FUNCTIONAL ANALYSIS OF FLUFFY,

A TRANSCRIPTIONAL REGULATOR FOR CONIDIAL DEVELOPMENT

IN *Neurospora crassa*

A Dissertation

by

PANAN RERNGSAMRAN

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

May 2004

Major Subject: Plant Pathology

FUNCTIONAL ANALYSIS OF FLUFFY,

A TRANSCRIPTIONAL REGULATOR FOR CONIDIAL DEVELOPMENT

IN *Neurospora crassa*

A Dissertation

by

PANAN RERNGSAMRAN

Submitted to Texas A&M University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Approved as to style and content by:

Daniel J. Ebbole (Chair of Committee)	
James L. Starr (Member)	Clint W. Magill (Member)
Susan S. Golden (Member)	Dennis C. Gross (Head of Department)

May 2004

Major Subject: Plant Pathology

ABSTRACT

Functional Analysis of Fluffy,

a Transcriptional Regulator for Conidial Development

in *Neurospora crassa*. (May 2004)

Panan Rerngsamran, B.S., KhonKaen University;

M.S., Chulalongkorn University

Chair of Advisory Committee: Dr. Daniel J. Ebbole

The *fluffy* gene of *Neurospora crassa* is required for asexual sporulation. It encodes an 88 kDa polypeptide containing a typical fungal Zn_2Cys_6 DNA binding motif. To identify the target genes on which FL may act, I sought to identify target sequences to which the FL protein binds. Several strategies were attempted to obtain purified FL protein. Purification was achieved by expressing the DNA binding domain of FL in *Escherichia coli* as a fusion with glutathione S-transferase followed by affinity purification using glutathione sepharose chromatography. DNA binding sites were selected by *in vitro* binding assays. Comparison of the sequences of selected clones suggested that FL binds to the motif 5'-CGG$(N)_9$CCG-3'. A potential binding site was found in the promoter region of the *eas* (*ccg-2*) gene, which encodes a fungal hydrophobin. *In vitro* competitive binding assays revealed a preferred binding site for FL in the *eas* promoter, 5'-CGGAAGTTTCCTCCG-3', which is located 1498 bp upstream of the *eas* translation initiation codon. *In vivo* experiments using a foreign DNA sequence tag confirmed that this sequence is a target site for FL regulation. Using

Saccharomyces cerevisiae as an experimental system, I demonstrated that the C-terminal portion of FL functions in transcriptional activation.

Microarray analysis was performed to study the role of *fl* in gene regulation on a large scale. mRNA levels in a *fl* mutant were compared to those in a strain overexpressing the *fl* gene. Experiments with cDNA microarray containing 13% of the total number of predicted *N. crassa* genes revealed 122 genes differentially expressed in response to overexpression of *fl*. Among these, *eas* displayed the greatest level of response. The cDNA microarray approach also revealed a number of genes that may be indirectly regulated by *fl* but may be involved in development. This information provides a foundation for further analysis of the role of *fl* in conidial development.

To my parents and relatives,

who, for years,

have awaited the completion of this part of my journey

ACKNOWLEDGMENTS

I would like to express my gratitude to many individuals who assisted me throughout the course of my doctoral research. First and foremost, I would like to thank my advisor, Dr. Daniel Ebbole, for the wonderful opportunity of participating in this exciting research endeavor, and for his scientific guidance, support, and patience.
I would also like to express my gratitude to my committee members: Dr. James Starr, Dr. Clint Magill, and Dr. Susan Golden, for their kind advice and valuable suggestions during the research and writing components of this dissertation.

I would like to acknowledge Dr. Michael Murphy and Dr. Sharon Doyle from the United States Department of Energy Joint Genome Institute for providing the SPR data.

I am especially grateful to all past and present members of Dr. Ebbole's lab, including Dr. Piotr Bobrowicz, Dr. Carlos Cortes, Dr. Guodong Lu, Dr. Hanno Wolf, Xin Xie, Cristina Filippe, Dan Li, Kiran Bhattarai, Yuan Jin, and Elena Kolomiets for their endless support and technical help. I would also like to acknowledge my fellow graduate students and the staff of the Department of Plant Pathology, whose assistance I have greatly appreciated.

Many thanks go to my wonderful friend Darlene Grzegorski for editing and proofreading this dissertation.

Lastly, I would like to thank the Royal Thai Government for providing financial support for my doctoral study and the National Science Foundation for funding this research project.

TABLE OF CONTENTS

Page

ABSTRACT……..…… iii

DEDICATION …….…… v

ACKNOWLEDGMENTS ……..…….……. vi

TABLE OF CONTENTS ……. vii

LIST OF FIGURES ……..……...……. ix

LIST OF TABLES ……..…….…… xi

CHAPTER

I INTRODUCTION……..…… 1

N. crassa…….…… 2
Conidiation in *N. crassa*.…… 4
FLUFFY, a Key Regulator for Conidial Development……...……... 7
Gene Regulation in Eukaryotes…….. 9
Zn_2Cys_6 Binuclear Zinc Finger Transcription Factors……... 12
Identifying the Target Gene of a Transcription Factor…….. 13
Objectives……... 16

II FLUFFY PROTEIN: OVEREXPRESSION AND PURIFICATION…… 19

Introduction…… 19
Materials and Methods……...…… 23
Results…… 35
Discussion……..…… 46

III FLUFFY, THE MAJOR REGULATOR OF CONIDIATION IN *Neurospora crassa*, DIRECTLY ACTIVATES A DEVELOPMENTALLY-REGULATED HYDROPHOBIN GENE….. 51

Introduction…… 51
Materials and Methods……...…… 55
Results…… 62
Discussion……..…… 87

CHAPTER Page

IV GENE EXPRESSION PROFILING OF *fl* MUTANT AND *fl* OVEREXPRESSION STRAINS…… 93

Introduction…… 93
Materials and Methods...…… 98
Results and Discussion…...…… 100

V SUMMARY……...…...……. 131

Future Directions…… 135

REFERENCES……...…… 136

VITA……..…… 154

LIST OF FIGURES

FIGURE		Page
1	Developmental timeline of *N. crassa*…….	6
2	Diagram of FL protein……	8
3	Construction of pRP72……...	28
4	Construction of pRP79……...	30
5	Morphology of *N. crassa* and expression of *fl* under *cpc-1* and *RP27* promoter……..	37
6	Northern blot analysis of *fl* in *S. cerevisiae* under *GAL1* promoter…...	39
7	Comparison of *fl* transcript in *N. crassa* and *S. cerevisiae*…….	40
8	Northern blot analysis of *fl* in *N. crassa* and *S. cerevisiae* with and without PEST sequences……	42
9	Comparison of expression level of FL_{1-371H} in *E. coli* under *trc* (lanes 2 and 3) and T7 (lanes 4 and 5) promoters……..……..	44
10	Extraction and purification of FL_{1-371H} from *E. coli* bearing pRP45 using Ni-NTA resin by native (A) and denaturing (B) condition……...	45
11	Expression and purification of GST::FL_{1-199}……..	47
12	*In vitro* binding of GST::FL_{1-199}…….	65
13	Localization of FL binding sites in the *eas* promoter…….	67
14	*In vitro* binding competition assay defining FL binding site preferences……..	68
15	Measurements of the binding of GST::FL_{1-199} to the consensus sequence found in the *eas* promoter (5'-TTCTCCTGCCCGGAAGTT TCCTCCGGCCATACTCAT-3') using surface plasmon resonance…	70
16	Northern blot analysis of *eas* and *eas-mpg1* tagged genes…….	72

FIGURE Page

17 Localization of a transcriptional activation domain of FL……. 74

18 Northern blot analysis of genes containing putative binding sites for FL…… 86

LIST OF TABLES

TABLE		Page
1	Plasmids used in this study……...	25
2	Oligonucleotide PCR primers used in this study…….…….	26
3	Oligonucleotides used in the competition assay……...	59
4	Identification of consensus sequence for FL binding……...	64
5	Potential FL binding sites in the *N. crassa* genome……..	76
6	Genes at least 2 fold differentially-expressed in *fl* mutant and FLOE strains…….	103
7	Gene expression patterns identified from northern blot analysis…..	111

CHAPTER I

INTRODUCTION

Filamentous fungi are valuable producers of industrial enzymes, chemicals, and pharmaceutical products (Bennett, 1998). Many of them play an important role in the manufacturing of common foodstuffs and beverages, such as bread, cheese, and beer. Some are valuable producers of pharmaceutical products, such as penicillin, cephalosporin, and cyclosporin. However, some filamentous fungi are pathogens of plants or humans, causing devastating diseases that have both social and economic impact. For example, up to 20% of the world rice crop are lost each year to rice blast, which is caused by the fungal pathogen *Magnaporthe grisea* (Eubanks, 2002). Though the majority of fungi are not harmful to humans, some fungi can cause severe disease, especially in immunocompromised patients. For example, *Coccidioides immitis*, the causal agent of human coccidioidomycosis, causes a mild respiratory infection, that in rare cases may precipitate a chronic pulmonary condition (Standaert *et al.*, 1995).

Many filamentous fungi disperse through the environment by means of conidia, which are produced in large numbers to facilitate dispersal to new hosts, leading to

---

This dissertation follows the style of *Molecular Microbiology*.

infection and disease. Understanding the process of conidial morphogenesis and mechanisms for regulation of conidial development in fungi will assist in promoting the development of strategies that may prevent or reduce the distribution of disease agents. The study of conidial development has been investigated most extensively in two filamentous fungi. *Neurospora crassa* and *Aspergillus nidulans* have been used as model genetic organisms for the study of developmental gene regulation during conidiation (Ebbole, 1996). Bioinformatic research suggests that the mechanism and process for conidial morphogenesis in these two fungi differs greatly (Galagan *et al.*, 2003). Although conidiation in *A. nidulans* is morphologically more complex than in *N. crassa*, it has been examined more extensively. Several key regulators for conidiophore development in *A. nidulans* have been cloned and characterized (Adams *et al.*, 1998). Conversely, despite relatively simple conidial development, conidiation in *N. crassa* is not well understood. One advantage for the study of conidiation in *N. crassa* is the wealth of mutants that have been isolated with conidiation defects and the availability of EST and genomic sequence databases. Utilizing these resources, this dissertation examines the process of conidiation in the model filamentous fungus *N. crassa*.

N. crassa

From 1843 to 1927 *N. crassa* was known as *Monilinia sitophila*. *N. crassa* is an orange mold commonly found on bread, carbohydrate-rich food, residues of sugar-cane processing, and the remains of burnt vegetation. In 1927, C. Shear and B. Dodge discovered the sexual fruiting bodies of this fungus and placed it in a new genus,

Neurospora (Davis and Perkins, 2002). C. Lindegren is considered the "father" of *Neurospora* genetics and later the "father" of yeast genetics in America (Hall and Linder, 1993). He produced the first genetic linkage maps of *N. crassa* and helped to establish *N. crassa* as a model system for genetic analysis (Lindegren, 1933). In 1941, G. Beadle and E. Tatum published the famous one-gene, one-enzyme hypothesis based on their work with *N. crassa* (Beadle and Tatum, 1941) for which they were subsequently awarded the Nobel Prize in Physiology and Medicine in 1958 (Davis, 2000). In 1945, Barbara McClintock, winner of the Nobel Prize for the discovery and study of transposons in corn, observed and described the meiotic cytology in *N. crassa* (McClintock, 1945). Because *N. crassa* is a heterothallic haploid fungus that can be grown in a simple medium, and provides a simple system for analysis of Mendelian genetic segregation of traits, *N. crassa* has long been used for genetic analyses (Davis and Perkins, 2002). The genome sequence of *N. crassa* was completed in 2002 (http://www.broad.mit.edu/annotation/fungi/neurospora/). Analyses revealed that the total chromosomal DNA length is approximately 43 Mb, and is predicted to encode 10,082 open reading frames (ORFs) with an average gene length of about 1.67 kb (Galagan *et al.*, 2003).

N. crassa grows vegetatively with haploid, multinucleate, filamentous hyphae that intertwine and anastomose to form a mycelium. It has three distinct sporulation pathways, leading to the formation of ascospores, macroconidia, and microconidia. Ascospores are the product of sexual reproduction, whereas macroconidia and microconidia are derived from asexual reproduction (Springer, 1993). Microconidia are

uninucleate and are smaller and less viable than macroconidia. The latter are the most efficient means for rapid dispersal of the fungus (Springer and Yanofsky, 1992). In this study I will focus on the development of macroconidia (here after, conidia).

CONIDIATION IN *N. crassa*

Multinucleate macroconidia result from repeated apical budding of aerial hyphae. Production of macroconidia is influenced by numerous factors, including, desiccation, light, carbon or nitrogen deprivation, carbon dioxide levels, and the circadian rhythm (Ebbole, 1996; Springer, 1993; Turian and Bianchi, 1972). Synchronous induction of the macroconidiation process occurs following exposure of the mycelium to air. Hyphal growth is directed away from the substrate, resulting in the formation of a mass of aerial conidiophores. Two hours later, apical elongation ceases and growth switches to a process of repeated apical budding (Springer and Yanofsky, 1989). The first several rounds of budding produce proconidial chains with interconidial diameters nearly as large as the diameter of the parent hypha. These chains are described as minor constriction chains. It has been reported that minor constriction chains can possibly revert to filamentous growth (Springer, 1993). Six hours after induction, budding growth becomes more pronounced giving rise to major constriction chains. At this point the developing conidiophores are unable to revert to either minor constriction chain growth or hyphal elongation. Therefore, the transition from minor to major constriction chain growth appears to be a committed step for differentiation of mature conidia (Springer, 1993). At about 12 hours after the induction or 4 hours after the major constriction

chains develop, budding growth stops and nuclei migrate into the proconidial chains. Subsequently, double crosswalls are laid down between each proconidium. Four hours later, the double crosswalls separate and conidia are easily released and dispersed by wind (Ebbole, 1996; Lee and Ebbole, 1998b; Springer and Yanofsky, 1989; Springer, 1993). Fig. 1 illustrates the morphological changes that occur during development.

A number of genes that are preferentially expressed during conidiation in *N. crassa* have been isolated by differential screening methods (Berlin and Yanofsky, 1985). These genes are known as *con*idiation-specific genes (*con* genes). Five of these genes *con-6*, *con-8*, *con-10*, *con-11*, and *con-13*, have been studied in detail (Corrochano *et al.*, 1995; Hager and Yanofsky, 1990; Roberts *et al.*, 1988; Roberts and Yanofsky, 1989; White and Yanofsky, 1993). Briefly, *con-6* encodes a 93 amino acid polypeptide thought to be involved in spore survival (White and Yanofsky, 1993). *con-8* and *con-10* encode 176 and 86 amino acid polypeptides, respectively (Corrochano *et al.*, 1995; Roberts and Yanofsky, 1989). *con-13* encodes a 340 amino acid residue polypeptide and its open reading frame is adjacent to that of *con-10* (Roberts and Yanofsky, 1989). Mutation of these genes has demonstrated that they are not essential for development (Berlin and Yanofsky, 1985). Another conidiation-specific gene that has been characterized in more detail is the *eas* (*easily wettable*) gene, which is also known as *clock-controlled gene-2* (*ccg-2*) and *blue-light inducible-7* (*bli-7*) (Bell-Pedersen *et al.*, 1992; Kaldenhoff and Russo, 1993). Previous studies have shown that the surfaces of conidia are coated by hydrophobin protein which is a product of the *eas* gene (Lauter *et al.*, 1992). Hydrophobins make the conidia hydrophobic and more easily dispersed by air

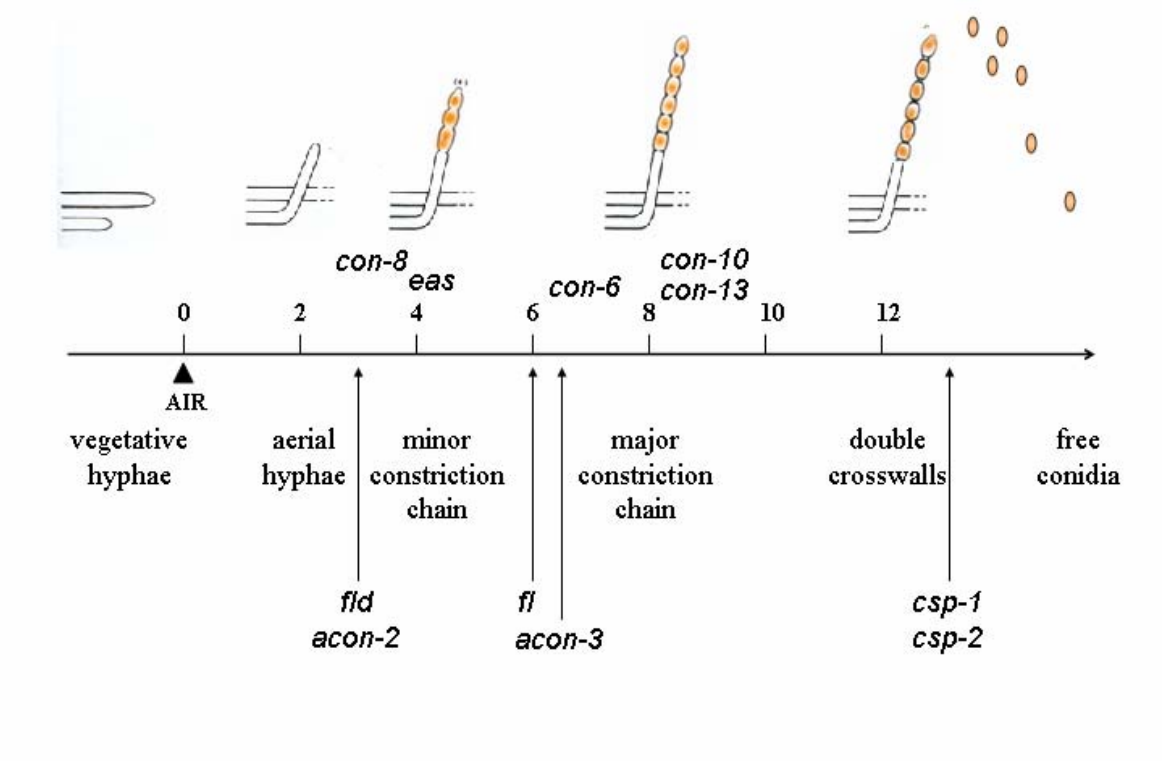


Fig. 1. Developmental timeline of *N. crassa*. Timeline (hr) showing conidiophore development and expression of conidiation specific genes after vegetative hyphae of *N. crassa* are induced to develop by exposure to air. Mutants that block development at the different morphological stages of conidiation are also shown. (Modified from Ebbole, 1996; and Turian and Bianchi, 1972).

(Ebbole, 1997). The products of these *conidiation*-specific genes have been found to be highly expressed during conidiation, with expression decreasing shortly after maturation of the conidia (Berlin and Yanofsky, 1985; Lauter *et al.*, 1992). The relative timing of *con* and *eas* gene expression is shown in Fig. 1.

Several morphological mutants of *N. crassa* were found to have effects on conidial development. Each mutant blocks development at a distinct stage (Matsuyama *et al.*, 1974; Springer and Yanofsky, 1989). As shown in Fig.1, *aconidiate-2* (*acon-2*) and *fluffyoid* (*fld*) mutants do not form minor constriction chains. They produce normal aerial hyphae but never bud to form minor constriction chains. *fluffy* (*fl*) and *acon-3* mutants are able to form minor but not major constriction chains. Two *conidial separation* (*csp-1*and *csp-2*) mutants are able to form both minor and major constriction chains and also double crosswall layers. However, the *csp* mutations block the maturation of double crosswalls, and thus, the conidia cannot be released (Ebbole, 1996; Madi *et al.*, 1994; Matsuyama *et al.*, 1974; Springer and Yanofsky, 1989). Previous studies have demonstrated that *fl* acts downstream of *acon-2* and upstream of *acon-3*. Furthermore, the product of *acon-2* is required for *fl* expression (Bailey-Shrode and Ebbole, in press; Correa and Bell-Pedersen, 2002). Of the genes represented by these mutants, the *fl* gene is the only gene that has been cloned (Bailey and Ebbole, 1998).

FLUFFY, A KEY REGULATOR FOR CONIDIAL DEVELOPMENT

Sequence analysis revealed that *fl* encodes an 88 kDa polypeptide, 792 amino acid residues in length. The Fluffy protein (FL) contains a Zn_2Cys_6 binuclear zinc cluster

domain that belongs to the Gal4p family. FL contains several motifs commonly found in this type of transcription factor family; a putative zinc finger DNA binding domain located at amino acid position 11-37, a coiled-coil motif amino acid position 45-64, and a middle homology region at amino acid position 325-358 (Fig. 2). The sequences in these three regions show high similarity to NIT4, a pathway specific regulator for nitrate assimilation in *N. crassa*; Cha4p, a positive regulator for serine/threonine utilization in *Saccharomyces cerevisiae*; and NCU09205.1, a predicted open reading frame from the *Neurospora* genome sequence (Bailey and Ebbole, 1998; Galagan *et al.*, 2003). However the amino acid sequences that lie outside these regions do not share homology with any other proteins in current databases.

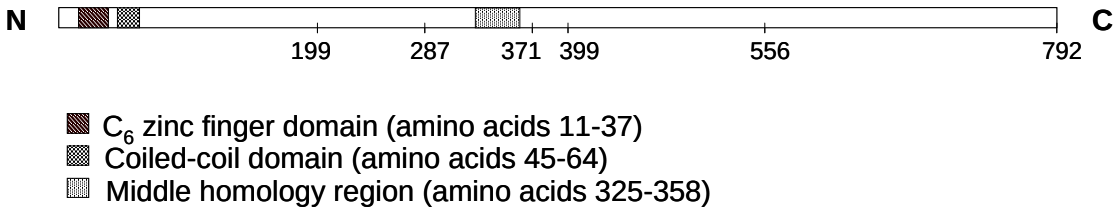


Fig. 2. Diagram of FL protein. The diagram illustrates important domains, and the relative amino acid positions. N, amino terminus; C, carboxy terminus.

Developmental induction of the *fl* transcript is tissue-specific with the highest expression in aerial hyphae (Bailey-Shrode and Ebbole, in press). It is expressed at a very low basal level in vegetative hyphae and substantially expressed at about 6 hours during the time that major constriction chains are formed (Bailey and Ebbole, 1998). A recent study also found that after developmental induction, *fl* also shows a weak

induction at about 15 min, and reaches a peak at 30 min. Following this peak in expression, the *fl* transcript level decreases, and subsequently reaches a strong second peak again at 6 hours (Correa and Bell-Pedersen, 2002). In later stages of development, transcript levels decrease to the pre-induction level. Overexpression of *fl* under the control of a heterologous promoter is sufficient to induce conidiation in a wild-type background (Bailey-Shrode and Ebbole, in press). By the time that *fl* is highly expressed, some *conidiation*-specific genes such as *con-6, con-10,* and *eas* are induced (Bailey and Ebbole, 1998; Bailey-Shrode and Ebbole, in press). The expression of *con*idiation-specific gene products has also been studied in different genetic backgrounds. It has been found that *fl* mutants have a block in the expression of *eas* (Bailey-Shrode and Ebbole, in press; Lauter *et al.*, 1992). In addition, it has been shown that elevated expression of *fl* induces the expression of the *eas* gene transcript (Bailey-Shrode and Ebbole, in press). Based on these findings, it is likely that FL may function as a transcriptional regulator of the gene(s) involved in conidial development in *N. crassa*, and in particular, may directly regulate the expression of the *eas* gene.

GENE REGULATION IN EUKARYOTES

Gene regulation in eukaryotes is a multi-stage process that occurs at various steps in a biochemical pathway and affects diverse components of the molecular machinery involved in cellular processes. Regulation may occur at the level of transcription, RNA processing, mRNA longevity, translation, and at post-translation levels. At the level of transcription, transcription factors (TF) play a major regulatory role. In general, the

expression of eukaryotic genes is initiated by the binding of basal transcription factors to promoter regions upstream of the genes. Two motifs in the promoter region are of particular importance. First, the TATA box, located approximately 25 bp upstream of the transcription start site (TSS), is a primary site for binding of general transcription factors. Second, a CAAT box, located at about -80 bp relative to the TSS, plays a strong role in determining the efficiency of the promoter (Kim and Roeder, 1994). The first step in the cascade of events that leads to gene transcription begins when TFIID binds to the TATA box. Subsequently, TFIIA, TFIIB and TFIIF join TFIID to form a complex spanning the region from -50 bp to +15 bp. TFIIF consists of two subunits, one with helicase activity which unwinds the DNA and a second subunit that strongly binds to RNA polymerase II. Once the RNA polymerase II has been recruited to the TSS, two additional transcription factors, TFIIE and TFIIH, join the complex and the transcription process is initiated (Reese, 2003; Warren, 2002).

Various cellular events, which occur prior to transcription initiation determine when and where certain genes will be expressed. Many genes respond to environmental and developmental cues, and are often only expressed in specific cells or tissues, in which case specific transcription factors are involved. A large number of these specific transcription factors have been identified. Despite significant variation in their molecular structures, transcription factors share a common characteristic in that they have two functional domains, one of which is a DNA-binding domain that recognizes the promoter of the target gene in a sequence specific manner, whereas the other is an activation domain that interacts with other proteins, which increases the rate of

transcription (Gadgil *et al.*, 2001). It has been shown that the activation domain of many transcription factors interacts with a subunit of TFIID, thereby promoting the assembly of the basal transcription factor machinery. Assembly in this manner may stabilize the basal transcription complex, which in turn speeds the process of transcription (Lieberman *et al.*, 1997; Thut *et al.*, 1995).

Specific transcription factors have been classified based on the structures of their DNA-binding domain. The main groups are helix-turn-helix, basic helix-loop-helix, leucine zipper, and several zinc finger motifs, including the Zn_2Cys_6 binuclear zinc finger proteins (Robert *et al.*, 2003). The first discovered DNA binding motif was the helix-turn-helix. It consists of two α helices connected to each other by a short turn of amino acids, where the C-terminal helix is responsible for DNA recognition (Rosinski and Atchley, 1999). The basic helix-loop-helix motif is composed of two α helices linked by a random-coil loop. A basic region located at the amino terminus of the first helix facilitates DNA binding (Borneman *et al.*, 2002). The leucine zipper or bZIP proteins have been identified in diverse groups of fungi, plants, and animals. One domain contains a heptad repeat of leucines called a leucine zipper region which mediates dimerization, whereas a second domain containing basic amino acid residues binds to DNA (Gadgil *et al.*, 2001). The zinc finger transcription group, which requires zinc ions in order to bind to the target DNA, is divided into three subgroups based on the DNA binding motifs. These motifs are Cys-X_{2-4}-Cys-X_{12}-His-X_{3-5}-His, Cys-X_2-Cys-X_{17}-Cys-X_2-Cys, and Cys-X_2-Cys-X_4, His-X_4-Cys, respectively (Robert *et al.*, 2003). The Zn_2Cys_6 binuclear zinc finger protein, which has consensus motif Cys-X_2-Cys-X_6-Cys-

X_{5-9}-Cys-X_2-Cys-X_{6-8}-Cys (Mamane *et al.*, 1998) is comprised of two α helices coordinated with two zinc ions to form a binuclear cluster structure (Todd and Andrianopoulos, 1997).

ZN_2CYS_6 BINUCLEAR ZINC FINGER TRANSCRIPTION FACTORS

The first identified Zn_2Cys_6 binuclear zinc finger motif was Gal4p of *S. cerevisiae*. This type of transcription factor has been found, thus far, only in the higher fungi (ascomycetes and basidiomycetes) (Todd and Andrianopoulos, 1997). Several studies of proteins with the Zn_2Cys_6 binuclear zinc cluster domain have shown that the DNA binding domain is located at the N-terminus, whereas the transcriptional activation region is located in the C-terminal half of the protein (Schjerling and Holmberg, 1996; Todd and Andrianopoulos, 1997). Besides DNA binding and activation domains, proteins in this family share two additional common regions. The first region, located adjacent to the C-terminus of the DNA binding domain, is called the coiled-coil region. Evidence indicates that this region mediates dimerization of the protein (Todd and Andrianopoulos, 1997). The function of the second region, called a middle homology region, has not been clearly defined. Removal of this region, in some cases, results in constitutive activation of the protein, however in other instances, this causes regulation to be abolished (Schjerling and Holmberg, 1996).

Most transcription factors that contain a Zn_2Cys_6 binuclear zinc cluster DNA binding domain in the Gal4p family recognize a CGG motif (Reece and Ptashne, 1993; Schjerling and Holmberg, 1996; Todd and Andrianopoulos, 1997). This CGG motif

recognized by the proteins has been classified into three different patterns; palindromic repeats (CGG_CCG; also called inverted repeats), direct repeats (CGG_CGG), or everted repeats (CCG_CGG) as found in Ppr1p, Hap1p, and Pdr3p respectively (Mamane *et al.*, 1998). In addition to the orientation of the CGG repeat, each individual protein in this family displays variation in the number of nucleotides between the two CGG repeats. For example, Gal4p and Lac9p bind to $CGGN_{11}CCG$, Put3p binds to $CGGN_{10}CCG$, and Ppr1p binds to $CGGN_{6}CCG$. In addition to the variation in number of nucleotides, sequence variation has been observed in the spacer region (Todd and Andrianopoulos, 1997).

IDENTIFYING THE TARGET GENE OF A TRANSCRIPTION FACTOR

In order to identify the target DNA sequence that is bound by a particular transcription factor, the transcription factor must be purified. Several methods have been successfully used to purify transcription factors such as ammonium sulfate precipitation, size exclusion chromatography, ion-exchange chromatography, and affinity chromatography (Gadgil *et al.*, 2001). There is no particular method that appears to provide a better yield than the others. The success of any method depends on the characteristics of each individual protein, such as stability and hydrophobicity. In some cases, a single step is sufficient to isolate pure protein, and in others, more than one method is required in order to obtain an adequate purity and retain activity (Robert *et al.*, 2003). Although nuclear extracts may be used as a source of protein to search for the target binding site, overexpression of recombinant proteins has facilitated research.

Several expression systems such as insect cells, yeast, *Escherichia coli*, and *in vitro* transcription/translation systems are available for synthesis of recombinant proteins (Kang *et al.*, 1993; Sheibani, 1999; Strauss *et al.*, 1998; Sudbery, 1996). In addition, the modification of the protein sequence with epitope tags is a powerful tool to facilitate convenient purification and detection of transcription factor proteins. Several types of epitope have been developed and used successfully for this purpose. Short peptides such as S-, FLAG-, and hexahistidine-tags (Thorn *et al.*, 2000) are commonly used and generally do not affect the function of the protein. Fusion of the target protein to larger proteins such as glutathione-S-transferase (GST), green fluorescent protein (GFP), or maltose binding (MAL) protein is also regularly used and has been found to improve the solubility of some target proteins (Terpe, 2003).

A common process used to search for the target binding site is based on a selection from a pool of random DNA or oligonucleotide, followed by enrichment of the target sequences by PCR. Several PCR-based methods have been developed such as SAAB (Selected And Amplified Binding site imprint assay), SELEX (Systematic Evolution of Ligands by EXponential enrichment), REPSA (Restriction Endonuclease Protection Selection and Amplification), CASTing (Cyclical Amplification and Selection of Targets), TDA (Target Detection Assay), MUST (Multiplex selection technique), and SELEX-SAGE (Systematic Evolution of Ligands by EXponential enrichment and Serail Analysis of Gene Expression) (Hardenbol and Van Dyke, 1996; Irvine et al., 1991; Nallur *et al.*, 1996; Roulet *et al.*, 2002; Thiesen and Bach, 1990; Wright *et al.*, 1991; Zhang *et al.*, 2003). A pool of randomized DNA is mixed with a purified transcription

factor or nuclear extract, and the DNA-protein complex is isolated from unbound DNA by one of several procedures, such as polyacrylamide gel electrophoresis (Wang *et al.*, 1998), immunoprecipitaion (Pollock and Treisman, 1990), nitrocellulose membrane filtration (Sakai *et al.*, 1998), affinity protein binding (Pierrou *et al.*, 1995), and surface plasmon resonance (Hao *et al.*, 2003). Many target genes of known transcription factors have also been identified by using recently developed technologies such as ChIP-chip array (Chromatin ImmunoPrecipitation-chip), and cDNA microarrays (Hellauer *et al.*, 2002; Iyer *et al.*, 2001; Kannan *et al.*, 2001; Lieb *et al.*, 2001; Ren *et al.*, 2000; Shalev *et al.*, 2002; Wells *et al.*, 2002; Wyrick *et al.*, 2001).

Once a potential candidate gene is identified, the relationship between the *cis*-acting sequence and the *trans*-acting element to the biological process must be measured. This characterization can be performed both *in vitro* and *in vivo*. *In vitro* analysis by electrophoretic mobility shift assay (EMSA) is based on the principle that a protein-DNA complex migrates through a native gel more slowly than the free radiolabeled DNA probe (Lane *et al.*, 1992). The migration of the complex and the free probe are detected by autoradiography or phosphoimager analysis. *In vitro* analysis by DNaseI footprinting is based on the principle that DNaseI can digest only naked but not protected DNA. The DNA is first end-labeled and incubated with the transcription factor. The reaction is then treated with DNaseI, resolved in polyacrylamide gel electrophoresis, and detected by autoradiograph or phosphoimager (Carey and Smale, 2000). The basic *in vivo* assay employs a translation fusion approach, which places a reporter gene under the control of the promoter sequence of interest, with or without the

cis-acting element. Many reporter genes such as luciferase, β-galactosidase, and chloramphenicol acetyltransferase (CAT) have been developed and used in this type of assay (Lewis *et al.*, 1998). This approach has been successfully used to characterize the *con-10* promoter (Corrochano *et al.*, 1995; Lee and Ebbole, 1998a) and the role of the AFLR transcription factor (Fernandes *et al.*, 1998). Another strategy is to use a foreign DNA sequence tag to examine regulation at the transcriptional level (Kaldenhoff and Russo, 1993). In this approach, the foreign DNA is inserted into the target gene and northern analysis with radiolabeled foreign DNA as probe is used to detect the level of regulation of the transgene.

OBJECTIVES

Evidence suggests that FL may function as a transcriptional regulator of the gene(s) involved in conidial development in *N. crassa* and, particularly, *eas* is possibly a direct target of FL (Bailey and Ebbole, 1998; Bailey-Shrode and Ebbole, in press; Correa and Bell-Pedersen, 2002; Lauter *et al.*, 1992). Therefore, my goal is to understand the role of *fl* in conidial development in *N. crassa*. The overall objectives of my research were to:

i) Determine whether FL is a transcription factor that contains both DNA binding and activation domains.

ii) Explore approaches to identify and characterize target genes regulated by FL.

