Copyright is owned by the Author of the thesis. Permission is given for a copy to be downloaded by an individual for the purpose of research and private study only. The thesis may not be reproduced elsewhere without the permission of the Author.

REGULATION OF THE *ASPERGILLUS NIDULANS* CYTOCHROME *C* GENE

A thesis presented in partial fulfilment of
the requirements for the degree of
Master of Science in Molecular Genetics
at Massey University, Palmerston North
New Zealand

Linda Joy McGlynn

1997

ABSTRACT

The filamentous fungus *Aspergillus nidulans* has been genetically and biochemically well-characterised and thus provides an attractive model for studies on the regulation of eukaroytic gene expression. This study was undertaken to investigate the factors affecting respiratory function in *A. nidulans*. Due to the central role of cytochrome *c* in oxidative respiration, this study was focused on the cytochrome *c* gene and primarily upon how oxygen availability affects its expression.

The *Aspergillus nidulans* cytochrome *c* gene (*cycA*) appears to be transcriptionally activated in response to oxygen availability (Raitt *et al.*, 1994). In the yeast *S. cerevisiae*, oxygen availability activates its cytochrome *c* genes via a heme-activated protein HAP1, which binds to the promoter region of each gene (Pfiefer *et al.*, 1989). Since heme is only synthesised in the presence of oxygen, activity of the HAP1 protein serves as an intracellular signal of oxygen availability.

In the upstream region of the *cycA* gene, a sequence with homology to the *S. cerevisiae* HAP1 binding site was present (Raitt, 1992). To determine the significance of the putative HAP1 binding site and the role of other promoter sequences in the *A. nidulans cycA* gene, a promoter-reporter vector was constructed. However, upon sequencing of the *cycA* promoter in the reporter vector, a sequencing error was discovered in the published *cycA* gene by Raitt *et al.* (1994) which affected the position of the major translational start site. Further examination of the *cycA* sequence also revealed a possible undetected intron (Intron I). To determine the number of introns in the *cycA* gene, RT-PCR was performed on *cycA* RNA. Sequencing of the RT-PCR amplified products showed that the previously undetected intron (Intron I) was present, and that the *cycA* gene contains three instead of two introns as published by Raitt *et al.* (1994). Since the published ATG start site was located within Intron I, a new translational start site was proposed. The major consequences of these changes to the *cycA* gene was that the putative HAP1 site was now located within the coding region of the gene, and therefore could not be a regulatory element. In addition only 247 bp of *cycA* promoter sequence remained cloned for analysis. To obtain additional promoter sequence, an *A. nidulans* genomic library was screened with a *Bam*HI cDNA probe containing 224 bp of the 5' region of the *cycA* gene. Three positive clones were obtained, of which λLM9 and λLM5 were identical, and λLM19 was an overlapping clone with λLM9 and λLM5. Restriction enzyme and Southern blott analysis of the two overlapping *cycA* clones, revealed that 2.1 kb *Eco*RI fragments from both clones contained the 5' region. The 2.1 kb *Eco*RI fragment from λLM9 was cloned into pUC18 and sequenced. The completed

upstream sequence of the *A. nidulans* cytochrome *c* gene was obtained, and putative regulatory signals including the HAP1 binding site were found, and compared with published promoter sequences which regulate the expression of respiratory-encoding genes from yeast.

ACKNOWLEDGEMENTS

Firstly, I would like to thank my supervisor Rosie Bradshaw, for keeping faith in me, and for her continual enthusiasm and encouragement, even when my thesis was not going according to plan! Her support was much appreciated.

I would also like to thank all past and present members of the Microbiology and Genetics Department. Thanks everyone for your friendship, you guys made it a great place to work in. In particular, I would like to thank our neighbouring lab, Scott base, who we were always borrowing things from, especially Carolyn Young, who somehow found time out of her busy research to give all of us MGU students loads of time saving advice, Mike Collett for answering all my questions so extensively and patiently, and Austen Ganley for being around to listen and to entertain us out at the farm. This also applies to Kate and Mike Collett who we shared many yummy dinners and good times with.

Many thanks goes to everyone from my lab, Dianne-for all her *Aspergillus* related advice, Paul-for providing lots of laughs, Carmel-thanks for being such a great friend, Anita-for her lively personality which helped keep us sane, Branwen-for your support especially during those hard times, Brendon-for all those friendly chats, and Bek-for technical support, I hope you have more luck with this project than I did! Also thanks to Tania who was always popping into our lab with lots of smiles.

A big thank you must also go to my family for their love and support. I would also like to thank my special friend Rowena. Thanks for putting up with me, especially during our four years flatting together, and for all your strength and encouragement, which went along way towards the accomplishment of this thesis.

Lastly, I have Richard to thank for many many things, I hardly know how to adequately express my gratitude and love. Thank you Rich for giving me your support, love, encouragement, guidance, cooking more than I did, and all those sacrifices you made because I was too busy! Also I must mention that your scientific advice was invaluable. I am very lucky to have you for a future husband!

TABLE OF CONTENTS

ABSTRACT ii

ACKNOWLEDGEMENTS iv

TABLE OF CONTENTS v

LIST OF TABLES xi

LIST OF FIGURES xii

CHAPTER 1. INTRODUCTION 1

1.1 THE ROLE OF THE CYTOCHROME C PROTEIN 1

1.2 WHY STUDY THE *ASPERGILLUS NIDULANS* CYTOCHROME *C* GENE? 1

1.3 REGULATION OF THE *SACCHAROMYCES CEREVISIAE* CYTOCHROME *C* GENES 2

1.3.1 HAP1 Regulation of *CYC1* and *CYC7* in Response to Oxygen Availability 3

1.3.1.1 Structure and Function of the HAP1 Protein 3

1.3.1.2 Multi-sequence Recognition by HAP1 4

1.3.2 Regulation of *CYC1* and *CYC7* in Response to Carbon Source by the Yeast CCAAT-binding Factor 8

1.3.2.1 The Yeast CCAAT-binding Factor (HAP2/3/4/5 Complex) 8

1.3.3 Other factors affecting *CYC7* expression 10

1.4 REGULATION OF THE *KLUYVEROMYCES LACTIS* CYTOCHROME *C* GENE (*KlCYC1*) 10

1.5 REGULATION OF THE *ASPERGILLUS NIDULANS* CYTOCHROME *C* GENE (*cycA*) 12

1.4.1 Regulation of the *cycA* gene in response to Environmental Parameters 12

1.4.2 Analysis of the 5' Region of the *cycA* gene 13

1.4.2.1 Putative HAP1 Binding Site....13
1.4.2.2 Putative Upstream Opening Reading Frame (uORF)....13

1.5 AIMS AND OBJECTIVES....13

CHAPTER 2. MATERIALS AND METHODS....17

2.1 FUNGAL AND BACTERIAL STRAINS, LAMBDA CLONES AND PLASMIDS....17

2.2 MEDIA....17
2.2.1 Luria Broth....17
2.2.3 MYG Media....17
2.2.4 SOC Medium....17
2.2.5 NZCYM....17
2.2.6 TB Top Agarose....17

2.3 COMMON BUFFERS AND SOLUTIONS....17
2.3.1 1 x TAE Buffer....19
2.3.2 1 x TBE Buffer....19
2.3.3 TE Buffer (10/1)....19
2.3.4 TE Buffer (10/0.1)....19
2.3.5 10 x Sequencing TBE Buffer....19
2.3.6 TES (10/1/100)....19
2.3.7 STET Buffer....19
2.3.8 SM Buffer:....19
2.3.9 Hybridisation Solution....19
2.3.10 50 x Denhardt's Solution....19
2.3.11 Tris-Equilibrated Phenol....19
2.3.12 20 x SSC....19
2.3.13 10 x Gel Loading Buffer....19
2.3.14 Standard DIG Buffer....19
2.3.15 DIG Buffer 1....20
2.3.16 DIG Buffer 2....20
2.3.17 DIG Buffer 3....20
2.3.18 Blocking Reagent stock solution....20
2.3.19 Alkaline Lysis Solution I....20
2.3.20 Alkaline Lysis Solution II....20

2.3.21 Alkaline Lysis Solution III..... 20

2.4 GROWTH AND MAINTENANCE OF CULTURES 20

2.5 DNA PREPARATIONS..... 20
2.5.1 Rapid Boiling Plasmid Preparation 20
2.5.2 Small Scale Alkaline Lysis Plasmid Preparation 20
2.5.3 Large Scale Alkaline Lysis Plasmid Preparation 21
2.5.4 Purification of Plasmid DNA by Cesium Chloride-Ethidium Bromide Gradient..... 22
2.5.5 Preparation of DNA from Fungal Cultures 22
2.5.6 Preparation of Lambda Phage DNA by Liquid Lysate 23
2.5.7 Purification of DNA by Phenol-Chloroform Extraction and Ethanol Precipitation..... 23

2.6 STANDARD PROCEDURES..... 24
2.6.1 Restriction Enzyme Digests..... 24
2.6.2 Agarose-gel Electrophoresis of DNA..... 24
2.6.3 Determination of Molecular Weights 24
2.6.4 Determination of DNA Concentration 25
2.6.4.1 Quantification of DNA Concentration by Electrophoresis 25
2.6.4.2 Spectrophotometric Determination of DNA Concentration and Purity..... 25

2.7 AMPLIFICATION OF DNA BY THE POLYMERASE CHAIN REACTION (PCR)..... 25

2.8 DNA CLONING PROCEDURES 26
2.8.1 Gel Purification of DNA 28
2.8.2 CAP Treatment of DNA..... 28
2.8.3 Ligation Reactions 28
2.8.4 Transformation of *Escherichia coli* 29
2.8.4.1 Preparation of Electroporation Competent *E. coli* Cells 29
2.8.4.2 Electroporation of *E. coli*..... 29
2.8.4.3 Preparation of CaCl2-Competent Cells 30
2.8.4.4 CaCl2 Transformation of *E. coli* 30

2.9 COLONY HYBRIDISATION USING THE DIG DETECTION SYSTEM..... 30
2.9.1 Colony Lifts 30
2.9.2 Colony Hybridisation 31

2.9.3 Preparation of DIG-Labelled Probe 31
2.9.4 Detection of Hybridised DIG-labelled Probe 32
2.10 DNA SEQUENCING 32
2.10.1 Sequenase Version 2.0 Sequencing Protocol 32
2.10.2 Amplicycle Sequencing Kit Protocol 33

2.11 RNA PROCEDURES 33
2.11.1 Precautions Taken for Working With RNA 33
2.11.2 Agarose-gel Electrophoresis of RNA 34
2.11.3 Determination of Fungal RNA Concentration and Purity 34

2.12 RNA PREPARATION 34
2.12.1 *Aspergillus nidulans* Spore Suspension Preparation 34
2.12.2 Isolation of RNA from *Aspergillus nidulans* 35

2.13 RT-PCR ANALYSIS OF RNA 35
2.13.1 Reverse Transcriptase Synthesis of cDNA 35
2.13.2 Amplification of cDNA by PCR 36

2.14 SOUTHERN BLOTTING AND HYBRIDISATION 36
2.14.1 Southern Blotting 36
2.14.2 Preparation of [a-32P]dCTP-Labelled DNA Probe 37
2.14.3 Southern Blot Hybridisation 37
2.14.4 Stripping Probes from Southern Blots 38

2.15 LIBRARY SCREENING BY PLAQUE HYBRIDISATION 38
2.15.1 Plating Phage λ 38
2.15.2 Plaque Lifts 39
2.15.3 Hybridisation of λ Phage DNA to [a-32P]dCTP-Labelled Probe 39

CHAPTER 3. CONSTRUCTION OF PROMOTER-REPORTER VECTORS AND DISCOVERY OF A SEQUENCING ERROR 40

3.1 INTRODUCTION 40

3.2 CLONING STRATEGY 40
3.2.1 The lacZ Reporter Vector 40
3.2.2 PCR Amplification of the *cycA* Promoter Fragments 40
3.2.3 General Preparation of pAN923-42B Vector and 0.68 kb *cycA* Promoter 43

3.3 LIGATIONS AND TRANSFORMATIONS OF THE 0.68 kb *cycA* PROMOTER 46
3.3.1 Ligation of 0.68 kb promoter directly into pAN923-42B 46
3.3.2 Ligation of 0.68 kb Promoter into T-Tailed pUC18 Vector 53
3.3.3 Ligation of 0.68 kb Promoter into Commercially CAP-treated pUC18 55
3.3.4 Ligation of 0.68 kb promoter into pAN923-42B 55

3.4 LIGATIONS AND TRANSFORMATIONS OF THE *cycA* Δ HAP1 PROMOTER 58

3.5 SEQUENCING OF THE 0.68 kb PROMOTER IN PLASMID R122 58

CHAPTER 3 DISCUSSION 64

3.6 PROBLEMS WITH CLONING INTO pAN923-42B VECTOR 64

3.7 IMPLICATIONS FROM THE DISCOVERY OF THE SEQUENCING ERROR IN THE PUBLISHED *cycA* GENE BY RAITT *et al* (1994) 64

CHAPTER 4. RT-PCR ANALYSIS OF *cycA* RNA 69

4.1 INTRODUCTION 69

4.2 DEVELOPMENT OF A RT-PCR STRATEGY 69

4.3 RT-PCR RESULTS 75
4.3.1 Sequencing of RT-PCR Products 78

CHAPTER 4 DISCUSSION 83

4.4 THE *cycA* GENE CONTAINS A THIRD INTRON AND HAS A CORRECTED N-TERMINAL REGION 83

4.5 ALL THREE INTRONS OF THE *cycA* GENE ARE LOCATED AT CONSERVED POSITIONS 86

CHAPTER 5. CLONING AND ANALYSIS OF THE 5' REGION OF THE *cycA* GENE 89

5.1 INTRODUCTION 89

5.2 SCREENING OF AN *A. NIDULANS* LIBRARY 89

5.3 SOUTHERN BLOTTING AND RESTRICTION MAPPING OF POSITIVE λ CLONES 89

5.4 SEQUENCE ANALYSIS OF CLONED λ DNA SHOWN TO HYBRIDISE TO THE *cycA* GENE 90
5.4.1 Subcloning and Sequencing of the 2.1 kb Fragment 90
5.4.2 Confirmation that the 2.1 kb Fragment Corresponds to Regions of the *A. nidulans* A18 Genome 100
5.4.3 Analysis of the cycA promoter sequence 100
5.4.3.1 Putative HAP1 UAS 100
5.4.3.2 Putative HAP2/3/4/5 (Yeast CCAAT-binding Factor) UAS 100
5.4.3.3 Putative CREA Binding Site 105
5.4.3.4 Other Regulatory Elements 105

CHAPTER 5 DISCUSSION 106

5.5 TATA BOX 106

5.6 THE ANCF (*A. NIDULANS* CCAAT BINDING FACTOR) COMPLEX MAY REGULATE THE *CYCA* GENE 106

5.7 THE *A. NIDULANS CYCA* GENE MAY BY REGULATED BY A HAP1-LIKE PROTEIN 108

CHAPTER 6. SUMMARY AND CONCLUSIONS 110

APPENDICES 112
Appendix 1.0 Vector Maps 112
Appendix 2.0 Sequence Data 114

REFERENCES 119

LIST OF TABLES

Table 2.1. Strains, λ clones and plasmids 18

Table 2.2. PCR and Sequencing Primers 27

Table 3.1. Transformations A, B & C (with pAN923-42B vector) 47

Table 3.2. Expected Restriction Digestion Fragment Profiles for Ligation of 0.68 kb Promoter into pAN923-42B 49

Table 3.3. Transformation D (with pAN923-42B vector) 52

Table 3.4. Expected Restriction Enzyme Profiles for Ligation of 0.68 kb Promoter into Plasmid pUC18 54

Table 3.5. Transformations E & F (with CAP-treated pUC18 vector) 56

Table 3.6. Transformation G (with pAN923-42B vector) 57

Table 4.1. PCR Products Expected From RT-PCR Analysis 74

Table 5.1. Restriction Enzyme Digestions of λ clones 95

LIST OF FIGURES

Figure 1.1. Projection of HAP1 binding at *CYC7* and *CYC1* onto DNA helices (Pfiefer et al., 1987)..... 6

Figure 1.2. Comparison of HAP1 binding sites in *S. cerevisiae* and *A. nidulans*..... 14

Figure 3.1. The lacZ reporter vector pAN923..... 41

Figure 3.2. PCR primer binding sites within the *cycA* promoter fragments..... 44

Figure 3.3. Restriction enzyme analysis of plasmids pAN923, R122 and R120..... 50

Figure 3.4. Comparison of the *cycA* nucleotide sequence from plasmid R122 with the published *cycA* sequence (Raitt, 1992)..... 60

Figure 3.5. Autoradiograph of sequence obtained from the *cycA* gene cloned into plasmid R41..... 60

Figure 3.6. The published nucleotide sequence of the *cycA* gene (Raitt *et al.*, 1994)..... 62

Figure 3.7. How the sequencing error affects the *cycA* amino-terminal sequence..... 66

Figure 4.1. Re-estimation of the *cycA* transcription start point..... 70

Figure 4.2. RT-PCR outline showing the relative positions of primers..... 72

Figure 4.3. RT-PCR analysis of total RNA from *A. nidulans* strain A18..... 76

Figure 4.4. cDNA sequence of the 596 bp RT-PCR product..... 79

Figure 4.5. The corrected nucleotide sequence of the *cycA* gene..... 81

Figure 4.6. Alignment of cytochrome c N-terminal protein sequences..... 84

Figure 4.7. Alignment of three cytochrome c protein sequences showing intron positions..... 87

Figure 5.1a. Restriction digestion profiles of λLM9 and λLM19..... 91

Figure 5.1b. Southern Blot of λLM9 and λLM19..... 91

Figure 5.2. Restriction maps of λLM9 and λLM19 isolated from an *A. nidulans* genomic library that hybridised to the *cycA* gene..... 93

Figure 5.3. Nucleotide sequencing strategy for the 5' region of the *cycA* gene..... 96

Figure 5.4. Nucleotide sequence of the 5' region of the *cycA* gene 98

Figure 5.5a. Restriction digestions of *A. nidulans* A18 genomic DNA..... 101

Figure 5.5b. Southern blot of *A. nidulans* A18 genomic DNA 101

Figure 5.6. Comparison of known yeast and putative *A. nidulans* HAP1 UASs..... 103

CHAPTER 1
INTRODUCTION

1.1 THE ROLE OF THE CYTOCHROME *C* PROTEIN

Cytochrome *c* is a highly conserved and well characterised protein which forms a key component of the respiratory pathway. The essential function of this process is oxidative phosphorylation from which energy is derived in the form of ATP. Although this pathway operates in the mitochondria, cytochrome *c* and some other respiratory components are encoded in the nucleus. Once synthesised these nuclear-encoded components are transported to the inner mitochondrial membrane for assembly of the respiratory chain. Respiratory components are thus encoded by both the nuclear and mitochondrial genomes. Coordinate expression of nuclear and mitochondrial genes is therefore required to achieve a balanced production of protein for assembly of respiratory complexes in the mitochondria. Coordinated regulation of these genes also occurs in response to mitochondrial and environmental conditions of the cell, for example, oxygen availability and respiratory status.

1.2 WHY STUDY THE *ASPERGILLUS NIDULANS* CYTOCHROME *C* GENE?

Extensive studies of the relationship between respiratory activity and environmental conditions have been carried out with the yeast *Saccharomyces cerevisiae,* a facultative anaerobe. In particular the effect of oxygen availability and carbon source on the expression of the cytochrome *c* genes from *S. cerevisiae* has been well characterised (for a review see Forsburg and Guarente, 1989a; Zitomer and Lowry, 1992). In contrast, limited research in this field has been conducted on the filamentous fungus, *Aspergillus nidulans*, and it is not known whether the cytochrome *c* gene of this obligate aerobe is regulated through similar *cis* and *trans* regulatory signals.

Recent research performed on the aerobic respiratory yeast *Kluyveromyces lactis,* has revealed that its cytochrome *c* gene (*KlCYC1*), is transcriptionally regulated by oxygen in a similar manner to the *S. cerevisiae* cytochrome *c* genes, despite differences in their respiratory metabolism (Freire-Picos *et al.,* 1995). However, it is not clear whether the *KlCYC1* gene is regulated similarly by carbon source (Freire-Picos *et al.,* 1993; Mulder *et al.,* 1994; Freire-Picos *et al.,* 1995; NGuyen *et al.,* 1995; Mulder *et al.*, 1995). Thus it seems likely that control of respiration in *A. nidulans* will be mediated by common regulatory pathways, such as those established for *S. cerevisiae*. It will be interesting to determine how regulation of the *A. nidulans* cytochrome *c* gene differs to that of the *S.*

cerevisiae cytochrome *c* genes, and may provide insight into the obligate aerobic nature of some fungi. *A. nidulans* is an ideal candidate for this comparison, as it provides a well established model system, where many genes have been studied in detail at the molecular and biochemical level (Arst and Scazzocchio, 1985; Davis and Hynes, 1991).

Environmental factors found to affect the regulation of the cytochrome *c* genes of *S. cerevisiae* and *A. nidulans* (such as carbon source, cell growth rate, and oxygen availability), alter the respiratory state and cellular metabolism of the organism (Forsburg and Guarente, 1989; Pillar and Bradshaw, 1991; Zitomer and Lowry, 1992; Raitt *et al.*, 1994). In some *Aspergillus* species, the control of the organism's respiratory state is important for the production of useful secondary metabolites. For example, if oxidative respiration is restricted, acetyl CoA accumulates and becomes available as a precursor for the biosynthesis of secondary metabolites. For one particular *Aspergillus* species, *A. parasiticus*, a highly toxic metabolite can be produced from acetyl CoA called aflatoxin. Biosynthesis of aflatoxin has been demonstrated to be enhanced under conditions of low oxygen and high glucose levels, which minimise oxidative respiration and enhance glycolysis (Shih & Marth, 1974). In addition, oxygen limitation can be a serious problem in large-scale chemostat cultures of obligately aerobic fungi, (for example; penicillin-producing *Penicillium chysogenum*), where high energy input is required to sparge and mix air in the vessel.

Therefore, the study of the regulation of the cytochrome c gene of *A. nidulans* should prove to be both interesting and of commercial value.

1.3 REGULATION OF THE *SACCHAROMYCES CEREVISIAE* CYTOCHROME *C* GENES

The yeast, *S. cerevisiae*, is unusual in that it has two forms of cytochrome *c* encoded by the independent genes, iso1-cytochrome *c* (*CYC1*, Smith *et al.*, 1979) and iso2-cytochrome *c* (*CYC7*, Montgomery *et al.*, 1980). *CYC1* and *CYC7* are isofunctional with 78% identity at the amino acid level (Montgomery *et al.*, 1980). *CYC1* accounts for 95% of the cytochrome *c* which is expressed in aerobic conditions, with *CYC7* making up the remaining 5% (Creusot *et al.*, 1988). The minor isoform of cytochrome *c*, *CYC7*, is regulated coordinately with *CYC1* in response to various factors. The two principal signals that affect the respiratory ability of yeast are carbon source and oxygen. The response to these environmental parameters occurs through a specific interaction of heme-responsive regulatory proteins at upstream activation sites (UAS) located in the promoter region of each cytochrome *c* gene. *CYC1* contains two functional UAS sequences,

UAS1 and UAS2 (Guarente *et al.,* 1984), whereas *CYC7* contains one UAS (Prezant *et al.,* 1987).

1.3.1 HAP1 Regulation of *CYC1* and *CYC7* in Response to Oxygen Availability

Both *CYC1* and *CYC7* are transcriptionally activated by oxygen, with *CYC1* induced 200-fold (Guarente and Mason, 1983) and *CYC7* induced about 20-fold (Prezant *et al.,* 1987). Under anaerobic conditions, *CYC7* is weakly repressed by lack of oxygen, whereas *CYC1* is strongly repressed. The *CYC7* gene product is also believed to have a longer half-life in anaerobic conditions which may poise the cell for a transition from anaerobic to aerobic growth conditions (Pillar & Bradshaw, 1991).

The *CYC1* and *CYC7* genes (and other genes encoding respiratory functions or oxidative damage repair functions), are not activated directly by molecular oxygen. Instead intracellular levels of heme serve as intermediate signalling molecules to convey information about oxygen levels. Since heme requires oxygen for its biosynthesis, heme levels ultimately reflect oxygen levels (Zitomer & Lowry, 1992). Heme exerts its regulatory role on *CYC1* and *CYC7* through interaction with a heme-activated protein, HAP1. HAP1 (also called CYP1), binds to UAS1 of *CYC1* and to the UAS of *CYC7* to activate transcription in the presence of heme (Pfiefer *et al.,* 1987).

1.3.1.1 Structure and Function of the HAP1 Protein

The HAP1 protein is a member of the zinc cluster proteins (Creusot *et al*, 1988; Pfiefer *et al*, 1989), and has a general function as a coordinator of the activity of cytochrome-encoding and related genes involved in oxygen regulation (Lodi and Guiard, 1991). Deletion analysis of HAP1 revealed three functional domains: the DNA binding domain, the activation domain and the heme-responsive domain (Pfiefer *et al*, 1989).

The activation domain is located at the carboxyl end of HAP1 between residues 1308-1483. This region is involved in transcriptional activation and is highly acidic, which is consistent with other activators such as GAL4.

The DNA binding domain lies between residues 1 and 148, and specifies binding to UAS1 of *CYC1* and *CYC7*. It contains a zinc finger known as a binuclear zinc cluster, which has extensive homology to the well characterised yeast GAL4 zinc cluster protein. The cluster is involved in non-specific DNA binding, with specific amino acids surrounding the cluster involved with specific recognition of each cytochrome *c* gene

(Pfeifer *et al.,* 1989). For example, serine-63 at the base of the zinc cluster is essential for binding to the UAS1 of *CYC1*.

Mutational analysis of the HAP1 DNA binding domain (Turcotte and Guarente, 1992), found that HAP1 uses a cofactor at *CYC7* and potentially a different cofactor at *CYC1*. They proposed that the difference in sequence of both UAS sites could specify which cofactor is bound.

The heme-responsive domain has been proposed to mediate heme control of HAP1 activity and is located between residues 245-445. This domain contains seven adjacent repeat units of a conserved K/R C P I/V D H amino acid motif, resembling a metal or heme-binding site, and is therefore believed to be the site for the direct binding of heme (Creusot *et al*, 1988; Pfiefer *et al.,* 1989).

In the absence of heme, an unknown cellular factor or factors was found to bind to the heme domain at the K/R C P I/V D H amino acid motif of HAP1 to inhibit DNA binding (Fytlovich *et al.* 1993; Zhang and Guarente, 1994a). This interaction with the cellular repressor(s) may weaken DNA binding by blocking HAP1 dimerisation, the active binding form of the protein (Zhang *et al.*, 1993). Consistent with this result, Zhang and Guarente (1994) found that HAP1, like steroid hormone receptors, was sequestered in a high molecular weight complex by one or more cellular factors in the absence of heme, thereby blocking dimerisation. However, the dissociation of the high molecular weight complex to the dimer form produced only partially active HAP1, and was therefore suggested to be only one of the events leading up to the activation of the HAP1 protein.

Recently, HAP1 mutants were obtained which were independent of heme when assayed for activation (Haldi and Guarente, 1995). These mutants define an additional region of HAP1 which is heme-responsive, located near the activation domain. Another heme-independent mutant has also been discovered, but the mutation is located near the heme-responsive domain (Ushinsky and Keng, 1994). These multiple heme-responsive domains have been suggested to allow a fine-tuned response to oxygen availability.

1.3.1.2 Multi-sequence Recognition by HAP1

The UAS regions of *CYC1* and *CYC7* bound by HAP1 share no obvious sequence similarity. Both sites compete for binding to HAP1 with comparable affinities (Pfiefer *et al*, 1987; Ha *et al.,* 1996), but HAP1 activates transcription at *CYC1* 10 fold more efficiently than at *CYC7* (Turcotte & Guarente, 1992). HAP1 binds to these UAS sites as a homodimer (Zhang *et al.,* 1993).

Despite the dissimilarity in sequence between the two HAP1 binding sites, overall HAP1 contacts both sites similarly (Figure 1.1; Pfiefer *et al*, 1987). For both sites, binding of HAP1 occurs primarily to one side of the DNA helix, protecting 23 base pairs. Crucial contacts are made in both the major and minor grooves with a single major groove contact on the reverse face of the helix at one end of each binding site. However, due to the dissimilar nature of the two sites, the actual positions of the major and minor groove contacts in *CYC1* and *CYC7* differ. The best alignment of the HAP1 binding sites from this study was obtained by inverting the *CYC7* sequence. This results in a 7 nucleotide match from the 23 nucleotides protected by HAP1.

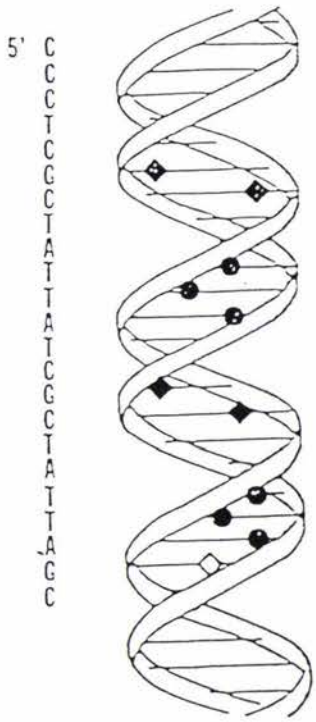
Further information on the binding specificity of HAP1 was obtained by selecting random HAP1 binding sites using a random oligonucleotide-PCR selection technique (Zhang and Guarente, 1994b). This study showed that HAP1 unlike the other zinc cluster proteins, including the GAL4 protein, binds to a direct repeat of two CGG triplets rather than a palindromic sequence. In addition, the HAP1 site contains a 6 bp spacer between the CGG triplets, with an asymmetrically positioned TA sequence critical for HAP1 binding. However, the UAS of *CYC7* does not fit this consensus as it contains two copies of CGC rather than CGG, and a more specific 6 bp spacer of TATTAT. The spacing between the CGG triplet repeats has been proposed to be important in controlling the binding specificity of protein members of the zinc cluster family. For example, studies on GAL4 (which contains a 11 bp spacer) showed that binding was greatly reduced with a decrease in spacing from 11 to 10 bp (Vashee *et al.,* 1993). Therefore the 6 bp spacing observed for the HAP1 binding site was suggested to be crucial for HAP1 binding.

Recently, random and site-directed mutagenesis of the *CYC1* and *CYC7* HAP1 binding sites showed that all HAP1 sites, including those from *CTT1* (catalase T), *CYB2* (cytochrome b_2) and *CYT1* (cytochrome c_1) genes, are related imperfect direct repeats with the 'optimal' sequence CGG N_3 TA N CGG N_3 TA (Ha *et al.,* 1996). Mutation of either CGG triplet in *CYC1* decreased *in vivo* activity and *in vitro* binding of HAP1, indicating that the CGG triplets are important for HAP1 binding. However, changing the CGC triplets in *CYC7* to CGG, increased HAP1 binding and activation, suggesting that HAP1 can recognise a CGC repeat, but less efficiently. In agreement with the earlier study by Zhang and Guarente (1994b), TA repeats were found to be important for HAP1 binding, but were found to be more essential for the imperfect HAP1 sites, such as the weaker CGC motif of *CYC7*. Thus, Ha *et al.* (1996) proposed that the sequence of the HAP1 binding sites influences the conformation of the protein to allow the activity of HAP1 to be altered.

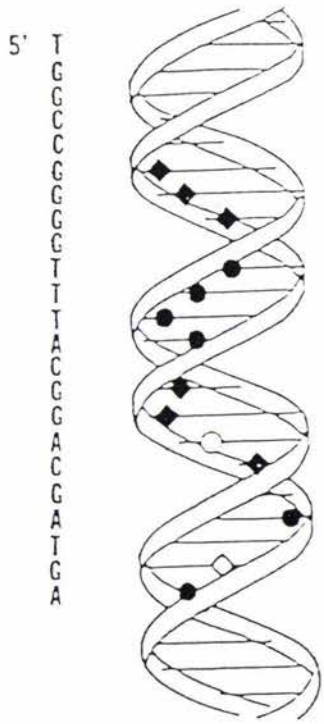
Figure 1.1 Projection of HAP1 binding at *CYC7* and *CYC1* onto DNA helices (Pfiefer *et al.*, 1987).

The DNA helices are drawn with 10.5 base pairs per helical turn. DNaseI protected sequences are shown (-251 to -229 for *CYC7* and -249 to -247 for *CYC1*). Guanine (major groove) contacts are shown as diamonds; adenine (minor groove) contacts are shown as circles. Open symbols represent contacts on the opposite side of the helix. The orientation of the *CYC7* sequence relative to the TATA box has been inverted to give maximal alignment of the binding motif.

HAP1 ON CYC7



HAP1 ON UAS1



Regulatory proteins that coordinate two or more genes, generally recognise a consensus sequence in the promoter region of the genes. The UAS1 of *CYC1* and UAS of *CYC7* appeared to be obvious exceptions to this, however it now seems that these binding sites are degenerate forms of the consensus sequence CGG N3 TA N CGG N3 TA (Ha *et al.*, 1996).

