstream of the promoter into contact with $E\sigma^{54}$ bound at the promoter. Levels of IHF within the cell vary according to growth phase and there is a 10 fold increase in IHF levels during stationary phase. It had been noticed in initial plate tests that colour development of colonies containing plasmid pPC228 (expressing the whole NifA protein) and pPC187-A (the *nifH:lacZα* reporter plasmids) on Xgal plates was very slow requiring 72 hours at 30 °C for significant colour development to occur. It was also noted that colour development progressed from the centre of colonies out towards the edges of the colonies. It was therefore thought that initiation of *nifH:lacZα* transcription by the whole NifA protein occurred primarily during stationary phase of growth due to increased levels of IHF upon which activation by bound NifA depends. In contrast the NifA NC domain (expressed from plasmid pPC229-A) cannot bind DNA and activates transcription of the *nifH:lacZα* reporter on pPC187-A from solution and is therefore not dependent on IHF to activate transcription. There would therefore be a continuous accumulation of colour development of colonies containing the Prey plasmid pPC229-A (expressing the NifA NC domain) and pPC187-A throughout the growth of colonies on plates when plates

were incubated at 30 °C. On the other hand, significant induction of the *nifH:lacZα* reporter gene on pPC187-A by the whole NifA (expressed from pPC228) would not occur until colonies started to enter into stationary phase. It was hoped that accumulating colony colour development produced by NifA NC domian (expressed from pPC229-A) would be prevented by an initial incubation at 37 °C for 24 hours. Incubation at this temperature would promote the fast growth of colonies and thus the stationary phase of growth would be achieved more rapidly than it would using incubation at 30 °C. At the same time any expressed NifA NC domain would be kept inactive and prevented from inducing expression of the *nifH:lacZα* reporter. Activation by of the *nifH:lacZα* reporter by the whole NifA (expressed from pPC228) would also be prevented at this temperature. However, after 24 hours at 37 °C plates would be switched to 30 °C for a further 24 hours to permit colony colour development. Since colonies on the plates would be entering into stationary phase after the initial incubation at 37 °C it was hoped that any newly synthesized NifA would readily activate the *nifH:lacZα* reporter to produce strong colony colour development. Obviously any newly synthesized NifA NC domain would also begin to activate expression of the *nifH:lacZα* reporter but at a much lower level than that of NifA. It was hoped that by suppressing activation of *nifH:lacZα* expression by the NifA NC domain until colonies had entered into the stationary phase that the contrast between the colour intensity of colonies on positive control plates (containing plasmids pPC228 and pPC187-A) and colonies on negative control plates (containing plasmids pPC229-A and pPC187-A) would be enhanced. Unfortunately this was not to be the case as can be seen from the results presented in Chapter 6. In general, sets of plates incubated at 37 °C for 24 hours followed by incubation at 30 °C for another 24 hours mirrored those obtained for sets of plates incubated at 30 °C for 72 hours. The only desirable effect obtained by using an initial incubation at 37 °C for 24 hours was that results were obtained more quickly.

A third approach used to try and optimize the contrast between positive pPC228/pPC230 and their corresponding negative pPC229-A/pPC231-A controls involved the manipulation of IPTG levels used for inducing expression of the whole NifA and the NifA NC domain. As described in Chapter 6 it was expected that lower levels of induction would produce the best contrast between positive pPC228/pPC230 controls and their corresponding pPC229-A/pPC231-A negative controls. The theory behind this was that the NifA NC domain activated transcription from the *nifH* promoter

from solution as it did not possess a C-terminal DBD. The level to which NifA NC domain induced expression from the *nifH* promoter was therefore expected to be dependent on its concentration in the cell and thus the lower its concentration in the cell the lower the level of induced expression from the *nifH* promoter. On the other hand, the level of expression induced from the *nifH* promoter by the whole NifA protein was not expected to be as dependent on its concentration in the cell due to its ability to bind at the *nifH* promoter. This binding activity would be expected to increase the local concentration of the whole NifA protein in the vicinity of the *nifH* promoter. Results obtained were completely opposite to what was expected. It was found that increasing the concentration of IPTG from 0.5mM to 2mM caused a decrease in the colour intensity of colonies containing either pPC185-A or pPC187-A plasmids on their own. This was mirrored by a decrease in the colour intensity of colonies containing pPC229-A with pPC187-A allowing a clear distinction to be made between them and colonies containing plasmids pPC228 with pPC187-A. It is believed that the induced higher levels of transcription of the *nifA* gene and the sequence for the NifA NC domain from strong *ptac* promoters on plasmids pPC228 and pPC229-A respectively sequesters of core-RNA polymerase away from σ^{54}. Although one might argue that such sequestering of core RNA polymerase into involvement in the transcription of *nifA* and NifA NC domain sequences might have a more general effect on bacterial cells causing them to grow more slowly it must be noted that σ^{54} is present in cells at much lower levels than σ^{70} and that it interacts with core RNA polymerase relatively weakly in comparison with σ^{70}. Any sequestering of core RNA polymerase due to enhanced levels of transcription of *nifA* and NifA NC domain sequences would therefore have a more pronounced and noticeable effect on σ^{54}-dependent promoters like the *nifH* promoter of the *nifH:lacZα* reporter. Although NifA and NifA NC domain concentrations within the cell increase with increasing IPTG concentration as a result of higher levels of transcription there is a reduced amount of $E\sigma^{54}$ bound at the *nifH* promoter available for activation. Therefore there is a reduction in the amount of both leaky and induced transcription from the *nifH* promoter of the *nifH:lacZα* and a corresponding reduction in the colour intensity of colonies on Xgal plates. In support of this theory, increasing levels of NifA or NifA NC domain within the cell by increasing translational efficiency rather than by increasing levels of transcription does result in increased levels of induced *nifH:lacZα* expression. This can be seen by comparing the colour intensity of

colonies given by pPC230 with those given by pPC228 for any one level of IPTG. The colour intensity of colonies given by pPC230 with its manipulated *ptac* promoter region for more efficient translation of NifA is greater than that observed for pPC228 for the same level of IPTG.