To accomplish my first objective, I used oligonucleotide random-site selection together with a yeast-one hybrid system. I determined that FL binds to the DNA motif

5’-CGG$(N)_9$CCG-3’ in a sequence specific manner. The consensus binding site of FL was identified as 5’-CGGA[A/C/G]NN[A/C/T]NNC[C/T]CCG-3’. A segment of FL required for transcriptional activation is located between amino acid residues 400 to 556.

To accomplish my second objective, I used two approaches to identify target genes. First, I used computer software to search available consensus sequences in the *N. crassa* database. I identified 123 ORFs containing the FL consensus sequence, which were predicted to encode proteins that are involved in a wide range of functions. This finding suggested that *fl* may regulate a much broader range of genes beyond those involved in conidiation, or that there may be a large number of promoters containing FL binding sites that are not subject to regulation by FL. One of the ORFs identified in the search was the *eas* gene. A consensus binding site of FL, 5’-CGGAAGTTTCCTCCG-3’, was located 1498 bp upstream of the *eas* translation initiation codon. *In vivo* analysis confirmed that this sequence is required for FL regulation, providing additional evidence that *eas* is a direct target gene of FL. Of an additional set of 10 genes examined, expression of only four could be detected by northern blot analysis and none of these appeared to be developmentally regulated.

The second approach used to find target gene(s) of FL involved microarray analysis of approximately 1,300 genes, which is about one-seventh of *N. crassa*’s genes (Lewis *et al.*, 2002). Microarray assays also verified that *eas* is a target gene of FL. Comparison of the profiles derived from a *fl* mutant and a *fl* overexpression strain revealed that 51 unique ORFs were induced more than two-fold in the *fl* overexpression strain. These genes included *fl*, *eas*, *ccg-1*, *con-6*, and *con-10*. Conversely, there were 71

unique ORFs that were down-regulated more than two-fold in the *fl* overexpression strain. Microarray analysis revealed that at the 18 hr time point, the *fl* mutant strain favored glycolysis, whereas *fl* overexpression strain favored gluconeogenesis. Among the genes with regulated expression in response to overexpression of *fl*, *eas* was the only ORF identified to contain the consensus binding site of FL within 2000 nucleotides upstream of the predicted translation initiation codon.

I conclude that knowledge of the DNA sequence to which FL binds provides only a limited ability to predict which genes are regulated by FL. It appears that in addition to FL binding sites, other elements in promoters are likely important for conidiation-specific regulation.

CHAPTER II

FLUFFY PROTEIN: OVEREXPRESSION AND PURIFICATION

INTRODUCTION

Asexual sporulation of *Neurospora crassa* has been used as a model to study cell differentiation. The synchronous process of conidial development occurs after induction by environmental signals; such as nutrient deprivation, blue light, and desiccation; or internal signal such as the circadian clock (Ebbole, 1996). Subsequent to the induction aerial hyphae are produced, follow by minor and major constriction chain formation, and then cross-walls are laid down at each constriction resulting in a chain of conidia (Springer, 1993). Several mutants such as *aconidiate-2* (*acon-2*), *acon-3, conidial separation* (*csp-1*), *csp-2*, and *fluffy* (*fl*) have been found to block the development (Ebbole, 1996; Matsuyama *et al.*, 1974; Springer, 1993). Among these, only the *fl* gene has been cloned (Bailey and Ebbole, 1998).

The *fl* gene was identified as a necessary component of conidial development in *N. crassa*. Recent cloning of the gene has enabled examination the relationship of its expression to the change of conidiophore morphogenesis (Bailey and Ebbole, 1998). The expression level of the *fl* transcript has been found to be tissue-specific. It is highly expressed in aerial hyphae at the time that major constriction chains are formed during development. As the Fluffy polypeptide (FL) contains a Zn_2Cys_6 binuclear zinc finger DNA binding domain similar to other members of Gal4p family of transcriptional

factors, it has been postulated to be a transcriptional regulator for conidial morphogenesis (Bailey and Ebbole, 1998; Bailey-Shrode and Ebbole, in press).

A general feature of transcription factors is that they contain two functional domains, a DNA binding domain and an activation domain (Carey and Smale, 2000). The DNA binding domain binds to a specific sequence within the promoter of the target gene. The activation domain recruits the transcriptional apparatus to the promoter, thereby activating the expression of the target genes (Carey and Smale, 2000).
A fundamental step of understanding the role of a particular transcription factor is the identification of the direct target genes. One of the methods used is to identify the specific sequence bound by the transcription factor (Blackwell, 1995; Zhang *et al.*, 2003). In order to perform an essential assay to identify the target binding site, a sufficient quantity and quality of the putative transcription factor is required.

The source of the protein may be direct, for example, a nuclear protein fraction obtained from the host organism. Alternatively a purified recombinant protein may be used (Carey and Smale, 2000). Although ideal, isolating the target protein from a natural source may be challenging. Most transcription factors are expressed at a very low level. Furthermore, the transcription factors, especially developmentally regulated factors, are expressed only at a certain time, and in a specific tissue (Carey and Smale, 2000), thereby make the extraction of sufficient protein for further analysis difficult.

To overcome these difficulties, overexpression of the recombinant protein is widely used. Overexpressions of the proteins using *in vitro* transcription/translation, and *in vivo* transcription/translation in heterologous hosts have been used successfully. The

in vitro transcription and translation protocol typically provides yields of approximately 1%, whereas the *in vivo* expression using yeast, insect cells, or *Escherichia coli* system can provide yields of approximately 1-5%, 5-20%, and 10-30%, respectively (Carey and Smale, 2000). Both *in vivo* and *in vitro* approaches have been employed to overexpress *N. crassa* protein. Several heterologous promoters such as constitutive promoters like a *trpC* promoter from *Aspergillus nidulans* and a *cross pathway control-1* (*cpc-1*) promoter from *N. crassa*, or inducible promoters like *qa-2* and *bli-4* promoter for quinic acid and blue-light induction, respectively, have been used to increase the level of *N. crassa* proteins *in vivo* (Bailey-Shrode and Ebbole, in press; Campbell *et al.*, 1994; Mautino *et al.*, 1996; Pietschmann *et al.*, 1991). Heterologous hosts such as *Saccharomyces cerevisiae* have been used successfully to produce *N. crassa* protein with full activity (Mahanty *et al.*, 1994). Using purified protein from insect cell culture and *in vitro* expression systems, functional analysis of QA-1F transcription factor was performed (Baum *et al.*, 1987). Several transcription factors have been extensively studied by expressing them in *E. coli*, such as NIT4 which plays a role in nitrate utilization in *N. crassa* (Fu *et al.*, 1995), and AFLR which is a component of the aflatoxin biosynthesis pathway in *A. nidulans* (Ehrlich *et al.*, 1999). Although production of recombinant proteins in *E. coli* provides the highest yield compared to other systems, two important factors limit the expression of eukaryotic proteins in *E. coli*. First, *E. coli* lacks the ability to carry out complex post-translational modifications which is common for eukaryotic protein synthesis. Second, the heterologous proteins synthesized by *E. coli* may be misfolded or aggregated and localized to inclusion bodies in an insoluble

form (Carey and Smale, 2000). However, it has been reported that the use of fusion proteins or epitope tags is not only useful for purification and detection but also is a generally useful approach to improve both the level of expression and solubility (Higgins and Hames, 1999).

Other factors that may interfere with the expression of a protein are its inherent solubility and stability. Proteins with high hydrophobicity such as membrane proteins are difficult to dissolve in aqueous solution unless suitable detergents are added (Higgins and Hames, 1999). Some proteins are very short-lived, and thus highly unstable, due to their rapid turnover rate. These proteins are often found to contain a motif enriched in proline (P), glutamate (E), serine (S), and threonine (T), known as the PEST sequence, which targets the protein for rapid degradation (Rogers *et al.*, 1986). Stability of the proteins containing the PEST sequence may be enhanced by removing or mutating the PEST motifs (Gorl *et al.*, 2001; Medintz *et al.*, 2000).

The objective of this work was to generate a sufficient quantity of functional FL protein for subsequent use in identifying its target binding site. The identification of the specific binding site of FL will aid in identifying its target genes and defining the role of *fl* as a transcription factor during development. Several approaches were used to obtain purified FL protein. I experimented with various expression systems, promoter types, and epitope-tagged types. Furthermore, I examined the effect of putative PEST sequences on FL protein accumulation in *N. crassa*. Here, I present my research effort to overexpress FL in *N. crassa* and *S. cerevisiae*. I could readily detect *fl* expression at the transcriptional level, but was unable to detect FL protein. Thus, despite expression from

a moderately strong promoter, FL protein must be present at low levels and I concluded that FL protein is likely rapidly degraded *in vivo*. Removal of two PEST motifs from FL did not enhance the stability of FL sufficiently to allow accumulating to detectable levels. This suggested that the PEST motifs of FL do not play a major role in determining the stability of the protein. Attempts to express and purify the DNA binding domain segment of FL in *E. coli* with a hexahistidine tag resulted in non-specific copurification, and protein aggregation. Finally, application of the GST fusion system resulted in the successful purification of FL protein under native conditions, which permitted further biochemical studies described in Chapter III.

MATERIALS AND METHODS

Strains, media, and transformation. The *N. crassa* wild-type strain 74OR23-1V*A* (FGSC 2489), and *fl a* mutant (FGSC 46) were obtained from the Fungal Genetics Stock Center (FGSC; Department of Microbiology, University of Kansas Medical Center). General growth medium was Vogel's minimal medium supplemented with 2% sucrose (Davis and de Serres, 1970). Transformation was performed by means of protoplasts for aconidial strains (Vollmer and Yanofsky, 1986) or electroporation for conidial strains using 0.2 cm gap BioRad cuvette and BioRad electroporator set to 1.5 kV/ 600 ohms/ 25 μF (Margolin *et al.*, 1997). Selection of transformants was carried out in Vogel's medium with 0.05% fructose, 0.05% glucose, and 2% sorbose in presented of 200μg/ml hygromycin.

S. cerevisiae strains FY73 [*MATa, trp1Δ63, leu2Δ63, ura3-52*], and W303 [*MATa, trp1-1, leu2-3, 112, ura3-1, his3-11-15, can1-100*] (Clontech) were used for yeast protein expression. Strain W303-1A [*MATa, ade2-12, trp1-1, leu2-3, 112, ura3-1, his3-11-15, can1-100*] was used for the yeast homologous recombination cloning technique (Bourett *et al.*, 2002). General yeast media were prepared and molecular cloning techniques were performed according to the Yeast Protocols Handbook (Clontech).

E. coli strain DH10B was cultured in LB medium (Sambrook and Russell, 2001) and used for general molecular cloning. Strains JM109 and BL21(DE3) were used for protein expression. *E. coli* transformation was performed by electroporation using 0.1 cm gap BioRad cuvette and BioRad electroporator set at 1.8 kV/ 200 ohms/ 25 μF.

Plasmid constructions for FL protein expression in *N. crassa*. Standard recombinant DNA techniques were carried out according to Sambrook and Russell (Sambrook and Russell, 2001). Plasmid descriptions and all oligonucleotides used were listed in Table 1 and Table 2, respectively.

To construct a plasmid for overexpression of a full-length of FL with hexahistidine tag under the regulation of the *cpc-1* promoter, the *Sma*I fragment from pLBS7 (Bailey-Shrode and Ebbole, in press) containing genomic DNA of *fl* was cloned into pBluescriptSK$^-$ (Stratagene) to yield pRP1. A double-stranded oligonucleotide with *Nar*I overhangs, encoding arginine (R), glycine (G), serine (S) and six amino acid residues of histidine (RGSH6) was generated by annealing equimolar amounts of two oligonucleotides, PS1 and PS2. This fragment was inserted into the *Nar*I site located at

Table 1. Plasmids used in this study.

Plasmid name	Detail
pBluescriptSK$^-$	*E. coli* cloning vector (Stratagene)
pCB1004	fungal cloning vector (Sweigard *et al.*, 1997)
pCR2.1-TOPO	PCR cloning vector (Invitrogen)
pET-30 Ek/LIC	*E. coli* expression vector, T7 promoter (Novagen)
pKK388-1	*E. coli* expression vector, *trc* promoter (Clontech)
pLBS6	fungal cloning vector, *cpc-1* promoter (Bailey-Shrode and Ebbole, in press)
pLBS7	*fl* gene in pLBS6 (Bailey-Shrode and Ebbole, in press)
pRP1	*fl* gene in pBluescriptSK$^-$
pRP2	*fl* gene with RGHSH6 in pBluescriptSK$^-$
pRP5	$fl_{1\text{-}782H6}$ under the control of *cpc-1* promoter
pRP17	RGSH6 in pKK388-1
pRP18	$fl_{1\text{-}371H6}$ under *trc* promoter
pRP40	$fl_{1\text{-}782H6}$ in TOPO vector
pRP41	$fl_{1\text{-}782H6}$ under *GAL1* promoter
pRP45	$fl_{1\text{-}371H6}$ under T7 promoter
pRP66	pCB1004 with *RP27* promoter
pRP67	pRP66 with β-tubulin terminator
pRP72	$fl_{1\text{-}782H6}$ under control of *RP27* promoter
pRP79	*fl* without PEST motifs under control of *RP27* promoter
pRP83	*fl* without PEST motifs under control of *GAL1* promoter
pRP92	$fl_{1\text{-}199}$ with GST fusion
pSM565	*E. coli*, yeast, fungal shuttle vector (GenBank AY142483)
pYES2	yeast expression vector, *GAL1* promoter (Clontech)

Table 2. Oligonucleotide PCR primers used in this study.

Primer	Sequence (5’ ---> 3’)
PS1	CGAGAGGATCCCATCACCATCACCATCACGGGG
PS2	CGCCCCGTGATGGTGATGGTGATGGGATCCTCT
PS6	CTGGCTGTTGCGAATTCTTG
PS7	CTAGAGGATCTCATCACCATCACCATCACTGCA
PS8	GTGATGGTGATGGTGATGAGATCCT
PS24	GACGACGACAAGATGCCAAGACAACACCTA
PS25	GAGGAGAAGCCCGGTTAAGACACTGCCCTGGCT
PS32	CTCACTCGAGGGCGAATTGGG
PS33	CGAAAGCTTAGATCTGTTAGGC
PS34	CCTCCGCGGTGGAATGCGCCG
PS35	GACGAGCTCATCATCATGCAAC
PS45	AACCCAATCTTCAAAATGCCAAGACAACAC
PS46	TCCTCGTTCAATGCGCATGAAAAAGGGACT
PS47	AGTCCCTTTTTCATGCGCATTGAACGAGGA
PS48	CTCCGAAGCATCATGGCGATTGCCACAGCT
PS49	AGCTGTGGCAATCGCCATGATGCTTCGGAG
PS50	TGAGTGGAATGATTAGCTGTGATGGTGATG
PS51	AAGGTACCATGCCAAGACAACACCTA
PS52	CGGAGCTCTTAGCTGTGATGGTGATG

the 3' end of *fl* in pRP1, to create pRP2. The plasmid was sequenced to confirm that the tagged sequences were in-frame with *fl*. The *Sma*I fragment containing $FL_{1\text{-}782H6}$ from pRP2 was cloned into the *Sma*I site of pLBS6 (Bailey-Shrode and Ebbole, in press) generating $FL_{1\text{-}782H6}$ under the regulation of the *cpc-1* promoter and *trpC* terminator. This plasmid, which also contains chloramphenicol and hygromycin resistance markers, was designated pRP5.

A plasmid containing $FL_{1\text{-}782H6}$ under the control of *Magnaporthe grisea* ribosomal protein 27 (*RP27*) promoter and *N. crassa* β-tubulin terminator was constructed in three steps as shown in Fig. 3. First, a PCR product bearing the *RP27* promoter was amplified from *M. grisea* DNA using primers PS32 and PS33. The PCR product was digested with *Xho*I and *Hind*III and inserted directionally into the *Xho*I-*Hind*III sites of pCB1004 (Sweigard *et al.*, 1997) generating plasmid pRP66. Second, a PCR product containing the β-tubulin terminator was generated using primers PS34 and PS35. The amplified product was digested with *Sac*II and *Sac*I and inserted directionally at the *Sac*II-*Sac*I sites of pRP66. The recombinant plasmid was named pRP67. Third, a *Sma*I fragment of $FL_{1\text{-}782H6}$ from pRP5 was inserted into the *Sma*I site of pRP67. The resulting plasmid, pRP72, contains $FL_{1\text{-}782H6}$ under the control of the *RP27* promoter and β-tubulin terminator, with chloramphenicol and hygromycin resistant genes as selectable markers.

Plasmid pRP79, which contains *fl* lacking PEST sequences, was constructed by a yeast homologous recombination method using pSM565 vector (GenBank AY142483)

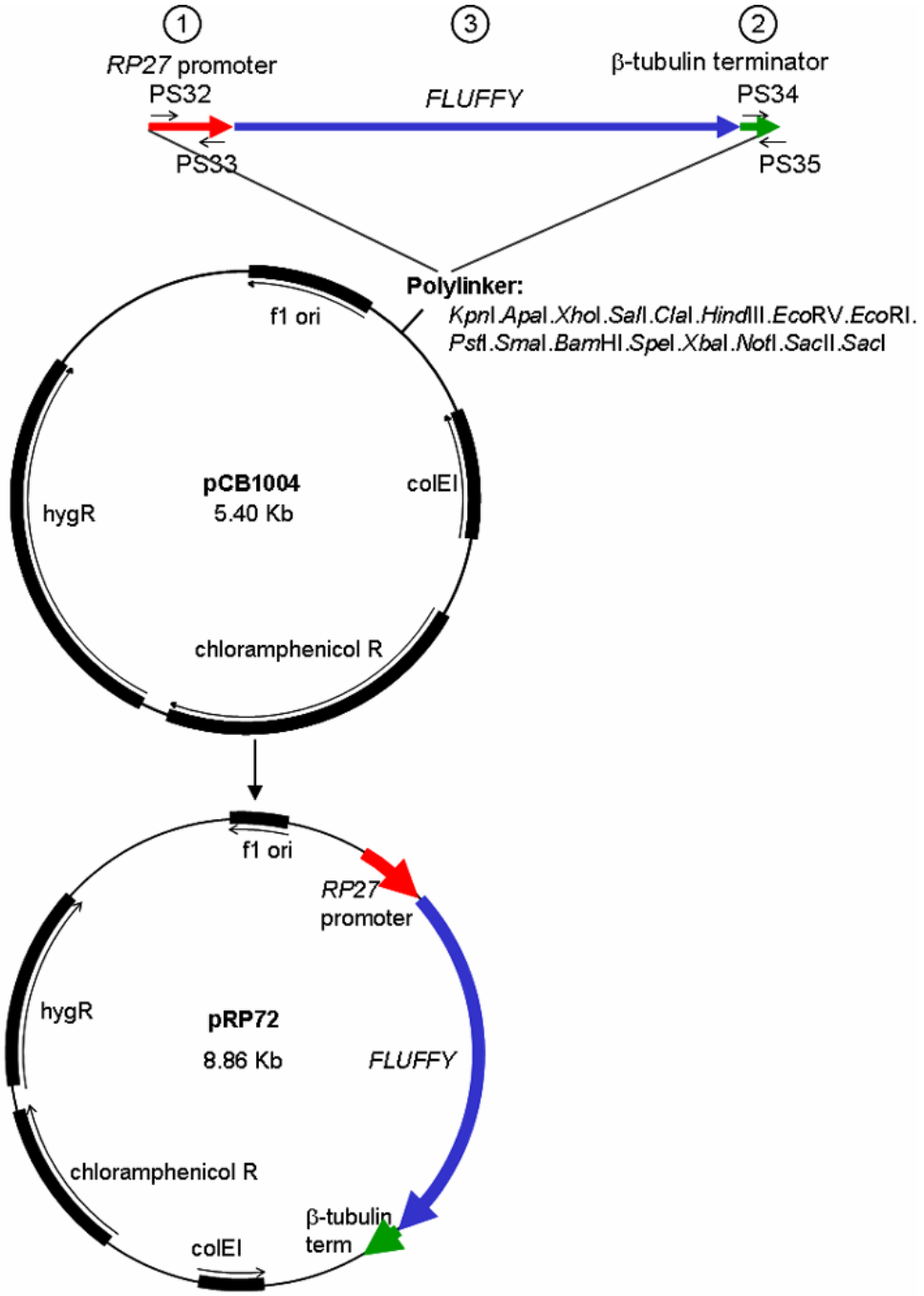


Fig. 3. Construction of pRP72. PCR product of *RP27* promoter (step 1, red) using primers PS32 and PS33, and β-tubulin terminator (step 2, green) using primers PS34 and PS35 was inserted into pCB1004. Full-length genomic DNA of *fl* with hexahistidine tag (step 3, blue) was placed under the control of *RP27* promoter.

as shown in Fig. 4. In this construct, *fl* without PEST sequences was placed under the control of *RP27* promoter. Three PCR fragments of *fl* were generated using *fl* cDNA with RGSH6 tagged as template. The first fragment was amplified using primers PS45 and PS46. Primer PS45 contains 15 nucleotides of the 3'end of *RP27* promoter, and also 15 nucleotides of the 5' end of the *fl* gene, while primer PS46 is a reverse strand of the nucleotides from position 1201-1215 and 1264-1278 of *fl* (Fig. 4). The second PCR fragment was amplified using primers PS47 and PS48. Primer PS47 complements to the primer PS46, and PS48 is a reverse strand of the nucleotides from position 2008-2022 and 2062-2076 of *fl*. The third PCR fragment was amplified using primers PS49 and PS50. Primer PS49 is a complementary strand of the primer PS48, and PS50 is a reverse strand containing 15 nucleotides of the hexahistidine tag at the 3' end of *fl* and also 15 nucleotides of 5' end of β-tubulin terminator gene presented in the vector pSM565. The plasmid pSM565 was digested with *Xho*I, filled-in with Klenow fragment, and mixed with the three PCR products. The DNA mixture was introduced into competent yeast cells, strain W303-1A, in the presence of 30% polyethylene glycol 3350 and 0.09 M lithium acetate pH8.6 (Bourett *et al.*, 2002). The reaction was incubated at 30^{o}C for 30 min, then heat shocked at 42^{o}C for 20 min, and plated on yeast selective medium without tryptophan. Plasmids were extracted from yeast cells and introduced into *E. coli*. Positive clones were identified by restriction enzymes digestion and sequencing.

Plasmid construction for FL protein expression in *S. cerevisiae*. To construct plasmid for overexpression a full-length of FL with hexahistidine tag in *S. cerevisiae*, a full-length sequence of *fl* cDNA was generated. The mRNA from *N. crassa* harboring

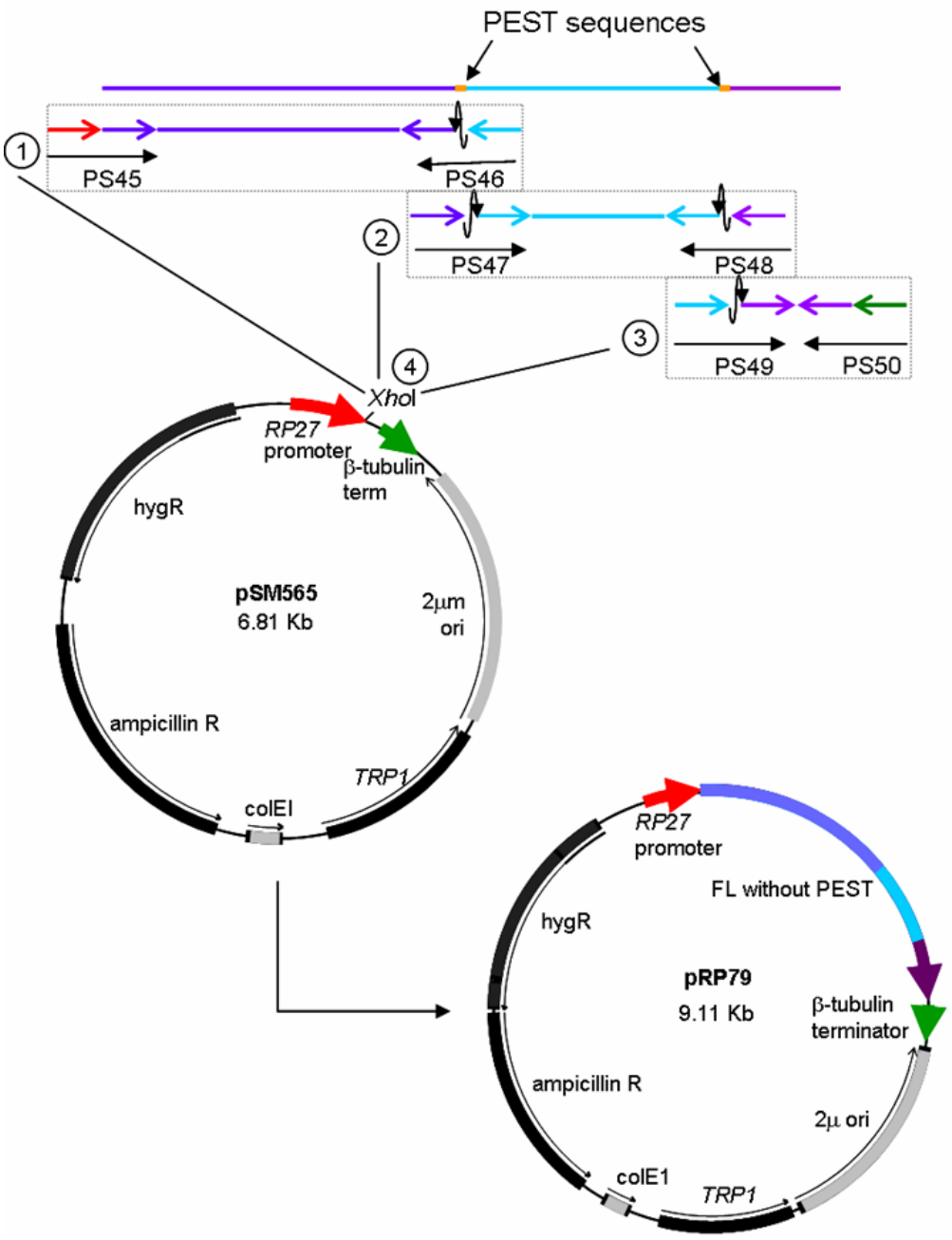


Fig. 4. Construction of pRP79. Three segments of FL (step 1, 2, and 3) lacking the PEST sequences were amplified using primers PS45-PS50 as indicated. The fragments were cloned into *Xho*I site of pSM565 (step 4) by yeast homologous recombination method. *fl* was expressed under control of the *RP27* promoter (red) and β-tubulin terminator (green).

pRP5 was extracted from total RNA using FastTrack® 2.0 mRNA Isolation Kit (Invitrogen). The mRNA was used to make cDNA with First-Strand cDNA Synthesis Kit (Amersham). Primers PS5 and PS8 were used to amplify *fl* cDNA from the cDNA pool using VENT DNA polymerase (New England Biolabs), generating a 2.37 kb full-length cDNA fragment of $FL_{1\text{-}782H6}$. The amplified product was initially cloned into a PCR cloning vector (Invitrogen) and the inserted sequence was analyzed. This plasmid was named pRP40. The *Kpn*I-*Xho*I fragment containing full-length $FL_{1\text{-}782H6}$ from pRP40 was inserted into the *Kpn*I-*Xho*I sites of pYES2 vector (Clontech). The recombinant plasmid, pRP41, contained full-length $FL_{1\text{-}782H6}$ under regulation of the *GAL1* promoter and *CYC1* terminator, with ampicillin resistant and uracil (*URA3*) gene as selectable markers in *E. coli* and yeast, respectively.

In order to express *fl* lacking PEST sequences in *S. cerevisiae*, the PCR product was amplified from pRP79 using primers PS51 and PS52. The PCR product was digested with *Kpn*I-*Sac*I and directionally cloned into the *Kpn*I-*Sac*I site of pYES2 vector. The recombinant plasmid was designated pRP83.

Plasmid constructions for FL protein expression in *E. coli*. Three different vector systems were used to express the putative FL binding domain. In the first system, pKK388-1 vector (Clontech) was used to generate a recombinant plasmid in two steps. First, a double stranded oligonucleotide coding for RGSH6 with *Xba*I and *Pst*I overhangs PS7 and PS8 was inserted at the *Xba*I-*Pst*I sites of pKK388-1 vector. This plasmid was named pRP17. Next, the PCR product coding for *fl* DNA binding domain, including the middle homology part of FL located in amino acid residues 1-371, was

amplified from cDNA template using primers PS5 and PS6. To create a good ribosomal binding site and start codon in *E. coli*, the first C after ATG of *fl* was changed to G generating an *Nco*I site. This procedure led to a changing of the second amino acid from proline to alanine. The PCR product was digested with *Nco*I and *EcoR*I and inserted into the *Nco*I-*EcoR*I sites of pRP17. This plasmid, pRP18, contained $FL_{1\text{-}371H}$ under the control of *trc* promoter.

The second system was created by expression of FL DNA binding domain under T7 promoter using the pET-30 Ek/LIC vector (Novagen) system. Primers PS24 and PS25 were used to amplify the *fl* DNA binding domain located in amino acid residues 1-378, which includes the middle homology part of FL. This PCR product was treated with T4 DNA polymerase in the present of only dATP to create specific 13- and 14-nucleotide single-stranded overhangs, which are compatible with the overhangs found in the vector. The reaction was extracted once with phenol/chloroform, ethanol precipitation, and then annealed to the pET30 Ek/LIC vector, generating the plasmid pRP36 containing $FL_{1\text{-}378H}$. In order to produce a protein construct with the ability to interact with antibody to RGSH6, *EcoR*I-*Xho*I fragment containing FL amino acid 369-378 and a part of vector was removed, and replaced by a fragment containing FL amino acid 369-371 and RGSH6 from pRP18. The new construct, encoded $FL_{1\text{-}371H}$, was named pRP45.

The third system expressed FL as a fusion protein to glutathione-S-transferase (GST) protein in the pGEX4T-1 vector (Amersham). To create the GST::$FL_{1\text{-}199}$ fusion protein, an *Nco*I-*Xba*I filled-in fragment containing the putative FL binding domain and

coiled-coil region of *fl* from pRP18 was ligated into the *Sma*I site of pGEX4T-1 vector. The direction of the cloned fragment was verified by restriction enzyme digestion. The recombinant plasmid was designated pRP92.

FL protein expression, extraction, and purification. Protein extraction from *N. crassa* was performed as follows. Vogel's medium was inoculated with conidia to a final concentration of $1x10^6$ conidia/ml. The culture was incubated at 34^{o}C, 200 rpm, for 16 hr. Mycelia were filtered through Whatman paper, frozen in liquid nitrogen and kept at -80^{o}C until used. The mycelia were ground to a powder in liquid nitrogen and the proteins were extracted in lysis buffer [20 mM HEPES pH 7.9, 100mM KCl, 2 mM EDTA, 10 mM DTT, 1 mM PMSF, and 1/20 v/w of protease inhibitor cocktail (Sigma)] using a mini beadbeater (Biospec Products). Crude protein was separated from cell debris by centrifugation at 15,000 x g for 5 min at 4^{o}C. Glycerol was added to 15%, and the sample was stored at -80^{o}C.

S. cerevisiae was first pre-cultured in a synthetic defined (SD) medium with glucose as a carbon source, at 30^{o}C, 200 rpm, for approximately 20 hr. The cells were pelleted, washed twice with sterile water, and transferred to minimal medium with galactose as the carbon source. The cells were harvested three hours after the induction with galactose. Protein was extracted as described (Yeast Protocols Handbook, Clontech).

E. coli was cultured in 2xYT medium containing 2% (w/v) glucose and 100 μg/ml ampicillin at 37^{o}C, overnight. The overnight culture was diluted 50 times with the same medium and was grown until the OD_{600} was approximately 0.6. Isopropyl-β-D-

thiogalactoside (IPTG) was added to 1 mM, and cells were grown at 30^{o}C for 3 hr. Cells were harvested by centrifugation at 2,500 x g for 10 min at 4^{o}C, and washed with cold phosphate-buffered saline (PBS). Native and denaturing extraction conditions as well as renaturing by stepwise dialysis were performed as described (Melcher, 2000). For GST::FL_{1-199}, the cell pellet was resuspended in PBS buffer and disrupted by sonication. Cell debris was removed by centrifugation at 12,000 x g for 15 min at 4^{o}C.

Purification of histidine-tagged proteins was performed using Ni-NTA agarose (Qiagen), whereas GST::FL fusion protein was fractionated through an affinity column packed with Glutathione Sepharose 4 Fast Flow (Pharmacia), as described by the manufacturer's protocol. Glycerol was added to the eluted proteins to a final concentration of 15%. The samples were aliquoted and stored at -80^{o}C.

Protein detection. Proteins were separated by 12% SDS-polyacrylamide gel electrophoresis (SDS-PAGE), stained with 0.025% Coomassie Brilliant blue R 250, destained, and vacuum dried. For western analysis, the proteins from SDS-PAGE were transferred onto PVDF membrane (Millipore). The membrane was blocked with 3% BSA in TBS buffer (10 mM Tris-HCl, pH7.5 and 150 mM NaCl) and incubated with antibodies specific to either the 6xHis tag (Qiagen) or FL_{1-371H} fusion protein (Quality Controlled Biochemical, Inc.), followed by incubation with either anti-mouse or anti-rabbit conjugated with alkaline phosphatase, respectively. Enzyme activity was detected using BCIP/NBT alkaline phosphatase substrate (Sigma) according to the manufacturer's instructions.

Rabbit-antibody raised against FL was performed using purified protein derived from *E. coli* bearing pRP45 (Quality Controlled Biochemical Inc). Mouse-antibody for RGSH6 was purchased from Qiagen. Anti-rabbit IgG and anti-mouse IgG alkaline phosphatase conjugate were obtained from Sigma.

Southern and northern blot analysis. *N. crassa* conidia were inoculated in Vogel's medium to a final concentration of $1x10^6$ conidia/ml. Mycelia of the *fl* mutant strain were blended and used as inoculum. The culture was performed at 34^{o}C, 200 rpm, for 16 hr. Mycelia were harvested onto Whatman paper and frozen in liquid nitrogen. Genomic DNA was isolated as described previously (Vollmer and Yanofsky, 1986). RNA samples were prepared by miniprep-RNA extraction as described (Madi *et al.*, 1994). DNA and RNA were blotted onto Zeta-probe blotting membrane (Bio-Rad), UV-cross-linked, baked at 80^{o}C for 30 min, and hybridized according to the membrane manufacturer's instructions. Probes were radio-labeled with [α^{-32}P]CTP using the Rediprime II system (Pharmacia).

