INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

Bell & Howell Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
800-521-0600

UMI[®]

Characterization of the *sbr* gene in *Neurospora crassa*

Yanhua Yan

**A thesis submitted to the Faculty of Graduated Studies and Research
In partial fulfillment of the requirements for the degree of
Master of Science**

**Ottawa-Carleton Institute of Biology
Carleton University
Ottawa, Ontario
Canada**

January 2001

National Library
of Canada

Acquisitions and
Bibliographic Services

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque nationale
du Canada

Acquisitions et
services bibliographiques

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file *Votre référence*

Our file *Notre référence*

The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

L'auteur conserve la propriété du droit d'auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

0-612-57792-9

Canada

The undersigned hereby recommend to the Faculty of Graduate Studies and Research
acceptance of the thesis

“Characterization of the *sbr* gene in *Neurospora crassa*”

submitted by Yanhua Yan
in partial fulfillment of the requirements for the degree of
Master of Science

Thesis Supervisor

Chairman, Department of Biology

Carleton University

Jan 18, 2001
Date

Abstract

Insertional mutation of the *Neurospora crassa* (*N. crassa*) small brown (*sbr*) gene results in a striking, colonial morphology. This gene was cloned previously, but only one 5' truncated cDNA sequence had been recovered by conventional screening of several libraries (Campsell, 1998). The cloned *sbr* gene was proposed as a regulatory gene because it encodes a helix-turn-helix (HTH), a putative zinc finger, and a glutamine-rich region. This study completed the *sbr* gene sequence and mapped the gene to linkage group III. The *sbr* gene has one intron and encodes a protein of 612 amino acids with two glutamine-rich regions at the N-terminal. Complementation with a *sbr* genomic clone recovered the wild type phenotype. SBR protein was expressed and purified under native condition. In a protein-DNA binding assay in which total genomic DNA digested by restriction enzymes was allowed to bind with SBR protein, three genomic fragments from size 0.3kb to 0.8kb were isolated from protein-DNA complexes. Electrophoretic mobility shift assay indicated SBR-DNA binding was specific. Protein from an expression clone with deletion of the putative zinc finger lost its DNA binding ability. Searching the *N. crassa* genome database retrieved a sequence containing the 0.3kb fragment. The possibility that the isolated fragments contain the SBR responsive regulatory elements has been suggested with bioinformatic programs in prediction of genes associated with these fragments and in searching transcription factor binding sites. Some expression plasmids with deletion at different part of the *sbr* gene were cloned for further study in determining functional domains.

To my wife **Zaiyu**

Acknowledgments

First of all, I wish to express my deep gratitude to my excellent supervisor Dr. P. John Vierula for his guidance and assistance during the development of this thesis.

I thank my committee members, Dr. Myron Smith and Dr. Doug Johnson, for their involvement and suggestions.

Thanks also go to grad students Weiping Chen and Alex Valencia who shared the lab and gave me help whenever available.

Table of content

TITLE.....	I
ACCEPTANCE PAGE	II
ABSTRACT	III
ACKNOWLEDGMENTS.....	V
TABLE OF CONTENT	VI
LIST OF FIGURES.....	IX
LIST OF TABLES.....	X
LIST OF ABBREVIATIONS	XI
INTRODUCTION	1
<i>NEUROSPORA CRASSA</i> AS A MODEL FOR MOLECULAR GENETIC STUDIES	1
MORPHOLOGICAL MUTANTS OF <i>N. CRASSA</i>	2
A NEW MUTANT OF <i>N. CRASSA</i>	3
TRUNCATED cDNA SEQUENCE OF THE <i>sbr</i> GENE IDENTIFIES A POSSIBLE REGULATORY	
GENE.....	6
STUDYING REGULATORY PROTEINS.....	9
A. Expression and purification of histidine-tagged regulatory protein.	9
B. Electrophoretic mobility shift assay.....	11
C. Footprinting.....	12
D. Identifying the DNA-binding region of a DNA binding protein.....	13
E. Identifying the DNA sequence bound to a putative transcription factor	14
OBJECTIVES OF THIS STUDY.....	16
MATERIALS AND METHODS	18
CLONING <i>sbr</i> GENOMIC DNA INTO PLASMID.....	18
A. Extracting DNA from phage clone isolated from genomic library.....	18
B. Subcloning <i>sbr</i> genomic DNA.	19
TRANSFORMATION OF <i>E. COLI</i> CELLS.....	19
A. Chemical transformation.....	19
B. Electroporation.	20
EXTRACTION OF PLASMID DNA FROM <i>E. COLI</i>	21
DNA SEQUENCING	22
COMPLEMENTATION STUDY: TRANSFORMING <i>sbr</i> MUTANT	23

A. Preparation of <i>sbr</i> spheroplasts.....	23
B. Electroporation of <i>sbr</i> :.....	24
GENE MAPPING.....	25
A. Growing map strains.....	25
B. Extraction of genomic DNA	25
C. Hybridization	26
SOUTHERN BLOT	26
RADIO LABELING DNA PROBE.....	27
HYBRIDIZATION AND FILM DEVELOPMENT.....	28
RT-PCR.....	29
A. Extraction of total RNA.....	29
B. Reverse transcription.....	30
PCR PROTOCOL	31
PURIFICATION OF SBR PROTEIN UNDER NATIVE CONDITION.....	31
A. Transformation of bacteria.....	31
B. Expression of SBR protein.....	32
C. Protein purification.....	32
SDS-POLYACRYLAMIDE GEL ELECTROPHORESIS (PAGE)	33
WESTERN BLOT AND IMMUNOSTAINING	34
ISOLATING SBR BINDING TARGET DNA ELEMENTS.....	35
ELECTROPHORESIS MOBILITY SHIFT ASSAY (EMSA).....	37
A. preparation of probe DNA:.....	37
B. Preparation of non-specific competitor.....	38
C. Protein-DNA binding.....	38
D. Electrophoresis	38
DNA ANALYSIS WITH BIOINFORMATIC PROGRAMS	38
RESULTS.....	41
GENOMIC CLONE AND SEQUENCE.....	41
<i>sbr</i> cDNA SEQUENCE.....	44
A. Amplification of <i>sbr</i> cDNA 5' end.....	44
B. Sequence of the RT-PCR product and the translation start site.....	48
AMINO ACID COMPOSITION OF SBR PROTEIN	49
CONSTRUCTION OF <i>sbr</i> ORFs AND EXPRESSION PLASMIDS:	50
SBR PROTEIN EXPRESSION AND PURIFICATION.....	54
DNA BINDING STUDY	57
A. Isolation of the SBR binding fragments from genomic DNA.....	57
B. Electrophoresis mobility shift assay (EMSA).....	58
SEQUENCE ANALYSIS OF THE ISOLATED FRAGMENTS.....	61
A. Sequence analysis of SBRBD300	61
B. Sequence analysis of SBRBD600.....	63
C. Sequence analysis of SBRBD800	63
D. Binding sites search.....	68

COMPLEMENTATION WITH <i>sbr</i> GENOMIC DNA	70
<i>Confirmation of sbr transformation.....</i>	<i>70</i>
GENE MAPPING.....	73
DISCUSSION.....	74
CDNA SEQUENCE AND TRANSLATION START SITE.	74
No <i>sbr</i> mRNA IN THE MUTANT STRAIN	76
COMPLEMENTATION WITH PLASMID GENOMIC CLONE	76
PUTATIVE TRANSCRIPTION ACTIVATION DOMAINS IN SBR PROTEIN.	77
SBR IS A DNA BINDING PROTEIN	79
<i>A. DNA binding property.....</i>	<i>79</i>
<i>B. The putative zinc finger mediates DNA binding?.....</i>	<i>82</i>
TARGET GENE PREDICTION AND BINDING SITE(S) ANALYSIS	85
<i>A. Gene prediction.....</i>	<i>85</i>
<i>B. SBR target binding sites analysis</i>	<i>87</i>
CONCLUSION	90
REFERENCE.....	92

List of Figures

Figure 1.	Morphology of wild type <i>N. crassa</i> and the <i>sbr</i> mutant.....	5
Figure 2.	Homology between the predicted SBR protein and regulatory proteins	8
Figure 3.	Different genomic clones.....	42
Figure 4.	Sequence and translation of the 5 end of <i>sbr</i> gene	43
Figure 5.	RT-PCR results.....	45
Figure 6.	The <i>sbr</i> gene and flanking sequence at 5 end	46
Figure 7.	Diagram of constructing the completed <i>sbr</i> ORF.....	51
Figure 8.	Diagram of constructing expression clones.....	52
Figure 9.	Examples showing solubility and purification of SBR proteins.....	55
Figure 10.	Purified SBR proteins.....	56
Figure 11.	The sequences of three isolated genomic fragments	59
Figure 12.	Results of EMSA.....	60
Figure 13.	Gene prediction for SBRBD300 and downstream sequence	64
Figure 14.	Gene prediction for SBRBD300 and upstream sequence.....	65
Figure 15.	Analysis of SBRBD600.....	66
Figure 16.	Analysis of SBRBD800.....	67
Figure 17.	Schematic diagram confirming the specificity of complementary study	71
Figure 18.	Mapping result for <i>sbr</i> gene.....	72

List of Tables

Table 1. The amino acid composition of SBR protein.....	47
Table 2. Details of constructing different <i>sbr</i> expression clones	53
Table 3. DNA binding sites of six transcription factors.....	69

List of Abbreviation

bp	base pair
BSA	bovine serum albumin
dATP	deoxyadenosine triphosphate
dCTP	deoxycytidine triphosphate
dGTP	deoxyguanosine triphosphate
dH ₂ O	distilled water
DMSO	dimethyl sulfoxide
dNTP	dinucleotide triphosphate
dTTP	deoxythymidine triphosphate
EDTA	N, N-ethylenediaminetetraacetic acid
EMSA	electrophoretic mobility shift assay
hr	hour
IPTG	isopropyl thio beta-D-galactosidase
kb	kilobase pair
kD	kilodalton
mw	molecular weight
min	minutes
ng	nanogram
Nt	nucleotides
OD	optical density
ORF	open reading frame
PAGE	polyacrylamide gel electrophoresis
PCI	phenol:chloroform:isoamyl alcohol (25:24:1 in v/v)
PEG	polyethylene glycol
PTC	40% PEG/50 mM Tris, pH 8/50 mM CaCl ₂
rpm	revolution per minute
RACE	rapid amplification of cDNA ends
RT	reverse transcription
SDS	sodium dodecyl sulfate
sec	seconds
SSC	standard sodium citrate
STC	sorbitol/tris/CaCl ₂ solution
TBE	Tris/borate/EDTA buffer
TBS	tris buffered saline
TE	10mM Tris, 1 mM EDTA
TEMED	N,N,N',N'-tetramethylethylenediamine
Tris	Tris(hydroxymethyl) aminomethane
μl	microlitre
v	volts
VMM	Vogel's minimal medium
W	watts
X-gal	5-bromo-4-chloro-3-indolyl-β-D-galactosidase

Introduction

***Neurospora crassa* as a model for molecular genetic**

One of the principle characteristics of filamentous fungi is that they grow as threads of cells (hyphae) and are propagated by generative and, or vegetative spores. Spores germinate with the outgrowth one or more germ tubes which extend only at their tips. These hyphae branch repeatedly and, in most fungal groups, are periodically subdivided by septae (transverse cell walls). Fusions occur readily between hyphal branches derived for the same spore, or between genetically compatible individual strains. Hyphae formed by the fusion of genetically distinct but compatible nuclei are referred to as heterokaryons. The network of hyphae resulting from these processes of tip growth, branching and fusion is called a mycelium.

One of the most extensively studied of all filamentous fungi is *Neurospora crassa*, a member of the division *Ascomycotina*. It was first named and described in detail by Shear and Dodge (1927) and was used by Beadle and Tatum for their development of the one-gene one-enzyme hypothesis. It has a life cycle encompassing both a vegetative (asexual) stage and a sexual stage (reviewed by Springer 1993). The tubular hyphae have cell walls composed of a protein layer sandwiched between an outer amorphous network of glycoproteins and glucans and an inner layer of chitin. *Neurospora* hyphae have cross walls (septae) containing central pores that allow transport of cytoplasm, mitochondria, and even nuclei between cell compartments. Each compartment usually has multiple nuclei and the tip cell contains at least one apical

nucleus which migrates in step with the growing tip (reviewed by: Collinge and Trinci, 1974; Trinci 1978; Davis, 2000).

During the asexual cycle, aerially disseminated, multinucleate, macroconidial spores are produced (reviewed by Springer and Yanofsky 1989). Sexual reproduction occurs when macroconidia of one mating type fuse with specialized hyphae, or trichogynes, associated with the perithecium, a mating structure consisting of differentiated, coiled hyphae, produced by the other mating type. Fertilization occurs when the nuclei from each strain are united in the perithecium, to form a dikaryon (reviewed by Springer, 1993). The nuclei undergo cycles of division and meiosis until the final product is created: numerous asci each containing 8 haploid ascospores. The ascospores are then ejected from the perithecium to complete the cycle.

N. crassa is easily manipulated using standard microbiological techniques, has simple nutritional requirements and is readily transformable with a number of selectable markers. It has a relatively small, 47 Mbp genome arranged into 7 linkage groups (chromosomes), with only about 8% repetitive sequence (Krumlauf and Marzluf, 1980). In addition to approximately 1,000 classically mapped genetic loci, approximately 15,000 ESTs and two chromosomes have already been sequenced. The remainder of the genome is expected to be completed within 1 year. Therefore, it is a very good model organism for genetic studies and amenable to a wide range of approaches.

Morphological mutants of *N. crassa*

Since the days of Beadle and Tatum, many *N. crassa* morphological mutants have been produced by chemical or UV mutagenesis and selection for aberrant growth

phenotypes. They usually exhibit a restricted colonial growth, aberrant conidiation, or abnormal morphology in sexual development stages (reviewed by: Springer and Yanofsky, 1989; Springer 1993; Russo and Pandit 1992). Many morphological mutants share certain characteristic abnormalities including an increased frequency of branching and slower rate of elongation at the hyphal tip. Some of these mutants have changes in the amount of cell wall components such as carbohydrates or peptides (de Terra and Tatum, 1963, Wrathall and Tatum, 1974). For example, the colonial mutant, *col-1*, contains 11% less glucose and nearly 250% more glucosamine per unit weight of cell wall compared the wild type (de Terra and Tatum, 1961). A host of potential defects in primary metabolic pathways, cytoskeletal assembly and function and protein secretion could affect hyphal wall synthesis and morphogenesis.

Like genetic disruptions, many chemical agents are also capable of inducing colonial growth by targeting different biological components of *Neurospora* morphogenesis (for review see Mishra, 1977). This indicates that morphology is a result of many processes from biochemical pathways to structural and regulatory components of cell wall biosynthesis, and mutation in a single gene may affect many aspects of morphological development.

A new mutant of *N. crassa*.

An alternative approach to discovering new genes and studying their functions is insertional mutagenesis. This technique disrupts the activity of a gene, usually producing a null phenotype, and also tags the mutation, facilitating subsequent cloning and characterization of the gene (Paietta and Marzluf, 1985, Staben et al., 1989). Insertional mutagenesis with transposable elements has been used extensively in both bacterial and

plant genetics. Although suitable, fungal transposons have not yet been developed, insertional mutagenesis has been adapted by relying on the propensity of dominant selectable markers to integrate relatively randomly into the genome by non-homologous recombination.

One of these new mutants of *N. crassa*, created by insertional mutagenesis with a hygromycin resistance construct, has an extreme, colonial morphology (Campsall, 1998). Unlike wild type, which rapidly extends hyphae radially from the inoculation site on the solid minimal media, this mutant exhibited slow, highly restricted mycelial growth. A single colony of this strain forms rough-edged circular mound marked with a glistening, irregular surface. The orange-brown color of the mutant, which develops as the colonies mature, differs from the pink-orange color of wild type. Therefore the new mutant was named *sbr* (small, brown) (Figure 1).

The phenotype and development of *sbr* are distinct from most other colonial strains. Its hyphae are swollen and spherically shaped, much wider than the thin tubular wild type hyphae. In addition *sbr* appears to lack conidiophores and macroconidia. Under Scanning electronic microscopy, wild type hyphae are thin and tubular with diameters of 4 μm , while *sbr* colonies consist of fat, rounded hyphae with diameters ranging from 20 μm to 80 μm . From these tuberous projections arise smaller yeast-like buds (Campsall, 1998).

Microtubules of wild type are parallel to the direction of hyphal growth and extend from the hyphal tip along its entire length. In contrast, the *sbr* strain has numerous, randomly arranged, short microtubules with no discernible pattern. In addition, while wild type, shows strong chitin staining primarily at septa, the *sbr* strain

Figure 1. Morphology of wild type *N. crassa* and the *sbr* mutant. Pictures were adapted from Campsall (1998). (A) the mutant shows highly restricted hyphae growth and forms a small colony while the wild type (wt) hyphae grow radially over the plate. (B) scanning electron microscopy shows tubular hyphae (B1) and conidiophore ('c' in B2) of the wild type and the swelling spherical hyphae of the mutant.

A

B

shows intense chitin deposition in both walls and septa (Campsall, 1998).

This mutation also impacts germination of sexual spores. Mating between strain 987 (type A) and *sbr* (type a) produced only ascospores which germinated and grew normally, but more than half of all spores failed to germinate (Campsall, 1998).

Truncated cDNA sequence of the *sbr* gene identifies a possible regulatory gene.

A genomic DNA fragment containing the insertional tag sequence and its flanking *sbr* genomic sequence was cloned by screening the *sbr* genomic library with a probe from the insertional tag sequence (Campsall, 1998). The flanking *sbr* genomic sequence was then used as probe to screen a *N. crassa* cDNA library and then the *sbr* cDNA was cloned. However, the sequence and northern blot indicated the cloned cDNA was truncated (Campsall, 1998).

Sequence analysis indicated that there is no overall homology between the deduced SBR amino acid sequence and other known sequences. However, there are three regions with sequences common to other regulatory proteins: the helix-turn-helix DNA binding motif, glutamine-rich region, and cysteine-rich region (Figure 2) (Campsall 1998).

The helix-turn-helix DNA binding motif has been found to occur in numerous regulatory proteins including homeobox containing transcription factors (Kornberg, 1993, reviewed by Latchman 1998). It has three conserved residues (alanine/glycine, glycine, and valine/isoleucine) that are believed to be important for the stability of the bi-helical complex (Brennan and Mathews, 1989).

The middle glycine residue corresponds to a turn in the secondary structure whereas the other two conserved residues are each part of an α -helix that surrounds the turn. Not only does SBR have conservation in these residues, but it also has the expected secondary structure. The motif was first proposed for a small number of regulatory proteins expressed by phage but it has since been identified in regulatory proteins of other organisms (for review, see Brennan and Mathews, 1989, Harrison and Aggarwal 1990). However, the presence of a motif does not give definitive proof as to its function. For example, the *S. cerevisiae* protein DAL81 also contains this motif but deletion of the sequence did not affect the function of the gene's product (Bricmont et al. 1991).

Structures similar to the poly-glutamine stretch and cysteine-rich residues of SBR are also found in a large number of gene products including other fungal proteins such as WC-1 (Ballario et al, 1996), NIT4 (Yuan et al., 1991), and DAL81 (Bricmont and Cooper, 1989). WC1 is a transcription factor that binds to and affects light regulation of the albino-3 gene of *N. crassa* (Ballario et al, 1996). The NIT4 protein of *N. crassa* appears to regulate structural genes at the transcriptional level via the nitrogen regulatory circuit (Yuan et al 1991). Deletions of these domains result in loss of function.

In yeast, DAL81 is believed to be a regulator of the allantoin degradation pathway at the transcription level and also encodes both of these domains (Bircmont et al, 1991). Deletion of one of its polyglutamine stretches results in 50% loss of the activity whereas deletion of zinc finger cysteine residues does not influence function (Bircmont et al, 1991).

Figure 2. Homology between the predicted SBR protein and regulatory proteins containing conserved domains of (A) glutamine-rich residues, (B) cysteine-rich residues, and (C) helix-turn-helix motif. Highlighted residues indicate conserved amino acids. NIT4 and WC1 are *N. crassa* regulatory proteins (Yuan et al., 1991; Ballario et al., 1996). DAL81 is a yeast regulatory protein (Bricmont et al., 1991). DOF is a transcription factor in plants (Yanagisawa and Schmidt, 1999). Others (λ REP, λ CRO, CAP, 434R, 434CRO) are from phage or *E. coli* (see review in Harrison and Aggarwal, 1990).

A

WC1	QHQMHHQHQQQQQQQQQQQQQQQQQQQQQQQQQQQQQHQQQQQ
NIT4	QQQQQRQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQ
DAL81	QQQQHQQQQQQQQQQQQQQQQQQQQ
SBR	QHQQHQHQQQQHQQHHQQQQQQQHQQQQQHQQQQQQ

B

NIT4	CIACRRRKSKCDGALPSCAACASVYGTEC
WC1	CANCHTRNTPEWRRGPSGNRDLNCS
SBR	CKNCHKGHGPWTSCIVVDGQMCWSCANCWFNASGARC
DOF	CPRCGSRDTKFCYNNYNTSQPRHLCKSC

C

λREP	Q E S V A D K M G M G Q S G V G A L
λCRO	Q T K T A K D L G V T Q S A I N K A
CAP	R Q E I G Q I V G C S R E T V G R I
434R	Q A E L A Q K V G T T Q Q S I E Q L
434CRO	Q T E L A T K A G V K Q Q S I Q L I
DAL81	D K T C A L N E G R Q S H L I L G R
SBR	T E D M A H E S G Y G Q L N V E S A

The cysteine-rich residues from the above proteins specify putative zinc finger DNA binding proteins that belong to different classes as determined by their amino acid sequence and secondary folding structure. WC1 encodes a single class IV zinc finger protein in which a single zinc ion is linked to four cysteines (Ballario et al 1996). Class III members include GATA factors, which are characterized by the occurrence of two zinc finger proteins, which bind a total of two zinc ions (Sanchez-Garcia and Rabbitts, 1994). In contrast, both NIT4 and DAL81 are considered class III zinc fingers, also known as the $\text{Zn}(11)_2\text{Cys}_6$ binuclear cluster DNA binding motif. They are believed to form a cloverleaf structure by arrangement of their six cysteine residues around two zinc ions (Todd and Andrianopoulos, 1997). This binding is unique to over 80 fungal proteins, which generally act as transcription activators.