One final approach to try and optimize the contrast between positive pPC228/pPC187-A and negative pPC229-A/pPC187-A controls involved sealing plates with parafilm. Due to the slow colony colour development observed when plates were incubated at 30 °C it was felt that incubation under anaerobic conditions simply would not be feasible. The theory behind wrapping plates in parafilm was that bacteria would initially have sufficient oxygen to grow and form visible colonies but that as they grew they would gradually utilise all of the available oxygen within the sealed agar plate. Micro-aerobic conditions, under which NifA works optimally, would ensue and it was hoped that NifA induced colony colour development on Xgal plates would occur more rapidly. Unfortunately wrapping plates in parafilm had the undesirable effect of enhancing the colour development of colonies containing only the pPC187-A plasmid to a level that did not allow a clear distinction to be made between the positive pPC228/pPC187-A and negative pPC229-A/pPC187-A controls. One possible reason for this increased colony colour development was not that there was increased expression of *lacZα* from the *nifH:lacZα* reporter on pPC187-A but that there was increased expression of the complementing β-galactosidase background from the F' factor in ET6016*rif*F'. This may have been caused as a result of rapid exhaustion of carbon sources due to anaerobic respiration followed by the CAP protein activating expression of the β-galactosidase complementing background from the F' factor. The increased concentration of complementing background within the cell would result in an increase in colony colour development.

The end result of the optimization experiments was the identification of the optimal set of constructed plasmids for the P2H system and the growth conditions under which they were to be used. The optimal set of plasmids consisted of the pPC228 Control plasmid, the pPC229-A Prey plasmid and the pPC187-A Bait plasmid. The optimal conditions were determined to involve growth on media containing 2mM IPTG and incubation at 30 °C for 72 hours under aerobic conditions. Having optimized the P2H system the next step was to test the ability of the system to detect protein-protein interactions *in vivo*.

8.1.6: Testing Of The NifA Based P2H System Using The Model Interaction Of The Tir and Intimin Proteins.

Having constructed the plasmids for the novel NifA based P2H system, developed suitable *E. coli* strains in which they could be used and optimized the growth conditions under which the assay would be performed the final step was to test the systems ability to detect protein-protein interactions *in vivo*. In order to do this a pair of proteins known to interact and that had been demonstrated to interact in the Y2H system were chosen. These proteins were the Tir and Intmin 280α (Int280α) proteins of enteropathogenic *E. coli*. As described in Chapter7, *tir* was cloned into the Bait plasmid pPC187-A to form plasmid pPC190 that would express it as a fusion with the DBD of NifA. The sequence for Int280α and a mutant form of the protein called Int280C/Aα were cloned into the Prey plasmid pPC229-A to form plasmids pPC233 and pPC234 respectively that express them as fusions with the NifA NC domain. Each of these plasmids was introduced into the host *E. coli* strain ET6016*rif*F' with plasmid pPC190 to see if an interaction between them and the Tir hybrid protein would be detected. Unfortunately the expected interaction between wild type Int280α and Tir was not detected.

There are several reasons as to why the expected interaction may not have been detected. The first of these is that although *tir* is expressed in enteropathogenic *E. coli* it is translocated to host epithelial cells where it becomes phosphorylated and then integrated into the host cell plasma membrane. The Tir protein may only fold correctly in eukaryotic cells or perhaps the phosphorylation of the protein in host cells is required for it to adopt a conformation capable of binding to Int280α. Consider that both the Tir and the Intimin proteins are expressed inside the same *E. coli* cell the very fact that Tir doesn't bind to Int280α inside *E. coli* might in itself be important since such binding would be detrimental to the cell. Another explanation for the failure to detect an interaction between the Tir and Int280α proteins might simply be because of occluson of the binding site on one of the proteins by NifA domains. If this is the case changing the polarity of the fusions i.e. fusing Int280α to the NifA DBD and Tir to the NifA AD, may overcome the problem and permit the detection of an interaction between Tir and Intimin. This problem is sometimes encountered in the Y2H system. It is important to note that due to the use of PCR for the amplification of *tir* and *int280α* sequences the

failure to detect an interaction between Int280α and Tir may have been due to mutations introduced in the sequence for either protein during their amplification.

An interesting result was obtained for the Int280C/Aα-NifA NC domain protein expressed from plasmid pPC234. This protein generated a strong positive result, stronger than that given by the whole NifA protein, even in the absence of the Tir hybrid protein. This indicated that the Int280C/Aα protein interacted non-specifically with some protein bound at the *nifH* promoter, perhaps with $E\sigma^{54}$ itself, to induce strong activation of the *nifH:lacZα* reporter gene. While the interaction detected was not the desired interaction with the Tir protein it does suggest that a protein-protein interaction can bring the NifA NC domain into proximity with the *nifH* promoter to activate transcription of the *nifH:lacZα* reporter gene. This shows that in principle the P2H system should work.

The fact that a positive interaction detected in the P2H system would generate a level of *nifH:lacZα* transcription exceeding that observed for the NifA protein is not surprising. As mentioned earlier the NifA protein is sensitive to oxygen which inhibits both its ability to bind to DNA and to activate transcription. Once separated from the remainder of the protein the C-terminal DNA binding domain of NifA is not oxygen sensitive and can bind to it cognate DNA sequence. Therefore any hybrid protein fused to the DBD of NifA would bind at the *nifH* promoter irrespective of oxygen levels. Activation of transcription is then dependent on interaction between the protein fused to the NifA DBD and another protein fused to the NifA NC domain and is not dependent on oxygen tension.