RESULTS

FL protein expression in *N. crassa* and *S. cerevisiae*. Purified FL protein is needed to identify the DNA recognition sequence of FL. Knowledge of the binding sequence will permit identification of potential target sites in the genome and thereby define genes that might be regulated by *fl*. In order to obtain FL protein from *N. crassa*, the full-length genomic DNA of *fl* was engineered to contain a tag composed of the tripeptide sequence RGS followed by six residues of histidine ($FL_{1\text{-}782H6}$). Antibodies

recognizing this $RGSH_6$ sequence are commercially available. Two constructs of $FL_{1\text{-}782H6}$ were created; pRP5 which has $FL_{1\text{-}782H6}$ under the regulation of *cpc-1* promoter, and pRP72 which has $FL_{1\text{-}782H6}$ under the control of *RP27* promoter. When the *cpc-1* driven $FL_{1\text{-}782H6}$ plasmid was introduced into the *fl* mutant strain, transformants were able to conidiate properly, indicating that the construct could complement the *fluffy* phenotype. In addition, when the plasmid was introduced into wild-type background conidiation was induced in submerged cultures (Fig. 5). Northern blot analysis showed elevated levels of *fl* from these transformants (Fig. 5, lane 3-6) compared to the wild-type, and *fl* mutant (Fig. 5; lane 1 and 2, respectively). These results indicate that the $FL_{1\text{-}782H6}$ in both constructs functioned properly, and elevated levels of FL could induce conidiation in liquid culture. The mRNA expression levels of *fl* driven by *cpc-1* and *RP27* promoter were similar. Proteins were extracted from these transformants, separated by SDS-PAGE, detected by Coomassie blue staining, and western blot analysis using antibody to $RGSH_6$. The expected size for $FL_{1\text{-}782H6}$ is approximately 89 kDa. However, no $FL_{1\text{-}782H6}$ protein was detected with antibody to RGSH6 (data not shown), indicating that there was a very low level of FL protein or that the protein was degraded rapidly after extraction.

Because of the unsuccessful attempt to detect FL protein from *N. crassa*, I used *S. cerevisiae* as an alternate expression system. A full-length *fl* cDNA was produced from mRNA of *N. crassa* bearing pRP5 by PCR and cloned into the pYES2 vector to drive expression of the protein from the *GAL1* promoter. The recombinant plasmid, pRP41, was introduced into yeast strains FY73 and W303. Total RNA and proteins were

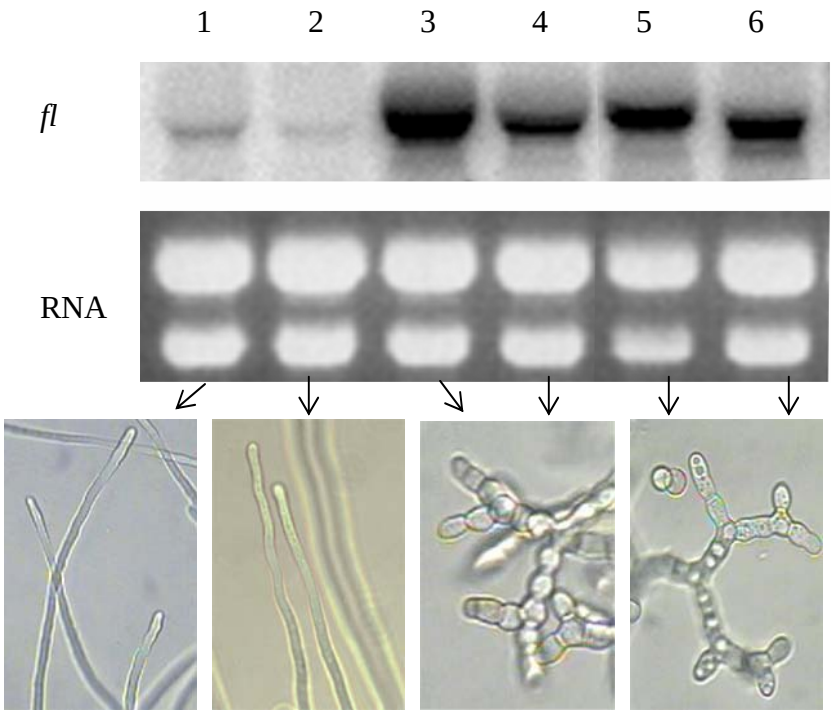


Fig. 5. Morphology of *N. crassa* and expression of *fl* under *cpc-1* and *RP27* promoter. *N. crassa* was grown in Vogel's liquid medium for 16 hr. Total RNA and protein were extracted from the mycelia of wild-type (lane 1), *fl* mutant (lane 2), and transformants bearing pRP5 (lanes 3-4), which contain *cpc-1*-FL_{RGSH6}, or pRP72 (lanes 5-6), which contain *RP27*-FL_{RGSH6}.

extracted after induction by galactose. It was found that *fl* transcript was detected in both yeast strains harboring pRP41 (Fig. 6, lane 1 and 3) but not in control strains with empty vector (Fig. 6, lane 2 and 4). The transcript levels of $FL_{1\text{-}782H6}$ under control of the *GAL1* promoter in *S. cerevisiae* and the *cpc-1* promoter in *N. crassa* revealed that the level of *fl* mRNA produced in yeast was much greater than that in *N. crassa* (Fig. 7, lane 4 and 3, respectively). The *fl* transcript in yeast was detected at a high level even when the blot was exposed for a short period of only six hours, whereas a longer exposure time was required to detect the *fl* transcript in *N. crassa* overexpression strain. The proteins from yeast overexpressing $FL_{1\text{-}782H6}$ were extracted. However, no protein corresponding to the expected size of $FL_{1\text{-}782H6}$ (89 kDa) was detected using $RGSH_6$ or FL-specific antibodies in any yeast transformants (data not shown). In summary, expression of $fl_{1\text{-}782H6}$ in both *N. crassa* and *S. cerevisiae* can be detected at the RNA level but not at the protein level. One possible explanation for this result is rapid turnover of the FL protein.

It has been reported that motifs known as PEST sequences, that are rich in proline, glutamate, serine, and threonine, serve as signals for proteolytic degradation (Rechsteiner and Rogers, 1996; Rogers *et al.*, 1986). The possibility that PEST sequences were responsible for the difficulties associated with isolating FL protein was investigated using the algorithm PESTFind (http://www.at.embnet.org/embnet/tools/bio/PESTfind/). The PESTFind search provides a score ranging from -50 to +50, with a score over zero indicating a possible PEST region. A value greater than +5 indicates a highly probable PEST region. According to this computer-based search, it was found that FL polypeptide contains two potential

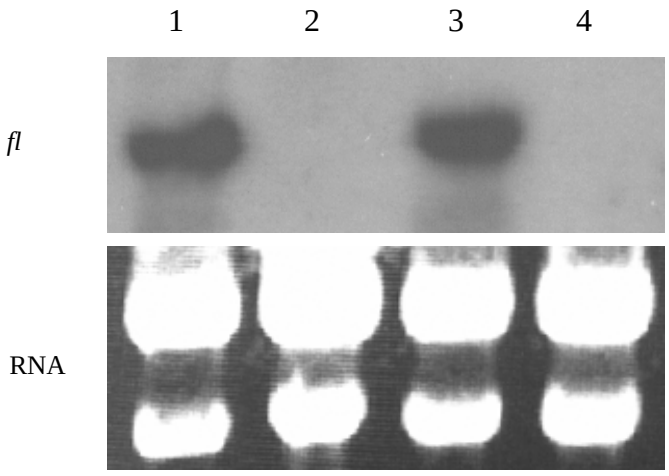


Fig. 6. Northern blot analysis of *fl* in *S. cerevisiae* under *GAL1* promoter. Total RNA (10 μg) was derived from yeast strain FY73 (lanes 1 and 2), and W303 (lanes 3 and 4) harboring pRP41 (lanes 1 and 3) and empty vector (lanes 2 and 4) after three hours induction with galactose.

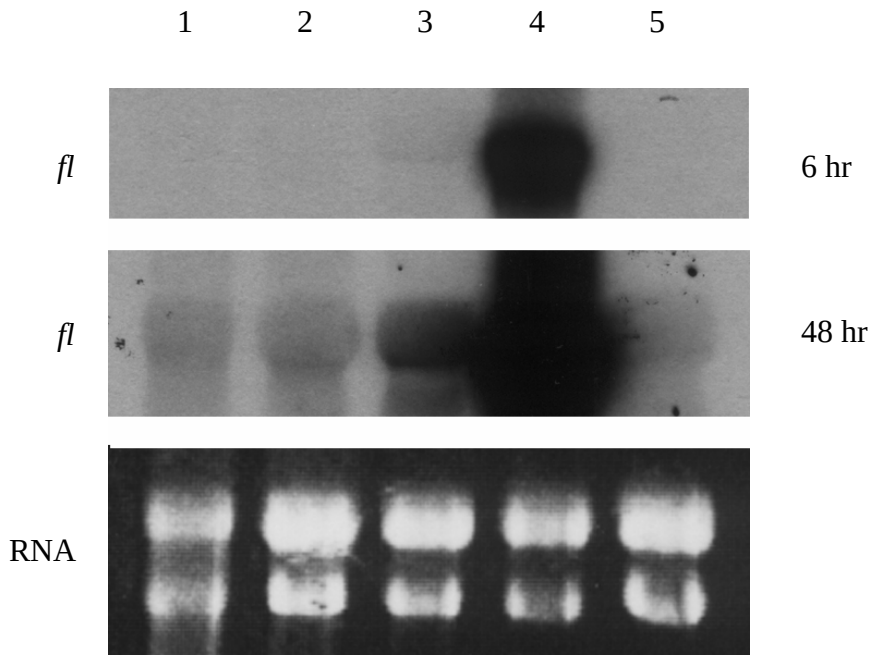
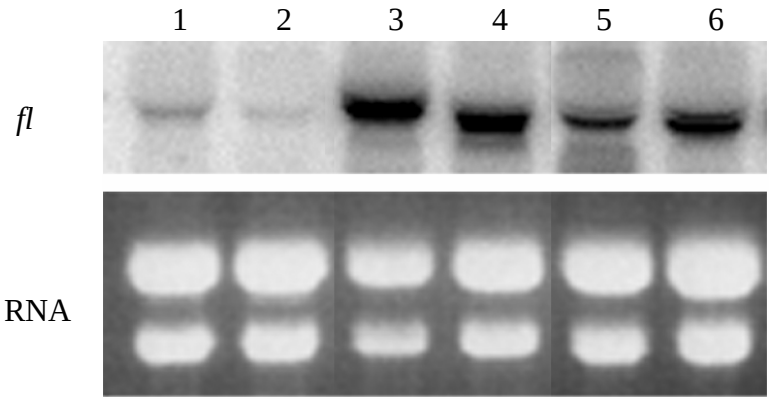


Fig. 7. Comparison of *fl* transcript in *N. crassa* and *S. cerevisiae*. Lane 1, *fl*; lane 2, *N. crassa* wild-type; lane 3, *N. crassa* harboring pRP5 (*cpc-1*-FL_{RGSH6}); lane 4, *S. cerevisiae* harboring pRP41 (*GAL1*-FL_{RGSH6}); and lane 5, *S. cerevisiae* harboring pYES2. The blot was exposed for 6 hr and 48 hr as indicated.

PEST sequences. These sequences are located at amino acids 406-422 (RLYQGSEETPEDEPQNR), and 675-688 (RPSDSFTSSTLTSH) with a PEST score of +9.26 and +7.25, respectively. These sequences may be responsible for targeting the FL protein for degradation and may account for the difficulty in detecting FL protein in *N. crassa* or *S. cerevisiae*. To test this hypothesis, both PEST sequences were removed from the FL polypeptide by construction of in-frame deletions. The final constructs, pRP79 and pRP83, contain FL without PEST sequences under the control of the *RP27* and the *GAL1* promoter for expression in *N. crassa* and *S. cerevisiae*, respectively. In both organisms, overexpression of *fl* lacking the PEST sequences showed elevated level of *fl* transcript compared to the host strain (Fig. 8). The levels observed in transformants lacking PEST sequences, however, appeared to be lower than in transformants containing PEST motifs in both *N. crassa* and *S. cerevisiae* (Fig. 8). No FL protein was detected by western blot analysis using antibodies to the $RGSH_6$ or FL-specific antibodies of transformants lacking the PEST sequences (data not shown) suggesting that these sequences were not solely responsible for the low level of FL protein. In addition, *N. crassa* harboring pRP79 did not conidiate in liquid culture indicating that FL protein that lacks the regions containing PEST sequences is non-functional.

FL protein expression in *E. coli*. Because of the difficulty in expressing FL in *N. crassa* and *S. cerevisiae*, an *E. coli* system was employed to express FL protein for purification. Several transcription factors have been successfully studied by expression of the proteins in *E. coli* (Carey and Smale, 2000; Ehrlich *et al.*, 1999; Fu *et al.*, 1995). $FL_{1\text{-}371H}$ was expressed under the control of two types of promoter, *trc* and T7.

A



B

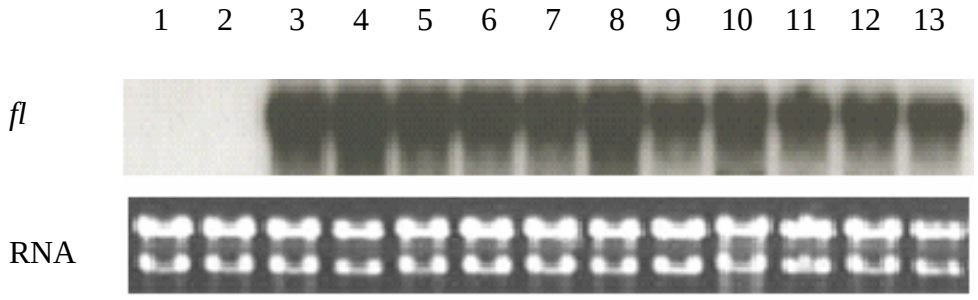


Fig. 8. Northern blot analysis of *fl* in *N. crassa* and *S. cerevisiae* with and without PEST sequences. A) *N. crassa* was grown in Vogel's liquid culture for 16 hr and total RNA was extracted. Lane 1, wild-type; lane 2, *fl* mutant; lanes 3-4, transformants harboring pRP72 (with PEST sequences); and lane 5-6, transformants bearing pRP79 (without PEST sequences). B) Total RNA was extracted from *S. cerevisiae* after induction. Lane 1, host strain; lane 2, transformant with empty vector; lanes 3-8, with pRP41 (with PEST sequences); and lanes 9-13, with pRP83 (without PEST sequences).

Crude cell lysates before and after induction were separated by SDS-PAGE to verify the level of expression (Fig. 9). This experiment showed that the expression level of $FL_{1\text{-}371H}$ under the T7 promoter was higher than the expression level observed using the *trc* promoter (Fig. 9, lane 3 and 5, respectively). Thus, the T7 promoter system was selected for expression and purification of $FL_{1\text{-}371H}$.

Native extraction and purification was initially attempted. However, several proteins co-purified with FL (Fig. 10A). In addition, the $FL_{1\text{-}371H}$ was mainly localized in the cell pellet (data not shown). Varying the parameters of growth conditions, such as temperature and IPTG concentration did not improve the efficiency of protein purification (data not shown). Thus, extraction and purification under denaturing condition followed by stepwise renaturing dialysis was performed, which yielded soluble purified protein (Fig. 10B). Antibody to the histidine tag recognizes a protein of the predicted size (48 kDa) present in *E. coli* harboring $FL_{1\text{-}371H}$ (lanes 5-7) but not in *E. coli* transformed with empty vector (lanes 1-3). Large scale growth for protein extraction and purification under denaturing condition from 3 liters of culture yielded approximately 16 g of wet cell weight which produced approximately 539 mg of total protein and 16 mg of $FL_{1\text{-}371H}$ protein, respectively. This amount of $FL_{1\text{-}371H}$ protein accounted for about 3% of the total protein. One fraction of the protein was used to raise antibody, while another fraction was processed through step-wise dialysis which was subsequently used in attempts to identify specific FL binding sequences using the oligonucleotide random site selection method. However, attempts to use the dialyzed $FL_{1\text{-}371H}$ protein in

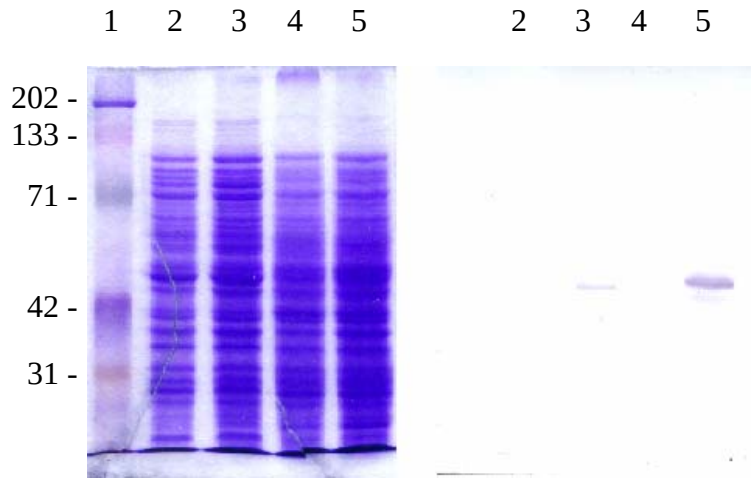
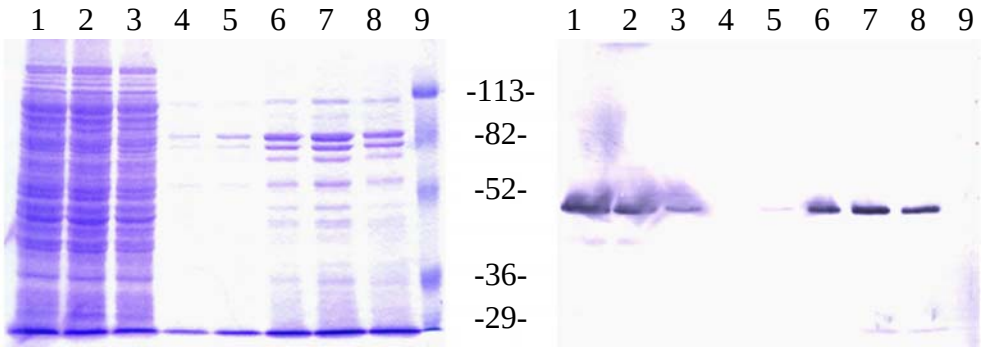


Fig. 9. Comparison of expression level of $FL_{1\text{-}371H}$ in *E. coli* under *trc* (lanes 2 and 3) and T7 (lanes 4 and 5) promoters. Crude proteins before (lanes 2 and 4) and after (lanes 3 and 5) induction with 10 mM IPTG were separated on12% SDS-PAGE, and stained with coomassie blue (left) and performed western analysis using antibody to RGSH6 (right). Lane 1, molecular weight marker (kDa).

A



B

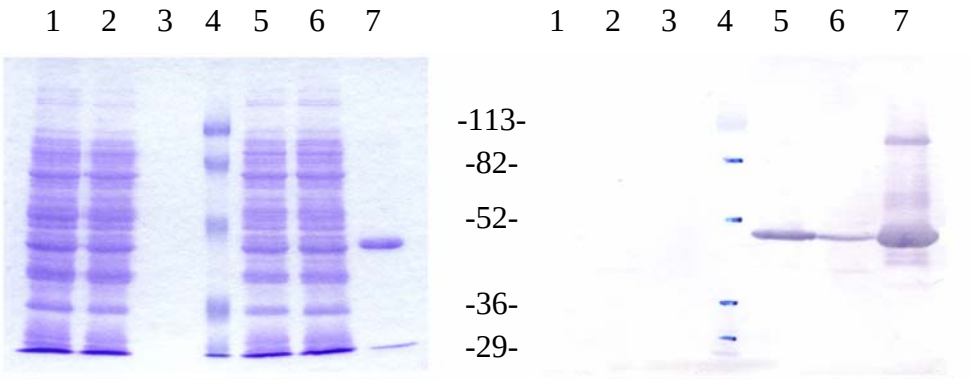


Fig. 10. Extraction and purification of $FL_{1\text{-}371H}$ from *E. coli* bearing pRP45 using Ni-NTA resin by native (A) and denaturing (B) condition. A) Lane 1, crude extract; lane 2, flow-through; lanes 3-5, washes; lanes 6-8, eluates, and lane 9, molecular weight marker (kDa). B) Protein derived from *E. coli* with empty vector (lanes 1-3), and with pRP45 (lanes 5-7). Lanes 1 and 5, crude extract; lanes 2 and 6, flow-though; lanes 3 and 7, eluate; and lane 4, molecular weight marker (kDa). Left, coomassie-stained gel; right, western analysis using antibody to RGSH6.

random oligonucleotide selection did not yield any detectable protein-DNA complexes (data not shown). Therefore, an alternate approach using GST::FL fusions was employed.

The putative binding domain and coiled-coil region (amino acid residues 1-199), was fused in-frame to the C-terminus of the GST coding region of pGEX4T-1 vector generating a GST::FL_{1-199} fusion protein. The resulting protein was extracted and purified under native conditions. *E. coli* with the pGEX4T-1 control vector produced a 29 kDa GST protein (Fig. 11, lane 2), whereas GST::FL_{1-199} gave a 50 kDa fusion protein (Fig. 11, lane 4, arrow). Polyclonal antibodies to FL were used to probe a western blot and showed a positive reaction with the FL fusion protein (Fig. 11, lane 4, arrow) but not with GST protein (Fig. 11, lane 2). The antibody also interacted with several apparently truncated forms of FL protein (Fig. 11, lane 4). No signal for the 50 kDa fusion protein was observed when antibody to FL was used with crude protein extract (Fig. 11, lane 3) indicating that GST::FL_{1-199} protein was produced at low levels. However, purification by glutathione-sepharose provided sufficient quantities of protein for further analysis.

DISCUSSION

Obtaining a sufficient quantity and purity of FL protein is required for study its biochemical function as a transcription factor, including its DNA binding capability, specific consensus sequence identification, and identification of binding sites in the promoters of target genes. In this study, I showed that overexpression of $FL_{1-782H6}$ under

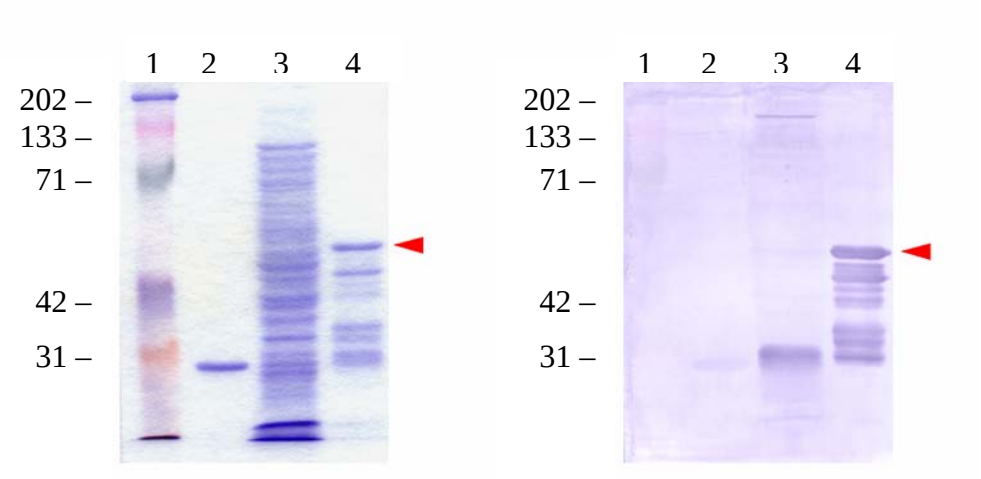


Fig. 11. Expression and purification of GST::$FL_{1-199.}$ Coomassie blue-stained gels (left) and western blot analysis using a rabbit polyclonal antibody against FL (right). Crude lysates were purified using GST-sepharose and proteins were separated in 12% SDS-PAGE. Lane 1, molecular weight marker (kDa); lane 2, purified GST; lane 3, crude extract from *E. coli* harboring GST::FL_{1-199}; lane 4, purified GST::FL_{1-199}. Arrow indicates the 50 kDa GST::FL_{1-199} fusion protein.

heterologous promoters in *N. crassa*, was observed only at the mRNA and phenotypic level. In yeast, the *fl* transcript was induced to a high level; however, the protein was not detected. Overexpression of the FL DNA binding domain in *E. coli* under the T7 promoter produced a large amount of the recombinant protein, but the protein was mainly localized in the cell pellet. Attempts to purify FL under native condition resulted in the co-purification of many non-specific proteins. Conversely, purification under denaturing conditions produced enough purified protein to use for antibody production; the protein obtained by this method, however, did not yield successful results in screening for the FL DNA binding site using degenerate oligonucleotides. Fusion of FL with GST protein improved the solubility of the protein, which finally facilitated purification of a sufficient quantity and quality of FL for further functional analysis.

Elevated expression of *fl* under the *cpc-1* promoter could complement the *fl* mutation. A construct expressing *fl* from the *RP27* promoter was also produced. Expression of *fl* from both the *cpc-1* promoter and the *RP27* promoter induced submerged conidiation in the wild-type background. These results suggest that the FL protein from these constructs is functional. However, purification of the protein from *N. crassa* expressing these transgenes was not successful. Detection of FL with antibody to the $RGSH_6$ and antibody to FL did not yield a signal in *Neurospora* suggesting that the protein is expressed at very low levels and/or is rapidly degraded. In yeast, the mRNA level of *fl* was very high, but nonetheless, no protein was detected. Likewise, this observation suggested that the protein may be translated from the mRNA at very low levels and/or is rapidly degraded after translation.

The stability of FL was evaluated by examining the potential role of PEST sequences. This type of motif has been found to act as a signal for rapid recognition by intracellular proteases resulting in rapid degradation of the protein (Rechsteiner and Rogers, 1996). It was found that FL contains two possible PEST motifs located at amino acids 406-422, and 675-688. Removing these motifs from FL, however, did not improve the stability of the FL protein. From the failure to purify the protein from either *N. crassa* or *S. cerevisiae*, it was concluded that the PEST sequences in FL do not play a major role in controlling the stability of the protein. In addition, overexpression of FL without the PEST sequences under control of the *RP27* promoter could not induce submerged conidiation in the wild-type background. These results suggest that the putative PEST motifs in FL may not be involved in mediating protein degradation, and furthermore, that the putative PEST sequences were required for function of FL. Several studies have shown that motifs resembling the PEST sequence motif may play a role in protein-protein interaction (Chen and Clarke, 2002; Hirose *et al.*, 2003; Lety *et al.*, 2002; Soloviev, 2000). Using a yeast one-hybrid system (Chapter III), the activation domain of FL was identified and localized to amino acids 400-556. Interestingly, the activation domain contains the first putative PEST sequence. It is possible that the PEST sequence in FL has a role in protein-protein interaction rather than signaling for degradation. Indeed, the first PEST motif, RLYQGSEETPEDEPQNR, has a high proportion of acidic residues. As discussed in Chapter III, acidic domains can function as transcriptional activators. As transformants that lacked the PEST sequences yielded less amount of *fl* mRNA than those containing the putative PEST sequence, it is also possible that the

PEST motifs in FL may potentially play a role in autoregulation of FL. Further analysis such as immunoprecipitation and amino acid substitution will aid in identifying partner proteins and ascertaining the role of putative PEST motifs in the function of FL.

Expression of FL in *E. coli* from a strong promoter (T7) gave higher protein levels compared to expression from the *trc* promoter. Unfortunately, the protein was mainly localized in cell pellet. This observation is a typical problem for expression of heterologous proteins in *E. coli* (Carey and Smale, 2000). Attempts to purify protein under native conditions led to co-purification of some *E. coli* proteins. As shown in lane 5, Fig. 10A, dissociation of the histidine tagged fusion protein from Ni-NTA occurring during the column washing steps. It is possible that the native conformation of FL protects the exposure of RGSH6 epitope, thereby causing low binding affinity to Ni-NTA. Extracting the protein by mild denaturant such as urea resulted in improved purification efficiency; however several co-purified proteins were still observed (data not shown). Using strong denaturant such as 6 M guanidine hydrochloride, that completely denatures the protein, led to successful purification of the protein (Fig. 10B). Overexpression of FL using the GST fusion tagged construct, gave lower expression levels of FL, however, the GST fusion tag improved the solubility of FL, allowing the GST::FL_{1-199} to be purified under native conditions. In addition, the GST::FL_{1-199} retained functional DNA binding activity as demonstrated by the ability of the protein to recognize and bind to the target binding site (Chapter III).

CHAPTER III

FLUFFY, THE MAJOR REGULATOR OF CONIDIATION IN *Neurospora crassa,* DIRECTLY ACTIVATES A DEVELOPMENTALLY-REGULATED HYDROPHOBIN GENE

INTRODUCTION

N. crassa grows vegetatively as haploid, multinucleate hyphae that intertwine and anastomose to form a mycelium. The mycelium has three developmental choices, leading to the formation of two types of asexual spores (macroconidia and microconidia) and sexual development leading to the formation of fruiting bodies capable of producing ascospores (Springer, 1993). Macroconidia are very hydrophobic due to a rodlet layer of hydrophobin protein that allows efficient dispersal of the fungus through air (Beever and Dempsey, 1978; Springer and Yanofsky, 1992). Production of macroconidia is influenced by desiccation, light, carbon or nitrogen deprivation, carbon dioxide levels, and the circadian rhythm (Ebbole, 1996; Springer, 1993; Turian and Bianchi, 1972). Synchronous induction of conidiation by exposure of the mycelium to air results in the formation of a mass of aerial conidiophores. During the first two to four hours, polar filamentous growth into the air is observed. After about four hours, apical elongation ceases and growth switches to a process of repeated apical budding (Springer and Yanofsky, 1989). The first several rounds of budding produce proconidial chains with interconidial diameters nearly as large as the diameter of the parent hypha. These chains

are described as minor constriction chains. Eight hours after induction, budding growth becomes more pronounced giving rise to major constriction chains. At this point the developing conidiophores are unable to revert to either minor constriction chain growth or hyphal elongation. Therefore, the transition from minor to major constriction chain growth appears to be a committed step for differentiation of mature conidia (Springer, 1993). At about 12 hours after the induction, budding growth stops and nuclei migrate into the proconidial chains. Double crosswalls are laid down between each proconidium. Four hours later the double crosswalls separate and conidia are easily released and dispersed through air (Ebbole, 1996; Springer and Yanofsky, 1989; Springer, 1993).

A number of genes that are preferentially expressed during conidiation in *N. crassa* were isolated by differential screening methods (Berlin and Yanofsky, 1985). These genes are known as *conidiation* specific genes (*con* genes). Five of these genes *con-6, con-8, con-10, con-11,* and *con-13,* have been studied (Corrochano *et al.*, 1995; Hager and Yanofsky, 1990; Roberts *et al.*, 1988; Roberts and Yanofsky, 1989; White and Yanofsky, 1993). Another conidiation-specific gene that has been characterized in detail is the hydrophobin-encoding *eas* (*easily wettable*) gene, which is also known as *clock-controlled gene-2* (*ccg-2*) and *blue-light inducible-7* (*bli-7*) (Bell-Pedersen *et al.*, 1992; Kaldenhoff and Russo, 1993; Lauter *et al.*, 1992). Expression of *eas* occurs within the first four hours after induction of conidiation and expression of *fl* is necessary and sufficient for this induction.

Several morphological mutants of *N. crassa* were found to have effects on conidial development without generally affecting growth rate or development of aerial

hyphae. Each mutant blocks development at a distinct stage. The *aconidiate-2* (*acon-2*) and *fluffyoid* (*fld*) mutants do not form minor constriction chains. *fluffy* (*fl*) and *acon-3* mutants are able to form minor but not major constriction chains. Two *conidial separation* (*csp-1*and *csp-2*) mutants are able to form both minor and major constriction chains and also double crosswall layers, however, the *csp* mutations block the maturation of double crosswalls so the conidia are not released (Ebbole, 1996; Matsuyama *et al.*, 1974; Springer and Yanofsky, 1989). Previous studies demonstrated that *fl* acts downstream of *acon-2* and upstream of *acon-3*. In addition, the product of *acon-2* is required for *fl* expression (Bailey-Shrode and Ebbole, in press; Correa and Bell-Pedersen, 2002). Of the genes represented by these mutants, the *fl* gene is the only gene that has been cloned (Bailey and Ebbole, 1998).

Sequence analysis revealed that *fl* encodes a 792 amino acid polypeptide containing a Zn_2Cys_6 binuclear zinc cluster domain that belongs to the Gal4p family. The amino acids in the zinc finger domain, coiled-coil, and middle homology regions show high similarity to NIT4, Cha4p, and NCU09205.1 (Bailey and Ebbole, 1998; Galagan *et al.*, 2003), whereas the amino acid sequences that lie outside these regions do not share strong sequence similarity with any other protein in current databases. Developmental induction of the *fl* transcript is tissue specific and expressed in aerial hyphae (Bailey and Ebbole, 1998). It is expressed at a low basal level in vegetative hyphae and induced during the time that major constriction chains are formed. In later stages of development, the transcript level decreases to the preinduction level. Overexpression of *fl* under control of a heterologous promoter is sufficient to induce

conidiation in a wild-type background (Bailey-Shrode and Ebbole, in press). By the time that *fl* is highly expressed, some conidiation-specific gene products such as *con-6, con-10,* and *eas* are induced. The expression of the conidiation specific gene products has also been studied in different genetic backgrounds. It has been found that the *fl* mutant background blocks the expression of *eas* (Bailey-Shrode and Ebbole, in press; Lauter *et al.*, 1992). In addition, elevated expression of *fl* induced the expression of the *eas* gene transcript (Bailey-Shrode and Ebbole, in press). From these findings I hypothesize that Fluffy protein (FL) may function as a transcriptional regulator of genes, such as *eas,* involved in conidial development.

Proteins that belong to the Gal4p family of transcription factors share a conserved DNA binding region of $CX_2CX_6CX_{5-9}CX_2CX_{6-8}C$ which binds to target sequences having CGG triplets (Mamane *et al.*, 1998). The typical CGG triplet is classified into three types; inverted repeat ($CGGX_nCCG$), everted repeat ($CCGX_nCGG$), and direct repeat ($CGGX_nCGG$) (Mamane *et al.*, 1998). A number of DNA binding proteins in this family have been found to be involved in a wide range of biological processes such as primary and secondary metabolism, drug resistance, chromosome segregation, and fruiting body development (Masloff *et al.*, 1999; Todd and Andrianopoulos, 1997). In this work, I demonstrate that FL is a transcriptional activator of the *eas* gene and identify the major *cis*-acting element through which *fl* acts. Knowledge of the FL target sequence allowed me to search the *N. crassa* genome sequence for genes whose promoters have apparent strong FL binding sites and to test several of them to determine if FL regulates their expression.