The cysteine-rich region of SBR does not correspond precisely to either of these classes nor to any other class of zinc finger protein. Despite this discrepancy, it is interesting that the cysteine residues in SBR are all clustered together in the same region. Since the sequence of this region does not match any known zinc fingers precisely, it may represent a non-functional motif or belong to a new class of zinc finger DNA binding proteins (Campsall, 1998).

Studying regulatory proteins

A. Expression and purification of histidine-tagged regulatory protein.

To study the biological features of a regulatory protein such as the structure, function, and the mode of its interaction with DNA and/or other regulatory proteins, the first important step is to express and purify the protein. Since transcription factors are not

usually very abundant, they can be over-expressed in heterologous cells such as bacteria, yeast or insect cells. The recombinant protein can then be used in DNA binding, protein-protein interaction, and *in vitro* transcription assays. Although classical purification protocols may be used to recover recombinant proteins following expression, recombinant DNA technology also permits the engineering of fusion proteins bearing specific affinity tags. These tags greatly simplify the purification of recombinant proteins, which usually involves a single chromatography step with an appropriate affinity resin.

Metal chelate affinity chromatography has been known to be a useful technique for protein purification since 1975 (Porath et al, 1975). Over the last decade the development of improved chelating agents for immobilizing metal ions on solid supports (Hoculi et al, 1987) has resulted in widespread application of the technology. The method is based on the interaction between histidine residues and electropositive transition metals, such as Ni^{2+} , Co^{2+} and Zn^{2+} (reviewed by Porath, 1992). Six consecutive histidine residues will bind very tightly to these metals under condition of physiological pH, even in the presence of strong denaturing agents. Since virtually no naturally occurring proteins contain multiple, neighboring histidine residues, a single purification step removes most contaminants from recombinant proteins bearing a six-histidine tag.

Histidine-tagged proteins can be purified following synthesis in prokaryotic or eukaryotic systems. Since the metal-histidine interaction is conformation independent, the purification can be carried out under native or denaturing conditions. The tag itself is small, uncharged at physiological pH and poorly immunogenic, so it does not interface with structure, function, or secretion of the recombinant protein and need not to be removed by cleavage (see the handbook of QIAexpression system, QIAGEN). For

example, when an yeast proliferating cell nuclear protein was expressed with and without a six-residue histidine tag at its amino-terminus, both forms of the recombinant protein were found to be biologically active and no significant differences were observed between the two forms (Biswas et al., 1995). Other examples of using histidine-tagged protein in DNA binding assay include Kato et al. (1994), Wollscheid et al. (1997) and Dabrowski and Kur (1999).

B. Electrophoretic mobility shift assay

Many methods have been developed in the study of DNA-binding proteins. Among these methods the most common technique is probably the electrophoretic mobility shift assay (EMSA, or gel mobility shift assay) (Revzin 1987, Fried 1989). It can be used with crude protein mixtures or purified proteins in studies of, for example, identification and characterization of binding proteins, the DNA sequence requirements of binding, kinetics of binding, and cofactor requirements. Using radioactively labeled DNA probes, detection of proteins present in subfemtomolar amounts is readily possible. The principal of the assay is very simple. DNA fragments and proteins are mixed in a suitable buffer and binding is allowed to occur. The mixture is then separated by nondenaturing gel electrophoresis; stable complexes of DNA and protein are generally significantly retarded in mobility in comparison with the free DNA and the separated complexes are normally viewed by autoradiography or phosphorimaging.

Despite the basic simplicity of the method, there are many factors in addition to the need to optimize basic buffer requirements (such as ionic strength, pH, gel matrix concentration) which are important to its successful application (reviewed by

Varshavsky, 1987). In all reactions, especially those involving crude cellular or nuclear extracts, there are a number of competing reactions involving both the specific DNA probe and its protein target. When using a specific, labeled DNA as bait to identify its target protein in crude cell extracts, the DNA probe may be bound non-specifically by other proteins. This may result in nonspecific bands, smearing of the probe or trapping in the sample well. To identify the targets of a DNA binding protein, each putative target has to be tested individually. If a specific DNA binding protein is used as bait to interact with a pool of labeled DNAs, the protein may bind with variable affinity to nonspecific and specific DNA in the reaction. This may result in nonspecific bands or smearing bands and lead to difficulty in identifying the DNA binding property of the protein and finding its target DNA. For each protein-DNA binding reaction, excess, unlabeled competitor DNA must be used as a competitor. The most common test for specificity of complex is the ability of excess, unlabeled competitor DNA to reduce the amount of complex formed by specific DNA-protein interaction. This is based on the fact that specific DNA substrate binds the protein more tightly than any other competitor.

C. Footprinting

Once a DNA-binding protein has been identified and isolated, its specific binding site sequence needs to be determined. Footprinting is used to identify the specific protein binding site in a given DNA fragment. Therefore it can be used both to identify the DNA binding ability of a protein to a DNA fragment and to determine binding site sequence. A few footprinting techniques have been used such as DNase I footprinting (Schmitz and Galas, 1978), Exonuclease III footprinting (Wu, 1985) etc. DNase I footprinting is the

most common technique of all the footprinting techniques. In this technique, a suitable uniquely end-labeled DNA fragment is allowed to interact with a given DNA-binding protein and then the complex is partially digested with DNase I. The bound protein protects the region of DNA from DNase digestion. Since the DNase I molecule is relatively large compared to other footprinting agents such as hydroxyl radicals (Tullius et al., 1987), its attack to DNA is more readily protected by steric hindrance. Unfortunately, DNase I does not cut the DNA indiscriminately. Some sequences are very rapidly attacked whereas others remain unscathed even after extensive digestion (Drew, 1984). This results in a rather uneven “ladder” of digestion products after electrophoresis, limiting the resolution. However, when the protein-protected and naked DNA ladders are run alongside each other, the footprints are normally quite apparent.

D. Identifying the DNA-binding region of a DNA binding protein.

The isolation of cDNA of the DNA binding protein provides a means of obtaining large amount of the corresponding protein for functional studies. One of the methods, described above, is expression in expression vectors and purification by histidine tagging. Because the cDNA clone can be readily cut into fragments and each fragment can be expressed as a protein in isolation, particular features exhibited by the intact protein can readily be mapped to a particular region by testing the ability of different regions to bind to the appropriate DNA sequence. For example, experiments have mapped the octamer-binding proteins Oct-1 (Sturm et al, 1987) and Oct-2 (Clerc et al, 1988) to a specific short region of the protein. The transcriptional activity of a *Drosophila* nuclear protein called GAGA factor has been mapped to its C-terminal glutamine-rich domain (Vaquero

et al, 2000). Once this has been done, particular bases in the DNA encoding the DNA-binding domain of the factor can then be mutated to alter its amino sequence, and the effect of these mutations on DNA binding can be assessed by expression of the mutant protein and measuring its ability to bind DNA.

E. Identifying the DNA sequence bound to a putative transcription factor

It is common for a transcription factor to be identified on the basis of its binding to a known regulatory DNA sequence (Garner and Revzin 1981; Varshavsky, 1987; Taylor et al., 1994). It is also possible for a novel regulatory gene to be cloned on the basis of changes in expression in response to a particular stimulus or gene mutation. For example, in *N. crassa*, WC-1 and WC-2 genes which regulate expression of light-responsive genes were cloned and characterized by UV mutagenesis or integrational mutagenesis (Paietta and Sargent, 1983, Linden and Macino, 1997). Inspection of the DNA sequence and its predicted protein sequence can often help to identify transcription factors because of similarity to known transcription factors or domains known to mediate DNA-binding or transcriptional activation. Nit-2 gene, a major nitrogen regulatory gene in *N. crassa* was found containing a zinc finger motif and defined as a transcription factor later (Facklam and Marzluf, 1978, Fu and Marzluf, 1990a, 1990b).

Unlike transcription factors identified on the basis of their ability to bind to a known DNA sequence, a gene for a putative regulatory protein cloned by mutation, for example, gives no information on the DNA sequence to which its protein binds. It is essential for the further study of this novel factor that such sequences are identified to

permit an analysis of the interactions of the factor on artificial promoters carrying its binding site and identification of its target genes.

Pollock and Treisman (1990) used a method in which oligonucleotides containing a randomized, central 26 base pair sequence flanked by two, defined 25 base pair sequences were prepared. The sequences were mixed with transcription factor protein, and an antibody to that protein was used to immunoprecipitate the DNA-protein complex. The immunoprecipitated sequences were amplified by PCR using primers corresponding to the defined flanking sequence. Repeated cycles of binding, precipitating, and PCR were carried out to purify the binding sequences. Ultimately, the oligonucleotides which bind to the protein were cloned and subjected to sequence analysis to identify the common sequence they contain and which sequence is the binding site for the factor. One example of using this method was identification of the DNA binding site for the Brn-3 POU family transcription factors (Gruber et al, 1997). Identified binding sites can be linked to promoter of reporter gene and introduced into cells with the transcription factor to determine if the factor functions as a repressor or activator of gene expression.

Kinzler and Vogelstein (1989) devised a more direct way to identify target genes for an uncharacterized factor. This method is referred to as the whole genome PCR method. It is essentially the same as that of Pollock and Treisman (1990), except it uses total genomic DNA instead of random oligonucleotides. The genomic DNA is digested with a restriction enzyme and small defined DNA sequences are added to the ends of the fragments. The DNA-protein complex is immunoprecipitated, resulting in purification of pieces of genomic DNA containing the binding site. DNAs are PCR amplified as before by using primers corresponding to the defined sequence. This method is more technically

difficult than the use of oligonucleotides because of the complexity of genomic DNA. The advantage is that DNA binding sites obtained are linked to sequences to which they are normally joined in the genome rather than in isolation. Hence the sequences can immediately be characterized and used to identify a target gene of the factor. Examples using this method are the identification of the target genes of the nuclear receptor transcription factor family in mouse cells (Caubin et al., 1994).

Another method similar to the whole genome PCR method has been successfully used in human cells (Inoue et al. 1991, 1993). In this method, a large amount of genomic DNA digested with two enzymes (usually one cut to generate a 5' extension, the other to generate a 3' extension) was allowed to interact with a DNA binding protein. The protein-DNA complex was then trapped by filtration and the bound DNA eluted. The eluted DNA was cloned into a vector digested with the same enzymes. The cloned plasmid was subjected to 5 rounds of binding selection. DNA from the final selection was sequenced and the binding specificity further confirmed by gel mobility shift assay. This method has the advantage of avoiding cloning the same binding site target in many different lengths as that in the previous PCR methods and therefore reduces the laborious of screening later. More targets elements were identified using this method than the genome PCR method (Inoue et al., 1991).

Objectives of this study

The previous study demonstrated that the striking morphological abnormality in *sbr* strain resulted from an insertional mutation by a hygromycin resistance construct (Campsall 1998). Analysis of a partial cDNA and amino acid sequence suggested that *sbr*

codes for a transcriptional activator. The hypothesis being tested here is that the *sbr* gene encodes a DNA binding protein which regulates hyphal morphogenesis. The goals of this research project have been: confirmation of insertional disruption of the *sbr* gene, cloning and sequencing *sbr* genomic DNA, mapping the gene to a chromosome, cloning of the complete *sbr* cDNA, expression and isolation of SBR protein, testing the DNA binding ability of the SBR protein, isolating and cloning SBR target DNA fragment(s), and constructing *sbr* expression open reading frames with different deletions for further study in determining functional domains.

Materials and Methods

Cloning *sbr* genomic DNA into plasmid

A. Extracting DNA from phage clone isolated from genomic library.

A volume of 20 μ l of phage clone suspension λ J1T6, previously isolated by Katrina Campsall (unpublished) from λ BARGEM7 *N. crassa* genomic library (Pall and Brunelli, 1994) was used to infect 200 μ l Q358 host cell ($OD_{600}=0.5$) at 37°C for 20 min. The mixture was added to 4 ml top LB agarose (0.7%) and poured onto a 90 mm LB plate containing 1.5% agarose. After incubating the plate overnight at 37°C, 3 ml of λ diluent (10mM Tris-Cl; pH7.5, 10 mM $MgSO_4$) were added to the plate surface. The plate was incubated at RT for 1.5 hrs with gentle shaking. The λ diluent was harvested and transferred to two 15 ml polypropylene tubes and spun at 5700 rpm in an SS-34 rotor at 4°C in order to pellet bacterial cell debris. The supernatant was transferred to a microfuge tube and 1 μ l of RNase A (10 mg/ml) and 1 μ l of DNase I (5000U/ml) were added before incubation at 37°C for 30 min. 600 μ l of a solution containing 20% PEG 8000 and 2 M NaCl in λ diluent were added and the mixture was incubated in ice water for 60 min. The mixture was centrifuged at 12000 rpm for 10 min at 4°C and the resulting bacteriophage pellet was resuspended in 500 μ l TE (pH8.0) by vortexing. Any debris incapable of resuspension was discarded. The mixture was incubated at 68°C for 6 min after the addition of 5 μ l of 10% (w/v) SDS. 10 μ l of 5 M NaCl was then added and the bacteriophage DNA was purified by PCI extraction. Isopropanol (500 μ l) was used to precipitate the DNA at 4°C for 30-50 min. DNA was recovered by centrifugation at

14,000 rpm and the pellet was washed with 70% ethanol. The final DNA pellet was resuspended in 50 µl TE.

B. Subcloning *sbr* genomic DNA.

The extracted phage DNA was digested with *Sa*I and a 6.6 kb fragment, which hybridized to the *sbr* probe, was isolated and purified by silica bead absorption. The pBluescript SK⁺ plasmid was digested with *Sa*I and the *sbr* DNA fragment was inserted into pBluescript vector with T4 DNA ligase at 16°C overnight. The ligation reaction included 150 ng *sbr* DNA, 50 ng pBluescript, 1/5 final volume of 5X T4 DNA ligation buffer (NEB), 0.5 µl T4 DNA ligase (5 U/µl) and water up to 15 µl.

Transformation of *E. coli* cells

A. Chemical transformation.

Chemical transformation protocols from Sambrook et al (1989) were followed and competent cells were prepared as follows. A single colony of DH5α *E. coli* from a fresh plate culture was inoculated into 2-3 ml of LB broth (per litre: 10 g NaCl, 10 g bactotryptone, 5 g yeast extract, pH 7.5) and grown overnight at 37°C with shaking. A volume of 1 ml of this culture was added to 100 ml LB broth and grown at 37°C (for 90-120 min) until the OD₆₀₀ reached 0.5. Cells were then transferred into a cold centrifuge bottle. The bottle was kept on ice for 10 min and then centrifuged at 4000x g in a GSA rotor for 5 min at 4°C. The supernatant was discarded and the pellet was suspended in 33 ml of cold buffer (pH 5.8) containing 100 mM RbCl, 50 mM MnCl₂·4H₂O, 30 mM potassium acetate, 10 mM CaCl₂·2H₂O and 15% (w/v) glycerol. Cells were kept on ice

for 15 min, then centrifuged again at 4000x g for 5 min. The pelleted cells were resuspended in 8 ml of cold buffer (pH 6.8) containing 10 mM MOPS, 10 mM RbCl, 75 mM $\text{CaCl}_2 \cdot 2\text{H}_2\text{O}$, and 15% glycerol and kept on ice for 15 min before being dispensed into cold microfuge tubes (200 μl aliquots) and transferred to -70°C freezer.

About 50 ng of transforming DNA, in less than 10 μl was added to a 200 μl aliquot of competent cells. The suspension was mixed gently and placed on ice for 30 min. The mixture was then placed in 42°C water in a heating block for exactly 90 sec and immediately transferred onto ice. After chilling on ice for 2 min, 800 μl SOC medium (2% bactotryptone, 0.5% yeast extract, 100 mM NaCl, 25 mM KCl, 10 mM MgCl_2 , 10 mM MgSO_4 , 20 mM glucose) was added and the transformation mixture was cultured at 37°C for 50 min. 50-200 μl of the culture were plated on a 90 mm LBA (LB medium with 100 $\mu\text{g/ml}$ ampicillin and 1.5% agar) plate containing 15 μl 0.1M IPTG and 50 μl 2% X-gal dissolved in DMSO. Plates were incubated at 37°C for 14-16 hrs and white colonies were selected and inoculated onto another LBA plate for further identification of the correct clone.

B. Electroporation.

Competent cells were prepared as follows: 4 ml from a 10 ml, overnight culture was diluted into 200 ml of LB and incubated at 37°C until the OD_{600} reached 0.6-0.7 (approximately 2 hrs). Cells were pelleted in a GSA bottle by centrifugation at 4000g for 5 min at 4°C , washed twice in 200 ml cold sterile water, followed in 100 ml of 10% cold sterile glycerol. After the final wash, cells were pelleted by centrifugation at 4000x g for

10 min. The cells were resuspended in 0.5 ml of 10% cold sterile glycerol and the suspension was distributed into 1.5 ml size tube at 40 µl/each and stored at -70°C.

For electroporation, transforming DNA (<5 ng) was mixed with 40 µl of host cell and transferred into a 0.2 cm cuvette on ice. Electroporation was done with the following parameters: 2500 V, 200 Ohms, and 25µF. After electroporation, 950µl of SOC was added immediately and the cells were incubated at 37°C for 50-60 min before plating on to appropriate LB plate.

Extraction of plasmid DNA from *E. coli*

For preliminary analysis, DNA mini preps were prepared from 2-3 ml of LB culture with cells recovered by centrifugation at 4000x g for 5 min. The cell pellet was suspended in 100 µl of TE buffer, followed by addition of 200 µl of solution consisting of 0.2M NaOH, 1% SDS. After 5 min at room temperature, the cells were completely lysed and then 150 µl of 3M potassium acetate solution (pH 4.8-5.0) was added for neutralization. After 5 min of neutralization on ice, the microfuge tubes were centrifuged at 12000x g for 5 min. The pellet was discarded and two volumes of absolute ethanol (900 µl) was added to the supernant. DNA was pelleted by centrifuging at 12000x g for 5 min, rinsed with 70% ethanol, dried in a desiccator and finally resuspended in 50 µl TE buffer.

For midi preparation of plasmid DNA, 30 ml of over night LB culture was centrifuged at 4000x g for 5min. The DNA pellet was resuspended in 3 ml TE buffer and 5ml of lysis solution (0.2M NaOH, 1% SDS) was added and mixed thoroughly (but gently). After 10 min at room temperature, 3 ml of cold 3M potassium acetate was added

and mixed well and the mixture was left for 10 more min on ice before centrifugation at 10000x g for 15 min. The supernatant was filtered through one layer of miracloth into a new tube and 0.6 volume of isopropanol was added and mixed. The tube was stored at room temperature for 10 min and centrifuged at 10000x g for 15 min. The pellet was briefly rinsed with 70% ethanol and dried in a fume hood for 10 min before being resuspended in 500 µl TE buffer. The DNA solution was transferred into a microfuge tube and re-precipitated with 500 µl of 1.6 M NaCl containing 13% (w/v) polyethylene glycol (PEG 8000) by centrifuging at 12000x g for 5 min. The supernatant was removed and the pellet was resuspended in 400 µl TE buffer containing 50 µg/ml RNase A. After at least 1 hr at room temperature, the solution was extracted with 400 µl PCI followed by 400 µl chloroform:isoamyl alcohol (24:1). The DNA was precipitated with 2 volume of absolute ethanol in the presence of 1/10 3 M sodium acetate (pH 5.2) and dried in a desiccator before being dissolved in 100-200 µl TE buffer.

DNA Sequencing

Both commercial machine sequencing and manual sequencing were used. Commercial sequencing was performed by Canadian Molecular Research Service Inc.

Manual sequencing reactions were performed using the T7 Sequenase v2.0 sequencing kit (Amersham) by following the manufacture's instructions. DNA used for sequencing was extracted with the midi prep method described before.

The larger bottom sequencing gel plate was washed with 0.1% SDS, rinsed with water, wiped with 0.5M NaOH , rinsed with water and finally wiped with 95% ethanol. The smaller top plate was washed with 0.1% SDS, rinsed with water, wiped dry, and then

silanized with dichlorodimethyl silane. A 6% standard gel (in the 80 ml gel, solution contained 33.6g urea, 40% acrylamide 11.5 ml, 10X TBE 8 ml, H₂O 32.3 ml) was used and 600 µl of freshly made 10% ammonium persulphate and 25 µl of TEMED were added before pouring the gel. After the gel was polymerized, the plates was assembled in the BRL Model S2 sequencing apparatus and 1X TBE buffer was added into the top and bottom chamber. Air bubbles between plates in top and bottom chambers were flushed out before inserting the “shark’s tooth” combs.

A 30 minute pre-run was performed with 60W constant power, 1500-1900V and 30-45mA controlled by the Model 3000Xi Programmable Power Supply (BioRad). The preheated sample (5µl) was then loaded into the flushed wells between “shark teeth”. A second loading was added 4 hrs after the first loading. The total running time was 7-8 hrs.