1.3.2 Regulation of *CYC1* and *CYC7* in Response to Carbon Source by the Yeast CCAAT-binding Factor

CYC1 and *CYC7* genes are both regulated by carbon source at the transcriptional level. *CYC1* expression is repressed 10-20 fold with glucose as the carbon source (Zitomer & Lowry, 1992), and induced 50-fold when shifted from glucose to the non-fermentable carbon source, lactate (Guarente *et al.*, 1984). *CYC7* expression is induced 4- to 8-fold with a shift from glucose to lactate medium (Prezant *et al.*, 1987).

1.3.2.1 The Yeast CCAAT-binding Factor (HAP2/3/4/5 Complex)

A non-fermentable carbon source such as lactate induces transcription of the *CYC1* gene through the binding of a conserved CCAAT-binding factor to a CCAAT box with the consensus TNATTGGT (the CCAAT box is encoded on the other strand). This is located in the upstream activation site, UAS2. In *S. cerevisiae*, the CCAAT-binding factor has been shown to be a heteromeric transcription factor composed of the heme-dependent regulatory proteins, HAP2, HAP3, HAP4 and HAP5. The subunits HAP2/3/4 were identified by mutations which abolished the transcriptional activation of the *CYC1* gene from the UAS2 element, and thus suppressed growth on a non-fermentable carbon source (Olesen *et al.*, 1987; Forsburg and Guarente, 1989b). HAP5 was recently identified by using the two-hybrid screening system (McNabb *et al.*, 1995). Two other undefined proteins were also obtained from the hybrid screen and may be required for the complete CCAAT-binding factor, or may encode ancillary factors for CCAAT-binding transcriptional activation (McNabb *et al.*, 1995).

None of the subunits of the yeast CCAAT-binding factor contain any homology to other known classes of DNA-binding motifs. Therefore, the precise mechanism for subunit association is unknown and will most likely prove to be novel. HAP4 has been shown to encode the principle transcriptional activation domain (Forsburg and Guarente, 1989b; Olesen and Guarente, 1990). It contains a highly acidic domain responsible for transcriptional activation which can be functionally exchanged with the GAL4 acidic activation domain. Furthermore HAP4, unlike HAP2 and HAP3, can independently activate transcription at high levels through a lexA operator-driven promoter. The

HAP2/3/5 subunits have been shown through mobility shift experiments to each be essential for DNA binding activity *in vitro*, thus representing the first heterotrimeric transcription factor (McNabb *et al.*, 1995).

Mutational analysis of the HAP2 protein has identified an essential 60 amino acid core region required for all functions of the HAP2 subunit (Olesen and Guarente, 1990). This core region contains an 18 amino acid subunit association domain (SAD) and a DNA binding domain of 21 amino acids (Olesen and Guarente, 1990; Xing *et al.*, 1993). Amino acid residues critical for HAP2 subunit association have been identified (Xing *et al.*, 1994), and structural studies of this region indicate that the SAD forms an alpha-helical structure which interacts with another HAP subunit (Xing *et al.*, 1994). HAP2 has been shown to bind to the HAP3 subunit (Olesen and Guarente, 1990), and most likely also binds to the HAP5 protein, which was isolated through protein interactions with the HAP2 protein (McNabb *et al.*, 1995). Through these studies it was suggested that HAP2 and HAP3 proteins interact through a weak interaction involving their respective DNA binding domains, and through a strong interaction between the SADs. The HAP5 subunit was suggested to be central to the complex, and acts by bringing HAP2 and HAP3 together through the binding of their SADs. This action would allow the weak DNA binding domains of HAP2 and HAP3 to come together and interact with the CCAAT box for cytochrome *c* activation.

The *S. cerevisiae* HAP2, HAP3 and HAP5 proteins appear to be highly conserved through yeast to humans. Functional mammalian equivalents (Chodosh *et al.*, 1988) and homologs of HAP2 and HAP3 have been isolated from *Schizosaccharomyces pombe* (Olesen *et al.*, 1991; Xing *et al.*, 1993), *Kluyveromyces lactis* (NGuyen *et al.*, 1995; Mulder *et al.*, 1994), mouse (Hooft van Huijsduijnen, 1990), and rat (Maity *et al.*, 1990). A homolog of HAP2 has been isolated from humans (Becker *et al.*, 1991), and a HAP3 homolog (*hapC*) and a HAP5 homolog (*hapE*) have been isolated from *A. nidulans* (Papagiannopoulos *et al.*, 1996a; Papagiannopoulos *et al.*, 1996b). Studies with the rat CCAAT-binding complex (CBF) suggest that the CBF-C component might represent the mammalian HAP5 homolog (Maity *et al.*, 1992).

A number of yeast genes required for respiration (for example *HEM1*, *COX4*, and *CYT1* genes) are activated by the yeast CCAAT-binding factor. Therefore, the HAP2/3/5 heterotrimer has been suggested to be a global yeast gene regulator (Zitomer and Lowry, 1992; Chodosh *et al.*, 1988). In addition, since only homologs of HAP2/3/5 have been obtained, it was suggested that this heterotrimer corresponds to the general transcription factor found in other species.

1.3.3 Other factors affecting *CYC7* expression

Minor factors affecting transcription of the yeast *CYC7* gene but not the *CYC1* gene are heat shock, stationary-growth phase and cAMP levels (Pillar & Bradshaw, 1991). During stationary-growth phase, transcription of *CYC7* is strongly induced (40-70 fold), whereas transcription of *CYC1* is greatly decreased. Heat shock increases the level of *CYC7* transcription by 4-6 fold aerobically and 30 fold anaerobically. Low cAMP levels (which have also been shown to induce several heat shock proteins) also induce the *CYC7* gene. Therefore the response of the *CYC7* gene to these environmental factors may possibly reflect a unique role for this gene.

The heme-dependent regulatory protein HAP1 which activates transcription of *CYC7* via the UAS is unlikely to be involved in both heat shock and stationary induction of the *CYC7* gene. HAP1 only activates *CYC7* in aerobic conditions and therefore cannot be involved in heat shock induction which occurs both aerobically and anaerobically. Stationary growth phase is probably not heme mediated, because *CYC1* which is activated by HAP1 is repressed during these conditions.

1.4 REGULATION OF THE *KLUYVEROMYCES LACTIS* CYTOCHROME *C* GENE (*KlCYC1*)

The aerobic respiratory yeast *K. lactis*, has one cytochrome *c* gene, *KlCYC1*, which produces two transcripts of different sizes (Freire-Picos *et al.*, 1995). This is in contrast to the fermentative yeast *S. cerevisiae* which produces a single transcript for each of *CYC1* and *CYC7*. Studies with this aerobic yeast may help to provide further understanding into the regulation of respiration in aerobic organisms, such as *A. nidulans*.

Consensus sequences for the recognition of the *S. cerevisiae* regulatory HAP2/3/4/5 complex and for the HAP1 transcription factor have been found in the upstream region of the *KlCYC1* gene (Freire-Picos *et al.*, 1993).

The *KlCYC1* gene was found to be transcriptionally activated by oxygen through the control of heme in the same way as found for the *S. cerevisiae CYC1* gene (Freire-Picos *et al.*, 1995). In addition, heterologous expression of the *KlCYC1* gene in a *hap1* mutant strain of *S. cerevisiae* decreased the level of cytochrome *c* expression in the mutant strain compared to wild type (Freire-Picos *et al.*, 1995). These results support the hypothesis that *KlCYC1* expression is regulated by the presence of oxygen through a factor similar to HAP1.

Functional homologs of the *S. cerevisiae HAP2* and *HAP3* genes have been isolated and cloned from *K. lactis* (NGuyen *et al.*, 1995; Mulder *et al.*, 1994). However, the significance of putative HAP2/3 regulatory sites remains unclear due to apparently conflicting results from two groups.

One group showed from the heterologous expression of *KlCYC1* in a *S. cerevisiae hap2* mutant, that the *HAP2* gene was essential for the expression of the *K. lactis* cytochrome *c* gene as it is for that of *S. cerevisiae* (Freire-Picos *et al.*, 1995). In addition the *K. lactis* strain Y1140 was found to be subject to glucose repression (Freire-Picos *et al.*, 1995), thus supporting the function of the HAP2/3/4/5 complex in *KlCYC1* expression.

Another group's results showed that similarly to *S. cerevisiae*, expression of the *KlCYC1* gene is significantly induced on a non-fermentable carbon source (Mulder *et al.*, 1995). However, in contrast to *S. cerevisiae*, they found that the *KlCYC1* gene was not glucose repressed, because expression levels of *KlCYC1* were not affected with the addition of glucose to a non-fermentable carbon source. However, glucose repression has been indicated to be a strain-dependent feature in *K. lactis*. The *K. lactis* nuclear genes *KlQCR7* and *KlQCR8* (encoding subunits VII and VIII respectively of the mitochondrial bc_1--complex) were also not repressed by glucose (Mulder *et al.*, 1995). These results are in agreement with observations made in 1966 by De Deken, who observed that various *K. lactis* respiratory enzymes were not affected by glucose repression, but contrasts to the findings above by Freire-Picos *et al.* (1995). However, both groups employed different strains of *K. lactis*, which may explain the different results obtained.

Further work by Mulder *et al.* (1994) and NGuyen *et al.* (1995), indicated that the *K. lactis HAP2* and *HAP3* genes were not essential for expression of the mitochondrial respiratory system. Both *KlHAP2* and *KlHAP3* genes were inactivated in *K. lactis*, but in contrast to the situation in *S. cerevisiae*, these *hap* null mutants did not alter the ability of *K. lactis* to respire and grow on non-fermentable carbon sources. Interestingly, the *KlHAP2* gene was able to be correctly expressed when placed in *S. cerevisiae* and to complement a HAP2 deficiency (NGuyen *et al.*, 1995). Similar results were obtained for the cytochrome *c* gene from the starch fermenting yeast *Schwanniomyces occidentalis* ($CYC1_o$), which was not repressed by glucose in native *S. occidentalis* cells, but was glucose repressed when placed in *S. cerevisiae* strains (Amegadzie *et al.*, 1990). Both these results tend to suggest that the HAP2/3/4/5 complex has no role in glucose repression in either *K. lactis*, or *S. occidentalis* cytochrome *c* genes, but will regulate these genes in response to carbon source when placed in a *S. cerevisiae* background.

In summary, the *KlCYC1* gene appears to be transcriptionally activated in response to oxygen through the binding of a HAP1-like transcription factor to its upstream regulatory region. However, the role of the HAP2/3/4/5 complex in the regulation of the *KlCYC1* is more unclear. The disruptions of the *KlHAP2* and *KlHAP3* genes performed in *K. lactis* by Mulder *et al.* (1994) and NGuyen *et al.* (1995), are more meaningful than the *KlHAP2* disruption performed in *S. cerevisae* by Freire-Picos *et al.* (1995) because the gene disruptions were carried out under the true environmental influences of the organism. Therefore these results suggest that either the HAP2/3/4/5 complex has a different function in *K. lactis,* or perhaps that additional controls or factors (specific to respiratory or fermentative yeasts) may be required to adjust the general function of the HAP2/3/4/5 complex.

1.5 REGULATION OF THE *ASPERGILLUS NIDULANS* CYTOCHROME *C* GENE (*cycA*)

Like most lower eukaroytes, *A. nidulans* contains only one copy of the cytochrome *c* gene (*cycA*) (Raitt *et al.,* 1994). The predicted nucleotide sequence has approximately 72% identity to the nucleotide sequence of the *S. cerevisiae CYC1* gene (Raitt *et al,* 1994). Like the *S. cerevisiae* cytochrome *c* genes, expression of the *cycA* gene is transcriptionally regulated by oxygen availability, and is induced by heat shock conditions (Raitt, 1992). However, in contrast to *S. cerevisiae*, *cycA* expression is not affected by glucose repression. The *cycA* gene may be transcriptionally activated by oxygen in a similar manner to that established for the cytochrome *c* genes of *S. cerevisiae* (see Section 1.3), but further experimental evidence is required to determine this.

1.4.1 Regulation of the *cycA* gene in response to Environmental Parameters

Transcription of the *cycA* gene is activated by oxygen with a 10 fold increase observed during the transition from anaerobic to aerobic conditions (Raitt *et al.,* 1994). Similar expression levels of the *S. cerevisiae CYC7* gene occur in aerobic conditions (Prezant *et al.,* 1987).

CycA transcript levels remain constant when *A. nidulans* is grown in the presence of fermentable (glucose) or non-fermentable (galactose) carbon sources (Raitt *et al.*, 1994). Therefore the *cycA* gene is not glucose repressed as observed for the *S. cerevisiae* cytochrome *c* genes. However, it may be that glucose repression in *A. nidulans* is a strain-dependent feature, as found for *K. lactis.* However, the lack of glucose repression is not surprising since *A. nidulans* is an obligate aerobe, where reduced activity of the

respiratory chain may compromise the organism's ability to generate ATP, and to support cell growth.

Under heat shock conditions, a 3-4 fold induction of the *cycA* transcript occurs (Raitt *et al.*, 1994). A similar response is observed for the *S. cerevisiae CYC7* gene (Pillar and Bradshaw, 1991).

1.4.2 Analysis of the 5' Region of the *cycA* gene

1.4.2.1 Putative HAP1 Binding Site

A putative HAP1 binding site was located 382-400 nucleotides upstream of the published translation initiation site, and was found to contain the greatest homology with the *CYC7* HAP1 binding site (Raitt, 1992). The sequences responsible for HAP1 mediated oxygen regulation of the yeast *CYC1* and *CYC7* genes, are shown aligned to the *cycA* putative HAP1 binding sequence (Figure 1.2; Raitt, 1992). Ten out of eleven nucleotides which contact HAP1 in the *CYC7* promoter are also conserved in the *cycA* promoter region. Due to these sequence similarities, it was proposed that a HAP1-type gene product may be involved in the oxygen induced transcriptional activation of the *A. nidulans cycA* gene.

1.4.2.2 Putative Upstream Opening Reading Frame (uORF)

At 120 nucleotides before the translational start point, an upstream open reading frame was found which could encode a 10 amino acid polypeptide including five consecutive serines (Raitt *et al.,* 1994). This uORF may be involved in a translational regulatory mechanism, similar to that observed in some genes of *S. cerevisiae* such as *GCN4*. The *GCN4* gene is a transcriptional activator of many genes involved in the biosynthesis of ten amino acids. It contains four ORFs in its mRNA leader sequence which are involved in translational control. In response to amino acid availability, these uORFs can control ribosome flow so initiation occurs at the correct codon (Abastado *et al.,* 1990).

1.5 AIMS AND OBJECTIVES

The aim of this study was to investigate the role of promoter sequences, in particular, the putative HAP1 binding site, in the regulation of the *A. nidulans* cytochrome *c* gene.

In vitro methods for promoter analysis include gel mobility shift experiments and DNaseI footprinting. However, promoter-reporter vectors present a powerful *in vivo* method to

Figure 1.2 Comparison of HAP1 binding sites in *S. cerevisiae* and *A. nidulans*

Nucleotides shown in blue are directly involved in the binding of the HAP1 protein. Red nucleotides represent those conserved between the *CYC7* and the *cycA* promoters. The numbering of the nucleotide is relative to the ATG initiation codon.

```
                    -320        -330        -340
                      I           I           I
S. Cerevisiae   5'  GTCATCGTCCGTAAACCCCGGCCAA 3'
CYC1            3'  CAGTAGCAGGCATTTGGGGCCGGTT 5'

                    -250        -240        -230
                      I           I           I
S. Cerevisiae   5'  AGCTAATAGCGATAATAGCGAGGGC 3'
CYC7            3'  TCGATTATCGCTATTATCGCTCCCG 5'

                  -400        -390        -380
                    I           I           I
A. nidulans     5'  GGCTAAGGGCGGTGACAGCTACTCT 3'
cycA            3'  CCGATTCCCGCCACTGTCGATGAGA 5'
```

investigate the functional role of promoter elements involved in gene regulation. This type of analysis, unlike the *in vitro* methods mentioned above, allows gene expression to be investigated under the true environmental influences of the organism. I therefore chose to use this method to study the putative HAP1 motif in the *cycA* gene.

For the filamentous fungus *A. nidulans*, integration vectors have been constructed containing the *lacZ* reporter gene of *Escherichia coli* (van Gorcom *et al,* 1986) to allow investigation of promoter sequences involved in gene regulation. Transcriptional or translational gene fusions can be obtained, and quantitative analysis of expression signals is attainable, since these vectors preferentially integrate at a unique place in the genome (the *arg*B locus), in one copy (van Gorcom *et al*, 1986). Streatfield *et al* (1991), have successfully used this type of *lacZ* gene fusion vector to carry out a functional analysis of the expression of the 3`-phosphoglycerate kinase (*pgk*) gene in *A. nidulans.*

In this project my initial aim was to prepare vectors which contained the *Escherichia coli lacZ* gene translationally fused to either the entire *cycA* promoter region, or the entire promoter region with deletion of the putative HAP1 binding site. Following transformation of these constructs into *A. nidulans,* and confirmation by Southern blotting that they had integrated into the *arg*B locus, the expression of the *lacZ* reporter would be assessed. This assessment would involve ß-galactosidase assays and Northern blots to measure transcription and translation of the *lacZ* reporter under different environmental conditions. This study should allow us to determine the significance of the putative HAP1 binding site for the regulation of the *A. nidulans* cytochrome *c* gene. In addition I hoped to confirm that observed differences in mRNA levels under low and high oxygen levels are due to transcriptional regulation and not simply due to differences in mRNA half-life. Genuine transcriptional regulation should be reflected in the promoter reporter fusions whilst differences in mRNA half-life would not. Information regarding possible translational regulation by the uORF should also be elucidated from this study.

CHAPTER 2
MATERIALS AND METHODS

2.1 FUNGAL AND BACTERIAL STRAINS, LAMBDA CLONES AND PLASMIDS

These are listed in Table 2.1.

2.2 MEDIA

All media was prepared with Milli-Q water and sterilised by autoclaving at 121°C for 15 minutes. Both solid and liquid media was cooled to approximately 50°C before the addition of media supplements.

2.2.1 Luria Broth (LB) Medium (g/L): tryptone, 10.0; yeast extract, 5.0; NaCl, 5.0; 1% D-glucose. The pH was adjusted to 7.0 before autoclaving. For solid media, agar was added to 1.5%. When required, ampicillin, isopropylthio-β-D-galactoside (IPTG) and 5-bromo-4-chloro-3-indolyl-β-D-galactoside in dimethylformamide (X-gal) were added at a final concentration of 100 μg/ml, 30 μg/ml and 60 μg/ml respectively.

2.2.3 MYG Media (g/L): malt extract, 5.0; yeast extract, 2.5; 2% glucose; $CuSO_4.5H_2O$, trace; agar, 20.0.

2.2.4 SOC Medium: 2% (w:v) Tryptone, 0.5% (w:v) Yeast Extract, 10 mM NaCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$, 20 mM Glucose.

2.2.5 NZCYM (g/L): NZ amine (type-A hydrolysate of casein), 10.0; casamino acids, 1.0; NaCl, 5.0; yeast extract, 5.0; $MgSO_4.7H_2O$, 2.0. The pH was adjusted to pH 7.5 before autoclaving.

2.2.6 TB Top Agarose (g/L): tryptone, 10.0; NaCl, 5.0; agarose, 8.0. After autoclaving, the media was cooled to 45-50°C and supplemented with 1 M $MgSO_4.7H_2O$.

2.3 COMMON BUFFERS AND SOLUTIONS

2.3.1 1 x TAE Buffer: 40 mM Tris-acetate, 11.4 ml/L glacial acetic acid, and 1 mM EDTA; pH 8.5.

Table 2.1 Strains, λ clones and plasmids

Strain, Plasmid or l Clone	Relevant Characteristics	Source or Reference
Fungal Cultures		
A.nidulans		
A18	(R153) *w*A3, *pyro*A4	Waldron and Roberts (1973)
Bacterial Cultures		
E. coli		
XL-1	*sup*E44 *hsd*R17 *rec*A1 *end*A1 *gyr*A96 *thi*-1*rel*A1 *lac*$^-$F'[*pro* AB$^+$ *lac*I^q *lac*ZDM15 Tn10(*tet*^{r})]	Bullock *et al.* (1987)
DH1	F' *sup*E44 *rec*A1 *end*A1 *gyr*A96 *thi*1 *rel*A1 *hsd*R17 l$^-$	Hanahan (1983)
LE392	*sup*E44 *sup*F58 *hsd*R514 *gal*K2 *gal*T22 *met*B1 *trp*R55 *lac*Y1	Borck *et al.* (1976)
Plasmids		
R40	0.9 kb *Hin*dIII-*Xho*I fragment of *A. nidulans cycA* gene contained in pICI9R	Raitt (1992)
R41	2.3 kb *Pst*I-*Hin*dIII fragment of *A. nidulans cycA* gene contained in pBS	Raitt (1992)
R109	(pAN923-42B) *lacZ* Ap^r ^{t}*trp*C	Van Gorcom *et al.* (1986)
R116	(pAN5-d1) *gpd*A *lacZ* Ap^r ^{t}*trp*C *arg*B-*Bgl*II	Punt *et al.* (1990)
R117	(pAN923-42B$_{BglII}$) *lacZ* Ap^r ^{t}*trp*C *arg*B-*Bgl*II	Punt *et al.* (1990)
R120	0.68 kb *Bam*HI fragment of *A. nidulans cycA* gene contained in R117 (wrong orientation)	this study
R121	0.68 kb *Bam*HI fragment of *A. nidulans cycA* gene contained in pUC18	this study
R122	0.68 kb *Bam*HI fragment of *A. nidulans cycA* gene contained in R117 (correct orientation)	this study
R127	2.1 kb *Eco*R1 fragment of *A. nidulans cycA* gene contained in pUC18	this study
pUC18	Ap^r *lacZ*'	Yanish-Perron *et al.* (1995)
λ clones		
λGEM-11	The arms are derived from EMBL3A and the stuffer fragment is derived from λ2001	Frischauf *et al.* (1983); Karn (1984)
λLM5	λGEM-11 clone containing *cycA* from A18	this study
λLM9	λGEM-11 clone containing *cycA* from A18	this study
λLM19	λGEM-11 clone containing *cycA* from A18	this study

2.3.2 1 x TBE Buffer: 89mM Tris HCl, 89 mM boric acid, 2.5 mM Na_2EDTA.

2.3.3 TE Buffer (10/1): 10 mM Tris-HCl, and 1 mM Na_2EDTA.

2.3.4 TE Buffer (10/0.1): 10 mM Tris-HCl and 0.1 mM Na_2EDTA.

2.3.5 10 x Sequencing TBE Buffer: (g/l): Tris, 162; Na_2EDTA, 9.5; and boric acid, 27.5.

2.3.6 TES (10/1/100): 10 mM Tris-HCl (pH 8.0), 1 mM Na_2EDTA (pH 8.0), and 100 mM NaCl.

2.3.7 STET Buffer: 8% sucrose, 5% Triton X-100, 50 mM Na_2EDTA (pH 8.0), and 50 mM Tris-HCl (pH 8.0).

2.3.8 SM Buffer: (g/L): NaCl, 5.8; $MgSO_47H_2O$, 2.0; 1M Tris-HCl (pH 7.5), 50 ml; and 2% gelatin, 5 ml.

2.3.9 Hybridisation Solution: (per litre): 150 ml of 20 x SSC (Section 2.3.12), 20 ml of 50 x Denhardt's Solution (Section 2.3.10), 30 ml of 10% SDS, 5 ml of salmon sperm DNA (at 50 μg/ml).

2.3.10 50 x Denhardt's Solution: (g/L): ficoll, 10.0; polyvinylpyrolidone, 10.0; bovine serum albumin, 10.0. Filter sterilised.

2.3.11 Tris-Equilibrated Phenol was prepared by melting solid phenol at 50°C followed by the addition of 0.1 M Tris-HCl (pH 8.0) and thorough mixing. After 15-30 minutes, the buffer was decanted and the procedure repeated twice. Hydroxyquinoline was added to a final concentration of 0.1% (w/v) and the equilibrated phenol stored under 100 mM Tris-HCl (pH 8.0) at 4°C.

2.3.12 20 x SSC: 3 M NaCl and 0.2 M tri-sodium citrate.

2.3.13 10 x Gel Loading Buffer: 2 M Urea, 50% glycerol, 50 mM Tris acetate, 0.4% Bromophenol Blue, 0.4% Xylene cyanol.

2.3.14 Standard DIG Buffer: 5 x SSC, 1.0% (w/v) blocking reagent stock solution (Section 2.3.18), 0.1% N-lauroylsarcosine, and 0.02% SDS.

2.3.15 DIG Buffer 1: 100 mM maleic acid, and 150 mM NaCl: pH 7.5.

2.3.16 DIG Buffer 2: 1% blocking reagent stock soluton (Section 2.3.18) diluted in Buffer 1 (Section 2.3.15).

2.3.17 DIG Buffer 3: 100 mM Tris-HCl (pH 9.5), and 100 mM NaCl.

2.3.18 Blocking Reagent stock solution: 10% (w/v) blocking reagent (Boehringer Mannheim DIG kit) diluted in DIG buffer 1 (Section 2.3.15).

2.3.19 Alkaline Lysis Solution I: 50 mM glucose, 25 mM Tris HCl pH 8.0, and 10 mM EDTA : pH 8.0.

2.3.20 Alkaline Lysis Solution II: 0.2 M NaOH, and 1% SDS.

2.3.21 Alkaline Lysis Solution III: 60ml 5 M potassium acetate, 11.5ml glacial acetic acid, and 28.5ml H_2O.

2.4 GROWTH AND MAINTENANCE OF CULTURES

Bacterial cultures were grown at 37°C on LB agar plates, supplemented with antibiotic when appropriate, and sealed with parafilm for storage at 4°C. Fungal cultures were grown on MYG plates, either at 37°C overnight or at 30°C for 2 days, and sealed with parafilm for storage at 4°C.

2.5 DNA PREPARATIONS

2.5.1 Rapid Boiling Plasmid Preparation

This method for preparing plasmid DNA was performed as described by Holmes and Quigley, (1981).

2.5.2 Small Scale Alkaline Lysis Plasmid Preparation

The pAN923-42B vector (Table 2.1) was prepared by the small scale alkaline lysis method followed by phenol-chloroform extraction and ethanol precipitation (Section 2.5.7).

This method was based on that of Sambrook *et al.*, (1989). 3 ml of LB was inoculated with a single bacterial colony and incubated overnight at 37°C with vigorous shaking. 1.5 ml of the culture was pelleted by centrifugation at 12000 rpm (in a microcentrifuge) for 30 seconds. After removing the supernatant the pellet was resuspended in 100 µl of solution I (Section 2.3.19). 200 µl of freshly prepared solution II (Section 2.3.20) was added and mixed by inverting the tube rapidly five times and placing on ice. 150 µl of solution III (Section 2.3.21) was added, followed by gentle vortexing, and incubated on ice for 5 minutes. After centrifugation at 12000 rpm (in a microcentrifuge) for 5 minutes, the supernatant was transferred to a fresh tube. An equal volume of phenol:chloroform was added, the tube vortexed and centrifuged for 2 minutes. The aqueous layer was transferred to a fresh tube, and the DNA was precipitated with 2 volumes of 100% ethanol. The DNA was pelleted by centrifugation for 5 minutes at 12,000 rpm (in a microcentrifuge) and then dried under vacuum before resuspension in TE buffer (10/0.1) or H_2O.

2.5.3 Large Scale Alkaline Lysis Plasmid Preparation

The modified pAN923-42B$_{BglII}$ vector (Table 2.1) was prepared by the large scale alkaline lysis method and purified by using a CsCl gradient (Section 2.5.4).

This method was based on that of Ish-Horowicz and Burke (1981). 250 ml of LB was inoculated with a 1/100 dilution from a 3 ml overnight culture of *E. coli*, and incubated overnight at 37°C on a shaking platform (at 250 rpm). Cells were harvested by centrifugation at 8000 rpm (centrifugation was carried out in a GSA rotor throughout) for 10 minutes, then washed in 100 ml of TE (10/0.1) by resuspension and pelleted by centrifugation at 8000 rpm for 10 minutes. The cells were resuspended in 30 ml of solution I (Section 2.3.19), 3 ml of lysozyme (at 50 mg/ml) was added and the tube incubated at room temperature for 10 minutes. 60 ml of solution II (Section 2.3.20) was added, mixed by inversion and the tube placed on ice for 10 minutes. 45 ml of solution III (Section 2.3.21) was added, mixed by inversion followed by incubation on ice for 10 minutes. After centrifugation at 8000 rpm for 10 minutes, the supernatant was transferred to a fresh tube and the DNA precipitated by the addition of 0.6 volumes of cold isopropanol followed by incubation at room temperature for 20 minutes. DNA was pelleted by centrifugation at 12000 rpm for 10 minutes, washed with 75 ml of 70% ethanol, and air dried before resuspension in 3.5 ml of TE buffer (10/0.1) with heating for 10 minutes at 65°C.

2.5.4 Purification of Plasmid DNA by Cesium Chloride-Ethidium Bromide Gradient.

CsCl was added to the plasmid DNA solution in a ratio of 1.05 g/ml, and ethidium bromide (at 10 mg/ml) was added in a ratio of 75 µl/ml. The mixture was vortexed and put at 4°C overnight. After this time, the solution was centrifuged at 12000 rpm in a SS34 rotor for 10 minutes to remove precipitated material, and the refractive index adjusted to between n=1.3860-1.3920 by the addition of either TE buffer or CsCl. Plasmid DNA was purified by centrifugation at 55000 rpm in a Sorvall combi TV865 rotor for 5 hours. The plasmid band was removed with an 18 gauge hypodermic needle and syringe under long wavelength UV light. Ethidium bromide was removed from the DNA sample by repeated extraction with equal volumes of SSC saturated isopropanol (prepared by stirring equal volumes of 20 x SSC and isopropanol for several hours). The DNA was dialysed at 4°C against TES (10/1/100) (Section 2.3.6) with 4 changes of buffer and stored at -20°C. DNA purity and concentration were calculated by spectroscopy as in Section 2.6.4.2.

2.5.5 Preparation of DNA from Fungal Cultures

This method for isolating DNA from fungi was based on that of Raeder and Broda (1985). Mycelium was grown on MYG plates overlaid with sterile cellophane disks. 50 µl of spore suspension (Section 2.11.1) was spread on to the cellophane disks and after incubation for 20 hours at 37°C mycelium was scraped off and freeze dried. Approximately 400 mg (dry weight) of mycelium was ground to a fine powder with liquid nitrogen, 4 ml of extraction buffer (22 mM Tris pH 8.5, 250 mM NaCl, 25 mM EDTA, 0.5% SDS) was added and the mixture vortexed. At this stage 2.8 ml of phenol (equilibrated with extraction buffer) and 1.2 ml of chloroform was added, followed by centrifugation at 12000 rpm, (centrifugations were performed in a GSA rotor at 4°C through out) for 40 minutes. Supernatant was transferred to a fresh tube and the sample was re-extracted with phenol and chloroform as before, followed by centrifugation at 12000 rpm for 15 minutes. To remove RNA, 100µg/ml of RNase was added to the supernatant and incubated at 37°C for 30 minutes. An equal volume of phenol:chloroform was added and the tube centrifuged for 15 minutes at 12000 rpm. Chloroform was added to the aqueous phase and centrifuged for 20 minutes at 12000 rpm. DNA was precipitated with 0.54 volumes of isopropanol and centrifuged for 5 minutes at 12000 rpm. The DNA pellet was washed with 70% ethanol, dried and resuspended in TE buffer (10/0.1).

2.5.6 Preparation of Lambda Phage DNA by Liquid Lysate

λ phage DNA was prepared using a modification of the liquid lysate method from Current Protocols in Molecular Biology (1994). 3 ml of LB supplemented with 0.2% maltose and 10 mM $MgSO_4.7H_2O$, was inoculated with a single bacterial colony of LE 392 and incubated overnight at 37°C. 100 µl of overnight bacterial culture was combined with 100 µl of titred diluted phage to give 10^6 or 10^7 phage/100 µl. The phage mixture was transferred to 50 ml of NZCYM and shaken at 300 rpm at 37°C until lysis occurred, indicated by the culture clearing. 200 µl of chloroform was added upon clearing to lyse any remaining cells and the culture shaken at 37°C for a few minutes. The culture was then immediately harvested by centrifugation at 10000 rpm (in a GSA rotor) for 10 minutes. The lysate was transferred into a fresh tube and stored at 4°C. RNase and DNase was added to the liquid lysate to give a final concentration of 10 µg/ml and incubated at 37°C for 30 minutes. Phage were precipitated by the addition of NaCl (to 0.5 M), and PEG 6000 (to 10% (w/v). Following incubation on ice for 1 hour, the phage were pelleted at 5000 rpm (in a GSA rotor) for 10 minutes at 4°C, and the supernatant discarded. The pellet was resuspended in 1 ml of SM buffer (Section 2 3.8) and the solution transferred to an eppendorf tube. Remaining bacterial debris was pelleted by centrifugation for 10 minutes at 12000 rpm, the supernatant transferred to a fresh tube and proteinase K added to a final concentration of 0.1 mg/ml. After incubation for 30 minutes at 37°C, the phage suspension was extracted with phenol:chloroform, vortexed for 20 minutes on an eppendorf platform, and centrifuged for 5 minutes at 10000 rpm (in a microcentrifuge). Phenol:chloroform extractions were repeated, then an equal volume of chloroform was added, vortexed for 5 minutes, and centrifuged for 5 minutes as above. The upper aqueous phase was taken and DNA precipitated by the addition of 2 volumes of ethanol and 0.3 M NH_4Ac. DNA was pelleted by centrifugation for 10 minutes, washed with 70% ethanol, vacuum dried and resuspended in TE buffer (10/0.1).

