Analysis of gene expression in *Aspergillus*

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad van doctor aan de Universiteit van Amsterdam,
op gezag van de Rector Magnificus Prof. Dr. J.J.M. Franse
ten overstaan van een door het college van dekanen ingestelde commissie in het
openbaar te verdedigen in de Aula der Universiteit

op donderdag 30 januari 1997
te 15.00 uur

door

Robertus Franciscus Maria van Gorcom

Geboren te Rijswijk ZH

Promotor:	Prof. Dr. P.H. Pouwels
Co-promotor:	Dr. C.A.M.J.J. van den Hondel
Leden:	Prof. Dr. B.J.C. Cornelissen Prof. Dr. K. van Dam Prof. Dr. R. van Driel Prof. Dr. W. Harder Prof. Dr. R.F. Hoekstra Dr. P.J. Punt

The research described in this thesis was carried out in the department of Molecular Genetics and Gene Technology of the former Medical Biological Laboratory TNO and the TNO Nutrition and Food Research Institute, under supervision of Dr. C.A.M.J.J. van den Hondel.

Publication of this thesis is financially supported by TNO Nutrition and Food Research Institute and the University of Amsterdam.

Eburon Publishers, the Netherlands
ISBN 90-5166-559-8

Contents

CHAPTER 1

Introduction

Filamentous Fungi

Filamentous fungi are eukaryotic microorganisms which are found almost everywhere in nature. These multicellular microorganisms grow in large branching filaments and are characterized by the presence of multiple nuclei per cell. Cells are not completely separated from each other but are in contact via large pores present in the separation walls, the septa. For reproduction and survival purposes filamentous fungi are able to form differentiated structures which are involved in the production of spores (several species have different structures for a-sexual and for sexual spores). Together with yeasts, filamentous fungi form a separate kingdom, called Mycota.

The main role of filamentous fungi in nature is the degradation of all kinds of complex natural polymeric substances, e.g. materials present in higher organisms. They play a major role in the mineralization of wood. Whereas hardly any living organism is able to degrade wood material, filamentous fungi are equipped with a large spectrum of enzymes whose specific function is to degrade the chemically complex material that is characteristic for wood. Therefore their main habitat is on dead plant material and in the soil, but specific fungi are also found in aquatic habitats and even in air. Apart from their role as biodegraders, several fungi are well known as plant-, insect- and animal pathogens. Plant pathogenic fungi are a major pest for several important crops and form a serious economic threat in agriculture. Consequently, enormous amounts of agrochemicals are used everywhere in the world to reduce the damage that fungal infections cause to the crop. Filamentous fungi which are pathogenic for insects are becoming popular as a biological weapon in the treatment of pests in agriculture caused by insects. Also for the human body some filamentous fungal species can be pathogenic but normally they do not cause serious problems. However, for immunocompromised people, such as patients in a post-transplantation period and HIV infected patients fungal pathogens are a major cause of death.

Industrial application of fungi

Filamentous fungi have already been used by men for several thousand years for the preparation and conservation of foods, especially in Asian countries. Fermentation by fungi of several agricultural products results in better digestible, more tasty and better conservable food products. In the western countries, some species are used for the preparation of specific types of cheese. More recently, industry has started to exploit the potential of fungi to secrete large amounts of products. To enable their role as biodegraders fungi secrete a very large number of substances that degrade or help to degrade the biopolymers. Some of these substances, primary metabolites such as citric and oxalic acid, help the fungus to enter the dead material and create an environment in which the next step(s) of the degradation can be performed optimally. These and other primary metabolites are currently produced with fungi, very efficiently and pure, at a very large scale and are used in many different kinds of food and non-food applications and in the chemical industry.

In the biodegradation processes in particular enzymes play a major role. Fungi secrete many enzymes which are each specialized in the degradation of a specific type of biopolymer (or even a specific bond in a biopolymer). The diversity of biopolymers and the complexity of some biopolymers is matched by a very large spectrum of enzymes that are produced by filamentous

fungi. Over the past 40 years fungal strains, which overproduce one specific enzyme or a set of enzymes, have been selected and used. These enzymes are sold for many food and non-food applications. The emerge of the recombinant DNA technology has even further expanded the range of fungal enzymes that can be produced in a safe and efficient way.
The third group of interesting compounds which are secreted by fungi are secondary metabolites. Some of these compounds are produced by fungi for their defense against competing (micro)-organisms, others are used in a pathogenicity process or for other purposes. Well-known representatives of this class of products are the penicillins and cephalosporins, which have found profound medical application.
Over the past five decades several industries have been set up to exploit the capabilities of fungi for the production of these products for commercial purposes. Therefore, large plants have been built all over the world for the production of organic acids, enzymes, antibiotics etc. Until the beginning of the 1980's, the very impressive production improvements which have been realized by these industries were accomplished only by classical genetics. With the emergence of the recombinant DNA technology new possibilities arose and consequently new technology specific for these filamentous fungi had to be developed to further increase their industrial potential (for a review in the industrial use of fungi see Finkelstein and Ball, 1992).