After electrophoresis, the silanized plate was removed and the gel was fixed in a solution containing 10% methanol, 10% acetic acid for 20 min. The fixed gel was removed by adhering to a sheet of 3MM filter paper and then pulling away from the plate. The gel-paper were covered with plastic wrap and dried in a BioRad Model 583 vacuum gel dryer at 80°C for 1 hour. The plastic wrap was removed and the gel was exposed to Kodak XAR film at room temperature for 3 days before developing.

Complementation study: Transforming *sbr* mutant

A. Preparation of *sbr* spheroplasts

Mycelial fragments from crumbling *sbr* colonies on 8-10 days old plates were detached using an inoculating loop and wash-transferred into 50 ml of VMM [per litre:

15 ml 50x Vogel's stock medium (per litre: 127 g $\text{Na}_3\text{ citrate.2H}_2\text{O}$, 250 g KH_2PO_4 , 100 g NH_4NO_3 , 10 g $\text{MgSO}_4.7\text{H}_2\text{O}$, 5 g $\text{CaCl}_2.2\text{H}_2\text{O}$ predissolved in 25 ml dH_2O added dropwise), 1 ml trace elements (per litre: 50 g citric acid.1 H_2O , 50 g $\text{ZnSO}_4.7\text{H}_2\text{O}$, 10 g $\text{Fe}(\text{NH}_4)_2(\text{SO}_4)_2.6\text{H}_2\text{O}$, 2.5 g $\text{CuSO}_4.5\text{H}_2\text{O}$, 0.5 g $\text{MnSO}_4.1\text{H}_2\text{O}$, 0.5 g HBO_3 , 0.5 g $\text{Na}_2\text{MoO}_4.2\text{H}_2\text{O}$ dissolved in dH_2O), 1 ml biotin stock (5 mg biotin in 100 ml of 50% v/v ethanol)] to culture for 2-3 days. The granular *sbr* colonies in liquid medium were pelleted by centrifugation and washed with 50 ml of water followed by a wash in 15 ml of 1M sorbitol. The pellet was resuspended in 5 ml of 1M sorbitol and then 2ml of 5mg/ml Novozyme stock solution (in 1M sorbitol, sterilized by filtration) was added and incubated at 30 °C for 2 hrs with shaking. The efficiency of spheroplasting was checked by observing the swelling and lysis of the spheroplasts in water under microscope. Spheroplasts were pelleted by centrifugation at 500g for 5 min and washed 2 times in 15ml 1M sorbitol, and then once in 15ml of STC (50mM Tris pH 8, 50mM CaCl_2 , and 1M sorbitol). Spheroplasts were resuspended in 8ml of STC and followed by addition of 2 ml of PTC (40% PEG, 50mM Tris pH 8, and 50mM CaCl_2). Aliquots of 1 ml were stored at -70°C.

B. Electroporation of *sbr*:

Spheroplasts from a frozen stock were first washed with 1M sorbitol. To 50 μl of washed spheroplasts, a volume of 10 μl of transforming DNA (0.2-0.4 $\mu\text{g}/\mu\text{l}$) was added. The mixture was kept on ice for 10 min before electroporation at 1500 volts, 200 Ohms, and 25 μF in a 0.2 cm cuvette. After electroporation, 950 μl of 1M sorbitol was added

and 250 µl of the suspension was plated directly on VMM plates containing hygromycin B.

Gene mapping

A. Growing map strains.

The small cross with twenty mapping strains was used to map the *sbr* gene according to the method of Metzenberg et al., (1984). VMM liquid cultures in a volume of 50 ml were grown at room temperature for 2-3 days with shaking at 200 rpm. Mycelia were collected by vacuum filtration and frozen at -70°C .

B. Extraction of genomic DNA

Frozen mycelia were dried in a -70°C vacuum dryer for 8 hrs. The dried mycelium (about 250 mg) was ground using a pestle, with acid-washed sea sand, in a mortar. The powder was resuspended in rapid DNA extraction buffer containing 0.2 M Tris (pH8.5), 0.25 M NaCl, 25 mM EDTA, and 0.5% SDS. The suspension was transferred into a 12 ml tube and 0.7 volume of equilibrated phenol was added and mixed, followed by adding 0.3 volume of chloroform and mixing. The solution was centrifuged at 13,000 rpm for 1 hour. The aqueous phase was transferred into a new tube and 125 µl of RNase A (20 mg/ml) was added. After 30 min of incubation at room temperature, the solution was extracted with an equal volume of chloroform:isoamyl alcohol (24:1) and centrifuged at 13,000 rpm for 10 min. The aqueous phase was transferred into a new tube and a 0.54 volume of isopropanol added and mixed, resulting in a visible DNA precipitate. The supernatant was removed with a pipet after leaving the tube at room temperature for 10 to 20 min. Remaining supernatant was removed after

centrifuging the sample for 5 sec. The DNA pellet was rinsed with 70% ethanol and resuspended in 300-500µl TE buffer after vacuum drying.

C. Hybridization

DNA from the 20 mapping strains was digested with *KpnI* enzyme and subjected to hybridization (described below) with the *sbr* gene probe. The hybridization patterns were compared with published gene mapping results (Metzenberg and Grotelueschen, 1992).

Southern blot

The protocol for Southern transfer was adapted from Sambrook et al (1989). After agarose gel electrophoresis of the DNA samples, the gel was soaked in 1 M NaCl, 0.4 M NaOH at RT for 45 min with gentle rocking to denature and separate DNA into single strands. This was followed by neutralization in 1 M Tris-Cl (pH 7.4), 1.5 M NaCl for another 45 min.

A support for the transfer was constructed by placing a 1 cm high Plexiglas block wrapped in 3MM filter paper into a 3 L glass baking dish. Enough 20x SSC (3 M NaCl, 0.3 M Na₃ citrate, pH7.0) was added to the dish to just reach the top of the support and a piece of nylon membrane (Amersham), slightly larger than the gel, was set on top. Two pieces of 3MM filter paper, the same size as the gel, were stacked on top of the membrane followed by a 10-15 cm high stack of paper towels that were slightly smaller than the gel. The capillary transfer of the 20x SSC buffer through the gel and up into the paper towels was allowed to proceed for a minimum of 14 hrs. After transfer, the membrane was briefly soaked in 6x SSC to remove residual agarose and then it was dried

in an 80°C oven for 80 min. The transfer efficiency was checked using UV light before being wrapped in foil for storage at room temperature.

Radio labeling DNA probe

A “Ready to go” DNA labeling bead (minus dCTP) kit (Pharmacia) was used following the manufacture’s instruction. Briefly, about 50 ng linear DNA in less than 45 µl TE buffer (pH 8) or H₂O was denatured at 95-100°C for 5 min and quickly chilled on ice for 5 min. The denatured DNA was added to the tube containing the “ready to go” reaction bead followed by adding 5 µl of α-³²P labeled dCTP. The final volume was brought up to 50 µl with water. The reaction was gently mixed and incubated at 37°C for 30 min. A 50 µl stop solution containing 1 µl of 0.5 M EDTA, 47 µl of STE (10 mM Tris pH 8, 100 mM NaCl, 1 mM EDTA) , 2 µl yeast tRNA was added to terminated the labeling reaction.

Alternatively, 0.1-0.2 µg of template DNA and 80 ng of dp(N)₆ oligonucleotides were mixed and brought to 10 µl with water. After denaturing at 95°C 5 min, the mixture was quickly chilled on ice, followed by adding 5 µl of 5x labeling buffer (0.5M Tris pH6.9, 100 mM MgSO₄, 25% PEG 6000, 1 mM DTT), 2.5 µl of 5mM dNTP (minus dCTP), 2.5 µl of α-³²P labeled dCTP, 4 µl water, and 1 µl (3.82 unit/µl) Klenow Enzyme (Feinberg and Vogelstein, 1984). The reaction was mixed and incubated at room temperature or 37°C for 1-2 hrs and then stopped by adding 50 µl stop solution.

Spun column chromatography (Sambrook et al 1989) was used to remove unincorporated nucleotides. The bottom end of a 1 ml syringe was plugged with glass wool and it was filled with Sephadex G-50 beads previously equilibrated in 1x STE (100

mM NaCl, 20 mM Tris-Cl pH 7.5, 10 mM EDTA). The Sephadex G-50 matrix was packed down to a volume of 0.9 ml by repeated spins at 3000 rpm in a swinging-bucket centrifuge. The radiolabeling reaction was loaded onto the top of the spun column. The column was centrifuged at 3000 rpm for 4 min forcing the radiolabeled probe to be eluted from the column into a centrifuge tube. The probe was denatured at 95°C for 5 min before use.

Hybridization and film development.

The nylon membrane with transferred DNA was inserted into a glass screw-cap hybridization tube. Hybridization solution containing 6x SSC, 5x Denhardts [50X stock: 1% (w/v) Ficoll, 1% (w/v) polyvinylpyrrolidone, 1% BSA], and 0.5% (w/v) SDS was added at 0.05 ml per cm² of membrane. The tube was loaded into the rotating Hybridization Incubator Model 310 (Robbins Scientific) and incubated at 65-68°C for one hour.

The DNA probe (500,000 cpm/ml in hybridization solution) was combined with salmon sperm DNA (20 ug/ml hybridization solution) and it was denatured at 95-100°C for 5 min. The mixture was quick-chilled and then added to the hybridization tube.

Hybridization was performed at 65-68°C for a minimum of 12 hrs. Stringency washes were performed at the conclusion of the incubation. The hybridization solution was discarded, 50 ml of 2xSSC were added to the hybridization tube and incubated at 68°C for 15 min. The procedure was repeated using 2x SSC, 0.1% SDS for 60 min followed by a wash of 50 ml of 0.1x SSC for 15 min.

The membrane was allowed to air dry slightly prior to exposure to Kodak autoradiographic film. The membrane and film were stored at -20°C in a light-tight cassette until the desired signal was reached.

After sufficient exposure of the film to a labeled membrane, the film was developed for 1-5 min in GBX developer and replenisher (Kodak). The film was rinsed in water for 1-2 min, soaked for 5 min in GBX fixer and replenisher (Kodak) and then rinsed for 5 min in water before it was left to dry.

RT-PCR

A. Extraction of total RNA.

Glassware and solutions were prepared following the protocol of Sambrook (1989). Other items were soaked in 1% DEPC for 30 min. Frozen mycelium was pulverized by grinding in liquid nitrogen. The frozen mycelial powder was added to a 40ml centrifuge tube containing 10ml of 65°C extraction buffer (0.2 M sodium acetate pH 5, 1% SDS, 10 mM EDTA) and 10ml of unbuffered phenol and mixed quickly. The mixture was kept in 65°C water bath for 20 min before centrifuging at 5000 rpm for 15 min. The aqueous phase was mixed with 10 ml of chloroform:isoamyl alcohol (24:1) and left in room temperature for 10 min before another centrifugation at 5000 rpm for 15 min. The aqueous phase was recovered and the volume measured in a sterile Falcon tube. For each 2.5 ml of the solution, 1 g of CsCl was added and dissolved by mixing. This solution was overlaid to a 1.2 ml cushion (5.7M CsCl, 100 mM EDTA, 0.2% DEPC, pH 7.0) in a 11.3 ml polyallomar tube (Backman). The tube was filled up with a solution of 1 g CsCl/2.5 ml DEPC treated water to 0.5 cm below the edge. Ultra-centrifugation was

in the SW41 swinging bucket Beckman rotor and centrifuging at 33,000 rpm for 22 hrs at room temperature.

After ultra-centrifugation, the supernatant was poured off and the pellet was rinsed with DEPC treated water. The RNA pellet was dissolved in 100 μ l of DEPC treated water for more than 12 hrs before transferring to microcentrifuge tubes and storing at -70°C.

B. Reverse transcription.

Before reverse transcription (RT), the total RNA extract was further purified using RNeasy plant mini kit (QIAGEN) following the manufacture's instruction. A total of 20 μ l of cleaned RNA (15-20 μ g) was heated to 85°C for 10 min, chilled on ice for 5 min, spun and added to a chilled mixture containing 3 μ l of 10 $\times$ AMV-RT buffer (Sigma), 2 μ l of 10 mM dNTP, 3 μ l (75pmol) of specific primer, and 1 μ l of RNase inhibitor. Then 1 μ l of AMV-RT (Sigma) was added, mixed, and centrifuged briefly. The RT reaction was carried out with first temperature step at 42°C or 50°C for 40 min, followed by 54°C, 58°C, and 62°C each for 30 min. First strand cDNA was tailed following the protocol of Frohman (1990). Briefly, after the RT reaction, primers were removed using a Centricon 100 spin filter (Amicon Corp.). The reaction mixture was diluted in 2 ml of 0.1X TE and centrifuged at 1000x g for 20 min. After washing twice with 0.1XTE, the retained liquid was collected and concentrated in 10 μ l TE using vacuum centrifuge dryer. The tailing reaction was carried out at 37°C for 5 min after adding 4 μ l of 5X tailing buffer (GibCO

BRL), 4 μ l of 1 mM dATP, and 10 units of recombinant terminal transferase (GIBCO BRL). From this tube, 1-5 μ l was used for PCR amplification.

PCR protocol

Standard PCR reactions were performed using a Perkin Elmer Cetus 480 DNA Thermal Cycler and GIBCO BRL reagents. Typical PCR reactions contained: 2.5 U of Taq DNA Polymerase (added after 95°C denaturation), 5 μ l of 10x PCR buffer, 1.5 μ l $MgCl_2$ (50 mM), 1 μ l dNTP (10mM), primers (25 pmoles of each), variable amount of template DNA (for RT-PCR, 1.5 μ l RT reaction used), and sterile dH_2O , up to 50 μ l. After thorough mixing by pipette, a 50 μ l overlay of sterile mineral oil was added.

The standard reaction conditions were: denaturation at 95°C for 5 min, followed by 35-40 cycles of 50 sec denaturation at 95°C, 50 sec annealing at optimal temperature for primer, extension at 72°C for 1-3 min (depending on the predicted size, estimated 1000 bp/minute). A final 10 min of extension at 72°C was included. The optimal primer annealing temperature was calculated by adding 4°C for each G/C and 2°C for each A/T present in the primer; this value was then reduced by 0-3°C depending on PCR results.

Purification of SBR protein under native condition

A. Transformation of bacteria.

The *sbr* ORFs, cloned into the QIAGEN pQE expression vector were propagated in XL1-blue cells in the presence of 100 μ g/ml ampicillin and the plasmid was extracted by midi prep method. The plasmid DNA was sequenced to confirm the correct open

reading frame before transforming the M15 cells (QIAGEN) which had been prepared according to the manufacture's instructions. Transformed M15 cells were plated on LB plate containing 100 µg/ml ampicillin and 25 µg/ml kanamycin.

B. Expression of SBR protein.

The transformed M15 cells were cultured in LB medium containing 100 µg/ml ampicillin and 25 µg/ml kanamycin. 2.5 ml from 10ml of 37°C overnight culture was added in to 50 ml medium and continued growing at 37°C with vigorous shaking until the OD₆₀₀ was 0.5-0.7 (approximately 30-60 min). IPTG was added to a final concentration of 1 mM and the culture was grown for 5 hrs at 37°C. Cells were harvested by centrifugation at 4000x g for 20 min and the cell pellet was frozen at -20°C for one or a few days.

C. Protein purification.

All purification steps were done at 4°C or on ice. The cell pellet was thawed and resuspended in lysis buffer (50 mM NaH₂PO₄, pH8.0; 300 mM NaCl; 10 mM imidazole) at 5ml per gram wet weight. The suspension was incubated for 60 min with lysozyme at 1mg/ml, then sonicated 6X for 10 sec alternating with 10 sec cooling intervals by using a sonicator equipped with a microtip. The lysate was incubated with RNase A (10 µg/ml) and DNase I (5 µg/ml) for 15 min before centrifugation at 10,000x g for 30 min to pellet the cellular debris. The clear supernatant (soluble crude extract) was collected into a new tube and the pellet was re-suspended in 5 ml lysis buffer (insoluble crude extract). To assess SBR protein solubility, 5 µl of each crude extract was used for SDS-PAGE analysis. Purification of SBR was carried out as follows: To 4ml of soluble crude extract, 1 ml 50% Ni-NTA agarose slurry (QIAGEN) was added and the mixture was incubated

for 60 min with gentle shaking on a shaker and occasional hand mixing. The mixture was loaded into a column and the column was then washed with 4 ml wash buffer (50 mM NaH_2PO_4 , pH 8.0; 300 mM NaCl; 20 mM imidazole), followed by 4 ml wash buffer containing 50% glycerol. The protein was eluted with 4X 0.5 ml elution buffer (50 mM NaH_2PO_4 , pH 8.0; 300 mM NaCl; 250 mM imidazole, 50% glycerol) into 4 tubes and then 2.5 μl of each elution was analyzed by SDS-PAGE. For the DNA binding assay, eluted protein was dialysed, in 200 ml of dialysis buffer containing same components except imidazole, at 4°C for two days to reduce imidazole. Protein concentration was determined by measuring the OD_{280} and using BSA of known concentration as standard.

SDS-Polyacrylamide Gel Electrophoresis (PAGE)

Two 0.75 mm thick polyacrylamide gels were made using the Mini-PROTEAN^R II Dual Slab Cell (BioRad). The instructions accompanying the apparatus for preparation of the 10% resolving gel and 4% stacking gel were followed. A comb (10 wells, 0.75 mm thick) was used to make the wells. 5 μl of broad-range protein molecular marker (BioLab, New England) was loaded into a well of each gel as a molecular weight standard. The gel was run at 180V for 50-60 min powered by the Bio-Rad Model 200/2.0 power supply.

Protein gels were stained in Coomassie (0.1% Coomassie Brilliant Blue R-250) in fixative (10% acetic acid, 40% methanol). Background staining was removed with destain (10% acetic acid, 40% methanol). Gels were dried at 80°C for 1 hr in a Model 583 Vacuum Gel Dryer (BioRad) while sandwiched between 2 sheets of cellophane that were equilibrated in destain with 3% (v/v) glycerol.

Western blot and immunostaining

Western transfer: The gel and nitrocellulose membrane (Protran) were equilibrated in transfer buffer (25 mM Tris, 192 mM glycine, 20% (v/v) methanol) for 30 min. Using the Transblot^R SD Semi-Dry Electrophoretic Transfer Cell (BioRad), 3 pieces of 3 MM filter paper (the same size as the gel) saturated in transfer buffer were stacked on the anode. The membrane (slightly larger than the gel) was layered next followed by the gel. An additional 3 pieces of filter paper were stacked on top and bubbles were removed at each layer. The cathode was placed on top and transfer was performed at 15V for 40 min. The membrane was stored at 4°C in plastic wrap.

Immunostaining: The nitrocellulose blots were blocked with 3% BSA in TBS (10 mM Tris.Cl, pH7.5; 150 mM NaCl) for 60 min with shaking at RT. Membranes were washed 2X for 10 min in TBS-Tween/Triton buffer (20 mM Tris.Cl, pH7.5; 500 mM NaCl; 0.05% Tween 20; 0.2% Triton X-100) and once in TBS for 10 minute, then incubated with a 1:2,000 dilution of mouse RGS-His antibody (QIAGEN) in TBS with 3% BSA at room temperature for 60 min, washed twice in TBS-Tween/Triton buffer and once in TBS for 10 min/each. The membrane was incubated with 1:20,000 goat anti mouse antibody conjugated with peroxidase in TBS/10% skim milk powder, followed by washing 4X for 10 min in TBS-Tween/Triton buffer.

The Chemiluminescence Blotting System (Roche) was used for detection of membrane-bound proteins labeled by antibodies. Both the kits reagents and instructions were used. Membranes were incubated in a mixture of substrate solution A and starting

solution (100:1) for 60 sec. The blot was wrapped in plastic wrap. Autoradiographic film was exposed to the membrane for 4 sec to 2 minute and developed immediately.

Isolating SBR binding target DNA elements

N. crassa genomic DNA was digested with *Hind*III and *Pst*I and then purified by absorption on silica beads. In a 2 ml centrifuge tube, 20 µl of the 50% Ni-NTA agarose slurry (QIAGEN) was added to 120 µl of lysis buffer (described in the protein purification protocol) containing 5 µl SBR protein and incubated on ice with shaking for 60 min. 100 µl cold 1X binding buffer (50 mM Tris.Cl, pH7.5, 100 mM NaCl, 10 mM β-mercaptoethanol, 0.1 mM EDTA, 0.5 mg/ml BSA, 10% glycerol, 10 mM imidazole) was added to exchange the buffer. The protein-Ni-NTA mixture was pelleted by centrifugation at 14000x g for 3 sec. After carefully removing the supernatant, the digested genomic DNA (20 µg) in 100 µl 1X binding buffer was added to the protein-Ni-NTA and mixed by gentle shaking. Protein-DNA binding was performed on ice for 30 min with shaking/hand mixing, followed by 4 washes (each wash with 2 ml of cold binding buffer and centrifugation for 3 sec at 14000x g). The bound DNA was eluted by adding 100 µl high salt elution buffer (1.5 M NaCl; 50 mM NaH₂PO₄, pH 8.0). To remove the salt, silica beads were used with one wash performed as follows: 300 µl cold 6M NaI was added followed by 5 µl of glass milk. After 5 min of incubation on ice, glass beads-DNA were pelleted by centrifugation at 14000g for 3 sec and the supernatant was discarded. Additional centrifugation was performed to ensure complete removal of any remaining liquid. The pellet was resuspended first in 1 ml of New Wash buffer (50 mM NaCl, 10 mM Tris.Cl, pH7.5, 2.5 mM EDTA, 50% ethanol), then the tube was filled and

mixed with another 1 ml of New Wash buffer. Glass beads were pelleted by centrifugation for 3 sec at 14000g and the remaining liquid was completely removed by additional centrifugation. DNA was recovered in 10 µl sterile distilled water by warming at 50°C for 3 min and centrifugation at 14000x g for 2-3 min.