MATERIALS AND METHODS

Fungal strains, media, and transformation. The *N. crassa* wild-type strain 74OR23-1V*A* (FGSC 2489), and *fl a* mutant (FGSC 46) were obtained from the Fungal Genetics Stock Center (FGSC; Department of Microbiology, University of Kansas Medical Center). Vogel's medium was prepared as described (Davis and de Serres, 1970). Protoplast transformation was performed for aconidial strains (Vollmer and Yanofsky, 1986). Transformation was performed by electroporation of conidia for other strains (Dev and Maheshwari, 2002).

The *S. cerevisiae* strain AH109 [*MATa, trp1-901, leu2-3, 112, ura3-52, his3-200, gal4Δ, gal80Δ, LYS2::GAL1*$_{UAS}$*-GAL1*$_{TATA}$*-HIS3, MEL1, GAL2*$_{UAS}$*-GAL2*$_{TATA}$*-ADE2, URA3::MEL1*$_{UAS}$*-MEL1*$_{TATA}$*-lacZ*] (Clontech) was used in the yeast one hybrid assay. Minimal media with appropriate requirements were prepared as described in the Yeast Protocols Handbook (Clontech).

FL fusion proteins. Two constructs of FL fusion proteins were generated as described in Chapter II, pRP45 and pRP92. The plasmid pRP45 contains a DNA segment encoding *fl* putative DNA-binding domain amino acid residues 1-371 fused with six histidine residue tag ($FL_{1\text{-}371H}$), whereas pRP92 bears a DNA coding for a putative DNA-binding domain amino acid residues 1-199 fused at C-terminus of GST protein (GST::$FL_{1\text{-}199}$).

Random binding-site selection. The single stranded 76-mer oligonucleotide, FLB8 (5'-GACGAATTCGCATGAGCTAGGTACC$(N)_{26}$GCTGCAGATGCTACTGAA GCTTCAC-3'), was designed such that it would contain a 26 nucleotide randomized

internal region and 25 nucleotide flanking sequences. Two additional 25-mer oligonucleotides, FLB9 (5'-GACGAATTCGCATGAGCTAGGTACC-3'), and FLB10 (5'-GTGAAGCTTCAGTAGCATCTGCAGC-3'), were synthesized for use as primers with FLB8. Approximately 6 μg of a pool of double-stranded randomized oligonucleotides were prepared by Klenow fill-in (Sambrook and Russell, 2001). The binding-site selection was performed as described (Pierrou *et al.*, 1995) with slight modifications. Briefly, the pool of double-stranded DNA was mixed in 100 μl binding buffer (20 mM HEPES pH7.9, 50 mM KCl, 2 mM $MgCl_2$, 0.5 mM EDTA, 10% glycerol, 2 mM DTT, 1 mM PMSF, 0.2 μg/μl poly(dI:dC), and 20 μM $ZnCl_2$). Ten ng of GST::FL_{1-199} fusion protein was added to the mixture and incubated at room temperature for 10 min. Fifty μl of a 10% slurry of glutathione-sepharose in 1x binding buffer was added to the reaction and incubated for 5 min. The glutathione-sepharose beads with bound protein-DNA complexes were pelleted by centrifugation at 2,660 x g for 1 min. The pellet was washed in 1 ml of ice-cold 1x binding buffer five times. The washed pellet was resuspended in 100 μl of PCR mix [1x PCR buffer (New England Biolabs), 0.5 mM dNTPs, 2 μM of each primer, and 1 unit VENT polymerase] and amplified for 15 cycles. Each cycle consisted of 1 min at 94^{o}C, 50 sec at 55^{o}C, and 50 sec at 72^{o}C. The PCR reaction was precipitated with 5 M NaCl and ethanol, and resuspended in 100 μl of water. The sepharose beads were removed by centrifugation at 15,000 x g for 1 min, and the supernatant was transferred to a new tube.

The second round of selection to enrich for the specific sequence was performed by repeating the steps described above with 10 μl of the PCR product. After the last

wash, the sepharose beads were resuspended in 100 μl of PCR mix, without VENT polymerase, and boiled for 4 min to release DNA from the complex. After centrifuging and transferring the supernatant to a new tube, VENT polymerase was added and PCR was performed as indicated for the first round.

Following five enrichment cycles of selection and amplification, the PCR product was extracted with phenol/chloroform and precipitated with ethanol. The DNA was then subjected to 5'end labeling and used for further selection by electrophoretic mobility shift assay (EMSA). The shifted protein-DNA bands were excised, eluted from the gel using a QiaII kit (Qiagen), and subjected to PCR amplification as described above with the exception that the TAQ DNA polymerase was added to the last cycle of the reaction to generate an A-overhang. The PCR products were cloned into the TOPO-PCR cloning vector (Invitrogen), and transformed into *E. coli* strain DH10B. Recombinant plasmids were identified by blue-white screening and sequenced.

Electrophoretic mobility shift assay. Double stranded oligonucleotides and DNA fragments were radiolabeled by T4 polynucleotide kinase using [γ^{-32}P]ATP and used in binding reactions. The binding reaction consisted of 10,000 cpm of labeled DNA with or without 1 μg of GST::FL_{1-199} fusion protein in 20 μl binding buffer. The reaction was incubated on ice for 15 min. A 4% polyacrylamide gel (80:1) was prepared in 0.5x TBE buffer with 2.5% glycerol and pre-run at 4^{o}C at 100V (10V/cm) for 1 hr, followed by 150V for 30 min. The protein-DNA complexes in the reaction were resolved by electrophoresis at 150V for 2 hr. The gels were vacuum-dried to Whatman paper at 80^{o}C for 1 hr and analyzed by Fuji BAS-1800II Phosphoimager system.

In the competition gel shift assay, fixed concentrations of wild-type DNA (0.1 pmole) and GST::FL_{1-199} protein (1 μg) were used. Titrations were done using unlabeled oligonucleotides that compete for protein binding with labeled oligonucleotide. The amount of competitors used were 0.1, 1, 5, 10, 20, and 40 pmole, representing 1x, 10x, 50x, 100x, 200x, and 400x, respectively. The reactions were made in binding reaction buffer with protein being added last. After a 15 min incubation on ice, the mixtures were subjected to electrophoresis as described above. Oligonucleotides used in the competition assay are listed in Table 3.

Surface plasmon resonance (SPR). SPR experiments were carried out with a Biacore® X instrument. The biotinylated double stranded oligonucleotide representing the putative FL binding site in the *eas* promoter, 5'-TTCTCCTGCCCGGAAGTTTCCTCCGGCCATACTCAT-3', was generated and diluted to 0.5 ng/μl in HBS-P (10 mM HEPES pH 7.5, 150 mM NaCl, 0.005% surfactant P20). Biotinylated dsDNA was immobilized to one flow cell of a captavidin (Molecular Probes) sensor chip surface to obtain approximately 120 resonance units (RU). A second flow cell containing captavidin served as a reference for non-specific interactions and bulk-shift responses which were subtracted. GST::FL_{1-199} protein was buffer exchanged into HBS-P and binding was measured at eight concentrations of GST::FL_{1-199} analyte; 50, 100, 175, 250, 350, 750, 1000, 1200 nM at a flow rate of 20 μl/min. Complete regeneration of the captavidin surface was achieved with two 30 sec injections of 50 mM NaOH, 1 M NaCl at a flow rate of 50 μl/min. The equilibrium dissociation constant (K_D) was calculated globally by fitting to the simultaneous ka/kd model and from the steady-

Table 3. Oligonucleotides used in the competition assay

Type of repeat	Nucleotide sequences 5' ---> 3'
eas -1498 region	
CCGG_9_CCGG	TTCTCCTGCCCGGAAGTTTCCTCCGGCCATACTCAT
Inverted	
CGG_9_CCG	TTCTC**a**TG**at**CGGAAGTTTCCTCCG**ta**CATACTCAT
CGG_10_CCG	TTCTC**a**TG**at**CGG**t**AAGTTTCCTCCG**ta**CATACTCAT
CGG_11_CCG	TTCTC**a**TG**at**CGG**t**AAGTTTCCT**t**CCG**ta**CATACTCAT
Everted	
CCG_9_CGG	TTCTC**a**TG**a**CCGAAGTTTCCTCGG**a**CATACTCAT
CCG_10_CGG	TTCTC**a**TG**a**CCG**t**AAGTTTCCTCGG**a**CATACTCAT
CCG_11_CGG	TTCTC**a**TG**a**CCG**t**AAGTTTCCT**t**CGG**a**CATACTCAT
Direct-1	
CGG_9_CGG	TTCTC**a**TG**at**CGGAAGTTTCCTCGG**a**CATACTCAT
CGG_10_CGG	TTCTC**a**TG**at**CGG**t**AAGTTTCCTCGG**a**CATACTCAT
CGG_11_CGG	TTCTC**a**TG**at**CGG**t**AAGTTTCCT**t**CGG**a**CATACTCAT
Direct-2	
CCG_9_CCG	TTCTC**a**TG**a**CCGAAGTTTCCTCCG**ta**CATACTCAT
CCG_10_CCG	TTCTC**a**TG**a**CCG**t**AAGTTTCCTCCG**ta**CATACTCAT
CCG_11_CCG	TTCTC**a**TG**a**CCG**t**AAGTTTCCT**t**CCG**ta**CATACTCAT

state affinity model with each response fitted locally using the Biacore evaluation software. Binding of GST::FL_{1-199} at 1000 nM to double stranded oligonucleotide 5'-TTCTCGGTTGGTCTCTGGCGGAGCCAAACCTTATCTGGTGTTACGTCACAATTCTTGTGTATGATTCGCTTTTCCC-3' (negative control) was negligible.

***EAS-MPG1* tagged gene construction.** Plasmid pCB1532 (Wang *et al.*, 2003) was digested with *Sal*I in the polylinker, filled in, and self-ligated to destroy the *Sal*I site. A PCR product spanning the *eas* gene (-1660 to +1540 of translation start site) was amplified from *N. crassa* genomic DNA by using primers PS58 (5'-CTAAAGCTTAGGGATATCCATGCAG-3') and PS59 (5'-CCGGATCCTTGAGATTAAGAGTC-3'). This fragment was cloned into *Hind*III/*Bam*HI site of the *Sal*I-destroyed pCB1532. The recombinant plasmid was digested with *Sal*I, and dephosphorylated with calf thymus alkaline phosphatase (New England Biolabs). A PCR product containing *mpg1* from *Magnaporthe grisea* was amplified from *M. grisea* DNA using primers PS78 (5'-CAATCGTCGACATGTTCTCCCTCAAG-3') and PS79 (5'-GATAGGAGTCGACTGCATATCGACC-3'), digested with *Sal*I, and ligated to the dephosphorylated plasmid above. The resulting plasmid has the *mpg1* coding region fused in-frame with the *eas* coding region. The orientation and fusion junction of the construct was confirmed by restriction enzyme digestion and sequencing. This plasmid was designated pRP109. To generate a plasmid lacking the putative binding site for FL, a *Hind*III-*Nru*I fragment of pRP109 was replaced with the *Hind*III-*Nru*I fragment of the PCR product derived from primers PS118 (5'-GCCAAGCTTATAAACCCGCAATC-

3') and PS119 (5'-CAAAGACATCAAACCACGACCAAC-3') using *N. crassa* DNA as template. This PCR product contains DNA sequence downstream of the FL binding site to the *Nru*I site in the *eas* promoter. This construct, which lacked the binding site for FL, was designated pRP111.

Northern blot analysis. For northern blot analysis, *N. crassa* ($1x10^6$ conidia/ml) were grown in Vogel's liquid media at 200 rpm, 34^{o}C, 20 hr. The mycelium was harvested on Whatman filter paper, and placed face up on a 0.45% Vogel's medium agar plate containing another piece of Whatman paper. The open plates were incubated at 34^{o}C. The mycelia were harvested at time 0 and 8 hr after exposure to air. mRNA was isolated from mycelia as described previously, and electrophoresed on a 1.5% agarose gel (Madi *et al.*, 1994). Northern blots were performed using standard techniques (Sambrook and Russell, 2001). The signal was analyzed using a Fuji BAS-1800II Phosphoimager system.

Fusion of *N. crassa* FL with *S. cerevisiae* GAL4 DNA-binding domain ($GAL4_{BD}$). The carboxy-terminal portion of FL, consisting of amino acids 60-780 and lacking the DNA binding domain, was amplified from *fl* cDNA using primers FLTH-1 (5'-ACATCGATACGAGCAACGGCAGG-3'), and FLTH-2 (5'-CCTCTCGAGCTCTGCAGACCAAG-3'). The PCR product was digested with *Cla*I and *Xho*I, gel purified, and cloned into the *Cla*I-*Xho*I site of pGADT7 vector (Clontech). The *Sma*I-*Pst*I fragment containing FL_{60-780} was excised, and inserted into the *Sma*I-*Pst*I site of pGBKT7 (Clontech). The FL_{60-780} region was fused in-frame with $GAL4_{BD}$ in the pGBKT7 vector and expressed under control of the *ADH1* promoter. Several constructs

with different lengths at the carboxy terminus of FL were generated using restriction enzyme digestion. The $GAL4_{BD}$::FL_{60-780} construct was digested with *Pst*I. The resulting fragment was then digested with either *Sca*I, *Spe*I, or *Stu*I, and then treated with Klenow fragment to generate blunt ends, gel purified, and self-ligated, generating $GAL4_{BD}$::FL_{60-556}, $GAL4_{BD}$::FL_{60-399}, and $GAL4_{BD}$::FL_{60-287} respectively.

RESULTS

FL binds to an inverted CGG repeat with a nine nucleotide spacer sequence. To determine the target binding site of FL, the GST::FL_{1-199} protein (Chapter II) was used. I performed five rounds of random oligonucleotide selection and one round of EMSA to screen for target binding sites of FL from a pool of 76 bp double-stranded oligonucleotides containing a central 26 bp region of degenerate sequence. The PCR products obtained after selection were cloned and 157 clones were sequenced in order to identify potential FL binding site motifs.

Most transcription factors of the GAL4p class bind to either CGG direct, inverted, or everted repeat motifs with different spacing between motif sequences. I specifically looked for these motifs by counting the number of sequences matching potential GAL4p-type binding sequences. The majority of sequences derived from the enrichment protocol contained a CGG_CCG inverted repeat with a nine nucleotide spacing. There were 56 clones that have this type of pattern. I conclude that the basic consensus sequence for FL DNA binding is $CGG(N)_9CCG$. Approximately 101 clones did not contain a $CGG(N)_9CCG$ motif. These clones may contain weak binding sites that

do not match this consensus or represent a background of sequences that co-purified with the specific FL-DNA complexes.

To identify a more accurate consensus sequence for FL, I examined the spacer nucleotides in the 56 sequences containing the CGG$(N)_9$CCG motif. I noted a very high frequency of A residues adjacent to the CGG at one end of the consensus (53/56 sequences). Using this CGGA sequence as the basis for alignment I determined a consensus sequence of CGGA(A/C/G)NN(A/C/T)NNC(C/T)CCG for these 53 sequences (Table 4). A double stranded oligonucleotide, CGGAGGTTACCTCCG, representing this consensus sequence was used to confirm the binding specificity of FL. FL strongly binds to this consensus sequence (Fig. 12, lane 3), and does not bind to an unrelated sequence (Fig. 13).

FL binds to an inverted CGG triplet in the *eas* promoter. Because expression of *fl* is necessary and sufficient to induce *eas* expression (Bailey-Shrode and Ebbole, in press), I hypothesized that FL may directly bind to the *eas* promoter. A previous analysis of *eas* expression demonstrated that nitrogen starvation-induced expression of *eas* could be observed with a promoter segment extending to position -1614 with respect to the translation initiation codon (Kaldenhoff and Russo, 1993). Our studies indicated that nitrogen starvation-induced expression of *eas* is primarily due to induction of conidiation rather than direct nutritional regulation (Bailey-Shrode and Ebbole, in press). To define FL binding sites in the *eas* promoter, four segments spanning the 1.66 kb *eas* promoter were amplified by PCR for EMSA analysis. Each fragment was incubated with either GST::$FL_{1\text{-}199}$ or $FL_{1\text{-}371H}$. FL binds strongly to a 400 bp of the *eas* promoter located

Table 4. Identification of consensus sequence for FL binding

	1[a]	2	3	4	5	6	7	8	9
A	100[b]	47	38	21	42	26	21	2	6
C	0	30	19	34	21	19	28	72	40
G	0	21	23	15	8	23	26	6	13
T	0	2	21	30	30	32	25	21	42
C[c]	**G**	**G**	**A**	**A/C/G**	**N**	**N**	**A/T/C**	**N**	**N**	**C**	**T/C**	**C**	**C**	**G**

[a] Position in spacer region

[b] Nucleotide frequency (%)

[c] Consensus nucleotide

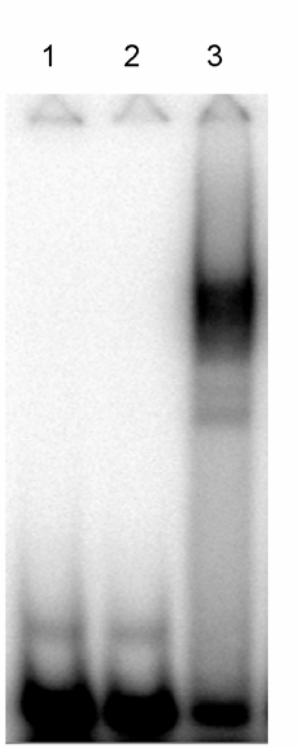


Fig. 12. *In vitro* binding of GST::FL_{1-199}. Gel retardation of the FL consensus binding site (CGGAGGTTACCTCCG) by FL. Radio-labeled-double stranded oligonucleotides were incubated alone (lane 1), with purified GST (lane 2), and with purified GST::FL_{1-199} (lane 3).

upstream of position -1200 (Fig. 13). A strong match to the consensus FL binding site, CGGAAGTTTCCTCCG, was found in this region at position -1498. An oligonucleotide fragment containing this sequence was bound by FL *in vitro* (see below). I also noted a strong consensus FL binding site, CGGACCATTTGTCCG, located -2573 bp upstream of the *eas* coding region. An oligonucleotide DNA fragment containing this sequence was also strongly bound by FL *in vitro* (not shown). There was also weak binding of FL to fragments c and d when GST::FL_{1-199} was used (Fig. 13, fragment c and d, lane 2), suggesting the occurrence of weak binding sites for FL in these promoter segments.

To further probe the sequence preferences for FL binding, I conducted an oligonucleotide competition assay. Several oligonucleotides differing at the CGG triplet and with either 9, 10, or 11 nucleotide spacing for each type of triplet were synthesized. When 0.1 pmole of oligonucleotide representing the FL binding site at position -1498 (control) in the *eas* promoter was labeled and allowed to compete with itself, it was found that about 1.8 pmole of the cold oligonucleotide was required in order to compete for 50% of the DNA probe complex (Fig. 14A (a) and 14B). An oligonucleotide fragment containing the everted repeat, CCG_CGG, with 9, 10, or 11 nucleotide spacing did not compete for binding of FL to the control probe, indicating that an everted repeat is not a binding site for FL (Fig. 14B). A direct repeat, CGG_CGG, was a better competitor than the everted repeat; however, more than 40 pmole was required to compete for 50% of the control sequence (Fig. 14B). A second direct repeat, CCG_CCG, is equivalent to the CGG_CGG repeat except for orientation, thus affecting the

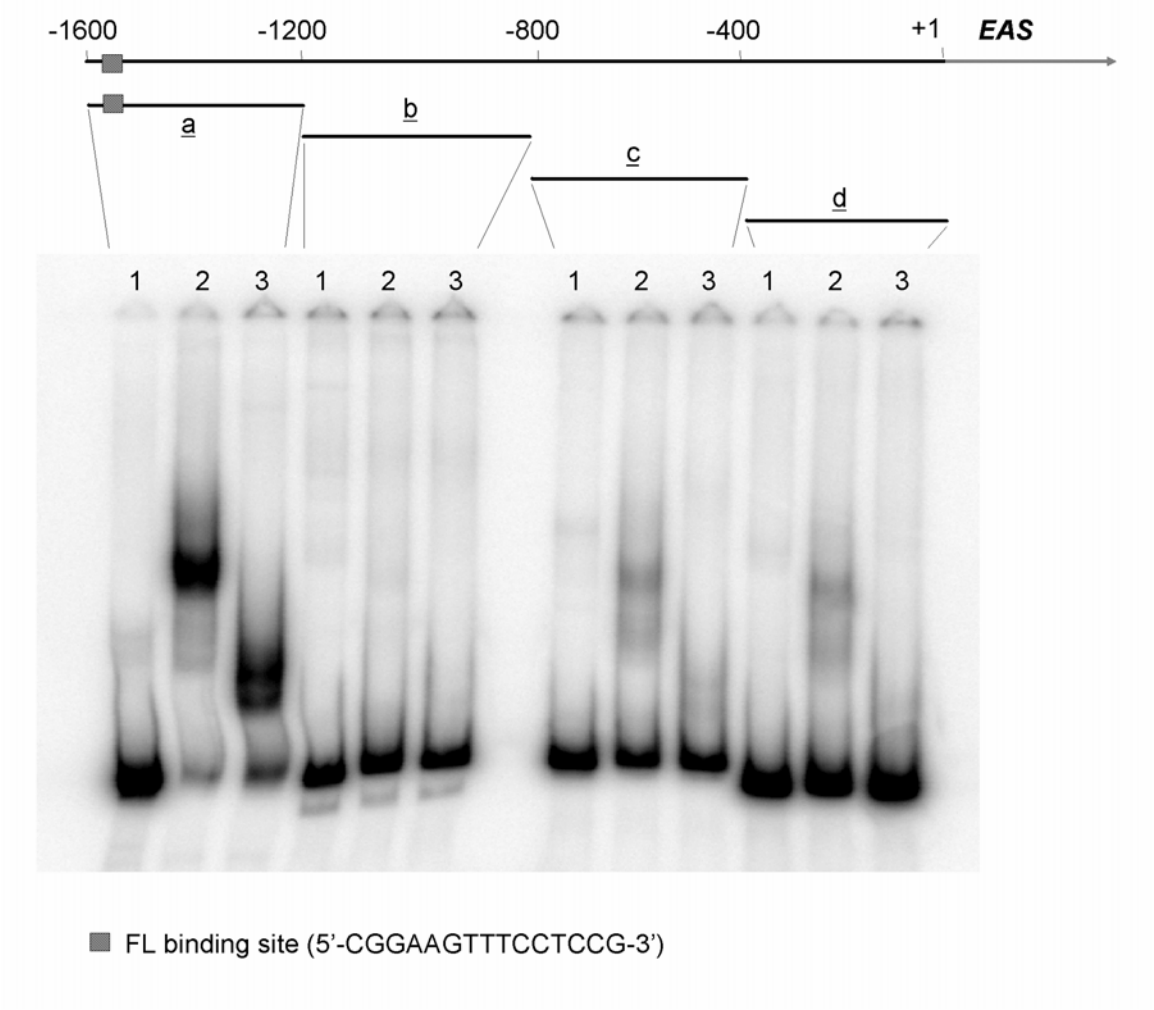


Fig. 13. Localization of FL binding sites in the *eas* promoter. Probes (fragments a to d) were generated by PCR amplification, radiolabeled, and used in EMSA. Positions of the probes are indicated relative to the translational start site of *eas*. Lane 1, probe alone; lane 2, with GST::$FL_{1\text{-}199}$; and lane 3, with $FL_{1\text{-}371H}$.

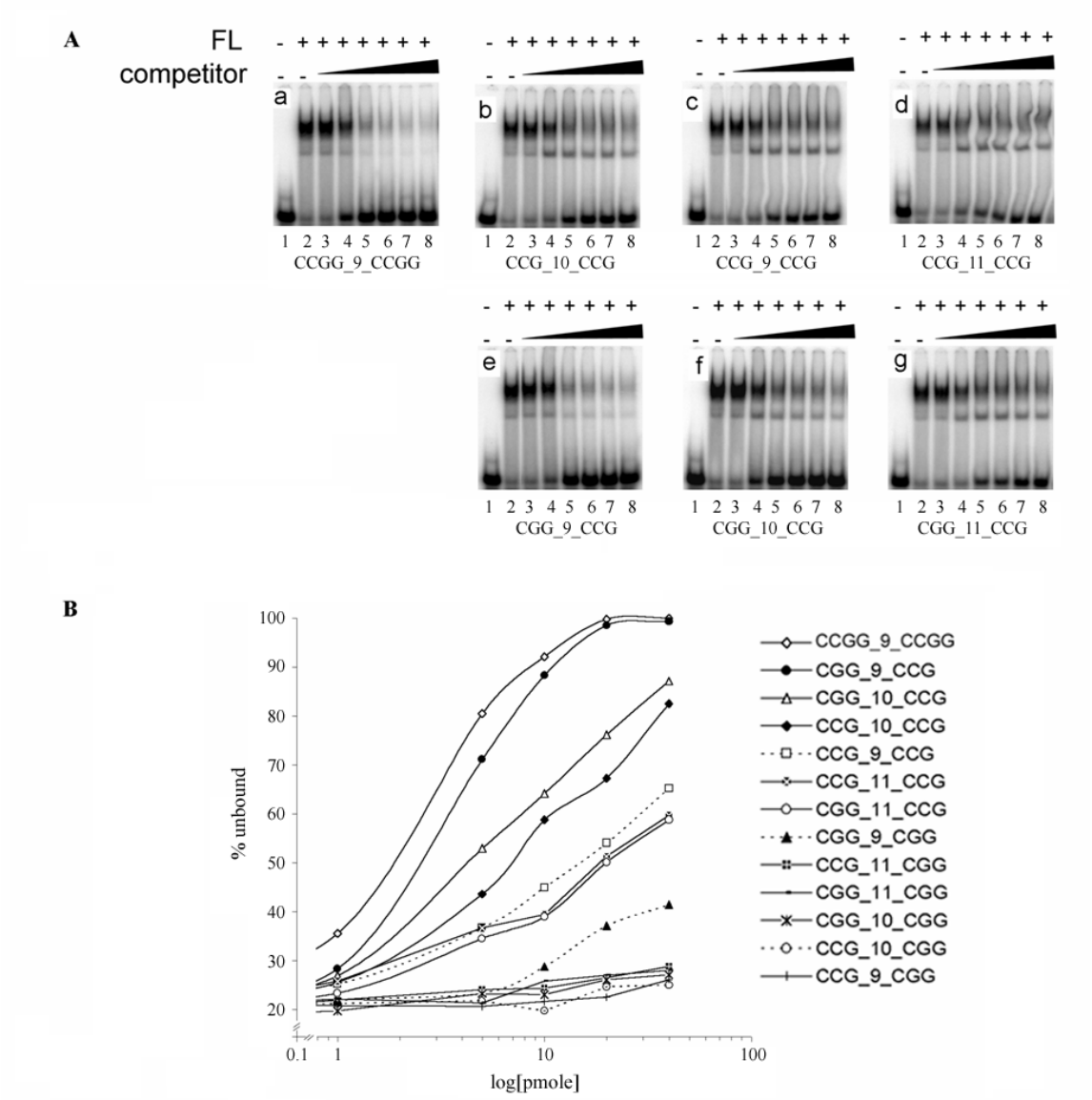


Fig. 14. *In vitro* binding competition assay defining FL binding site preferences. (A) EMSA competition assay. The wild-type control probe was end-labeled and pre-incubated with different amounts of unlabeled oligonucleotide before 1 μg of GST::FL_{1-199} was added to the reaction and analyzed by EMSA. Lane 1, probe alone; lane 2, with GST::FL_{1-199} protein no competitor; lanes 3-8, with indicated competitor at 0.1, 1, 5, 10, 20, and 40 pmole, respectively. a) wild-type sequence (CCGG_CCGG); b-d) direct repeat (CCG_CCG) with 10, 9 and 11 spacing, respectively; e-g) inverted repeat (CGG_CCG) with 9, 10 and 11 spacing, respectively. (B) Plot of the percent of unbound ^{32}P -labeled DNA (Y-axis) against the concentration of unlabeled competitor DNA (X-axis).

orientation of the spacer sequence (Table 3). This sequence was found to be a better competitor with 6.5, 13, and 18 pmole of $CCG(N)_{10}CCG$, $CCG(N)_9CCG$, and $CCG(N)_{11}CCG$, respectively, being required to compete for binding of 50% of the control sequence (Fig. 14A (b, c, and d) and 14B). With the inverted repeat, CGG_CCG, 2.1, 4 and 20 pmole of the 9, 10, and 11 nucleotide spaced triplet oligonucleotides were required, respectively (Fig. 14A (e, f, and g) and 14B). This indicates that the $CGG(N)_9CCG$ inverted repeat is the best competitor and preferred sequence for the FL binding site with a preference order of $CGG(N)_9CCG > CGG(N)_{10}CCG > CCG(N)_{10}CCG > CCG(N)_9CCG > CCG(N)_{11}CCG > CGG(N)_{11}CCG$.

In order to both validate the EMSA results and to further investigate the interaction kinetics of FL protein binding to a strong target sequence, I used surface plasmon resonance (SPR) technology. Biotinylated double-stranded oligonucleotide containing the *eas* -1498 binding sequence was bound to captavidin, a derivative of streptavidin which enables reversible binding of biotin, on the surface of a biosensor chip. The affinity was determined by injecting various concentrations of $GST::FL_{1-199}$ analyte over the ligand-bound sensorchip. The results are presented in the form of a sensorgram (Fig. 15A) that shows the change in resonance units over time and in a plot of the steady-state kinetics (Fig. 15B) that shows the binding response at equilibrium versus the concentration of analyte. The interaction was found to be protein concentration-dependent as indicated by the increase in resonance units in parallel with increasing protein concentration. Binding of $GST::FL_{1-199}$ to a negative control sequence at a high concentration was negligible (data not shown). Analysis of the sensorgram

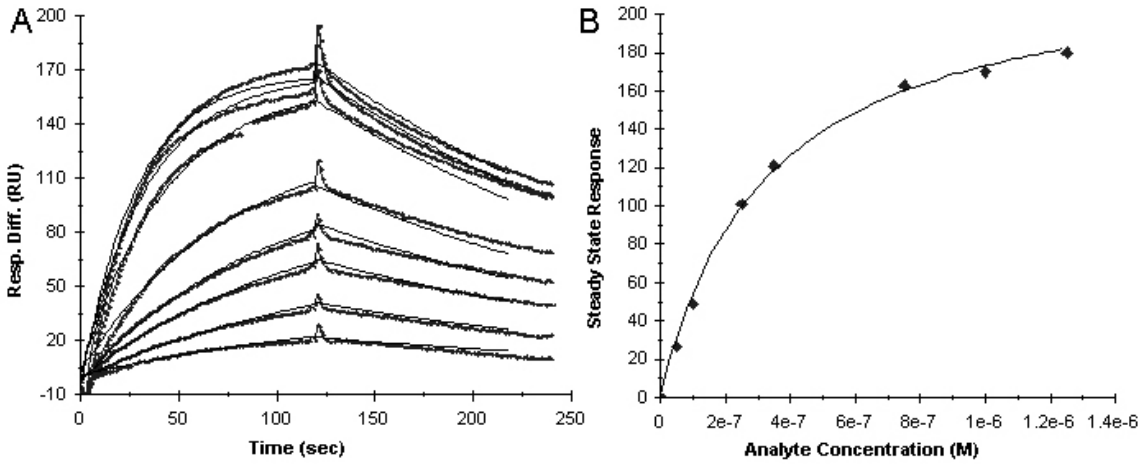


Fig. 15. Measurements of the binding of GST::FL_{1-199} to the consensus sequence found in the *eas* promoter (5'-TTCTCCTGCCCGGAAGTTTCCTCCGGCCATACTCAT-3') using surface plasmon resonance. (A) Sensorgrams of the binding response of GST::FL_{1-199} to the immobilized dsDNA ligand for a concentration series of GST::FL_{1-199} ranging from 50 to 1200 nM. The K_D is 180 nM when determined by a global fit of the kinetic simultaneous k_a/k_d model, assuming 1:1 binding stoichiometry. (B) Plot of the steady-state affinity using Req values determined from local fits of each measured response. The K_D is 330 nM when determined from the steady-state affinity model.

revealed that GST::FL_{1-199} binds to the putative binding sequence in the *eas* promoter with an affinity, as described by the K_D, of approximately 180 nM.

***eas* is regulated by FL during development.** An *in vivo* approach was used to examine the role of the -1498 FL binding site in *eas* regulation. I used a foreign DNA coding sequence, *mpg1* from *M. grisea*, inserted in-frame in the *eas* gene to provide a hybridization tag that will not cross-hybridize with *N. crassa* sequences. I made two constructs to probe the function of the FL binding site at position -1498. A control construct with the full (1660 bp) *eas* promoter was prepared (pRP109), along with a construct truncated just downstream of the FL binding site (1487 bp) (pRP111). These constructs were introduced into three different genetic backgrounds, wild-type (74A), *fl* null mutant (*fl*), and *fl* overexpression (FL-OE) strains. RNA was extracted from mycelia before and after mycelial pads were exposed to air to induce conidiation to assess developmental induction of *eas* expression.