2.5.7 Purification of DNA by Phenol-Chloroform Extraction and Ethanol Precipitation

To the DNA sample an equal volume of Tris-equilibrated phenol:chloroform (1:1) was added, vortexed, and centrifuged at 13000 rpm for 5 minutes. The aqueous phase was transferred to a fresh tube, and an equal volume of chloroform added. The mixture was vortexed and then centrifuged for 5 minutes and the aqueous phase transferred to a fresh tube. To ethanol precipitate the DNA, 1/10 volume of 0.3M sodium acetate plus 3 volumes of 95% ethanol was added to the aqueous phase and mixed. The DNA was pelleted by centrifugation for 15 minutes. The ethanol was poured off, and the pellet

washed with 70% ethanol by centrifugation for 3 minutes, dried under vacuum for 5 minutes and resuspended in TE buffer (10/0.1) or H_2O.

2.6 STANDARD PROCEDURES

2.6.1 Restriction Enzyme Digests

DNA digests were carried out with the appropriate restriction enzyme buffer at 37°C, for 1.5-3 hours, in a total volume at least 10 times the combined volume of DNA and enzyme. When possible, commercial buffers specifically matching the restriction enzyme were used. RNA was removed by adding RNase to a final concentration of 0.5 μg/μl and incubating for a further 10 minutes at 37°C. Digests were checked for completion by electrophoresis on a 1-2.5% agarose mini-gel.

2.6.2 Agarose-gel Electrophoresis of DNA

DNA fragments were size fractionated by electrophoresis through a gel ranging from 0.8-2.5% agarose (depending on the expected fragment sizes), in TAE or TBE buffer at 80-100 V. Large preparative gels were usually electrophoresised overnight at 25-30 V. Agarose gels were stained in ethidium bromide, 10 minutes for mini-gels or 30 minutes for preparative gels, and then destained in Milli-Q water, observed under short wave UV light and photographed with polaroid 667 film.

2.6.3 Determination of Molecular Weights

DNA fragments were sized by running the DNA sample alongside any of the following standards:

(i) *Hind*III/*Eco*RI double digest of lambda DNA
(ii) 1 kb DNA ladder (BRL)
(iii) *Hinf*I digest of pBR322

Relative mobilities of DNA fragments were measured, and the molecular weights determined using the GEL 1.01 software program (1993 Jean-Michel Lacroix).

2.6.4 Determination of DNA Concentration

2.6.4.1 Quantification of DNA Concentration by Electrophoresis

A series of Lambda DNA or pBR322 DNA concentration standards were run on a 1.0-2.5% agarose minigel alongside the DNA sample. DNA sample concentration was estimated by comparing the intensity of fluorescence to that of the DNA concentration standards.

2.6.4.2 Spectrophotometric Determination of DNA Concentration and Purity

This method was used to the estimate the DNA concentration of DNA solutions of high concentration and purity. The DNA sample was appropriately diluted and the absorbance measured at both 260 nm and 280 nm. The DNA concentration was calculated on the assumption that an absorbance of 1.0 at 260 nm is equivalent to 50 μg/ml. Purity was indicated by the 260 nm/280 nm ratio, where pure DNA has an OD_{260}/OD_{280} ratio of 1.8-2.0 (Current Protocols in Molecular Biology, 1989; Glasel, 1995).

2.7 AMPLIFICATION OF DNA BY THE POLYMERASE CHAIN REACTION (PCR)

PCR reactions were set up on ice using a cocktail which contained all common reagents used for n+1 PCR reactions. Uncommon reagents were aliquoted out separately. The PCR reactions were carried out in either 25 μl or 50 μl volumes in 0.2 ml strip tubes (Biotek). The final concentrations of each component in 1 reaction were, 1 x Tfl buffer (Boehringer Mannheim), 1.5 mM $MgCl_2$, 1.25 mM dNTPs, 1 unit/25μl of Tfl polymerase (Boehringer Mannheim), 2 μM forward primer, 2 μM reverse primer, 0.1 ng to 1 ng of template DNA, and MilliQ water up to the total reaction volume. A negative control containing water instead of DNA template was included in each PCR run. After mixing, the PCR reaction tubes were placed in a thermal cycler (Corbett FTS-960) preheated to 95°C.

For PCR amplification of the *A. nidulans cycA* promoter region from cloned genomic DNA (Raitt *et al.* (1994), published version), the following PCR cycle conditions were used:

Initial Step	95°C	2.5 minutes	
Denaturation	95°C	1 minute	x 30
Annealing	60°C	1 minute	
Elongation	72°C	2 minutes	
Elongation	72°C	5 minutes	
Cool	4°C	1 minute	

For repeated PCR amplification of the 596bp and 298bp *A. nidulans* cDNA products, the following optimised PCR cycle conditions were used:

Initial Step	95°C	2.5 minutes	
Denaturation	95°C	30 seconds	x 25
Annealing	65°C	30 seconds	
Elongation	72°C	1 minute	
Elongation	72°C	5 minutes	
Cool	4°C	1 minute	

On completion of PCR, the reactions were checked on a 2.5% agarose minigel. Primers used in the PCR reactions are listed in Table 2.2.

2.8 DNA CLONING PROCEDURES

Fragments to be cloned were gel purified (Section 2.8.1), ligated into the appropriate CAP-treated vector (Sections 2.8.2 and 2.8.3), and transformed into a suitable *E. coli* host strain (Section 2.8.4, Table 2.1). Transformants were screened for recombinants by one of the following methods. When blue/white selection of recombinants was available (eg. when using pUC18 vector), the preferred method involved restriction digest analysis (Section 2.6.1) of plasmid DNA obtained by the rapid boiling method (Section 2.5.1). When there was no blue/white selection and the colony numbers were large, colonies were screened either by colony hybridisation (Section 2.9) or PCR (Section 2.7) using primers designed to specifically detect the presence of insert (refer to Table 2.2).

Table 2.2 PCR and Sequencing Primers

Primer	Size	Tm°C*	Sequence (5' to 3')	Source
RB1	24 mer	72	gTggATCCTACCTCTA TTggTCA	Saunders (1993)
RB2	24 mer	74	gTggATCCTCgAAgAgC TTAgCAC	Saunders (1993)
RB3	26 mer	76	gTgTCgACTCTCCTggT AAgTAgTTT	Saunders (1993)
RB4	25 mer	72	gTgTCgACAAgggTTATg TAAAgAg	Saunders (1993)
LG1	24 mer	72	gTAgATCTCgTCTAATC gCgAACg	this study
LG2	25 mer	74	CCAATCAgAggTAggAT GATTCAgC	this study
LG3	24 mer	76	gACgCTggTTACTTggAg CTCTgC	this study
LG4	24 mer	70	AgTgCAgAgCTCCAAgT AAccAgC	this study
LM5	23 mer	76	CgAggTgTTCCCTCAgC ACgAgC	this study
LM6	25 mer	76	CAggAATAggATTgCTC gTgCTgAg	this study
LM7	25 mer	72	ggTAgTTAACgACTTTA CCATTgCg	this study
CYC8	23 mer	66	gTggTgAAgCTAAATgT TgCAAg	this study
pUC/M13 Forward	22 mer	70	gCCAgggTTTTCCCAgT CACgA	Perkin Elmer
pUC/M13 Reverse	24 mer	70	gAgCggATAACAATTTC ACACAgg	Perkin Elmer

* Calculated as Tm°C=2(A+T)+4(G+C) from Itakura *et al.*, (1984).

2.8.1 Gel Purification of DNA

DNA fragments to be purified from either a PCR reaction or restriction enzyme digest, were run on a 1% sea-plaque low melting point agarose gel in TAE buffer. After staining in ethidium bromide, the DNA fragments were visualised under long wave UV light, and the appropriate band cut out with a sterile scalpel blade. The DNA was extracted from the agarose using the following commercial kits according to the manufacturers instructions:

(i) Gibco BRL Glassmax DNA Isolation Spin Cartridge System (Life Technologies Ltd).
(ii) Bio 101 Geneclean Kit (Bio 101).
(iii) Wizard PCR Preps DNA Purification System for Rapid Purification of DNA Fragments (Dade Diagnostics Ltd).

2.8.2 CAP Treatment of DNA

Vector DNA (5-10 μg) was digested to completion with the desired restriction enzyme, and the enzyme in-activated by heat (if applicable) or by phenol and chloroform extraction (Section 2.5.8). 1 unit of calf alkaline phosphatase (CAP) was added to the DNA and incubated for 30 minutes at 37°C. This was repeated once, then proteinase K and SDS were added to a final concentration of 50 μg/ml and 1%, respectively followed by incubation for 1 hour at 37°C. The DNA was subsequently purified as described in Section 2.5.7.

2.8.3 Ligation Reactions

Ligation reactions usually contained 1-3 times the molar ratio of insert:vector, (using a minimum of 20 ng of vector), 2.0 μl of 10 x ligation buffer, 1.0 μl of undiluted T4-DNA ligase, and Milli-Q water to a total volume of 10 μl. Ligation was carried out at room temperature for at least 1 hour and then overnight at 4°C, (or ligated at room temperature only). For ligations that did not work under these conditions, reactions were incubated at 14°C for 2 hours and than at 4°C overnight. Ligations were checked by taking a 1 μl aliquot before the addition of ligase, and 1 μl after the ligation incubation period. Both samples were run on an agarose minigel and compared. The presence of higher molecular weight bands and the disappearance of insert DNA was indicative of successful ligation. Transformation of the ligation mixture was then carried out as described in Section 2.8.4.

2.8.4 Transformation of *Escherichia coli*

Preparation of electro-competent *E. coli* cells (Section 2.8.4.1) and electroporation of *E. coli* (Section 2.8.4.2) was based on the method by Dower *et al.* (1988). Electroporation was the preferred method for transformation of *E. coli*, as the procedure could be performed both rapidly and efficiently. However, the $CaCl_2$ transformation method (Sections 2.8.4.3 and 2.8.4.2) was advantageous to use when difficulties were experienced with arcing (due to the DNA sample having a high ionic strength) during electroporation. Dialysis could be used to overcome this problem, but it was found to be an unreliable method. In addition, large volumes of ligation reactions could be easily transformed by the $CaCl_2$ method to avoid ethanol precipitation of dilute DNA samples (particularly small PCR fragments, ranging from 190-680 bp), and possible complete loss of DNA.

2.8.4.1 Preparation of Electroporation Competent *E. coli* Cells

One litre of LB broth was inoculated with 10 ml of an overnight culture of the desired *E. coli* strain (XL-1 or DH1), and grown at 37°C with vigorous shaking to mid-log phase (OD_{600} 0.5-1.0, about 3 hours). The cells were chilled on ice for 20 minutes then harvested by centrifugation for 10 minutes at 5000 rpm in a GSA rotor. The cells were washed sequentially (by resuspension, centrifugation at 5000 rpm to pellet the cells, then draining the supernatant) in 1.0 litre of ice cold sterile water, 500 ml of water, 20 ml of ice cold 10% glycerol, then finally resuspended in 4 ml of 10% glycerol. The cells were stored at -70°C in 200 µl aliquots.

2.8.4.2 Electroporation of *E. coli*

2-5 µl of DNA (ligation mixture or controls) was added to 50-100 µl of electro-competent *E. coli* cells in an ice cold eppendorf and gently mixed. The DNA/cell mixture was immediately transferred to an ice cold 0.2 cm cuvette, tapped to the bottom and electroporated in a Biorad gene pulser set at 25 µF, 2.5 kV and 200 Ω. A time constant of 4-5 msec was indicative of successful electroporation and cells were immediately resuspended in 1 ml of SOC medium (Section 2.2.4) and incubated for 1 hour at 37°C. A positive (circular plasmid DNA) and negative (water only) control were always included. Cells were spread on LB agar plates supplemented with ampicillin (100 µg/ml) and incubated overnight at 37°C.

2.8.4.3 Preparation of $CaCl_2$-Competent Cells

E. coli cells were made competent by $CaCl_2$ treatment as described by Cohen *et al.* (1972). LB (25 ml) was inoculated 1:50 with an overnight culture of *E. coli* and grown at 37°C on a shaking platform at 225 rpm to mid log phase (A_{550}=0.5, usually 2-4 hours). The cells were then pelleted by centrifugation (in a SS34 rotor) for 10 minutes at 5000 rpm at 4°C, resuspended in 10 ml of 60mM $CaCl_2$, and incubated on ice for 20 minutes, pelleted again and resuspended in 250 µl of 60 mM $CaCl_2$.

2.8.4.4 $CaCl_2$ Transformation of *E. coli*

DNA (diluted with water to a volume of 50 µl) was added to 250 µl of $CaCl_2$ competent *E. coli* cells and incubated on ice for 30 minutes. The transformation mix was heat shocked for 2.5 minutes at 42°C, then 900 µl of LB broth (prewarmed to 37°C) was added. The cells were shaken at 250 rpm for 1 hour at 37°C, then spread on LB plates supplemented with ampicilin (100 µg/ml) and incubated overnight at 37°C.

2.9 COLONY HYBRIDISATION USING THE DIG DETECTION SYSTEM

This method was based on the Boehringer manual: The DIG System Users Guide for Filter Hybridisation (1994).

2.9.1 Colony Lifts

Colonies to be screened were picked and patched onto duplicate nylon filters (Biotechnology Systems, 82 mm disks), placed in plates containing selective LB media, and incubated at 37°C for approximately 5 hours (or until the colonies were 1-2 mm in diameter). Each filter was marked asymmetrically with a needle to record its orientation on the plate. Filters were lifted off, and placed colony side up on 3 layers of blotting paper, soaked in 1.5 M NaCl, 0.5 M NaOH for 2 minutes, then transferred to blotting paper soaked in 1.5 M NaCl, 0.5 M Tris-HCl (pH 7.2), 1 mM EDTA for 5 minutes, and finally transferred to 2 x SSC for 2 minutes. Filters were air dried on blotting paper, then the DNA was fixed by UV-crosslinking for 2 minutes (DNA side up) on a transilluminator. Filters were washed in 3 x SSC to remove bacterial debris, dried and then stored at room temperature until required.

2.9.2 Colony Hybridisation

Filters were placed in a small plastic container and prehybridised in 25 ml of standard DIG buffer (Section 2.3.14) for 2 hours. After prehybridisation, buffer was discarded, and 7 ml of fresh standard DIG buffer was added. Denatured DIG-labelled probe (Section 2.9.3) was added to a final concentration of 5-25 ng/ml and hybridisation in a rotating oven at 65°C overnight was carried out. Filters were washed with 2 x wash solution (2 x SSC containing 0.1% SDS) at room temperature (2 x 5 minutes), then washed with 0.1 x wash solution (0.1 x SSC containing 0.1% SDS) at 68°C (2 x 15 minutes). After washing, the DIG detection procedure (Section 2.9.4) was performed to develop the hybridisation signal.

2.9.3 Preparation of DIG-Labelled Probe

DNA was labelled with Digoxigenin-11-dUTP (DIG) using the random primer method (Boehringer manual: The DIG Systems Users Guide for Filter Hybridisation, 1994). Approximately 10 ng of DNA was denaturated for 5 minutes in a boiling water bath, and immediately placed on ice to cool. The following reagents were added to the DNA template: 4 µl of high prime reaction mix (Boehringer, vial 1), 4 µl of DIG DNA labelling mix (containing 1 mM dATP, dCTP, dGTP, 0.65 mM dTTP, and 0.35 mM DIG-dUTP), and Milli-Q water to a total volume of 20 µl. The reaction was incubated at 37°C for 20 hours and stopped by the addition of 2 µl of 0.2 M EDTA. The labelled DNA was than precipitated by adding 0.1 volumes of 4 M LiCl, 3 volumes of 100% ethanol and incubating at -20°C overnight. The DNA was pelleted by centrifugation at 13000 g for 15 minutes, washed with 70% ethanol and then dried under vacuum. The probe was resuspended in 50 µl of TE buffer (10/1) and stored at -20°C.

To estimate the concentration of the DIG-labelled probe and to check the efficiency of the labelling reaction, 10-fold serial dilutions (in TE buffer) of labelled control DNA (Boehringer) and DIG-labelled probe were performed. 1 µl of each of these dilutions were spotted on to a piece of membrane (Hybond-N, Amersham) marked so each dilution could be identified. The DNA was fixed onto the membrane by UV crosslinking for 1 minute. After the detection procedure was carried out (Section 2.9.4), the concentration of the labelled probe was estimated by comparing the spot intensities of the control DNA (of known concentration) with the DIG-labelled probe DNA dilutions.

2.9.4 Detection of Hybridised DIG-labelled Probe

The filters that had been hybridised to the DIG-labelled probe (Sections 2.9.2 and 2.9.3) were washed in DIG buffer 1 (Section 2.3.15) for 5 minutes, than in DIG buffer 2 (Section 2.3.16) for 30 minutes at room temperature with gentle shaking. After this time buffer 2 was discarded and replaced with the anti-DIG-alkaline phosphatase conjugate (antibody, Boehringer) diluted 10^{-4} in DIG buffer 2. After incubation for 30 minutes at room temperature the antibody solution was poured off, and the filter placed in a fresh, clean container where it was washed at room temperature with DIG buffer 1 for 2 x 15 minutes. The filter was incubated for 2 minutes in DIG buffer 3 (Section 2.3.17) at room temperature then placed on a sheet of acetate. 100 µl drops of CSPD (chemiluminescent alkaline substrate) diluted 10^{-2} in DIG buffer 3, (using a total volume of 0.5 ml/100 cm^2 of filter) was spotted onto the filter. Another sheet of acetate was lowered onto the filter and the CSPD was spread evenly over the filter by gently wiping the surface of the acetate. The CSPD was left in contact with the filter for 5 minutes at room temperature and then incubated with the edges of the acetate sheets sealed for 15 minutes at 37°C. The acetate sealed filter was placed in a X-ray cassette with intensifying screens, either Fuji Medical or Kodak Scientific Imaging X-ray film, and exposed at room temperature for 20-30 minutes. After exposure the film was developed using standard procedures.

2.10 DNA SEQUENCING

Initial DNA sequencing work was performed using the Sequenase Version 2.0 Kit from USB. The majority of sequencing used the Amplicycle Kit from Perkin Elmer. Both methods were based on the Sanger method of chain termination (Sanger *et al.,* 1977). The Amplicycle system also utilises the Polymerase Chain Reaction (PCR), enabling small amounts of template DNA to be labelled quickly and efficiently, hence this kit was ideal for directly sequencing small quantities of PCR products.

2.10.1 Sequenase Version 2.0 Sequencing Protocol

To denature the template, 5-8 µg of template DNA was incubated with, 0.2 M NaOH and 0.2 mM EDTA for 30 minutes at 37°C. 0.1 volumes of 3M sodium acetate was added to neutralise the DNA and mixed well. DNA was precipitated by adding 2-4 volumes of 95% ethanol and was pelleted by centrifugation for 10 minutes. The DNA pellet was washed in 2 volumes of 70% ethanol, dried under vacuum and resuspended in 7 µl of water, 2 µl of Sequenase reaction buffer and 1 µl of the appropriate sequencing primer (0.5-1 pmol/µl). Annealing of the primer to the template DNA, was accomplished by warming the reaction to 65°C for 2 minutes, and slowly cooling to 30°C (over 30

minutes), then placing on ice. During the annealing reaction, 2.5 µl of each of the G,A,T,C termination mixes was aliquoted out per sequencing reaction to a labelled microtitre tray (Nunc). Upon completion of the annealing reaction, 1.0 µl of 0.1 M DTT, 2.0 µl of dGTP labelling mix (diluted 5x in Milli-Q water), 0.5 µl of [a-^{35}S]dATP, and 2.0 µl of Sequenase (diluted 8x in enzyme dilution buffer) were added to the annealed template and mixed. These labelling reactions were incubated for 2-5 minutes at room temperature. After incubation, 3.5 µl of the labelling reaction was added to each of the four termination mixes in the microtitre tray. A further incubation at 37°C for 5 minutes was carried out and stopped by the addition of 4 µl of stop solution. Sequencing reactions could then be stored at -20°C.

2.10.2 Amplicycle Sequencing Kit Protocol

For each DNA sequencing sample, 2 µl of each of the G,A,T,C termination mixes were aliquoted into four labelled 0.2 ml strip tubes (for use with the Corbett FTS-960 thermal cycler). A cocktail was prepared for each DNA sequencing template as follows: 1.0 µl of 20 µM primer, 1.0 µl of [a-^{33}P]dATP or dCTP, 4.0 µl of 10x cycling mix, template DNA (10 fmol of PCR product, or approximately 1-2 µg of Rapid Boil plasmid DNA [Section 2.6.1]), and Milli-Q water to 30 ml. 6 ml of this cocktail was then dispensed to each of the four termination tubes previously prepared. The tubes were then placed in a Corbett FTS-960 thermal cycler preheated to 95°C. The following cycling conditions were used:

Initial Step	95°C	1 minute	
Denaturation	95°C	30 seconds	x 25
Annealing	65°C	30 seconds	
Elongation	72°C	1 minute	
Elongation	72°C	1 minute	
Cool	4°C		

Within 45 minutes of PCR completion, 4 µl of stop solution was added to the tubes, and the reactions stored for up to 1 week at -20°C.

2.11 RNA PROCEDURES

2.11.1 Precautions Taken for Working With RNA

Glassware was acid washed in a chromic acid bath and baked for at least 2 hours at 180°C. Previously unopened bags of disposable tips and eppendorfs were autoclaved in

RNA treated glassware. Fresh disposable gloves were always worn when handling both RNA equipment and during RNA experiments. RNA solutions were only prepared from unopened stocks of reagents. A set of gilson pipettes dedicated to RNA work were always used.

2.11.2 Agarose-gel Electrophoresis of RNA

Electrophoresis equipment dedicated to RNA work was used to avoid contamination with RNases. The gel box, gel combs and glass plates were presoaked overnight in 1% SDS and wiped clean with ethanol. A surface tension 1% agarose TAE gel was poured onto the glass plate. Approximately 7 μg of RNA (prepared as in Section 2.13.2) was loaded in 1x RNA loading dye. The gel was electrophoresised at 80 V at room temperature until the dye had migrated 3/4 of the gel length. The gel was stained in ethidium bromide for 10 minutes and the RNA observed under UV light and photographed.

2.11.3 Determination of Fungal RNA Concentration and Purity

The RNA sample was appropriately diluted and the absorbance measured by a Shimazdu spectrophotometer at a wavelength of 260 nm. The RNA concentration was calculated on the assumption that an absorbance of 1.0 at 260 nm is equivalent to 40 μg/ml. RNA purity is indicated by the 260 nm/280 nm ratio, where pure RNA has an OD_{260}/OD_{280} ratio of 2.0 (Current Protocols in Molecular Biology, 1989; Glasel, 1995).

2.12 RNA PREPARATION

2.12.1 *Aspergillus nidulans* Spore Suspension Preparation

MYG (Section 2.2.3) plates were inoculated with *A. nidulans* spores and incubated at 37°C for 2-3 days or until the culture was sporulating well. 5 ml of Tween 80/Saline solution (5 ml of tween 80, 1.6 g NaCl, Milli-Q water to 200 ml), was pipetted onto the surface of the MYG plate, and rubbed over the surface with a glass spreader to release spores. This spore suspension was pipeted onto another *A. nidulans* culture and the procedure repeated. The final spore suspension was passed through sterile cheese cloth to remove debris and pelleted by centrifugation at 3 K (Heraeus Sepatech Megafuge 1.0) for 5 minutes. The spore pellet was then resuspended in 2 ml of Saline (0.8% NaCl) and aliquoted into eppendorfs. The spore concentration was estimated by counting, using a haemocytometer.

2.12.2 Isolation of RNA from *Aspergillus nidulans*

Total RNA was prepared for RT-PCR (Section 2.13) from strain A18 (Table 2.1) based on the method of Shirzadegan *et al.* (1991). Due to degradation of RNA and time constraints, total RNA (A18) was obtained from laboratory stocks prepared by R. Bradshaw. Approximately 400 mg of dry weight mycelium was ground to a fine powder with liquid nitrogen in an ice cold mortar and pestle. The powder was added to 4 ml of 80°C phenol extraction buffer (containing 0.1 M Tris HCl pH 8.0, 0.01 M EDTA, 1% SDS, in an equal volume of Tris-equilibrated phenol), immediately mixed by vortexing and incubated for 5 minutes at 80°C. An equal volume of chloroform was added, mixed by vortexing and centrifuged at 11 K for 15 min at 4°C. The top aqueous phase was then extracted with phenol/chloroform and centrifuged at 11 K for 15 minutes (centrifugations were performed in a SS34 rotor throughout). The aqueous phase was extracted with an equal volume of chloroform and centrifuged again as above. To precipitate the RNA, 1/2 volume of 8 M LiCl was added to the aqueous phase, and left at -20°C overnight. The RNA was pelleted at 12 K for 30 min at 4°C and then washed with 1 volume of 2 M LiCl, 0.05 M EDTA and centrifuged at 12 K for 30 min. Finally, the pellet was washed in 3 M NaOAc (pH 5.5) and resuspended in 200 µl of TE (10/0.1).

2.13 RT-PCR ANALYSIS OF RNA

2.13.1 Reverse Transcriptase Synthesis of cDNA

Before the RNA was transcribed into cDNA, it was DNaseI treated to remove any traces of DNA which may be preferentially amplified during the PCR step. The method for DNaseI treatment was based on that of Collett (1995). To 10 µg of RNA (Section 2.12.2), 1.0 µl of 100mM dithiothreitol (DTT) in 200 µl of DNaseI buffer (100 mM sodium acetate, 5 mM magnesium sulphate, pH 5), 50 U of RNase inhibitor, and 20 U of DNaseI (free of RNase) were added and incubated for 30 minutes at 37°C. Another 20 U of DNaseI was added and incubated for 30 minutes at 37°C. The RNA was then Phenol/chloroform extracted and ethanol precipitated (as in Section 2.5.7, except 3 M NH_4Ac was used instead of 3 M NaOAc). RNA was resuspended in 20 µl of TE (10/0.1).

The DNaseI treated RNA and untreated RNA was reverse transcribed into cDNA using the method based on that by Kawasaki (1990). 1 µg of RNA was denaturated at 90°C for 5 minutes, then immediately put on ice for 5 minutes before it was added to the PCR cocktail. The cocktail contained the following: 2 µl of 10 x PCR buffer (500 mM KCl,

200 mM Tris-HCl [pH 8.4 at 23°C], 25 mM $MgCl_2$ and 1 mg/ml bovine serum albumin), 1 mM of each dNTP, 100 pmol of random primers (BRL), 40 units of RNase inhibitor, 200 units of Superscript reverse transcriptase (BRL) and Milli-Q water to 20 µl. Upon addition of denatured RNA to the cocktail, the reaction was mixed and incubated for 10 minutes at room temperature, at 42°C for 60 minutes, and then denatured at 95°C for 5 minutes. The reaction was then cooled on ice for 5 minutes and stored at -20°C.

2.13.2 Amplification of cDNA by PCR

10 µl of the cDNA reaction (prepared in Section 2.13.1) was added to the following: 4 µl of 10 x PCR buffer, 2 µM 'upstream' primer, 2 µM 'downstream' primer, 2.5 units of Taq DNA Polymerase, and Milli-Q water to 50 µl. The cDNA reaction tubes were then placed in a thermal cycler preheated to 95°C. The following cycle conditions were used to amplify the cDNA:

Initial Step	95°C	2 minutes	
Denaturation	95°C	30 seconds	x35
Annealing	60°C	45 seconds	
Elongation	72°C	60 seconds	
Elongation	72°C	7 minutes	
Store	4°C		

2.14 SOUTHERN BLOTTING AND HYBRIDISATION

2.14.1 Southern Blotting

This method was based on that of Southern (1975). DNA to be blotted and hybridised was electrophoresed through a 1% agarose gel, stained and photographed as described in Section 2.6.2. The wells were cut off the gel and the gel dimensions measured. The gel was gently agitated in a tray containing depurination solution (250 mM HCl) for 15 minutes, then gently shaken in denaturating solution (500 mM NaOH, 500 mM NaCl) for 30 minutes and finally gently shaken in neutralisation solution (500 mM Tris (pH 7.4), 2 M NaCl) for 30 minutes. The gel was washed in 2 x SSC for 2 minutes.

While the gel was being treated, the blotting apparatus was assembled. Two sheets of Whatman 3 MM chromatography paper soaked in 20 x SSC (Section 2.3.12), were placed over a plastic trough with large wells at each end, so that the ends of the paper

dipped into these wells. The wells were then filled with 20 x SSC. The blotting trough was overlaid with gladwrap, and a grid was marked on the gladwrap of dimensions 2 mm less than the gel size. The grid was then cut out and the treated gel was placed, with well side down over the grid, so that the edges of the gel are on top of the gladwrap. A piece of nylon membrane (Hybond-N, Amersham; or Sartolan Blotting membranes) was cut out 2 mm less than the gel size, soaked in 2 x SSC, and placed over the gel. Any bubbles between the gel and filter were carefully removed. Two pieces of Whatman 3 MM were cut to 2 mm less than the gel, soaked in 2 x SSC and placed on top of the nylon filter. Another 2 sheets of dry Whatman 3MM of the same dimensions as above were placed on top of the wet 3 MM sheets. A pile of paper towels (5-8 cm deep) were then placed on top of the dry 3 MM sheets, followed by a 250 ml glass bottle filled with liquid as a weight. After blotting overnight, the blotting apparatus was disassembled and the nylon filter washed in 2 x SSC for 5 minutes. The filter was left to air dry between blotting paper and then baked under vacuum at 80°C for 2 hours.

2.14.2 Preparation of [a-^{32}P]dCTP-Labelled DNA Probe

DNA was labelled using the Ready-To-Go DNA Labelling Kit (Pharmacia). 25-50 ng of DNA (in a volume up to 45 µl), was denatured in a boiling water bath for 3 minutes, then immediately put on ice for 2 minutes and centrifuged briefly. The DNA solution was then added to the tube of reagent mix and mixed by pippeting the solution up and down. 3-5 µl of [a-^{32}P]dCTP (3000 Ci/mmol) and water to a final volume of 50 µl was added. The reaction was incubated for 30 minutes at 37°C.

Unincorporated nucleotides were removed from the labelled probe by using the ProbeQuant G-50 Micro Columns (Pharmacia Biotech) as described by the manufacturers.

2.14.3 Southern Blot Hybridisation

The Southern blot (Section 2.14.1) to be probed was prehybridised in a sealed glass hybridisation tube for a minimum of 2 hours at 65°C in hybridisation solution (Section 2.3.9). After prehybridisation, the hybridisation solution was discarded, and replaced with 7 ml of fresh hybridisation solution. The [a-^{32}P]dCTP labelled probe (Section 2.14.2) was denatured by boiling for 10 minutes immediately before it was added to the hybridisation tube. Hybridisation was carried out in a rotating oven at 65°C for a minimum of 6 hours, but usually overnight.

After hybridisation, the hybridisation fluid was discarded and about 100 ml of wash solution (3 x SSC, 0.2% SDS) was added. The filter was washed 3 times for 20 minutes at 65°C, and then wrapped (still wet) in gladwrap. The filter was placed in a X-ray cassette with intensifying screens against either Fuji Medical or Kodak Scientific Imaging X-ray film, and exposed for an appropriate period of time at -70°C. After exposure the film was developed (using standard procedures) in the dark room by placing in developing solution for 3 minutes, then in fixing solution for 3 minutes.

2.14.4 Stripping Probes from Southern Blots

Filters to be stripped were gently shaken in a container containing a solution of boiling 0.1% SDS. The solution was allowed to cool to room temperature, the procedure repeated and the filter autoradiographed (Section 2.14.3), to check that the hybridised probe had been removed. If probe was still present, this process was repeated.

2.15 LIBRARY SCREENING BY PLAQUE HYBRIDISATION

This method was based on that of Sambrook *et al.* (1989). Three rounds of library screening were performed to select the final positive plaques. Each round of screening involved plating the recombinant phage (Section 2.15.1), taking plaque lifts (Section 2.15.2), hybridisation of the filters to labelled probe (Section 2.15.3), and the selection of positive plaques (Section 2.15.3). λ phage DNA was then prepared from the positive plaques as detailed in Section 2.5.6.

2.15.1 Plating Phage λ

The phage population to be screened was diluted in SM buffer (Section 2.3.8) at 3 x 10^4 pfu/ml for the primary screen, or 3x 10^2 to 3x 10^3 pfu/ml for the secondary and tertiary screens. 100 µl of diluted phage was then added to 100 µl of an overnight bacterial culture of LE 392 grown in LB (Section 2.2.1) supplemented with 0.2% maltose and 10 mM $MgSO_4.7H_2O$. The phage/bacteria mix was incubated for 30 minutes at 37°C, added to 3 ml of TB top agarose (Section 2.2.6) at 45-50°C, then poured onto fresh LB agar plates. The top agarose was allowed to set, then the plates were incubated at 37°C until small plaques formed (about 7 hours). Plates were stored at 4°C for a minimum of 1 hour or overnight before plaque lifts were taken.