The recovered DNA was inserted into the pBluescript SK⁺ vector, previously digested with *Hind*III and *Pst*I, with T4 DNA ligase at 16°C overnight in a 15 µl reaction. The ligated DNA was cleaned as described above and recovered into 10 µl sterile distilled water. 5 µl of the DNA was used to transform *E. Coli* competent cell DHα5 by electroporation as described in the *E. coli* cell transformation protocol. All the cells were plated on LBA plate and incubated at 37°C for 18-20 hrs.

All the colonies were washed off with 5 ml LB and poured into a 30ml LB containing ampicillin, and incubated for midi prep extraction of the plasmid pool. The extracted plasmids (20 µg) were subjected to protein-DNA binding selection again and PCR testing as well with T3/T7 primers. The selection was repeated 5 times, and after the final selection, 20 colonies were randomly selected for DNA extraction and PCR with T3 and T7 primers to screen for any insertion. PCR amplified DNA was further tested by *Hind*III/*Pst*I digestion, and selected DNAs were sequenced.

To ensure completeness of washing during DNA binding, an additional wash with 100 µl of binding buffer after the 4th wash was used to transform competent cells. Nonspecific binding to vector DNA was assayed by replacing the genomic DNA with vector DNA.

Electrophoresis mobility shift assay (EMSA)

A. Preparation of probe DNA:

A protocol described by Molloy (2000) was followed with minor changes. Briefly, 2 µg of probe DNA prepared by PCR was digested with *HindIII* and *AvaI* (*AvaI* is next to *PstI* on the vector. This site is good for dCTP labeling by end filling) in 40 µl reaction. After digestion, the following components were added in order: 1 µl of 5 mM mixture of dATP, dTTP and dGTP (final 125 µM); 0.5µl digestion buffer; 5µl α -³²P-dCTP, 2 unit of large fragment of DNA polymerase (GIBCO BRL). After mixing, the reaction was incubated at room temperature for 15 min. Unlabeled dCTP was added to 125 µM and incubated for a further 5 minute to ensure complete filling. Bromophenol blue was added to a final concentration of 0.02%. The reaction was loaded in a 10 mm wide well on a 1.2 mm thick, 5% acrylamide, and 0.25% bis-acrylamide gel prepared in 0.5X TBE buffer (Sambrook, 1989). The probe was electrophoresed at 10 V/cm until the bromophenol blue was near the bottom of the gel. One of the glass plates was removed and the gel was covered with plastic wrap and exposed to autoradiographic film for 5 min. The film was marked for alignment with the gel and then developed. The position of the labeled band was identified and the gel slice at the right position was excised with a scalpel. The probe DNA was eluted in 100 µl TE buffer by shaking on a rocking platform at room temperature. The radio labeled probe was stored at -20°C in the presence of 1mM β -mercaptoethanol to limit radiolytic breakdown.

B. Preparation of non-specific competitor

The cloning vector plasmid used in isolating the target DNA was used as a competitor. Plasmid was cut with *EcoRV* to generate blunt end DNA. DNA was purified by phenol:chloroform extraction and precipitated with ethanol.

C. Protein-DNA binding

In a 15 μ l reaction, the following was added in order: sterile water, 1.5 μ l 10X binding buffer (stock: 500 mM Tris.Cl, pH7.5, 1 M NaCl, 100 mM β -mercaptoethanol, 1 mM EDTA, 5 mg/ml BSA), competitor DNA (0 to 11 μ l), probe DNA (1 μ l), SBR protein (1-3 μ l). Since SBR was stored in 50% glycerol, when less protein was added, the glycerol concentration was adjusted to a final 10-15%. Reaction was mixed and incubated at room temperature for 20 min.

D. Electrophoresis

4-5% acrylamide gels (stock: 39% acrylamide, 1% bis-acrylamide) were prepared with the running buffer (6.7 mM Tris.Cl, pH8.0, 3.3 mM sodium acetate, 1 mM EDTA,). After polymerization, the gel was pre-run for 1-2 hrs at 10V/cm. The binding reaction was loaded on gel in the cold room and electrophoresed at the 10V/cm for 3-4 hrs. The gel was fixed in 10% methanol and 10% acetic acid for 20 min, transferred to a 3MM paper, covered with plastic wrap, and vacuum dried at 80°C for 1 hour before exposing to Kodak autoradiographic film for 2 days at -20°C.

DNA analysis with bioinformatic programs

The following bioinformatic program were used in DNA analysis:

- a) NCBI BLAST--- NCBI's sequence similarity search tool was designed to support analysis of nucleotide and protein databases: <http://www.ncbi.nlm.nih.gov/BLAST/>.
- b) MIPS *Neurospora crassa* database (MNCDB)-- it also provides BLAST search tool sequence similarity analysis: <http://www.mips.biochem.mpg.de/proj/neurospora/>.
- c) NCBI ORF program -- The ORF (Open Reading Frame Finder) program is a graphical analysis tool which finds all open reading frames of a selectable minimum size in a user's sequence or in a sequence already in the database. This tool identifies all open reading frames using the standard or alternative genetic codes. <http://www.ncbi.nlm.nih.gov/gorf/gorf.html>
- d) TRANSFAC--- The TRANSFAC database is maintained by GBF research group "Bioinformatics" at Braunschweig. It compiles data about gene regulatory DNA sequences and protein factors binding to and acting through them. Programs are developed that help to identify putative promoter or enhancer structures and to suggest their features. Transcription factors information and binding sites and related sources such consensus and related publications are available: <http://transfac.gbf.de/TRANSFAC/>.
- e) MatInspector V2.2--- a TRANSFAC program based on TRANSFAC 4.0, searching for binding sites with different core and matrix score, detailed information of binding factors are directly associated with TRANSFAC database: <http://transfac.gbf.de/cgi-bin/matSearch/matsearch.pl>.
- f) GENSCAN: a gene identification program developed at MIT. The program predicts the locations and exon-intron structures of genes in genomic sequences from a variety of organisms. This server can now accept sequences up to 1 million base

pairs (1 Mbp) in length: <http://genes.mit.edu/GENSCAN.html>. This program is one of the top three accurate programs (Singh, 2000).

- a) CGG (computational genomics group) DNA Analysis -- the program is to identify putative genes. The program is maintained at the Sanger Center (UK): <http://genomic.sanger.ac.uk/gf/gf.html>.
- g) MZEF: Michael Zhang's Exon Finder -- an internal exon prediction program developed by Michael Zhang's group at Cold Spring Harbor Laboratory: <http://sciclio.cshl.org/genefinder/>. This program is one of the top three accurate programs (Singh, 2000).
- h) Motif search: a collection of different search programs including PROSITE, BLOCKS, ProDom, PRINTS, Pfam was directed to address: <http://www.motif.genome.ad.jp/>.

Results

Genomic clone and sequence

A phage λ clone harboring the *sbr* gene was isolated from a genomic library by Campsall (1998). DNA of this clone was extracted and subjected to enzyme digestion for subcloning. DNA fragments selected with the *sbr* cDNA probe after specific enzyme digestion was ligated into plasmid pBluescript and then introduced into DH5 α *E. coli* competent cells. Southern blots of digestions with *Bam*HI, *Kpn*I and *Sal*I showed hybridization with *sbr* probe (data not shown) at size 3.0 kb, 4.8 kb, and 6.6 kb respectively. The 3.0 kb fragment from *Bam*HI digest was first cloned into pBluescript vector and sequenced by using designed primers or the universal forward and reverse primers (Figure 3). This fragment contained all the sequence of the truncated *sbr* cDNA with about 500 bp more at both 5' and 3' ends. A 64 bp intron with *Hind*III site in it was identified (Figure 3).

Since sequencing of this *Bam*HI clone revealed no start codon upstream of the truncated cDNA (Figure 4), the larger (4.8 kb) fragment, from the *Kpn*I digest was cloned (Figure 3). However, restriction enzyme analysis and sequencing results showed that the *Kpn*I site at the 5' end was very close to the *Bam*HI site (68 bp upstream) (Figures 3,4). An ATG start codon was found between the *Kpn*I and *Bam*HI sites (Figure 4) but it is not in a consensus sequence context for *N. crassa* C>NNNCAMVATGGC (M=A or C, V=A, or C, or G) (Bruchez et al., 1993a). Therefore, the 6.6 kb fragment from *Sal*I digest was cloned at the *Sal*I site (Figure 3) and it added 1.7 kb to the 5' end of the *Kpn*I clone. This 1.7 kb segment was divided into two fragments

Figure 3. Primers used in manual sequencing and different genomic clones harboring or flanking the *sbr* cDNA. (A) is a 6.6 kb *Sal*I clone (sbr6.6k), (B) a 5.3kb *Xba*I-*Sal*I clone (sbr5.3k), (C) a 4.8kb *Kpn*I clone (sbr4.8k), (D) a 3.0kb *Bam*HI clone (sbr3.0k), (E) a 1.3kb *Sal*I-*Xba*I clone (sbr1.3k), and (F) a 0.6kb *Xba*I and *Bam*HI clone (sbr0.6k). The relative location of the truncated cDNA clone is indicated underneath the *Bam*HI clone that was first cloned and sequenced. All clones are in pBluescript (SK⁺) vector. The orientation of insertion is indicated by *Kpn*I and *Sac*I at the end sides in each clones.

Figure 4. Sequence and translation of the 5' end of the *sbr* gene and its flanking region. The top is the translation overview of the *sbr* harboring region from the *Sa*II site to *Bam*HI site as indicated in figure 3. The sequence from *Xba*I to *Not*I covers the 5' end of the truncated cDNA. Digestion sites are underlined. Highlighted sequences are the two primers sbr-001 and sbr-002 used in RT-PCR. Closely downstream from primer sbr-002 are the *Kpn*I and *Bam*HI sites and ATG start codon (boxed).

by *Xba*I digestion and further subcloned as *Sal*I-*Xba*I and *Xba*I-*Bam*HI clones (Figure 3). The sequence of the *Xba*I-*Bam*HI subclone was confirmed by sequencing from both directions. The genomic sequence from the *Xba*I site to the truncated cDNA was sequenced twice, due to the overlapping of *Kpn*I and *Bam*HI clones (Figures 3, 4). A stop codon is just upstream of the *Kpn*I site, suggesting that the *sbr* ORF begins from downstream of this site or an intron exists in this region (Figure 4).

***sbr* cDNA sequence**

A. Amplification of *sbr* cDNA 5' end.

In order to identify the transcription start site and the completed *sbr* cDNA 5' end sequence, RACE primers of sequence GACTCGAGGACATCGATTTTTTTTTTTTTTTTTT and GACTCGAGGACATCG (Frohman 1990) were used. A primer named *sbr*-yy9 (Figure 5) spanning the intron was used for reverse transcription (RT) with RNA (total RNA) as template. The RT product was tailed with dATP using terminal transferase and subjected to PCR amplification with the RACE primers and the specific primer *sbr*-yy9. Nested PCR with an interior specific primer *sbr*-004 (Figure 5) was also performed. However, no specific result was obtained (data not shown).

Based on genomic sequence, two more primers named *sbr*-001 and *sbr*-002 (Figures 4, 5) near the possible cDNA 5' end were designed to replace the above RACE primers in RT-PCR. The RT reaction was performed with primer *sbr*-yy9 as described above.

Figure 5. RT-PCR results and map of primers. (A) indicates the relative locations of used primers on or near *sbr* gene. Primer *sbr-yy9* was used for RT reaction, primers *sbr-001* and *sbr-002* (also in figure 4) were used for PCR after RT. Primer *sbr-004* was used to replace *sbr-yy9* for nested PCR. (B): lane 1, RT-PCR with primers *sbr-001* and *sbr-yy9*. Lane 2, RT-PCR with primers *sbr-002* and *sbr-yy9*. Lane 3, the same primers as lane 2, but using mutant RNA in RT. Lane 4, nested PCR of lane 1. Lane 5, nested PCR of lane 2, 1.3kb was expected. Lane 6, nested of lane 3. Lane 8, control RT-PCR reaction from mutant RNA using *ro-4* primers (Robb et al., 1995). Lane 9 is a control RT-PCR reaction for *ro4* primers using genomic DNA as template. It generated larger fragment since there were two introns between the primers.

A

B

Figure 6. The *sbr* gene and flanking sequence at the 5' end. The asterisks indicate stops codons before and after translation. The *sbr* gene encodes a protein of 612 amino acids with two glutamine-rich regions (highlighted), a helix-turn-helix domain (underlined), and cysteine-rich region (double underlined). The 64bp intron at nucleotide 3276 to 3340 is shown in lower case letters.

1 gtcgacgcctctgctgtggaccgacgatggcagcatggcgccgcacgccgaacatggacgt
61 aactacttttgggtcggcggagatgtcgcgggcnattttnagggagtggctggcaccggaa
121 aacagcacggnacccaacggagagccgaagcacttgggtatttancnncnnaatgttggcg
181 ctttttgcnattnntnggctatggcccgtaaccccaaccaaatggcgctgagaggactg
241 gcggataccctggcgatggaggtcaattactaccagataaaccgggtcaaggtgcacatt
301 gtgtatcccggcacgattgtatcacccgggtacgagaggggaaaaatcagacgaagccggac
361 atcacggtggagctggaaaaggacgagccagcggaagcccggtacggtagcaaggagg
421 gcgattgccgggctggaggctggcaagtattttgtcgatgtgtcattcctgggaaggctg
481 atgcagtgcgggatcatggcggggtcgccgaggaacaactgggtgctggacacgttgatg
541 ggggtggctcataccgatcatctacttcttcgtgttgagggggatgaactcaacgattgtg
601 aagtgggcgagggaaaaaaggggcatcctttttacgcatccccaagaagaagtgtgcgat
661 ggtttggatccttgttctttttgtatggatgtgtatatgatctatgaaacataaaaaat
721 acctgtaataacttttgttgttgaatagaccgatatcagcctcaagacaacctggacaag
781 tcccccgctctccttttgtgagacactcaccttgaagtcaaagcactacgtcttcaccgcg
841 atccggatgcgttttgacaccaagaatgagcaatccatgcttagttgggtgacaacgact
901 ttaaccctgcttgtgttgtccagcaataagtgtatattccgacgtcaagatgtttggag
961 caccacttgggtgagtgcggagacagaaagatgactacctgctaaacccatgggtatgcccc
1021 cgttcgagactgtatgaatccttctggcagaacctgctcttactgtatactagacttacgg
1081 attgtatcgatggaatgtgtgagcttcggttcagagcttttgaccttcgccagtatacat
1141 gtggataggtatcagtcggtgaannaattagctcggaactccatccattccavcacgtcc
1201 cgtttctacgattatcgggccagcacctgtracagacatgcatcatatcaatactatagt
1261 catsattcccacscatctggaststbtaatacgtccatttctagactttgggggaaac
1321 attgatatttccctacaccgggcgatttactccacttttttgatggcacttccggttgacc
1381 atctacgaggatcgcgggcgggagggggcagaaggcacggataaacatttcaggattggacca
1441 ttttcttcgttgtcagtggtttgtggcttgtgcgccctgacatagcaggtggcgctgtgt
1501 gcacagttgcacaaagtcgagtgatggacactggctcgagttgccgccacctcttgggt
1561 gctggttgcctgccggtgactgtccgccagctaaaatgtcgcgctccgtcgagcgggtgcc
1621 cgccgatccagcggttccttggacaccgcacacttctatcctctcctccccaaacttgt
1681 cgccccgacctttggaacaccagctgatgacagctgctcaacaccaactcaccctaccaag
1741 cgccattaccagcccatacagcgacgaccagctcgcgacacctggtaaccgcctgtcatt

1801 gctgccaaactgggtaccagcaacccgggtccccagcggacgagtATGAAACCATCGCCAATC
M K P S P I

1861 CGCCCTGTCGCTGGCCCCGGATCCGGTGTCCAGCGCGGTGCTGTGGGTGAGGTTCCCGGA
7 R P V A G P G S G V Q R G A V G Q V P G

1921 GCTGTGGGCCAGGTCCAGGTGGGCTCCGTGGGACTTCAGGCCGGCTCTGCCCCGGCCAA
27 A V G Q V Q V G S V G L Q A G S A P G Q

1981 GGCCAAGTCCCAGCCTCGCGACTGACTCCGGGCCCCGGCCCAACCAAAGTGACCAAGCCG
47 G Q V P A S R L T P G P G P T K V T K P

2041 ACCACCCCGGCACGTTCTCGCACTGGCGCCAGTCCTCACTCGCTGGTTGGAACAGCCAA
67 T T P A R S R T G A S P H S L V R N S Q

2101 CTTGTCCAGCAGCTGGCTGCTCACGAAGCTCAGCTCCAGCAGTTCCAACAGCAACAGCAG
87 L V Q Q L A A H E A Q L Q Q F Q Q Q Q Q

2161 CTGCAACAACAGCAACAACAACAACAACACCAACAGCACCAACAACAACAACAAAATCTG
107 L Q Q Q Q Q Q Q Q H Q Q H Q Q Q Q Q N L

2221 GCTACCAACTTGAGCCCCGACCAGGACCCCTAACCACGTTAACCAGATCCATCCCCAACTT
127 A T N L S P D Q D P N H V N Q I H P Q L

2281 CAGGCCAGCTTCAGCTCCAGCATCAGCAGGCTCAGGAGCACGACCTCCACAGCCATCAC
147 Q A Q L Q L Q H Q Q A Q E H D L H S H H

2341 CTGCAACAGCATCAGCCAGATCAAACCCAGGATCATCTTCAGCATCAACAGGAGCAACAT
 167 L Q Q H Q P D Q T Q D H L Q H Q Q E Q H
 2401 CAGGAGCCTGATATTGACCAGGATTCCCTTAAACCAGACCCGTCCGCCTACGACGACCCT
 187 Q E P D I D Q D S L K P D P S A Y D D P
 2461 ACGGCCAGGATGGGCACTGAAGAGTATGCGGCCGAGCTGCTGCTGCCGTCATGGAGACG
 207 T A R M G T E E Y A A A A A A A V M E T
 2521 GGTCTTGGTCAACCCGGCTCCCTGAATGATGGCACTGGCGAGCCGGATGTTGACGCCGAG
 227 G L G Q P G S L N D G T G E P D V D A E
 2581 ATGGAGGACCAAGATGCCTCTGCCGCTGCGAACTCGGGACTCGGCGGCCATGTTGACCTT
 247 M E D Q D A S A A A N S G L G G H V D L
 2641 AGTGGTGCCGGCATGCTAGGTAACCCAGGTGCCCTGGTGGTGTGCCTCAACCCTCTCCC
 267 S G A G M L G N P G A P G G V P Q P S P
 2701 CAGGACCACTCTCAACATCAACAACACCAGCACCAGCAGCAGCAACATCAACACCATCAG
 287 Q D H S Q H Q Q H Q H Q Q Q Q H Q H H Q
 2761 CAACAACAGCAGCAGCAGCATCAACAGCAACAGCAGCATCAACAGCAACAGCAGCAGGGC
 307 Q Q Q Q Q Q H Q Q Q Q Q H Q Q Q Q Q Q G
 2821 CGATATGCTTCAGGCCTTGGACAGCCCCAGGCCCATCCTCATCAACATCAGCACCAACCC
 327 R Y A S G L G Q P Q A H P H Q H Q H Q P
 2881 CCTGTTGCCCCCATGGGCCAGCCCCAGCAGATGCAGCCGTCCCCTCATCAGCACCCAATG
 347 P V A P M G Q P Q Q M Q P S P H Q H P M
 2941 GACCCTCAGCTATCCAAGAGCACGGAGGATATGGCTCATGAATCCGGTTATGGCCAGCTA
 367 D P Q L S K S T E D M A H E S G Y G Q L
 3001 AACGTCGAGAGTGCTCTTGCCAAGAGATTGGCTCGTGAGCCGGGTCATCGTCTTGCCCAG
 387 N V E S A L A K R L A R E P G H R L A Q
 3061 CAGAGGCGCCCTGAGCAAGTGCTGAACTTGGCACGTAGAAGCAATGTTGAAGCTCTGTTCC
 407 Q R R P E Q V L N L A R R S N V E A L F
 3121 GCTCACATTGCTGGTGAGCCGGCTCGCGTCCCTTGTAAGAATTGCCACAAGGGCCACGGT
 427 A H I A G E P A R V P C K N C H K G H G
 3181 CCATGGACCAGCTGTATTGTGGTGGACGGCCAGATGTGCGGCAGCTGTGCCAATTGCTGG
 447 P W T S C I V V D G Q M C G S C A N C W
 3241 TTTAATGCCAGTGGTGCCAGATGCAGTTTTTACGgtaagctttcctcagatgcttgcaaa
 467 F N A S G A R C S F H
 3301 tgcgacgatgagctacaatgctaacatcacggtctcagAAACGCGCAACCCCCAAAAAT
 478 E T R N P Q N
 3361 TCTCAACAGCAGCAGCACAATGCGGCGATCCTGCCGAACGGGGCGAACGGTGGAATTAAT
 485 S Q Q Q Q H N A A I L P N G A N G G I N
 3421 ATTTACAGCTGACCCCAATGCCTTTAGGTTTGGCACTCCTCACCCGCTTGGTACTCATGCT
 505 I S A D P N A F R F G T P H P L G T H A

3481 GCCCTTGCTGCTCCTGCTCCCGTGAGTGCTATCGCTCCAGCAGCTCCCTCGATGAGCAAC
 525 A L A A P A P V S A I A P A A P S M S N

3541 CCAATGCTTCAGCACATGTTGAATCGGGCCCTGTCCGAGGTACGGAGCGCTGACAAGGCT
 545 P M L Q H M L N R A L S E V R S A D K A

3601 ACCCGGCAACTGATTCTCATGGAATGCGCTGCCAAAAAACTGGCGTTGAAGATGGTCGAT
 565 T R Q L I L M E C A A K K L A L K M V D

3661 TATGAGGAGCTTATTAGCTCCCAGGAGCAGTCAGGGAACGCCGGAGGACCTGGCCAGCAG
 585 Y E E L I S S Q E Q S G N A G G P G Q Q

3721 GGCTTGGGTGAAGATGCGGGTGCATAGGaaagaaaacgctcctgatttacaaatggagaaac
 605 G L G E D A G A *

3781 ttctcactcgttccagacttagaggactttatgggcatacatatggcccgtcggcttgca
 3841 tccctgatcgtgcggttcttttcangacgcaaaagtnttcccgtcctgcaactngacca
 3901 tcaatncaggacccagcatgggtcatgactgtccgcacaaccacgcagttgcgtggatat
 3961 actggttgctacccgattcgcgatgtgaaggatcatcaccccatatttaacctttcgacat
 4021 atcacagcagaaaatttgattcctttcgaaggatcaagggtcatgaagagccttatccaca
 4081 tagaaccactaccaacaccttttcgtttcttcgtttgggggcttcctagagaatttatca
 4141 ttaggctgattggatggaatgggggaaaagacaaatcttagccgggacagatactttgat
 4201 atttttTcttcagactcgtatgtcgatagtggcacggatgcaaactctaccagcgataccc
 4261 aagtcccgcgcccaaaccctcctcttcaaggcaggaagcaatttgacaatgcctccagcct
 4321 aaaaaaa

Table 1. The amino acids composition of the SBR protein. Glutamine constitutes nearly 20% of total amino acids. Four amino acids (alanine, glutamine, glycine and proline) make 50% of total amino acids. 77.7 % (87 out of 112) of glutamines and 70% (29 out of 41) histidines are found in the two glutamine-rich regions. Proline or leucine content is also high in either region.