In wild-type strains, *eas* is expressed by 8 hr post-induction (Fig. 16). In contrast, there is no developmental induction of *eas* in the *fl* mutant background. In *fl* overexpression strains, *eas* is constitutively expressed. Similar results were obtained with the pRP109 and pRP111 transformant strains using the *eas* probe; there is developmental induction of *eas* mRNA levels in the wild-type background, lack of expression in the *fl* mutant, and high level constitutive expression in the *fl* overexpression strain (Fig. 16). In the *fl* overexpression strain transformed with the pRP109 construct, the *eas* probe detected both intact *eas* (lower band) and the *eas-mpg1*

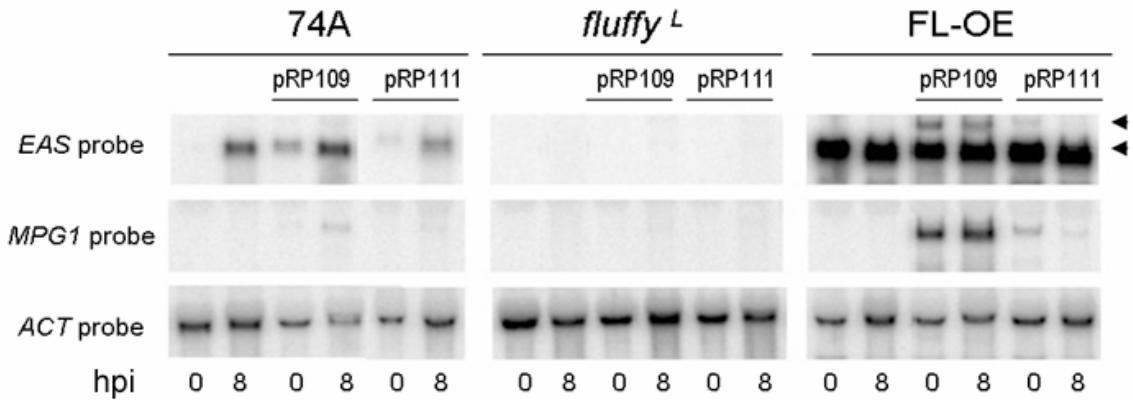


Fig. 16. Northern blot analysis of *eas* and *eas-mpg1* tagged genes. The *mpg1* tagged-gene was transcribed under the control of the *eas* promoter with (pRP109) or without (pRP111) the strong FL binding site at position -1,498. The constructs were introduced into wild-type (74A), *fluffy* mutant (*fl*), and *fluffy* overexpression (FL-OE) strains. RNA was extracted from the mycelial pad after exposure to air for 0 or 8 hours post induction (hpi). Each lane contains 10 μg of total RNA. Full-length *eas* and *mpg1* were used as probes, and *actin* (*act*) was used as a control for mRNA loading. Arrow heads show the *eas-mpg1* transgene (upper band), and native *eas* (lower band) transcripts.

transgene (upper bands). Although transcript levels of the *eas-mpg1* transgene are lower than the endogenous *eas* gene, the level of expression was sufficient to assess expression patterns.

As expected for the *fl* strains, no transgene was detected using *mpg1* as the probe (Fig. 16). For pRP109 in the wild-type background, the transcript for *mpg1* was induced during development, whereas in the *fl* overexpression background there was constitutive expression of the transgene. With the construct that lacked the -1498 binding site for FL (pRP111), weak signals were found in both wild-type and *fl* overexpression backgrounds, suggesting that this site for FL binding plays a major role in activation of *eas* but indicated that weak FL-dependent activation of *eas* remained.

FL contains a transcriptional activation region. Many Zn_2Cys_6 binuclear zinc cluster proteins contain DNA binding motifs at the N-terminus, and a transcriptional activation domain at the C-terminus. To determine the region of FL responsible for transcriptional activation, various fragments of *fl* (Fig. 17) were fused in-frame with the GAL4 DNA binding domain ($GAL4_{BD}$). The chimeric proteins were expressed in *S. cerevisiae* under the control of the *ADH1* promoter from plasmid pGBKT7. These constructs were introduced into the yeast strain AH109, which contains three reporter genes regulated by the GAL4 promoter.

Strains containing FL from amino acid 60 to 780 and 60 to 556 in the $GAL4_{BD}$::FL plasmid constructs both grew on medium lacking histidine and adenine (Fig. 17). These results indicate that FL in both constructs contains a transcriptional expression of the HIS3 and ADE2 genes under control of the GAL4 promoter was not

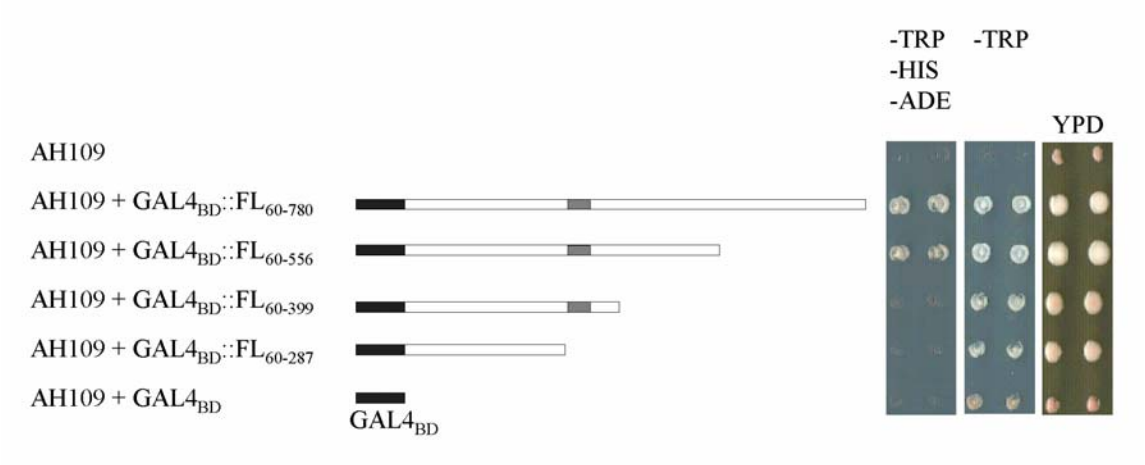


Fig. 17. Localization of a transcriptional activation domain of FL. Yeast strain AH109 contains reporter genes for the biosynthesis of histidine and adenine regulated by GAL4 UAS sequences. This strain was transformed with plasmids coding for the GAL4-DNA binding domain ($GAL4_{BD}$) fused to different region of FL. The chimeric fusion proteins are shown schematically. The portion of FL represented in each construct is shown by subscripts corresponding to the amino acid position in the FL polypeptide. AH109 harboring different constructs of truncated FL were grown on YPD or minimal media with or without supplementation; tryptophan (TRP), histidine (HIS), adenine (ADE), as indicated.

activation domain. However, in the $GAL4_{BD}$::FL_{60-399} and $GAL4_{BD}$::FL_{60-287} constructs sufficient to allow growth, suggesting that a transcriptional activation domain is absent. Together, these experiments indicate that a transcriptional activation domain is located between amino acid residues 400 to 556. A further deletion to residue 287 that removed the middle homology region also failed to provide transcriptional activation.

FL binding sites in the *N. crassa* genome. Because the complete genome sequence of *N. crassa* is available (http://www.broad.mit.edu/annotation/fungi/neurospora/), I searched for potential FL binding sites using two search patterns; CGGA[A/C/G]NN[A/T]NN[C/G/T]TCCG (consensus 1) and CGGA[A/C/G]NN[A/C/T]NNC[C/T]CCG (consensus 2) in the 2000 bp region located upstream of all 10,082 predicted open reading frames (ORFs) of *N. crassa*. These sequences were chosen as the best palindromic sequences matching the consensus and the best overall fit to the consensus, respectively. Using the EMBOSS FUZZNUC program (Olson, 2002), (ftp://ftp.uk.embnet.org/pub/EMBOSS/) I found 74 and 70 predicted ORFs, respectively with 21 ORFs in common (Table 5). One of the common ORFs identified was for the *eas* gene. Including *eas*, eleven of the genes matching consensus 1, five of which also match consensus 2 were amplified by PCR and used as probes for northern blot analysis of transcript levels in *fl*, wild-type and *fl* overexpression strains grown in minimal medium cultures. Transcripts for only four genes were detected; *eas*, NCU00957.1, NCU07288.1, and NCU08093.1 (Fig. 18). Only the *eas* transcript showed obvious regulation by FL.

Table 5. Potential FL binding sites in the *N. crassa* genome

Consensus 1 [a]	Consensus 2 [a]	Sequence [b]	Location of the sequence [b]			NCU number [c]	Detail [d]
			contig#	position	LG		
x		CGGAGAGACAGTCCG	3.6	84884-84870	III, VI	NCU00104.1	heat shock protein CLPA [*Paracoccidioides brasiliensis*]
x	x	CGGACTGTCTCTCCG	3.6	84870-84884	III, VI	NCU00105.1	Nip7p [*Aspergillus fumigatus*], for efficient 60S ribosome subunit biogenesis
x		CGGAGTTAGAGTCCG	3.11	71524-71538	III, VI	NCU00219.1	cytoskeleton actin-modulating protein [*Mus musculus*]
x		CGGAGTGTTGTTCCG	3.12	93410-93424	III, VI	NCU00246.1	set of predicted proteins
	x	CGGAGGGTAACCCCG	3.13	124740-124754	III, VI	NCU00397.1	GDSL-motif lipase/hydrolase protein [*Arabidopsis thaliana*]
	x	CGGAACACAGCCCCG	3.16	24090-24104	III, VI	NCU00486.1	predicted protein
	x	CGGACACCCACTCCG	3.24	44848-44834	I	NCU00712.1	3-hydroxy-3-methylglutaryl-coenzyme A reductase [*Penicillium citrinum*]
x [e]		CGGAGTGAGATTCCG	3.38	54387-54401	I	NCU00957.1	extracellular matrix protein precursor [*Fusarium oxysporum f. sp. lycopersici*]
	x	CGGAACACGTCTCCG	3.41	124935-124921	V	NCU01042.1	set of hypothetical proteins [*Neurospora crassa*]
x [f]	x	CGGACTCTCTCTCCG	3.44	688-702	V	NCU01154.1	NsdD (GATA-factor) [*Emericella nidulans*]
x	x	CGGACTGATGCTCCG	3.45	117336-117350	V	NCU01192.1	probable DNA-directed DNA polymerase III [*Schizosaccharomyces pombe*]
x [f]	x	CGGAATCTTTCTCCG	3.57	33583-33569	V	NCU01427.1	geranylgeranyl pyrophosphate synthetase, *al-3* [*N. crassa*]

Table 5. Continued

Consensus 1 [a]	Consensus 2 [a]	Sequence [b]	Location of the sequence [b]			NCU number [c]	Detail [d]
			contig#	position	LG		
	x	CGGACCATACCCCCG	3.80	15325-15339	II, V	NCU01830.1	4-hydroxyphenylpyruvate dioxygenase [*Magnaporthe grisea*]
x		CGGAAGCAAGGTCCG	3.87	38233-38219	I	NCU01971.1	predicted protein
x		CGGACCTTGCTTCCG	3.87	38219-38233	I	NCU01972.1	predicted protein
x		CGGACCGTATTTCCG	3.88	79881-79867	I	NCU02008.1	predicted protein
	x	CGGACTTCTGCTCCG	3.89	18036-18050	I	NCU02028.1	set of two hypothetical proteins [*N. crassa*]
x		CGGAAATTTGGTCCG	3.95	38563-38577	I	NCU02111.1	myosin I myoA, heavy chain [*E. nidulans*]
x [f]		CGGAGGTTCCGTCCG	3.101	48997-48983	I	NCU02160.1	hypothetical protein (AF385833) RAC1 [*Rattus norvegicus*]
x	x	CGGACGGAACCTCCG	3.101	48983-48997	I	NCU02161.1	serine rich hypothetical protein [*S. pombe*]
	x	CGGACCACTCCTCCG	3.104	6239-6225	VII	NCU02202.1	serine/threonine kinase IRE1 [*Hypocrea jecorina*]
	x	CGGAGTCCGCCTCCG	3.108	201371-201357	VII	NCU02301.1	predicted protein
	x	CGGAGCGAATCCCCG	3.112	18337-18351	VII	NCU02347.1	unnamed protein product [*Podospora anserina*]
x		CGGAGCCAGCTTCCG	3.113	21192-21206	VII	NCU02355.1	unnamed protein product [*P. anserina*]
x		CGGAGGCAGTTTCCG	3.117	13666-13680	VII	NCU02417.1	N-acetyltransferase [*S. pombe*]

Table 5. Continued

Consensus 1 [a]	Consensus 2 [a]	Sequence [b]	Location of the sequence [b]			NCU number [c]	Detail [d]
			contig#	position	LG		
x		CGGAATTACATTCCG	3.138	42058-42044	I	NCU02659.1	predicted protein
x	x	CGGAGCCAAGCTCCG	3.138	170040-170054	I	NCU02694.1	hypothetical protein [*P. anserina*]
x	x	CGGAGGGTTTCTCCG	3.144	30495-30509	I	NCU02792.1	predicted protein, 51 amino acids
x		CGGAGGAAGGGTCCG	3.148	5012-5026	I	NCU02824.1	hypothetical protein B14D6.400 [*N. crassa*]
	x	CGGACTCTTACCCCG	3.150	44897-44883	I	NCU02852.1	cytochrome P450 [*Coprinopsis cinerea*]
	x	CGGAAACAAGCCCCG	3.161	27379-27365	I	NCU03075.1	predicted protein, 28 amino acids
x		CGGAAATAATTTCCG	3.172	38176-38190	I	NCU03306.1	predicted protein
x		CGGACGGACTTTCCG	3.181	100-114	II, V	NCU03453.1	set of hypothetical proteins [*N. crassa*]
	x	CGGACGACTCCTCCG	3.199	18705-18691	V	NCU03594.1	predicted protein, 88 amino acids
	x	CGGAGCCAAGCCCCG	3.199	44319-44305	V	NCU03601.1	set of two hypothetical proteins [*N. crassa*]
x [f]		CGGAAGCAGGGTCCG	3.200	42166-42152	V	NCU03661.1	probable GTPase activating protein [*N. crassa*]
x		CGGACCCTGCTTCCG	3.200	42152-42166	V	NCU03663.1	set of two hypothetical proteins [*N. crassa*]
	x	CGGAAACACCCCCCG	3.201	11784-11798	V	NCU03669.1	pmt2 rRNA methyltransferase [*S. pombe*]

Table 5. Continued

Consensus 1 [a]	Consensus 2 [a]	Sequence [b]	Location of the sequence [b]			NCU number [c]	Detail [d]
			contig#	position	LG		
x		CGGAGACAGTGTCCG	3.201	247870-247884	V	NCU03740.1	set of two conserved hypothetical proteins [*N. crassa*]
	x	CGGAGTCCAACTCCG	3.203	16443-16457	V	NCU03792.1	probable inositol polyphosphate phosphatase [*S. pombe*]
x [f]	x	CGGAATCTTTCTCCG	3.203	70593-70579	V	NCU03809.1	related to RNA polymerase II transcriptional regulation mediator MED6 [*N. crassa*]
x		CGGAGAAAGATTCCG	3.203	70579-70593	V	NCU03811.1	hypothetical protein
	x	CGGACGGCCGCCCCG	3.212	136852-136838	I, VI	NCU03935.1	probable homoserine dehydrogenase [*N. crassa*]
	x	CGGAAGCCGGCCCCG	3.223	38186-38172	V	NCU04316.1	carboxypeptidase; kex1 [*E. nidulans*]
	x	CGGACGGCAACCCCG	3.228	155557-155543	IV, VII	NCU04419.1	ethylene-forming enzyme [*Pseudomonas syringae*]
	x	CGGAATACTCCTCCG	3.228	195819-195833	IV, VII	NCU04431.1	putative secreted glucosidase [*Streptomyces coelicolor* A3 (2)]
	x	CGGAATCATCCCCCG	3.228	226144-226130	IV, VII	NCU04439.1	probable ATP-dependent RNA helicase [*A. fumigatus*]
	x	CGGAATGCCTCTCCG	3.230	89469-89483	IV, VII	NCU04486.1	predicted protein
x		CGGAGTGAATGTCCG	3.231	11473-11459	IV, VII	NCU04506.1	predicted protein
x	x	CGGACATTCACTCCG	3.231	11459-11473	IV, VII	NCU04507.1	ATP(GTP)-binding protein [*Homo sapiens*]
	x	CGGACCACTGCCCCG	3.239	8280-8294		NCU04599.1	predicted protein, 58 amino acids

Table 5. Continued

Consensus 1 [a]	Consensus 2 [a]	Sequence [b]	Location of the sequence [b]			NCU number [c]	Detail [d]
			contig#	position	LG		
	x	CGGAAAGCGGCTCCG	3.264	11604-11590	VI	NCU04824.1	predicted protein
x	x	CGGAAGTTCTCTCCG	3.266	57218-57204		NCU04862.1	short-chain dehydrogenase/reductase family protein [*A. thaliana*]
x		CGGAGAGAACTTCCG	3.266	57204-57218		NCU04864.1	predicted protein, 74 amino acids
x [f]	x	CGGAGGGTGACTCCG	3.269	11936-11922		NCU04883.1	endochitinase precursor [*Stachybotrys elegans*]
	x	CGGAAACCTTCCCCG	3.271	25539-25525		NCU04896.1	predicted protein
	x	CGGACCCAGACCCCG	3.282	8369-8355	II, VI	NCU05002.1	predicted protein, 46 amino acids
	x	CGGAGCTCCACTCCG	3.286	82112-82098	II, VI	NCU05065.1	predicted protein, 86 amino acids
x	x	CGGACTTTCTCTCCG	3.286	82867-82853	II, VI	NCU05065.1	predicted protein, 86 amino acids
x		CGGAGACATCGTCCG	3.295	7482-7496	IV	NCU05237.1	hypothetical protein [*S. pombe*]
	x	CGGAACCCAACTCCG	3.311	132532-132546	VI	NCU05521.1	pathogenicity protein PATH531 [*Magnaporthe grisea*]
x		CGGAGCCATTGTCCG	3.312	148443-148457	VI	NCU05583.1	predicted protein
x		CGGAATTAGAGTCCG	3.322	22838-22852	III	NCU05717.1	Mitochondrial ribosomal protein MRP17 [*Saccharomyces cerevisiae*]

Table 5. Continued

Consensus 1 [a]	Consensus 2 [a]	Sequence [b]	Location of the sequence [b]			NCU number [c]	Detail [d]
			contig#	position	LG		
x	x	CGGACCAAGTCTCCG	3.332	53499-53485	VII	NCU05861.1	liver regeneration-related protein Ab1-133 [*Rattus norvegicus*]
x		CGGAGACTTGGTCCG	3.332	53485-53499	VII	NCU05862.1	Major Facilitator Superfamily (MSF) transporter [*S. pombe*]
x		CGGACATTTCTTCCG	3.332	104640-104626	VII	NCU05879.1	predicted protein, 28 amino acids
	x	CGGACAGCTTCTCCG	3.337	10322-10308		NCU05928.1	predicted protein
x		CGGAATAAGGGTCCG	3.338	53230-53244		NCU05947.1	predicted protein
x	x	CGGACCTTTTCTCCG	3.364	5851-5837	III	NCU06267.1	predicted protein
x		CGGAGAAAAGGTCCG	3.364	5837-5851	III	NCU06268.1	vacuolar protein sorting-associated protein [*S. pombe*]
	x	CGGAGACCTCCTCCG	3.366	66173-66159	IV	NCU06305.1	Major Facilitator Superfamily (MSF) transporter [*Beauveria bassiana*]
x		CGGAGTGATTTTCCG	3.366	162578-162564	IV	NCU06333.1	translocation protein sec62 [*Yarrowia lipolytica*]
x	x	CGGAAAATCACTCCG	3.366	162564-162578	IV	NCU06334.1	predicted protein
x		CGGAAGAAGGGTCCG	3.367	1337-1351	IV	NCU06336.1	N2,N2-dimethylguonasine tRNA methyltransferase [*S. pombe*]
	x	CGGACTGATCCCCCG	3.371	57028-57014	III, V	NCU06427.1	hypothetical protein with serine/threonine protein kinase domain [*N. crassa*]

Table 5. Continued

Consensus 1 [a]	Consensus 2 [a]	Sequence [b]	Location of the sequence [b]			NCU number [c]	Detail [d]
			contig#	position	LG		
x	x	CGGACCAACTCTCCG	3.367	19200-19214	IV	NCU06520.1	probable ATP-dependent RNA helicase [*S. pombe*]
	x	CGGAAGGACTCCCCG	3.381	49950-49964	IV	NCU06588.1	predicted protein
x		CGGAAGATAAGTCCG	3.386	16371-16357	V	NCU06664.1	translin-like protein [*S. pombe*]
x		CGGACTTATCTTCCG	3.386	16357-16371	V	NCU06665.1	related to 4-coumarate-CoA ligase [*N. crassa*]
	x	CGGAAACCTCCTCCG	3.386	33562-33576	V	NCU06669.1	set of predicted proteins
	x	CGGACTGTTGCCCCG	3.387	39021-39035	V	NCU06689.1	set of predicted proteins
x		CGGACGCAATTTCCG	3.415	11557-11571	VI	NCU07119.1	peptide synthetase [*Aureobasidium pullulans*]
x		CGGAACAAGAGTCCG	3.421	26545-26531	IV	NCU07258.1	putative protein [*A. thaliana*]
x [e]		CGGACGGACTTTCCG	3.422	31296-31282	IV	NCU07288.1	predicted protein
x		CGGAAAGTCCGTCCG	3.422	31282-31296	IV	NCU07289.1	transcription initiation factor TFIIB [*S. pombe*]
	x	CGGACAGCGACTCCG	3.422	54530-54544	IV	NCU07296.1	Serine/threonine-protein kinase COT-1 [*N. crassa*]
x		CGGACGGATATTCCG	3.423	150377-150363	IV	NCU07343.1	integral membrane protein [*Clostridium acetobutylicum*]
x		CGGAGGAAATTTCCG	3.426	18949-18963		NCU07359.1	predicted protein

Table 5. Continued

Consensus 1 [a]	Consensus 2 [a]	Sequence [b]	Location of the sequence [b]			NCU number [c]	Detail [d]
			contig#	position	LG		
	x	CGGAATTTATCCCCG	3.435	24497-24511	I	NCU07464.1	set of predicted proteins
x		CGGAATCTAATTCCG	3.448	66915-66901	III	NCU07601.1	set of predicted proteins
	x	CGGAAATATGCCCCG	3.450	17033-17019	III	NCU07612.1	predicted protein
	x	CGGACCTTTGCCCCG	3.450	24026-24012	III	NCU07614.1	predicted protein
	x	CGGAACGCAACTCCG	3.461	101795-101781	IV	NCU07712.1	putative ATP-dependent RNA helicase [*A. fumigatus*]
	x	CGGAGCCCCACCCCG	3.466	46548-46534		NCU07782.1	predicted thioesterase [*Vibrio vulnificus* CMCP6]
x		CGGAGTATCTTTCCG	3.468	29714-29700		NCU07805.1	selenocysteine lyase [*Thermobifida fusca*]
x		CGGAGCCTGCTTCCG	3.475	14312-14298	III	NCU07879.1	protein involved in mitochondrial iron accumulation; Mmt2p [*S. cerevisiae*]
x [f]	x	CGGAAGCAGGCTCCG	3.475	14298-14312	III	NCU07880.1	protein kinase NPKA [*E. nidulans*]
x		CGGACAATCATTCCG	3.481	126844-126830	IV	NCU08005.1	ferredoxin-NADP reductase [*S. pombe*]
x		CGGAATGATTGTCCG	3.481	126830-126844	IV	NCU08006.1	predicted protein
x		CGGACCATCTTTCCG	3.484	145050-145036	I	NCU08092.1	NIPSNAP1 protein, vesicular transport protein [*Danio rerio*]
x [e]		CGGAAAGATGGTCCG	3.484	145036-145050	I	NCU08093.1	V-type ATPase; ATP synthase j chain, mitochondrial [*S. pombe*]

Table 5. Continued

Consensus 1 [a]	Consensus 2 [a]	Sequence [b]	Location of the sequence [b]			NCU number [c]	Detail [d]
			contig#	position	LG		
x [e]	x	CGGAAGTTTCCTCCG	3.510	34034-34020	II	NCU08457.1	rodlet protein, EAS, *eas*/*bli-7* gene product [*N. crassa*]
	x	CGGAATAATACCCCG	3.520	51135-51121	III	NCU08576.1	predicted protein
x		CGGAATAAACGTCCG	3.531	122517-122503	I	NCU08642.1	PpaB [*Danio rerio*], contains F-box, and cNMP binding domains
	x	CGGAGGACATCCCCG	3.537	62365-62379	IV	NCU08685.1	phosphoribosylformylglycinamidine synthase [*S. pombe*]
x	x	CGGACATTTGCTCCG	3.541	62206-62192	II	NCU08714.1	predicted protein
x		CGGAGCTATCGTCCG	3.548	39726-39712	IV	NCU08836.1	set of predicted proteins
x	x	CGGACCCTCGCTCCG	3.548	64782-64768	IV	NCU08843.1	predicted protein, 96 amino acids
	x	CGGAGACCTGCTCCG	3.565	13410-13424		NCU09022.1	set of predicted proteins
	x	CGGACCCAAGCCCCG	3.565	21106-21092		NCU09023.1	laccase 2 [*Botryotinia fuckeliana*]
	x	CGGAAACCAGCCCCG	3.566	31078-31092		NCU09054.1	predicted protein
x	x	CGGAGAAACTCTCCG	3.598	10569-10583		NCU09392.1	predicted protein, 74 amino acids
	x	CGGAGCGCGCCCCCG	3.603	11385-11399		NCU09451.1	predicted protein
x		CGGAGGCAATTTCCG	3.607	12488-12474	II	NCU09493.1	predicted protein, 39 amino acids

Table 5. Continued

Consensus 1 [a]	Consensus 2 [a]	Sequence [b]	Location of the sequence [b]			NCU number [c]	Detail [d]
			contig#	position	LG		
	x	CGGACGCAGCCCCCG	3.633	20940-20926		NCU09703.1	alpha-acetolactate decarboxylase [*Methanosarcina mazei* Goe1]
	x	CGGAGCGCCGCTCCG	3.666	16329-16343	I	NCU09805.1	alpha amylase (Taka Amylase) (E.C.3.2.1.1)
x	x	CGGAAACAGACTCCG	3.666	32668-32654	I	NCU09808.1	dynamin-related protein DNM1 [*S. cerevisiae*]
x		CGGAGTCTGTTTCCG	3.666	32654-32668	I	NCU09809.1	very hypothetical protein [*S. pombe*]
x		CGGAGTCTGTTTCCG	3.666	32654-32668	I	NCU09810.1	succinyl-CoA ligase [GDP-forming] alpha-chain
x		CGGACTTTCTTTCCG	3.679	51538-51524		NCU09887.1	glucose-6-phosphate/phosphate translocator-related [*A. thaliana*]
	x	CGGAGCTATACCCCG	3.689	49655-49641		NCU09966.1	set of predicted proteins
	x	CGGACGCCCGCTCCG	3.706	24399-24385		NCU09992.1	K12H4.7 protein [*Caenorhabditis elegans*]
x		CGGACTTAAAGTCCG	3.773	6472-6458		NCU10067.1	part of 26S proteasome complex [*S. cerevisiae*]

[a] Two search patterns represent consensus 1 (CGGA[A/C/G]NN[A/T]NN[C/G/T]TCCG) and consensus 2 (CGGA[A/C/G]NN[A/C/T]NNC[C/T]CCG) were used to screen 2 kb upstream ORFs of *N. crassa* database (http://www.broad.mit.edu/annotation/fungi/neurospora/) using EMBOSS FUZZNUC program. The searches identified 74 and 70 predicted ORFs for consensus 1 and consensus 2, respectively.

[b] Sequence, location [contig number, position on the contig, position on linkage group (LG)], and ORF number (NCU#) are listed.

[c] NCU number from *N. crassa* database.

[d] Best match derived from BLASTP search at *e*-value at least 10^{-5}.

[e] Northern blot analysis with signal, see Fig. 18 for the pattern.

[f] Northern blot analysis with no signal.

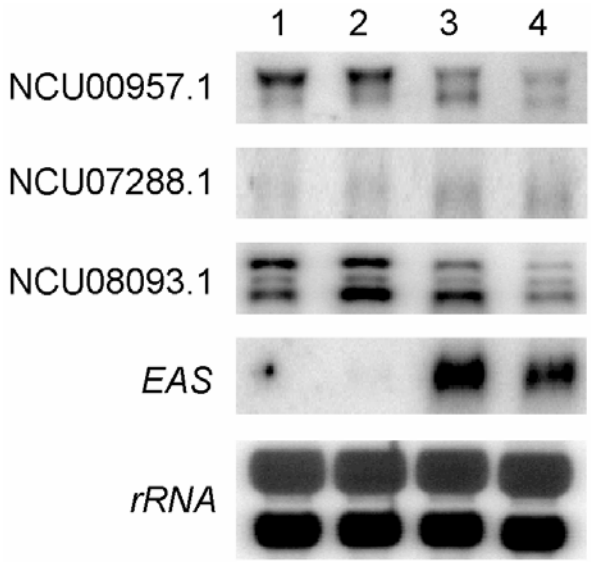


Fig. 18. Northern blot analysis of genes containing putative binding sites for FL. Lanes 1, 2, and 4 were the RNA from *fl* mutant, wild-type, and *fl* overexpression strain cultured in Vogel's liquid culture for 18 hr, respectively. Lane 3 was the RNA from *fl* overexpression strain cultured in Vogel's liquid culture for 12 hr. No signal was obtained for NCU01154.1, NCU01427.1, NCU02160.1, NCU03661.1, NCU03809.1, NCU04883.1, and NCU07880.1.

DISCUSSION

The *fl* gene has long been known to be necessary for conidial morphogenesis in *N. crassa* (Matsuyama *et al.*, 1974; Springer and Yanofsky, 1989). The gene product was also reported to be essential for the production of hydrophobin rodlet protein, the product of the *eas* gene (Bailey-Shrode and Ebbole, in press; Lauter *et al.*, 1992). Subsequent cloning and sequencing of *fl* showed that it encodes an 88 kDa polypeptide that contains a Zn_2Cys_6 binuclear zinc cluster DNA binding domain, suggesting that FL acts as a transcription factor (Bailey and Ebbole, 1998). Here I show that FL is capable of binding to the DNA sequence $CGG(N)_9CCG$ and possesses a functional transcriptional activation domain. I also demonstrate that FL binds to the *eas* promoter and that *eas* is regulated directly by FL.

The binding sites of several other binuclear zinc cluster transcription factors have been identified, and most of them bind to CGG triplets separated by 0 to 18 nucleotides (Schjerling and Holmberg, 1996; Todd and Andrianopoulos, 1997). By using the randomized oligonucleotide selection technique described by Pierrou *et al*. (Pierrou *et al.*, 1995), it was found that FL recognized the inverted CGG motif separated by 9 bp with additional sequence preferences indicated by alignment of 53 binding sites containing this motif. Over 150 oligonucleotide fragments that co-purified with FL were sequenced and many of these did not contain obvious FL binding sites. I conclude that these additional sequences either contain weak FL binding sites or simply were contaminants obtained with the FL-DNA complexes. Supporting this conclusion, it was found that FL weakly bound to the oligonucleotides containing the $CGG(N)_{10\text{-}11}CGG$

and $CCG(N)_{9\text{-}11}CCG$ as shown in the oligonucleotide competition experiment. In addition, FL may also weakly bind to the other sequences resembling these binding sites. Including these weaker binding sites, 72 of the 157 oligonucleotides have FL binding sites. Increasing the number of selection cycles and/or the stringency of the binding and washing steps during the selection might improve the selection for strong consensus binding sites of FL.

In a search of the *N. crassa* genome sequence, I found that there are almost 9,000 regions in the *N. crassa* genome that match the $CGG(N)_9CCG$ pattern. About 4,600 sites are located within a 2.0 kb region upstream of predicted open reading frames (ORFs). Based on our analysis of FL binding sequences, a more refined search pattern was used to prioritize ORFs containing putative binding sites for FL. These predicted proteins are involved in a wide range of functions suggesting that *fl* regulates a much broader range of genes than just those involved in conidiation, or that there are a large number of promoters containing FL binding sites that are not subject to regulation by FL. Only a small fraction of the promoter regions containing FL binding sites are associated with ORFs represented in the available *N. crassa* EST dataset (Nelson *et al.*, 1997), suggesting that many of these genes are expressed at low levels in the several conditions used to generate libraries for EST analysis. A survey of 10 of the novel genes by northern blot analysis revealed that seven were expressed at low levels (not detected) and three additional transcripts were detected but were not convincingly regulated by FL. Although this is not a comprehensive analysis, it does not encourage a further effort to characterize expression of genes containing putative FL binding sites in their

promoter regions by northern blot analysis. However, future microarray analysis with representatives of all predicted genes in the genome will be valuable in identifying genes that are regulated by FL for correlation to the occurrence of FL binding sites.

A binding site for FL in the *eas* promoter is located at 1498 bp upstream of the translation initiation codon of *eas*. The EMSA and SPR experiments (Fig. 13 and Fig. 15) presented here demonstrate that FL specifically binds to this sequence. *In vivo* expression analysis (Fig. 16) indicates that this sequence is a direct target regulated by FL. Removing of the FL binding site (-1498) from the *eas* promoter results in a significant loss of the expression of *eas-mpg1* transgene products. The *eas* mutation in *N. crassa* strain UCLA191 was generated by ethyl methane sulfonate treatment (Selitrennikoff, 1976). It was later found that the mutation in this strain resulted from an insertion between 1000 and 1500 bp upstream of the transcription initiation site (Lauter *et al.*, 1992). Furthermore, a deletion of this region of the promoter results in the loss of *eas* transcript in response to nutrient starvation (Bell-Pedersen *et al.*, 1996; Kaldenhoff and Russo, 1993). The binding site for FL in the *eas* promoter that I have identified is located in this region. Thus, the loss of *eas* expression in the UCLA191 allele likely results from the elimination of the FL binding site from the *eas* promoter region. In previous studies (Bell-Pedersen *et al.*, 1996; Kaldenhoff and Russo, 1993), induction of *eas* expression during development was observed in constructs containing 1900 bp upstream of the transcription initiation site. A construct that deleted DNA upstream of position -625 was not developmentally regulated but had a high basal level of expression. This demonstrated the presence of a negative element for *eas* expression

during mycelial growth but perhaps prevented identification of the *fl* binding site as a positive element for developmental induction of *eas*.

I noted that there was a low level of residual FL-dependent expression of the *eas-mpg1* transgene in the strains with the pRP111 construct (lacking the FL binding site). In the EMSA binding to the *eas* promoter (Fig. 13), I observed that there was weak binding of FL to promoter fragments c and d. Thus, it is likely that FL can weakly bind to sequences in these regions allowing weak induction of transcription. These two promoter fragments contain sequences, 5'-TGGAGGTTTGCGCCG-3' and 5'-CGGAGAACTTGGCCA-3', respectively, that resemble FL binding sites. It has been reported for other binuclear zinc cluster transcription factors that the factor can synergistically bind to more than one site in the promoter in order to regulate target genes. For example; AlcR, a transcription factor controlling ethanol metabolism in *A. nidulans* (Panozzo *et al.*, 1997), can bind to three different sites in the *alcA* promoter to regulate the expression of alcohol dehydrogenase. Deletion of one of these sites resulted in decreased expression, but not complete loss, and all of the sites are required in order to fully activate *alcA*.