Molecular genetic research on filamentous fungi

Besides the needs from the enzyme and metabolite producing industries, there was another important incentive to extend the genetic research with filamentous fungi that already started in the beginning of this century with the new tools of molecular biology and molecular genetics. Filamentous fungi are eukaryotic organisms that are relatively easy to grow on simple (liquid and solidified) media and for which good genetic methodology and knowledge is available, but in contrast to most yeasts, filamentous fungi form several kinds of specialized differentiated structures thus allowing the study of mechanisms of differentiation. Differentiation is also studied in many higher organisms. But the advantage of filamentous fungi is that also genetic research can be included.
Consequently at several universities and research institutes and at some industrial laboratories research aimed at different aspects of the development of molecular genetic tools was initiated.

Analysis of gene expression

When the first basal tools for the genetic modification of filamentous fungi had been worked out it became also necessary to develop techniques that allowed the quantification of expression of individual genes in a more standardized way. Even when enzymatic assays of the products of the genes to be analyzed are available and easy to perform, it is difficult to readily compare expression levels of different genes. In bacteria, yeasts and also in animal cells, for which molecular genetic methodology had been developed before, (in most cases bacterial) reporter genes had shown their usefulness. Therefore we tried to adapt this technology for use in filamentous fungi. At the time we started this research in 1983, the *Escherichia coli*

whether bacterial genes could be expressed in filamentous fungi, the *lacZ* gene had been successfully expressed in yeast and cells of higher organisms. Therefore, we started to construct *lacZ* gene fusions with one of the few fungal genes available at that moment, the *Aspergillus nidulans trpC* gene. The results of those studies, which are presented in chapter 2, were published in 1985.
The transformation technology as it was developed in filamentous fungi forced us to work out a different strategy compared to the systems set up for other organisms. Stable transformation in filamentous fungi could only be obtained by integration of donor DNA into the genome of the recipient strain. In our first study (chapter 2) we found that both the copy number of the integrated fusion gene and the site of integration into the genome strongly influenced the results obtained. Consequently, a proper quantitative comparison of the reporter gene expression levels driven by different gene expression signals seemed impossible. Therefore, we designed a strategy which allowed us to obtain, with a high efficiency, transformants that contain only one copy of the reporter cassette integrated in the same orientation at a specific site in the genome. This strategy indeed resulted in transformants whose reporter gene expression could be used to quantify and compare gene expression driven by the expression control regions of different genes. This work, described in chapters 3 and 4, was published in 1986 and 1988.

Different types of reporter genes

At the time we made the choice for the reporter gene to be used in our studies there was not much to choose from. Although several genes were used in *Escherichia coli* as a reporter gene, the only gene that was used successfully in several prokaryotic and eukaryotic organisms and for which a convenient activity assay was available, was the *Escherichia coli* β-galactosidase gene (*lacZ*). For this enzyme, the biochemistry of which had been extensively studied, several chromogenic substrates were available, which allowed a very easy *in vitro* quantification of the amount of enzyme present in an extract. In the meantime we also developed a qualitative *in vivo* assay. In cases where the amount of enzyme was very low also substrates which allowed a very sensitive qualitative and quantitative comparison were available.
Since the time we made our choice for the *lacZ* gene, several other reporter genes have been studied. However, apart from the *E. coli uidA* gene, encoding the enzyme β-glucuronidase, most genes have disadvantages that are not encountered with the *lacZ* reporter gene. β-Galactosidase and β-glucuronidase do not need additional cellular components such as ATP or NAD(P)H to become enzymatically active, this in contrast to other reporter genes such as luciferase. Furthermore, both enzymes are relatively stable under a variety of conditions, can be assayed with a simple spectrophotometer and show an activity that is proportional to the amount of enzyme over a large range of concentrations.
Recently the use of fluorescent reporter proteins has attracted broad attention. The advantage of these new reporters is that their expression can be measured without disruption of the cell, allowing the quantification of cell (or even organelle) specific gene expression. The first results with the Green Fluorescent Protein (GFP) from the jellyfish *Aequorea victoria* in filamentous fungi were reported (Bottin et al. 1996) but in some organisms expression problems occur (Kassow,

Punt and van den Hondel, personal communication). However, it seems likely that improved, synthetic versions of the GFP gene (Cormack et al. 1996; Reichel et al 1996) and/or other fluorescent protein encoding genes will be attractive additions to the spectrum of reporter genes available for use in filamentous fungi.