8.2: Conclusions And Future Prospects.

Although an interaction between the Tir and Int280α proteins was not detected in the novel NifA based P2H system there are numerous reasons as to why the interaction may not have been detected. However the results obtained for the Int280C/Aα hybrid protein do provide some encouragement that the system will work in theory. It may simply be the case that the model protein-protein interaction initially chosen to test the novel P2H system was not a suitable one although as stated above the fact that no interaction

between Int280α and Tir was detected might in itself by an important finding. Future work will obviously involve testing the system with other model protein-protein interactions in order to determine its worth. It will also involve testing and optimization of the *nifH:gent*^{r} which although constructed has not been subject to any testing as yet. The primary reason for this was that it was felt that until the P2H system had been sufficiently optimized to permit clear distinction between positive and negative controls that the *nifH:gent*^{r} would be of little use.

There are several ways in which the P2H system can be improved. First of all the NC domain of the *K. pneumoniae* which is not oxygen sensitive could be used. This would help to promote more rapid colour development under aerobic conditions. In addition to this the *nifH* promoter of *K. pneumoniae* could be used in place of the *S. meliloti* 2011 *nifH* promoter. It is now known that this promoter binds σ^{54} more weakly than the S. *meliloti* 2011 *nifH* promoter because it lacks a tract of thymine residues between -15 and -17. Because of this the concentration of bound $E\sigma^{54}$ at the promoter is lower and activation of transcription from the promoter from solution is thus lower. The promoter exhibits a greater dependency for bound NifA protein and so using this promoter rather than the *S. meliloti nifH* promoter would help to overcome the problems associated with activation of reporter genes from solution by the NifA NC domain. Using this promoter and the *K. pneumoniae* NifA NC domain should enhance the performance of the P2H system considerably. In addition to these changes additional binding sites for NifA might be incorporated upstream of the *nifH* promoter to improve the sensitivity of the system while MCS's could be expanded upon to make the plasmids more useful.

Assuming the basic NifA based P2H system described in this thesis can be shown conclusively to be capable of detecting protein-protein interaction *in vivo* there is obviously the potential for the development of a whole range of systems for detecting other types of molecular interaction such as prokaryotic one-hybrid (P1H) systems for the detection of protein-DNA interactions and prokaryotic reverse two-hybrid (PR2H) systems for detecting disruption of protein-protein interactions.

Bibliography

1. **Aelst, L.V., Barr, M., Marcus, S., Polverino, A., Wigler M. 1993.** Complex formation between RAS and RAF and other protein kinases. *Proc. Natl. Acad. Sci. U.S.A.* **88:** 9578-9582.

2. **Agron, P.G., Ditta, G.S. and Helinski, D.R. 1993.** Oxygen regulation of *nifA* transcription *in vitro. Proc. Natl. Acad. Sci. USA* **90**: 3506-3510.

3. **Allen, J.B., Walberg, M.W., Edwards, M.C. and Elledge, S.J. 1995.** Finding prospective partners in the library: the two-hybrid system and phage display find a match. *Trends In Biotechnology* **20**: 511-516.

4. **Aronheim, A., Zandi, E., Hennemann, H., Elledge, S.J. and Kurin, M. 1997.** Isolation of an AP-1 repressor by a novel method of detecting protein-protein interactions. *Mol. Cell Biol.* **17:** 3094-3102.

5. **Austin, S. and Dixon, R. 1992.** The prokaryotic enhancer binding protein NtrC has an ATPase activity which is phosphorylation and DNA dependent. *EMBO J.* **11**:2219-2228.

6. **Austin, S., Kundrot, C. and Dixon, R. 1991.** Influence of a mutation in the putative nucleotide binding site of the nitrogen regulatory protein NtrC on its positive control function. *Nucleic Acids Res.* **19**: 2281-2287.

7. **Bai, C. and Elledge, S.J. 1996.** Gene identification using the yeast two-hybrid system. *Methods Enzymol.* **273:** 331-347.

8. **Bartel, P.L. and Fields, S. 1995**. Analyzing protein-protein interactions using the two-hybrid system. *Methods Enzymol.* **254:** 241 - 263.

9. **Bartel, P.L., Roecklein, J.A., SenGupta, D. and Fields, S. 1996.** A protein linkage map of *Escherichia coli* bacteriophage T7. *Nat. Genet.* **12:** 72-77.

10. **Beringer, J.E. 1974.** R factor transfer in *Rhizobium leguminosarum. J. Gen. Micobiol.* **84:** 188-198.

11. **Beranger, F., Aresta, S., Gunzburg, J. and Camonis, J. 1997.** Getting more from the two-hybrid system: N-teminal fusions to LexA are efficient and sensitive baits for two-hybrid studies. *Nucleic Acids Res.* **25**: 2035-2036.

12. **Boder, E. and Wittrup, K.O. 1997.** Yeast surface display for screening combinatorial polypeptide libraries. *Nat. Biotechnology.* **15:** 553-557.

13. **Brent, R. and Finley, R.L.Jr. 1997.** Understanding gene and allele function with two-hybrid methods. *Ann. Rev. Genet.* **31:** 663-704.

14. **Brent, R. and Ptashne, M. 1985.** A eukaryotic transcriptional activator bearing the DNA binding specificity of a prokaryotic repressor. *Cell.* **43:** 729-736.