As observed for many other binuclear zinc cluster proteins, the DNA binding and activation domains of FL are located at the amino and carboxy terminus, respectively. The binuclear zinc cluster resides in amino acids 11-37 and at least one potent transcriptional activation domain is located in the C-terminal half of the protein between amino acids 400-556. One class of activation domains consists of an abundance of acidic amino acids (Ptashne, 1988). I found that approximately 14% of the amino acids in the

400-556 region of FL are acidic. This region may activate target genes by recruiting the transcriptional machinery to the promoter, or may promote a protein-protein interaction with another protein that is involved in regulation, as has been observed in Gal4p (Lue *et al.*, 1987; Melcher and Johnston, 1995).

Binuclear zinc finger transcription factors have thus far only been found in ascomycete and basidiomycete fungi (Todd and Andrianopoulos, 1997), where they regulate a wide variety of functions. In *N. crassa*, there are at least 76 predicted ORFs that contain an apparent fungal zinc binuclear domain (http://mips.gsf.de/proj/neurospora/). Among these, only five proteins have been studied in detail. NIT4 is a pathway-specific regulatory protein involving nitrate utilization (Yuan *et al.*, 1991), QA-1F is an activator of quinic acid metabolism (Geever *et al.*, 1989), ACR-2 is a regulator for acriflavine resistance (Akiyama and Nakashima, 1996), ACU-15 is a regulator for acetate utilization (Bibbins *et al.*, 2002), and FL (Bailey and Ebbole, 1998). FL has greatest similarity to NIT4 protein (Bailey and Ebbole, 1998); however, the similarities are restricted in the DNA binding and coiled-coil domains (30% identity), and middle homology region (21% identity). With the complete genome sequence of *N. crassa* I identified a hypothetical protein (NCU09205.1) that has a higher percent identity to FL; 47% in the DNA binding and coiled-coil domain and 36% identity in the middle homology region. As this is the best candidate for a paralogous gene to FL, it will be of interest to determine if this protein has any role related to development.

It has been shown that overexpression of *fl* is sufficient to induce conidial development and expression of *eas*, a gene induced early in the conidiation process. In addition to its role in conidiation, *fl* also appears to play a role in ensuring progression of development. *fl* mutants form minor constriction chains but not major constriction chains (Matsuyama *et al.*, 1974) and *fl* is highly expressed at the time that major constriction chains are formed (Bailey and Ebbole, 1998). Thus, FL appears to play two roles, one involved in induction of gene expresssion early in development and an additional role in the transition from minor to major constriction chains.

CHAPTER IV

GENE EXPRESSION PROFILING OF *fl* MUTANT AND *fl* OVEREXPRESSION STRAINS

INTRODUCTION

The fungus *Neurospora crassa* reproduces through sexual and asexual sporulation cycles. The sexual cycle results in the production of dark brown to black ascospores; whereas the asexual cycle gives rise to uninucleate microconidia and multinucleate macroconidia (Davis, 2000). The major means of fungal distribution occurs by the dispersal of spores throughout the environment. Macroconidiation (here after, conidiation) can be synchronously induced by certain environmental cues, such as nutrient deprivation, carbon dioxide levels, exposure to air or blue light, and circadian rhythm (Ebbole, 1996; Springer, 1993; Turian and Bianchi, 1972). The process includes i) the development of aerial hyphae ii) repeated apical budding of the aerial hyphae to form minor constriction chains iii) the formation of major constriction chains from the minor constrictions and iv) the ultimate establishment of double crosswalls between each proconidium (Springer, 1993). Finally, the conidia undergo further biochemical maturation over several days that is necessary for subsequent germination (Springer and Yanofsky, 1989).

Several developmental mutants that block the pathway of conidiophore morphogenesis have been isolated. The formation of minor constriction chains is

blocked in the *aconidiate-2* (*acon-2*) and *fluffyoid* (*fld*) mutants. *acon-3* and *fluffy* (*fl*) mutants can form minor but not major constriction chains. Mutation at the double crosswall maturation step occurs in *conidial separation* mutants, *csp-1* and *csp-2* (Ebbole, 1996; Springer, 1993). Despite the number of mutant strains that have been identified, only *fl* gene has been cloned. This gene has been shown to encode a transcription factor that contains a binuclear zinc finger DNA binding motif similar to the one found in the Gal4p transcription factor family (Bailey and Ebbole, 1998).

Transcription factors in the Gal4p family recognize CGG triplets in the target genes they regulate. Specificity among different Gal4p family members is determined by the types of the triplet each transcription factor recognizes and binds. The repeats have been classified according to their orientation as inverted repeat (CGG_CCG), direct repeat (CGG_CGG), or everted repeat (CCG_CGG). In addition, variation in the sequence of nucleotides between the triplets confers further specificity (Reece and Ptashne, 1993; Schjerling and Holmberg, 1996; Todd and Andrianopoulos, 1997).

In vitro binding assays were used to identify a consensus binding site of FL. The binding site was determined to consist of an inverted CGG repeat with nine nucleotide spacing, 5'-CGGA(A/C/G)NN(A/C/T)NNC(C/T)CCG-3' (Chapter III). The binding site, 5'-CGGAAGTTTCCTCCG-3', was located 1498 bp upstream of the start codon of the *easily wettable* (*eas*) gene, an allele of the *clocked-control gene-2* (*ccg-2*) and *blue-light inducible-7* (*bli-7*) (Bell-Pedersen *et al*., 1992; Kaldenhoff and Russo, 1993). *eas* codes for hydrophobin rodlet protein that coats the outside of mature conidia, making conidia hydrophobic, and thus easy to disseminate by air.

Supporting its role in conidiation, *fl* transcript is present at a low basal level in mycelia, and accumulates to higher level in aerial hyphae during conidial development (Bailey and Ebbole, 1998). To address the role of *fl* in conidiation, the gene was overexpressed under control of a constitutive promoter. It was found that overexpression of *fl* induced conidiation in submerged liquid culture (Bailey-Shrode and Ebbole, in press). Similar observations were made during investigation of the well-studied C_2H_2 zinc finger transcription factor, *bristleA* (*brlA*), which plays a central role in the conidiation process in *Aspergillus nidulans* (Adams *et al*., 1998). Unlike in *A. nidulans* where genetic and molecular studies have identified several target genes that are affected by *brlA*, only one direct target of FL, the *eas* gene, has been found (Chapter III).

A number of genes that are differentially expressed during conidiation in *N. crassa* have been isolated and characterized (Berlin and Yanofsky, 1985; Corrochano *et al*., 1995; Roberts and Yanofsky, 1989; White and Yanofsky, 1993). These genes are known as *con*idiation specific (*con*) genes. They are highly induced during conidiation and transcript levels decline shortly after conidia start to germinate.

Having identified binding site for FL in the *eas* gene, other developmentally regulated genes were examined for the presence of the putative binding site for FL. In addition to *eas*, *con-6*, and *con-10* are not expressed in the *fl* mutant (Bailey and Ebbole, 1998), and thus were investigated as potential targets of FL. However, the consensus binding site of FL was found only in the *eas* promoter and not in *con-6* and *con-10*, indicating that these two genes are most likely not direct targets regulated by *fl*. Furthermore, a search of the *N. crassa* genome sequence for the putative binding site of

FL was performed. Using the EMBOSS FUZZNUC program (Olson, 2002), (ftp://ftp.uk.embnet.org/pub/EMBOSS/), a 2 kb region of promoter from the 10,082 ORFs of *N. crassa* were analyzed for the presence of a strong FL binding site. The search revealed 123 ORFs containing a putative binding site for FL (Table 5, Chapter III). Northern blot analysis, however, revealed that only four out of eleven ORFs showed expression, and among these, only *eas* showed regulation by FL.

Thus far, little is yet known about the genes involved in conidial morphogenesis of *N. crassa*. Further advancement in this field of study may be facilitated by the recent development of cDNA microarray technology, a powerful tool for examining the expression of thousands of genes in a single hybridization. This approach has been utilized to identify differential gene expression and thereby correlate gene activity with particular steps in biological pathways. It has also proven useful in the discovery of novel target genes and biological processes (Schulze and Downward, 2001). With continued improvement, the technique has been used widely for a broad spectrum of organisms and cell lines, including human, rat, mouse, fruit fly, plant, yeast, filamentous fungi, bacteria, tumor cells, etc. The technique is also valuable for the study in *N. crassa*, which has the advantage of a considerable library of Expressed Sequence Tags (ESTs) and a completely sequenced genome (Nelson *et al*., 1997, http://www-genome.wi.mit.edu/annotation/fungi/neurospora). Microarray technology has been successfully employed in *N. crassa* for expression profiling and identification of novel genes that participate in; for example; glucose starvation, nutrient-dependent variation, and circadian rhythm pathways (Aign and Hoheisel, 2003; Correa *et al.*, 2003; Lewis *et*

al., 2002; Xie, 2003). cDNA microarrays were used in this study to examine expression profiles and to identify target and/or downstream genes that are regulated by *fl*.

In this study, mRNA levels in a *fl*-overexpression (FLOE) strain were compared to the levels in a *fl* mutant strain using cDNA microarrays composed of 1,343 unique genes, representing approximately 13% of the total *N. crassa* gene content (Lewis *et al.*, 2002). The cDNA used in the arrays were derived from EST libraries obtained from *Neurospora* genome project at the University of New Mexico. These libraries were generated from mRNA that was isolated from germinating conidia, growing mycelium, and perithecial tissues (Nelson *et al.*, 1997). Microarray analysis revealed that 122 unique genes, or about 9% of the genes represented on the arrays, were differentially expressed at least two-fold between FLOE and *fl* strain. Of these 122 genes, 51 were up-regulated in response to overexpression of *fl*, while 71 genes were down-regulated. Some of the developmentally-regulated genes were identified in the up-regulated group. Among the genes that were up-regulated, *eas* displayed the greatest level of response to the allelic state of *fl*. Furthermore, expression profiling revealed that *fl*-induced cells displayed a pattern of gene expression suggestive of a higher rate of sugar consumption that led to ethanol production and utilization at an earlier time than in wild-type or *fl* mutant cells. This is consistent with previous physiological characterization of development (Turian and Bianchi, 1972).

MATERIALS AND METHODS

Fungal strains, and growth conditions. The *N. crassa* wild-type strain 74OR23-1V*A* (FGSC 2489), ORS-6*a* (FGSC 2490), and fl^{L} *a* mutant (FGSC 46) were obtained from the Fungal Genetics Stock Center (FGSC; Department of Microbiology, University of Kansas Medical Center). The *fl* overexpression strain bearing pRP5 (FLOE) is described in Chapter II. For microarray experiments, conidia of FLOE or mycelia of *fl* mutant were inoculated in 10 ml of Vogel's medium with 2% sucrose (Davis and de Serres, 1970). The cultures were grown at 34^{o}C, 200 rpm, for 18 hr. The mycelia were blended, transferred into 100 ml of Vogel's media, and grown under the same condition for 18 hr. The mycelia were harvested and frozen in liquid nitrogen. These steps were performed in triplicate and the resulting RNAs were combined for each strain to help reduce variation that may occur between experiments.

Microarray hybridization, scanning, and analysis. Total RNA from three triplicate cultures were extracted (Madi *et al*., 1994) and subsequently combined into one sample for each strain. Poly(A)$^{+}$ RNA was purified using a PolyATrack kit (Promega). Using 300 ng of each mRNA sample as template, cDNA was synthesized and labeled in a two-step method with 3DNA^{TM} Submicro EX Expression Array Detection Kit (Genisphere). The two-channel labeling probes, cy3 and cy5, were hybridized to the microarray, which encompassed 1,343 genes of *N. crassa* (Lewis *et al*., 2002), according to the manufacturer's instructions. Following the hybridization and washing steps, the intensities of the cy3 and cy5 signals were scanned (GSI

Luminomics), and data were collected using ScanAlyze software (http://rana.lbl.gov/EisenSoftware.htm).

Expression data analysis was performed using GeneSpring software (Silicon Genetics). Initially, raw data from each spot was subtracted from the background value, followed by intensity-dependent normalization using the locally weighted scatter plot smooth (LOWESS) normalization at 20% percentile for smoothing. Genes whose intensities were very similar to the overall background levels were filtered out, and the resulting genes that passed this criterion were analyzed further.

Northern blot analysis. RNA was electrophoresed on a 1.5% agarose formaldehyde gel. Northern blot and hybridization were performed using Zeta-probe blotting membrane (BioRad) as described by manufacturer's manual. DNA probes were prepared by PCR either directly from the EST clones using universal primers (T3 and T7), or from genomic DNA using specific primers. Probes were radiolabeled with [α-^{32}P]dCTP using the Rediprime II system (Pharmacia). The signal was analyzed using Fuji BAS-1800II Phosphoimager.

Overexpression and RIP mutation. To construct the plasmid for overexpress a NCU09049.1, a PCR product using primer PS162 (5'-CACCGAATTCAATCAATCAC-3'), and PS163 (5'-CAGTCTAGATTCCATGAGC-3') with *N. crassa* DNA as a template was generated and digested with *Eco*RI and *Xba*I. The fragment was directionally inserted into plasmid pRP67 (Chapter II). For NCU08751.1, PCR product using primers PS165 (5'-GTTCTAGACTATTATGCC-3'), and PS166 (5'-TACCTGCAGCTAACCATC -3') was digested with *Xba*I and *Pst*I and cloned into

pRP67. In these constructs NCU09049.1 and NCU08751.1 are under the regulation of the *RP27* promoter. The constructs were introduced into the wild-type strain, 74OR23-1V*A*. Southern analysis was performed as described (Sambrook and Russell, 2001). To generate RIP mutations (Selker *et al.*, 1987), transformant strains carrying an extra copy of the target gene were crossed with the wild-type and at least one hundred progeny from each cross were analyzed for conidiation phenotype. Restriction isoschizomer enzyme digestion with *Sau*3AI and *Dpn*II were used to detect DNA mutation and methylation.

RESULTS AND DISCUSSION

Comparing differential gene expression between FLOE and *fl* mutant. To identify genes that are regulated by *fl*, RNA was isolated from the FLOE strain in which *fl* was under the control of the *cpc-1* promoter (Chapter II) and from the fl^{L} strain, a strain carrying a deletion of the DNA binding domain of FL (Bailey and Ebbole, 1998; Matsuyama *et al.*, 1974). Gene expression in FLOE was compared with that of the *fl* mutant strain using cDNA microarrays which contain of 1,343 unique genes (Galagan *et al.*, 2003; Lewis *et al.*, 2002). The cDNA from FLOE and *fl* mutant was synthesized from RNA isolated from submerged cultures. Only vegetative hyphae were observed in the *fl* mutant but conidiation was induced in the FLOE strain (Bailey and Ebbole, 1998). To minimize variation between samples of the same strain, a pool of total RNA derived from three different cultures was used to synthesize the cDNA probe. Three independent cDNA probes were synthesized and used in three independent arrays. A total of 1,066 out of 1,791 probe spots passed the filtering criteria and were used for expression

analysis. The correlation coefficient between replicates was 0.712. A correlation coefficient of 0.8 and greater is considered to have high reproducibility. Therefore, my experiments display some variation that is due to technical variability rather than biological variation since the same RNA samples were used in replicates. This technical variation results from differences in DNA spots among different microarrays, different efficiencies in labeling reactions, differences in hybridization and washing, etc. The variation observed in the microarray data was such that it was necessary to confirm the microarray results using northern blot analysis (see below). Analysis of the microarray data indicated that 159 probe spots were at least two-fold differentially expressed in response to overexpression of *fl* (Table 6). Some of the genes are represented by more than one cDNA clone on the array. Analysis showed that these 159 clones represent a total of 122 unique genes, which accounted for approximately 9% of the genes represented on the microarray. The predicted open reading frames (ORFs) from the *N. crassa* gene predictions corresponding to the cDNA clones were compared against the available protein databases using BLASTP. Only alignments with an *e*-value less than or equal to $1x10^{-6}$ were annotated. Of the 122 unique ORFs, 48 genes are hypothetical or have no known homolog. An additional 74 genes have sequence similarity to proteins involved in a broad range of functional categories as shown in Table 6. Of the 159 differential clones, 79 clone spot pairs were up-regulated, and 80 clones were down-regulated in response to the overexpresson of the *fl* transcription factor. The 79 up-regulated clones correspond to 51 unique genes, whereas the 80 down-regulated clones represent 71 unique genes.

To confirm the microarray results, northern blot analyses of 26 cDNA probes were performed. The inserts from the EST clones were amplified using T3 and T7 universal primers. The PCR products were used to probe membranes which contained RNA from *fl* mutant, wild-type (74A), and FLOE growing in liquid culture for 18 hours (Table 7; lane 1, 2, and 4, respectively). RNA from an earlier time point, 12 hours, from FLOE was also included (Table 7, lane 3). RNA from *fl* mutant (lane 1) and FLOE (lane 4) are the same RNA samples that were used for the microarray experiments. In all cases, the northern blot analyses with FLOE (Table 7, lane 4) and *fl* mutant RNA (Table 7, lane 1) correlated well with the regulation pattern obtained from microarray analysis. These results indicate that the data obtained from microarray analysis is reliable despite the correlation coefficient being lower than desired.

Patterns of gene expression. Based on both microarray and northern blot results, I grouped 21 genes into three expression pattern groups (Table 7). Four genes were not detected by northern blot analysis and were not considered further. Group 1 represents RNAs for twelve genes that were up-regulated in the FLOE strain relative to both *fl* mutant and wild type. Of the 79 clones up-regulated in response to *fl* overexpression, eleven of the most highly responsive clones represented the *eas* (*ccg-2*) gene, which showed an 8 to 53-fold induction (Table 6). The *t*-test *p*-value among these 11 probe spots derived from the statistical analysis software from GeneSpring was less than 0.001, suggesting that the change in expression of this gene is significant and indicating that the expression of *eas* is *fl*-dependent. This result is consistent with the fact that *eas* is directly regulated by *fl* as shown in Chapter III. Three additional genes

Table 6. Genes at least 2 fold differentially-expressed in *fl* mutant and FLOE strains

Category/ ORF ID [a]	EST ID [b] or clone name	Blast match [c]	Expression ratio [d] FLOE/*fl*$^{-}$	t-test *p*-value [e]
Development				
NCU08457.1	CCG-2 1/8	hydrophobin (EAS), (*N. crassa*)	53.0	1.05E-08
NCU08457.1	CCG-2.7	hydrophobin (EAS), (*N. crassa*)	41.5	2.64E-06
NCU08457.1	CCG-2 1/2	hydrophobin (EAS), (*N. crassa*)	39.0	2.55E-06
NCU08457.1	SC2A2	hydrophobin (EAS), (*N. crassa*)	22.9	7.82E-04
NCU08457.1	CCG-2.8	hydrophobin (EAS), (*N. crassa*)	19.0	2.50E-05
NCU08457.1	CCG-2.6	hydrophobin (EAS), (*N. crassa*)	17.6	1.41E-04
NCU08457.1	CCG-2.4	hydrophobin (EAS), (*N. crassa*)	16.8	4.89E-04
NCU08457.1	CCG-2.5	hydrophobin (EAS), (*N. crassa*)	16.1	1.37E-04
NCU08457.1	CCG-2.1	hydrophobin (EAS), (*N. crassa*)	13.5	1.21E-03
NCU08457.1	NC5E3	hydrophobin (EAS), (*N. crassa*)	11.6	2.92E-05
NCU08457.1	CCG-2	hydrophobin (EAS), (*N. crassa*)	8.0	6.39E-06
NCU08457.1	SC1E10	CCG-2 [f]	34.0	8.12E-05
NCU08457.1	NM1C3	CCG-2 [f]	27.7	4.35E-04
NCU08457.1	NC1A11	CCG-2 [f]	27.3	3.90E-06
NCU08769.1	pCON 6	conidiation specific protein-6 (CON-6), (*N. crassa*)	18.0	1.22E-05
NCU08769.1	CON-6	conidiation specific protein-6 (CON-6), (*N. crassa*)	17.1	4.46E-05
NCU07325.1	CON-10	conidiation specific protein-10 (CON-10), (*N. crassa*)	4.1	4.96E-02
NCU08726.1	pRP6 Fluffy	Fluffy, (*N. crassa*)	5.0	7.11E-03
NCU01517.1	NM6B8	similar to glucan 1,4-alpha-glucosidase (*S. cerevisiae*)	4.1	6.46E-04
NCU08907.1	SC5B12	blastomyces yeast phase-specific protein 1 (*Ajellomyces dermatitidis* strain 81)	2.7	3.69E-04
NCU08907.1	SC1D4	blastomyces yeast phase-specific protein 1 (*Ajellomyces dermatitidis* strain 81)	2.5	1.04E-03
NCU08791.1	NM2E10	catalase A (*Podospora anserine*)	2.3	1.37E-02

Table 6. Continued

Category/ ORF ID [a]	EST ID [b] or clone name	Blast match [c]	Expression ratio [d] FLOE/*fl*$^{-}$	t-test *p*-value [e]
Central metabolism				
Glycolysis				
NCU01528.1	CCG-7	glyceraldehyde-3-phosphate dehydrogenase (CCG-7), (*N. crassa*)	0.42	1.76E-03
NCU01528.1	NC1G4	glyceraldehyde-3-phosphate dehydrogenase (CCG-7), (*N. crassa*)	0.41	3.04E-02
NCU01528.1	NC3C3	glyceraldehyde-3-phosphate dehydrogenase (CCG-7), (*N. crassa*)	0.39	2.22E-02
NCU06075.1	NC4C8	pyruvate kinase (ACE-8), (*N. crassa*)	0.48	9.41E-02
NCU06075.1	SC2A6	pyruvate kinase (ACE-8), (*N. crassa*)	0.25	1.51E-02
NCU08807.1	NM1B6	carbon catabolite repressor (CRE-1), (*N. crassa*)	0.52	5.94E-02
NCU09041.1	W10D10	NADP-dependent mannitol dehydrogenase (*Cladosporium fulvum*)	2.8	5.33E-05
Fermentation				
NCU02476.1	NP6D7	alcohol dehydrogenase I and III (*A. nidulans*)	6.4	1.96E-06
NCU01754.1	SM2D6	alcohol dehydrogenase I and III (*A. nidulans*)	3.4	6.92E-03
NCU09285.1	W01C6	alcohol dehydrogenase (*S. pombe*)	3.7	4.55E-06
NCU09285.1	NC3D8	alcohol dehydrogenase (*S. pombe*)	3.2	8.42E-05
TCA cycle and electron transport system				
NCU10008.1	NM1B3	fumarase (*S. cerevisiae*)	0.39	6.91E-02
NCU00959.1	NM7H6	succinate dehydrogenase (ubiquinone) iron-sulfur protein precursor (*S. cerevisiae*)	4.1	6.18E-04
NCU03031.1	SC5F1	succinate dehydrogenase (ubiquinone) iron-sulfur protein precursor (*S. pombe*)	2.6	4.61E-03
NCU08940.1	NM4C4	ubiquinol-cytochrome c reductase complex (complex III), (*S. pombe*)	3.5	1.15E-04

Table 6. Continued

Category/ ORF ID [a]	EST ID [b] or clone name	Blast match [c]	Expression ratio [d] FLOE/*fl*$^{-}$	t-test *p*-value [e]
NCU06606.1	SC7F10	ubiquinol-cytochrome c reductase complex (complex III), (*S. pombe*)	3.3	1.69E-05
NCU06606.1	NM7D8	ubiquinol-cytochrome c reductase complex (complex III), (*S. pombe*)	2.5	2.46E-04
Lipid metabolism				
NCU06783.1	SP1B2	probable ATP citrate lyase, subunit 2 (*Sordaria macrospora*)	0.46	3.23E-02
NCU03492.1	NP6F6	probable fatty acid hydroxylase (*S. cerevisiae*)	0.40	2.84E-02
NCU08976.1	W07H5	fatty acid elongation enzyme (*Isochrysis galbana*)	0.37	3.86E-02
NCU05259.1	SM4H12	delta-9 fatty acid desaturase (*Ajellomyces capsulatus*)	0.36	9.78E-03
NCU09770.1	NC4G6	acetyl-CoA hydrolase (ACU-8), (*N. crassa*)	0.28	2.90E-02
Amino acid metabolism				
NCU01757.1	NM5A10	similar to L-asparaginase (*S. pombe*)	5.4	4.10E-06
NCU00461.1	W01E10	NAD-glutamate dehydrogenase (GDH-1), (*N. crassa*)	3.2	2.73E-04
NCU03118.1	W17D3	putative saccharopine dehydrogenase Lysine forming enzyme (*A. nidulans*)	0.45	3.34E-02
NCU01666.1	NC5G2	acetolactate synthase (*S. cerevisiae*)	0.44	1.12E-01
NCU04050.1	W01D5	cross-pathway control protein 1 (CPC-1), (*N. crassa*)	0.30	1.06E-02
NCU08162.1	W17B9	argininosuccinate lyase (*Fusarium oxysporum*)	0.22	1.34E-02
Transporter				
NCU02887.1	NP2B6	probable K+ channel protein (*S. pombe*)	6.3	5.83E-05
NCU08858.1	SM1B7	probable alpha-glucoside transporter (*N. crassa*)	4.1	3.77E-04
NCU01470.1	NM5H5	putative vacuolar protein (*S. pombe*)	3.2	1.22E-02
NCU00667.1	W08A9	V-type ATPase subunit c' (*S. cerevisiae*)	0.45	2.09E-02
NCU08332.1	SP4G12	probable V-ATPase, 20K chain (*M. grisea*)	0.39	1.80E-02

Table 6. Continued

Category/ ORF ID [a]	EST ID [b] or clone name	Blast match [c]	Expression ratio [d] FLOE/*fl*$^{-}$	t-test *p*-value [e]
NCU08332.1	NC1F8	probable V-ATPase, 20K chain (*M. grisea*)	0.34	2.53E-02
NCU09698.1	W17H1	probable membrane transporter (*S. pombe*)	0.29	3.45E-02
NCU01065.1	W10D1	probable ammonium transporter (*A. nidulans*)	0.20	2.57E-02
Cytoskeleton				
NCU01204.1	NC4E9	related to tropomyosin TPM1 (*N. crassa*)	0.42	2.12E-02
NCU04054.1	NC3D5	tubulin beta chain (*N. crassa*)	0.41	2.31E-02
NCU04054.1	NC2E3	tubulin beta chain (*N. crassa*)	0.37	2.53E-02
NCU09488.1	SC2E4	related to SDA1, actin organization protein (*N. crassa*)	0.39	3.07E-02
NCU06397.1	NC3C2	profilin involved in actin polymerization (*S. pombe*)	0.37	2.05E-02
NCU02555.1	NM4E10	actin related protein (*S. pombe*)	0.27	1.59E-02
NCU01883.1	NM7F2	probable RVS167 protein, a component of actin cytoskeleton (*N. crassa*)	0.23	1.69E-02
Cell wall synthesis				
NCU07253.1	NM7B6	1,3-beta glucanosyltransferase (*Aspergillus fumigatus*)	2.1	4.51E-03
NCU07523.1	NP4B12	1,3-beta glucanase (*Trichoderma atroviride*)	2.0	3.14E-03
NCU05969.1	W07G9	endoglucanase IV precursor (*Hypocrea jecorina*)	0.40	6.60E-02
NCU02668.1	W06F9	cell wall synthesis protein (*Penicillium chrysogenum*)	0.27	2.25E-02
Signaling				
NCU01613.1	SP1F10	related to Shk1 kinase-binding protein (*N. crassa*) [f]	5.5	9.20E-05
NCU06249.1	W13G8	probable serine-threonine protein kinase (*S. cerevisiae*)	0.34	1.93E-02
NCU02738.1	NP4B8	calcium-binding protein (*Homo sapien*)	0.23	1.94E-02
Protein modification				
NCU06603.1	NC3C12	intracellular protease (*Xylella fastidiosa*)	4.0	1.09E-05
NCU02549.1	W06E1	mitochondrial processing peptidase (*N. crassa*)	2.5	2.43E-03

Table 6. Continued

Category/ ORF ID [a]	EST ID [b] or clone name	Blast match [c]	Expression ratio [d] FLOE/*fl*$^{-}$	t-test *p*-value [e]
NCU02635.1	SM2A8	probable mannosyltransferase for complex glycosylation (*Pichia angusta*)	2.3	6.74E-06
NCU04370.1	SP6F9	ubiquitin-activating enzyme (*S. cerevisiae*)	0.54	5.97E-02
NCU02289.1	NM1B5	ubiquitin conjugation enzyme (*Colletotrichum cingulata*)	0.42	1.03E-01
NCU05592.1	NP5B11	ubiquitin conjugating enzyme (*S. pombe*)	0.42	9.92E-02
NCU08605.1	SM2E10	proteasome component precursor (*S. pombe*)	0.45	2.26E-02
NCU02059.1	W07A10	aspartyl (acid) proteases (*Sordaria macrospora*)	0.41	6.76E-02
Stress reponse				
NCU09534.1	SM2F5	glutathione peroxidase (*Blumeria graminis*)	0.30	3.46E-03
NCU09559.1	CCG-9	probable trehalose synthase (CCG-9), (*N. crassa*)	0.29	5.82E-03
NCU06110.1	NC3B5	thiazole biosynthetic enzyme (*Fusarium solani*)	0.28	2.70E-02
NCU06031.1	W09E10	rehydrin protein homolog (*Candida albicans*)	0.26	1.76E-02
NCU07024.1	W08A7	osmotic sensitive-2 (OS-2), (*N. crassa*)	0.23	2.55E-02
Transcription/Translation process				
NCU01317.1	SC5G10	similar to 60S ribosomal protein L12 (*Rattus norvegicus*)	0.40	2.66E-02
NCU01438.1	NP5B7	probable nucleosome assembly protein (*N. crassa*)	0.39	2.86E-02
NCU07420.1	NC1D6	translation initiation factor eIF-4A (*N. crassa*)	0.39	2.65E-02
NCU07420.1	NP5E5	translation initiation factor eIF-4A (*N. crassa*)	0.36	1.55E-02
NCU03047.1	SC5G7	transcription initiation factor TFIIF subunit (*S. pombe*)	0.17	2.10E-02
Unknown function				
NCU09040.1	SM2B10	probable NADP-dependent oxidoreductase, (*S. pombe*)	16.0	2.89E-04
NCU09049.1	NM5G10	predicted protein	14.7	3.30E-03
NCU07267.1	SM4C4	BLI-3 protein (*N. crassa*)	11.0	3.05E-07
NCU08130.1	SM2H2	predicted protein	6.4	1.70E-07

Table 6. Continued

Category/ ORF ID [a]	EST ID [b] or clone name	Blast match [c]	Expression ratio [d] FLOE/*fl*^{L}	t-test *p*-value [e]
NCU08949.1	NM5B10	hypothetical protein	6.1	7.22E-08
NCU08949.1	NM1B8	hypothetical protein	3.0	6.90E-04
NCU08949.1	W08H11	hypothetical protein	2.7	1.48E-02
NCU08949.1	SM1H1	hypothetical protein	2.7	6.91E-03
NCU09917.1	SP1A8	hypothetical protein	6.1	3.05E-06
NCU02470.1	SP1C4	possible to PR protein, fruiting body precursor (*A. fumigatus*)	5.6	6.20E-07
NCU02470.1	SP1D5	related to PR protein, fruiting body precursor (*A. fumigatus*)	2.7	3.76E-04
NCU06977.1	NM5C1	predicted protein	5.5	4.93E-05
NCU06977.1	NP6G4	predicted protein	4.1	5.25E-05
NCU05768.1	SP4D12	predicted protein	4.8	4.43E-04
NCU03753.1	CCG-1.5	glucose-repressible gene protein (GRG-1), (*N. crassa*)	4.8	2.92E-05
NCU03753.1	NM6B6	glucose-repressible gene protein (GRG-1), (*N. crassa*)	4.1	6.01E-04
NCU03753.1	CCG-1.2	glucose-repressible gene protein (GRG-1), (*N. crassa*)	3.1	3.71E-03
NCU03753.1	NP5G7	glucose-repressible gene protein (GRG-1), (*N. crassa*)	3.1	3.04E-03
NCU03753.1	SP6C9	glucose-repressible gene protein (GRG-1), (*N. crassa*)	2.8	5.96E-03
NCU03753.1	CCG-1.1	glucose-repressible gene protein (GRG-1), (*N. crassa*)	2.0	1.40E-03
NCU04816.1	SP1D9	predicted protein	4.5	2.80E-04
NCU00922.1	SP6H10	predicted protein	4.1	5.67E-06
NCU00922.1	NM3F2	predicted protein	2.1	2.95E-02
NCU00595.1	NM3A8	hypothetical protein	3.5	8.08E-05
NCU08037.1	SP6B4	predicted protein	3.4	6.07E-02
NCU05812.1	SM1H5	predicted protein	3.2	1.11E-04
NCU09521.1	NP5B8	hypothetical protein	3.1	1.33E-03
NCU00432.1	W01D7	hypothetical protein	3.0	8.01E-06
NCU06688.1	SM1H4	predicted protein	2.9	3.12E-03
NCU08847.1	NP4B6	hypothetical protein	2.8	1.80E-03

Table 6. Continued

Category/ ORF ID [a]	EST ID [b] or clone name	Blast match [c]	Expression ratio [d] FLOE/*fl*^{L}	t-test *p*-value [e]
NCU07287.1	NP2E4	predicted protein	2.6	3.25E-05
NCU04667.1	SM1A4	predicted protein	2.2	4.74E-02
NCU02474.1	SP6H7	predicted protein	2.1	1.52E-03
NCU04493.1	SC3E5	predicted protein	2.1	4.59E-05
NCU05754.1	NM7A6	predicted protein	2.0	4.03E-02
NCU03205.1	NC3C8	predicted protein	1.6	4.35E-01
NCU00576.1	W08E5	hypothetical protein	0.54	8.95E-02
NCU07100.1	NP2C12	hypothetical protein	0.51	1.11E-01
NCU06910.1	SP1A6	hypothetical protein	0.51	5.13E-02
NCU03126.1	SM2D4	predicted protein	0.49	7.14E-02
NCU02736.1	NC5G4	predicted protein	0.48	7.51E-02
NCU01586.1	NM4B4	hypothetical protein	0.47	2.60E-02
NCU01297.1	NM7F1	predicted protein	0.44	2.32E-02
NCU07618.1	NM9B3	predicted protein	0.43	3.08E-02
NCU09355.1	SP1C12	predicted protein	0.42	4.63E-02
NCU00729.1	W06G5	predicted protein	0.41	2.03E-02
NCU00573.1	SM2G7	hypothetical protein	0.40	1.17E-01
NCU01418.1	CCG-6	clock-controlled gene-6 protein (CCG-6), (*N. crassa*)	0.39	5.34E-02
NCU01418.1	NC5G12	clock-controlled gene-6 protein (CCG-6), (*N. crassa*)	0.29	1.62E-02
NCU01418.1	NC4G5	clock-controlled gene-6 protein (CCG-6), (*N. crassa*)	0.20	1.29E-02
NCU06076.1	W07A7	hypothetical protein	0.38	2.04E-02
NCU09223.1	NC3C7	hypothetical protein	0.37	3.11E-02
NCU09931.1	SM2C9	hypothetical protein	0.37	2.41E-02
NCU00866.1	W09G9	predicted protein	0.36	1.58E-02
NCU03074.1	SP4E7	predicted protein	0.36	1.44E-02
NCU09133.1	W01H9	predicted protein	0.35	5.30E-02
NCU00935.1	SP4A10	predicted protein	0.35	4.90E-02
NCU04931.1	NP6A12	predicted protein	0.31	8.79E-02

Table 6. Continued

Category/ ORF ID [a]	EST ID [b] or clone name	Blast match [c]	Expression ratio [d] FLOE/*fl*[-]	t-test *p*-value [e]
NCU05881.1	NM1E5	hypothetical protein	0.31	2.09E-02
NCU07222.1	W10D8	predicted protein	0.30	3.81E-02
NCU04928.1	W01D8	predicted protein	0.27	3.95E-02
NCU05938.1	NM5G12	predicted protein	0.31	4.47E-02
NCU05023.1	NM5E8	predicted protein	0.28	2.66E-02
NCU07110.1	W08F10	predicted protein	0.27	2.27E-02
NCU06416.1	W13B2	probable oxidoreductase (*Pseudomonas aerugenosa*)	0.23	1.46E-02
NCU08621.1	W17G5	predicted protein	0.23	1.33E-02
NCU00811.1	W10D9	related to Allegen Aas f 4 (*A. fumigatus*)	0.18	1.34E-02
NCU04931.1	NM6A4	predicted protein	0.17	1.86E-02
NCU08751.1	NP3F7	hypothetical protein	0.11	1.17E-02

[a] The predicted ORFs derived from http://www-genome.wi.mit.edu/annotation/fungi/neurospora.