2.15.2 Plaque Lifts

Nylon filters (Biotechnology Systems, 82 mm disks) were marked asymmetrically with a needle for orientation purposes. Duplicate lifts were taken from each plate for both the primary and secondary rounds of library screening. The filter was placed on the LE 392 phage lawn and left for 1 minute for the first lift and 2 minutes for the duplicate lift, once the filter was completely wet. The filters were then placed DNA side up on 3 layers of filter paper soaked in 500 mM NaOH, 500 mM NaCl for 2 minutes, then 500 mM Tris-HCl (pH 7.4), 2 M NaCl for 5 minutes, and finally in 2x SSC for 1-2 minutes. The filters were air dried and the DNA fixed onto the filters by microwaving on high for 1 minute.

2.15.3 Hybridisation of λ Phage DNA to [a-^{32}P]dCTP-Labelled Probe

Library filters were hybridised to [a-^{32}P]dCTP-labelled DNA (Section 2.14.2) and autoradiographed as described in section 2.14.3, except plastic pots (10 cm diameter) were used (10 filters/pot), and the hybridisation and washing steps were performed in a shaking water bath.

Positive plaques were identified by a hybridisation signal on the autoradiogram at identical sites on duplicate filters. The corresponding plaques were located on the plate by alignment of the orientation markings. A plug of agar containing the positive plaques was removed by using a 1 ml Gilson pipette tip from which the end 5-10 mm had been cut off. The plugs were then put into an eppendorf containing 1 ml of SM buffer and 1 drop of chloroform and the phage left to elute into the buffer at 4°C overnight.

CHAPTER 3
CONSTRUCTION OF PROMOTER-REPORTER VECTORS AND DISCOVERY OF A SEQUENCING ERROR

3.1 INTRODUCTION

To analyse the promoter region of the *Aspergillus nidulans* cytochrome *c* (*cycA*) gene, in particular the role of the putative HAP1 binding site (Raitt, 1992), a promoter reporter vector was constructed containing the *cycA* promoter fused to the *lacZ* gene of *Escherichia coli*. The construction of a second vector which was to contain the *cycA* promoter with the putative HAP1 site deleted (Δ HAP1) fused to the *lac*Z gene, was also attempted.

3.2 CLONING STRATEGY

3.2.1 The *lac*Z Reporter Vector

The 11.5 kb reporter vector used in this study, pAN923-42B$_{BglII}$ (R117, Table 2.1), constructed by Van Gorcom *et al.* (1986), and modified to create a mutant *argB* gene by Punt *et al.* (1990), (Figure 3.1), allows in-phase translational fusion to the *E. coli lacZ* gene. pAN923 contains a unique *Bam*HI site in one of three possible translational reading frames (41B, 42B, 43B) in front of the *lacZ* gene. By insertion of the *cycA* 5' fragments into the *Bam*HI site at 42B, a translational fusion was obtained with the 0.68 kb fragment containing 578 bp of promoter and upstream region, along with 54 bp of coding sequence (hereafter called the 'promoter' region).

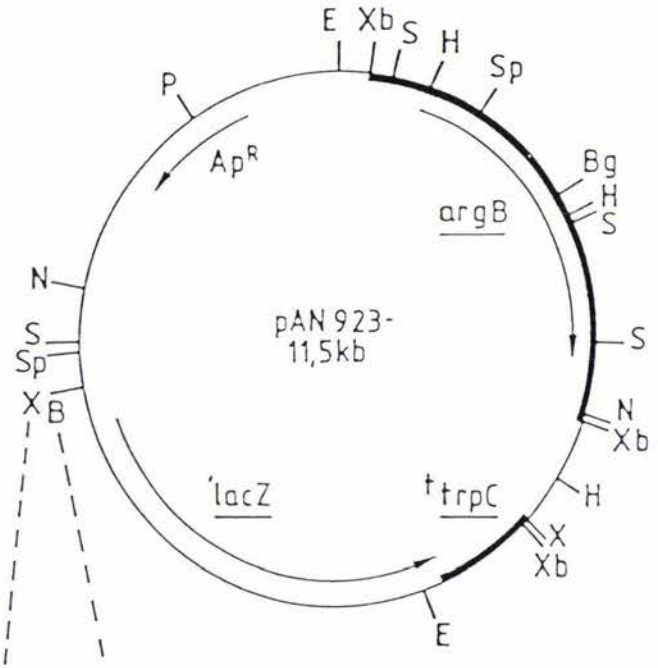
The reporter vector which contains a mutant *argB* gene, can be transformed into an *A. nidulans argB2* mutant strain and integrated at the *argB* gene locus. Complementation of the *argB* mutations to yield an *argB*$^+$ phenotype only occurs if homologous integration occurs, thus allowing selection for this event. After confirmation by Southern blotting of single copy integration, ß-galactosidase assays and Northern blots can be performed to determine *in vivo* promoter activity, and to measure transcriptional and translational *lacZ* reporter expression.

3.2.2 PCR Amplification of the *cycA* Promoter Fragments

PCR (Section 2.7) was used to amplify the *cycA* promoter fragments using plasmid R41 DNA (Table 2.1) which contains a 2.3 kb *cycA* genomic fragment (Raitt *et al.*, 1994). Primers for PCR were designed to incorporate a restriction enzyme site at each

Figure 3.1 The *lacZ* reporter vector pAN923

The *cycA* fragments containing the promoter and 5' coding region were inserted into the *Bam*HI (B) site of pAN923-42B. Thick lines represent *A. nidulans* DNA; thin lines represent *E. coli* DNA. B, *Bam*HI; Bg, *Bgl*II; E, *Eco*RI; H, *Hind*III; N, *Nru*I; P, *Pst*I; S, *Sal*I; Sp, *Sph*I; Ss, *Sst*I; X, *Xho*I; Xb, *Xba*I.



	X	B	S	P	H	'lacZ
41B	CT.CGA.GGA.TCC.GTC.GAC.CTG.CAG.CCA.AGC.TTG.GCA.CTG.GCC.GTC.					
42B					X / CTC.GAG.GAT.CCC.GTC.	B
43B	X / C.TCG.AGG.ATC.CCC.GGG.AAT.TCA.CTG.GCC.GTC.	B			E	

end of the PCR product (Figure 3.2). Primers RB1 and RB2 (Table 2.2) include *Bam*HI recognition sequences at the 5' end and were used to amplify the entire *cycA* promoter region (published by Raitt *et al.* (1994)), along with the 5' coding region to give a total of 0.68 kb. Both the vector and the 0.68 kb fragment were then digested with *Bam*HI, followed by ligation to allow the *cycA* fragment to be fused in frame at the *Bam*HI site of the *lac*Z gene. To obtain a translational fusion, the promoter fragment must be orientated so that the protein-coding region of the *cycA* gene is fused next to the protein-coding region of the *lac*Z gene. Diagnostic restriction digests were necessary to distinguish the orientation of the ligated promoter fragments (see Table 3.2, Figure 3.3).

To amplify the Δ HAP1 *cycA* promoter fragment for the second construct, primers RB1 and RB4 were used to amplify the region upstream of the Δ HAP1 site, producing a 190 bp fragment. Primers RB3 and RB2 were used to amplify the region downstream of the Δ HAP1 site, producing a 400 bp fragment. The internal primers, RB3 and RB4 both include a *Sal* I restriction site. After *Sal* I digestion both the 400 bp and 190 bp fragments can be ligated together to form the 0.59 kb Δ HAP1 promoter fragment which will contain a 54 bp deletion of the putative HAP1 binding region. The 0.59 kb fragment contains 536 bp of promoter region, along with 54 bp of coding region. Digestion of the Δ HAP1 promoter with *Bam*HI, allows this promoter to be cloned, as with the 0.68 kb promoter, into pAN923-42B$_{BglII}$.

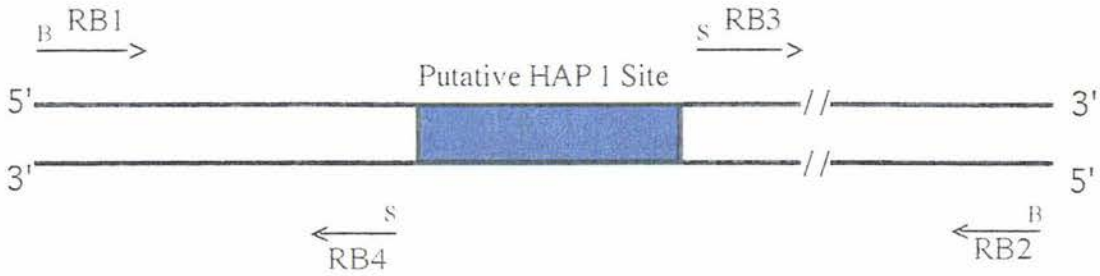
PCR conditions for the amplification of the *cycA* promoter fragments were established by Saunders, 1993. However, R41 plasmid (Table 2.1) DNA concentration was optimised to produce a high yield and purity of PCR product; a concentration of 0.1 ng/µl was found to be optimal as a template for amplification of the promoter fragments.

3.2.3 General Preparation of pAN923-42B Vector and 0.68 kb *cycA* Promoter

Plasmid pAN923-42B (R109, Van Gorcom *et al.*, 1986) was initially used as the cloning vector, before the modified pAN923-42B$_{BglII}$ (R117) vector (Punt *et al.* 1990) was available. After plasmid preparation, pAN923 vectors were digested to completion with *Bam*HI and CAP treated as described in section 2.8.2. This was necessary because pAN923 provides no effective selection system for recombinants in *E. coli* and CAP treatment prevents self-ligation of the vector, thus abolishing high vector background on selective LB ampicillin plates.

Figure 3.2 PCR primer binding sites within the *cycA* promoter fragments

The relative positions of the primer sites used for amplification of the *cycA* promoter fragments are shown. B and S represent *Bam*HI and *Sal*I restriction enzyme sites respectively. The blue coloured box identifies the putative HAP1 binding region.


B RB1
S RB3
Putative HAP 1 Site
5'
3'
3'
5'
S
RB4
B
RB2

Approximate Scale
0
200
300
700 bp

The 0.68 kb and Δ HAP1 promoter fragments obtained by PCR (Section 2.7), were gel purified (Section 2.8.1) followed by phenol/chloroform and ethanol precipitation purification (Section 2.5.7) to remove any residual Taq polymerase terminal transferase activity which can add adenosine bases to the digested 3' ends of PCR products (Clark, 1988; Bennet and Molenaar, 1994). After purification, the promoter fragments were digested with *Bam*HI and/or *Sal*I as appropriate (Section 2.6.1), resulting in the removal of a small number of base pairs from each end. Since the change of length would be too slight to detect on an agarose gel, a control digest (to check the enzymes) was performed on plasmid DNA to give known fragment sizes. Following restriction enzyme digestion (Section 2.6.1) the promoter fragments were purified (Section 2.5.7) to remove the enzyme activity.

3.3 LIGATIONS AND TRANSFORMATIONS OF THE 0.68 kb *cycA* PROMOTER

3.3.1 Ligation of 0.68 kb promoter directly into pAN923-42B

The efficiencies of vector digestion, CAP-treatment and ligation were checked by comparing yields of transformants from electroporation of the following samples; self-ligated *Bam*HI-digested and CAP treated pAN923 vector (Cut, Cap, Lig), self-ligated *Bam*HI digested pAN923 vector (Cut, Lig), *Bam*HI-digested and CAP treated pAN923 vector (Cut, Cap) and uncut pAN923 vector. The results of these controls are included in Tables 3.1, 3.3 and 3.6.

The controls for transformation experiments A, B, and C, involved in the ligation of the 0.68 kb promoter into pAN923-42B (Table 3.1), suggested that the vector was poorly CAP treated but efficently *Bam*HI digested (shown by the high numbers of transformants per μg of Cut, Cap, Lig vector DNA and low numbers for Cut, Cap vector respectively). However, the comparison of control samples to determine capping efficiency was inaccurate due to the low numbers of transformants produced per plate (2-10 colonies). The transformation of control samples with a higher DNA concentration would have been more accurate. Electroporation and ligation efficiencies were satisfactory (shown by the numbers of transformants for uncut vector and Cut, Lig vector respectively).

The first attempt at ligating the 0.68 kb promoter fragment into pAN923-42B and transforming into *E. coli* XL-1 cells was unsuccessful (Transformation A, Table 3.1). After transformation of the ligation reactions into *E. coli* XL-1 cells, 7 transformants were obtained (Table 3.1, (A)) and checked by *Bam*HI digestion (Section 2.6.1) of plasmid DNA prepared by the rapid boiling method (Section 2.5.1). If ligation was

Table 3.1 Transformations A, B & C (with pAN923-42B vector)

DNA transformed into *E. coli* XL-1 cells[a]	DNA/plate (ng)	Number of colonies per Plate	Number of transformants per µg DNA (mean/experiment)
A			
0.68 kb insert and pAN923 vector ligation (1:1)	1.2 1.5	0 0	0
0.68 kb insert and pAN923 vector ligation (2:1)	1.5 1.5 3.0 3.0	2 1 4 0	1.1×10^3
B			
0.68 kb insert and pAN923 vector ligation (1:1)[b]	1.6 1.6	2 2	1.3×10^3
0.68 kb insert and pAN923 vector ligation (2:1)	1.9 1.9 1.6 1.6	3 2 1 0	1.0×10^3
0.68 kb insert and pAN923 vector ligation (2:1)[c]	1.5 1.5	3 3	2.0×10^3
0.68 kb insert and pAN923 vector ligation (3:1)	2.0 2.0 1.7 1.7	0 1 3 1	8.0×10^2
C			
0.68 kb insert and pAN923 vector ligation (3:1)[d]	6.2	53	8.5×10^3
Controls			
Cut, Cap pAN923 vector	1.5 8.5	0 0	0
Cut, Cap, Lig pAN923 vector	1.4	2	1.4×10^3
Cut, Lig pAN923 vector	2.5	10	4.0×10^3
Uncut pAN923 vector	0.6	235	4.0×10^5

[a] numbers in brackets indicate insert:vector ratio used

[b] indicates re-electroporation of (1:1) ligation from section A

[c] indicates re-electroporation of (2:1) ligation from section A

[d] indicates re-electroporation of (3:1) ligation from section B

successful, the 11.5 kb vector band, and the 0.68 kb promoter band should both be present (Table 3.2). For all 7 transformants, only the 11.5 kb vector band was present, indicating that ligation of promoter into pAN923 had not occurred. The ligation reactions were repeated and both old and new ligations were electroporated into *E. coli* XL-1 cells (Transformation B, Table 3.1). In total, 21 transformants were obtained, and PCR (Section 2.7) was used to screen the transformants by using primer pair RB1 and RB2 (Table 2.2) to detect the 0.68 kb promoter. Two transformants contained a PCR product of approximately 0.68 kb and were further checked by *Bam*HI digestion of rapid boil plasmid DNA (Section 2.6.1) to confirm the presence of ligated promoter. Promoter orientation and copy number were determined by *Sph*I digestion. Restriction digest patterns showed that one transformant (R120, Table 2.1) contained the 0.68 kb promoter ligated into pAN923 in the wrong orientation (Table 3.2, Figure 3.3). The other transformant produced a 7.2 kb fragment with *Bam*HI digestion and 6.1 kb plus 1.3 kb fragments with *Sph*I digestion. This restriction profile was not expected for either vector only or vector with insert, and must therefore indicate either a contaminant or a deletion/re-arrangement event.

The remaining ligation mixture which produced the R120 plasmid from transformation B (Table 3.1), was electroporated into *E. coli* XL-1 cells (Section 2.8.4.2), resulting in 53 transformants (Transformation C, Table 3.1). These transformants were checked by PCR (Section 2.7) as described above, and all were shown to contain the promoter fragment. *Bam*HI digestions of 25 transformants showed that 23 contained insert, 1 contained vector only, and 1 contained no plasmid DNA. The 23 insert-containing transformants were checked by *Sph*I digestion to check promoter orientation. All transformants contained insert ligated into the vector in the wrong orientation which was surprising, since 50% of inserts would be expected to be ligated into the vector in the other orientation. Perhaps there was some unknown preference for the insert ligating in the wrong orientation. Another possibility was that all transformants had arisen from one transformation event so that all transformants were of clonal origin. Alternatively, these transformants may have arisen from contamination by plasmid R120, but this is unlikely since transformation control experiments were negative. In summary, time and resources seemed wasted by further checking of the remaining transformants.

Further pAN923/0.68 kb promoter ligations were performed with freshly prepared vector (Table 3.3). However, the vector was not cut to completion by *Bam*HI, (as shown by the number of transformants with cut, capped, unligated vector) and the cut ends were poorly capped (as shown by the increased number of transformants with capped and ligated vector). This resulted in many transformants to screen for the presence of insert (Transformation D, Table 3.3). Despite the large number it was thought worthwhile to

Table 3.2 Expected Restriction Digestion Fragment Profiles for Ligation of 0.68 kb Promoter into pAN923-42B

Orientation of promoter	Restriction Digest	Fragment Sizes (kb)
Correct orientation	*Bam*HI	11.5, 0.68
	*Sph*I	7.16, 4.15, 0.87
Reverse orientation	*Bam*HI	11.5, 0.68
	*Sph*I	7.68, 4.15, 0.37

Figure 3.3 Restriction enzyme analysis of plasmids pAN923, R122 and R120

Restriction digests of plasmids pAN923, R120 (from transformation B, Table 3.1), and R122 (from transformation G, Table 3.6) were carried out individually with *Bam*HI and *Sph*I. Lane 2: pAN923, *Bam*HI. Lane 3: pAN923, *Sph*I. Lane 4: R122, *Bam*HI. Lane 5: R122, *Sph*I. Lane 6: R120, *Bam*HI. Lane 7: R120, *Sph*I. Lane 1 contains a 1 kb ladder (Gibco BRL, Life Technologies). The size of the bands indicated are in kb.

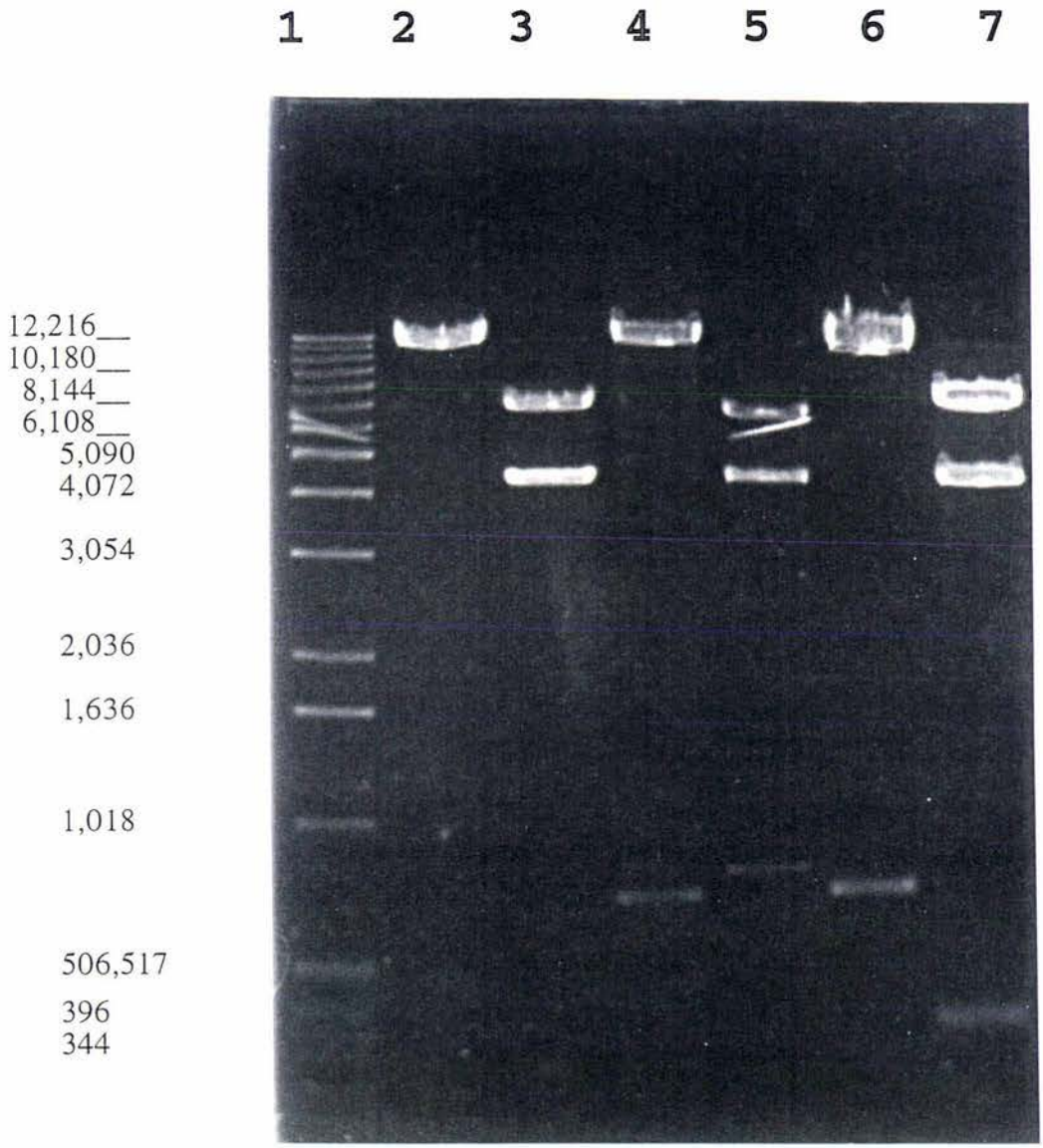

1
2
3
4
5
6
7
12,216
10,180
8,144
6,108
5,090
4,072
3,054
2,036
1,636
1,018
506,517
396
344

Table 3.3 Transformation D (with pAN923-42B vector)

DNA transformed into *E. coli* XL-1 cells	DNA/plate (ng)	Number of colonies per Plate	Number of transformants per µg DNA (mean/experiment)
D			
0.68 kb insert and	5.5	9	1.2×10^4
pAN923 vector	6.9	113	
ligation	9.6	179	
Controls			
Cut, Cap pAN923 vector	50	91	1.8×10^3
Cut, Cap, Lig pAN923 vector	5.0	800	1.6×10^5
Uncut pAN923 vector	0.5	360	7.2×10^5

screen for recombinants by colony hybridisation (Section 2.9.2). Unfortunately, there was a lot of background present on the filters affecting the detection of positive colonies. However ten positive colonies were selected for further investigation, but upon restriction enzyme analysis of rapid boil plasmid DNA (Section 2.5.1) were found to contain no insert.

3.3.2 Ligation of 0.68 kb Promoter into T-Tailed pUC18 Vector

Difficulties experienced with cloning the 0.68 kb promoter directly into pAN923 vector (Section 3.3.1) were due to a number of factors. One major problem was probably inefficient cutting of *Bam*HI sites on a large proportion of the PCR product ends, thus decreasing the ligation efficiency. The T-tailed vector cloning approach (Marchuk *et al.*, 1991) avoids this problem by making use of the adenosine base overhang generally added to PCR products by Taq polymerase, allowing ligation to a 3' T-Tail attached to the pUC18 vector (kindly provided by R. Johnson). Once the promoter is cloned into the T-Tailed vector, it can be digested using flanking sites in the polylinker and gel purified (Section 2.8.1) before ligation into pAN923.

Transformants obtained from the T-tail ligations produced unexpected restriction digestion profiles. Eighteen white transformants were obtained from three separate ligation experiments and were checked by PCR (Section 2.7) using primers RB1 and RB2 (Table 2.2) to detect insert. Eight transformants contained a 0.68 kb PCR product and were further checked by digestion individually with *Bam*HI and *Sph*I. All eight transformants produced a 5.4 kb fragment with *Bam*HI and a 4.1 kb and 1.3 kb fragment with *Sph*I (see Table 3.4 for expected restriction enzyme profiles). After considering all possible combinations of insert and vector, no possibilities matched the data observed for these transformants. The total construct size (5.4 kb) suggested that the transformants contained pUC18 (2.7 kb) ligated to four 0.68 kb promoters, or two ligated pUC18 plasmids, but the *Sph*I digestion results did not support this hypothesis. The alternative explanation was that DNA contamination had occurred.

Lack of success with the T-tailing cloning approach may be due to a number of reasons. The ligation efficiency may not have been optimal, causing a decrease in the number of white transformants obtained. Also, the terminal transferase activity of Taq polymerase is only weak (Parry, 1993) and thus may add low numbers of adenosine bases to the PCR fragments.

Table 3.4 Expected Restriction Enzyme Profiles for Ligation of 0.68 kb Promoter into Plasmid pUC18

Restriction Digest	Expected Fragment Sizes (kb)
*Bam*HI	2.7, 0.68
*Sph*I	2.8, 0.60
	OR*
	3.3, 0.13

* depending on orientation

3.3.3 Ligation of 0.68 kb Promoter into Commercially CAP-treated pUC18

Another approach to facilitate the cloning of the PCR-amplified 0.68 kb promoter into pAN923, was to digest with *Bam*HI and subclone it into commercially CAP treated pUC18 plasmid (BRL/Life Technologies). Successful ligation of insert in either orientation should show that both *Bam*HI ends of the promoter were intact and functional. Cloning into pUC18 was expected to be easier than cloning directly into pAN923 since it has the advantage of blue/white selection. Another advantage of ligating the promoter into pUC18, a multiple copy number plasmid, was that large quantities of digested promoter could be obtained, gel purified (Section 2.8.1), then ligated into pAN923 to increase the chances of success.

The pUC18 ligations with the PCR derived 0.68 kb promoter, produced either no transformants or they were not stable transformants (Transformation F, Table 3.5). However, ligations with pUC18 and the 0.68 kb promoter derived from plasmid R120 (in which the promoter had previously been cloned in reverse orientation, Table 2.1) produced 12 transformants (Transformation E, Table 3.5). Restriction enzyme analysis of the 12 transformants from transformation E showed that four contained an insert of the expected size for the promoter, two of which were ligated in each of the different orientations. One representative of the pUC18 clones was stored in glycerol (R121, Table 2.1) and used for all further sources of 0.68 kb promoter. Since the only difference between transformants E and F was the source of the 0.68 kb promoter, it was suggested that the problem with obtaining recombinants with transformation F (using the PCR amplified promoter), was due to inefficient digestion of the promoter ends. This may have been because the enzyme recognition site (incorporated by the PCR primers) was too close to the ends of the PCR product (2 bp) for efficient cutting.

3.3.4 Ligation of 0.68 kb promoter into pAN923-42B

A *Bam*HI digestion of plasmid R121, followed by gel purification, was carried out to obtain the 0.68 kb promoter fragment for ligation into pAN923-42B$_{BglII}$. Now that large quantities of insert with *Bam*HI ends were easily obtainable for ligation, larger scale ligation reactions comprising 50 ng of insert in a 3:1 or 1:1 insert to vector ratio were set up to increase ligation frequency. Due to large volume sizes (30-40 µl), these samples were transformed by the calcium chloride method (Section 2.8.4.3). Table 3.6 displays the transformation results.

Table 3.5 Transformations E & F (with CAP-treated pUC18 vector)

DNA transformed into *E. coli* XL-1 cells[a]	DNA/plate (ng)	Number of colonies per Plate	Number of transformants per μg DNA
E			
0.68 kb insert (from R120) and pUC18 vector ligation (1:1)	5.3	12	2.3×10^3
F			
0.68 kb insert (PCR) and pUC18 vector ligation (2:1)	7.5	0	0
0.68 kb insert (PCR) and pUC18 vector ligation (3:1)	8.8	0	0

[a] numbers in brackets indicate insert:vector ratio used

Table 3.6 Transformation G (with pAN923-42B vector)

DNA transformed into *E. coli* XL-1 cells[a]	DNA/plate (ng)	Number of colonies per Plate	Number of transformants per μg DNA
G			
0.68 kb insert and pAN923 vector ligation (3:1)	242	1280	5.3×10^3
0.68 kb insert and pAN923 vector ligation (1:1)	655	704	1.1×10^3
Controls			
Cut, Cap pAN923 Vector	20	48	2.4×10^3
Cut, Cap, Lig pAN923 vector	10	37	3.7×10^3
Uncut pAN923 Vector	0.5	360	7.2×10^5

[a] numbers in brackets indicate insert:vector ratio used

The CAP treatment controls for transformation G (Table 3.6) indicate that the vector has not been cut or CAP treated to completion. In addition, the number of transformants from the promoter ligation experiments are approximately equivalent to the number of transformants present in the cut, cap, lig control. Therefore the majority of transformants were expected to be circular pAN923 vector. However, two transformants from each of the pAN923 vector:insert ligations were checked to confirm the nature of the transformants. One of these transformants contained, in addition to the vector band, the 0.68 kb promoter when digested with *Bam*HI. An *Sph*I digest of this transformant demonstrated that the 0.68 kb promoter was ligated into pAN923 in the correct orientation with the following band sizes: 7.1 kb, 4.15 kb, 0.80 kb (Table 3.2, Figure 3.3). This construct was named R122 (Table 2.1).

3.4 LIGATIONS AND TRANSFORMATIONS OF THE *cycA* Δ HAP1 PROMOTER

All ligations and transformations of the Δ HAP1 promoter were unsuccessful (results not shown). As with the 0.68 kb promoter, ligations of the Δ HAP1 promoter were attempted with both pAN923 and commercially CAP treated pUC18 plasmids. The lack of success with the Δ HAP1 promoter ligations was probably due to the increasingly more difficult nature of three way ligations and hence a lower ligation frequency. Similar problems with the enzyme sites at the ends of the PCR products, (as discussed with the 0.68 kb promoter) were also expected to be the cause of unsuccessful ligation with the Δ HAP1 promoter fragments. Due to unexpected results arising in the sequence of the whole promoter (Section 3.5), cloning of the *cycA* Δ HAP1 was not pursued further.

3.5 SEQUENCING OF THE 0.68 kb PROMOTER IN PLASMID R122

Sequencing across the junction in the new R122 vector construct, containing the 0.68 kb promoter in the pAN923 reporter vector, was essential to determine the exact position of the *cycA* gene ATG start codon relative to the *Bam*HI insertion site in the MCS. This information was required to confirm that a functional (in-frame) translational fusion was obtained. In addition, the promoter was amplified by PCR using Taq polymerase, which has a mis-incorporation error rate of 8.9×10^{-5} errors/bp (Cariello *et al.*, 1991). Therefore, sequencing of the promoter was required to check there were no mutations present in the promoter sequence which could affect promoter expression studies, or in the coding sequence to put the *cycA* coding region out of frame.

pUC/M13 universal forward and RB1 primers (Table 2.2) were used to obtain double stranded sequence using the Sequenase Version 2.0 (Section 2.10.1) and Amplicycle

(Section 2.10.2) sequencing protocols. The preferred method of sequencing was the Amplicycle method, since it consistently produced high quality sequence using a quick and simple protocol. Plasmid DNA was prepared by small scale alkaline lysis preparation (Section 2.5.2) and purified by Wizard PCR preps DNA purification system (Section 2.8.1) for use with the Sequenase kit. Rapid boil DNA (Section 2.5.1) was used for sequencing with the Amplicycle kit.

Upon comparing the sequence obtained from plasmid R122 with the published sequence (Raitt *et al.* 1994), a discrepancy was noticed, with six thymidine bases obtained rather than the expected four (Figure 3.4). Unfortunately, the region of this discrepancy occured within the published coding region of the *cycA* gene, thus disrupting the reading frame. Either Taq polymerase had slipped during the amplification of the promoter, incorporating two extra thymidine bases, or the published sequence was incorrect. To determine which of these alternatives was correct, sequence was obtained from plasmid R41, which contains a sub-clone from a genomic library (Table 2.1), using primer RB2 (Table 2.2). Sequence from R41 over the region with the discrepancy clearly confirmed that six thymidine bases were present (Figure 3.5). Therefore the published sequence of the *cycA* gene by Raitt *et al.* (1994) was incorrect over this region. Figure 3.6 shows the published nucleotide sequence of the *cycA* gene as published by Raitt *et al.* (1994).