15. **Broder, Y.C, Katz, S. and Aronheim, A. 1998.** The *ras* recruitment system, a novel approach to the study of protein-protein interactions. *Curr. Biol.* **8**: 1121-1124.

16. **Buck, M. and Cannon, W. 1992a.** Activator independent formation of a closed complex between σ^{54}-haloenzyme and the *nifH* and *nifU* promoters *of Klebsiella Pneumoniae. Mol. Microbiol.* **6**: 1625-1630.

17. **Buck, M. and Cannon, W. 1992b.** Specific binding of the transcription factor sigma-54 to promoter DNA. *Nature* **358**:422-424.

18. **Buck, M., Cannon, W. and Woodcock, J. 1987.** Transcriptional activation of the Klebsiella pneumoniae nitrogenase promoter may involve DNA loop formation. *Mol. Microbiol.* **1**:243-249.

19. **Buck, M., Khan, H. and Dixon, R. 1985.** Site directed mutagenesis of the *Klebsiella pneumoniae nifL* and *nifH* promoters and *in vitro* analysis of promoter activity. *Nucleic Acids Res.* **13:** 7621 – 7638.

20. **Buck, M., Miller, S., Drummond, M. and Dixon, R. 1986.** Upstream activator sequences are present in the promoters of nitrogen fixation genes. *Nature* (London) 320: 374-378.

21. **Buikema, W.J., Szeto, W.W., Lemley, P.V., Orme-Johnson, W.H. and Ausubel, F.M. 1985.** Nitrogen fixation specific regulatory genes of *Klebsiella pneumoniae* and *Rhizobium meliloti* share homology with the general nitrogen regulatory gene *ntrC* of *K. pneumoniae*. *Nucleic Acids Res.* **13**: 4539-4555.

22. **Bunker, C.A. and Kingston, R.E. 1995.** Identification of a cDNA for SSRP1, an HMG-box protein, by interaction with the c-Myc oncoprotein in a novel bacterial expression screen. *Nucleic Acids Res.* **23**: 269-276.

23. **Cairns. M.T., Green, A.J., White, P.M., Johnston, P.G. and Brenner, S. 1997.** A novel bacterial vector system for monitoring protein-protein interactions in the cAMP-dependent protein kinase complex. *Gene* **185**: 5-9.

24. **Cannon, W. and Buck, M. 1992.** Central domain of the positive control protein NifA and its role in transcriptional activation. *J. Mol. Biol.* **225**: 271-286.

25. **Cannon, W., Claverie-Martin, F., Austin, S. and Buck, M. 1993.** Core RNA polymerase assists binding of the transcription factor σ^{54} to promoter DNA. *Mol. Microbiol.* 8: 287-298.

26. **Casaz, P. and Buck, M. 1997.** Probing the assembly of transcription initiation complexes through changes in σ^N protease sensitivity. *Proc. Natl. Acad. Sci. USA*. **94**: 12145 - 12150.

27. **Chen, Y.M., Bachman, K. and Magasanik, B. 1982.** Characterization of a gene, *glnL*, the product of which is involved in the regulation of nitrogen utilization in *Escherichia coli. J. Bacteriol.* **150**: 214-220.

28. **Chien, C.T., Bartel, P.L., Sternglanz, R. and Fields, S. 1991.** The two-hybrid system: A method to identify and clone genes for proteins that interact with a protein of interest. *Proc. Natl. Acad. Sci. USA* **88**: 9578-9582.

29. **Cho, R.J., Fromont-Racine, M., Wodicka, L., Feierbach, B., Stearns, T., Legrain, P., Lockhart, D.J. and Davis, R.W. 1998.** Parallel analysis of genetic selections using whole genome oligonucleotide arrays. *Proc. Natl. Acad. Sci. U.S.A.* **95:** 3752-3757.

30. **Claverie-Martin, F. and Magasanik, B. 1991.** Role of integration host facor in the regulation of the *glnHp2* promoter of *Escherichia coli. Proc. Natl. Acad. sci. USA* **88:** 1631-1635.

31. **Claverie-Martin, F. and Magasanik, B. 1992.** Positive and negative effects of DNA bending on activation of transcription from a distant site. *J. Mol. Biol.* **227**: 996-1008.

32. **Collado-Vides, J., Magasanik, B. and Gralla, J.D. 1991**. Control site location and transcriptional regulation in *Escherichia coli. Microbiol. Rev.* **55**: 371-394.

33. **Dixon, E.P., Johnstone, E.M., Liu, X. and Little, S.P. 1997.** An inverse mammalian two hybrid system for Beta secretase and other proteases. *Anal. Biochem.* **249:** 239-241.

34. **Dmitrova, M., Younes-Cauet, G., Oertel-Buchheit, P., Porte, D., Schnarr, M. and Granger-Schnarr, M. 1998.** A new LexA-based genetic system for monitoring and analyzing protein heterodimerization in *Escherichia coli. Mol. Gen. Genet.* **257**: 205-212.

35. **Drees, B.L. 1999.** Progress and variations in two-hybrid and three-hybrid technologies. *Curr. Opin. Chem. Biol.* **3:** 64-70.

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

37. **Drummond, M.H., Contreras, A. and Mitchenal, L.A. 1990.** The function of isolated domains and chimeric proteins constructed from the transcriptional activators NifA and NtrC of *Klebsiella pneumoniae. Mol. Microbiol.* **4**: 29-37.

38. **Durfee, T., Becherer, K., Chen, P-L., Yeh, S-H., Yang, Y. Kilburn, A.E., Lee, W-H. and Elledge, S.J. 1993.** The retinoblastoma protein associates with the protein phosphatase type I catalytic subunit. *Genes Dev.* **7:** 555-569.