Amino acid composition of SBR protein and the two glutamine-rich regions

Residue	Total protein		1 st Q-rich region(86-193)		2 nd Q-rich region(293-369)	
	Number	Percentage	Number	Percentage	Number	Percentage
A Ala	69	11.27	6	5.56	3	3.45
B Asx	0	0.00	0	0.00	0	0.00
C Cys	8	1.31	0	0.00	0	0.00
D Asp	25	4.08	7	6.48	2	2.30
E Glu	24	3.92	4	3.70	0	0.00
F Phe	6	0.98	1	0.93	0	0.00
G Gly	55	8.99	0	0.00	4	4.60
H His	41	6.70	14	12.96	15	17.24
I Ile	11	1.80	2	1.85	0	0.00
K Lys	12	1.96	0	0.00	0	0.00
L Leu	42	6.86	<u>12</u>	<u>11.11</u>	1	1.15
M Met	15	2.45	0	0.00	3	3.45
N Asn	24	3.92	4	3.70	0	0.00
P Pro	54	8.82	5	4.63	<u>12</u>	<u>13.79</u>
Q Gln	112	18.30	47	43.52	40	45.98
R Arg	22	3.59	0	0.00	1	1.15
S Ser	40	6.54	2	1.85	4	4.60
T Thr	18	2.94	2	1.85	0	0.00
V Val	27	4.41	2	1.85	1	1.15
W Trp	2	0.33	0	0.00	0	0.00
X Unk	0	0.00	0	0.00	0	0.00
Y Tyr	5	0.82	0	0.00	1	1.15
Z Glx	0	0.00	0	0.00	0	0.00
TOTAL	612	100.00	108	100.00	87	100.00
MW	65809.37 Da					

RT-PCR with the primer pair *sbr*-001+ *sbr*-yy9 generated no specific product, whilst RT-PCR with primer pair *sbr*-002+ *sbr*-yy9 resulted in specific amplification. These results were further confirmed by nested PCR with interior *sbr* primer *sbr*-004 (Figure 5). Therefore it was obvious that the primer *sbr*-yy1 was not within the RT product sequence.

Using RNA extracted from the *sbr* mutant, no specific product was obtained with the same primers in RT-PCR and nested PCR as that for the wild type (Figure 5). To test whether this might be due to the poor quality of the mutant RNA, two primers from the *ro4* gene (Robb et. al., 1995) which flank two introns, were used. The RT reaction was performed with the poly(T) containing RACE primer mentioned earlier. PCR of the RT product revealed a specific band lacking the intronic sequences of the product from the PCR of genomic DNA (Figure 5). This confirmed the integrity of the RNA from the *sbr* strain and therefore the absence of *sbr*- RNA in the mutant.

B. Sequence of the RT-PCR product and the translation start site

The RT-PCR product was directly cloned into PCRII-TOPO vector and sequenced. No additional introns were detected between primer *sbr*-002 and the 5' end of the previously sequenced truncated *sbr* cDNA (Figure 4). The cDNA sequence showed a stop codon upstream to the *KpnI* site and an ATG codon just downstream. The sequence of this region was confirmed by 4X sequencing of the RT-PCR product or genomic DNA.

A search for possible transcription start sites within the *SalI* genomic clone sequence upstream to primer *sbr*-002 revealed no strong match to the consensus,

TCATCANC (Bruchez et al 1993b). However, a similar sequence, TCTTCACC, was located 904 bp upstream of primer *sbr*-002. In addition, no close similarity to the Kozak fungal consensus sequence CNNNCAMVATGGC (Bruchez et al 1993a) was found.

Therefore the ATG codon just downstream of the *Kpn*I site was assumed to be the translation start site. Figure 5 shows the complete sequence of the *sbr* gene and 5' flanking region, and the translated protein sequence.

Amino acid composition of SBR protein

The complete *sbr* ORF is 1.84 kb, encoding 612 amino acids (Figure 6). This extends the 5' end of the truncated cDNA by approximately 600 bp. In addition to the glutamine-rich region from the truncated cDNA, there was one more glutamine-rich region, located at amino acid sequence 86-193. Overall, there are 112 glutamines for nearly 20% of total amino acids (table 1). The second most abundant amino acid is alanine at 11% of the total amino acids. The third and fourth most abundant residues are glycine (8.99%) and proline (8.82%) respectively. These four amino acids make up nearly 50% of the total amino acid composition. There were 75 positively charged amino acids (histidine, lysine, and arginine) and only 49 negatively charged ones (aspartic acid and glutamic acid).

Analysis of the glutamine-rich domains showed that each one consists of more than 43% of glutamine and they contain nearly 80% of this amino acid in SBR (table 1). Another interesting finding was that both regions contained high percentage of histidine. These two domains have 70% of all the histidines in SBR. A high percentage of leucine and proline were also found in the glutamine rich domains (table 1).

Construction of *sbr* ORFs and expression plasmids

Since the truncated *sbr* cDNA has a *NotI* site within the RT-PCR product of the 5' terminal cDNA (Figure 4), this digestion site was used to link the two pieces of cDNA to form a complete *sbr* ORF. Briefly, the 5' terminal cDNA was cloned into pBluescript at *KpnI* and *NotI* sites (Figure 7A). The truncated 3' terminal cDNA cloned at *EcoRI* site (Campsall, 1998) (Figure 7B) was then cut out by *NotI* and ligated to the 5' cDNA clone at the *NotI* site. After transforming DH5 α cells, colonies with the correct insertion orientation were determined by PCR with suitable primers and tested for different digestion sites within the sequence (Figure 7C).

Expression plasmids with different *sbr* ORFs were constructed in pQE vectors based on the locations of possible functional domains (Figure 8) to fuse the 6X histidine tags at the N- termini. Since the *sbr* start codon is between the *BamHI* and *KpnI* sites (Figure 4), the *BamHI* site was used to construct these expression clones in the correct orientation resulting in expression of the vector's start codon and the 6 histidines. After transformation, colonies with the correct ORF were selected by digestion with particular enzymes and by PCR with the pQE forward primer and *sbr* specific primers as shown in figure 3 and 8, and confirmed by sequencing. The cloning strategies involving the use of genomic clones, cDNA clones, PCR primers, and digestion sites are illustrated in figure 8

Figure 7. Diagram illustrating construction of the complete *sbr* ORF. The 5' terminal cDNA sequence was cloned into pBluescript at the *KpnI* and *NotI* sites (A). The truncated cDNA cloned at *EcoRI* site was released by *NotI* digestion and inserted into clone A. Correct insertion was selected by *AvaI* digestion and PCR with vector forward primer and *sbr* specific reverse primer (C). Other digestion sites are indicated. *BamHI* and *PstI* sites were used in construction of expression clones shown in figure 8.

A

B

Cut and link at NotI site

C

Figure 8. Diagram illustrating construction of expression clones. (A) Indicates primers, digestions sites and used source DNA from genomic or cDNA clone. (B) Five expression clones with different deletion of putative functional domains. The name of each clone is indicated at the left side, and the vector of the clone at right side.

A

B

■ cDNA sequence

▨ ORF

▤ Glutamine-rich

■ Helix-turn-Helix

▤ Cysteine-rich

Table 2. Details in construction of different expression clones. The names of the clones and primers are indicated in figure 8.

Name (see figure 7)	amino acids	vector	cloning details (see primer locations in figure 7)
PQE-SBR600	600	pQE-30	cDNA clone containing the entire <i>sbr</i> ORF was digested with <i>Bam</i> HI and <i>Pst</i> I. the released cDNA fragment was inserted into pQE-30 vector.
PQE-SBR470	468	pQE-30	genomic DNA from <i>Kpn</i> I clone (see figure 2) was digested with <i>Bam</i> HI and <i>Hind</i> III and the released DNA fragment was inserted into pQE-30 vector.
PQE-SBR420	421	pQE-32	cDNA clone containing the entire <i>sbr</i> ORF was used for PCR with the universal primer and <i>sbr</i> primer <i>sbr</i> -pQEyy containing <i>Bam</i> HI site. PCR product was cloned at <i>Bam</i> HI and <i>Pst</i> I sites.
PQE-SBR400	407	pQE-30	the <i>Sa</i> II clone was used for PCR with primers <i>sbr</i> -002 and <i>sbr</i> pQE4(containing <i>Pst</i> I site). PCR product was cloned into pQE-30 at <i>Bam</i> HI and <i>ps</i> tI sites.
PQE-SBR350	357	pQE-30	the <i>Sa</i> II clone was used for PCR with primers <i>sbr</i> -002 and <i>sbr</i> pQE3(containing <i>Pst</i> I site). PCR product was cloned into pQE-30 at <i>Bam</i> HI and <i>ps</i> tI sites.

with the details shown in table 2. The pQE-SBR600 (Figure 8) encodes almost the entire SBR protein, with only 12 amino acids deleted at N-terminus. The pQE-SBR470 construct also encodes all potential functional domains but with a deletion of C-terminal 144 amino acids. The pQE-SBR420 was constructed with a deletion of the first glutamine-rich region. The pQE-SBR400 lacks the putative zinc finger domain and its C terminal sequence, while the pQE-SBR350 has deletions of both putative DNA binding domains and their C-terminal region (Figure 8).

SBR protein expression and purification

The SBR expression constructs were introduced into M13 rep cells and induced to express the SBR fusion proteins by the addition of IPTG. The expression efficiency and specificity were examined by SDS-PAGE (Figures 9,10) and western blots with anti-histidine antibody (data not shown). No protein was expressed in the control strain transformed with the cloning vector alone (Figure 10). Proteins were lightly expressed in the control without IPTG (Figure 9A). The solubility of SBR proteins in M13 cells was examined by comparing the sample of total proteins with the sample pelleted after native lysis. Very little SBR protein remained in the pellet, indicating SBR proteins were soluble under native condition (Figure 9B). The 6 His-tagged SBR proteins were purified under native conditions with Ni-NTA agarose matrix and assessed by SDS-PAGE (Figures 9B, 10). Very little contamination by other protein was detected. In the 10-12% SDS-PAGE, the glutamine-rich SBR protein migrated slower than expected, due to high proportion of basic amino acids in SBR. For example, the full ORF protein was about 66

Figure 9. Examples showing solubility (A) and purification (B) of SBR proteins. A: three proteins (A1-A4: pQE-SBR600, B1-B4: pQE-SBR420, C1-C4: pQE-SBR470. see figure 8 for names) show high solubility. A1, B1 and C1 are control without IPTG. A2, B2 and C2 are samples induced with IPTG. A3, B3 and C3 are supernatants after lysis and centrifugation under native condition. A4, B4 and C4 are from the pellets. Arrows indicate expressed proteins. B: purification of pQE-SBR600. T, total sample before purification. FL, flow through of sample from the affinity matrix. W1 and W2, collected first and second wash buffer respectively. E1 to E4, first to fourth elution. M, marker protein.

A

B

Figure 10. Purified proteins on one gel. A, control of vector alone. B, pQE-SBR600. C, pQE-SBR470. D, pQE-SBR420. E, pQE-SBR400. F, pQE-SBR350. G, marker. Note pQE-SBR420 migrated faster than pQE-SBR400 and pQE-SBR350. See figure 8 for names.

kDa (see table 1), but it migrated as a 99 kDa protein (Figures 9, 10). Similarly, the fusion protein PQE-SBR420 has 421 amino acids but it migrated slower than PQE-SBR400 and PQE-SBR350 products (Figure 10).

DNA binding study

A. Isolation of the SBR binding fragments from genomic DNA.

To test whether the SBR protein is a DNA binding protein, about 20 µg of genomic DNA from *N. crasssa* wild type 988a was digested with *Hind*III and *Pst*I and allowed to interact with SBR protein of full size ORF, as described in the Methods and Materials. The SBR fusion protein and associated DNA were enriched by binding to Ni-NTA agarose matrix. Co-purified DNA fragments were cloned into pBluescript and the selection process was repeated 5 times. Three independent clones, 294 bp, 667 bp and 828 bp in length, were isolated and sequenced and are designated SBRBD300, SBRBD600, and SBRBD800 respectively (Figure 11).

To test if the selected DNA was due to incomplete washing after the binding reaction, a final wash buffer was collected as control to transform competent cells. No colonies grew after transformation. The vector alone was also used to bind SBR and it was subjected to the same treatment as the selected plasmids with genomic target DNA. After transformation, many but fewer colonies (about 70%) grew than that of genomic clones.

B. Electrophoresis mobility shift assay (EMSA).

The shortest fragment, SBRBD300, was radio-labeled with ^{32}P dCTP by end filling with Klenow fragment 1. Because the vector plasmid showed nonspecific binding with SBR, it was used as competitor in the EMSA. When no competitor was present, the protein-DNA complex did not migrate at all (Figure 12). When the ratio of competitor/probe was 30/1, two bands of DNA-protein complex appeared and the slower migrating band disappeared when competitor was increased to 80-fold. The other band did not disappear until the ratio was 220/1 (Figure 12B).

EMSA was also performed with mutant SBR protein pQE-SBR400, generated by deleting the cysteine-rich region and the downstream region (Figure 8). In contrast with the assay using wild type SBR, two bands of protein-DNA complex appeared when no competitor was present (Figure 12B). The slower migrating band indicated multiple SBR molecules in the complex. These two complexes disappeared when small amounts of competitor DNA (at 25-fold) was present.

The other two fragments were not tested for EMSA because large fragment could contribute more non-specific binding sites and limit the sensitivity.

Figure 11. The sequences of three isolated genomic fragments bound to SBR protein.
The name and size of each fragment are indicated.

Name: SBRBD300.

size = 294bp.

```
1      CTGCAGGATC GAACTAACAG CTGCAAGGTG AGATGATGGC GTTTTGGTGG CTTTGTAG
61     CGCGGGTTAT CACCAGGCCT CCGCACAGAT TCGGGGCATT CCTTTTCAGC TAAGTAAGCA
121    TCGCAGACTG CGCAGATCGC TTTACAAACC CAAACCCTAA GGCACCCAAG TTGAGTTTCT
181    GTAATCGCCA TTGGTAAAGC TGTCAGTCAG GCCCTGTCAG CTGAAAACGA ACACCCCGAG
241    TCCTTTCGGT ACCGGTTCCG TGCTCATCAG CTGGCGTCGT CAACAAGAAA AGCTT
```

Name: SBRBD600.

size = 667bp.

```
1      CTGCAGCAGA TGCTGAAAAG CATTGGCTAT GATGAAGCAA GGGAGGATGC ATACAGACTC
61     AAAGGCGTCC AGCTCATTGA TAGCGTTCGA CAAAGCATAA GCCTGTSCGG ACCCATCGCT
121    TCCAAGCCTT ACCTACGCAT TTTGCTGACG AACGCCACAG GCCGGTCAAG ACATTGATA
181    CAGCAGCGAC CTAATACCAC AGGTTCAGAA TACGGTTTCC GAGCGCCGAG TACAATACG
241    AAGATGTAGC TCTTGCGTCG TTGKTGTTG CATGCAAGGT GGAGGAKACG ATCAAGAAGT
301    CCAAGGATGT CCTATGTSCT GCGACAATA TCAGGCAGCC CCACGATCAA AGGACTCCGG
361    ATGACAAGGT TCGTCAATTA TCTTGTGGC GATCTACGGT GTACCTACTA ATATTGACCG
421    CGCAGATGTT CGATGGTCCA TTCCAAATTS ACAGTGGGCC TCGAGCGTKC ATATTCTGGA
481    GACCATTGCG CTTTCGACTT TCCAGGCTCA GTTATTCCCC AGAAGTTGCT GATTBAACAT
541    TACTTCGGAA GATGTTCCCA AAGGGCGGGA AAGGGGACAG TCCCGAGGTC AAGAAGTTCA
601    TCAGCGATGC ATACGACATG TCCATCGATC TTTACAAGAC GTTTGCGCCC TTGAAGCAAC
661    CAAGCTT
```

Name: SBRBD800.

size = 928bp.

```
1      CTGCAGTCTG GATGCCCAGA CTGGGCACGG CACGGCAAGG CAGCCTCTCG CAGCCCAGCY
61     AATCTSGGCT GTCACAGCTA GTCTCAMCYA TMATACACGA TATATCTCGA AATCCAGGAA
121    TTTCCCTATC AAATGCCGGC TCACGCGAGA MAGAMACCAT CGTCCTCGAT GGACCTAAAC
181    CTCCAAGGCT TCTCCGACAA GTKCCAGGCC TCCTCCGACA AATCTGTCCA TCCGGGATGC
241    TTCGGGGCGC ATCGGGATGC ACCAGCATTG CAGTGGTGAA RACGGTCAAG GTCCGGGGAG
301    TGCCGATGAA TTCCGGGAAG CTAGATRACA GCCAAGAAG GTGACCGGCA CTCGAGAAGG
361    TGACTGGCAG TGGCACTCGA GAAGGTAATC GGCAGTCTCT GGCAGTTGAG AGGGAGACTG
421    GCAGTCTCTG GTACTCGAGA AGGTGACCGG CAGTCTCCAG CACCCGAGAG GGTAAGTGGC
481    AGTCTCTGGC ACTCGAGAGG GTGGCTGGCA CTCGCTGGCA VTCGAGAGAS AGAGGCACTC
541    AAAAGTGCCA TTGCCGTAGA GAACGCTGGT TCTATGGATA AATAGAAAAG AAGAAAGRSA
601    AAGGGGTGTA AACGGGGGCT CTGGCTGGCA CCGGCTGGCA CGAGCBTCAG GGAGCTGACA
661    AGGGAGTTGA CACGGCTRTC AGAGCTGGTA AGGGRGCTGA CAGGGAAGKT GATATGGCTG
721    TCAGAGCTGK AAAGRGCTGG CAAGAGRGCT GGCAAGAGAG CTGGCAAGAG AGCTGGCAA
781    AGGGGGTGAC AGCTGTGTTG ACTGGTAATG CCAGTGTGTG TGAAGCTT
```


Figure 12. Results of EMSA using the isolated fragment SBRBD300 as probe and the cloning vector as competitor. (A): EMSA of pQE-SBR600. Lane 0, no protein added. Lane 1 to lane 5, competitor DNA at 0, 30, 50, 80, 100 fold. (B): EMSA comparison of pQE-SBR600 and pQE-SBR400. Lane 0, no protein added. Lane 1-4, pQE-SBR600 with competitor DNA at 0, 100, 160 and 220-fold. Lane 5-8, pQE-SBR400 with competitor DNA at 0, 25, 50, 100 fold. Arrows indicate specific DNA binding. Note, no binding for pQE-SBR400 was detected when competitor was present at 25-fold or higher ratios.

A

B

Sequence analysis of the isolated fragments

A. Sequence analysis of SBRBD300

A similarity search of the MIPS *Neurospora crassa* database and NCBI BLASTN retrieved a 77715 bp genomic DNA from *N. crassa* linkage group II. Position 63500 to 63844 contains the sequence of SBRBD300. The sequence from SBRBD300 and 2520bp downstream or upstream of it were analyzed for genes which might be responsive to SBR regulation. The upstream sequence was converted to reverse complementary sequence for descriptive convenience.

a) Sequence of SBRBD300 and downstream 2520bp (Figure 13).

Two putative genes, predicted by the BioMax Informatic program in Germany, were retrieved in the NCBI BLASTX search with the SBRBD300 + downstream 2520bp sequence as query. The first ORF encodes 138 amino acids and was derived from sequence 36-797 by joining four small exons (exon1=29bp, intron1=144bp, exon2=95bp, intron2=79bp, exon3=69bp, intron3=115bp, exon4=220bp). This putative gene was not detected by any other search algorithm.