Introduction to this thesis

This thesis describes in chapters 2, 3 and 4 the development of an expression analysis system for the filamentous fungi *Aspergillus nidulans* and *Aspergillus niger* based on the *E. coli lacZ* gene, enabling quantification and comparison of gene expression driven by expression signals from different genes. Vectors have been constructed that allow in phase fusion of the 5' regions of genes of interest (including the expression control sequences, the transcription start region, the sequences encoding the 5' untranslated region of the mRNA and the sequences encoding the start of translation) to the β-galactosidase encoding region.

Stable transformants in *Aspergillus* are usually obtained by integration of the exogenous DNA into one of the chromosomes of the recipient. To enable comparison of different expression control regions in fungal transformants, without the interference of effects due to different chromosomal integration sites, the expression analysis vectors were targeted to a specific locus of the chromosome.

Chapter 5 of this thesis describes the isolation and characterization of a gene encoding the enzyme that enables *Aspergillus niger* to grow on benzoate as the sole carbon source by activating the aromatic ring via hydroxylation at the para position. This benzoate para-hydroxylase gene (*bphA*) was shown to be a member of the large cytochrome P450 superfamily and classified as *cyp53A1* (Nelson et al, 1996). The messenger RNA of this gene, that is formed after induction by benzoate, contains two small open reading frames and an intron in the region preceding the structural gene. Both frames stop after only a few amino acid codons far upstream from the start of the cytochrome P450 encoding open reading frame. The analysis system described in the first part of this thesis has been used in experiments described in chapter 6, to analyze the possible effects of these two small open reading frames present in the *A. niger bphA* gene, on the level and/or regulation of its expression.

The presence of such upstream open reading frames in fungal genes is not often reported in the literature but we noticed that their presence has not always been recognized or mentioned by authors describing new DNA sequences of fungal genes. Therefore, we have performed a screening of all gene sequences available from species of the genus *Aspergillus* for the presence of such upstream open reading frames. The results of this screening and some further analysis of their characteristics are presented in chapter 7.

At the end of this thesis the reader will find summaries presented both in the English and Dutch language.

REFERENCES

Bottin, A., J. Kämper and R. Kahmann (1996) Characterization of a carbon source-regulated gene from *Ustilago maydis*. *Mol. Gen. Genet.*, in press.

Cormack, B.P., R.H. Valdivia and S. Falkow (1996) FACS-optimized mutants of the green fluorescent protein (GFP). *Gene*, in press.

Finkelstein, D.B. and C. Ball (Eds.) Biotechnology of filamentous fungi. *Biotechnology Series* **21**, (1992) pp 1-503, Buterworth-Heinemann, Boston, USA.

Nelson, D.R., L. Koymans, T. Kamataki, J.J. Stegeman, R. Feyereisen, D.J. Waxman, M.R. Waterman, O. Gotoh, M.J. Coon, R.W. Estabrook, I.C. Gunsalus and D.W. Nebert (1996) P450 superfamily: update on new sequences, gene mapping, accession numbers and nomenclature. *Pharmacogenetics* **6**, 1-42.

Reichel, C., J. Mathur, P. Eckes, K. Langenkemper, C. Koncz, J. Schell, B. Reiss and C. Maas (1996) Enhanced green fluorescence by the expression of an *Aequorea victoria* green fluorescent protein mutant in mono- and dicotyledonous plant cells. *Proc. Natl. Acad. Sci. USA*, **93**: 5888-5893.

CHAPTER 2

Expression of an *Escherichia coli* β-galactosidase fusion gene in *Aspergillus nidulans*

Robert F.M. van Gorcom, Peter H. Pouwels, Theo Goosen, Jaap Visser, Henk W.J. van den Broek, John E. Hamer, William E. Timberlake and Cees A.M.J.J. van den Hondel

Gene (1985) **40**, 99-106

GENE 1462

Expression of an *Escherichia coli* β-galactosidase fusion gene in *Aspergillus nidulans*

(Recombinant DNA; *trpC-lacZ* fusion; eukaryotic promoter; transformation)