[b] The ESTs of *N. crassa* are from the *Neurospora* genome project at the University of New Mexico.

[c] The Blast match was performed with the BLASTP search at NCBI database.

[d] The expression ratio is the average value from three replicates.

[e] The t-test *p*-value derived from GeneSpring software.

[f] Four EST clones which did not show signal by northern blot analysis.

Table 7. Gene expression patterns identified from northern blot analysis

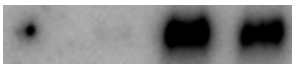
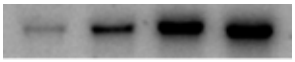
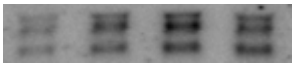
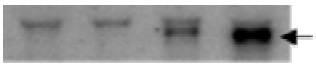
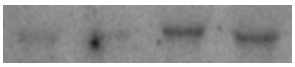
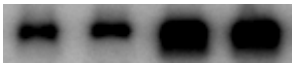
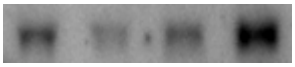
Group/ Gene or EST [a]	Expression ratio (microarray) [b]	Expression ratio (northern blot) [b]	Northern blot analysis 1[c] 2 3 4
Group 1			
eas	53.0	46.8	
NM2E10	2.3	13.3	
con-6	17.1	3.0	
con-10	4.1	1.7	
fl	5.0	3.2	
NM7H6	4.1	2.1	
SC7F10	3.3	2.5	
NC3D8	3.2	3.8	
NM5G10	14.7	9.2	
SM1A4	2.2	1.8	
NM1B8	3.0	5.1	
NM5C1	5.5	2.6	

Table 7. Continued

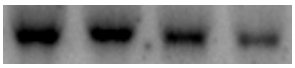
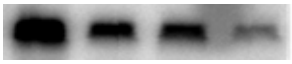
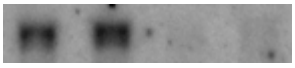
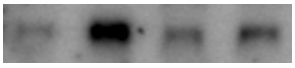
Group/ Gene or EST [a]	Expression ratio (microarray) [b]	Expression ratio (northern blot) [b]	Northern blot analysis 1[c] 2 3 4
Group 2			
actin	0.27	0.33	
NC4C8	0.48	0.14	
SC2A6	0.25	0.13	
SM2F5	0.30	0.40	
os-2	0.23	0.22	
W13G8	0.34	0.28	
ccg-6	0.39	0.41	
NM1E5	0.31	0.43	
NP3F7	0.11	0.36	
Group 3			
SC1D4	2.5	1.5	

Table 7. Continued

Group/ Gene or EST [a]	Expression ratio (microarray) [b]	Expression ratio (northern blot) [b]	Northern blot analysis 1[c]	2	3	4
Ungrouped						
NC1A11	28.3		ND [d]			
SC1E10	37.2		ND [d]			
NM1C3	32.8		ND [d]			
SP1F10	5.5		ND [d]			
Control rRNA						

[a] The EST IDs are derived from *Neurospora* genome project at the University of New Mexico.

[b] Expression ratio is a ratio of data obtaining from FLOE (lane 4) devided by *fl* (lane 1).

[c] Total RNA of *fl*, 74A, and FLOE derived from 18 hours of liquid culture (lane 1, 2 and 4, respectively); and FLOE from 12 hours liquid culture (lane 3).

[d] No signal detected by northern blot analysis.

displayed a pattern of expression resembling *eas* (28 to 34-fold increase in FLOE). These cDNAs, NC1A11, SC1E10, and NM1C3, were tested by northern blot analysis and no signal was detected. In examining the physical location of these clones in the microtitre plates containing DNA for microarray production, I found that each of them were adjacent to wells containing *eas* clones. This suggests to me that there was some contamination of these three DNAs with *eas* DNA prior to use for spotting the microarrays. The original plates containing the *E. coli* clones were checked and sequence analysis revealed that the clones were correct. This indicates that the contamination came at the time of PCR amplification or downstream DNA processing and is not due to contamination of the original clones or that the level of contamination is below the limit of detection by sequence analysis.

Three other identified developmentally regulated genes, *cat-1* (NM2E10/NCU08791.1), *con-6* and *con-10*, were also assigned to Group 1. In addition, the *fl* gene itself was expressed 5.85-fold higher in the FLOE strain than the *fl* mutant. Thus, the overexpression of *fl* in the FLOE strain could be detected in the microarray experiment.

Three genes related to energy metabolism, NM7H6, SC7F10, and NC3D8, were also found in Group 1 and are discussed below. The remaining four genes have no assigned function. EST NM5G10 corresponds to NCU09049.1 and encodes a predicted protein with an apparent homolog in *Magnaporthe grisea* (MG08535.4). The 172 residue *N. crassa* protein is predicted to be anchored in the membrane by SignalP-2.0 (http://www.cbs.dtu.dk/services/SignalP-2.0/) (Nielsen *et al.*, 1997), however, the

M. grisea protein is predicted to be secreted. A cell membrane or cell wall localization would be consistent with this gene playing a role in morphogenesis. SM1A4 corresponds to the 327 residue ORF NCU04667.1 with strong similarity (e^{-24}) to a hypothetical protein of *M. grisea* (MG09261.4). The *N. crassa* protein has five repeats of a motif (GPHGSRAANALDPRVDSDRDGSR) that is conserved in the *M. grisea* homolog. Different characteristic repeat motifs are found in *con-6* and *con-10*. Although the biochemical functions of these proteins are not known, this observation suggests that a variety of different proteins containing repeated sequence motifs may be characteristics of proteins involved in conidiation. NM1B8/NCU08949.1 encodes a 171 residue ORF that is unique to *N. crassa*. NM5C1/NCU06977.1 is predicted to be a 289 residue protein anchored in the membrane that has high similarity (e^{-101}) to the *M. grisea* ORF MG01461.4.

Group 2 is characterized by reduced expression in the FLOE strain (Table 7). Eight genes were found to follow this pattern of expression including actin. The microarray and northern blot results are consistent with strong regulation of actin transcription in response to induction of conidiation. This may be related to overall growth rate control, however, it is also likely that cytoskeleton rearrangement is central to the switch from polar to budding growth and a different relative need for actin during conidiation may result in a downshift in transcription. This finding points out the danger of relying on "housekeeping" genes as a constitutive control for genes that are not regulated. Pyruvate kinase (NC4C8 and SC2A6) was also affected and its significance with respect to energy metabolism will be discussed later. Glutathione peroxidase

(SM2F5) is involved in maintaining redox homeostasis and to prevent oxidative stress. Oxidative stress is thought to be an inducing signal for induction of conidiation during exposure of a mycelium to an air interface. Paradoxically, glutathione peroxidase, that should help prevent oxidative stress from occurring, is up-regulated in the *fl* mutant and is expressed very poorly in the FLOE strain despite the fact that catalase (*cat-1*), an enzyme that helps to protect cells from oxidative stress once it has already occurred, is induced in the FLOE strain. However, *cat-1* is likely induced as a developmental response rather than directly induced by oxidative stress (Michan *et al.*, 2002). Therefore, it is difficult to make any strong conclusions about a functional role for glutathione peroxidase or catalase in regulating conidiation. Nevertheless, it appears that glutathione peroxidase expression is inversely correlated to the conidiogenic potential of the cell.

Another clearly regulated member of Group 2 is a MAP kinase, *os-2*. This is the homolog of *S. cerevisiae HOG1*. *HOG1* is involved in the hyperosmotic stress response pathway and functional characterization of *os-2* in *N. crassa* also demonstrated a role in the hyperosmotic stress pathway (Zhang *et al.*, 2002). The *os-2* pathway is involved in glycerol accumulation in response to hyperosmotic stress but data also indicate an involvement of this pathway in cell wall integrity, cytoskeleton structure, or cell cycle progression. These activities may be particularly relevant to cells undergoing the morphological transition from filamentous to budding growth associated with conidiation. The altered expression of actin that I observed is consistent with this latter interpretation.

A second putative kinase, W13G8/NCU06249.1, encodes a 1221 residue ORF. There is a strong similarity to MG05220.4 along the entire length of the protein. The C-terminal 300 amino acids have similarity to a wide variety of serine/threonine kinases. The N-terminal region displayed conservation with the PAS domains of the PAS kinase *PSK1* of *S. cerevisiae*. PAS kinases are thought to respond to a variety of external and internal signals, such as redox state (Rutter *et al.*, 2001). Its expression was reduced in the FLOE strain at the 18 hr time point but not affected at the 12 hr time point. This most strongly resembles the pattern observed for actin. As with *os-2*, changes in the transcript levels of this kinase are likely to result from the consequences of induced development rather than direct transcriptional regulation by *fl*.

The clock controlled gene, *ccg-6*, was also down-regulated somewhat in the *fl* overexpression strain. This gene encodes a 142 amino acids polypeptide with no known homologue and the function is not known. NM1E5/NCU05881.1 encodes a 701 codon ORF with closest similarity to the *M. grisea* ORF MG00830.4. NP3F7/NCU08751.1 encodes a 334 codon ORF with partial sequence similarity to MG03420.4.

Group 3 consists of a single gene that is expressed at relatively high levels in the wild type mycelium but is poorly expressed in both the *fl* mutant and FLOE strains. This gene, represented by cDNA clone SC1D4/NCU08907.1 has sequence similar to blastomyces yeast phase-specific protein 1. This protein is associated with the budding growth phase of blastomyces. Northern blot analysis (Table 7, Group 3) revealed that although transcript of this gene was most highly expressed in wild-type, its transcript levels were lower in the *fl* mutant than the FLOE strain. One possibility is that this gene

may be required *fl* for expression during hyphal growth and during development its expression may be down-regulated by other developmental factors.

Other genes were tested by microarray analysis but not further verified by northern blot analysis. NM6B8, representing NCU01517.1, codes for a protein which has sequence similar to the yeast glucan 1,4-alpha-glucosidase (also known as glucoamylase). Although the activity of glucoamylase is for releasing glucose from polysaccharide, this protein has also been identified as a sporulation-specific protein in yeast (Pugh and Clancy, 1990). Both glucoamylase and yeast phase-specific protein 1 are involved in the transition from mycelial phase to sporulation and the yeast phase, respectively (Bono *et al.*, 2001; Pugh and Clancy, 1990). It is possible that these two genes may participate in the transition stage from vegetative growth to conidiation.

Besides the genes *eas, con-6* and *con-10,* microarray analysis did not show any other genes that respond to *fl* that were likely to be involved in development based on gene annotations. One possibility is that the EST libraries that were used to make the microarrays were derived from tissues that are underrepresented in genes expressed during conidial development. In addition, not every gene represented by clones in the cDNA library, such as those for *con-8* and *con-13*, were printed on the microarrays. Furthermore, some genes whose expression levels were very low, such as the genes that encode transcription factors, often did not pass the cut off value for signal intensity and therefore were not analyzed. Therefore, these microarray experiments do not include many of the genes associated with conidial development or may not be sensitive enough to reliably detect expression of some of the genes that are regulated by *fl*.

Effect of *fl* on central metabolism. Central metabolism is the main process that cells use to generate energy in the form of ATP. It involves glycolysis which takes place in cytosol; and the tricarboxylic acid (TCA) cycle and electron transport chain which occur in mitochondria. When glucose, which is the common entry molecule for glycolysis, is exhausted, gluconeogenesis and the glyoxylate cycle are utilized to generate glucose for biosynthesis (Alberts *et al.*, 1994).

Glycolysis

Previous microarray experiments demonstrated that when *N. crassa* was cultured in glucose-rich medium, transcripts of genes encoding enzymes involved in glycolysis and fermentation were induced, whereas in glucose deprivation medium expression of the genes involved in tricarboxylic acid (TCA) cycle and gluconeogenesis were induced (Xie, 2003). It is known that nutrient deprivation, particularly for carbon and nitrogen, is a major trigger to induce conidiation (Ebbole, 1998; Springer, 1993). Microarray analysis in my study showed that in the *fl* mutant strain, where cells were in the vegetative stage, mRNA for the enzymes glyceraldehyde-3-phosphate dehydrogenase (CCG-7) and pyruvate kinase (ACE-8), were two-fold higher than in the FLOE strain (Table 5). Glyceraldehyde-3-phosphate dehydrogenase catalyzes the conversion of glyceraldehyde-3-phosphate to glycerate-3-phosphate which is subsequently converted to phosphoenolpyruvate. The phosphoenolpyruvate is then converted to pyruvate by pyruvate kinase (Alberts *et al.*, 1994). The induction of these enzymes catalyzes the flow of the reaction from glucose to pyruvate through glycolysis. Northern blot analysis of NC4C8 and SC2A6 representing *ace-8* (Table 7, Group 2) from the cultures of the same

age (18 hr; lane 1, 2 and 4) showed that the mRNA level of this gene in the *fl* mutant was higher than in the wild-type and FLOE strains, and the level in wild-type was higher than in FLOE strain. However, at the 12 hr time point expression of *ace-8* in the FLOE strain was similar to the wild-type at 18 hr. Thus, the relatively high level of expression of this gene indicates that metabolism in the *fl* mutant is more glycolytic than in the wild-type and FLOE strains. In addition, the expression of this gene in FLOE at an earlier time (12 hr; lane 3) was higher than at later time (18 hr; lane 4). I interpreted this result to indicate that as development proceeds the rate of glycolysis decreases.

As mentioned, Xie (Xie, 2003) reported that transcripts of genes encoding enzymes involved in glycolysis were elevated when *N. crassa* was cultured in glucose-rich medium. Furthermore, it has been reported that the *fl* mutant consumed sugar (glycolysis) slower than wild-type (Bernardini and Turian, 1978). I interpret this to indicate that the amount of sugar remaining in the medium with the *fl* mutant, at the time point analyzed, may be higher than in wild-type or FLOE strains. Therefore, *fl* mutants may have a higher level of glycolytic enzymes to utilize the higher level of glucose remaining. My hypothesis is that the high expression level of *ace-8* mRNA in the *fl* mutant strain (Table 7, lane 1) was related to the remaining high amount of sugar in the culture, whereas the low expression level of *ace-8* in wild-type and FLOE strains (Table 7, lane 2 and lane 4) was related to the low amount of sugar left in the media. This speculation is in agreement with the result obtained by Bernadini and Turian (Bernardini and Turian, 1978) that the wild-type has higher a sugar consumption rate than the *fl* mutant. Moreover, the expression level of *ace-8* in FLOE was less than in

wild-type (Table 7, lane 4 and lane 2). This result suggested that overexpression of *fl* plays some roles in induction of sugar consumption, resulting in increasing the glycolysis rate. This would lead to faster exhaustion of sugar, an earlier time for ethanol production, and a shift to ethanol utilization and gluconeogenesis at the 18 hr time point. This result is consistent with the fact that conidiation is initiated under glucose deprivation.

The *cre-1* gene of *N. crassa* is homologous to *creA* in *A. nidulans*, which is a carbon catabolite repressor that is induced in the presence of glucose and auto-repressed by glucose depletion (Ebbole, 1998; Strauss *et al.*, 1999). A low expression level of *cre-1* in FLOE (Table 6) also suggested that the level of glucose in the media of FLOE culture was lower than in the *fl* mutant strain. The reduction of *cre-1*, *ace-8*, and *ccg-7* transcripts therefore is consist with a low level of glucose remaining in the media of the FLOE strain after 18 hr in culture. This result suggested that FLOE cells, which were at the conidiation stage, consumed sugar faster than *fl* mutant cells which were in the vegetative phase of growth. One possibility is that FLOE cells have a higher rate of glucose metabolism than *fl* cells.

Fermentation

Glycolysis generates pyruvate which is an intermediate metabolite for many biological processes. One branch pathway from pyruvate is the production of ethanol through fermentation where pyruvate is converted to acetaldehyde which is subsequently reduced to ethanol by alcohol dehydrogenase. In *A. nidulans*, alcohol dehydrogenase (ADH) proteins have been classified into three groups. ADH I and ADH II are under

carbon catabolite repression through CREA. ADH I is induced by ethanol and is repressed in the presence of glucose, whereas ADH II is repressed by ethanol. ADH III has very high sequence identity to ADH I (76% identity), however, it is not under carbon catabolite repression and it works under anaerobic conditions (Hunter *et al.*, 1996). It is necessary to note that the nomenclature of ADH I and ADH II of *A. nidulans* is not the same as ADH I and ADH II in *S. cerevisiae*. In *S. cerevisiae*, ADH I functions in anaerobic conditions to generate ethanol, and ADH II functions to utilize ethanol (Young *et al.*, 2000). In *N. crassa*, there are at least 28 ORFs in *N. crassa* genome that contain an alcohol dehydrogenase domain (http://www-genome.wi.mit.edu/annotation/fungi/neurospora), however, the function of these members so far has not been identified. From my microarray analysis, transcripts of three alcohol dehydrogenases were increased in FLOE culture (Table 6). Protein sequence alignments revealed that NCU02476.1 and NCU01754.1 have sequence similarity to ADH I and ADH III of *A. nidulans*, while NCU09285.1 is closely related to an alcohol dehydrogenase unidentified function in *Schizosaccharomyces pombe*. If the level of glucose in the FLOE culture was low, it is most likely that all or at least one of these up-regulated alcohol dehydrogenase transcripts may function similarly to the ADH I of *A. nidulans*. In a study of glucose regulation (Xie, 2003), under glucose starvation conditions, NCU01754.1 and NCU02476.1 were expressed at 7.7- and 1.1-fold higher levels, respectively. Additionally, a previous study with alcohol dehydrogenase in *N. crassa* found that there was one alcohol dehydrogenase produced during fermentation using sucrose as a carbon source, while there were at least two isozymes observed when

ethanol was used as a main substrate (Zink, 1969). Taken together, it is possible that the three alcohol dehydrogenases identified by microarray analysis may function as ADH I of *A. nidulans* that utilize ethanol as their substrate.

In the absence of a time course that measures gene expression, glucose concentration, and ethanol concentration I can only speculate about the events occurring in these cultures based on previous studies that do measure these parameters comparing wild type and *fl* mutant. Repeating such a study to include the FLOE strain is needed. Previous studies showed that *fl* mutants utilize sugar more slowly than wild type and produce less ethanol (Bernardini and Turian, 1978). I hypothesize that the FLOE strain uses glucose at an even greater rate, producing ethanol earlier than in the wild type. This is based on the elevated levels of alcohol dehydrogenase mRNA at the 12 and 18 hr time point and the lower level of glycolytic enzyme mRNAs at the 18 hr time point. I interpreted this result to mean that the FLOE strain utilized glucose more rapidly than wild type and *fl* mutant so that by 18 hr little glucose remained in the culture and genes for utlizing glucose were down-regulated. Although this interpretation is somewhat speculative, it provides a clear model for future experimental testing.

To summarize, northern blot analysis of *ace-8* (Table 7, group 2), would be consistent with sugar consumption in the FLOE strain being higher than in wild-type, and wild-type having a higher rate than in the *fl* mutant (sugar consumption: FLOE > wild-type > *fl* mutant). Furthermore, northern blot analysis of NC3D8 representing one of the putative ADH I genes (Table 7, group 1), revealed that the expression level of the putative ADH I in FLOE was higher than in the wild-type. The level of the ADH1

homolog was also higher in the wild-type than in *fl* mutant. Because ADH I is induced by ethanol in *A. nidulans*, this result could be used as a basis for speculating that the time at which ethanol is produced in the FLOE strain was earlier than in wild-type, and that wild-type produced ethanol earlier than the *fl* mutant (time of ethanol production: FLOE > wild-type > *fl* mutant). If this were true, it would be consistent with an interpretation that ethanol utilization and ADH I production of FLOE was greater than in wild-type or *fl* mutant at the 18 hr time point (ethanol utilization and ADH I production: FLOE > wild-type > *fl* mutant). Bernardini and Turian (Bernardini and Turian, 1978) found that in both wild-type and *fl* mutant, ethanol production was positively correlated with sugar utilization. This process should require the activity of enzymes involved in glycolysis and ADH II. Taken together, these results suggest that *fl* may exert an effect on glycolysis and ethanol production and utilization. However, none of the genes involved in the glycolysis pathway contain a consensus binding site of FL (not included in Table 5, Chapter III). This suggests that any effect of *fl* is probably indirect.

TCA cycle

The main pathway for production of pyruvate is conversion to acetyl-CoA for entry into the TCA cycle. From each molecule of acetate, in the form of acetyl-CoA (and 3 NAD^+, FAD, GDP, P_i, 2 H_2O), the TCA cycle generates two molecules of carbon dioxide, 3 NADH, $FADH_2$, 2 H^+ and GTP. The NADH and $FADH_2$ can be used to generate additional energy through the electron transport system. Microarray data comparing the *fl* mutant and FLOE strains showed that, except for the transcript for

fumarase, no genes that participate in the TCA cycle have differential expression. Therefore, the TCA cycle genes appear to be expressed in a fairly constitutive manner.

Electron transport chain

The electron transport chain is the final stage of cell respiration. The mitochondria electron transport chain of *N. crassa* comprised of four multisubunit membrane-bound protein complexes that perform a series of oxidation-reduction reactions to reduce oxygen to water (Davis, 2000; Joseph-Horne *et al.*, 2001). Complex I performs NADH ubiquinone oxidoreductase that transfers electron from NADH to ubiquinone (also called CoenzymeQ, CoQ) with translocation of protons across the membrane. The proton motive force across the membrane is used for ATP synthesis. Alternative NADH dehydrogenases were also found but the oxidation of NADH by the alternative dehydrogenases occurs without proton translocation (no ATP). Complex III is a cytochrome bc_1 complex that performs CoQ-cytochrome c oxidoreductase transferring electron from CoQ to cytochrome c. Complex IV contains cytochromes a and a3 with a cytochrome c oxidase that transfers electrons from cytochrome c to oxygen which is then converted to water. Complex II carries subunits of cytochrome b and a succinate dehydrogenase activity which is an important enzyme complex for both the tricarboxylic acid (TCA) cycle, and respiratory chains. It oxidizes succinate to fumarate generating $FADH_2$ and transfers electrons into the respiratory chain from $FADH_2$ to ubiquinone (Davis, 2000; Joseph-Horne *et al.*, 2001). Four ORFs which have sequences similarity to the proteins participating in the respiration pathway were increased in response to the overexpression of *fl*. These include NCU00959.1 and NCU03031.1 for Complex II,

NCU06601.1 and NCU08940.1 for Complex III (Table 6). In culture, conidiation occurs after glucose has been utilized. These are conditions that favor respiration and do not favor ethanol fermentation; rather they favor ethanol utilization through respiration and gluconeogenesis (Bernardini and Turian, 1978; Colvin *et al.*, 1973). If glucose is exhausted in the FLOE culture, the only means of generating energy is via respiration, whereas the wild type and *fl* mutant cultures may still be generating ATP via both glycolysis and respiration. Conceivably, this may partly explain the increase in transcripts for components of the respiratory chain. However, we do not know how the conidial morphogenesis that is occurring in the FLOE strain may affect the number of mitochondria/unit cell mass, and thereby effect expression of genes required for mitochondrial biogenesis.

Lipid and amino acid metabolism

Cells of *N. crassa* cultured in liquid minimal medium for 12-18 hr are at a late log phase (Colvin *et al.*, 1973). If there are excess amounts of nutrients in the medium at this stage, they are converted to storage forms such as fatty acid, sterol and glycogen (Davis, 2000). The transcript for ATP citrate lyase was increased in *fl* mutant culture (Table 6). This enzyme is responsible for the conversion of citrate to oxaloacetate and acetate. Acetate then couples with CoA, forming acetyl-CoA, and leaves the mitochondrion where it is subsequently converted to fat through lipogenesis (Nowrousian *et al.*, 2000). A transcript of *acu-8* gene, the gene coding for acetyl-CoA hydratase was also induced in this same condition. It has been suggested excess acetyl-CoA is toxic to cells, and that the product of *acu-8* is required to balance the level of

acetyl-CoA in the cytosol (Connerton *et al.*, 1992). In agreement with the induction of ATP citrate lyase, several mRNA representing enzymes that are associated with fatty acid biosynthesis were also induced in *fl* mutant strain as shown in Table 6. This result is consistent with microarrays results for *N. crassa* cultured in glucose-rich medium (Xie, 2003).

Amino acid biosynthesis appeared to be elevated in the *fl* mutant strain relative to FLOE as evidence by higher transcription levels of genes for lysine, leucine, and arginine biosynthesis and the general regulator of amino acid biosyntheses, *cpc-1* (Table 6). The general growth rate of the *fl* mutant strain at the 18 hr time point may faster than the FLOE. This is supported by the increased level of several genes involved in transcription/translation in *fl* mutant strain (Table 6).

Effect of *fl* on cell wall synthesis and the cytoskeleton. Fungal mycelia grow by extension from the tip of the hyphae. During mycelial elongation, active compartments such as the cytoskeleton, mitochondria, the endoplasmic reticulum, nuclei and vacuoles are found to accumulate at the tip area (Davis, 2000). A clustering of secretory vesicles called the Spitzenkörper is also found associated 3-4 μm behind the tip of the elongating hyphae. The Spitzenkörper is located at the site where growth is occurring (Bartnicki-Garcia *et al.*, 1995). The secretory vesicles of the Spitzenkörper are thought to contain materials that are required for cell wall synthesis and hyphal elongation. These vesicles are transported from the Spitzenkörper area to the tip by cytoskeleton motors which are divided into three types, kinesin, dynein, and myosin. The motors utilize ATP to carry their ‘cargo’ along the microfibrils such as microtubules

and F-actin. Kinesin and dynein move along the microtubules, while myosin uses F-actin for the transportation (Davis, 2000; Steinberg, 2000). The microtubule dependent movement was found to be distributed throughout the cytoplasm, whereas the actin dependent movement was found to be associated with the hyphal tip and cortex (Davis, 2000). Microarray analysis revealed that transcripts representing actin, tubulin, and their associated proteins were induced 2-4 times in *fl* mutant strain, where mycelial phase was taking place (Table 6). Northern blot analysis using actin (Table 7) also confirmed that its transcript was elevated in the mycelial phase (lane 1, and 2), but started to decrease when cells were initiating conidiation (lane 3), and sharply decreased once conidiation began (lane 4).

The cell wall not only determines the tubular shape of the hyphae, but also acts as a protective barrier from enzymes, mechanical forces, osmotic pressure, etc. The cell wall is composed of polysaccharide, glucan, and chitin (Farkas, 2003). Glucan synthase and chitin synthase are the major enzymes involved in the cell wall synthesis. In addition, during tip growth, some wall-degrading enzymes such as glucanases and chitinases are also required to soften the tip wall (Davis, 2000; Scott, 1976). On the other hand, conidial cell wall is rich in a hydrophobic rodlet protein which is composed of the hydrophobin, *eas*, a small protein with eight cysteine residues (Ebbole, 1997). Microarray analysis showed that two transcripts encoding cell wall synthesis enzymes, NM7B6 and W06F9, were induced in conidiating cells and vegetative cells, respectively. Furthermore, two mRNA representing cell wall degrading enzymes, NP4B12 and W07G9, were induced in conidiating cells and vegetative cells, respectively. These

results suggested that both cell wall synthesis and degrading enzymes probably are cell-type specific. In addition, as already mentioned, *eas* which encodes hydrophobin protein is directly regulated by *fl* therefore it was highly induced in conidiating cells (FLOE).

Identification of genes regulated by *fl*. In Chapter III, consensus sequence of FL was identified. A search for the appearance of this consensus binding sequence in a 2 kb region upstream of all 10,082 ORFs in *N. crassa* was performed using the EMBOSS FUZZNUC software. The search revealed 123 ORFs (Chapter III, Table 5) that contain the consensus sequence for a FL binding site in their promoter regions. The 123 ORFs were compared with the up- and down- regulated genes listed in Table 6. However, only the *eas* gene is present in both lists. This result suggests that, beside *eas*, all other genes in Table 6 are not direct targets of FL.

Northern blot analysis of two genes, NM5G10/NCU09049.1 and NP3F7/NCU08751.1 (Table 7) showed that NM5G10 was expressed only in FLOE, whereas NP3F7 was highly expressed in the *fl* mutant. To study the roles of these two genes, I attempted to mutate these genes using the RIP strategy (Selker *et al.*, 1987). A plasmid containing the DNA of each gene was introduced into the wild-type strain (74A). The transformants harboring an extra copy of the gene were verified by Southern blot analysis (Fig. 19 and Fig. 20, B). One transformant was selected to cross to the opposite mating type (ORS-6*a*). One hundred progeny from each cross were analyzed for a defect in conidiation and growth rate. However, all progeny showed the wild-type phenotype. To determine if RIP mutation occurred, Southern blot analysis of the genomic DNA digested with *Sau*3A1 and *Dpn*II was performed. Because RIP is

expected to occur in approximately 25% of the progeny, I analyzed 12 progeny from each cross. All twelve progeny from each cross showed a similar pattern using both enzymes indicating that no methylation or RIP mutation occurred in these progeny. Therefore, it is possible that RIP did not occur in these crosses, or that additional progeny need to be tested.

Clearly, a whole genome microarray coupled with an extensive set of time course experiments with wild-type, FLOE, and *fl* mutant strains will be useful in generating a complete picture of the role of *fl* on gene expression. Here, I have provided a snapshot of the effect of mutation or overexpression of *fl* to identify the role for *fl* in metabolism. Remarkably, only *eas* was identified as a direct target of *fl*, and this likely reflects the limitation of our cDNA microarray. However, several genes (NCU09049.1, NCU08751.1, NCU08907.1) were identified that are likely regulated during development, but not as direct targets of *fl*.

CHAPTER V

SUMMARY

Transcription factors control gene expression by recognizing and interacting with a *cis*-acting sequence in the promoter of specific genes to activate or repress expression at a specific time and/or in specific tissues. The consensus binding site of a particular transcription factor can be identified using the purified transcription factor protein and *in vitro* DNA-binding selection strategies.

A number of commercially available expression vectors can be used to facilitate the extraction and purification procedures, and also assist in achieving high levels of expression. However, as each protein has its own characteristics, an expression system that is appropriate for use with the target protein must be selected. For example, despite yielding low expression levels initially, GST fusion resulted in the successful isolation of recombinant FL protein. When the procedure was scaled up, a sufficient amount of recombinant FL with intact DNA-binding function was obtained for further study.

Using an *in vitro* selection approach, the binding site of FL was identified as a CGG inverted repeat with nine nucleotide spacing, 5'-CGG$(N)_9$CCG-3'. The consensus sequence, 5'-CGGAGGTTTCCTCCG-3', was found at -1498 bp in the promoter of the *eas* gene. *In vivo* sequence tag analysis using a fragment of the *eas* promoter up to -1660 bp indicated that the presence of this sequence is required for the expression of the *eas* gene. This fragment, however, is not sufficient for full expression of *eas*, as

compared to the intact gene, suggesting that sequences beyond this region play a role in the regulation of *eas*. In this upstream region, there may be either another binding site of *fl*, as found in AlcR (Panozzo *et al.*, 1997), or a binding site for another transcription factor that works with FL, as found in NIT2/NIT4 (Feng and Marzluf, 1998). A preliminary *in vitro* assay demonstrated strong binding of FL to the oligonucleotide representing the sequence at -2573 bp in the *eas* promoter. Additional analysis is required to verify if this sequence is necessary for regulation by FL, and whether this site can act efficiently with the -1498 bp site to induce expression of *eas* to a comparable level as the endogenous *eas* gene.

The EMBOSS FUZZNUC software identified 123 ORFs that contained FL binding sites. However, northern blot analysis of most of the genes did not reveal any expression signal. Two possibilities may account for this observation: the expression level was too low to be detected by northern analysis, or expression of these genes may require the synergistic function of another gene product, which is expressed only during the development, in addition to FL. As the mRNA that was used in these experiments was derived from liquid cultures, it is possible that the co-regulated gene was either not expressed or was inactive under liquid growth conditions, therefore resulted in no expression of the target gene. Northern blot analysis of mRNA derived from conidial development of wild-type aerial culture could provide further information on these 123 genes.