39. **Dujon, B. 1996.** The Yeast Genome project: What did we learn? *Trends Genet.* **12:** 263-270.

40. **Evangelista, C., Lockshon, D. and Fields, S. 1996.** The yeast two hybrid system: prospects for protein linkage maps. *Trends Cell. Biol.* **6:** 196-199.

41. **Fields, S., Kohara, Y. and Lockhart, D.J. 1999.** Functional Genomics. *Proc. Natl. Acad. Sci. U.S.A.* **96**: 8825-8826.

42. **Fields, S. and Song, O. 1989.** A novel genetic system to detect protein-protein interactions. *Nature* **340**: 245-246.

43. **Fields, S. and Sternglanz, R. 1994.** The two-hybrid system: an assay for protein-protein inteactions. *Trends In Genetics* **10**: 286-291.

44. **Finley, R.L. Jr. and Brent, R. 1994.** Interaction mating reveals binary and ternary connections between Drosophila cell cycle regulators. *Proc. Natl. Acad. Sci. U.S.A.* **91:** 12980-12984.

45. **Fischer, H.M., Brunderer, T. and Hennecke, H. 1988.** Essential and non-essential domains in the *Brandyrhizobium japonicum* NifA protein: identification of indispensible cysteine residues potentially involved in redox activity and/or metal binding. *Nucleic Acids Res.* **16**: 2207-2224.

46. **Fischer, H.M. and Hennecke, H. 1987.** Direct response of *Brandyrhizobium japonicum nifA*-mediated *nif* gene regulation to cellular oxygen status. *Mol. Gen. Genet.* **209**: 621-626.

47. **Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F., Kerlavage A.R., Bult C.J., Tomb J.F., Dougherty B.A., Merrick J.M., et al. 1995.** Whole-genome random sequencing and assembly of *Haemophilus influenzae Rd. Science* **28:** 468-470.

48. **Flint, A.J., Tiganis, T., Barford, D. and Tonks, N.K. 1997.** Development of substrate trapping mutants to identify physiological substrates of protein tyrosine phosphatases. *Proc. Natl. Acad. Sci. U.S.A.* **94:** 1680-1685.

49. **Flores, A., Briand, J.F., Gadal, O. and Andrau, J.C.** ***et al.*** 1999. A protein-protein interaction map of yeast RNA polymerase III. *Proc. Natl. Acad. Sci. USA* **96**: 7815-7820.

50. **Formosa, T., Barry, J., Alberts, B.M. and Greenblatt, J. 1991.** Using affinity chromatography to probe structure of protein machines. *Methods Enzy.* **208:** 24-45.

51. **Frederickson, R.M. 1998.** Macromolecular matchmaking: advances in two hybrid and related technologies. *Current Opinion In Biotecnology* **9**: 90-96.

52. **Fromont-Racine, M., Rain, J.C. and Legrain, P. 1997.** Toward a functional analysis of the yeast genome through exhaustive two-hybrid screens. *Nat. Genet.* **16**: 277-282.

53. **Gill, G. and Ptashne, M. 1987.** Mutants of GAL4 protein altered in an activation function. *Cell.* **51:**121-126.

54. **Gill, G. and Ptashne, M. 1988.** Negative effect of the transcriptional activator Gal4. *Nature.* **334:** 721-724.

55. **Golemis, E.A. and Khazak, V. 1997.** Alternative yeast two hybrid systems: The interaction trap and interaction mating. *Methods Mol. Biol.* **63:** 197-218.

56. **Gralla, J.D. and Collado-Vides, J.** Organization and function of transcription regulatory elements. *Escherichia coli And Salmonella, Cellular And Molecular Biology.* 2nd ed. **Vol 1.** 1232-1244.

57. **Gyuris, J., Golemis, E., Chertkov, H., and Brent, R. 1993.** Cdi1, a human G1 and S phase protein phosphatase that associates with Cdk2. *Cell.* **75:** 791-803.

58. **Helmann, J.D. and Chamberlin, M. 1988.** Structure and function of bacterial sigma factors. *Annu. Rev. Biochem.* **57**: 839-872.

59. **Holmes, D.S., and Quigley, M. 1981**. A rapid boiling method for the preparation of bacterial plasmids. *Anal. Biochem.* **114:** 193-197

60. **Holt, K.H., Olson, A.L., Moye-Rowley, W.S. and Pessin, J.E. 1994.** Phosphatidylinositol 3-kinase activation is mediated by high-affinity interactions between distinct domains within the p110 and p85 subunits. *Mol. Cell. Biol.* **14:** 42-49.

61. **Hoover, T.R., Santero, E., Porter, S. and Kustu, S. 1990.** The integration host factor stimulates interaction of RNA polymerase with NifA, the transcriptional activator of nitrogen fixation operons. *Cell* **63**: 11-22.

62. **Hope, I.A. and Struhl, K. 1986.** Functional dissection of a eukaryotic transcriptional activator protein, GCN4 of yeast. *Cell.* **46:** 885-894.

63. **Huala, E.. and Ausubel, F.M. 1989.** The central domain of *Rhizobium meliloti* NifA is sufficient to activate transcription from the *R. meliloti nifH* promoter. *J. Bacteriol.* **171**: 3354-3365.

64. **Huala, E., Stigter, J., and Ausubel, F.M. 1992.** The central domain of *Rhizobium leguminosarum* DctD functions independently to activate transcription. *J. Bacteriol.* **174:** 1428 - 1431.

65. **Huang, J. and Schreiber, S.L. 1997.** A genetic system for selecting small molecule inhibitors of protein-protein interactions in nanodroplets. *Proc. Natl. Acad. Sci. USA* **94**: 13396-13401.