Robert F.M. van Gorcom[a], Peter H. Pouwels[a], Theo Goosen[b], Jaap Visser[b], Henk W.J. van den Broek[b], John E. Hamer[c], William E. Timberlake[c] and Cees A.M.J.J. van den Hondel[a]*

[a] Medical Biological Laboratory TNO, P.O. Box 45, 2280 AA Rijswijk, Tel. (015) 138777; [b] Department of Genetics, Agricultural University, Gen. Foulkesweg 53, 6703 BM Wageningen (The Netherlands) Tel. (08370) 82150, and [c] Department of Plant Pathology, University of California, Davis, CA 95616 (U.S.A.) Tel. (916) 752-8712

(Received June 18th, 1985)
(Revision received September 4th, 1985)
(Accepted September 6th, 1985)

SUMMARY

We inserted in frame the *Escherichia coli lacZ* gene into the protein-coding region of the *Aspergillus nidulans trpC* gene and introduced the resultant fused gene into the *A. nidulans* genome. A functional βGal fusion protein was produced. Removal of the *trpC* transcription and translation initiation sequences from the fusion gene abolished production of the fusion protein, showing that expression is dependent on these sequences. Thus, *lacZ* fusions should be of use for estimating gene activity in *A. nidulans*.

INTRODUCTION

Regulation of gene expression in *A. nidulans* has been investigated most extensively by using genetic methods (Cove, 1979; Arst, 1981). Studies on nitrogen and carbon metabolism have shown that expression of genes is controlled by *trans*-acting regulatory elements, which may interact with controlling regions adjacent to structural genes (Hynes, 1979; Sharma and Arst, 1985). However, little is known about the molecular mechanisms controlling gene expression. Recent cloning of genes whose expression is regulated, such as the *amdS* gene (Hynes et al., 1983) and the *trpC* gene (Yelton et al., 1983) opens the possibility of studying these mechanisms in *A. nidulans*.

A convenient method for studying gene expression in prokaryotes (Casadaban et al., 1983) and certain eukaryotes (Casadaban et al., 1983; Lis et al., 1983; Nielsen et al., 1983) is the fusion of the promoter and/or other regulatory sequences of the gene of interest to the *lacZ* gene of *E. coli*. Studies in *E. coli*, *Saccharomyces cerevisiae* and animal cells have shown that the expression and regulation of a parti-

* To whom correspondence and reprint requests should be addressed.

Abbreviations: Ap, ampicillin; bp, base pair(s); βGal, β-galactosidase; Δ, deletion; kb, 1000 bp; Km, kanamycin; ONPG, *o*-nitrophenyl-β-D-galactopyranoside; PMSF, phenylmethylsulfonyl fluoride; ^{R}, resistant; ^{S}, sensitive; wt, wild type; XGal,

ly by assaying βGal activity (Casadaban et al., 1983; Hall et al., 1983). The use of *lacZ* fusions has the additional advantage that gene expression can be detected in vivo by using the chromogenic substrate XGal. The recent demonstration of the ability to transform *A. nidulans* (Ballance et al., 1983; Tilburn et al., 1983; Yelton et al., 1984) raises the possibility of also applying this methodology to *Aspergillus*.

In this paper we describe the construction of a fusion of the *E. coli lacZ* gene with the *trpC* gene of *A. nidulans*. The fusion gene was cloned into an *A. nidulans* vector (Wernars et al., 1985) containing the *amdS* gene as a selection marker. *A. nidulans* strains transformed with the fusion plasmid were analysed for *E. coli lacZ* gene expression. Our results show that the *E. coli lacZ* gene is expressed in *A. nidulans* as a functional fusion protein and that its expression is dependent on the transcriptional and translational signals of the *A. nidulans trpC* gene.

MATERIALS AND METHODS

(a) Strains and plasmids

A. nidulans strain MH1277 (*biA*1, *amdS*320, *amdI*18, *amdA*7, *niiA*4; Hynes et al., 1983) was used as recipient in transformation experiments. Plasmids were constructed and propagated in the *E. coli* K-12 strains JA221 (*Δ*[*trpE*]5, *leuB*6, *recA*, *hsdR*$^-$, *hsdM*$^+$; Clarke and Carbon, 1978) or AMA1004 (*Δ*[*lacIPOZ*]C29, *galU*, *galK*, *trpC*9830, *leuB*6, *hsdR*$^-$, *hsdM*$^+$, *strA*^{R}; Casadaban et al., 1983).