A second ORF, encoding 110 amino acids from Nt 1387-1719, was also detected to some extent by three other programs (see Material and Method section). GENSCAN found a possible ORF encoding 174 amino acids at region 986-1719. An overlapping ORF of 206 amino acids, detected by CGG program, is from Nt 735-1719. The MZEF program found a likely gene from Nt 1605-2090. In summary, an ORF between Nt 1387-1719 was identified by the three programs, and the region from 986-1719 was identified by two of them.

Gene predictions for the reverse strand by GENSCAN, CGG, and MZEF all predicted a gene overlapping Nt 452(-) to 2255(-) (Figure 13B). The ORF predicted by GENSCAN encodes 264 amino acids, from Nt 916(-) to 2255(-). CGG predicted a gene encoding 206 amino acids from Nt 730(-) to 1711(-). The gene from MZEF prediction was shorter, from Nt 452(-) to 1093(-).

Unfortunately, the BLASTX algorithm detected no significant homology between these predicted gene products and known genes or ESTs from other organisms. In searching for motifs, no significant similarity with high matching score was found.

b) sequence of SBRBD300 and upstream 2520bp (Figure 14)

No predicted genes have been reported at the MIPS facility for sequences 2520 bp upstream of SBRBD300. However, use of two of the alternative algorithms identified several possible genes in this region.

The sequence of SBRBD300 and the upstream 2520 bp was converted to its reverse complementary strand to maintain the fragment SBRBD300 at the 5' end. Several large ORFs were found in this sequence, the largest of which is 564 bp, from Nt 1603 to 2166 (Figure 14A). The second largest ORF in this region is 276bp, extending from Nt 698 to 973, encodes 91 amino acids. The MZEF program identified two possible genes: one Nt 698-1162, and overlapping two adjacent ORFs in frames 2 and 3. The other one, from Nt 1171-2166, spans two ORFs in the same frame. The largest ORF was detected by both of the predicted genes from CGG and MZEF program. No putative genes were identified by the GENSCAN program. On the complementary strand, a large gene was predicted by the CGG program but not by GENSCAN or MZEF (Figure 14B).

Unfortunately, no significant similarity with other known protein sequences was detected by BLASTX for these predicted genes.

B. Sequence analysis of SBRBD600

A BLASTN search found no sequences matching SBRBD600 in the *Neurospora* genome database and the gene identification programs predicted no gene for this short fragment. Translation with the DNAMANTM software package revealed an ORF of 80 amino acids in frame 3, with ATG site at 426(+) but no termination codon at the end of sequence (Figure 15). The translation of this ORF was displayed by NCBI ORF program in figure 15B. Another ORF of 34 amino acids from frame 1 was only 15bp up with a ATG codon at 307(+). In fact, in frame 3, there was no stop codon downstream the two consecutive stop sites at 78 and 81.

C. Sequence analysis of SBRBD800

Like the SBRBD600, no gene was predicted for this fragment. No stop codons are present from Nt -540 to -1, and an ATG start codon is present at position -268 (Figure 16A). The NCBI ORF program identified this region in frame -3, with 89 codons from the ATG start codon at -268 to -1 as a putative gene fragment (Figure 16B). Another DNAMAN ORF of 76 amino acids was found in the -3 frame from Nt -551 to -321 (Figure 16A).

Figure 13. Gene prediction for SBRBD300 and downstream sequence. The relative position of the SBRBD300 fragment is shown at left side underneath the scale ruler. Analysis was done for both forward strand (A) and reverse strand (B). For each strand, ORFs or genes are aligned from top to bottom in order as: DNAMAN translation overview, NCBI ORF display, and predicated genes by different programs. The orientation of ORFs and genes is indicated by arrows in a DNAMAN translation overview. Note that the largest DNAMAN ORF in forward strand (at 1387-1719) is overlapped by predicted genes from all programs. The reverse strand shows shorter and fewer ORFs.

A

B

Figure 14. Gene prediction for SBRBD300 and upstream sequences. The sequence was converted to the reverse strand with SBEBD300 oriented at the 5' end. Analysis was done for both forward strand (A) and reverse strand (B). For each strand, ORFs or putative genes are aligned from top to bottom in order as: DNAMAN translation overview, NCBI ORF display, and predicated genes by different programs. The relative position of the SBRBD300 fragment is shown at left side underneath the scale ruler. The orientation of ORFs is indicated by arrows in the DNAMAN translation overview. (A) shows longer ORFs than that in figure 13A. The largest ORF (1603 –2166) overlaps predicted genes identified by two different computer programs.

A

B

Figure 15. Analysis of SBRBD600. Since no gene was predicted for this short sequence, this figure shows ORFs from DNAMAN translation overview (A) and NCBI ORF display (B). The plus 3 frame in (A) has no stop codon after position 81. The sequence from 261 to the end was translated in (B). Note there is no stop codon at the end of translation.

A

B

Length: 135 aa

Accept

Alternative Initiation Codons

```

261 ttgtktgttgcattgcaaggtggaggakacgatcaagaagtccaag
   L X V A C K V E X T I K K S K
306 gatgtcctatgttctgcgcacaatatcaggcagccccacgatcaa
   D V L C X A H N I R Q P H D Q
351 aggactccggatgacaaggttcgtcaattatcttgttggcgatct
   R T P D D K V R Q L S C W R S
396 acggtgtacctactaatattgaccgcgcagatgttcgatggtcca
   T V Y L L I L T A Q M F D G P
441 ttccaaattsacagtgggcttcgagcgtkcatattctggagacca
   F Q I X S G P R A X I F W R P
486 ttgtgctttcgactttccaggctcagttattccccagaagtgtct
   F G F R L S R L S Y S P E V A
531 gattbaacattacttcggaagatgttcccaaagggcgggaaaggg
   D X T L L R K M F P K G G K G
576 gacagtcccagggtcaagaagtccatcagcgatgcatacgacatg
   D S P E V K K F I S D A Y D M
621 tccatcgatctttacaagacgtttgcgcccttgaagcaaccaagc
   S I D L Y K T F A P L K Q P S
666 == 567

```

Figure 16. Analysis of SBRBD800. Since no gene was predicted for this short sequence, this figure shows ORFs from the DNAMAN translation overview (A) and NCBI ORF display (B). The minus 2 frame in (A) has no stop codon from 540 to 1. This ORF is shown in minus 3 in (B). The sequence from 460 to the left end was translated in (B). Note there is no stop codon at the end of translation.

A

B

View 1 GenBank ▾ Redraw 100 ▾ SixFrames

Length: 153 aa

Accept

Alternative Initiation Codons

Frame from to Length

-3	1..460	460
+3	12..344	333
-2	321..551	231
+3	621..827	208
-3	647..826	180
+2	2..178	177
+1	133..279	147
-1	1..132	132
-1	721..827	108

```

460 ctcgagactgccggtcaccttctcaggtaccagagactgccagtc
    L E T A G H L L E Y Q R L P V
415 tcccttcaagtgccagagactgccgattaccttctcagtgcca
    S L S S A R D C R L P S R V P
370 ctgccagtcaccttctcaggtgccggtcaccttcttggctgtya
    L P V T F S S A G H L L M L X
325 tctagcttccccgaattcatcgccactccccggaccttgaccgty
    S S F P E F I G T P R T L T V
280 ttcaccactgcaatgctgggtgcattcccgatgcgccccgaagcacc
    F T T A M L V H P D A P R S I
235 ccggtatggacagatttgcggaggaggcctggmaccttgcggaga
    P D G Q I C R R R P G T C R R
190 agccttggagggtcttaggtccatcgaggacgatgggtkctctctcg
    S L G G L G P S R T M V S X S
145 cgtgagccggcatttgatagggaattcctggatttcgagatata
    R E P A F D R S I P G F R D I
100 tcgtgtatkatrgktgagactagctgtgacagccsagattrgctg
    S C X X X E T S C D S X D X L
55 ggcctgcgagaggctgccttgcctgcccgtgcccagctctgggcacc
    G C E R L P C R A V P S L G I
10 cagactgcag 1
    1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
  
```

D. Binding sites search.

In searching binding sites for transcription factors, the bioinformatics program MatInspector (see Materials and Methods section) was used. Many candidate binding sites were found for the SBRBD300, SBRBD600, or SBRBD800. (table 2). Only those with highest matching score and common to each isolated DNA fragment were listed in table 2. All binding sites listed had score of 1 (100% match) to core sequence. Two of the fungal transcription factor binding sites, NIT2 site and STUAP site, are listed. NIT2 is an activator of nitrogen-related genes (Fu and Marzluf 1990a). STUAP is *Aspergillus nidulans* Stunted protein (Dutton et al, 1997). Some binding sites from other species were found with very high matching scores. AP1 (activator protein 1) belongs to family of MADS box containing proteins in vertebrates, plants and yeast (Karin et al., 1997). DOF proteins are a family of transcription factor in plants (Yanagisawa and Schmidt, 1999). DFD proteins are deformed (Dfd) homeotic genes encoding transcriptional regulatory proteins in *Drosophila* (Egger et al., 1992). GATA are GATA binding factors found in vertebrates, insects and fungi (Ko and Engel, 1993; Lowry and Atchley 2000, reviewed by Scazzocchio, 2000).

In addition to the listed common sites, other sites with high scores but not common to all three fragments were found. An activator protein 4(AP4) site (consensus: CAGCTG. Hu et al., 1990) was found in SBRBD300 and SBRBD800. An EVI1 site (consensus: TGACAAGATAA. Perkins et al., 1991) was found only in SBRBD600

Table 3. DNA binding sites of six transcription factors. These six sites were found with high matching score in each of the isolated genomic fragments. Analysis was performed by MatInspector (see text for reference). Four of the binding factors were reported in fungi. Example sequences show good match to the consensus sequences.

A:

Binding sites in SBRBD300:

Transcription Factor	Matching at position # in query DNA (ordered by score)	score		Consensus *	Example & its position (core in uppercase)
		Core	matrix		
DOF	48-, 137-, 192-, 285+, 240-, 99-	1.000	0.986-0.900	WAAAGC	aAAAGc (48-)
NIT2	68+	1.000	0.967	TATCDH	TATCac (68+)
DFD	176-	1.000	0.964	ATTAMY	ATTAcA (176-)
GATA	65-	1.000	0.950	WGATAR	tGATAa (65-)
API	201-, 197-, 211-, 274-,	1.000	0.935-0.885	RSTGACTNMNW	ccTGACtgaca (201-)
STUAP	58-, 58-	1.000	0.902-0.872	WWCGCGWNM	agCGCGggt (58+)

B:

Binding sites in SBRBD600

Transcription Factor	Matching at position # in query DNA (ordered by score)	score		Consensus *	Example & its position (core in uppercase)
		Core	matrix		
DFD	533+, 403-, 371+	1.000	0.981-0.963	ATTAMY	ATTAct (533+)
NIT2	379+, 77-, 175-, 329+	1.000	0.980-0.937	TATCDH	TATCtt (379+)
DOF	13+, 488-, 88+, 345+, 627-, 326-, 556+	1.000	0.974-0.943	WAAAGC	aAAAGc (13+)
GATA	374-, 74+, 376-324-, 172+, 75+	1.000	0.974-0.910	WGATAR	aGATAa (374-)
STUAP	416+, 416-	1.000	0.918-0.877	WWCGCGWNM	acCGCGcag (416+)
API	144+, 360+	1.000	0.912-0.910	RSTGACTNMNW	gcTGACgaacg (144+)

C:

Binding sites in SBRBD800

Transcription Factor	Matching at position # in query DNA (ordered by score)	score		Consensus *	Example & its position (core in uppercase)
		Core	matrix		
NIT2	103+, 127+, 907-, 575-, 97-	1.000	0.995-0.947	TATCDH	TATCtc (103+)
DFD	380-, 800-	1.000	0.979-0.974	ATTAMY	ATTAcc (380-)
DOF	590+, 775+, 586+, 538+, 727+, 590+	1.000	0.959-0.945	WAAAGC	aAAAGg (775+)
GATA	122-, 706+, 576-	1.000	0.957-0.933	WGATAR	tGATAg (122-)
API	359+, 785+, 340+, 442+	1.000	0.941-0.918	RSTGACTNMNW	ggTGACtggca (359+)
STUAP	141-, 141+	1.000	0.932-0.924	WWCGCGWNM	ctCGCGtga (141-)

NIT2: activator of nitrogen-regulated genes in *N. crassa*. (Fu and Marzluf, 1990)

STUAP: *Aspergillus* stunted protein in *Aspergillus nidulans*. (Dutton et al., 1997)

GATA: GATA factors. Found in vertebrates, insects, and fungi. (Ko and Engel, 1993)

DFD: deformed homeotic genes encoding homeodomains, *Drosophila*. (Egger et al., 1992)

DOF: single zinc finger transcription factors in plants (Yanagisawa and Schmidt, 1999)

API: activator protein 1 in vertebrates, plant, yeast etc. (Karin et al., 1997)

+ forward strand, - reverse strand

* D = A or G or T

H = A or C or T

M = A or C

R = A or G

S = C or G

W = A or T

Y = C or T

while a STAT binding site (consensus: TTCCCRKAA, Horvath et al., 1995) was detected only in SBRBD800. Both sites had a score of 1.000 for core and matrix as well (data not shown).

Complementation with *sbr* genomic DNA

A 6.6 kb *SalI* fragment containing the entire *sbr* gene was used to transform spheroplasts of *sbr* mutant. In three electroporation experiments and two PEG-mediated transformations, only one transformant with a wild type phenotype was detected (Figure 17B). Three separate transformation experiments with a second genomic clone, containing a 5.4 kb, *XbaI* and *SalI* fragment (see figure 3B) did not yield any wild type transformants.

Confirmation of *sbr* transformation

To confirm that the complemented strain was a transformant, its genomic DNA was compared to both the wild type and the *sbr* mutant. This analysis employed: 1) PCR with primers which flank the original insertion site in the *sbr* gene, and 2) hybridization of a Southern blot with probes of the hygromycin B gene or *sbr* genomic DNA adjacent to the insertion site (Figure 17).

The Southern blot of genomic DNA from all three strains, each digested with *XbaI*, *HindIII*, *KpnI*, *EcoRI*, and *SalI* was first probed with the hygromycin B gene. All three strains had identical hybridization patterns (data not shown). In contrast, hybridization with the *sbr* probe, revealed the polymorphism between wild type and *sbr*

Figure 17. Schematic diagram of the experiments to confirm complementation of the *sbr* mutant with the wild type, *sbr*⁺ allele. (A) shows some digestion sites in the insertional plasmid pCSN43. (B) shows the some reference digestion sites, the location of probe used in (C), location of primers used in (D), and the original, mutagenic insertion site in the *sbr* gene. (C) Southern blot of *Sa*I digested, genomic DNA from the *sbr* mutant (lane 1), the transformant (lane 2) and wild type strain (lane 3). The probe used is shown in (B). Note lane 2 has the combination of patterns of lane 1 and lane 3. (D) PCR results for genomic DNA from wild type (lane 1), mutant (lane 2) and the transformant (lane 3). Primers are shown in (B). Note lane 3 has the combination of patterns of lane 1 and lane 2.

A

B

C

1 2 3

D

Figure 18. Mapping result for *sbr* gene. Southern blot of *Kpn*I digested genomic DNA from 20 mapping strains (left to right , #4411 to #4430) was hybridized with the 6.6kb probe, *Sal*I *sbr* genomic clone (see figure 3). Strain #4411 and #4416 are the parents of mapping strains. Patterns of hybridization are grouped as 'O' (4411 pattern, 'O' represents 'Oak Ridge' genetic background) or 'M' (4416 pattern, 'M' represents Mauriceville strain). For reference of mapping and the strains, see Metzenberg et al. (1984). The segregation of *sbr* gene in the 20 strains matches that of *trp-1* gene on linkage group III (Metzenberg and Grotelueschen, 1992).

4411

4416

4411	4416	4413	4414	4415	4416	4417	4418	4419	4420	4421	4422	4423	4424	4425	4426	4427	4428	4429	4430
O	M	O	M	M	M	O	O	M	M	O	O	M	M	O	M	O	M	O	M

strains caused by the original mutagenic event. Both bands are present in the transformed strain due to complementation by an ectopic copy of the *sbr*⁺ allele introduced by electroporation (Figure 17C).

This was confirmed with the PCR-based analysis in which the *sbr* mutant showed no specific amplification compared with the wild type using *sbr* specific primers (sbr-002, sbr-004) flanking the insertion site. Amplification of the complemented transformant's DNA resulted in combined PCR products of the wild type and *sbr* strains (Figure 17D).

Gene mapping

The small cross, RFLP mapping strains were used to map the chromosomal locus (Metzenberg et al., 1984). A Southern blot of genomic DNA from the parental strains (#4411 and #4416), digested with 8 different enzymes (*EcoRV*, *EcoRI*, *BamHI*, *KpnI*, *SalI*, *XbaI*, *SstI* and *ApaI*) was first probed with the 6.6 kb *sbr*⁺ clone (Figure 3A). A distinctive polymorphism between parental strains was detected in the *KpnI* digest. Subsequent hybridization of a Southern blot of *KpnI* digested DNA from all 20 mapping strains were then scored for the 4411 (O) or 4416 (M) banding pattern (Figure 18). The pattern of segregation among the 20 strains matches that of genes *con-7* and *trp-1* on linkage group III (Metzenberg and Grotelueschen, 1992). Therefore, the *sbr* gene is located on chromosome III, near genes *trp-1* and *con-7*.

Discussion

cDNA sequence and translation start site

The *sbr* gene was initially identified by insertional mutagenesis and cloned by marker rescue (Campsall, 1998). As evidenced by both Northern blotting experiments and the probing of multiple cDNA libraries, *sbr* mRNA abundance is very low (Campsall, 1998). The lone cDNA clone recovered to date was estimated to be truncated at the 5' end by about 600 to 800 Nt. Therefore, the initial challenge in this project was to determine the extent of transcribed sequences from the *sbr* locus.

To avoid genomic contamination in the total RNA extract, reverse transcription was done with primer *sbr-yy9* which spans the intron. The *sbr* gene's putative translation start codon was determined based on the following two facts. First, RT-PCR results (Figure 5) showed no amplification with primer pairs *sbr-001* and *sbr-yy9*, whereas a specific product was generated with primer pairs *sbr-002* and *sbr-yy9*. Three possibilities for amplification failure from primer *sbr-001* are: it is within an intron, it is upstream of the transcription start site, or the RT product is truncated at a location between primer *sbr-001* and *sbr-002*.

Since no amplification with primer pairs *sbr-001* and *sbr-yy9* occurred, the transcription start site is most likely between *sbr-001* and *sbr-002* (Figures 4,5) and genomic DNA contamination of the RNA can be excluded. This was further confirmed by nested PCR (Figure 5). As confirmed by 4X sequencing coverage of the RT-PCR product and genomic DNA, the presence of a stop codon just upstream of the putative

start codon, and lack of any nearby splice sites, supports the identification of the translation start site between the *KpnI* and *BamHI* sites (Figure 4) at Nt #1843 (Figure 6).

This ATG codon is located 540bp upstream from the truncated cDNA and is the only ATG codon between the stop codon and the truncated cDNA. According to Bruchez et al (1993a), the consensus or Kozak sequence in *N. crassa* is CNNNCAMVATGGC (M represents A or C and V represents A or C or G). Although, the sequence context of the putative *sbr* translation start site (CGGACGAGTATGAA; where underlined letters agree with the consensus) differs from the consensus, considerable variance have been reported for many other genes as well. For example, the following *N. crassa* genes all diverge from the consensus: the *pzl-1* gene (GAGAAATCATGGG; Szoor et al. 1998), the inner membrane mitochondrial NADH dehydrogenase gene (AACAAACAATGCT; Melo et al. 1999), the potassium transporter gene (AAAAAAAGATGGA; Haro et al., 1999), and a secretory pathway gene (CTATAAAGATGAA; Solscheid and Tropschug, 2000).

The RACE PCR procedure (Frohman, 1990) was used to attempt to map the transcription start site more precisely but this was not successful. Attempts were also made to amplify longer cDNA sequences from seven available cDNA libraries (conidial, mycelial, conidiating, perithecial, glutamine-grown, nitrate-grown and vegetative). None of these efforts yielded any specific *sbr* cDNA sequences beyond the lone, truncated cDNA clone. This supports other data that *sbr* transcripts are rare in cells grown under a variety of conditions.

The consensus +1 transcription site in *N. crassa* is TCATCANC (Bruchez et al 1993b). Searching for that consensus found no matching within 1700 bp upstream to primer *sbr*-002. The only similar sequence, TCTTCACC, was located 904 bp upstream

to primer sbr-002. This sequence is unlikely to be the transcription start site because most *N. crassa* genes have distance of 50-150 bp from the +1 site to translation start site and a few of them, 500-650 bp upstream (Bruchez et al 1993a). Also as discussed earlier, the primer sbr-001 is likely upstream to the +1 site.

No *sbr* mRNA in the mutant strain

The failure to amplify any 5' terminal, *sbr* cDNA from the available cDNA libraires prompted the use of RT-PCR for further transcript analysis. The primer, sbr-yy9, used to prime reverse transcription (RT) of mRNA from both wild type and mutant strains. The fact that RT-PCR was not able to amplify any specific DNA from the mutant indicates that no functional *sbr* mRNA was present in the RNA extract. To ensure that this was not the result of too little mRNA or poor quality mRNA, several control reactions were performed. The first possibility could be excluded by the fact that a specific, *sbr* amplification product was obtained using the same amount of wild type RNA. The integrity of the mRNA samples was assured by RT-PCR amplification of another low abundance mRNA species, *ro-4*. (Robb et al., 1995) (Figure 5, lanes 8,9).