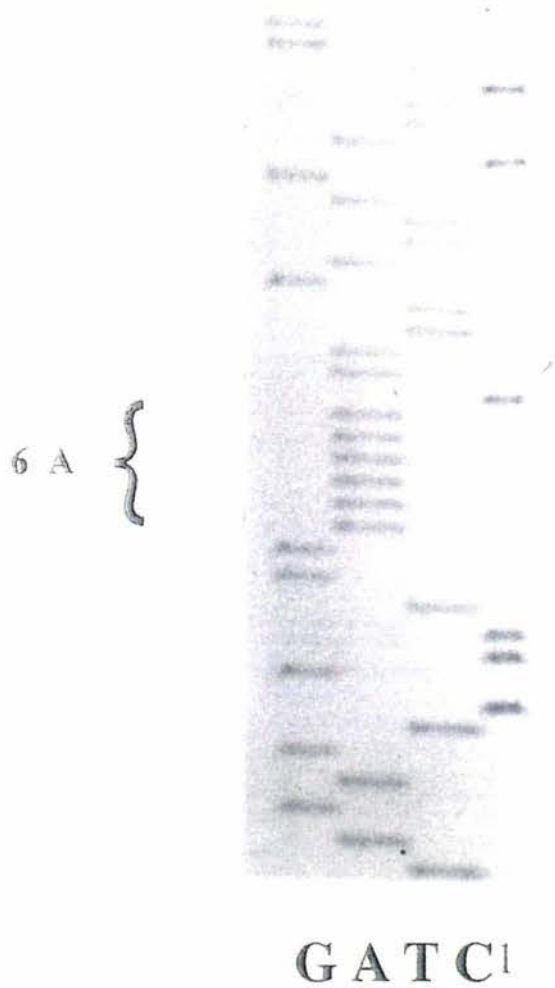
Figure 3.4 Comparison of the *cycA* nucleotide sequence from plasmid R122 with the published *cycA* sequence (Raitt, 1992)

The region spanning the sequencing discrepancy has been presented. The top strand represents sequence obtained from plasmid R122 (see Figure 4.5), whereas the bottom strand represents the published nucleotide sequence of the *cycA* gene (Figure 3.6, Raitt *et al.*, 1994). The boxed area in red highlights the region in which the sequencing error occured, and the published ATG start codon is shown in green.

Figure 3.5 Autoradiograph of sequence obtained from the *cycA* gene cloned into plasmid R41

Nucleotide sequence from the non-coding strand of plasmid R41 is presented. Six adenosine bases were observed in the A lane (lane 2) instead of the expected four.

```
R122     403 ATGCTAATCAATTGTTTTTTCCAGGCGACT 432
             :::::::::::::::::::  ::::::::::
cycA     +1  ATGCTAATCAATTGTTTT  CCAGGCGACT 28
```



[1] The sequence presented in this autoradiograph was obtained from the non-coding strand of plasmid R41 (Table 2.1), therefore this sequence represents the reverse and complement of the coding strand from plasmid R122 (Table 2.1) in Figure 3.4.

Figure 3.6 The published nucleotide sequence of the *cycA* gene (Raitt *et al.*, 1994)

The nucleotide sequence of the *cycA* gene as published by Raitt *et al.* (1994) is displayed. Nucleotides are numbered from the A of the published methionine translational start residue (+1). The alternative translational start site proposed in this study is indicated by the pink ATG. The major transcriptional start site (published by Raitt *et al.*, 1994) for the 2 intron model is indicated below the red asterisk. The two intervening regions (Introns I, II) are displayed in blue with the corresponding splice consensus sequences underlined. Consensus splice sequences for the proposed additional intron are underlined and coloured blue. The predicted amino acid sequence is shown below the coding strand using the standard one-letter code. The conserved CKQCH domain and the Gly 8 residue are indicated in green in the amino acid sequence. The position of the observed sequencing error is indicated in red. The putative HAP1 site proposed by Raitt (1992) is marked by green highlighting.

```
-649  TACTCTGAGCACAGTGCTGAATCATCCTACCTCTGATTGGTCAATTCCCAGATCACGGGT
-589  GTCGTCGGGGGCGCGACCAAGAAACCAGCTCTACAAATTCCCTCCAAGTTTTTTCTTCC
-529  CTTTGGCCAGTCCGCTTGACTTGAATTCTGTCTTTCATCTTCTCTGCTACATACAACTCT
-469  TTGTACTATACCACTTACCTCTTTACATAACCCTTTCTCTCTACCTCTTTATTTTTATC
-409  ACTCACAATGGCTAAGGGCGGTGACAGCTACTCTCCTGGTAAGTAGTTTGAATTCATCTC
-349  TTCGTTTCGGCTAGGCGCTCTCTGCTGGGGGTGATATTCACTCCACTGCGTGCATTGTCG
-289  AAGTGAACGATATGTGAGACACAGGCTGTGAATGATGTGTTGGGTCTGGTGAGAACATTG
-229  TCCGATCCAACACAAGCTCAAAGTTGCCCACTTCTGGATCGCCATTTGATCGCCAGCACA
-169  ATACAATTTCTACTTCTATCGCGCGCTGGCAATCCTCACTTTGCAGCGGATGCTTTATTC
-109  TTCATCGTCGTCGCCACAATGACGAGCTTCGACGCTTACTGCTTCACGACCACTCTCAGC
-49   ATGCGCGAAGCCGAACTGGAAGAGCTGGAGAAAGAGAAAGCGGACCAGAATGCTAATCAA
                                                      M  L  I  N
11    TTGTTTTCCAGGCGACTCTACCAAGGGTGCTAAGCTCTTCGAGACCCGTTGCAAGCAGTG
       C  F  P  G  D  S  T  K  G  A  K  L  F  E  T  R  C  K  Q  C
71    CCACACTGTCGAGAACGGCGGCGGCCACAAGGTCGGCCCCAACCTCCACGGTCTCTTCGG
       H  T  V  E  N  G  G  G  H  K  V  G  P  N  L  H  G  L  F  G
131   CCGTAAGACTGGTCAGGCTGGAGGCTACGCCTACACCGATGCCAACAAGCAGGCCGACGT
       R  K  T  G  Q  A  G  G  Y  A  Y  T  D  A  N  K  Q  A  D  V
191   CACCTGGGACGAGAACTCTCTGGTACAATCCCATGACAGCTCTAACAGCTCTGGGCCGAT
       T  W  D  E  N  S  L
251   TGCTAACTTTCTTTCACCAACAGTTCAAGTACCTCGAGAACCCCAAGAAGTACATCCCTG
                              F  K  Y  L  E  N  P  K  K  Y  I  P  G
311   GTACCAAGATGGCTTTCGGTGGTCTCAAGAAGACCAAGGAGAGGAACGATCTCATCACGT
        T  K  M  A  F  G  G  L  K  K  T  K  E  R  N  D  L  I  T
371   ATGTAACGCTGCTTACCACGGATAGGGCACATAGGCTAACAGGATGCACAGCTACCTCAA
                                                         Y  L  K
431   GGAGAGCACTGCTTAAATCGTTCGCGATTAGACGAGATAAACCGCCCCCCTGGGATTAGA
       E  S  T  A  end
491   CGAGGGCCTCTGGCTAGGTGACAGGCGGGTACTGTAACATTACACCTAGACCTGGTTTTG
551   AAGGTCGTCGGGACATGGAGGATATTATAGATCTTGTTTCCTTCGCCATCCTTGTCTATA
611   TCTTATTCTTTCCTTTACCTTGACGAGTGTTTTTTCAGCTTTGTGGTACC
```

CHAPTER 3
DISCUSSION

3.6 PROBLEMS WITH CLONING INTO pAN923-42B VECTOR

The difficulties experienced with cloning the *cycA* promoter fragments into pAN923-42B were numerous and are summarised as follows. Firstly, the cloning strategy required ligation to occur through *Bam*HI sticky ends. Since the *Bam*HI recognition site was incorporated into the PCR product (positioned 2 nucleotides from the terminus) by the PCR primers, inefficient digestion of the promoter fragments near the ends of the PCR molecules was probably the cause of reduced ligation efficiency. For example, sub-cloning of the 0.68 kb promoter into pUC18 (Section 3.3.3) was successful with promoter derived from the *Bam*HI digestion of plasmid R120 (Transformation E, Table 3.5), but unsuccessful (under identical conditions) when using PCR-amplified and *Bam*HI digested 0.68 kb promoter (Transformation F, Table 3.5). Thus, the enzyme was probably unable to efficiently digest the PCR product. Secondly, the pAN923 vector contains no effective selection system for recombinants in *E. coli*. The vector therefore required CAP treatment to abolish high vector background, which was usually inefficient. Attempts were made to increase the capping efficiency by using a fresh batch of CAP enzyme, altering CAP enzyme concentration, incubation times, and concentrations of proteinase K. However, no overall improvement was observed from these modifications. Thirdly, the requirement for a translational fusion to investigate a potential translational control mechanism required clones to be screened for insert orientation, thus in retrospect a directional cloning approach would have been desirable.

3.7 IMPLICATIONS FROM THE DISCOVERY OF THE SEQUENCING ERROR IN THE PUBLISHED *cycA* GENE BY RAITT *et al* (1994)

Discovery of the presence of six thymidine bases in the *cycA* nucleotide sequence rather than the expected four as published by Raitt *et al* (1994) had many implications. The addition of two extra thymidine bases in the coding region of the *cycA* gene disrupts the published translational reading frame. This means that the long sought after translational fusion of the *cycA* promoter with the *lacZ* gene of pAN923-42B (R122, Table 2.1) was not obtained. Moreover, the discovery of this sequencing error brought into question the validity of the open reading frames of the cytochrome *c* gene as published by Raitt *et al* (1994). The published amino acid sequence downstream from the sequencing error would now code for different amino acids since the reading frame has altered. However, the published amino acid composition of the *cycA* gene shows extensive homology to both the yeast *CYC1* gene and the *Neurospora crassa* cytochrome *c* gene and contains all

the functional features associated with the cytochrome *c* protein. These features include a conserved heme binding domain CKQCH (position 21 to 25 in the amino acid sequence, Figure 3.6), and conserved glycine residues required for winding of the polypeptide around the heme molecule (Raitt, 1992). One of these invariant glycine residues, Gly 8 (Figure 3.6), falls immediately after the sequencing error, and is therefore no longer in frame with the published ATG codon with the sequencing correction (Figure 3.7). Since all the above features are found and conserved in the *cycA* protein as published by Raitt *et al.* (1994), it would appear that the published reading frame is correct. However, the published ATG start codon is no longer in the correct reading frame to produce the conserved cytochrome c protein, and therefore cannot be the translational start site.

Closer scrutiny of the published *cycA* nucleotide sequence revealed that both the six thymidine bases and the published ATG codon may fall within an intron region not previously detected. Filamentous fungal consensus splice sites for an additional intron were noticed at the 5' site (5'-GTANGT-3'), a possible branch point (5'-PyGCTAAC-3') and the 3' site (5'-CAG-3') at nucleotide positions -371, -77, and +19, respectively as seen in Figure 3.6 (Gurr *et al.* 1987). Evidence of this intron would further establish that the published translational start point is not correct.

Further examination of the *cycA* nucleotide sequence revealed a new potential translational start point, upstream of the putative intron, at position -402 nt in Figure 3.6 (highlighted in pink). This ATG codon would give the correct reading frame to produce the conserved *cycA* protein after the splicing of the putative additional intron. Moreover, re-calulation of the transcription start point (by moving the transcription site point in the 5' direction to take account of the additional intron) indicates the mRNA transcript commences at -230 nt. The suggested ATG codon would also become the first AUG in the mRNA, and translation of fungal genes usually occurs from the first AUG (Kozak, 1991; Gurr *et al* 1987). The sequence context of the AUG triplet is also important as it affects the fidelity and efficiency of translation initiation. For example, the scanning 40 S ribosome may occasionally initiate at the second AUG codon if the context of the first AUG is weak (Kozak, 1991). The Kozak sequence or consensus sequence for initiation of translation for filamentous fungi is CNNNCA(A/C)N<u>ATG</u>GC (Jon *et al.* 1993). The surrounding nucleotides of the proposed new ATG codon in this study form a strong Kozak sequence of CACTCACA<u>ATG</u>GC, whereas the context of the published ATG codon is weak with the Kozak sequence of GGACCAGA<u>ATG</u>CT.

If the proposed additional third intron exists, the putative HAP1 binding site would be situated after the putative ATG start codon in the coding region of the gene. Consequently, this HAP1 site cannot be an upstream regulatory element for analysis! In

Figure 3.7 How the sequencing error affects the cycA amino-terminal sequence

The twelve amino acids from the amino-terminal of the predicted translated sequence is shown using the standard one-letter code. The top line represents the amino acid sequence produced from the published *cycA* gene (Raitt *et al.*, 1994). The bottom line shows how the correction for the two extra Thymidine bases affect the predicted amino acid sequence if the translational start site proposed by Raitt *et al.* (1994) was maintained. Homologous amino acid residues are indicated by the colour red. The invariant amino acid residue Gly 8 conserved in all cytochrome c proteins is shown in blue.

Published amino-terminal sequence: +1 M L I N C F P G D S T K G A K +12
without sequencing correction (4T)

Predicted amino-terminal sequence : +1 M L I N C F F Q A T L P R V L +12
after sequencing correction (6T)

addition, the 0.68 kb *cycA* promoter cloned into the R122 plasmid contains only a small portion of promoter sequence, and therefore this construct would be of no further use. Thus for the remainder of my project I endeavoured to confirm my suspicions about the presence of a third intron and thus an altered translation initiation site, and to re-examine the 'new' *cycA* promoter region.

CHAPTER 4
RT-PCR ANALYSIS OF *cycA* RNA

4.1 INTRODUCTION

RT-PCR analysis was performed to determine if the *cycA* gene contains two or three introns and therefore the ATG translational start codon proposed from this study (see Section 3.7), located upstream of the putative intron (position -402nt, Figure 3.6). The precise location of the introns was confirmed by sequencing the *A. nidulans* cDNA products over the intron/exon boundaries.

The transcriptional start site for two introns was deduced from primer extension analysis by Raitt *et al.* (1994). If there are three introns in the *cycA* gene as proposed in this study, a revised transcriptional start site can be calculated. The transcription start site proposed by Raitt *et al.* (1994), was deduced by primer extension analysis and does not account for the length of the third putative intron, which has been spliced from the RNA. Therefore the position of the revised transcriptional start site would be moved upstream by the length of the additional intron (395 bp) from the transcription start site calulated for two introns (Figure 4.1). Confirmation of the transcriptional start site is presently being determined in this laboratory.

4.2 DEVELOPMENT OF A RT-PCR STRATEGY

mRNA is reverse transcribed into single stranded cDNA by the reverse transcriptase enzyme. Since introns are spliced from the mRNA, the cDNA strand synthesised will contain only exon information and 5' and 3' non-translated regions. To determine whether the *cycA* gene contains two or three introns, PCR primers were used which would produce a PCR product from the cDNA strand if there were three introns as in the theory outlined above, but no PCR product if there were two introns as shown in the published *cycA* sequence (Table 4.1, Figure 4.2). If three introns were present in the *cycA* gene, primer pair RB1/LG1 would amplify the entire cDNA of 596 bp and primer pair RB1/RB2 would produce a 298 bp product (Table 4.1). However, if two introns were present, no PCR product would be expected with either of these primer combinations because the RB1 primer site is located upstream of the major transcriptional start site (deduced from primer extension analysis (Raitt *et al.* (1994)). Therefore, the RB1 primer site would not be included in the mRNA.

A third primer pair, RB3/LG1 was used as a negative control for the amplification of products from the cDNA. No RT-PCR product should be amplified from the cDNA for

Figure 4.1 Re-estimation of the *cycA* transcription start point

How the transcriptional start site for a *cycA* gene with three introns was estimated is shown. Asterisks with 3 and 2 represent the major transcriptional start point for three or two introns respectively. The proposed additional intron (Intron I), is shown by a large box. The published two introns (Introns II and III) are represented by blue blocks. Thin black lines represent exons and 5' and 3' untranslated regions. The approximate position of the primer originally used for primer extension analysis (Raitt, 1992) is indicated and the length of the mRNA product produced from this primer is shown.

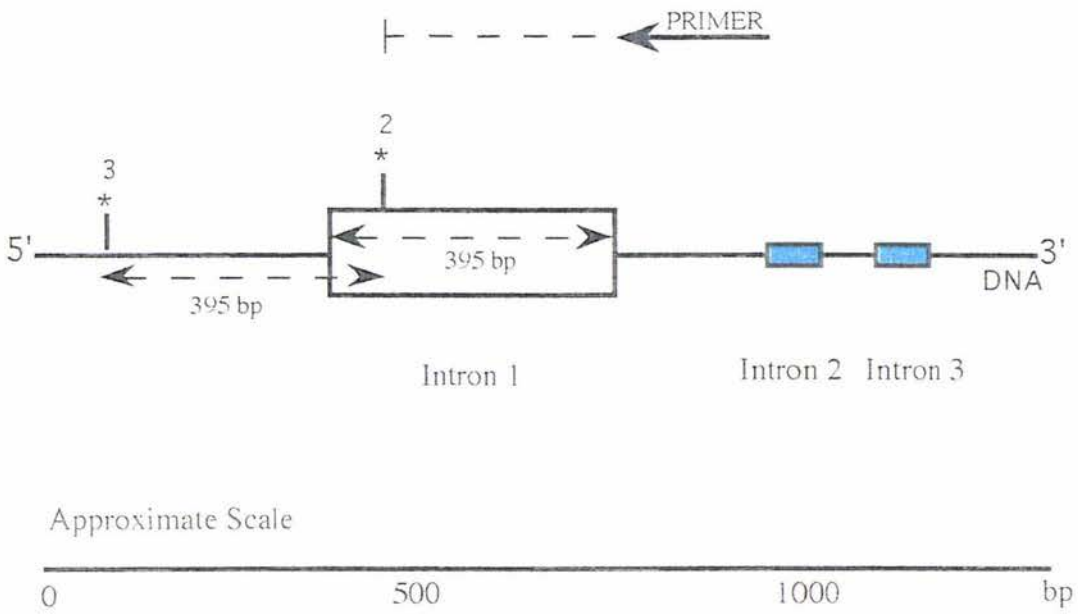

PRIMER
2
3
5'
3'
395 bp
395 bp
DNA
Intron 1
Intron 2
Intron 3
Approximate Scale
0
500
1000
bp

Figure 4.2 RT-PCR outline showing the relative positions of primers

The position of the primers (RB1, RB3, RB2, and LG1) used for RT-PCR are shown with respect to Introns I, II and III. The proposed third intron (Intron I) is shown by a red block, whereas the two published introns (Introns II and III) are indicated by blue blocks. Thin black lines represent exons and 5' and 3' untranslated regions. Asterisks represents the major transcriptional start point for 3 or 2 introns respectively.

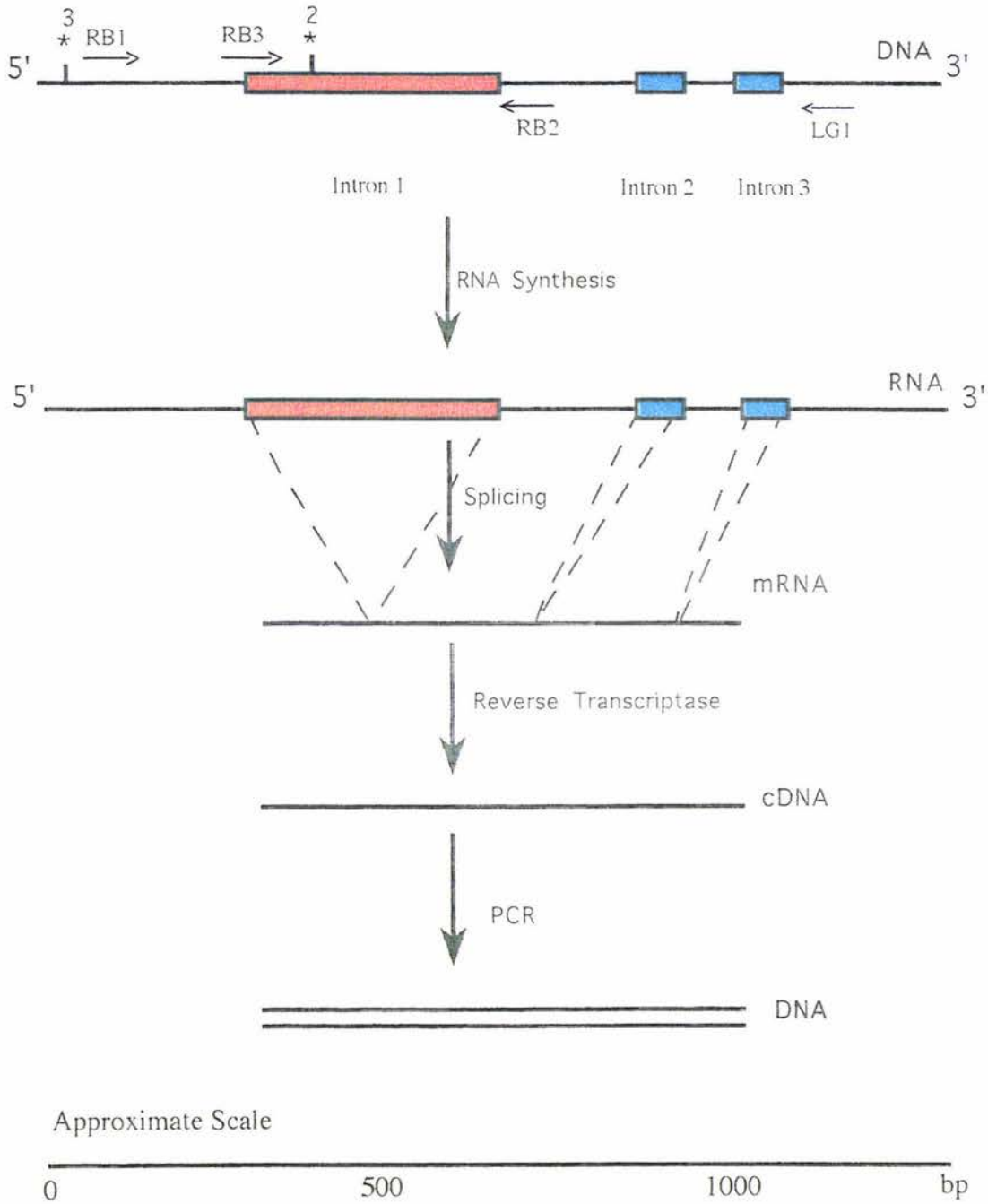

3
RB1
RB3
2
5'
DNA
3'
RB2
LG1
Intron 1
Intron 2
Intron 3
RNA Synthesis
5'
RNA
3'
Splicing
mRNA
Reverse Transcriptase
cDNA
PCR
DNA
Approximate Scale
0
500
1000
bp

Table 4.1 PCR Products Expected From RT-PCR Analysis

Primer Pairs	RT-PCR Products Expected From:		
	Splicing of 3 introns (bp)	Splicing of 2 introns (bp)	DNA template (bp)
RB1-LG1	596	0	1100
RB1-RB2	298	0	691
RB3-LG1	0	0	860

either two or three introns, since the RB3 primer site would fall within the proposed intron region for the three intron model, or would lie upstream of the transcriptional start point for the two intron model, and therefore would not be transcribed into cDNA.

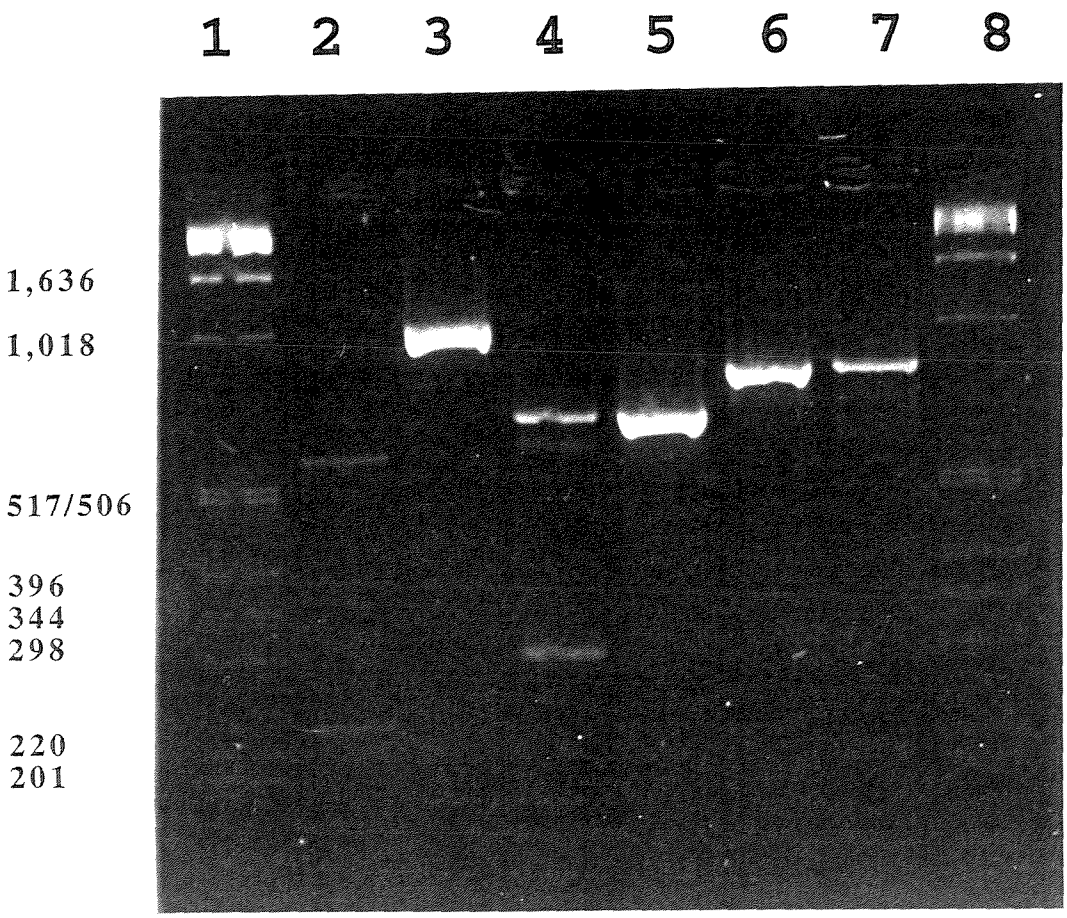
4.3 RT-PCR RESULTS

RT-PCR products are shown in the gel photograph in Figure 4.3. The 596 bp and the 298 bp RT-PCR products expected from the splicing of three introns were successfully amplified with the RB1/LG1 primer pair (Lane 2) and RB1/RB2 primer pair (Lane 4) respectively, indicating that the number of introns present in the *cycA* gene is three, rather than the published two (Raitt *et al.,* 1994). The 596 bp product contains the full length cDNA sequence of the published *cycA* gene, whereas the 298 bp product contains the upstream cDNA region of the *cycA* gene, including the region from which the third intron has been spliced. A non-specific product of approximately 240 bp was also amplified consistently with the RB1/LG1 primer pair (Lane 2). This amplification was probably as a result of the primers binding non-specifically to the cDNA, because the product was too small to represent amplification from either DNA or partially spliced cDNA. Although Raitt *et al.* (1994) found that *A. nidulans* contains only one cytochrome *c* gene, the 240 bp product may have been amplified from an undiscovered second *A. nidulans* cytochrome *c* gene. However, this was unlikely due to the result of recent work in this laboratory (Gardiner, unpublished) which confimed that *A. nidulans* contains only one cytochrome *c* gene.

The three sets of primer pairs (RB1/LG1, RB1/RB2, RB3/LG1) were also used with a DNA template as a positive control (from total RNA samples that were not DNaseI treated prior to RT-PCR). These samples contained enough residual DNA for PCR. Both the 596 bp and the 298 bp RT-PCR products were distinguishable from the DNA template amplified products of 1100 bp (Lane 3) and 691 bp (Lane 5) respectively. A product of 691 bp was also amplified from the DNaseI treated RT-PCR sample with primers RB1/RB2 (Lane 4), probably as a result of DNA amplification due to incomplete DNaseI treatment. In addition, with these primers a faint non-specific product of approximately 610 bp was also observed for both DNaseI and non-DNaseI treated templates (Lanes 4 and 5), and probably represents the primers binding non-specifically to the DNA. For primer pair RB3/LG1 (Lanes 6 and 7), no cDNA product was amplified. This was expected as the RB3 primer site is not transcribed into cDNA for PCR amplification. The expected DNA template derived product of 860 bp was, however, produced from PCR amplification with primer pair RB3/LG1 (Lane 6 and 7). A faint band of approximately 750 bp was also amplified with these primers for both the DNaseI and non-DNaseI treated templates (Lanes 6 and 7). This product may be produced from the non-specific

Figure 4.3 RT-PCR analysis of total RNA from *A. nidulans* strain A18

All reactions were amplified by PCR from reverse transcribed total RNA from A18. The templates for lanes 2, 4 and 6 were DNaseI treated before RT-PCR. The following PCR primer pairs were used. Lanes 2 and 3: RB1 and LG1. Lanes 4 and 5: RB1 and RB2. Lanes 6 and 7: RB3 and LG1. Lanes 1 and 8 contain the 1 kb ladder (Gibco BRL, Life Technologies), the size of the bands indicated are in bp.


1 2 3 4 5 6 7 8
1,636
1,018
517/506
396
344
298
220
201

amplification of DNA, or a product formed from the non-splicing of Intron I (to allow binding of the RB3 primer) with either the splicing of both Intron II and III (to give a 746 bp product) or the splicing of just Intron II (799 bp) or Intron III (807 bp). Another faint band of approximately 630 bp was also observed for the non-DNaseI treated sample with the RB3/LG1 primers, and was probably the result of non-specific amplification from the DNA template.

4.3.1 Sequencing of RT-PCR Products

The reverse transcribed and PCR amplified products of 596 bp and 298 bp were sequenced directly using the Perkin-Elmer Amplicycle Kit (Section 2.10.2) with primers RB1, RB2, and LG1 (Table 2.2). The PCR products were gel-purified (Section 2.8.1) before sequencing to remove additional PCR products present from the PCR reaction. To obtain sufficient quantities of DNA for sequencing, the cDNA fragments were further amplified by PCR using the PCR cycle conditions for the amplification of the *cycA* promoter region (Section 2.7). Initial sequence obtained from the cDNA fragments was generally of poor quality and often contained termination bands across all four lanes. Optimisation of the PCR reaction was performed to increase the purity of the PCR generated cDNA fragments for sequencing. Optimisation was performed by increasing the primer annealing temperature, decreasing the number of PCR cycles from 30 to 25, decreasing DNA concentration and decreasing the denaturation, annealing and elongation times (see Section 2.7 for optimised PCR cycle conditions). After PCR optimisation, increased quantities of DNA were obtained, and reduced DNA smearing (from the amplification of non-specific products) during agarose electrophoresis, indicated better quality DNA. Single stranded sequence of the 596 bp cDNA fragment was subsequently obtained successfully (Figure 4.4). Sequence from over the intron/exon boundaries (Figure 4.4) demonstated that three introns were spliced from the *cycA* gene, confirming that the *cycA* gene contains a total of three introns. The corrected nucleotide sequence for the *cycA* gene, including introns, is presented in Figure 4.5.

Figure 4.4 cDNA sequence of the 596bp RT-PCR product.

Nucleotides are numbered from the A of the proposed methionine translational start codon (+1). The predicted amino acid sequence is shown below the coding strand in blue using the standard one-letter code. The pink asterisks above the nucleotide sequence indicates where the introns (Intron I, II and III) were spliced.

```
     .         .         .         .         .         .
-67  TTGTACTATACCACTTACCTCTTTACATAACCCTTTCTCTCTACCTCTTTATTTTTTATC

     .         .         .         .         *.        .
-7   ACTCACAATGGCTAAGGGCGGTGACAGCTACTCTCCTGGCGACTCTACCAAGGGTGCTAA
            M  A  K  G  G  D  S  Y  S  P  G  D  S  T  K  G  A  K

     .         .         .         .         .         .
54   GCTCTTCGAGACCCGTTGCAAGCAGTGCCACACTGTCGAGAACGGCGGCGGCCACAAGGT
      L  F  E  T  R  C  K  Q  C  H  T  V  E  N  G  G  G  H  K  V

     .         .         .         .         .         .
114  CGGCCCCAACCTCCACGGTCTCTTCTGCCACACTGTCGAGAACGGCGGCGGCCACAAGGT
      G  P  N  L  H  G  L  F  C  H  T  V  E  N  G  G  G  H  K  V

     .         .         .         .         .         .
174  CGGCCCCAACCTCCACGGTCTCTTCGGCCGTAAGACTGGTCAGGCTGGAGGCTACGCCTA
      G  P  N  L  H  G  L  F  G  R  K  T  G  Q  A  G  G  Y  A  Y

     .         .         .         .         .         *.
234  CACCGATGCCAACAAGCAGGCCGACGTCACCTGGGACGAGAACTCTCTGTTCAAGTACCG
      T  D  A  N  K  Q  A  D  V  T  W  D  E  N  S  L  F  K  Y  L

     .         .         .         .         .         .
     TCGAAACCCCAAGAAGTACATCCCTGGTACCAAGATGGCTTTCGGTGGTCTCAAGAAGAC
      E  N  P  K  K  Y  I  P  G  T  K  M  A  F  G  G  L  K  K  T

     .         .         .   *     .         .         .
     CAAGGAGAGGAACGATCTCATCACCTACCTCAAGGAGAGCACTGCTTAAATCGTT
      K  E  R  N  D  L  I  T  Y  L  K  E  S  T  A  end
```

Figure 4.5 The corrected nucleotide sequence of the *cycA* gene.