66. **Hudson, J.R., Dawson, E.P., Rusing, K.L., Jackson, C.H., Lochshon, D., Conover, D., Lanciault, C., Harris, J.R., Simmons, S.J., Rothstein, R. and Fields, S. 1998.** The complete set of predicted genes from *S. cerivisiae* in readily usable form. *Genome Res.* **7:** 1169-1173.

67. **Inouye, C., Dhillon, N., Durfee, T., Zambryski, P.C. and Thorner, J. 1997.** Mutational analysis of STE5 in the yeast *Saccharomyces cerivisiae*; application of a differential interaction trap assay for examining protein-protein interactions. *Genetics.* **147:** 479-492.

68. **Inoue, H., Nojima, H., and Okayama, H. 1990.** High efficiency transformation of *Escherichia coli* with plasmids. *Gene*. **96:** 23-28.

69. **Ishihama, A. 1992**. Role of the RNA polymerase α subunit in transcriptional activation. *Mol. Microbiol*. **6**. 3283-3288.

70. **Iwabuchi, K., Li, B., Bartel, P.L. and Fields, S. 1993.** Use of the two-hybrid system to identify the domain of p53 involved in oligomerization. *Oncogene*. **8:** 1693- 1696.

71. **Johnson, N., and Varshavsky, A. 1994.** Split ubiquitin as a sensor of protein interactions *in vivo. Proc. Natl. Acad. Sci. U.S.A*. **91:** 10340 -10344.

72. **Johnston, M. 1987.** A model fungal gene regulatory mechanism: the GAL genes of Saccharomyces cerevisiae. *Microbiol. Rev*. **51:** 458-476.

73. **Karimova, G., Pidoux, J., Ullmann, A. and Ladant, D. 1998.** A bacterial two-hybrid system based on a reconstituted signal transduction pathway. *Proc. Natl Acad. Sci. USA* **95**: 5752-5756.

74. **Keegan, L., Gill, G. and Ptashne, M. 1986.** Separation of DNA binding from the transcription-activating function of a eukaryotic regulatory protein. *Science.* **231:** 699-704.

75. **Kustu, S., North, A.K. and Weiss. 1991.** Prokaryotic transcriptional enhancers and enhancer binding proteins. *Trends Biochem. Sci.* **16**: 397-402.

76. **Kustu, S., Santero, E., Keener, J., Popham, D. and Weiss, D. 1989.** Expression of σ^{54}(*ntrA*)-dependent genes is probably united by a common mechanism. *Microbiol. Revs.* **53**: 367-376.

77. **Laughon, A. and Gesteland, R. 1984**. Primary structure of the *Saccharomyces cerevisiae* GAL4 gene. *Mol. Cell Biol.* **4:** 260-267.

78. **Leanna, C.A. and Hannink, M. 1996.** The reverse two-hybrid system: a genetic scheme for selection against specific protein/protein interactions. *Nucleic Acids Res.* **24**: 3341-3347.

79. **Le Douarin, B., Pierrat, B., Baur, E., Chambon, P. and Losson, R. 1995.** A new version of the two-hybrid assay for detection of protein-protein interactions. *Nucleic Acids Res.* **23**: 876-878.

80. **Lee, H.S., Berger, D.K. and Kustu, S. 1993.** Activity of purified NifA, a transcriptional activator of nitroen fixation genes. *Proc. Natl. Acad. Sci. USA* **90**: 2266-2270.

81. **Li, B. and Fields, S. 1993.** Identification of mutations in p53 that affect binding to SV40 large T antigen using the Yeast Two-Hybrid System. *FASEB J.* **7:** 957.

82. **Li, J.J. and Herskowitz, I. 1993.** Isolation of ORC6, a component of the yeast origin recognition complex by a one-hybrid system. *Science*. **262:** 1870-1874.

83. **Licitra, E.J. and Liu, J.O. 1996.** A three-hybrid system for detecting small ligand-protein receptor interactions. *Proc. Natl. Acad. Sci. USA* **93**: 12817-12821.

84. **Lu, Z., Murray K.S., Van Cleave V., LaVallie E.R., Stahl M.L., McCoy J.M. 1995.** Expression of thioredoxin random peptide libraries on the Escherichia coli cell surface as functional fusions to flagellin: a system designed for exploring protein-protein interactions. *Bio/Technology*. **13:** 366-372.

85. **Luban, J. and Goff, S.P. 1995.** The yeast two-hybrid system for studying protein-protein interactions. *Current opinion in Biotechnology* **6**: 59-64.

86. **Luo, Y., Batalao, A., Zhou, H. and Zhu, L. 1997**. Mammalian two-hybrid system. A comlementary approach to the yeast two-hybrid system. *BioTechniques* **22**: 350-352.

87. **Luo, Y. and Vijaychanders, S. 1996.** Cloning and analysis of DNA-binding proteins by yeast one-hybrid and one-two-hybrid systems. *BioTechniques* **20**: 564-568.

88. **Ma, J. and Ptashne, M. 1987a.** Deletion analysis of GAL4 defines two transcriptional activating segments. *Cell*. **48:** 847-853.

89. **Ma, J. and Ptashne, M. 1987b.** A new class of yeast transcriptional activators. *Cell*. **51:** 113-119.

90. **Ma, J. and Ptashne, M. 1988.** Converting a eukaryotic transcriptional inhibitor into an activator. *Cell*. 55: 443-446.

91. **Magasanik, B. 1996.** Regulation Of nitrogen utilization. *Escherichia coli* and *Salmonella*: cellular and molecular biology, 2nd ed. **Vol I.** 1344-1354.