Complementation with plasmid genomic clone

To confirm that the phenotype of the mutant was due to insertional inactivation of the *sbr* gene, complementation experiments with the *sbr*⁺ allele were undertaken. The *sbr* mutant proved to be very resistant to transformation with exogenous DNA, most likely because of a thick cell wall which is recalcitrant to enzymatic removal (Campsall, 1998).

Only one transformant with a wild type phenotype of *N. crassa* was recovered from numerous electroporation and PEG-mediated uptake experiments. Analysis of this transformant confirmed that it had incorporated the *sbr*⁺ allele contained in the 6.6 kb, *Sa*I fragment. The results of Southern blotting experiments and PCR both indicated that complementation of the *sbr* mutant was attributable to non-homologous integration and expression of the wild type allele (Figure 17). The Southern blot of *Sa*I digested DNA revealed a 6.6 kb band in the wild type genome and shorter fragment in the mutant because of the internal *Sa*I site in the *hph*^R construct (Figure 17A). The transformant showed bands of hybridization of both wild type and mutant, indicating the successful introducing of the *Sa*I clone. The same hybridization pattern in both mutant and transformant by probe of *hygromycin B* gene also indicated that both strains have the pCSN43 insertion at the same position and the transformant was recovered from mutant. PCR did not amplify specific DNA from the mutant because the inserted PCSN43 is very large and beyond maximum size of PCR amplification, especially with limited primer extension time. The transformant showed specific amplification as in the wild type and nonspecific amplification as in the mutant. This provided further evidence that expression of the ectopic *sbr* gene resulted in complementation.

Putative transcription activation domains in SBR protein.

Transcription factors contain DNA binding domains and transcriptional activation domains. Glutamine-rich regions which function as transcriptional activation domains have been found in many eukaryotic transcription factors (Kadonaga et al., 1987; review in Mitchell and Tjian, 1989; Bricmont et al., 1991; Yuan et al., 1991; Persengiev et al.,

1995;). The *sbr* gene encodes a helix-turn-helix motif and a possible zinc finger motif with two glutamine-rich regions upstream of both the helix-turn-helix and zinc finger motif (Campsell, 1998, this study).

Regulatory proteins with two glutamine-rich regions have been found in many different species. For example, the human specificity protein 1 (Sp1) is a transcription factor that contains two glutamine rich domains upstream of the zinc fingers of the DNA binding region (Kadonaga et al., 1987). Both of the glutamine-rich domains are required for its synergistic functionality. Deletion of the first glutamine-rich domain at the 5' end reduces its functionality to a level of negligible transcriptional activation (Persengiev et al., 1995). In *Saccharomyces cerevisiae*, the DAL81 gene encodes a specific regulatory factor of the allantoin degradation pathway (Bricmont et al., 1989). It has two regions rich in glutamine-stretches flanking the zinc finger domain respectively. Deletion of the upstream glutamine domain resulted in loss of 50% DAL81 function, whilst deletion of the other glutamine domain did not effect its functionality (Bricmont et al., 1991). In *N. crassa*, the nit-4 gene which is a pathway-specific regulatory gene in the nitrogen circuit also codes for two glutamine-rich regions and a zinc finger domain. However, a protein with a deletion of the most C-terminal glutamine domain was still functional *in vivo* (Yuan et al., 1991). The two glutamine domains in the SBR protein are located, like Sp1 protein, N-terminal to the binding domain. The cAMP response element binding protein (CREB) also has two glutamine-rich activation domains upstream of its binding domain (Sassone-Corsi, 1998).

The high percentage of histidine in both glutamine-rich regions (table 1) is one unique difference between SBR and other glutamine-domain containing proteins

indicated above. The biological significance of high histidine content requires further study. Another feature is the high percentage of proline in the second glutamine-rich region. Activation domains could be classified into three groups: acidic domain (rich in acidic amino acids) such as GAL4, glutamine-rich domain, and proline-rich domain such as AP2, Jun, and Oct-2 (for review, see Mitchell and Tjian, 1989). A single transcription factor such as Oct-2 may have two activation domains of different types. The amino acid composition in an activation domain is associated with its function. For example, proline-rich domains could stimulate transcription when bound at either the proximal (as promoter) or distal (as enhancer) side of the gene, whilst a glutamine-rich domain has no function when bound at distal side (Seipel et al., 1992). It is unknown whether the SBR glutamine-rich region has some feature of proline domain mentioned above.

For further functional assays, different SBR ORFs were constructed with deletions of the first glutamine-rich region (pQE-SBR420) or the downstream sequence immediately from the second glutamine-rich region (pQE-SBR350) (Figure 8).

SBR is a DNA binding protein

A. DNA binding property

Since previous study suggested that *sbr* is a regulatory protein with putative DNA binding domains (Campsell, 1998), its DNA binding properties were tested. Digested, wild type, *N. crassa* genomic DNA was allowed to interact with purified SBR protein. Three genomic fragments bound to SBR were isolated (Figure 11)

The method described in this study to isolate target genomic DNA fragments of SBR protein has been successfully used in other studies in identifying human estrogen

receptor-binding sites and the estrogen-responsive gene (Inoue et al., 1991, 1993). The specificity of this method of isolation was improved by five rounds of selection. Each cycle of binding is a competitive selection of bound DNA fragments cloned from the previous cycle of binding. PCR of a pool of the cloned DNAs during each round of selection confirmed the utility of this approach, i.e. the more cycle of selection, the fewer bands of DNA on gel (data not shown).

The binding specificity was tested for the isolated 0.3kb fragment (SBRBD300) by EMSA. Using unlabeled competitor DNA is the key factor in evaluating the binding specificity (Garner and Revzin, 1981; Fried and Crothers, 1981; Strauss and Varshavsky, 1984). The competitor DNA should not contain any copies of the specific binding sites for the protein. Phage λ DNA, plasmid DNA, *E. coli* genomic DNA or calf thymus DNA have all been used as non-specific competitors (Shanblatt and Revzin, 1984; Varshavsky 1987). Unnatural polynucleotides with simple repeat sequences such as poly d(I-C), poly d(G-C), poly d(A-T) are favored in many EMSA studies, because it increases sensitivity in identifying binding protein from crude nuclear extract (Carthew et al., 1985; Singh et al., 1986). These studies with nuclear extracts used short probes with known DNA sequence to identify cognate binding protein.

However, in this study, the binding site sequence for SBR is unknown and so the pBluescript plasmid alone was used as a competitor because it was the cloning vector and it showed its binding to SBR. Using pBluescript plasmid as competitor has two advantages: first it could verify whether the isolated DNA was a result of vector-SBR binding. Secondly, it ensures that there is competition between the probe and competitor DNA.

The specific DNA binding properties of SBR are demonstrated in figure 12. Competitive EMSA showed no migration of probe DNA when no competitor was added. This is an indication of multiple SBR molecules binding nonspecifically with the probe molecule (Taylor et al., 1994). When competitor was added at 30- to 50-fold, there were still two different protein-DNA complexes. Only one complex was persistent when the competitor was 80-fold or higher. The migration speed of this complex was much slower than the free probe, implying more than one binding site in the 0.3kb probe. Many promoters have multiple binding sites for a regulatory factor. One example in *N. crassa* is the NIT-2 response regulatory sequence in Nit-3 gene. It has multiple NIT-2 binding sites at both directions on the two strands (Fu and Marzluf, 1990a). Previous studies reported that competitors were not able to compete specific protein-DNA bindings until they were present at 25- to 100-fold higher levels, depending on the binding protein and the competitor used.

Oligonucleotide competitor derived from mutation at the binding core sequence competed specific DNA binding of WC-2 protein at 25-fold higher than the probe DNA (Linden and Macino, 1997). *E. coli* genomic DNA could not compete with specific DNA binding of α -protein at 100-fold higher levels (Strauss and Varshavsky, 1984). The binding between estrogen receptor protein and its target fragments isolated with the same method as the present study was not competed by the cloning vector pUC18 at 100-fold (Inoue et al., 1991). With the presence of variable amounts (1-3 μ l) of SBR protein, SBR-DNA binding was not competed by the DNA cloning vector until it was 220-fold higher than the probe DNA. This strongly suggests a specific binding of SBR protein to the sequence of SBRBD300.

The other two larger fragments were not tested for EMSA because large fragment may contribute more nonspecific binding sites and limit the sensitivity of assay and much more competitor DNA is required. After determining the DNA sequence of SBR target site, the specificity of these two fragments will be clear.

B. The putative zinc finger mediates DNA binding?

Zinc finger motifs have been frequently reported as DNA binding domains of transcription factors. The competitive EMSA for mutant SBR protein (PQESBR400) with deletion of the cysteine-rich region and downstream sequence showed no DNA binding when competitor was present even at 25-fold higher concentrations. This suggests that the cysteine-rich region is involved in SBR-DNA binding shown for the wild type SBR protein.

However, this cysteine-rich region does not match any pattern of reported zinc finger motifs (Figure 2B). Different types of multiple-cysteine fingers have been found in transcription factors (reviewed in Evans and Hollenberg, 1988; in Klug and Schwabe, 1995). Thyroid hormone receptors have Cys₄-Cys₅ type fingers (Beato et al., 1995; Mangelsdorf and Evans, 1995). *N. crassa* WC-1 and Nit-2 genes have Cys₄ type finger which is C-X₂-C-X_{17,18}-C-X₂-C (Ballario et al., 1996, Fu and Marzluf, 1990b). Yeast DAL81 and Gal4 genes have Cys₆ type finger C-X₂-C-X₆-C-X_{6,8}-C-X₂-C-X₆-C (Johnston, 1987; Pan and Coleman 1990; Bricmont et al., 1991;). The sequence C-X₂-C-X₉-C-X₇-C-X₂-C-X₂-C-X₈-C in SBR is to some degree similar to the Cys₄ type zinc finger or Cys₆ type finger. The first two C-X₂-C patterns have an intervening 17 amino acids and this is the structure of Cys₄ type finger. The sixth amino acid proline (P)

conserved in the intervening region in many Cys₄ type fingers is also conserved in SBR protein. The conserved tryptophan (W) and glycine (G) are also present in the SBR protein but with a shift of one amino acid position forward and backward respectively (compared with example WC1 in figure 2B). These suggest a possibility of forming a 'finger' holding a zinc ion. If so, this amino acid sequence variation between SBR putative zinc finger and others in the intervening sequence may be responsible for its specific DNA sequence recognition, since in a zinc finger structure this intervening region makes direct contact with the DNA.

Detailed structural analysis has shown that the intervening amino acids in zinc fingers do not form a loop as proposed in the original model (for reviews, see Rhodes and Klug, 1993; Klug and Schwabe, 1995). Rather, the finger region forms a motif consisting of two anti-parallel β -sheets with an adjacent α -helix against one face of the β -sheet (Lee et al., 1989). Upon contact with DNA, the α -helix lies in the major groove of the DNA and makes sequence specific contacts with the bases of DNA, whilst the β -sheet lies further away from the helical axis of the DNA and contacts the DNA backbone.

The importance of amino acids in the α -helix has been confirmed in experiments in which the amino acids at different positions in the zinc finger were randomly altered and their interaction with a wide range of DNA sequence assessed (Choo and Klug, 1994; Rebar and Pabo, 1994). However, the SBR cysteine-rich region was predicted to contain small β -sheet structures and turns but lacking α -helical secondary structures (Campbell 1998). Despite this discrepancy, it is interesting that the cysteine residues in SBR are all clustered together in the same region. Only a single cysteine occurs outside of the region, 100 amino acids downstream of the cysteine-rich region. If this cysteine-

rich motif does contribute to DNA binding, it may represent a new type of structural pattern in DNA contacts. Single amino acid mutation and EMSA could provide more evidence for its DNA binding properties.

It is worth noting that the lack of α -helix in predicted second structure may not be a strong evidence to exclude a zinc finger motif. The DOF transcription factor in plant contains a single zinc finger with sequence C-X₂-C-X₇-C-X₁₃-C-X₂-C (Yanagisawa, 1995). A single mutation in this sequence abolished DNA binding (Yanagisawa and Schmidt, 1999). This sequence also does not match any typical zinc finger motif and somehow is similar to the SBR cysteine-rich sequence. I used the same methods to predict the secondary structure of the protein and no pattern of β -sheet-turn- β -sheet- α -helix was found in Dof either.

Alternatively, the DNA binding ability of SBR is not associated with the cysteine-rich region but other downstream amino acids deleted in PQESBR400. Although the majority of DNA-binding domains which have been identified fall into four classes (helix-loop-helix motif, helix-turn-helix motif, zinc finger motif, and leucine zipper and basic DNA binding domain), some transcription factors, such as AP-2 (activator protein 2, mediates gene activation in response to cAMP) and CTF/NF1 (CCAAT box transcription factor or nuclear factor 1) (reviewed in Mitchell and Tjian, 1989), have DNA binding domains distinct from the above known motifs. In SBR, although no specific motif was found in the region downstream the putative zinc finger motif, it is interesting that a long helical stretch involving 44 amino acids was predicted near the carboxyl terminal (Campbell 1998). Whether this domain is involved in DNA binding remains to be tested. The expression plasmid PQESBR470 has a deletion of 144

amino acids at the C- terminal. This segment of SBR retains the cysteine-rich region and could be used for further testing of the DNA binding ability of the cysteine-rich region.

Target gene prediction and binding site(s) analysis

A. Gene prediction

a) SBRBD300 and its flanking sequence

If SBR is a regulatory protein, a gene should be associated with each target DNA sequence isolated from specific SBR-DNA complexes. Two putative genes have been identified by the MIPS sequencing project in the region flanking the SBRBD300 sequence. These putative “synthetic” genes were derived by gene modeling with computer program developed by BioMax Informatic in Germany (<http://www.biomax.de/>). No gene modeling was available at this site for the complementary strand. In addition, no analysis of gene prediction accuracy were available for the BioMax software.

Therefore, three other bioinformatic programs were also exploited to predict putative genes. Among these programs, GENSCAN and MZEF were reported as being among the top three programs in terms of accuracy (Burset and Guigo, 1996; Singh, 2000).

One of the synthetic genes from the genome sequence provider is at position 36-979 (Figure 13A). It contains three introns and four exons (protein ID:CAB97295.1). NCBI ORF program displayed only one short ORF (111bp) at this region. No gene was predicted in this region by the other three programs. Therefore it is unlikely to be a real gene.

The other synthetic gene, a 333bp ORF located 1.3kb downstream of SBRBD300, overlaps genes predicted by the other three programs (Figure 13A). DNAMAN and NCBI ORF programs show six ORFs from different frames, with all of them clustered in a region from 735-1719. This region harbors the predicted genes from three programs, strongly suggest the existence of a real gene in this region. The transcription start site could be a 30 to 600bp or more (Bruchez et al., 1993a) away from the starting site of the predicted genes. A regulatory binding site could be as far as 1.2kb upstream to a transcription site (Fu and Marzluf, 1990a). Therefore, the SBRBD300 sequence could be in the regulatory region of this possible gene. SBR could bind to a promoter and / or an enhancer since enhancers can be located upstream, downstream, or even within a transcription unit (Hatzopoulous et al., 1988; Muller et al., 1988).

A similar analysis was performed for the complementary sequence (Figure 14). Sequences downstream of SBRBD300 were identified as possible genes by the CGG and MZEF programs only, but both predictions covered only a 564 bp ORF (Figure 14A). ORFs are clustered within the 698 to 2500 nucleotide region for all of the predicted genes, implying the existence of a real gene in this region under the control of the same element(s) in the SBRBD300 sequence.

In summary, SBRBD300 may contain an upstream regulatory element but Northern blotting and cDNA library screening will be required to provide more evidence of potential target genes for SBR.

b) SBRBD600 and SBRBD800

No protein coding sequence in the GenBank database were found to significantly match sequences of the other two isolated fragments SBRBD600 and SBRBD800. ORFs generated from DNAMAN and NCBI ORF programs imply the existence of possible genes, but they are too short for conclusive gene prediction. In the SBRBD600 fragment, +1 and +3 reading frames have relatively long ORFs (Figure 15), whilst the SBRBD800 fragments have relative long ORFs in reading frame +3, -2, and -3 (Figure 16). Genomic clones flanking these fragments are needed for performing additional gene predictions and northern blot assays.

B. SBR target binding sites analysis

If an isolated sequence is derived from the regulatory region of a gene (genes?), DNA sequence analysis should reveal the presence of consensus DNA binding site(s) for various factors.

Explosive amount of sequence information and numerous transcription factors have already been identified. Several bioinformatic programs have been developed to exploit this data and can be used to predict binding sites and identify promoters/enhancers (Frech et al., 1993, 1997a, 1997b; Quandt et al, 1995; Heinemeyer et al., 1998; Lavorgna et al., 1998; Klingenhoff et al., 1999). Transcription factors usually bind to multiple target sequences and regulate multiple genes. Thus, the intrinsic specificity of transcription factors is rather low compared with prokaryotic DNA-binding proteins, probably because the synergistic action of multiple transcription factors on the same promoter may be the strategy for the complex regulation of gene expression (for review, see Werner and Burley, 1997). Therefore, finding potential target sites in a vast

sequence space is a multiple-minimum problem. Many methods have been developed to predict the target sites. Currently, the sequence-based method is the most commonly used method for the target prediction. It relies on sequence information obtained from known binding sequences. Usually, consensus sequence patterns or weight matrices are used to scan the database. Prediction of binding sites using matrix database is usually considered to be of a higher quality. However, there are a limited number of matrices available and many binding sites are not represented by a weight matrix.

The sequences of the three isolated fragments were used to search for potential transcription factors binding sites with the program MatInspector2.2 (see Materials and Methods section). It provides matching scores for both the matrix and the core of the consensus sequence. The core sequence is the binding sequence without exception for different binding assays. The core sequence usually indicates the most critical part for the protein binding. The calculation of matrix similarity includes the information content of individual positions within the matrix (Quandt et al., 1995). For example, if the matrix contains a position where only an A occurs in all training sequences and the test sequence features a T nucleotide at the corresponding position, the overall score of this region will be dramatically reduced.

The consensus of six transcription factors have been found common in the three isolated fragments, with core score 1 (100% match) and matrix score >0.9 (table 2). Nit2, GATA, and STUAP have been reported in fungi (Fu et al., 1990a; Ballario et al., 1996; Dutton et al., 1997; Linden and Macino, 1997). In *N. crassa*, the Nit2 protein and GATA factor proteins such as WC1 and WC2 contain a Cys4 type DNA binding domain. STUAP was reported in *Aspergillus nidulans* (Dutton et al., 1997). It contains a helix-

loop-helix binding domain. DFD factors contain homeodomains encoded by the deformed homeotic genes. Homeodomain was proved containing helix-turn-helix motif (for review, see Treisman et al., 1992). DOF factors have been reported in plant. As discussed earlier, the DOF binding domain contains a specific single zinc finger which has a cysteine in the intervening sequence between zinc ion binding cysteines (two C-X₂-C). Since SBR putative zinc finger seems involving in DNA binding, its binding to any of the these zinc finger target sites is a greater possibility than to that of helix-loop-helix or helix-turn-helix target sites. Secondly, only Dof and AP1 consensus were found at multiple positions on fragment SBRBD300. The EMSA results suggested multiple-molecule binding to this fragment. AP-1 belongs to MADS box family (Karin et al., 1997) to which no similarity was found in SBR (Campbell, 1998). Therefore the target consensus sequence for DOF single zinc finger is the first choice of candidates for SBR binding sequence.

Not only can two factors of the same family of binding domain bind to a consensus sequence, two factors with no similar binding domains could bind to an identical sequence. For example, whilst transcription factors CTF/NF1 and C/EBP both can bind to identical sequence in the CAAT box, they do so via completely different DNA-binding domains. The C/EBP contains the leucine zipper motif with adjacent basic DNA-binding domain common to a number of transcription factors (Graves, et al., 1987), whilst CTF/NF1 contains a distinct DNA-binding motif as mentioned earlier in this discussion.

On the other hand, many transcription factors have the ability to bind to several dissimilar sequences using the same DNA binding domain. For example, the

glucocorticoid receptor binds to a specific DNA sequence in genes which are induced by glucocorticoid and to a distinct DNA sequence in genes which are repressed by this hormone, the same binding domain of the protein being used in each case (Sakai et al., 1988). Similarly, the yeast CYC1 and CYC2 genes contain entirely distinct sequence, both of which bind the HAP1 transcription factor allowing gene activation to occur (Pfeifer et al., 1987). Above complexity and the limited number of matrices available lead to difficulty in identifying the accurate SBR target site sequence.

Finding multiple consensus sites in each isolated fragment also favors the suggestion that each fragment contains regulatory response sequence. A single transcription factor binding site often does not encode a transcription function of its own (just binding of the protein) because biologic function usually requires cooperation of two or more transcription factors (Jones et al., 1988; Carey, 1998). In fact, other consensus sites with high matching scores were also found in each fragment. Ultimately, concrete evidence, such as DNA footprinting will be required to confirm any of these speculations of possible SBR binding sites.

Conclusion

Morphological mutation caused by insertion of plasmid pCSN43 into the *sbr* gene has been demonstrated by complementation and confirmed by RT-PCR. The *sbr* gene is located on linkage group III near *trp-1* and *con-7* genes. It codes for a protein of 612 amino acids with two glutamine-rich regions which may function as activation domains in regulating gene transcription.

The following results support the hypothesis that SBR is a regulatory protein: 1) three genomic fragments bound to SBR were cloned from SBR-DNA complexes, 2) EMSA results indicated specific SBR-DNA binding which is reduced by deletion of the cysteine-rich putative zinc finger, and 3) gene predictions and binding sites analysis suggest the isolated fragments contain multiple potential regulatory elements.

It is necessary to repeat EMSA with shorter DNA (for example, 30-50 bp) and other types of competitor DNA. In addition, DNase I footprinting with short DNA fragments is needed to further confirm DNA binding and clarify the binding site sequence. Confirmation that SBR is a regulatory factor will ultimately depend on a demonstration that some genes at or near the target sites respond to SBR activation or repression.