The nucleotide sequence of the *cycA* gene published by Raitt *et al.* (1994), has been presented here in its corrected form. Nucleotides are numbered from the A of the proposed methionine translational start residue (+1). The published translational start site (Raitt *et al.*, 1994) is indicated by the colour pink. The major transcriptional start site is indicated by a red asterisk. Three intervening regions (Introns I, II and III) are displayed in blue with the corresponding splice consensus sequences underlined. The predicted amino acid sequence is shown below the coding strand using the standard one-letter code. The observed sequencing error is indicated in red. The putative HAP1 site proposed by Raitt (1992) is marked in green.

```
      .         .         *  .         .         .         .
-247  TACTCTGAGCACAGTGCTGAATCATCCTACCTCTGATTGGTCAATTCCCAGATCACGGGT
      .         .         .         .         .         .         .
-187  GTCGTCGGGGGCGCGACCAAGAAACCAGCTCTACAAATTCCCTCCAAGTTTTTTCTTCC
      .         .         .         .         .         .         .
-127  CTTTGGCCAGTCCGCTTGACTTGAATTCTGTCTTTCATCTTCTCTGCTACATACAACTCT
      .         .         .         .         .         .         .
-67   TTGTACTATACCACTTACCTCTTTACATAACCCTTTCTCTCTACCTCTTTATTTTTTATC
      .         .         .         .         .         .         .
-7    ACTCACAATGGCTAAGGGCGGTGACAGCTACTCTCCTGGTAAGTAGTTTGAATTCATCTC
             M  A  K  G  G  D  S  Y  S  P  G
      .         .         .         .         .         .         .
54    TTCGTTTCGGCTAGGCGCTCTCTGCTGGGGGTGATATTCACTCCACTGCGTGCATTGTCG
      .         .         .         .         .         .         .
114   AAGTGAACGATATGTGAGACACAGGCTGTGAATGATGTGTTGGGTCTGGTGAGAACATTG
      .         .         .         .         .         .         .
174   TCCGATCCAACACAAGCTCAAAGTTGCCCACTTCTGGATCGCCATTTGATCGCCAGCACA
      .         .         .         .         .         .         .
234   ATACAATTTCTACTTCTATCGCGCGCTGGCAATCCTCACTTTGCAGCGGATGCTTTATTC
      .         .         .         .         .         .         .
294   TTCATCGTCGTCGCCACAATGACGAGCTTCGACGCTTACTGCTTCACGACCACTCTCAGC
      .         .         .         .         .         .         .
354   ATGCGCGAAGCCGAACTGGAAGAGCTGGAGAAAGAGAAAGCGGACCAGAATGCTAATCAA
      .         .         .         .         .         .         .
414   TTGTTTTTTTCCAGGCGACTCTACCAAGGGTGCTAAGCTCTTCGAGACCCGTTGCAAGCAG
                       D  S  T  K  G  A  K  L  F  E  T  R  C  K  Q
      .         .         .         .         .         .         .
474   TGCCACACTGTCGAGAACGGCGGCGGCCACAAGGTCGGCCCCAACCTCCACGGTCTCTTC
      C  H  T  V  E  N  G  G  G  H  K  V  G  P  N  L  H  G  L  F
      .         .         .         .         .         .         .
534   GGCCGTAAGACTGGTCAGGCTGGAGGCTACGCCTACACCGATGCCAACAAGCAGGCCGAC
      G  R  K  T  G  Q  A  G  G  Y  A  Y  T  D  A  N  K  Q  A  D
      .         .         .         .         .         .         .
594   GTCACCTGGGACGAGAACTCTCTGGTACAATCCCATGACAGCTCTAACAGCTCTGGGCCG
      V  T  W  D  E  N  S  L
      .         .         .         .         .         .         .
654   ATTGCTAACTTTCTTTCACCAACAGTTCAAGTACCTCGAGAACCCCAAGAAGTACATCCC
                                F  K  Y  L  E  N  P  K  K  Y  I  P
      .         .         .         .         .         .         .
714   TGGTACCAAGATGGCTTTCGGTGGTCTCAAGAAGACCAAGGAGAGGAACGATCTCATCAC
       G  T  K  M  A  F  G  G  L  K  K  T  K  E  R  N  D  L  I  T
      .         .         .         .         .         .         .
774   GTATGTAACGCTGCTTACCACGGATAGGGCACATAGGCTAACAGGATGCACAGCTACCTC
                                                                Y  L
      .         .         .         .         .         .         .
834   AAGGAGAGCACTGCTTAAATCGTTCGCGATTAGACGAGATAAACCGCCCCCCTGGGATTA
      K  E  S  T  A  end

894   GACGAGGGCCTCTGGCTAGGTGACAGGCGGGTACTGTAACATTACACCTAGACCTGGTTT

954   TGAAGGTCGTCGGGACATGGAGGATATTATAGATCTTGTTTCCTTCGCCATCCTTGTCTA

1014  TATCTTATTCTTTCCTTTACCTTGACGAGTGTTTTTTCAGCTTTGTGGTACC
```

CHAPTER 4
DISCUSSION

4.4 THE *cycA* GENE CONTAINS A THIRD INTRON AND HAS A CORRECTED N-TERMINAL REGION

RT-PCR was performed to determine if the *cycA* gene contained two or three introns. Amplified products of 596 bp and 298 bp, expected from the splicing of three introns were produced using two different pairs of primers. Subsequent sequencing of both these products confirmed that the *cycA* gene contained three introns instead of two (Raitt *et al.,* 1994). The new intervening sequence (Intron I, 395 nt) is positioned from +32 nt to +426 nt in Figure 4.5. A consequence of the detection of Intron I was that the N-terminal region of the published ORF was affected. The published ATG start codon was positioned within the new intron region and thus could not be the correct translational start point. An ATG start codon was proposed and is located upstream of Intron I at nt position +1, Figure 4.5.

The new ORF (Figure 4.5) encodes a polypeptide of 112 amino acid residues (instead of 109 as published by Raitt *et al.,* 1994), and contains a new N-terminal region, with the first 6 amino acid residues of the published cycA protein replaced by a new set of 9 amino acid residues (Figure 4.6). The N-terminal region of the cycA protein (as proposed in this study from the new ATG) was aligned with those from other fungi (including the predicted *cycA* sequence published by Raitt *et al.* (1994)), the plant *Arabidopsis thaliana*, and animals (Figure 4.6). The cytochrome *c* N-terminal region appears to be highly conserved within closely related groups. For example, within the yeast group, cytochrome *c* from the three yeasts *Candida albicans*, *Schwanniomyces occidentalis* and *Kluyveromyces lactis*, have identical amino-terminal regions. Within the animal kingdom, the human, rat, chicken and mouse cytochrome *c* genes encode only one amino acid variation within this region. However, the cytochrome *c* genes of the yeast *S. cerevisiae* encode slightly different amino acids at the N-terminal when compared with the other yeasts. Although the new *A. nidulans cycA* N-terminal portion (this study) is not completely homologous with the filamentous fungus *Neurospora crassa,* it provides a better match than the published cycA protein (with only 4 from the 9 new amino acids not identical) to any other cytochrome *c* genes sequenced so far. It also has the greatest amino terminal conservation with *N. crassa,* and with the *CYC7* gene of *S. cerevisiae*. Interestingly, the new *cycA* ORF contains the same number of amino acid residues as the *CYC7* gene. Once the cytochrome *c* genes from other filamentous fungi are sequenced, a closer match may be obtained within this group.

Figure 4.6 Alignment of cytochrome *c* N-terminal protein sequences

The N-terminal cytochrome *c* protein sequences of several different organisms are aligned with respect to *A. nidulans* (as proposed in this study), using the standard one-letter code. Alignment at conserved glycine residues shows different lengths of N-terminal sequence. Amino acid residues identical to *A. nidulans* (this study) are indicated by green. Dashes represent gaps introduced into the sequence to allow for optimal alignment. The highly-conserved glycine residue is indicated in red.

Filamentous Fungi

A. nidulans (this study)	M A K G G D S Y S P G D S T K G A...
A. nidulans (Raitt *et al.* 1994)	- - - M L I N C F P G D S T K G A...
N. crassa	- - - - - M G F S A G D S K K G A...

Yeast

S. cerevisiae CYC7	M A K E S T G F K P G S A K K G A...
S. cerevisiae CYC1	- - - - M T E F K A G S A K K G A...
C. albicans	- - - M P A P Y E K G S E K K D A...
S. occidentalis	- - - M P A P Y K K G S E K K G A...
K. lactis	- - - M P A P Y K K G S E K K G A...
S. pombe	- - - - - M P Y A P G D E K K G A...

Plant

A. thaliana	- M Q V A D I S L Q G D A K K G A...

Animal

Human/Mouse/Rat/Chicken (Somatic)	- - - - - - - - - - M G D $^{V}_{I}$ E K G K...

4.5 ALL THREE INTRONS OF THE *cycA* GENE ARE LOCATED AT CONSERVED POSITIONS

In the cytochrome *c* genes of the higher eukaryotes, mouse (Limbach and Wu, 1985), chicken (Limbach and Wu, 1983), and rat (Scarpulla and Wu, 1981), the position of intervening sequences appear to be conserved, with all of the coding sequences interupted after Lysine 55. The corresponding position in the *cycA* gene is at Alanine residue 64, nine amino acids upstream from the *A. nidulans* Intron II region (Figure 4.5 and Figure 4.7). None of the yeast cytochrome *c* genes sequenced so far contain intervening sequences, although introns are present in other yeast mitochondrial respiratory-chain proteins, for example the cytochrome *b* protein of *S. cerevisiae* (Lazowska *et al.,* 1988). Figure 4.7 shows the cycA protein sequence aligned with the cytochrome *c* proteins of *N. crassa* and *A. thaliana*. The third intron shown to be present in the *cycA* gene by RT-PCR and subsequent sequencing (termed Intron I) is located in the coding sequence after the first nucleotide in the triplet coding for the conserved glycine 10 residue (Figure 4.7). This intron is also found in an identical position in the *N. crassa* cytochrome *c* gene (Bottorff *et al.,* 1994). Intron III from *A. nidulans* is also conserved, being at an identical position in the *N. crassa* cytochrome *c* gene (at Thymidine 105 in the cycA protein, Figure 4.7). The plant cytochrome *c* gene from *Arabidopsis thaliana* (Kemmerer *et al.,* 1991a) was found to cluster with *N. crassa* and other yeasts rather than with other higher plants. In accordance with this observation by Kemmerer *et al.* (1991a) Intron II, from *A. thaliana,* interrupts the coding sequence after Lysine 73, at the same position as for Intron II from *A. nidulans* (Figure 4.7). However, a recent paper by Hudson (1995) suggests that the *A. thaliana* cytochrome *c* sequence may originate from contaminating fungal material present in plant tissue, and thus is really a fungal cytochrome *c* sequence. The only other cloned and sequenced plant cytochrome *c* gene is from *Oryza sativa* (Rice) (Kemmerer *et al.,* 1991b), which contains two introns located in unique positions after Glycine 38 and Lysine 82 in the cycA protein sequence.

In summary, all three intron positions from the *cycA* gene are conserved, and are found at identical positions with either *N. crassa* or *A. thaliana* cytochrome *c* genes. The length of these conserved introns at identical positions was also found to be similar but not exact (as shown in Figure 4.7).

Figure 4.7 Alignment of three cytochrome *c* protein sequences showing intron positions

The alignment of cytochrome *c* protein sequences from *Aspergillus nidulans*, *Neurospora crassa*, and *Arabidopsis thaliana* are shown using the standard one-letter code for amino acid residues. Dashes represent gaps introduced into the sequence for optimal alignment. The glycine residue found to be conserved in all cytochrome *c* genes sequenced is indicated by red.

```
A. nidulans     Met A K G G D S Y S P G (Intron I  395 bp) D S   12
N. crassa       Met --------- G F S A G (Intron I  547 bp) D S   7
A. thaliana       Met Q V A D I S L Q G ------------------ D A   11

13   T K G A K L F E ---------------- T R C K Q C H T V E N G   32
8    K K G A N L F K ---------------- T R C A Q C H T L E E G   28
12   K K G A N L F K (Intron I  78 bp) T R C A Q C H T L K A G   32

33   G G H K V G P N L H G L F G R K T G Q A G G Y A Y T D A N   61
29   G G N K I G P A L H G L F G R K T G S V D G Y A Y T D A N   57
33   E G N K I G P E L H G L F G R K T G S V A G Y S Y T D A N   61

62   K Q A D V T W D E N S L (Intron II 61 bp) F K Y L E N P K   81
58   K Q K G I T W D E N T L ---------------- F E Y L E N P K   77
62   K Q K G I E W K D D T L (Intron II 59 bp) F E Y L E N P K   81

82   K Y I P G T K M A F G G L K K T K E R N D L I T (Intron III 105
78   K Y I P G T K M A F G G L K K D K D R N D I I T (Intron II  101
82   K Y I P G T K M A F G G L K K P K D R N D L I T ---------- 105

106  53 bp) Y L K E S T A   112
102  74 bp) F M K E A T A   108
106  ------ F L E E E T K   112
```

CHAPTER 5
CLONING AND ANALYSIS OF THE 5' REGION OF THE *cycA* GENE

5.1 INTRODUCTION

The *cycA* gene as published by Raitt *et al.* (1994), was found from this study to have the major translational ATG start codon positioned incorrectly (see Chapters 3 and 4). Consequently, due to the new location of the ATG codon, and the detection of one extra intron (Intron I), the clone of the *cycA* gene in plasmid R41 (Table 2.1, Figure 4.5), only contains 247 bp of cloned and sequenced promoter region (instead of 649 bp as published). It was therefore of interest to obtain additional promoter sequence in which to look for interesting regulatory regions, such as the yeast HAP1 motif involved in oxygen regulation. Therefore, library screening was performed to enable the cloning and characterisation of the 5' region of the *cycA* gene.

5.2 SCREENING OF AN *A. NIDULANS* LIBRARY

To isolate the 5' region of the *cycA* gene, 10 filters containing approximately 30,000 pfu from a λGEM-11 *A. nidulans* genomic amplified library (kindly donated by Michael Hynes, University of Melbourne), prepared from *E. coli* MH2088, were screened by plaque hybridisation (Section 2.15). A 298 bp *A. nidulans* cDNA *Bam*H1 fragment (Section 2.15.3) which contains 224 bp of the 5' region of the *cycA* gene was used as a probe. Hybridisation and washing conditions are as in Section 2.15.3.

Thirty-five positive plaques were detected from the primary screen, and five plaques were selected for second round screening. During second round screening four out of five plaques remained positive, and after third round screening three plaques remained positive. These were designated λLM5, λLM9 and λLM19 (Table 2.1). DNA was prepared from these clones as described in Section 2.5.6.

5.3 SOUTHERN BLOTTING AND RESTRICTION MAPPING OF POSITIVE λ CLONES

The three positive λ clones were digested (Section 2.6.1) with double and single combinations of *Eco*RI, *Hin*dIII, and *Xho* I. These enzymes were chosen as they cut once or twice within the published region of the *cycA* gene (Raitt *et al.*, 1994). In addition, *Eco*RI does not cut within the λGEM-11 arms, *Xho*I cuts cloned inserts out of the λGEM-11 arms, and *Hin*dIII cuts within the right arm. Digested λ DNA was

separated on a 1% agarose gel (Section 2.6.2), and the sizes of fragments obtained are summarised in Table 5.1. Clones λLM5 and λLM9 produced identical restriction enzyme patterns when digested with single and double combinations of *Eco*RI, *Hin*dIII, and *Xho*I, suggesting that λLM5 and λLM9 were identically isolated λ clones. However, λLM19 contained only a few identical restriction fragments with λLM5 and λLM9 (see Table 5.1) using the same combinations of enzymes, and thus appears to be an overlapping clone. A Southern blot of the two different clones, λLM9 and λLM19, was carried out to determine which restriction fragment contained the 5' region of the *cycA* gene. This was performed by hybridising a blot of the gel containing λLM9 and λLM19 clones digested with single and double combinations of *Eco*RI, *Hin*dIII, and *Xho* I (Figure 5.1a), to the 298 bp cDNA probe containing 224 bp of 5' *cycA* region (Figure 5.1b). Results from the Southern blot enabled a restriction map of both λ clones to be obtained (Figure 5.2), and confirmed that λLM9 and λLM19 are overlapping clones at the 5' region, with λLM19 containing less of the 3' *cycA* region. These results indicated that the 2.1 kb *Eco*RI fragment present in λLM9 and λLM19 contained the 5' region of the *cycA* gene.

5.4 SEQUENCE ANALYSIS OF CLONED λ DNA SHOWN TO HYBRIDISE TO THE *cycA* GENE

5.4.1 Subcloning and Sequencing of the 2.1 kb Fragment

To obtain additional promoter sequence of the *cycA* gene, the 2.1 kb *Eco*RI fragment that hybridised to the 298 bp cDNA probe in both λLM9 and λLM19 (Table 5.1, Figures 5.1a and 5.1b), was subcloned from λLM9 into pUC18 to give plasmid R133 (Table 2.1).

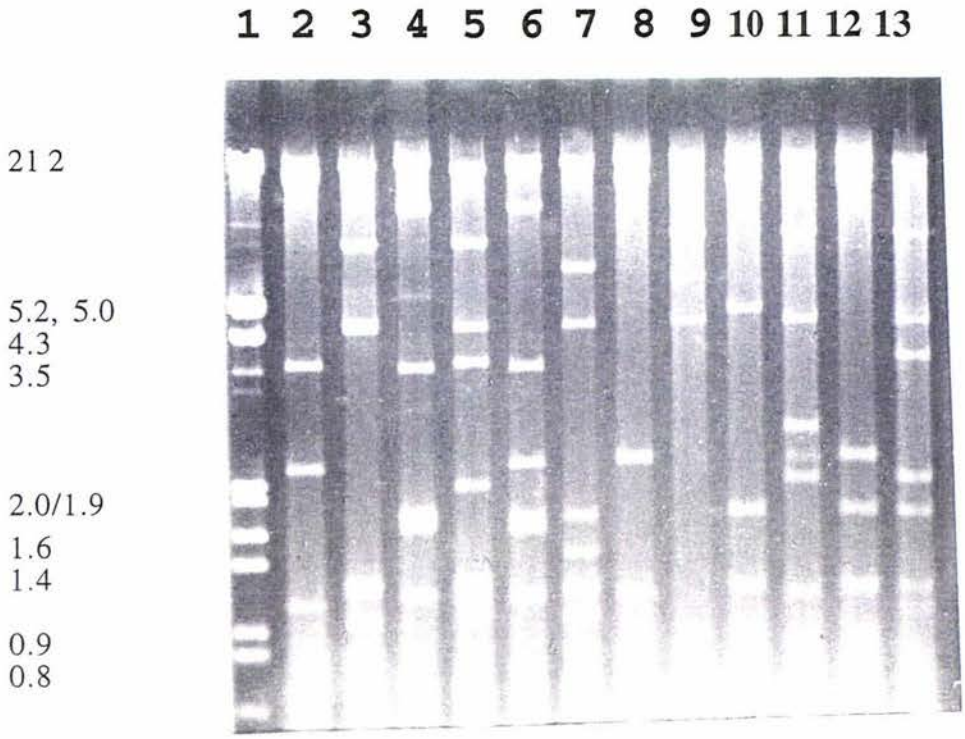
DNA from plasmid R133 was prepared by the rapid boiling method (Section 2.5), and sequenced by the Amplicycle sequencing system (Section 2.10.2). pUC/M13 universal forward (Perkin Elmer), pUC/M13 universal reverse (Perkin Elmer), and LG2 primers (Table 2.2) were used to obtain initial insert sequence. This information was used to design other primers (LG3, LG4, LM5, LM6, LM7, CYC8, Table 2.2) for directed sequencing. Figure 5.3 shows the sequencing strategy used to obtain the nucleotide sequence of the 5' region of the *cycA* gene from the cloned 2.1 kb *Eco*RI λ fragment. The nucleotide sequence is shown in Figure 5.4. Assembly of contigs and analysis of the 5' region of the *cycA* gene was performed using Sequencer Version 2.0.6 (Gene Codes Corporation).

Figure 5.1a Restriction digestion profiles of λLM9 and λLM19

1 kb Gibco BRL (Life Technologies) ladder (lane 1). Lanes 2-7 are digests of λLM9: *Eco*RI (lane 2), *Hin*dIII (lane 3), *Xho*I (lane 4), *Eco*RI and *Hin*dIII (lane 5), *Eco*RI and *Xho*I (lane 6), *Xho*I and *Hin*dIII (lane 7). Lanes 8-13 are digests of λLM19: *Eco*RI (lane 8), *Hin*dIII (lane 9), *Xho*I (lane 10), *Eco*RI and *Hin*dIII (lane 11), *Eco*RI and *Xho*I (lane12), *Xho*I and *Hin*dIII (lane 13).

Figure 5.1b Southern Blot of λLM9 and λLM19

Autoradiograph of a Southern blot of the gel shown above in Figure 5.1a, hybridised to [a-^{32}P]dCTP labelled 298 bp cDNA *Bam*HI fragment. Lanes as in Figure 5.1a.


1 2 3 4 5 6 7 8 9 10 11 12 13
21 2
5.2, 5.0
4.3
3.5
2.0/1.9
1.6
1.4
0.9
0.8

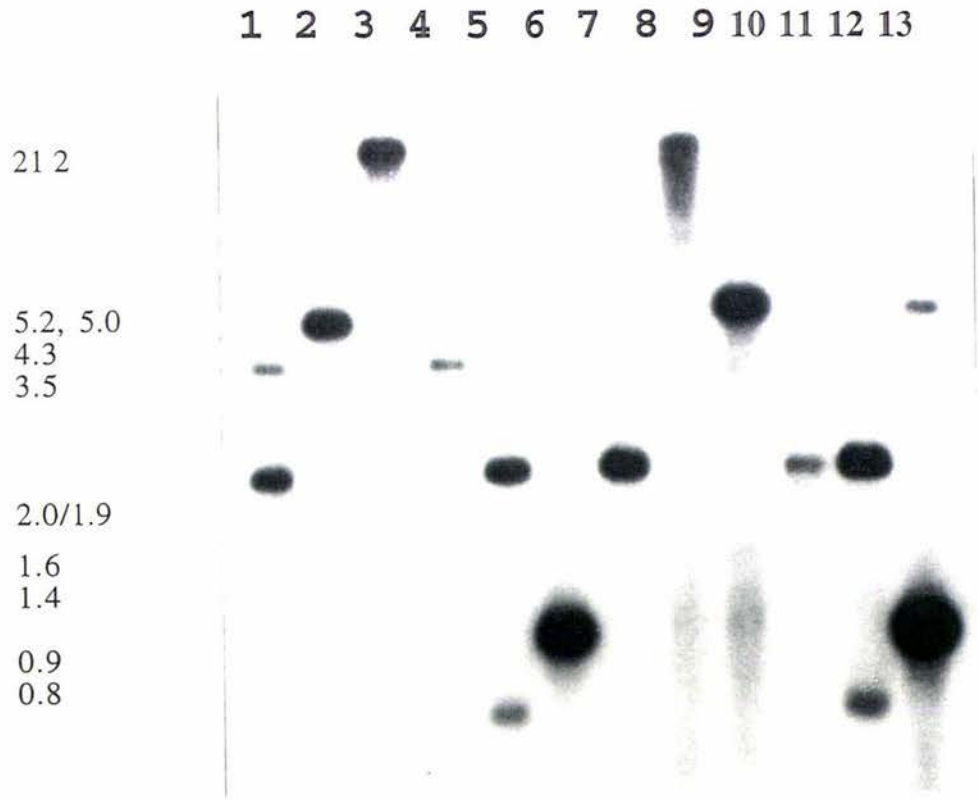

1 2 3 4 5 6 7 8 9 10 11 12 13
21 2
5.2, 5.0
4.3
3.5
2.0/1.9
1.6
1.4
0.9
0.8

Figure 5.2 Restriction maps of λLM9 and λLM19 isolated from an *A. nidulans* genomic library that hybridised to the *cycA* gene

Restriction maps of two *cycA* clones from *A. nidulans* strain A18. Arrow between two fragments indicates that the position of these fragments may be the reverse of that shown.

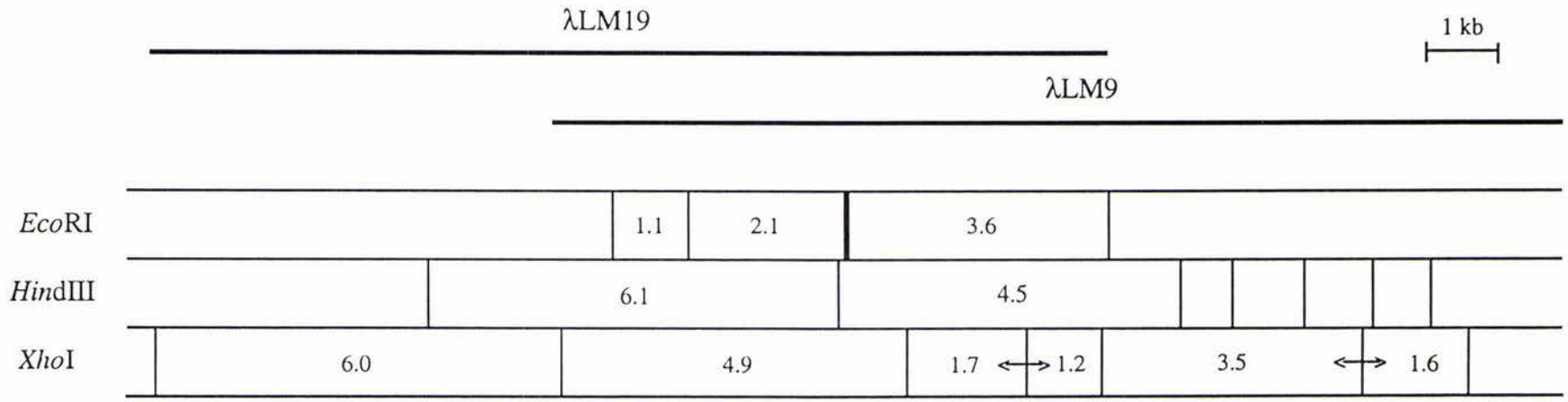

λLM19
1 kb
λLM9
EcoRI
1.1
2.1
3.6
HindIII
6.1
4.5
XhoI
6.0
4.9
1.7
1.2
3.5
1.6

Table 5.1 Restriction Enzyme Digestions of λ clones

λ Clone	Restriction digest	Fragment Size (KB)
λLM9	*Eco*RI	22.7, 17.5, 14.9, 3.6[a], 2.1[ab], 1.1[b]
	*Hin*dIII	20.0, 8.1, 4.5[a], 4.5[b], 1.2, 1.0, 0.8, 0.7
	*Xho*I	22.7[a], 10.8, 3.5, 1.7[b], 1.6, 1.2[b]
	*Eco*RI and *Hin*dIII	20.0, 8.1, 4.5[b], 3.6[a], 1.9[b], 1.2, 1.1[b], 1.0, 0.8, 0.7
	*Eco*RI and *Xho* I	22.7, 18.6, 10.8, 3.5, 2.1[ab], 1.7[b], 1.6, 1.2[b], 1.1[b], 0.7[ab]
	Xho I and *Hin*dIII	20.0, 6.4, 4.5, 1.7[b], 1.4, 1.2[b], 1.0[ab], 0.7, 0.6
λLM19	*Eco*RI	22.6, 15.7, 2.1[ab], 1.1[b]
	*Hin*dIII	22.7[a], 8.4, 6.1, 4.5
	*Xho*I	24.4, 20.0, 4.9[a], 1.7[b], 1.2[b]
	*Eco*RI and *Hin*dIII	22.7, 8.4, 4.5[b], 2.6, 2.1[a], 2.0, 1.2[b]
	*Eco*RI and *Xho*I	24.2, 20.0, 14.9, 2.1[c], 1.9[b], 1.7[b], 1.2[b], 1.1[b], 0.7[ab], 0.6
	*Xho*I and *Hin*dIII	20.0, 8.4,4.6[c], 4.5, 3.6, 2.0, 1.7[b], 1.2[b], 1.0[ab]

[a] indicates strong hybridisation signal with 298 bp cDNA probe.

[b] indicates identical restriction enzyme fragments present in both λLM9 and λLM19.

[c] indicates partial restriction fragment hybridised with 298 bp cDNA probe.

Figure 5.3 Nucleotide sequencing strategy for the 5' region of the *cycA* gene

The sequencing stratergy used to obtain additional promoter sequence from the 2.1 kb *Eco*RI fragment is shown, primer name is indicated above each arrow.

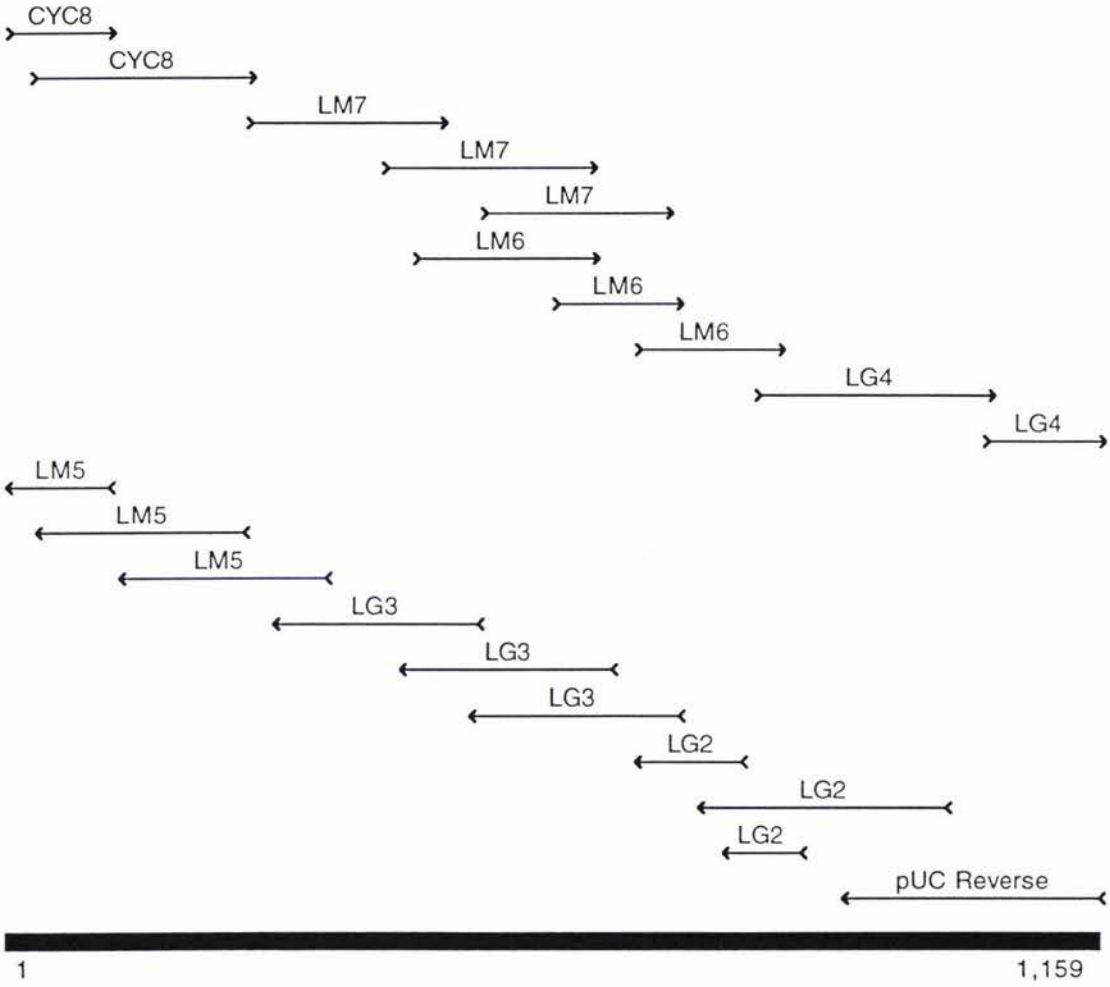

CYC8
CYC8
LM7
LM7
LM7
LM6
LM6
LM6
LG4
LG4
LM5
LM5
LM5
LG3
LG3
LG3
LG2
LG2
LG2
pUC Reverse
1
1,159

Figure 5.4 Nucleotide sequence of the 5' region of the *cycA* gene

The nucleotide sequence of the 5' region of the *A. nidulans cycA* gene is presented here. Nucleotides are numbered from the A of the proposed methionine translational start residue (+1) (this study, Chapter 4). The HAP1 consensus sequences are indicated by red characters, with nucleotides matching the CGG N3 TA N CGG N3 TA consensus underlined. The putative HAP2/3/4/5 binding site is shown in blue, TATA motifs are coloured green, and CREA sequences are indicated by pink.

```
                                .         .         .         .
-1247                           TCACATAGCTCCCAACCCAGAAAGCAGTTTGCGGGTAAAT
      .         .         .         .         .         .         .
-1207 GAGTACGCACAAAAGCAATCCAGACATGAATCCACCGACTCGTCAAAAACCGAAACATGA
      .         .         .         .         .         .         .
-1147 CCGTCCCCTCGGGCGGGAACATATTCGGGTACTTCTTTTTTGCCCGCTCCGCCTCTTCTT
      .         .         .         .         .         .         .
-1087 TCTCAGAGAACTTGGGACCGGGGTAGTTAACGACTTTACCATTGCGTGTTGGAACGACGC
      .         .         .         .         .         .         .
-1027 GGCGCGCAGGATGTGAGGGCGGGCGAAATTTATCTGGTTGCGCGAGGACGCGCGGATTTT
      .         .         .         .         .         .         .
-967  GGCTGTTGTCACTTGAGGAATTTGTTGATGCGCAGCGGTGCGTCGGTGGGAGCCTGCCGA
      .         .         .         .         .         .         .
-907  TGCGGGAGGACCGGGACAGGAATAGGATTGCTCGTGCTGAGGGAACACCTCGGGGAAGGA
      .         .         .         .         .         .         .
-847  GGAATGGGGGAATGGCTGATGGGGGCATTCTGTACGAATTGCTGGTTCTTGGCTGCGATT
      .         .         .         .         .         .         .
-787  TCTCTATATGCTAGCTTCTGGTCGCCGGCATAACATTTTGGTGTTGATATAATCATGTGA
      .         .         .         .         .         .         .
-727  CTTCTGCCGCCGGGAATAAGGCATCAAGGCATCAAGGCACAAACACATTTTTTCTAATGG
      .         .         .         .         .         .         .
-667  CCGCTAAGGCATCAGGCCACTTCGGATTAGGGCCGGGGAGAGCGGGAAAAACTCGCCATG
      .         .         .         .         .         .         .
-607  ACTAGGCGAATGAAAGGATGCAGATTTGTTATTACGGGGAGGGCTACTCCGGCCTCCGTA
      .         .         .         .         .         .         .
-547  GCCCACCGTTGCCCATTCCCCGAGACAGACAGTGCAGAGCTCCAAGTAACCAGCGTCTCT
      .         .         .         .         .         .         .
-487  ATGCGCTGGAATGAGGTCATGCCATGACGGCATGCAGAATCACCAATCATCTTTTACAGT
      .         .         .         .         .         .         .
-427  ATAGTAGTTAGGCTCTATGATAGATAGATGTCATAGAAGGTGCATTGTTGCTATCTAGAG
      .         .         .         .         .         .         .
-367  CTGCATAACTGAGCCCTTAGACGTAGTATATAGGATTACAATAGTCTCTAAATAAAGCTT
      .         .         .         .         .         .         .
-307  CATCCAGCCAGGCGTTATTTGCAGTGACCGATTCCTGACCTCAGACCCGGCAGCGCCCGT
      .         .         *         .         .         .         .
-247  TACTCTGAGCACAGTGCTGAATCATCCTACCTCTGATTGGTCAATTCCCAGATCACGGGT
      .         .         .         .         .         .         .
-187  GTCGTCGGGGGCGCGACCAAGAAACCAGCTCTACAAATTCCCTCCAAGTTTTTTCTTCC
      .         .         .         .         .         .         .
-127  CTTTGGCCAGTCCGCTTGACTTGAATTCTGTCTTTCATCTTCTCTGCTACATACAACTCT
      .         .         .         .         .         .         .
-67   TTGTACTATACCACTTACCTCTTTACATAACCCTTTCTCTCTACCTCTTTATTTTTTATC
      .         .         .         .
-7    ACTCACAATGGCTAAGGGCGGTGACAGCTACTCTCCTG
             M  A  K  G  G  D  S  Y  S  P  G
```

5.4.2 Confirmation that the 2.1 kb Fragment Corresponds to Regions of the *A. nidulans* A18 Genome

To verify that the 2.1 kb *Eco*RI fragment present in both λLM9 and λLM19, and isolated from λLM9, truly represents the 5' region of the *cycA* gene, Southern blots were performed against *A. nidulans* A18 (Table 2.1) genomic DNA with the 2.1 kb *Eco*RI λ fragment as a probe (Figures 5.5a and 5.5b). Southern blot analysis of *A. nidulans* A18 genomic DNA digested with *Eco*RI, and both *Eco*RI and *Hin*dIII, revealed that the λLM9 2.1 kb *Eco*RI and the 1.9 kb *Eco*RI-*Hin*dIII fragments (which were shown to hybridise to the 298 bp cDNA probe in Figure 5.1b) were also present in the *A. nidulans* A18 genome, thus confirming that the sequenced 2.1 kb *Eco*RI fragment represents the 5' region of the *cycA* gene as found in the genome. Unfortunately partial fragments were also detected by Southern blotting with *Eco*RI, and both *Eco*RI and *Hin*dIII digests, but the resulting fragment sizes were found to match the expected sizes for partial fragments using the information obtained from the restriction map of λLM9 (Figure 5.2).