92. **Mather, M.W., Keightley, J.A. and Fee, J.A. 1993.** Recovery and cloning of genomic DNA fragments from dried agarose gels. *Methods In Enzymology*. **218:** 695-704

93. **Marsolier, M.C., Prioleau, M.N. and Sentenac, A. 1997.** An RNA polymerase II-based two-hybrid system to study RNA polymerase II transcriptional regulators. *J. Mol. Biol.* **268**: 243- 249.

94. **McKnight, J., Kristie, T. and Rolzman, B. 1987.** Binding of the virion protein mediating alpha gene induction in herpes simplex virus 1-infected cells to its cis site requires cellular proteins. *Proc. Natl. Acad. Sci. U.S.A.* **84:** 7061-7065.

95. **Miller, J.H. (1972).** In: Experiments in molecular genetics. Cold Spring Laboratory Harbour Laboratory Press, New York. pp452-455.

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

97. **Morett, E. and Buck, M. 1989.** *In Vitro* studies on the interaction of DNA polymerase σ^{54} with the *Klebsiella pneumoniae* and *Rhizobium meliloti nifH* promoters. The role of NifA in the formation of an open promoter complex. *J. Mol.Biol.* **210**:65-77.

98. **Morett, E., Cannon, W. and Buck, M. 1988.** The DNA binding domain of the transcription activator protein NifA resides in its carboxy terminus, recognises the upstream activatior sequences of *nif* promoters and can be separated from the positive control function of NifA. *Nucleic Acids Res.* **16**: 11469-11488.

99. **Morett, E., Fischer, H.M. and Hennecke, H. 1991.** Influence of oxygen on DNA binding, positive control, and stability of the *Brandyrhizobium japonicum* NifA regulatory protein. *J. Bacteriol. 173*: 3478-3487.

100. **Morett, E. and Segovia, L. 1993.** The σ^{54} bacterial enhancer-binding protein family: Mechanism of action and phylogenic relationship of their functional domains. *J. Bacteriol.* **175**:6067-6074.

101. **Ninfa, A.J., Reitzer, L.J. and Magasanik, B. 1987.** Initiation of transcription at the bacterial glnAp2 promoter by purified *E. coli* components is facilitated by enhancers. *Cell* **50**: 1039-1046.

102. **North, A.K., Klose,K.E., Stedman, K.M. and Kustu, S. 1993.** Prokaryotic enhancer binding proteins reflect eukaryotic like modularity: the puzzel of nitrogen regulatory protein C. *J Bacteriol.* **175**: 4267-4273.

103. **North, A.K. and Kustu, S. 1997.** Mutant forms of the enhancer-binding protein NtrC can activate transcription from solution. *J. Mol. Biol.* **267**:17- 36.

104. **Osborne, M.A., Dalton, S. and Kochan, J.P. 1995.** The yeast tribrid-system: genetic detection of trans-phosphorylated ITAM-SH2-interactions. *Biotechnology* **13**: 1474-1478.

105. **Pahel, G., Rothstein, D.M. and Magasanik, B. 1982.** Complex *glnA-glnL-glnG* operon of *Escherichia coli. J. Bacteriol.* **150**: 202-213.

106. **Phizicky, E.M. and Fields, S. 1995.** Protein-protein interactions: methods for detection and analysis. *Microbiological Rev.* **59**: 94-123.

107. **Popham, D.L., Szeto, D., Keener, J. and Kustu, S. 1989.** Function of a bacterial activator protein that binds to transcriptional enhancers. *Science* **243**: 629-635.

108. **Putz, U., Skehel, P. and Kuhl, D. 1996.** A tri-hybrid system for the analysis and detection of RNA- protein interactions. *Nucleic Acids Res.* **24**: 4838-4840.

109. **Reitzer, L.J. and Magasanik, B. 1986.** Transcription of *glnA* in *E. coli* is stimulated by activator bound to sites far from the promoter. *Cell* **45**: 785-792.

110. **Reitzer, L.J., Movsas, B. and Magasanik, B. 1989.** Activation of glnA transcription by nitrogen regulator I (NRI)-phosphate in *Escherichia coli:* evidence for a long-range physical interaction between NRI-phosphate and RNA polymerase. *J Bacteriol.* **171**: 5512-5522.

111. **Roder, K.H., Wolf, S.S. and Schweizer, M. 1996.** Refinement of vectors for use in the yeast two-hybrid system. *Analytical Biochem.* **241**: 260-262.

112. **Ross, W., Gosink, K.K. and Salomon, J. 1993.** A third recognition element in bacterial promoters: DNA binding by the alpha subunit of RNA polymerase. *Science* **262**: 1407-1413.

113. **Rossi, F., Charlton, C.A., Blau, H.M. 1997.** Monitoring protein-protein interactions in intact eukaryotic cells by β-galactosidase complementation. *Proc. Natl. Acad. Sci. U.S.A.* **94:** 8405-8410.

114. **Sadowski, I., Bell, B., Broad, P. and Hollis, M. 1992.** GAL4 fusion vectors for expression in yeast or mammalian cells. *Gene* **118**: 137-141.

115. **Santero, E., Hoover, T.R., North, A.K., Berger, D.K., Porter, S.C. and Kustu, S. 1992.** Role of integration host factor in stimulating transcription from the σ^{54}-dependent *nifH* promoter. *J. Mol. Biol.* **227**: 602-620.

116. **Sasse-Dwight, S., and Gralla, J.D. 1990.** Role of eukaryotic-type functional domains found in the prokaryotic enhancer receptor factor σ^{54}. *Cell* **62**: 945-954.