Several of the SBR constructs described here should prove useful in future studies of the role of *sbr* in morphogenesis.

Reference

- Ballario, P., P. Vittorioso, A. Magrelli, C. Talora, A. Cabibbo, and G. Macino (1996). White collar-1, a central regulator of blue light responses in *Neurospora*, is a zinc finger protein. *EMBO J.* 15:1650-1657.
- Beato, M., P. Herrlich and G. Schutz (1995) Steroid hormone receptors: many actors in search of a plot. *Cell* 83:852-857.
- Biswas, E.E., P.H. Chen, and S.B. Biswas (1995) Overexpression and rapid purification of biologically active yeast proliferating cell nuclear antigen. *Protein Expr. Purif.* 6(6):763-770.
- Brennan, R.D. and B.W. Matthews (1989) The helix-turn-helix DNA binding motif. *J. Bio. Chem.* 264:1903-1906.
- Bricmont P.A., and T.G. Cooper (1989) A gene product needed for induction of allantoin system of genes in *Saccharomyces cerevisiae* but not for their transcriptional activation. *Mol. Cell. Biol.* 9:3869-3877.
- Bricmont P.A., J.R. Daugherty, and T.G. Cooper (1991) The DAL81 gene product is required for induced expression of two differently regulated nitrogen catabolic genes in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* 11:1161-1166.
- Bruchez, J.J.P., J. Eberle and V.E.A. Russo (1993a) Regulatory sequences involved in translation of *Neurospora crassa* mRNA: Kozak sequence and stop codons. *Fungal Genetic Newsletter* 40:85-88.
- Bruchez, J.J.P., J. Eberle and V.E.A. Russo (1993b) Regulatory sequences in the transcription of *Neurospora crassa* genes: CAAT box, TATA box, introns, poly(A) tail formation sequences. *Fungal Genetic Newsletter* 40:89-96.
- Burset, M. and R. Guigo (1996) Evaluation of gene structure prediction programs. *Genomics* 34:353-367.
- Campsall, K.D. (1998) Cloning and characterization of *sbr*: a new colonial mutant of *Neurospora crassa*. Thesis of Master's Degree. Carleton University. Ottawa, Canada.
- Carey, M. (1998) Enhanceosome and transcriptional synergy. *Cell.* 92: 5-8.
- Carthew, R. W., L. A. Chodosh, and P.A. Sharp (1985) An RNA polymerase II transcription factor binds to an upstream element in the adenovirus major late promoter. *Cell.* 43:439-448.

- Caubin, J., T. Iglesias, J. Bernal, A. Munoz, G. Marquez, J.L. Barbero and A. Zaballos (1994) Isolation of genomic DNA fragments corresponding to genes modulated *in vivo* by transcription factor. *Nucleic Acids Res.* 22: 4132-4138.
- Choo, Y. and A. Klug (1994) Toward a code for the interaction of zinc fingers with DNA: selection of randomized fingers displayed on phage. *Proc. Natl. Acad. Sci. USA* 91:11163-11167.
- Clerc, R.G., L.M. Corcoran, J.H. LeBowitz, D. Baltimore, and P.A. Sharp (1988) The B-cell specific Oct-2 protein contains POU box and homeo box type domains. *Gene Dev.* 2: 1570-1581.
- Collinge, A.J. and A.P.J. Trinci (1974) Hyphal tips of wild type and spreading colonial mutants of *Neurospora crassa*. *Arch. Microbiol.* 99:353-368.
- Dabrowski S., and J. Kur (1999) Cloning, overexpression, and purification of the recombinant His-tagged SSB protein of *Escherichia coli* and use in polymerase chain reaction amplification. *Protein Expr. Purif.* 16: 96-102.
- Davis R.H. (2000) *Neurospora*. Contribution of a model organism. Oxford University press. New York.
- de Terra, N. and E. L. Tatum (1961) Colonial growth of *Neurospora*. *Science* 134: 1066-1068
- de Terra, N. and E. L. Tatum (1963) A relationship between cell wall structure and colonial growth in *Neurospora crassa*. *Amer. Bot.* 50:669-677.
- Drew, H. R. (1984) Structural specificities of five commonly used DNA nucleases. *J. Mol. Biol.* 176: 535-557.
- Dutton, J.R., S. Johns, B.L. Miller (1997) StuAp is a sequence-specific transcription factor that relulates developmental complexity in *Aspergillus nidulans*. *EMBO J.* 16: 5710-5721.
- Ekker, S. C., D. P. von Kessler and P. A. Beachy (1992) Differential DNA sequence recognition is a determinant of specificity in homeotic gene action. *EMBO J.* 11: 4059-4072.
- Evans, R.M. and S.M. Hollenberg (1988) Zinc fingers: gilt by association. *Cell.* 52:1-3.

Facklam T.J., and G.A.Marzluf (1978) Nitrogen regulation of amino acid catabolism in *Neurospora crassa*. *Biochem Genet* 16:343-54.

Feinberg, A.P., and B. Vogelstein (1984) "A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity". Addendum. *Anal. Biochem.* 137(1):266-7

Frech, K., G. Herrmann, and T. Werner (1993) Computer-assisted prediction, classification, and delimitation of protein binding sites in nucleic acids. *Nucleic Acids Res.* 21:1655-1664.

Frech, K., J. Danescu-Mayer, and T. Werner (1997a) A novel method to develop highly specific models for regulatory units detects a new LTR in GenBank which contains a functional promoter. *J. Mol. Biol.* 270: 674-687.

Frech, K., K. Quandt, and T. Werner (1997b) Finding protein-binding sites in DNA sequence: the next generation. *Trends Biochem. Sci.* 22:103-104.

Fried, M. (1989) Measurement of protein-DNA interaction parameters by electrophoresis mobility shift assay. *Electrophoresis* 10:366-376.

Fried M. and D.M. Crothers (1981) Equilibria and kinetics of lac repressor-operator interactions by polyacrylamide gel electrophoresis. *Nucleic Acids Res.* 9:6065-6525.

Frohman M. A. (1990) RACE: rapid amplification of cDNA ends. In *PCR Protocols. A guide to methods and applications.* Academic press, Inc. San Diego. 28-38.

Fu, Y.-H. and G. A. Marzluf (1990a) *nit-2*, the major positive-acting nitrogen regulatory gene of *Neurospora crassa*, encodes a sequence-specific DNA-binding protein. *Proc. Natl. Acad. Sci. USA* 87:5331-5335.

Fu, Y.-H. and G. A. Marzluf (1990b) *nit-2*, the major positive-acting nitrogen regulatory gene of *Neurospora crassa*, encodes a protein with a putative zinc finger DNA-binding domain. *Mol. Cell Biol.* 10:1056-1065.

Garner, M. and A. Revzin (1981) A gel electrophoresis method for quantifying the binding of proteins to specific region: application to components of the *Escherichia coli* lactose operon regulatory system. *Nucleic Acids Res.* 9:3047-3060.

Graves, B.J., P.F. Johnson, and S. L. McKnight (1987) Homologous recognition of a promoter domain common to the MSV, LTR, and the HSV tk gene. *Cell* 44: 565-576.

Gruber, C.A., J.M. Rhee, A. Gleiberman, and E.E. Turner (1997) POU domain factors of the Brn-3 class recognize functional DNA elements which are distinctive, symmetrical and highly conserved in evolution. *Mol. Cell. Biol.* 17: 2391-2400.

Haro, R., L. Sainz, F. Rubio, and A. Rodriguez-Navarro (1999) Cloning of two genes encoding potassium transporters in *Neurospora crassa* and expression of the corresponding cDNAs in *Saccharomyces cerevisiae*. *Mol. Microbiol.* 31:511-520.

Harrison, S. C. and A.K. Aggarwal (1990) DNA recognition by proteins with the helix-turn-helix motif. *Ann. Rev. Biochem.* 59:933-969.

Hatzopoulous, A.K., U. Schlokot, and P. Gruss (1988) Enhancers and other cis-acting sequence. In: *Transcription and Splicing* (Hames, B.D. and Glover, D.M. eds), pp 43-96. Oxford. IRL Press.

Heinemeyer, T., E. Wingender, I. Reuter, H. Hermjakob, A.E. Kel, O.V. Kel, E.V. Ignateva, E.A. Ananko, O.A. Podkolodnaya, F.A. Kolpakov, N.L. Podkolodny, and N.A. Kolchanov (1998). Database on transcriptional regulation: TRANSFAC, TRRD and COMPEL. *Nucleic Acids Res.* 26:362-367.

Hoculi, E. H. Dobeli, and A. Schacher (1987) New metal chelate adsorbent selective for proteins and peptides containing neighbouring histidine residues. *J. Chromatog.* 411:177-184.

Horvath, C.M., Z. Wen, and J.E. Jr Darnell (1995) A STAT protein domain that determines DNA sequence recognition suggests a novel DNA-binding domain. *Genes Devel.* 9:984-994.

Hu Y.F., R. Luscher, A. Admon, N. Mermod, and R. Tjian (1990) Transcription factor AP-4 contains multiple dimerization domains that regulate dimer specificity. *Genes Devel.* 4(10):1741-1752.

Inoue, S., A. Orimo, T. Hosoi, S. Kondo, H. Toyoshima, T. Kondo, A. Ikegami, T. Ouchi, H. Orima, and M. Muramatsu (1993) Genomic binding-site cloning reveals an estrogen-responsive gene that encodes a RING finger protein. *Proc. Natl. Acad. Sci. USA.* 90: 11117-11121.

Inoue, S., S. Kondo, M. Hashimoto, T. Kondo, and M. Muramatsu (1991) Isolation of estrogen receptor-binding sites in human genomic DNA. *Nucleic Acids Res.* 19:4091-4096.

Johnston, M. (1987) A model fungal gene regulatory mechanism: the GAL genes of *Saccharomyces cerevisiae*. *Microbiol. Rev.* 51:458-476.

- Jones, N.C., P. W. J. Rigby, and E. B. Ziff (1988) Trans-acting protein factors and the regulation of eukaryotic transcription. *Genes Devel.* 2:267-281.
- Kadonaga, J.T., K.R. Carner, F.R. Masiarz, and R. Tjian. (1987) Isolation of cDNA encoding transcription factor Sp1 and functional analysis of the DNA binding domain. *Cell* 51:1079-1090.
- Karin, M., Z.G. Liu, and E. Zandi (1997) AP-1 function and regulation. *Curr. Opinion in Cell Biol.* 9:240-246.
- Kato, K., Y. Makino, T. Kishimoto, J. Yamauchi, S. Kato, M. Muramatsu, and T. Tamura (1994) Multimerization of the mouse TATA-binding protein (TBP) driven by its C-terminal conserved domain. *Nucleic Acids Res.* 22:1179-1185.
- Kinzler, K., and B. Vogelstein (1989) Whole genome PCR: application to the identification of sequences bound by gene regulatory proteins. *Nucleic Acid. Res.* 17: 3645-3653.
- Klingenhoff, A, K. Frech, K. Quandt and T. Werner (1999) Functional promoter modules can be detected by formal models independent of overall nucleotide sequence similarity. *Bioinformatics* 15:180-186.
- Klug, A. and J.R. Schwabe (1995) Zinger fingers. *FASEB J.* 9:597-604.
- Ko, L.J., and J.D. Engel (1993) DNA-binding specificities of the GATA transcription factor family. *Mol. Cell. Biol.* 13:4011-4022.
- Kornberg, T.B. (1993) Understanding the homoedomain. *J. Bio. Chem.* 268: 26813-26816.
- Krumlauf, R and G. A. Marzluf (1980) Genome organization and characterization of repetitive and inverted repeat DNA sequences in *Neurospora crassa*. *J. Bio. Chem.* 255: 1138-1145.
- Latchman, D.S. (1998) Eukaryotic transcription factors. Academic Press. San Diego.
- Lavorgna, G, E. Boncinelli, A. Wagner, and T. Werner (1998) Detection of potential target genes in silico? *Trends Genet.* 14(9): 375-376
- Lee, M.S., G.P. Gippert, K.V. Soman, D.A. Case, and P.E. Wright (1989) Three-dimensional solution structure of a single zinc finger DNA binding domain. *Science* 245: 635-637.

Linden, H. and G. Macino (1997) White collar 2, a partner in blue-light signal transduction, controlling expression of light-regulated genes in *Neurospora crassa*. EMBO J. 16:98-109.

Lowry JA, and W.R. Atchley (2000) Molecular evolution of the GATA family of transcription factors: conservation within the DNA-binding domain. J. Mol. Evol. 50(2):103-115

Mangelsdorf, D.J. and R.M. Evans (1995) The RXR heterodimers and orphan receptors. Cell 83: 841-850.

Melo, A.M., M. Duarte, and A. Videira (1999) Primary structure and characterisation of a 64 kDa NADH dehydrogenase from the inner membrane of *Neurospora crassa* mitochondria. Biochim. Biophys. Acta 1412 (3):282-287.

Metzenberg, R.L., J. N. Stevens, E.U. Selker, and E. Morzycka-Wroblewska (1984) A method for finding the genetic map position of cloned fragments. Fungal Genetic Newsletter. 39:50-58.

Metzenberg, R.L. and J. Grotelueschen (1992) Restriction polymorphism maps of *Neurospora crassa*: update. Neurospora Newsletter. 47: 35-39.

Mishra, N.C. (1977) Genetics and biochemistry of morphogenesis in *Neurospora*. Adv. Genet. 19: 341-405.

Mitchell, P.J. and R. Tjian (1989) Transcriptional regulation in mammalian cells by sequence specific DNA binding proteins. Science 245: 371-378.

Molloy, P.L. (2000) Electrophoretic mobility shift assay. In Methods in Molecular Biology. Vol. 130: transcription factor protocols. Edited by M.J. Tymms. Humana Press Inc. Totowa, New Jersey. 235-246.

Muller, M.M., T. Gerster, and W. Schaffner (1988) Enhancer sequences and the regulation of gene transcription. Eur. J. Biochem. 176:485-495.

Paietta, J., and G.A. Marzluf (1985) Gene disruption by transformation in *Neurospora crassa*. Mol. Cell. Biol. 5:1554-1559.

Paietta, J., and M.L. Sargent (1983) Isolation and characterization of light insensitive mutants of *Neurospora crassa*. Genetics 164:11-21.

Pall, M. L. and J.P. Brunelli (1994) New plasmid and λ plasmid hybrid vectors and a *Neurospora crassa* genomic library containing the bar selectable marker and the Crelox site-specific recombination system for use in filamentous fungi. Fungal Genetics Newsletter 41: 63-65.

Pan, T., and J.E. Coleman (1990) GAL4 transcription factor is not a "zinc finger" but forms a $\text{Zn(II)}_2\text{Cys}_6$ binuclear cluster. *Proc. Natl. Acad. Sci. USA.* 87:2077-2081.

Perkins A. S., R. Fishel, N.A. Jenkins, and N.G. Copeland (1991) Evi-1, a murine zinc finger proto-oncogene, encodes a sequence-specific DNA-binding protein. *Mol. Cell. Biol.* 11:2665-2674.

Persengiev, S. P., J. D. Saffer, and D.L. Kilpatrick (1995) An alternatively spliced form of the transcription factor Sp1 containing only a single glutamine-rich transactivation domain. *Proc. Natl. Acad. Sci. USA.* 92:9107-9111.

Pfeifer, K., I. Prezant, and L. Guarente (1987) Yeast HAP1 activator binds to two upstream sites of different sequence. *Cell* 49:19-27.

Pollock, R. and R. Treisman (1990) A sensitive method for the determination of protein-DNA binding specificities. *Nucleic Acids Res.* 18:6197-6204.

Porath, J. (1992) Immobilised metal ion affinity chromatography. *Prot. Express. Purif.* 3:263-281.

Porath J., J. Carlsson, I. Olsson, and G. Belfrage (1975) Metal chelate affinity chromatography, a new approach to protein fractionation. *Nature* 258:598-599.

Quandt, K., K. Frech, H. Karas, E. Wingender, and T. Werner (1995) MatInd and MatInspector: new fast and versatile tools for detection of consensus matches in nucleotide sequence data. *Nucleic Acids Res.* 23: 4878-4884.

Rebar, E.J. and C.O. Pabo (1994) Zinc finger phage: affinity selection of fingers with new DNA-binding specificities. *Science* 263:671-673.

Revzin, A. (1987) Gel electrophoresis assays for DNA-protein interactions. *Bio. Tech.* 7: 346-355.

Rhodes, D., and A. Klug (1993) Zinc finger structure. *Scientific American* 268: 32-39.

Robb, M.J., M.A. Wilson, and P.J. Vierula (1995) A fungal acin-related protein involved in nuclear migration. *Mol. Gen. Genet.* 247:583-590.

Russo, V.E.A., and N.N. Pandit (1992) Development in *Neurospora crassa*. In *Development; the molecular genetic approach*. Springer-Verlag. Berlin. 88-102.

Sakai, D.D., S. Helms, J. Carlstedt-Duke, J.A. Gustafsson, F.M. Rottman, and K.R. Yamamoto. (1988) Hormone-mediated repression: a negative glucocorticoid

response element from the bovine prolactin gene. *Genes and Development*. 2:1144-1154.

Sambrook, J., E.F. Fritsch, and T. Maniatis (1989) *Molecular cloning: a laboratory manual*, 2nd edition. Cold Spring Laboratory Press. Cold Spring Harbor.

Sanchez-Garcia, I., and T.H. Rabbitts (1994) The LIM domain: a new structure motif found in zinc-finger proteins. *Trends Genet.* 10:315-320.

Sassone-Corsi, P. (1998) Transcription factors responsive to cyclic ATP. *Annu. Rev. Cell Dev. Biol* 30:167-170..

Scazzocchio, C. (2000) The fungal GATA factors. *Current Opin. Microbiol.* 3:126-131.

Schmitz, A. and D.J. Galas (1978) DNase I footprinting: a simple method for the detection of protein-DNA binding specificity. *Nucleic Acids Res.* 5: 3157-3170.

Seipel, K. O. Georgiev, and W. Schaffner (1992) Different activation domains stimulate transcription from the remote ('enhancer') and proximal ('promoter') position. *EMBO J.* 11:4961-4968.

Shanblatt, S. and A. Revzin (1984) Kinetics of RNA polymerase promoter complex formation: effects of nonspecific DNA-protein interactions. *Nucleic Acids Res.* 12: 5287-5306.

Shear, C.L., and B.O. Dodge (1972) Life histories and heterothallism of the red bread-mold fungi of the *Molilia sitophila* group. *J. Agricultural Res.* 34:1019-1042.

Singh, G. B. (2000) Computational approaches for gene identification. In *Methods in Molecular Biology*. Vol 132: Bioinformatics methods and Protocols. Edited by S. Misener and S. A Krawetz. Humana Press Inc. Totowa, N.J. page: 351-364.

Singh, H., R. Sen, D. Baltimore and P. A. Sharp (1986) A nuclear factor that binds to a conserved sequence motif in transcriptional control elements of immunoglobulin genes. *Nature* 319 (6049):154-158.

Solscheid, B, and M. Tropschug (2000) A novel type of FKBP in the secretory pathway of *Neurospora crassa*. *FEBS Lett.* 480:118-22.

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

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

Strauss, F and A. Varshavsky (1984) A protein binds to satellite DNA repeat at three specific sites that would be brought into mutual proximity by DNA folding in the nucleosome. *Cell* 37:889-901.

Staben, C., B. Jensen, M. Singer, J. Pollock, M. Schechtman, J. Kinsey, and E. Selker (1989) Use of bacterial hygromycin B resistance gene as a dominant selectable marker in *Neurospora crassa* transformation. *Fungal Genetics newsletter*, 36: 79-81.

Sturm, R., T. Baumruker, R.Jr. Franza, and W. Herr (1987) A 100-kD Hela cell octamer binding protein (OBP) interacts differently with two separate octamer-related sequence within the SV40 enhancer. *Genes Devel.* 1: 1147-1160.

Szoor, B., Z.S. Feher, T. Zeke, P. Gergely, E. Yatzkan, O. Yarden and V. Dombradi (1998) pzl-1 encodes a novel protein phosphatase-Z-like Ser/Thr protein phosphatase in *Neurospora crassa*. *Biochim. Biophys. Acta* 1388 (1): 260-266.

Taylor, J. D., A. J Achroyd, and S. E. Halford (1994) The gel shift assay for the analysis of DNA-protein interactions. In *Method in Molecular Biology*. Vol 30: DNA-protein interactions: principle and protocols. 263-279.

Todd, R.B. and A. Andrianopoulos (1997) Evolution of a fungal regulatory gene family: the Zn(II)₂Cys₆ binuclear cluster DNA binding motif. *Fungal Genetics and Biology*. 21: 388-405.

Treisman, J., E. Harris, D. Wilson, and C. Desplan (1992) The homeodomain; a new face for the helix-turn-helix. *Bio. Essays* 14: 145-150.

Trinci, A.P.J. (1978) Wall and hyphal growth. *Scientific Progress, Oxford* 65:75-99.

Tullius, T. D., B. A. Dombroski, M.E. A. Churchill, and L. Kam (1987) Hydroxyl radical footprinting: a high resolution method for mapping protein-DNA contacts. *Meth. Enzymol.* 155: 537-558.

Vaquero, A, M.L., Espinas, F. Azorin, and J. Bernues (2000) Functional mapping of the GAGA factor assigns its transcriptional activity to the C-terminal glutamine-rich domain. *J. Biol. Chem.* 275:19461-8.

Varshavsky, A. (1987) Electrophoretic assay for DNA-binding proteins. *Meth. Enzymol.* 151:551-565.