5.4.3 Analysis of the *cycA* promoter sequence

5.4.3.1 Putative HAP1 UAS

The *cycA* upstream sequence was searched to detect sequences which were similar to the 'optimal' HAP1 binding site CGG N_3 TA N CGG N_3 TA (Ha *et al.*, 1996). It was anticipated that only imperfect versions of the 'optimal' consensus sequence would be found in the *cycA* gene, like all other known HAP1 binding sites (Ha *et al.*, 1996). The potential *A. nidulans* HAP1 sites were aligned with other known Yeast HAP1 binding sites (Figure 5.6). From this alignment, the HAP1 binding site at -634 nt from *A. nidulans* appears to best match the optimal sequence, however the other HAP1 sites cannot be disregarded (see Section 5.4.6 for discussion of the *A. nidulans* HAP1 regulatory sequences).

5.4.3.2 Putative HAP2/3/4/5 (Yeast CCAAT-binding Factor) UAS

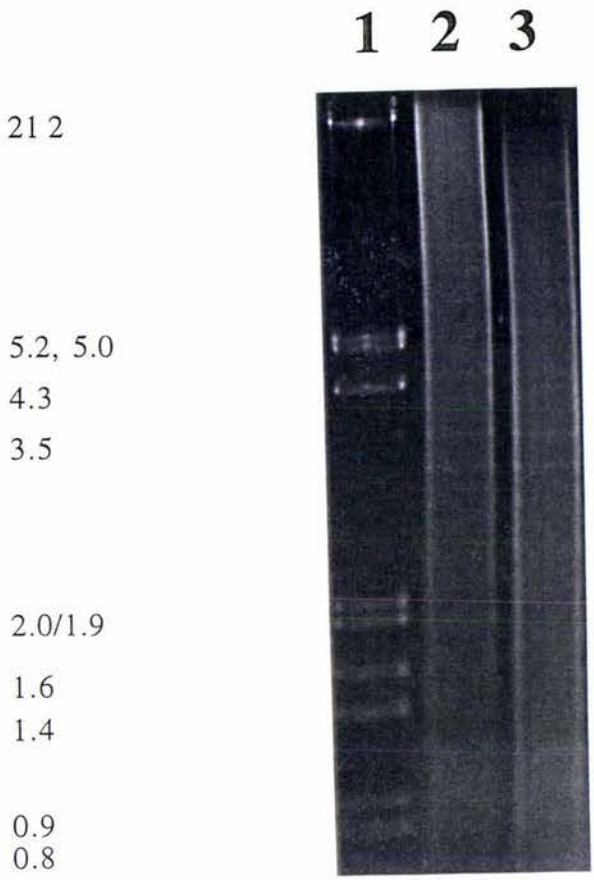
The region upstream of the *cycA* gene was also screened for the HAP2/3/4/5 complex (or Yeast CCAAT-binding factor) regulatory sequence. In the yeast *S. cerevisae,* this complex of proteins binds CCAAT sequences and activates many genes required for respiration (Zitomer and Lowry, 1992 for a review). An *A. nidulans* sequence was located at position -446 (Figure 5.4) that was 100% identical to the HAP2/3/4/5 consensus binding site TNATTGGT (or ACCAATNA).

Figure 5.5a Restriction digestions of *A. nidulans* A18 genomic DNA

Lanes 2 and 3 contain restriction digestions of *A. nidulans* A18 genomic DNA: *Eco*RI (lane 2), *Eco*RI and *Hin*dIII (lane 3). Lane 1 contains λ DNA digested with *Eco*RI and *Hin*dIII.

Figure 5.5b Southern blot of *A. nidulans* A18 genomic DNA

Autoradiograph of a Southern blot of the gel shown above in Figure 5.5a, hybridised to [a-^{32}P]dCTP labelled 2.1 kb *Eco*RI fragment from λLM9: *Eco*RI (lane 1), *Eco*RI and *Hin*dIII (lane 2).


1 2 3
21 2
5.2, 5.0
4.3
3.5
2.0/1.9
1.6
1.4
0.9
0.8

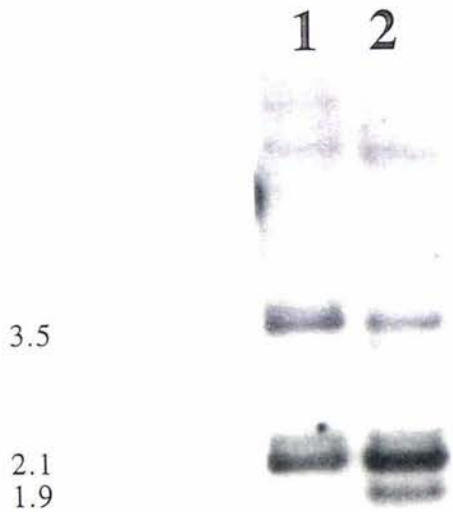

1 2
3.5
2.1
1.9

Figure 5.6 Comparison of known yeast and putative *A. nidulans* HAP1 UASs.

The known (functional) HAP1 binding sites from yeast are aligned with the putative HAP1 binding sites from the *A. nidulans cycA* gene. Characters in red indicate nucleotides which match the conserved CGG triplets or TA repeats given in the 'optimal' HAP1 sequence.

Known Yeast UASs of HAP1:

CYC1	TGGC CGG GGTTTA CGG ACGATGA
CYC7	CCCT CGC TATTAT CGC TATTAGC
CTT1	GGAA TGG AGATAA CGG AGGTTCT
CYB2	GGCA AGG AGATAT CGG CAGGCTT
CYT1	CCGC CGG AAATAC CGG CCGCCCA
CYT1 (reverse)	CGGC CGG TATTTC CGG CGGCCAA
K1CYC1 (reverse)	ATTT CGG GAACAT CGG TCAAGAC

A. nidulans putative HAP1 UAS:

CYCA (-634)	CCGC CGG GGAGAG CGG GAAAAGG
CYCA (-669)	TAAT GGC CGCTAA GGC ATCAGGC
CYCA (-905)	GATG CGG GAGGAC CGG GACAGGA
CYCA (+3, original)	CAAT GGC TAAGGG CGG TGACAGC
'OPTIMAL'	CGG NNNTAN CGG NNNTA

Genes with HAP1 sites:

CYC1: Iso-1-cytochrome *c* from *S. cerevisiae* (Pfeifer *et al.*, 1987).
CYC7: Iso-2-cytochrome *c* from *S. cerevisiae* (Prezant *et al.*, 1987).
CTT1: Catalase T from *S. cerevisiae* (Wrinkler *et al.*, 1988).
CYB2: Cytochrome b_2 from *S. cerevisiae* (Lodi and Guiard, 1991).
CYT1: Cytochrome c_1 from *S. cerevisiae* (Schneider and Guarente, 1991).
K1CYC1: Cytochrome *c* from *K.lactis* (Freire-Picos *et al.*, 1995).
CYCA: Cytochrome *c* from *A. nidulans* (+3, original; Raitt, 1992)

5.4.3.3 Putative CREA Binding Site

CREA protein (encoded by the *creA* gene) is a negative regulator involved in glucose repression in *A. nidulans* (Cubero and Scazzocchio, 1994). CREA is a GC box-binding protein (Ronne, 1995), and binds sequences with the consensus SYGGRG (Cubero and Scazzocchio, 1994). The binding ability of CREA is also context dependent, and seems to require an AT rich 5' context *in vitro* (Cubero and Scazzocchio, 1994) Three candidate CREA sites were found in the *cycA* gene at positions -561 (reverse and complement position), -635, and -1070 (Figure 5.4). Multiple CREA regulatory elements have been found in the promoters of other *A. nidulans* genes (Mathieu and Felenbok, 1994).

5.4.3.4 Other Regulatory Elements.

Three possible TATA boxes were present in the upstream region of the *cycA* gene at -293, -319, -341 nt positions (Figure 5.4). No significant pyrimidine rich elements were found, although an 18 bp pyrimidine tract extending from -124 to -139 nt (Figure 5.4), was located downstream from the proposed transcripitonal start site (Section 4.1). Therefore, either this element has no significance, or the proposed major transcriptional start site has been calculated incorrectly. Confirmation of the *cycA* transcription start site is currently under study in this laboratory.

The following *cis*-elements were not found in the *cycA* gene; a heat shock protein element, an iron responsive element (IRE), and a Yeast Ras/cAMP responsive element (PDS).

CHAPTER 5
DISCUSSION

An *A. nidulans* genomic library was sucessfully screened to obtain the 5' region of the *cycA* gene. The nucleotide sequence upstream of the *cycA* gene revealed a number of putative regulatory elements. The relevance of some of these *cis* regulatory elements will be discussed below.

5.4.4 TATA Box

Three possible TATA boxes were present in the 5' region of the *cycA* gene at 64, 90, and 112 nucleotides upstream of the estimated major transcriptional start site (Figure 5.4). Usually, TATA consensus sequences are found within 30-120 bp of the major transcriptional start site in fungal genes (Punt and van den Hondel, 1992). All three putative TATA boxes present in the *A. nidulans* promoter are located within this range, and therefore any one of them might represent the TATA box.

However, not all filamentous fungal promoters contain TATA boxes and, when they are present, promoters usually contain only one copy (Punt and van den Hondel, 1992; Gurr *et al.,* 1987). In contrast, in *S. cerevisiae*, multiple TATA boxes are often found at various positions (Gurr *et al.,* 1987). This is true for the *S. cerevisiae CYC1* gene, which contains three TATA boxes (Smith *et al.,* 1979). The *cycA* gene of *A. nidulans* therefore might be unusual for a filamentous fungal gene and contain three functional TATA boxes, although this seems unlikely.

In filamentous fungi, no regulatory proteins have been identified which bind to the TATA elements, thus the regulatory role of these elements in fungal genes is uncertain. However, deletion analysis of *A. nidulans trpC* promoter has suggested that the TATA box may be important in transcription iniation (Hamer and Timberlake, 1987). Whether one or all three of the TATA like boxes present in the *A. nidulans cycA* promoter are functional, will need to be determined experimentally.

5.4.5 The AnCF (*A. nidulans* CCAAT binding Factor) complex may regulate the *cycA* gene

A CCAAT motif was found in the upstream region of the *cycA* gene at position -446 nt (Figure 5.4), which shows complete homology to the binding site of the *S. cerevisae* HAP2/3/4/5 complex (or yeast CCAAT-binding factor). In the yeasts *S. cerevisiae* and

S. pombe, this heteromeric complex is involved in the activation of respiratory genes, such as those encoding cytochrome *c*, in response to a non-fermentable carbon source.

Research into the regulation of the *A. nidulans cycA* gene has revealed that its expression is not repressed by glucose (Raitt, 1992; Raitt *et al.*, 1994). Since the HAP2/3/4/5 complex is responsible for glucose repression, the homology to this regulatory site may only be fortuitous. However, this study in *A. nidulans* was not performed on a number of different strains to determine if *A. nidulans* is like the aerobic yeast *K. lactis*, whose cytochrome *c* gene (*KlCYC1*) was only glucose repressed in some strains (Freire-Picos *et al.*, 1995; Mulder *et al.*, 1995).

The *A. nidulans amds* gene, which encodes an acetamidase enzyme, is controlled by both carbon and nitrogen regulation, and contains a CCAAT sequence in the 5' regulatory region which is required for setting basal levels of expression (Bonnefoy *et al.*, 1995; van Heeswijck and Hynes, 1991). A protein complex named AnCF (A. nidulans CCAAT binding Factor) was found to bind to the CCAAT sequence (van Heeswijck and Hynes, 1991). Recent results suggest that the AnCF complex is analogous to the yeast HAP2/3/4/5 complex (Bonnefoy *et al.*, 1995; Papagiannopoulos *et al.*, 1996a; Papagiannopoulos *et al.*, 1996b). Firstly, the yeast complex was found to recognise and activate expression of the *amds* gene through the *amds* CCAAT sequence (Bonnefoy *et al.*, 1995). Secondly, homologs of the *S. cerevisiae* HAP3 and HAP5 components of the yeast CCAAT-binding factor have been isolated from *A. niduilans* (Papagiannopoulos *et al.*, 1996a; Papagiannopoulos *et al.*, 1996b). Deletion of the *A. nidulans* HAP3 homolog (*hapC*), resulted in decreased levels of *amds* expression under carbon limiting conditions, suggesting that it may encode a component of the AnCF complex. The HAP5 homolog (*hapE*) has yet to be disrupted. Interestingly, sequences other than the CCAAT sequence in the *amds* 5' regulatory region were required to achieve a response to carbon limitation, suggesting a difference in the response of the HAP2/3/4/5 complex to glucose repression in *A. nidulans* compared to that seen in *S. cerevisiae*.

Research on the *K. lactis HAP2* and *HAP3* genes has suggested that these genes, in contrast to their *S. cerevisiae* counterparts, were not essential for the expression of the *KlCYC1* gene (Mulder *et al.*, 1994; NGuyen *et al.*, 1995). Thus, the equivalent of the *S. cerevisiae* HAP2/3/4/5 complex in *K. lactis* does not seem to be involved in the response to glucose repression, (which only occurs in some strains), but will probably play a more general role in gene regulation, as found for the *A. nidulans amds* gene. In addition, the human HAP2 homolog, CP1, in combination with related factors, binds to CCAAT sequences in promoter regions of human genes to play a constitutive and cell type specific role in gene expression (Becker *et al.*, 1991).

In summary, it appears that the yeast HAP2/3/4/5 complex has been conserved from humans (Becker *et al.*, 1991) through to filamentous fungi (Bonnefoy *et al.*, 1995; Papagiannopoulos *et al.*, 1996a; Papagiannopoulos *et al.*, 1996b), due to the discovery of an analogous complex in *A. nidulans* (AnCF complex). The AnCF complex and the CCAAT-binding complex in *K. lactis* and humans seems to differ from the *S. cerevisiae* HAP2/3/4/5 complex by having no role in the control of glucose repression. Instead, the AnCF complex is required to set the basal level of *amds* transcription. Thus, the cellular role of the CCAAT-binding transcription factors seems to have diverged through evolution. Due to the modular nature of the HAP proteins, it is possible that the loss of carbon control by the CCAAT binding factors, is through the evolution of the HAP4 subunit which confers the yeasts complex's response to glucose repression.

Regardless of whether the *cycA* gene is found to be regulated by carbon source in some strains of *A. nidulans*, if the AnCF complex acts via the CCAAT sequence, it will probably affect the expression of the *cycA* gene by setting a basal level of transcription. Verification of this will have to await further experimental studies on the *cycA* gene of *A. nidulans*.

5.4.6 The *A. nidulans cycA* gene may by regulated by a HAP1-like Protein.

Three candidate HAP1 binding sites were found in the *cycA* 5' region, and are aligned with the known yeast HAP1 UASs in Figure 5.6. Due to the diversity of natural HAP1 binding sites (Ha *et al.*, 1996), it is difficult to determine which one, if any, are the most likely HAP1 binding site(s) before performing further experiments. The original HAP1 site chosen at the start of this study (on the basis of available results by Pfiefer *et al.*, 1987), and later determined to be located in the coding region of the gene (Section 3.7), matches least well with the 'optimal' HAP1 consensus sequence. In the light of recent results by Ha *et al.* (1996), this putative HAP1 site would not have been chosen as a candidate sequence. The study by Ha *et al.* (1996) showed that HAP1 will only bind to <u>direct</u> CGG repeats with a 6 bp spacer. If the CGG repeats are not conserved (ie. are degenerate forms), the TA repeats positioned asymmetrically in the spacer region are then essential for HAP1 binding. Thus the original HAP1 site at position +3 (Figure 5.6) would not have been chosen primarily because it has an <u>inverted</u> repeat of the GC-rich triplets.

The *A. nidulans* candidate HAP1 site at -669 forms a strong HAP1 binding site, it has a direct repeat of GGC, and has the TA sequence to stabilise protein binding. The other putative HAP1 site at -634 forms a possibly stronger HAP1 site, it has a direct repeat of

the optimal CGG triplet, but does not have the TA sequence, (which was determined by Ha *et al.* (1996) not to be necessary when the CGG triplet repeats were maintained). The sequence at -905 is virtually identical to the -634 site, but is missing one preferred A nucleotide when compared to the -634 site. However, the -905 site is situated 900 bp upstream from the ATG codon, whereas most regulatory binding sites are usually found closer to the ATG. Thus the -634 HAP1 site is proposed to be the most likely binding site for HAP1, but the possibility exists that both the -634 and -669 sites, only 35 bp apart, may both be HAP1 sites. Other regulatory proteins have been found to have two closely located sites which are both required, for example the *A. nidulans* CREA GC-rich binding protein (Cubero and Scazzocchio, 1994). In addition, only sites which have a 6 bp spacer between the CGG triplet repeats have been selected as putative HAP1 sites, but it is interesting to note that there were sequences present in the *cycA* promoter which had CGG repeats with a 5 bp spacer.

Since there are two HAP1 sites in the *cycA* gene which closely fit the HAP1 'optimal' consensus sequence, I speculate that a HAP1-like protein homologous to those which regulate oxygen transcriptional activation of many yeast respiratory genes, will probably also regulate the *A. nidulans cycA* gene. The functional significance of these sites in the regulation of the *cycA* gene is currently being investigated in this laboratory.

CHAPTER 6
SUMMARY AND CONCLUSIONS

A promoter-reporter vector containing the *A. nidulans cycA* promoter fused to the *lacZ* gene of *Escherichia coli* was constructed to allow investigation of the role of upstream regulatory elements, primarly the putative HAP1 binding site, in the regulation of the *A. nidulans cycA* gene. However, sequencing of the *cycA* promoter in the new R122 reporter vector revealed a sequencing discrepancy when compared to the published *cycA* sequence (Raitt *et al.,* 1994). Confirmation that the published sequence was incorrect was obtained by sequencing the R41 plasmid, which contains a sub-clone of the *cycA* gene from an *A. nidulans* genomic library. These results clearly confimed that six thymidine bases, rather than the published four, were present in the coding region of the *cycA* gene. Discovery of a sequencing error in the coding region of the *cycA* gene brought into question the validity of the open reading frames of the published *cycA* gene by Raitt *et al.* (1994). However, by comparison to other cytochrome *c* sequences it was considered that the deduced published amino acid sequence downstream from the sequencing error was correct. This implies that the published translation start codon was not correct, because it was no longer in the correct reading frame to produce the highly conserved cytochrome c protein. Further examination of the *cycA* gene showed that both the sequencing error and the published translation start codon may fall within a previously undetected intron region. To determine if an additional intron was present in the *cycA* gene, and thus provide further confirmation that the published translational start point was incorrect, RT-PCR analysis was performed on *cycA* RNA.

Two amplified products of 596 bp and 298 bp, indicative for the splicing of three introns, were produced by RT-PCR using two different pairs of primers. Subsequent sequencing of both these products proved that the *cycA* gene contained three introns instead of the two as published by Raitt *et al.* (1994).

Consequently, the published ATG initiation codon falls within the region of the previously undetected intron (Intron I). Therefore, a new translational start site was proposed, which contained a strong Kozak sequence for initiation of translation, and was the first AUG in the mRNA. Thus, the *cycA* gene has a corrected N-terminal region which was found to be very similar with the N-terminal region of the *S. cerevisae CYC7* gene. In addition, the position of the additional intron (Intron I) was conserved, being found at an identical position to that of an intron in the *Neurospora crassa* cytochrome *c* gene.

Due to the re-location of the translational ATG codon, the *cycA* promoter cloned into the reporter vector would now contain only a small portion of promoter sequence, and therefore this construct would be of no further use. In addition, the putative HAP1 binding site (Raitt, 1992) was now known to be situated in the coding region of the gene, and hence is not an upstream regulatory element for analysis.

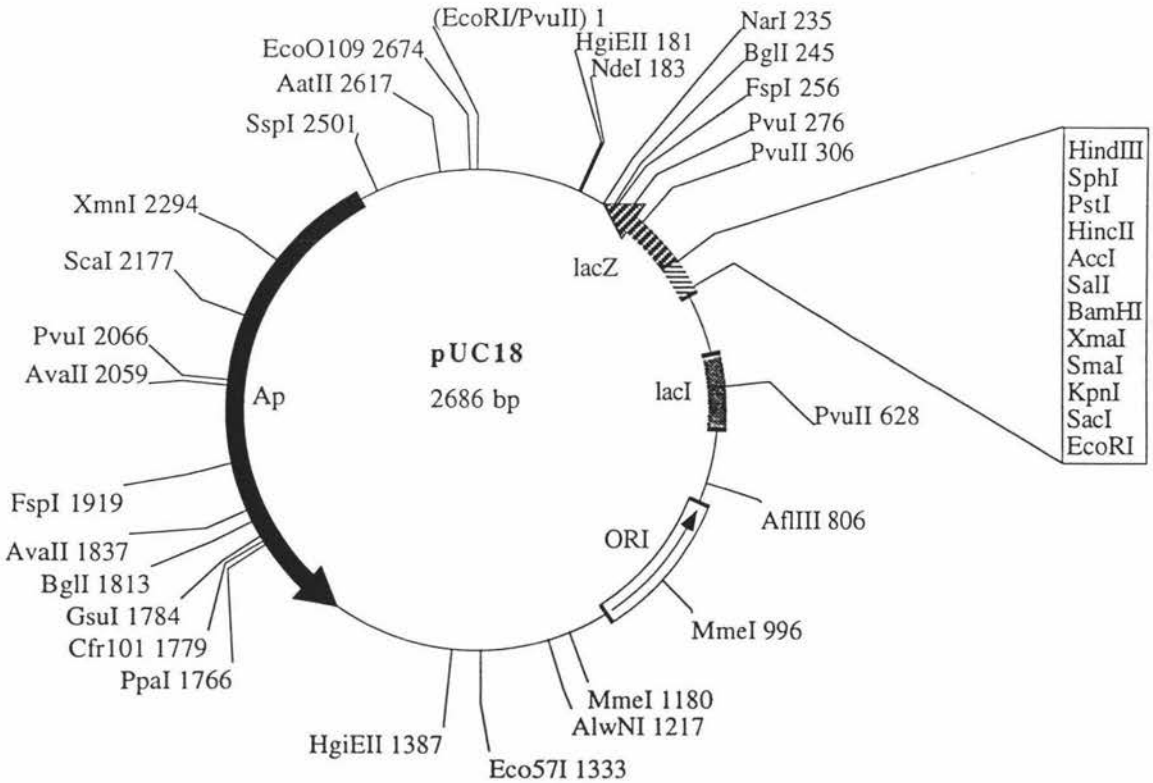
To obtain the entire *cycA* promoter region for further analysis, an *A. nidulans* genomic library was screened with the 298 bp *A. nidulans* cDNA *Bam*H1 fragment, which contains 224 bp of the 5' region of the *cycA* gene. Three *cycA* clones were isolated, and from restiction enzyme analysis of these three clones, the λLM9 clone was found to be identical to the λLM5 clone (since they contained the same restriction fragments), and the λLM9 and λLM19 clones were thought to be overlapping. Restriction enzyme and Southern blotting analysis of the two overlapping *cycA* clones, showed that a 2.1 kb *Eco*RI lambda fragment contained the 5' region of the *cycA* gene. This fragment was then subcloned from λLM9 into pUC18 to give plasmid R133. Sequencing of plasmid R133 (Table 2.1) enabled 1247 bp of *cycA* promoter sequence to be obtained. Analysis of this region revealed possible consensus sequences for the HAP1, AnCF complex, and the CREA transcription factors. *In vitro* and *in vivo* analysis of the *cycA* promoter is currently being performed in this labarotory to determine the significance of these sites. *In vitro* analysis will include gel mobility shift assays, where fragments of the *cycA* promoter will be tested with nuclear extracts from both *A. nidulans* and *S. cerevisae*. *In vivo* analysis in *A. nidulans* will be performed using reporter vector constructs containing either the entire *cycA* promoter, or the *cycA* promoter with appropriate deletions.

APPENDICES

Appendix 1.0 Vector Maps

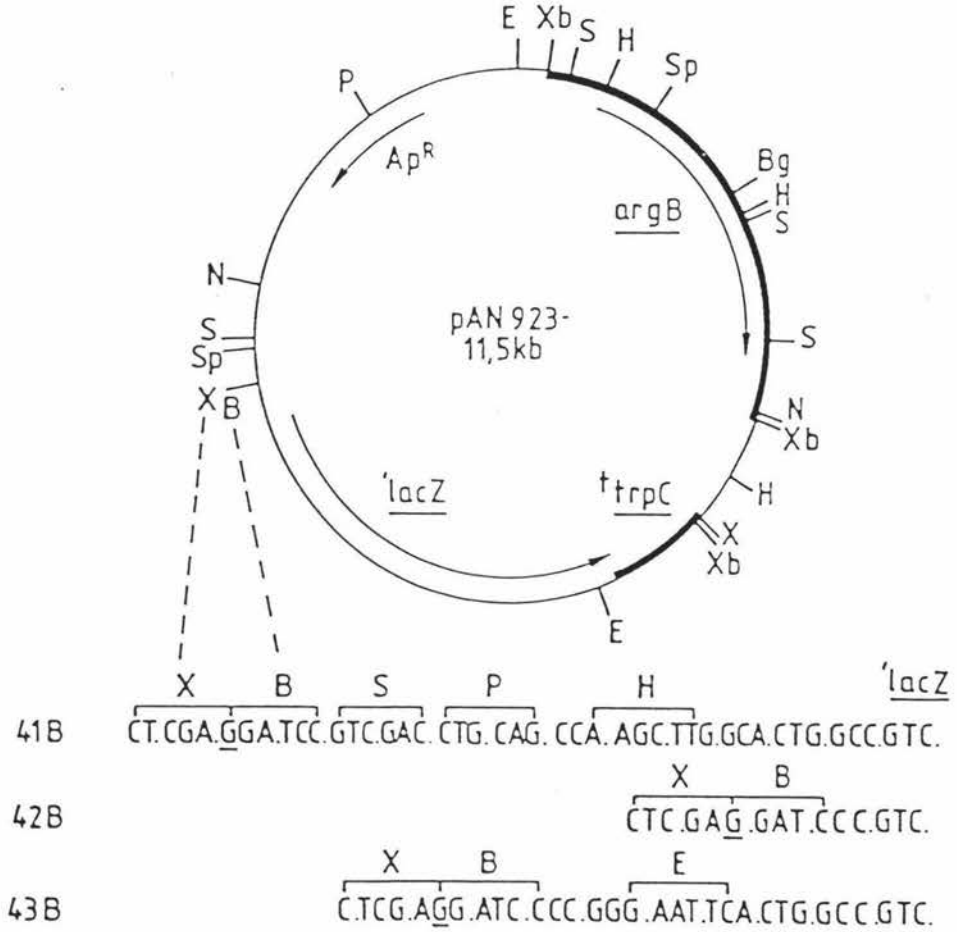
Restriction map of pUC18 showing all restriction enzyme sites.

Plasmid R121 was created by insertion of the 0.68 kb *Bam*HI fragment of *A. nidulans* into the *Bam*HI site of pUC18 (Section 3.3.3). Plasmid R127 was constructed by insetion of the 2.1 kb *Eco*RI fragment of *A. nidulans* into the *Eco*RI site of pUC18 (Section 5.4.1).



Restriction map of pAN923-41B, 42B (R109) and 43B, showing sites for B, *Bam*HI; Bg, *Bgl*II; E, *Eco*RI; H, *Hind*III; N, *Nru*I; P, *Pst*I; S, *Sal*I; Sp, *Sph*I; Ss, *Sst*I; X, *Xho*I; Xb, *Xba*I.

Plasmids R120 (Section 3.3.1) and R122 (Section 3.3.4) were created by the insertion of the 0.68 kb *Bam*HI fragment of *A. nidulans* into the *Bam*HI site of pAN923-42B in the wrong and correct orientation respectively.