117. **SenGupta, D.J., Zhang, B., Kraemer, B., Pochart, P., Fields, S. and Wickens, M. 1996.** A three-hybrid system to detect RNA-protein interactions *in vivo. Proc. Natl. Acad. Sci. USA* **93**: 84965-8501.

118. **Serebriiskii, I., Khazak, V. and Golemis, E.A. 1999.** A two-hybrid dual bait system to discriminate specificity of protein interactions. J. Biol. Chem. 274: 17080-17087.

119. **Shih, H.M., Goldman, P.S., DeMaggio, A.J., Hollenberg, S.M., Goodman, R.H. and Hoekstra, M.F. 1996.** A positive genetic selection for disrupting protein-protein interactions: Identification of CREB mutations that prevent association with the coactivator CBP. *Proc. Natl. Acad Sci. U.S.A.* **93:** 13896-13901.

120. **Silver, P.A., Keegan, L.P. and Ptashne, M. 1984**. Amino terminus of the yeast GAL4 gene product is sufficient for nuclear localization. *Proc. Natl. Acad. Sci. U.S.A.* **81:** 5951-5955.

121. **Smith, G.P. 1985.** Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. *Science.* **228:** 1315-1317.

122. **Smith, D.B. and Johnson, K.S. 1988.** Single step purification of polypeptides expressed in *Escherichia coli* as fusions with glutathione S-transferase. *Gene.* **67:** 31-40.

123. **Soares, M.B., Bonaldo, M.F., Jelene, P., Su, L., Lawton, L. and Efstratiadis, A. 1994.** *Proc. Natl. Acad. Sci. U.S.A.* **91:** 9228 - 9232.

124. **Spada, S. and Plunkthun, A. 1997.** Selective infective phage (SIP) technology: A novel method for in vivo selection of interacting protein-ligand pairs. *Nature Medicine.* **3:** 694-696.

125. **Spencer, D.M., Wandless, T.J., Schreiber, S.L. and Crabtree, G.R. 1993.** Controlling signal transduction with synthetic ligands. *Science.* **262:** 1019-1024.

126. **Su, W., Porter, S., Kustu, S. and Echols, H. 1990.** DNA looping and enhancer activity: association between DNA bound NtrC activator and RNA polymerase at the bacterial *glnA* promoter. *Proc. Natl. Acad. Sci. USA* **87**:5504-5508.

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

128. **Tintut, Y., Wang, J.T. and Gralla, J.D. 1995.** A novel bacterial transcription cycle involving σ^{54}. *Genes & Development.* **9:** 2305 - 2313.

129. **Tirode, F., Malaguti, C., Romero, F., Attar, R., Camonis, J. and Egly, J.M. 1997.** A conditionally expressed third partner stabilizes or prevents the formation of a transcriptional activator in a three-hybrid system. *J. Mol. Biol.* **272**: 22995-22999.

130. **Vidal, M., Brachmann, R., Fattaey, A., Harlow, E. and Boeke, J.D. 1996.** Reverse two-hybrid and one-hybrid systems to detect dissociation of protein-protein and DNA-protein interactions. *Proc. Natl. Acad. Sci. USA* **93**: 10315-10320.

131. **Vidal, M. Braun, P., Chen, E., Boeke, J.D. and Harlow, E. 1996.** Genetic characterisation of a mammalian protein-protein interaction domain by using a yeast reverse two hybrid system. *Proc. Natl. Acad. Sci. U.S.A.* **93:** 10321-10326.

132. **Vidal, M. and Endoh, H. 1999.** Prospects for drug screening using the reverse two-hybrid system. *Trends Biotechnol.* **17**: 374-381.

133. **Vidal, M. and Legrain, P. 1999.** Yeast forward and reverse 'n'-hybrid systems. *Nucleic Acids Res.* **27**: 919-929.

134. **Vojtek, A.B., Hollenberg, S.M., and Cooper, J.A. 1993.** Mammalian Ras interacts directly with the serine/threonine kinase Raf. *Cell.* **74:** 205-214.

135. **Wang, M.M. and Reed, R.R. 1993.** Molecular cloning of the olfactory neuronal transcription factor Olf-1 by genetic selection in yeast. *Nature.* **364:** 121-126.

136. **Warbrick, E. 1997.** Twos company, three's a crowd: the yeast two hybrid system for mapping molecular interactions. *Structure* **5**:13-17.

137. **Weiss, D.S., Batut, J., Kiose, K.E., Keener, J. and Kustu, S. 1991.** The phosphorylated form of the enhancer-binding protein NtrC has an ATPase activity that is essential for activation of transcription. *Cell* **67**:155-167.

138. **White, M.A. 1996.** The yeast two hybrid system: Forward and reverse. *Proc. Natl. Acad. Sci. U.S.A.* **93:** 10001-10003.

139. **Wilson, T.E., Fahrner, T.J., Johnston, M. and Milbrandt, J. 1991.** Identification of the binding site for NGFI-B by genetic selection in yeast. *Science.* **252:** 1296-1300.

140. **Wilson, D.R. and Beveridge, T.J. 1993.** Bacterial flagellar filaments and their component flagellins. *Can. J. Microbiol.* **39:** 451-472.

141. **Wolf, S.S., Roder, K. and Schweizer, M. 1996.** Construction of a reporter plasmid that allows expression libraries to be exploited for the one-hybrid system. *BioTechniques* **20**: 568-574.

142. **Wootton, J.C. and Drummond, M.H. 1989.** The Q-linker: a class of interdomain sequences found in bacterial multidomain regulatory proteins. *Protein Eng.* **2**: 535-543.

143. **Zhang, J. and Lauter, S. 1996.** A yeast three-hybrid method to clone ternary protein complex components. *Anal. Biochem.* **242:** 68-72.