Although the consensus binding sites of several transcription factors have been identified by using the *in vitro* selection method, this approach may sometimes lead to

inaccurate predictions. For example, the consensus binding sequence for RAP1 was determined by the SELEX method. A target gene list was produced based on the consensus binding sequence (Graham and Chambers, 1994). However, not all known target genes for RAP1 were found in this list (Eriksson *et al.*, 2000; Schuller *et al.*, 1994). These observations indicate that the *in vitro* assay does indeed provide fundamental information; however, because *in vitro* and *in vivo* binding conditions are certainly different, differences in binding specificity may be observed. An *in vivo* binding assay such as Chromatin ImmunoPrecipitation (ChIP) may facilitate more accurate results. In *S. cerevisiae,* where the genome size is small (about 12 Mb), ChIP was integrated with microarray technology, called ChIP-chip, to identify the target genes of the transcription factors of interest (Lieb *et al.*, 2001; Ren *et al.*, 2000; Wyrick *et al.*, 2001). However, as the genome size of *N. crassa* is more triple the size of *S. cerevisiae*, it may be difficult to generate a whole genome array. Thus, the traditional ChIP technique may be more suitable.

The activation domain of FL was identified and localized to amino acid residues 399-556. In addition, PEST motifs were found at positions 406-422 and 675-688. These motifs may play an important role in protein-protein interactions as discussed in Chapter II. Therefore, further study of a potential partner protein that may interact with FL is an interesting issue that warrants further study.

In Chapter IV, microarray analysis revealed that transcripts of enzymes involved in glycolysis in the *fl* mutant strain were higher than in wild-type and FLOE. Furthermore, previous research found that the *fl* mutant utilized sugar slower than wild-

type (Bernardini and Turian, 1978). Taken together I interpreted that the high level of expression of genes for glycolysis in the *fl* mutant was due to a higher sugar level in the medium at that time point. In the FLOE strain the rate of sugar utilization was faster, therefore, sugar was exhausted by the 18 hr time point, resulting in the low expression level of glycolysis enzymes. The deprivation of sugar initiated gluconeogenesis, and ethanol was utilized by alcohol dehydrogenases. Based on previous findings (Zink, 1969) and sequence similarity, the three alcohol dehydrogenases that were induced in FLOE may function similarly to ADH I of *A. nidulans* to utilize ethanol.

Several genes were differentially expressed in the *fl* mutant and FLOE strains. However, this microarray approach did not aid in identifying new target genes of *fl* in addition to *eas*. This could be because the genes present on the microarray correspond to only one-seventh of the total *N. crassa* genome. In addition, the genes on the microarray may under-represent the actual genes involved in conidial development. A new microarray representing all of the genes of *N. crassa*, which is being developed, will be a useful tool for identification of target genes of *fl*. A comparison of wild-type and *fl* mutant mRNA derived from aerial hyphae during development would be a good starting point, since my experiments comparing *fl* mutant and FLOE strains in liquid culture demonstrated an apparent indirect, rather than direct, effect of *fl* on the expression of genes in cultures grown for 18 hr.

Overall, I conclude that using the *in vitro* binding site selection approach provides necessary fundamental knowledge regarding binding site properties of a transcription factor, but may not provide enough information to predict the correct target

gene. The *eas* gene (and probably other genes among the 123 ORFs potentially regulated by FL) contain the binding site of FL; however, it appeares that additional elements and/or factors are important for conidiation-specific regulation. The role of *fl* during conidial development is, at the least, to induce the expression of the *eas* gene in order to produce hydrophobin proteins, which must be deposited at the conidial surface in order for conidia to mature and disseminate properly.

FUTURE DIRECTIONS

To fully define the network and processes required to produce conidia, there are likely to be many genes upstream and downstream of *fl* that need to be identified and characterized: for example, *acon-2*, *acon-3*, *csp-1* and *csp-2*. The identification of addition partner proteins that interact with FL, if any, will also lead to a better understanding of conidial development. Since the mid-1980's, progress in understanding conidial development in *N. crassa* has been slow. This is primarily because of the inefficiency of direct complementation of aconidial strains. With the complete genome sequence of *N. crassa*, together with genomic technologies, it is expected that cloning and characterization of genes involved in conidiation will be greatly accelerated. The work described here has demonstrated that once genes are identified and cloned, rapid progress can be made in characterizing their functions in asexual development in *N. crassa*.

REFERENCES

Adams, T.H., Wieser, J.K., and Yu, J.H. (1998) Asexual sporulation in *Aspergillus nidulans*. *Microbiol Mol Biol Rev* **62**: 35-54.

Aign, V., and Hoheisel, J.D. (2003) Analysis of nutrient-dependent transcript variations in *Neurospora crassa*. *Fungal Genet Biol* **40**: 225-233.

Akiyama, M., and Nakashima, H. (1996) Molecular cloning of the *acr-2* gene which controls acriflavine sensitivity in *Neurospora crassa*. *Biochim Biophys Acta* **1307**: 187-192.

Alberts, B., Bray, D., Lewis, J., Raff, M., Roberts, K., and Watson, J.D. (1994) *Molecular Biology of the Cells.* New York: Garland Publishing.

Bailey, L.A., and Ebbole, D.J. (1998) The *fluffy* gene of *Neurospora crassa* encodes a Gal4p-type C_6 zinc cluster protein required for conidial development. *Genetics* **148**: 1813-1820.

Bailey-Shrode, L., and Ebbole, D.J. (in press) The *fluffy* gene of *Neurospora crassa* is necessary and sufficient to induce conidiophore development. *Mol Microbiol*

Bartnicki-Garcia, S., Bartnicki, D.D., Gierz, G., Lopez-Franco, R., and Bracker, C.E. (1995) Evidence that Spitzenkörper behavior determines the shape of a fungal hypha: a test of the hyphoid model. *Experimental Mycology* **19**: 153-159.

Baum, J.A., Geever, R., and Giles, N.H. (1987) Expression of qa-1F activator protein: identification of upstream binding sites in the *qa* gene cluster and localization of the DNA-binding domain. *Mol Cell Biol* **7**: 1256-1266.

Beadle, G.W., and Tatum, E.L. (1941) Genetic control of chemical reactions in *Neurospora*. *Proc Natl Acad Sci USA* **27**: 499-506.

Beever, R.E., and Dempsey, G.P. (1978) Function of rodlets on the surface of fungal spores. *Nature* **272**: 608-610.

Bell-Pedersen, D., Dunlap, J.C., and Loros, J.J. (1992) The *Neurospora* circadian clock-controlled gene, *ccg-2,* is allelic to *eas* and encodes a fungal hydrophobin required for formation of the conidial rodlet layer. *Genes Dev* **6**: 2382-2394.

Bell-Pedersen, D., Dunlap, J.C., and Loros, J.J. (1996) Distinct *cis*-acting elements mediate clock, light, and developmental regulation of the *Neurospora crassa eas* (*ccg-2*) gene. *Mol Cell Biol* **16**: 513-521.

Bennett, J.W. (1998) Mycotechnology: the role of fungi in biotechnology. *J Biotechnol* **66**: 101-107.

Berlin, V., and Yanofsky, C. (1985) Isolation and characterization of genes differentially expressed during conidiation of *Neurospora crassa*. *Mol Cell Biol* **5**: 849-855.

Bernardini, D., and Turian, G. (1978) Studies on the conidial differentiation of *Neurospora crassa*. 7. Alcoholic regulation and conidiogenous capacity (wild-type and *fluffy* mutant). *Annales de Microbiologie* **129 B**: 551-559.

Bibbins, M., Crepin, V.F., Cummings, N.J., Mizote, T., Baker, K., Mellits, K.H., and Connerton, I.F. (2002) A regulator gene for acetate utilisation from *Neurospora crassa*. *Mol Genet Genomics* **267**: 498-505.

Blackwell, T.K. (1995) Selection of protein binding sites from random nucleic acid sequences. *Methods Enzymol* **254**: 604-618.

Bono, J.L., Jaber, B., Fisher, M.A., Abuodeh, R.O., O'Leary-Jepson, E., Scalarone, G.M., and Smith, L.H., Jr. (2001) Genetic diversity and transcriptional analysis of the *bys1* gene from *Blastomyces dermatitidis*. *Mycopathologia* **152**: 113-123.

Borneman, A.R., Hynes, M.J., and Andrianopoulos, A. (2002) A basic helix-loop-helix protein with similarity to the fungal morphological regulators, Phd1p, Efg1p and StuA, controls conidiation but not dimorphic growth in *Penicillium marneffei*. *Mol Microbiol* **44**: 621-631.

Bourett, T.M., Sweigard, J.A., Czymmek, K.J., Carroll, A., and Howard, R.J. (2002) Reef coral fluorescent proteins for visualizing fungal pathogens. *Fungal Genet Biol* **37**: 211-220.

Campbell, J.W., Enderlin, C.S., and Selitrennikoff, C.P. (1994) Vectors for expression and modification of cDNA sequences in *Neurospora crassa*. *Fungal Genet Newsl* **41**: 20-21.

Carey, M., and Smale, S.T. (2000) *Transcriptional Regulation in Eukaryotes: Concepts, Strategies, and Techniques.* Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Chen, E.Y., and Clarke, D.M. (2002) The PEST sequence does not contribute to the stability of the cystic fibrosis transmembrane conductance regulator. *BMC Biochemistry* **3**: 29.

Colvin, H.J., Sauer, B.L., and Munkres, K.D. (1973) Glucose utilization and ethanolic fermentation by wild type and extrachromosomal mutants of *Neurospora crassa*. *J Bacteriol* **116**: 1322-1328.

Connerton, I.F., McCullough, W., and Fincham, J.R. (1992) An acetate-sensitive mutant of *Neurospora crassa* deficient in acetyl-CoA hydrolase. *J Gen Microbiol* **138**: 1797-1800.

Correa, A., and Bell-Pedersen, D. (2002) Distinct signaling pathways from the circadian clock participate in regulation of rhythmic conidiospore development in *Neurospora crassa*. *Eukaryot Cell* **1**: 273-780.

Correa, A., Lewis, A.Z., Greene, A.V., March, I.J., Gomer, R.H., and Bell-Pedersen, D. (2003) Multiple oscillators regulate circadian gene expression in *Neurospora*. *Proc Natl Acad Sci USA* **100**: 13597-13602.

Corrochano, L.M., Lauter, F.R., Ebbole, D.J., and Yanofsky, C. (1995) Light and developmental regulation of the gene *con-10* of *Neurospora crassa*. *Dev Biol* **167**: 190-200.

Davis, R.H. (2000) *Neurospora: Contributions of a Model Organism.* New York: Oxford University Press.

Davis, R.H., and de Serres, F.J. (1970) Genetic and microbiological research techniques for *Neurospora crassa*. *Methods Enzymol* **17A**: 79-143.

Davis, R.H., and Perkins, D.D. (2002) Timeline: *Neurospora*: a model of model microbes. *Nat Rev Genet* **3**: 397-403.

Dev, K., and Maheshwari, R. (2002) Transformation in heterokaryons of *Neurospora crassa* is nuclear rather than cellular phenomenon. *Curr Microbiol* **44**: 309-313.

Ebbole, D.J. (1996) Morphogenesis and vegetative differentiation in filamentous fungi. *J Genet* **75**: 361-374.

Ebbole, D.J. (1997) Hydrophobins and fungal infection of plants and animals. *Trends Microbiol* **5**: 405-408.

Ebbole, D.J. (1998) Carbon catabolite repression of gene expression and conidiation in *Neurospora crassa*. *Fungal Genet Biol* **25**: 15-21.

Ehrlich, K.C., Montalbano, B.G., and Cary, J.W. (1999) Binding of the C_6 zinc cluster protein, AFLR, to the promoters of aflatoxin pathway biosynthesis genes in *Aspergillus parasiticus*. *Gene* **230**: 249-257.

Eriksson, P., Alipour, H., Adler, L., and Blomberg, A. (2000) Rap1p-binding sites in the *Saccharomyces cerevisiae* GPD1 promoter are involved in its response to NaCl. *J Biol Chem* **275**: 29368-29376.

Eubanks, M. (2002) Rice and rice blast: new bases for comparison. *Environ Health Perspect* **110**: A740.

Farkas, V. (2003) Structure and biosynthesis of fungal cell walls: methodological approaches. *Folia Microbiologica* **48**: 469-478.

Feng, B., and Marzluf, G.A. (1998) Interaction between major nitrogen regulatory protein NIT2 and pathway-specific regulatory factor NIT4 is required for their synergistic activation of gene expression in *Neurospora crassa*. *Mol Cell Biol* **18**: 3983-3990.

Fernandes, M., Keller, N.P., and Adams, T.H. (1998) Sequence-specific binding by *Aspergillus nidulans* AflR, a C_6 zinc cluster protein regulating mycotoxin biosynthesis. *Mol Microbiol* **28**: 1355-1365.

Fu, Y.H., Feng, B., Evans, S., and Marzluf, G.A. (1995) Sequence-specific DNA binding by NIT4, the pathway-specific regulatory protein that mediates nitrate induction in *Neurospora*. *Mol Microbiol* **15**: 935-942.

Gadgil, H., Jurado, L.A., and Jarrett, H.W. (2001) DNA affinity chromatography of transcription factors. *Anal Biochem* **290**: 147-178.

Galagan, J.E., Calvo, S.E., Borkovich, K.A., Selker, E.U., Read, N.D., Jaffe, D., *et al.* (2003) The genome sequence of the filamentous fungus *Neurospora crassa*. *Nature* **422**: 859-868.

Geever, R.F., Huiet, L., Baum, J.A., Tyler, B.M., Patel, V.B., Rutledge, B.J., *et al.* (1989) DNA sequence, organization and regulation of the *qa* gene cluster of *Neurospora crassa*. *J Mol Biol* **207**: 15-34.

Gorl, M., Merrow, M., Huttner, B., Johnson, J., Roenneberg, T., and Brunner, M. (2001) A PEST-like element in FREQUENCY determines the length of the circadian period in *Neurospora crassa*. *EMBO J* **20**: 7074-7084.

Graham, I.R., and Chambers, A. (1994) Use of a selection technique to identify the diversity of binding sites for the yeast RAP1 transcription factor. *Nucleic Acids Res* **22**: 124-130.

Hager, K.M., and Yanofsky, C. (1990) Genes expressed during conidiation in *Neurospora crassa*: molecular characterization of *con-13*. *Gene* **96**: 153-159.

Hall, M.N., and Linder, P. (1993) *The Early Days of Yeast Genetics.* Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Hao, D., Ohme-Takagi, M., and Yamasaki, K. (2003) A modified sensor chip for surface plasmon resonance enables a rapid determination of sequence specificity of DNA-binding proteins. *FEBS Lett* **536**: 151-156.

Hardenbol, P., and Van Dyke, M.W. (1996) Sequence specificity of triplex DNA formation: analysis by a combinatorial approach, restriction endonuclease protection selection and amplification. *Proc Natl Acad Sci USA* **93**: 2811-2816.

Hellauer, K., Akache, B., MacPherson, S., Sirard, E., and Turcotte, B. (2002) Zinc cluster protein Rdr1p is a transcriptional repressor of the PDR5 gene encoding a multidrug transporter. *J Biol Chem* **277**: 17671-17676.

Higgins, S.J., and Hames, B.D. (1999) *Protein Expression: a Practical Approach.* Oxford: Oxford University Press.

Hirose, S., Nishizumi, H., and Sakano, H. (2003) Pub, a novel PU.1 binding protein, regulates the transcriptional activity of PU.1. *Biochem Biophys Res Commun* **311**: 351-360.

Hunter, G.D., Jones, I.G., and Sealy-Lewis, H.M. (1996) The cloning and sequencing of the *alcB* gene, coding for alcohol dehydrogenase II, in *Aspergillus nidulans*. *Curr Genet* **29**: 122-129.

Irvine, D., Tuerk, C., and Gold, L. (1991) SELEXION. Systematic evolution of ligands by exponential enrichment with integrated optimization by non-linear analysis. *J Mol Biol* **222**: 739-761.

Iyer, V.R., Horak, C.E., Scafe, C.S., Botstein, D., Snyder, M., and Brown, P.O. (2001) Genomic binding sites of the yeast cell-cycle transcription factors SBF and MBF. *Nature* **409**: 533-538.

Joseph-Horne, T., Hollomon, D.W., and Wood, P.M. (2001) Fungal respiration: a fusion of standard and alternative components. *Biochim Biophys Acta* **1504**: 179-195.

Kaldenhoff, R., and Russo, V.E.A. (1993) Promoter analysis of the *bli-7/eas* gene. *Curr Genet* **24**: 394-399.

Kang, T., Martins, T., and Sadowski, I. (1993) Wild type GAL4 binds cooperatively to the GAL1-10 UASG in vitro. *J Biol Chem* **268**: 9629-9635.

Kannan, K., Amariglio, N., Rechavi, G., Jakob-Hirsch, J., Kela, I., Kaminski, N., *et al.* (2001) DNA microarrays identification of primary and secondary target genes regulated by p53. *Oncogene* **20**: 2225-2234.

Kim, T.K., and Roeder, R.G. (1994) Proline-rich activator CTF1 targets the TFIIB assembly step during transcriptional activation. *Proc Natl Acad Sci USA* **91**: 4170-4174.

Lane, D., Prentki, P., and Chandler, M. (1992) Use of gel retardation to analyze protein-nucleic acid interactions. *Microbiol Rev* **56**: 509-528.

Lauter, F.R., Russo, V.E., and Yanofsky, C. (1992) Developmental and light regulation of *eas*, the structural gene for the rodlet protein of *Neurospora*. *Genes Dev* **6**: 2373-2381.

Lee, K., and Ebbole, D.J. (1998a) Analysis of two transcription activation elements in the promoter of the developmentally regulated *con-10* gene of *Neurospora crassa*. *Fungal Genet Biol* **23**: 259-268.

Lee, K., and Ebbole, D.J. (1998b) Tissue-specific repression of starvation and stress responses of the *Neurospora crassa con-10* gene is mediated by RCO1. *Fungal Genet Biol* **23**: 269-278.

Lety, M.A., Frehel, C., Berche, P., and Charbit, A. (2002) Critical role of the N-terminal residues of listeriolysin O in phagosomal escape and virulence of *Listeria monocytogenes*. *Mol Microbiol* **46**: 367-379.

Lewis, J.C., Feltus, A., Ensor, C.M., Ramanathan, S., and Daunert, S. (1998) Applications of reporter genes. *Anal Chem* **70**: 579A-585A.

Lewis, Z.A., Correa, A., Schwerdtfeger, C., Link, K.L., Xie, X., Gomer, R.H., *et al.* (2002) Overexpression of White Collar-1 (WC-1) activates circadian clock-associated genes, but is not sufficient to induce most light-regulated gene expression in *Neurospora crassa*. *Mol Microbiol* **45**: 917-931.

Lieb, J.D., Liu, X., Botstein, D., and Brown, P.O. (2001) Promoter-specific binding of Rap1 revealed by genome-wide maps of protein-DNA association. *Nat Genet* **28**: 327-334.

Lieberman, P.M., Ozer, J., and Gursel, D.B. (1997) Requirement for transcription factor IIA (TFIIA)-TFIID recruitment by an activator depends on promoter structure and template competition. *Mol Cell Biol* **17**: 6624-6632.

Lindegren, C.C. (1933) The genetics of *Neurospora*-III. Pure bred stocks and crossing over in *N. crassa*. *Bull Torrey Bot Club* **60**: 133-154.

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.

Madi, L., Ebbole, D.J., White, B.T., and Yanofsky, C. (1994) Mutants of *Neurospora crassa* that alter gene expression and conidia development. *Proc Natl Acad Sci USA* **91**: 6226-6230.

Mahanty, S.K., Rao, U.S., Nicholas, R.A., and Scarborough, G.A. (1994) High yield expression of the *Neurospora crassa* plasma membrane H(+)-ATPase in *Saccharomyces cerevisiae*. *J Biol Chem* **269**: 17705-17712.

Mamane, Y., Hellauer, K., Rochon, M.H., and Turcotte, B. (1998) A linker region of the yeast zinc cluster protein Leu3p specifies binding to everted repeat DNA. *J Biol Chem* **273**: 18556-18561.

Margolin, B.S., Freitag, M., and Selker, E.U. (1997) Improved plasmids for gene targeting at the *his-3* locus of *Neurospora crassa* by electroporation. *Fungal Genet Newsl* **44**: 34-36.

Masloff, S., Poggeler, S., and Kuck, U. (1999) The *pro1*(+) gene from *Sordaria macrospora* encodes a C_6 zinc finger transcription factor required for fruiting body development. *Genetics* **152**: 191-199.

Matsuyama, S.S., Nelson, R.E., and Siegel, R.W. (1974) Mutations specifically blocking differentiation of macroconidia in *Neurospora crassa*. *Dev Biol* **41**: 278-287.

Mautino, M.R., Barra, J.L., and Rosa, A.L. (1996) *eth-1*, the *Neurospora crassa* locus encoding S-adenosylmethionine synthetase: molecular cloning, sequence analysis and *in vivo* overexpression. *Genetics* **142**: 789-800.

McClintock, B. (1945) *Neurospora*-I. Preliminary observations of the chromosomes of *Neurospora crassa*. *Am J Bot* **32**: 671-678.

Medintz, I., Wang, X., Hradek, T., and Michels, C.A. (2000) A PEST-like sequence in the N-terminal cytoplasmic domain of *Saccharomyces* maltose permease is required for glucose-induced proteolysis and rapid inactivation of transport activity. *Biochemistry* **39**: 4518-4526.

Melcher, K. (2000) A modular set of prokaryotic and eukaryotic expression vectors. *Anal Biochem* **277**: 109-120.

Melcher, K., and Johnston, S.A. (1995) GAL4 interacts with TATA-binding protein and coactivators. *Mol Cell Biol* **15**: 2839-2848.

Michan, S., Lledias, F., Baldwin, J.D., Natvig, D.O., and Hansberg, W. (2002) Regulation and oxidation of two large monofunctional catalases. *Free Radic Biol Med* **33**: 521-532.

Nallur, G.N., Prakash, K., and Weissman, S.M. (1996) Multiplex selection technique (MuST): an approach to clone transcription factor binding sites. *Proc Natl Acad Sci USA* **93**: 1184-1189.

Nelson, M.A., Kang, S., Braun, E.L., Crawford, M.E., Dolan, P.L., Leonard, P.M., *et al.* (1997) Expressed sequences from conidial, mycelial, and sexual stages of *Neurospora crassa*. *Fungal Genet Biol* **21**: 348-363.

Nielsen, H., Engelbrecht, J., Brunak, S., and von Heijne, G. (1997) Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Engineering* **10**: 1-6.

Nowrousian, M., Kuck, U., Loser, K., and Weltring, K.M. (2000) The fungal *acl1* and *acl2* genes encode two polypeptides with homology to the N- and C-terminal parts of the animal ATP citrate lyase polypeptide. *Curr Genet* **37**: 189-193.

Olson, S.A. (2002) EMBOSS opens up sequence analysis. European Molecular Biology Open Software Suite. *Brief Bioinform* **3**: 87-91.

Panozzo, C., Capuano, V., Fillinger, S., and Felenbok, B. (1997) The zinc binuclear cluster activator AlcR is able to bind to single sites but requires multiple repeated sites for synergistic activation of the *alcA* gene in *Aspergillus nidulans*. *J Biol Chem* **272**: 22859-22865.

Pierrou, S., Enerback, S., and Carlsson, P. (1995) Selection of high-affinity binding sites for sequence-specific, DNA binding proteins from random sequence oligonucleotides. *Anal Biochem* **229**: 99-105.

Pietschmann, S., Eberle, J., Lauter, F.R., Pandit, N.N., and Russo, V.E.A. (1991) Co-regulation of two tandem genes by one blue-light element in *Neurospora crassa*. *Fungal Genet Newsl* **38**: 85-86.

Pollock, R., and Treisman, R. (1990) A sensitive method for the determination of protein-DNA binding specificities. *Nucleic Acids Res* **18**: 6197-61204.

Ptashne, M. (1988) How eukaryotic transcriptional activators work. *Nature* **335**: 683-689.

Pugh, T.A., and Clancy, M.J. (1990) Differential regulation of *STA* genes of *Saccharomyces cerevisiae*. *Mol Gen Genet* **222**: 87-96.

Rechsteiner, M., and Rogers, S.W. (1996) PEST sequences and regulation by proteolysis. *Trends Biochem Sci* **21**: 267-271.

Reece, R.J., and Ptashne, M. (1993) Determinants of binding-site specificity among yeast C_6 zinc cluster proteins. *Science* **261**: 909-911.

Reese, J.C. (2003) Basal transcription factors. *Curr Opin Genet Dev* **13**: 114–118.

Ren, B., Robert, F., Wyrick, J.J., Aparicio, O., Jennings, E.G., Simon, I., *et al.* (2000) Genome-wide location and function of DNA binding proteins. *Science* **290**: 2306-2309.

Robert, A.M., Harry, W.J., and Suchareeta, M. (2003) Methods for transcription factor separation. *J Chromatogr B* **797**: 269–288.

Roberts, A.N., Berlin, V., Hager, K.M., and Yanofsky, C. (1988) Molecular analysis of a *Neurospora crassa* gene expressed during conidiation. *Mol Cell Biol* **8**: 2411-2418.

Roberts, A.N., and Yanofsky, C. (1989) Genes expressed during conidiation in *Neurospora crassa*: characterization of *con-8*. *Nucleic Acids Res* **17**: 197-214.

Rogers, S., Wells, R., and Rechsteiner, M. (1986) Amino-acid-sequences common to rapidly degraded proteins - the Pest hypothesis. *Science* **234**: 364-368.

Rosinski, J.A., and Atchley, W.R. (1999) Molecular evolution of helix-turn-helix proteins. *J Mol Evol* **49**: 301-309.

Roulet, E., Busso, S., Camargo, A.A., Simpson, A.J., Mermod, N., and Bucher, P. (2002) High-throughput SELEX SAGE method for quantitative modeling of transcription-factor binding sites. *Nat Biotechnol* **20**: 831-835.

Rutter, J., Michnoff, C.H., Harper, S.M., Gardner, K.H., and McKnight, S.L. (2001) PAS kinase: an evolutionarily conserved PAS domain-regulated serine/threonine kinase. *Proc Natl Acad Sci USA* **98**: 8991-8996.

Sakai, Y., Nakagawa, R., Sato, R., and Maeda, M. (1998) Selection of DNA binding sites for human transcriptional regulator GATA-6. *Biochem Biophys Res Commun* **250**: 682-688.

Sambrook, J., and Russell, D.W. (2001) *Molecular Cloning: a Laboratory Manual.* Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Schjerling, P., and Holmberg, S. (1996) Comparative amino acid sequence analysis of the C_6 zinc cluster family of transcriptional regulators. *Nucleic Acids Res* **24**: 4599-4607.

Schuller, H.J., Schutz, A., Knab, S., Hoffmann, B., and Schweizer, E. (1994) Importance of general regulatory factors Rap1p, Abf1p and Reb1p for the activation of yeast fatty acid synthase genes FAS1 and FAS2. *Euro J Biochem* **225**: 213-222.

Schulze, A., and Downward, J. (2001) Navigating gene expression using microarrays-a technology review. *Nature Cell Biol* **3**: E190-E195.

Scott, W.A. (1976) Biochemical genetics of morphogenesis in *Neurospora*. *Annual Rev Microbiol* **30**: 85-104.

Selitrennikoff, C.P. (1976) *Easily-wettable,* a new mutant. *Neurospora Newsl* **23**: 23.

Selker, E.U., Cambareri, E.B., Jensen, B.C., and Haack, K.R. (1987) Rearrangement of duplicated DNA in specialized cells of *Neurospora*. *Cell* **51**: 741-752.

Shalev, A., Pise-Masison, C.A., Radonovich, M., Hoffmann, S.C., Hirshberg, B., Brady, J.N., and Harlan, D.M. (2002) Oligonucleotide microarray analysis of intact human pancreatic islets: identification of glucose-responsive genes and a highly regulated TGFbeta signaling pathway. *Endocrinology* **143**: 3695-3698.

Sheibani, N. (1999) Prokaryotic gene fusion expression systems and their use in structural and functional studies of proteins. *Prep Biochem Biotechnol* **29**: 77-90.

Soloviev, M.M. (2000) Detection of a surface-exposed PEST like sequence in the metabotropic glutamate receptor mGluR1α. *Bioinformatics* **16**: 837-838.

Springer, M.L. (1993) Genetic control of fungal differentiation: the three sporulation pathways of *Neurospora crassa*. *BioEssays* **15**: 365-374.

Springer, M.L., and Yanofsky, C. (1989) A morphological and genetic analysis of conidiophore development in *Neurospora crassa*. *Genes Dev* **3**: 559-571.

Springer, M.L., and Yanofsky, C. (1992) Expression of *con* genes along the three sporulation pathways of *Neurospora crassa*. *Genes Dev* **6**: 1052-1057.

Standaert, S.M., Schaffner, W., Galgiani, J.N., Pinner, R.W., Kaufman, L., Durry, E., and Hutcheson, R.H. (1995) Coccidioidomycosis among visitors to a *Coccidioides immitis*-endemic area: an outbreak in a military reserve unit. *J Infect Dis* **171**: 1672-1675.

Steinberg, G. (2000) The cellular roles of molecular motors in fungi. *Trends Microbiol* **8**: 162-168.

Strauss, J., Horvath, H.K., Abdallah, B.M., Kindermann, J., Mach, R.L., and Kubicek, C.P. (1999) The function of CreA, the carbon catabolite repressor of *Aspergillus nidulans*, is regulated at the transcriptional and post-transcriptional level. *Mol Microbiol* **32**: 169-178.

Strauss, J., Muro-Pastor, M.I., and Scazzocchio, C. (1998) The regulator of nitrate assimilation in ascomycetes is a dimer which binds a nonrepeated, asymmetrical sequence. *Mol Cell Biol* **18**: 1339-1348.

Sudbery, P.E. (1996) The expression of recombinant proteins in yeasts. *Curr Opin Biotechnol* **7**: 517-524.

Sweigard, J.A., Chumley, F., Carroll, A., Farrall, L., and Valent, B. (1997) A series of vectors for fungal transformation. *Fungal Genet Newsl* **44**: 52-53.

Terpe, K. (2003) Overview of tag protein fusions: from molecular and biochemical fundamentals to commercial systems. *Appl Microbiol Biotechnol* **60**: 523-533.

Thiesen, H.J., and Bach, C. (1990) Target Detection Assay (TDA): a versatile procedure to determine DNA binding sites as demonstrated on SP1 protein. *Nucleic Acids Res* **18**: 3203-3209.

Thorn, K.S., Naber, N., Matuska, M., Vale, R.D., and Cooke, R. (2000) A novel method of affinity-purifying proteins using a bis-arsenical fluorescein. *Protein Sci* **9**: 213-217.

Thut, C.J., Chen, J.L., Klemm, R., and Tjian, R. (1995) p53 transcriptional activation mediated by coactivators TAFII40 and TAFII60. *Science* **267**: 100-104.

Todd, R.B., and Andrianopoulos, A. (1997) Evolution of a fungal regulatory gene family: the $Zn(II)_2Cys_6$ binuclear cluster DNA binding motif. *Fungal Genet Biol* **21**: 388-405.

Turian, G., and Bianchi, D.E. (1972) Conidiation in *Neurospora*. *Bot Rev* **38**: 119-154.

Vollmer, S.J., and Yanofsky, C. (1986) Efficient cloning of genes of *Neurospora crassa*. *Proc Natl Acad Sci USA* **83**: 4869-4873.

Wang, Z., Yang, P., Fan, B., and Chen, Z. (1998) An oligo selection procedure for identification of sequence-specific DNA-binding activities associated with the plant defence response. *Plant J* **16**: 515-522.

Wang, Z.Y., Thornton, C.R., Kershaw, M.J., Debao, L., and Talbot, N.J. (2003) The glyoxylate cycle is required for temporal regulation of virulence by the plant pathogenic fungus *Magnaporthe grisea*. *Mol Microbiol* **47**: 1601-1612.

Warren, A.J. (2002) Eukaryotic transcription factors. *Curr Opin Struct Biol* **12**: 107–114.

Wells, J., Graveel, C.R., Bartley, S.M., Madore, S.J., and Farnham, P.J. (2002) The identification of E2F1-specific target genes. *Proc Natl Acad Sci USA* **99**: 3890-3895.

White, B.T., and Yanofsky, C. (1993) Structural characterization and expression analysis of the *Neurospora* conidiation gene *con-6*. *Dev Biol* **160**: 254-264.

Wright, W.E., Binder, M., and Funk, W. (1991) Cyclic amplification and selection of targets (CASTing) for the myogenin consensus binding site. *Mol Cell Biol* **11**: 4104-4110.

Wyrick, J.J., Aparicio, J.G., Chen, T., Barnett, J.D., Jennings, E.G., Young, R.A., *et al.* (2001) Genome-wide distribution of ORC and MCM proteins in *S. cerevisiae:* high-resolution mapping of replication origins. *Science* **294**: 2357-2360.

Xie, X. (2003) Sugar sensing and regulation of conidiation in *Neurospora crassa*. Ph.D. Dissertation. Texas A&M University, College Station.

Young, E.T., Sloan, J., Miller, B., Li, N., van Riper, K., and Dombek, K.M. (2000) Evolution of a glucose-regulated ADH gene in the genus *Saccharomyces*. *Gene* **245**: 299-309.

Yuan, G.F., Fu, Y.H., and Marzluf, G.A. (1991) *nit-4*, a pathway-specific regulatory gene of *Neurospora crassa*, encodes a protein with a putative binuclear zinc DNA-binding domain. *Mol Cell Biol* **11**: 5735-5745.

Zhang, L.H., Liu, D.P., and Liang, C.C. (2003) Finding regulatory sequences. *Int J Biochem Cell Biol* **35**: 95-103.

Zhang, Y., Lamm, R., Pillonel, C., Lam, S., and Xu, J. (2002) Osmoregulation and fungicide resistance: the *Neurospora crassa os-2* gene encodes a HOG1 mitogen-activated protein kinase homologue. *Appl Environ Microbiol* **68**: 532-538.

Zink, M.W. (1969) Regulation of NAD-specific alcohol dehydrogenases in *Neurospora crassa*. *Can J Microbiol* **15**: 265-271.

VITA

Panan Rerngsamran

Address: 207/17 Tiwanon 32, Tiwanon Road, Muang, Nonthaburi, 11000 Thailand

Education:

1998-2004 Texas A&M University, College Station
Ph.D. in plant pathology, May 2004

1988-1992 Chulalongkorn University, Bangkok, Thailand
M.S. in industrial microbiology, June 1992

1984-1988 KhonKaen University, KhonKaen, Thailand
B.S. in biology, March 1988

Experience:

1996-1998 Lecturer, Department of Microbiology, Faculty of Science
Chulalongkorn University, Bangkok Thailand

1995-1996 Trainee, The UNESCO International Post-Graduate University Courses in Microbiology, Osaka University, Osaka, Japan

1992-1995 Researcher, Soil Microbiology Research Group, Division of Soil Science, Bangkok, Thailand