Appendix 2.0 Sequence Data

Summary of sequence obtained from the 2.1 kb *Eco*RI *cycA* fragment

```
CYC8    TCACATATTA GTCACATAGC TCCCAACCCA GAAAGCAGTT
CYC8                                 CCA GAAAGCAGTT
LM5     TCACATATTA GTCACATAGC TCCCAACCCA GAAAGCAGTT
LM5                                      GAAAGCAGTT
1       ----------------------------------------------------------
        TCACATATTA GTCACATAGC TCCCAACCCA GAAAGCAGTT

CYC8    TGCGGGTAAA TGAGTACGCA CAAAAGCAAT CCAGACATGA
CYC8    TGCGGGTAAA TGAGTACGCA CAAAAGCAAT CCAGACATGA
LM5     TGCGGGTAAA TGAGTACGCA CAAAAGCAAT CCAGACATGA
LM5     TGCGGGTAAA TGAGTACGCA CAAAAGCAAT CCAGACATGA
41      ----------------------------------------------------------
        TGCGGGTAAA TGAGTACGCA CAAAAGCAAT CCAGACATGA

CYC8    ATCCACCGAC TCGTCAAAAA CCGAAACATG AC
CYC8    ATCCACCGAC TCGTCAAAAA CCGAAACATG ACCGTCCCCT
LM5     ATCCACCGAC TCGTCAAAAA CCGAAACATG AC
LM5     ATCCACCGAC TCGTCAAAAA CCGAAACATG ACCGTCCCCT
81      ----------------------------------------------------------
        ATCCACCGAC TCGTCAAAAA CCGAAACATG ACCGTCCCCT

CYC8    CGGGCGGGAA CATATTCGGG TACTTCTTTT TTGCCCGCTC
LM5     CGGGCGGGAA CATATTCGGG TACTTCTTTT TTGCCCGCTC
LM5     CGGGCGGGAA CATATTCGGG TACTTCTTTT TTGCCCGCTC
121     ----------------------------------------------------------
        CGGGCGGGAA CATATTCGGG TACTTCTTTT TTGCCCGCTC

CYC8    CGCCTCTTCT TTCTCAGAGA ACTTGGGACC GGGGTAGTTA
LM5     CGCCTCTTCT TTCTCAGAGA ACTTGGGACC GGGGTAGTTA
LM5     CGCCTCTTCT TTCTCAGAGA ACTTGGGACC GGGGTAGTTA
161     ----------------------------------------------------------
        CGCCTCTTCT TTCTCAGAGA ACTTGGGACC GGGGTAGTTA

CYC8    ACGACTTTAC CATTGCGTGT TGGAACGACG CGGCGCGCAG
LM5     ACGACTTTAC CATTGCGTGT TGGAACGACG CGGCGCGCAG
LM5     ACGACTTTAC CATTGCGTGT TGGAACGACG CGGCGCGCAG
201     ----------------------------------------------------------
        ACGACTTTAC CATTGCGTGT TGGAACGACG CGGCGCGCAG
```

```
CYC8    GATGTGAGGG CGGGCGAAA
LM7                    CGAAAT TTATCTGGTT GCGCGAGGAC
LM5     GATGTGAGGG CGGG
LM5     GATGTGAGGG CGGGCGAAAT TTATCTGGTT GCGCGAGGAC
241     ----------------------------------------------------------
        GATGTGAGGG CGGGCGAAAT TTATCTGGTT GCGCGAGGAC

LM7     GCGCGGATTT TGGCTGTTGT CACTTGAGGA ATTTGTTGAT
LM5     GCGCGGATTT TGGCTGTTGT CACTTGAGGA ATTTGTTGAT
LG3       GCGGATTT TGGCTGTTGT CACTTGAGGA ATTTGTTGAT
281     ----------------------------------------------------------
        GCGCGGATTT TGGCTGTTGT CACTTGAGGA ATTTGTTGAT

LM7     GCGCAGCGGT GCGTCGGTGG GAGCCTGCCG ATGCGGGAGG
LM5     GCGCAGCGGT GCGTCGGTGG
LG3     GCGCAGCGGT GCGTCGGTGG GAGCCTGCCG ATGCGGGAGG
321     ----------------------------------------------------------
        GCGCAGCGGT GCGTCGGTGG GAGCCTGCCG ATGCGGGAGG

LM7     ACCGGGACAG GAATAGGATT GCTCGTGCTG AGGGAACACC
LM7                                             CACC
LG3     ACCGGGACAG GAATAGGATT GCTCGTGCTG AGGGAACACC
361     ----------------------------------------------------------
        ACCGGGACAG GAATAGGATT GCTCGTGCTG AGGGAACACC

LM7     TCGGGGAAGG AGGAATGGGG GAATGGCTGA TGGGGGCATT
LM7     TCGGGGAAGG AGGAATGGGG GAATGGCTGA TGGGGGCATT
LM6                                      GGGGGCATT
LG3     TCGGGGAAGG AGGAATGGGG GAATGGCTGA TGGGGGCATT
LG3                  TGGGG GAATGGCTGA TGGGGGCATT
401     ----------------------------------------------------------
        TCGGGGAAGG AGGAATGGGG GAATGGCTGA TGGGGGCATT

LM7     CTGTACGAAT TGCTGGTTCT T
LM7     CTGTACGAAT TGCTGGTTCT TGGCTGCGAT TTCTCTATAT
LM6     CTGTACGAAT TGCTGGTTCT TGGCTGCGAT TTCTCTATAT
LG3     CTGTACGAAT TGCTGGTTCT TGGCTGCGAT TTCTCTATAT
LG3     CTGTACGAAT TGCTGGTTCT TGGCTGCGAT TTCTCTATAT
441     ----------------------------------------------------------
        CTGTACGAAT TGCTGGTTCT TGGCTGCGAT TTCTCTATAT

LM7     GCTAGCTTCT GGTCGCCGGC ATAACATTTT GGTGTTGATA
LM7                           ATAACATTTT GGTGTTGATA
LM6     GCTAGCTTCT GGTCGCCGGC ATAACATTTT GGTGTTGATA
LG3     GCTAGCTTCT GGTCGCCGGC A
LG3     GCTAGCTTCT GGTCGCCGGC ATAACATTTT GGTGTTGATA
LG3              T GGTCGCCGGC ATAACATTTT GGTGTTGATA
481     ----------------------------------------------------------
```

```
LM7     TAATCATGTG ACTTCTGCCG CCGGGAATAA GGCATCAAGG
LM7     TAATCATGTG ACTTCTGCCG CCGGGAATAA GGCATCAAGG
LM6     TAATCATGTG ACTTCTGCCG CCGGGAATAA GGCATCAAGG
LG3     TAATCATGTG ACTTCTGCCG CCGGGAATAA GGCATCAAGG
LG3     TAATCATGTG ACTTCTGCCG CCGGGAATAA GGCATCAAGG
521     ----------------------------------------------------------
        TAATCATGTG ACTTCTGCCG CCGGGAATAA GGCATCAAGG

LM7     CATCAAGGCA CAAACACATT TTTTCTAATG GCCGCTAAGG
LM7     CATCAAGGCA CAAACACATT TTTTCTAATG GCCGCTAAGG
LM6     CATCAAGGCA CAAACACATT TTTTCTAATG GCCGCTAAGG
LM6                        ATT TTTTCTAATG GCCGCTAAGG
LG3     CATCAAGGCA CAAACACATT TTTTCTAATG GCCGCTAAGG
LG3     CATCAAGGCA CAAACACATT TTTTCTAATG GCCGCTAAGG
561     ----------------------------------------------------------
        CATCAAGGCA CAAACACATT TTTTCTAATG GCCGCTAAGG

LM7     CATCAGGCCA CTTCGGATTA G
LM7     CATCAGGCCA CTTCGGATTA GGGCCGGGGA GAGCGGGAAA
LM6     CATCAGGCCA CTTCGGATTA GGGC
LM6     CATCAGGCCA CTTCGGATTA GGGCCGGGGA GAGCGGGAAA
LG3     CATCAGGCCA CTTCGGATTA GGGCCGGGGA GAGCGGGAAA
LG3     CATCAGGCCA CTTCGGATTA GGGCCGGGGA GAGCGGGAAA
601     ----------------------------------------------------------
        CATCAGGCCA CTTCGGATTA GGGCCGGGGA GAGCGGGAAA

LM7     AACTCGCCAT GACTAGGCGA ATGAAAGGAT GCAGATTTGT
LM6     AACTCGCCAT GACTAGGCGA ATGAAAGGAT GCAGATTTGT
LM6                                   GGAT GCAGATTTGT
LG3     AACTC
LG3     AACTCGCCAT GACTAGGCGA ATGAAAGGAT GCAGATTTGT
LG2                                   GGAT GCAGATTTGT
641     ----------------------------------------------------------
        AACTCGCCAT GACTAGGCGA ATGAAAGGAT GCAGATTTGT

LM7     TATTACGGGG AGGGCTACTC C
LM6     TATTACGGGG AGGGCTACTC CGGCCTCCGT A
LM6     TATTACGGGG AGGGCTACTC CGGCCTCCGT AGCCCACCGT
LG3     TATTACGGGG AGGGCTACTC CGGCCTCCGT AGCCCA
LG2     TATTACGGGG AGGGCTACTC CGGCCTCCGT AGCCCACCGT
681     ----------------------------------------------------------
        TATTACGGGG AGGGCTACTC CGGCCTCCGT AGCCCACCGT
```

```
LM6          TGCCCATTCC CCGAGACAGA CAGTGCAGAG CTCCAAGTAA
LG2          TGCCCATTCC CCGAGACAGA CAGTGCAGAG CTCCAAGTAA
LG2                     CGAGACAGA CAGTGCAGAG CTCCAAGTAA
LG2                                                  AA
     721     ----------------------------------------------------------
             TGCCCATTCC CCGAGACAGA CAGTGCAGAG CTCCAAGTAA

LM6          CCAGCGTCTC TATGCGCTGG AATGAGGTCA TGCCATGACG
LG4                                             CCATGACG
LG2          CCAGCGTCTC TATGCGCTGG
LG2          CCAGCGTCTC TATGCGCTGG AATGAGGTCA TGCCATGACG
LG2          CCAGCGTCTC TATGCGCTGG AATGAGGTCA TGCCATGACG
     761     ----------------------------------------------------------
             CCAGCGTCTC TATGCGCTGG AATGAGGTCA TGCCATGACG

LM6          GCATGCAGAA TCACCAATC
LG4          GCATGCAGAA TCACCAATCA TCTTTTACAG TATAGTAGTT
LG2          GCATGCAGAA TCACCAATCA TCTTTTACAG TATAGTAGTT
LG2          GCATGCAGAA TCACCAATCA TCTTTTACAG TATAGTAGTT
     801     ----------------------------------------------------------
             GCATGCAGAA TCACCAATCA TCTTTTACAG TATAGTAGTT

LG4          AGGCTCTATG ATAGATAGAT GTCATAGAAG GTGCATTGTT
LG2          AGGCTCTATG ATAGATAGAT GTCATAGAAG GTGCATTGTT
LG2          AGGCTC
     841     ----------------------------------------------------------
             AGGCTCTATG ATAGATAGAT GTCATAGAAG GTGCATTGTT

LG4          GCTATCTAGA GCTGCATAAC TGAGCCCTTA GACGTAGTAT
LG2          GCTATCTAGA GCTGCATAAC TGAGCCCTTA GACGTAGTAT
pUC Reverse        TAGA GCTGCATAAC TGAGCCCTTA GACGTAGTAT
     881     ----------------------------------------------------------
             GCTATCTAGA GCTGCATAAC TGAGCCCTTA GACGTAGTAT

LG4          ATAGGATTAC AATAGTCTCT AAATAAAGCT TCATCCAGCC
LG2          ATAGGATTAC AATAGTCTCT AAATAAAGCT TCATCCAGCC
pUC Reverse  ATAGGATTAC AATAGTCTCT AAATAAAGCT TCATCCAGCC
     921     ----------------------------------------------------------
             ATAGGATTAC AATAGTCTCT AAATAAAGCT TCATCCAGCC

LG4          AGGCGTTATT TGCAGTGACC GATTCCTGAC CTCAGACCCG
LG2          AGGCGTTATT TGCAGTGACC GATTCCTGAC CTCAGA
pUC Reverse  AGGCGTTATT TGCAGTGACC GATTCCTGAC CTCAGACCCG
     961     ----------------------------------------------------------
             AGGCGTTATT TGCAGTGACC GATTCCTGAC CTCAGACCCG
```

```
LG4            GCAGCGCCCG TTACTCTGAG CACAGTGCTG AATCATCCTA
LG4                                                TCATCCTA
pUC Reverse    GCAGCGCCCG TTACTCTGAG CACAGTGCTG AATCATCCTA
     1001      ----------------------------------------------------------
               GCAGCGCCCG TTACTCTGAG CACAGTGCTG AATCATCCTA

LG4            CC
LG4            CCTCTGATTG GTCAATTCCC AGATCACGGG TGTCGTCGGG
pUC Reverse    CCTCTGATTG GTCAATTCCC AGATCACGGG TGTCGTCGGG
     1041      ----------------------------------------------------------
               CCTCTGATTG GTCAATTCCC AGATCACGGG TGTCGTCGGG

LG4            GGCGCGACCA AGAAACCAGC TCTACAAATT CCCTCCAAGT
pUC Reverse    GGCGCGACCA AGAAACCAGC TCTACAAATT CCCTCCAAGT
     1081      ----------------------------------------------------------
               GGCGCGACCA AGAAACCAGC TCTACAAATT CCCTCCAAGT

LG4            TTTTTTCTTC CCTTTGGCCA GTCCGCTTGA CTTGAATTC
pUC Reverse    TTTTTTCTTC CCTTTGGCCA GTCCGCTTGA CTTGAATTC
     1121      ----------------------------------------------------------
               TTTTTTCTTC CCTTTGGCCA GTCCGCTTGA CTTGAATTC
```

REFERENCES

Abastado, J., Miller, P.F., Jackson, B.M. and Hinnebusch, A.G. (1990). Suppression of ribosomal reinitiation at upstream open reading frames in amino acid-starved cells forms the basis for GCN4 translational control. *Mol Cell Biol* **11:** 486-496.

Amegadzie, B.Y., Zitomer, R.S., Hollenberg, C.P. (1990). Characterization of the cytochrome *c* gene from the starch-fermenting yeast *Schwanniomyces occidentalis* and its expression in Baker's yeast. *Yeast* **6**:429-40.

Arst, H.N. and Scazzocchio, C. (1985). Formal genetics and molecular biology of the control of gene expression in *Aspergillus nidulans*. In: Bennet J.W., Lasure L.L. (eds). Gene Manipulation in fungi. Academic Press, New York, pp 306-343.

Becker, D.M., Fikes, J.D. and Guarente, L. (1991). A cDNA encoding a human CCAAT-binding protein cloned by functional complementation in yeast. *Proc Natl Acad Sci* **88:** 1968-1972.

Bennet, B.L. and Molenaar, A.J. (1994). Cloning of PCR products can be inhibited by Taq Polymerase carryover. *Biotechniques* **16:** 32-36.

Bonnefoy, N., Copsey, J., Hynes, M.J., Davis, M.A. (1995). Yeast proteins can activate expression through regulatory sequences of the *amdS* gene of *Aspergillus nidulans*. *Mol Gen Genet* **246:** 223-227.

Borck, K., Beggs, J.D., Brammar, W.J., Hopkins, A.S., and Murray, N.E. (1976). The construction *in vitro* of transducing derivatives of phage lambda. *Mol Gen Genet* **146:** 199-207.

Bottorff, D. A., Parmaksizoglu, S., Lemire, E. G., Coffin, J. W., Bertrand. H., Nargang, F. E. (1994). Mutations in the structural gene for cytochrome *c* result in deficiency of both cytochromes *aa3* and *c* in *Neurospora crassa*. *Curr Genet* **26:** 329-335.

Bullock, W.O., Fernandez, J.M., and Short, J.M. (1987). XL-1-Blue: a high efficiency plasmid transforming *recA Escherichia coli* strain with beta-galactosidase selection. *Biotechniques* **5:** 376-378.

Cariello, N.F., Swenberg, J.A., and Skopek, T.R. (1991). Fidelity of *Thermococcus Litoralis* DNA polymerase (Vent) in PCR determined by denaturing gradient gel electrophoresis. *Nucleic Acids Res* **19:** 4193-4198.

Chodosh, L.A., Olesen, J., Baldwin, A. S., Guarente, L. and Sharp, P.A. (1988). A yeast and a human CCAAT-binding protein have homologous subunits that are functionally interchangeable. *Cell* **53:** 25-35.

Clark, J. M. (1988). Novel non-templated nucleotide addition reactions catalyzed by procaryotic and eucaryotic DNA polymerases. *Nucleic Acids Res* **16:** 9677-9686.

Cohen, S.N., Chang, A.C.Y., Hsu, L. (1972). Non-chromosomal antibiotic resistance in bacteria: Genetic transformation of *E. coli* by R-factor DNA. *Proc Natl Acad Sci* **69:** 2110.

Collett, M.A. (1995). Cloning, characterisation and evolutionary relationships of two *pyr4* genes from an *Acremonium* endophyte of perennial ryegrass. Ph.D. Thesis, Massey University, Palmerston North, New Zealand.

Creusot, F., Verdiere, J., Gaisne, M. and Slonimski, P.P. (1988). Regulator of oxygen-dependent gene expression in yeast. *J Mol Biol* **204:** 263-276.

Cubero, B. and Scazzocchio, C. (1994). Two different, adjacent and divergent zinc finger binding sites are necessary for CREA-mediated carbon catabolite repression in the proline gene cluster of *Aspergillus nidulans. EMBO J* **13:** 407-415.

Davis, M.A., Hynes, M.J. (1991). Regulatory circuits in *Aspergillus nidulans*. In: Bennet J.W., Lasure L.L. (eds). More gene manipulations in fungi. Academic Press, San Diego, California, pp 151-189.

De Dekan, R.H. (1966). The Crabtree effect: a regulatory system in yeast. *J Gen Microbiol* **23:** 149-156.

Dower, W.J., Miller, J.F., and Ragsdale, C.W. (1988) High efficiency transformation of *Escherichia coli* by high voltage electroporation. *Nucleic Acids Res* **16:** 6127-6145.

Forsburg, S.L. and Guarente, L. (1989a). Communication between mitochondria and the nucleus in regulation of cytochrome genes in the yeast *Saccharomyces cerevisiae. Ann Rev Cell Biol* **5:** 153-180.

Forsburg, S.L. and Guarente, L. (1989b). Identification and characterisation of HAP4: a third component of the CCAAT-bound HAP2/HAP3 heteromer. *Genes Dev* **3:** 1166-1178.

Freire-Picos, M.A., Rodriguez Torres, A.M., Ramil, E., Cerdan, M.E., Breunig, K. D., Hollenberg, C.P., Zitomer, R. S. (1993). Sequence of a cytochrome *c* gene for *Kluveromyces lactis* and its upstream region. *Yeast* **9:** 201-204.

Freire-Picos, M.A., Hollenberg, C.P., Breunig, K.D., Cerdan, M.E. (1995). Regulation of cytochrome *c* expression in the aerobic respiratory yeast *Kluyveromyces lactis. Febs Lett* **360:** 39-42.

Frischauf, A., Lehrach, H., Poustka, A., and Murray, N. (1983). Lambda replacement vectors carrying polylinker sequences. *J Mol Biol* **170:** 827-842.

Fytlovich, S., Gervais, M., Agrimonti, C. and Guiard, B. (1993). Evidence for an interaction between the CYP1 (HAP1) activator and a cellular factor during heme-dependent transcriptional regulation in the yeast *Saccharomyces cerevisiae. EMBO J* **12:** 1209-1218.

Glasel, J.A. (1995). Validity of nucleic acid purities monitored by 260nm/280nm absorbance ratios. *Biotechniques* **18:** 62-63.

Guarente, L. and Mason, T. (1983). Heme Regulates Transcription of the *CYC1* Gene of *S. cerevisiae* via an Upstream Activation Site. *Cell* **32:** 1279-1286.

Guarente, L., Lalonde, B., Gifford, P., Alani, E. (1984). Distinctly regulated tandem upstream activation sites mediate catabolite repression of the *CYC1* gene of *S. cerevisiae. Cell* **36:** 503-11

Gurr, S., Unkles, S., Kinghorn, J. (1987). The structure and organization of nuclear genes of filamentous fungi. Society of General Microbiology (special publication) Chapter **5:** 93-139.

Ha, N., Hellauer, K. and Turcotte, B. (1996). Mutations in target DNA elements of yeast HAP1 modulate its transcriptional activity without affecting DNA binding. *Nucleic Acids Res* **24:** 1453-1459.

Haldi, M. L. and Guarente (1995). Multiple domains mediate heme control of the yeast activator HAP1. *Mol Gen Genet* **248:** 229-235.

Hamer, J.E. and Timberlake, W.E. (1987). Functional organisation of the *Aspergillus nidulans trpC* promoter. *Mol Cell Biol* **7:** 2352-2359.

Hanahan, D. (1983) Studies on transformation of *Escherichia coli* with plasmids. *J Mol Biol* **166:** 557-580.

Hudson, A. (1995). Fungal cytochrome *c* genes from plants. *J Mol Evol* **41**:1170-1171.

Holmes, D.S. and Quigley, M. (1981). A rapid boiling method for the preparation of bacterial plasmids. *Anal Biochem* **114:** 193-197.

Hooft van Huijsduijnen, R., Li, X.Y., Black, D., Matthes, H., Benoist, C. and Maithis, D. (1990). Co-evolution from yeast to mouse: cDNA cloning of the two NF-Y (CP1/CBF) subunits. *EMBO J* **9:** 3119-3127.

Ish-Horowicz, D. and Burke, J.F. (1981). Rapid and efficient cosmid cloning. *Nucleic Acids Res* **9:** 2989-2998.

Itakura, K., Rossi, J.J. and Wallace, R.B. (1994). Synthesis and use of synthetic oligonucleotides. *Ann Rev Biochem* **53:** 323.

Jon J.P. Bruchez, Eberle, J. and Russo, V.E.A. (1993). Regulatory sequences involved in the translation of *Neurospora crassa* mRNA: Kozak sequences and stop codons. Fungal Genetics Newsletter, No. 40.

Kawasaki, E.S. (1990). Amplification of RNA. In: *PCR Protocols: A Guide to Methods and Applications.* Eds. Innis, M.A., Gelfand, D.H., Sninsky, J.J., White, T.J. San Diego, Academic Press, Inc. 21-27.

Karn, J., Matthes, H.W., Gait, M.J. and Brenner, S. (1984). A new selective phage cloning vector, lambda 2001, with sites for *Xba*I, *Bam*HI, *Hind*III, *Eco*RI, *Sst*I and *Xho*I. *Gene* **32:** 217-224.

Kemmerer, E.C., Lei, M. and Wu, R. (1991a). Structure and molecular evolutionary analysis of a plant cytochrome *c* gene: Surprising implications for *Arabidopsis thaliana. J Mol Evol* **32:** 227-237.

Kemmerer, E.C., Lei, M. and Wu, R. (1991b). Isolation and molecular evolutionary analysis of a cytochrome *c gene* from *Oryza sativa* (Rice). *Mol Biol Evol* **8:** 212-226.

Kozak, M. (1991). Structural features in eukaryotic mRNAs that modulate the initiation of translation. *J Biol Chem* **266:** 19867-19870.

Lazowska, J., Claisse, M., Gargouri, A., Kotylak, Z., Spyridakis, A. and Slonimski, P. (1989). *J Mol Biol* **205:** 275-289.

Lue, N.F., Chasman, D.I., Buchman, A.R. and Kornberg, R.D. (1987). Interaction of GAL4 and GAL80 gene regulatory proteins *in vitro. Mol Cell Biol* **7:** 3446-3451.

Limbach, K.J. and Wu, R. (1983). Isolation and characterisation of two alleles of the chicken cytochrome *c* gene. *Nucleic Acids Res* **11:** 8931-8950.

Limbach, K.J. and Wu, R. (1985). Characterisation of a mouse somatic cytochrome *c* gene and three cytochrome *c* pseudogenes. *Nucleic Acids Res* **13:** 617-630.

Lodi, T. and Guiard, B. (1991). Complex transcriptional regulation of the *Saccharomyces cerevisiae CYB2* gene encoding cytochrome b_2: CYP1 (HAP1) activator binds to the *CYB2* upstream activation site UAS1-B2. *Mol Cell Biol* **11:** 3762-3772.

Maity, S.N., Vuorio, T. and de Crombrugghe, B. (1990). The B subunit of a rat heteromeric CCAAT-binding transcription factor shows striking sequence identity with the yeast HAP2 transcription factor. *Proc Natl Acad Sci* **87:** 5378-5382.

Maity, S. N., Sinha, S., Ruteshouser, E. C. and de Crombrugghe, B. (1992). Three different polypeptides are necessary for DNA binding of the mammalian heteromeric CCAAT binding factor. *J Biol Chem* **267:** 16574-16580.

Marchuk, D., Drumm, M., Saulino, A. and Collins, F.S. (1990). Construction of T-vectors, a rapid and general system for direct cloning of unmodified PCR products. *Nucleic Acids Res* **19:** 1154.

Mathieu, M. and Felenbok, B. (1994). The *Aspergillus nidulans* CREA protein mediates glucose repression of the ethanol regulon at various levels through competition with the ALCR-specific transactivator. *EMBO J* **13:** 4022-4027.

McNabb, D. S., Xing, Y. and Guarente, L. (1995). Cloning of yeast *HAP5:* a novel subunit of a heterotrimeric complex required for CCAAT binding. *Genes Dev* **9:** 47-58.

Montgomery, D.L., Leung, D.W., Smith, M., Shalit, P., Faye, G. and Hall, B.D. (1980). Isolation and sequence of the gene for iso-2-cytochrome *c* in *Saccharomyces cerevisiae. Proc Natl Acad Sci* **77:** 541-545.

Mulder, W., Scholten, I.H.J.M., de Boer, R.W., Grivell, L.A. (1994). Sequence of the HAP3 transcription factor of *Kluyveromyces lactis* predicts the presence of a novel 4-cysteine zinc-finger motif. *Mol Gen Genet* **245:** 96-106.

Mulder, W., Scholten, I.H.J.M., Grivell, L.A. (1995). Carbon catabolite regulation of transcription of nuclear genes coding for mitochondrial proteins in the yeast *Kluyveromyces lactis. Curr Genet* **28:** 267-273.

NGuyen, C., Bolotin-Fukuhara, M., Wesolowski-Louvel, M., Fukuhara, H. (1995). The respiratory system of *Kluyveromyces lactis* escapes from HAP2 control. *Gene* **152:** 113-115

Oechsner, U., Hermann, H., Zollner, A., Haid, A., Bandlow, W. (1992). Expression of yeast cytochrome c1 is controlled at the transcriptional level by glucose, oxygen and haem. *Mol Gen Genet* **232:** 447-59.

Olesen, J., Hahn, S., Guarente, L. (1987). Yeast HAP2 and HAP3 activators both bind to the *CYC1* upstream activation site, UAS2, in an interdependent manner. *Cell* **51:** 953-61.

Olesen, J.T. and Guarente, L. (1990). The HAP2 subunit of yeast CCAAT transcriptional activator contains adjacent domains for subunit association and DNA recognition: model for the HAP2/3/4 complex. *Genes Dev* **4:** 1714-1729.

Olesen, J. T., Fikes, J. D. and Guarente, L. (1991). The *Schizosaccharomyces pombe* homolog of *Saccharomyces cerevisiae* HAP2 reveals selective and stringent conservation of the small core protein domain. *Mol Cell Biol* **11:** 611-619.

Papagiannopoulos, P., Andrianopoulos, A, Sharp, J., Davis, M.A. and Hynes, M.J. (1996a). The *hap* genes of *Aspergillus nidulans* play a role in the regulation of *amdS* expression. Genetics Society of Australia, 43rd Annual Conference.

Papagiannopoulos, P., Andrianopoulos, A., Sharp, J.A., Davis, M.A. and Hynes, M.J. (1996b). The *hapC* gene of *Aspergillus nidulans* is involved in the expression of CCAAT-containing promoters. *Mol Gen Genet* **251:** 412-421.

Parry, S. K. (1993). Construction and utilisation of a bidirectional reporter vector in the analysis of two *nod*-boxes in *Rhizobium loti.* Masters thesis, Massey University, Palmerston North, New Zealand.

Pfeifer, K., Prezant, T. and Guarente, L. (1987). Yeast HAP1 Activator Binds to Two Upstream Activation Sites of Different Sequence. *Cell* **49:** 19-27.

Pfeifer, K., Kim, K., Kogan, S. and Guarente, L. (1989). Functional dissection and sequence of yeast HAP1 activator. *Cell* **56:** 291-301

Pillar, T.M. and Bradshaw, R.E. (1991). Heat shock and stationary phase induce transcription of the *Saccharomyces cerevisiae* iso-2 cytochrome *c* gene. *Curr Genet* **20:** 185-188.

Prezant, T., Pfeifer, K. and Guarente, L. (1987). Organisation of the regulatory region of the yeast *CYC7* gene: Multiple factors are involved in regulation. *Mol Cell Biol* **7:** 3252-3259.

Punt. P. and van den Hondel, C. (1992). Analysis of transcription control sequences of fungal genes. In: (eds) Molecular Signals in Plant-Microbe Communications. CRC Press, Florida, Chapter 1, 29-48.

Punt, P.J., Dingemanse, M.A., Kuyvenhoven, A., Soede, R.D., Pouwels, P.H. and van den Hondel, C.A. (1990). Functional elements in the promoter region of the *Aspergillus nidulans gpdA* gene encoding glyceraldehyde-3-phosphate dehydrogenase. *Gene* **93:** 101-9.

Raeder, U. and Broda, P. (1985). Rapid preparation of DNA from filamentous fungi. *Lett Appl Microbiol* **1:** 17-20.

Raitt, D.R. (1992). The cytochrome *c* gene of *Aspergillus nidulans*. Ph.D. Thesis, Leicester Biocentre, United Kingdom.

Raitt, D.R., Bradshaw, R.E. and Pillar, T.M. (1994). Cloning and characterisation of the cytochrome *c* gene of *Aspergillus nidulans. Mol Gen Genet* **242:** 17-22.

Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989). *Molecular Cloning: A laboratory Manual.* New York: Cold Spring Harbour Laboratory Press.

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

Saunders, K. (1993). The regulation of the *Aspergillus nidulans* cytochrome *c* gene. Honours Thesis, Massey University, Palmerston North, New Zealand.

Scarpulla, R.C., Agne, K.M. and Wu, R. (1981). Isoaltion and structure of a rat cytochrome *c* gene. *J Biol Chem* **256:** 6480-6486.

Schneider, J.C. and Guarente, L. (1991). Regulation of the yeast *CYT1* gene encoding cytochrome c_1 by HAP1 and HAP2/3/4. *Mol Cell Biol* **11:** 4934-4942.

Shih, C.N. and Marth, E.H. (1974). Aflatoxin formation, lipid synthesis, and glucose metabolism by *Aspergillus parasititicus* during incubation with and without agitation. *Biochem Biophys Acta* **338:** 286-296.

Shizadegan, M.,Christie, P. and Seemann, J.R. (1991). An efficient method for isolation of RNA from tissue cultured plant cells. *Nucleic Acids Res* **19:** 6055.

Smith, M., Lueng, D.W., Gillam, S. and Astell, C.R. (1979). Sequence of the gene for iso-1-cytochrome *c* in *Saccharomyces cerevisiae. Cell* **16:** 753-761.

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

Streatfield, S.J., Toews, S. and Roberts, C.F. (1992). Functional analysis of the expression of the 3'-phosphoglycerate kinase *pgk* gene in *Aspergillus nidulans. Mol Gen Genet* **223:** 231-240.

Turcotte, B. and Guarente, L. (1992). HAP1 positive control mutants specific for one of two binding sites. *Genes Dev* **6:** 2001-2009.

Ushinsky, S.C. and Keng, T. (1994). A novel allele of HAP1 causes uninducible expression of HEM13 in *Saccharomyces cerevisiae. Genetics* **136:** 819-31

Van Gorcom, R.F.M., Punt, P.J., Pouwels, P.H., van den Hondel, C.A.M.J.J. (1986). A system for the analysis of expression signals in *Aspergillus. Mol Gen Genet* **223:** 17-23.

Van Heeswijk, R. and Hynes, M.J. (1991). The *amdR* product and a CCAAT-binding factor bind to adjacent, possibly overlapping DNA sequences in the promoter region of the *Aspergillus nidulans amdS* gene. *Nucleic Acids Res* **19:** 2655-2660.

Vashee, S., Xu, H., Johnston, S.A. and Kodadek, T. (1993). How do "Zn2 cys6" proteins distinguish between similar upstream activation sites? Comparison of the DNA-binding specificity of the GAL4 protein *in vitro* and *in vivo. J Biol Chem* **268:** 24699-24706.

Waldron,C. and Roberts, C.F. (1973). Cytoplasmic inheritance of a cold sensitive mutant in *Aspergillus nidulans. J Gen Microbiol* **78:** 379-381.

Xing, Y., Fikes, J.D. and Guarente, L. (1993). Mutations in yeast HAP2/3 define a hybrid CCAAT box binding domain. *EMBO J* **12:** 4647-4655.

Xing, Y., Zhang, S., Olesen, J. T., Rich, A. and Guarente, L. (1994). Subunit interaction in the CCAAT-binding heteromeric complex is mediated by a very short α-helix in HAP2. *Proc Natl Acad Sci* **91:** 3009-3013.

Yanisch-Perron, C., Vieira, J. and Messing, J. (1985). Improved M13 phage cloning vectors and host strains: Nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene* **33:** 103-119.

Zhang, L., Bermingham-McDonogh, O., Turcotte, B. and Guarente, L. (1993). Antibody-promoted dimerisation bypasses the regulation of DNA binding by the heme domain of the yeast transcriptional activator HAP1. *Proc Natl Acad Sci USA* **90:** 2851-2855.

Zhang, L. and Guarente, L. (1994a). The yeast activator HAP1-a GAL4 family menber-binds DNA in a directly repeated orientation. *Genes Dev* **8:** 2110-2119.

Zhang, L. and Guarente, L. (1994b). HAP1 is nuclear but is bound to a cellular factor in the absence of heme. *J Bio Chem* **269:** 14643-14647.

Zitomer, R.S. and Lowry, C.V. (1992). Regulation of gene expression by oxygen in *Saccharomyces cerevisiae. Microbiol Rev* **56:** 1-11.