Plasmid p3SR2, containing the *amdS* gene within a 5.0-kb *Eco*RI-*Sal*I fragment cloned in pBR322, was obtained from Dr. M.J. Hynes (Hynes et al., 1983). pGW315 is described by Wernars et al. (1985). pHY101, containing the *A. nidulans trpC* gene within a 4.15-kb *Xho*I fragment cloned in the *Xho*I site of pACYC177 was described by Yelton et al. (1984). pMC1871, containing a nearly complete copy of the *E. coli lacZ* gene within a 3.1-kb *Bam*HI fragment, was obtained from Dr. M.J. Casadaban (Casadaban et al., 1983).

(b) *A. nidulans* transformation, growth conditions and DNA isolation

Transformation of the *A. nidulans* strain MH1277 was carried out as described by Wernars et al. (1985). The normal growth medium was minimal medium (Pontecorvo et al., 1953) in which nitrate was replaced by 10 mM acetamide and containing 1% glucose and 15 nM D(+)-biotin. Cells were tested for βGal activity by their capacity to form blue colonies on 1% agar plates with M9 medium (Maniatis et al., 1982) supplemented with 1% glucose, galactose or lactose, 10 mM acetamide, 15 nM D(+)biotin and 40 μg/ml XGal. Cultures were grown at 37°C.

DNA was isolated from liquid-nitrogen-frozen mycelium as described by Yelton et al. (1984).

RESULTS AND DISCUSSION

(a) βGal activity of *A. nidulans*

Studies on the expression of the βGal gene (*bgaA*) of *A. nidulans* have shown that the gene is repressed by glucose or sucrose and is induced by lactose and galactose (Fantes and Roberts, 1973). Induction of the *Aspergillus bgaA* gene was confirmed by growing the organism on medium containing lactose or galactose and XGal. Under these conditions blue colonies were observed, whereas growing the organism on medium containing glucose or sucrose and XGal resulted in colourless colonies (results not shown). These results enabled us to test for in vivo expression of the *E. coli lacZ* gene in *A. nidulans* without interference from endogenous βGal activity.

(b) Construction of an *A. nidulans trpC-E. coli lacZ* gene fusion

The *Bam*HI fragment of plasmid pMC1871 (Casadaban et al., 1983), containing the *lacZ* gene lacking the first seven codons, was inserted into the *Bam*HI site of the *Aspergillus trpC* gene (Mullaney et al., 1985; Fig. 1). DNA sequence analysis of a fragment containing the junction of the *trpC* gene and the N-terminal part of the *lacZ* gene confirmed that the translational reading frames of the genes were maintained (Kalnins et al., 1983; Mullaney et al., 1985). The fusion gene was inserted in either orientation into the *A. nidulans amdS* vector pGW315 [= p3SR2 (Hynes et al., 1983) containing the 3.6-kb

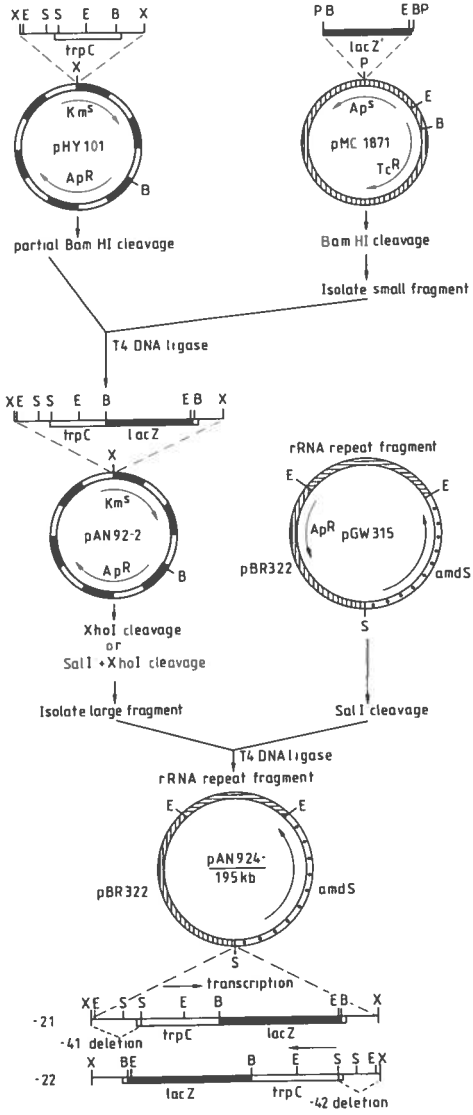


Fig. 1. Diagrammatic representation of the steps used to construct the *trpC-lacZ* fusion gene plasmids presented in these studies.

The 3.1-kb *Bam*HI fragment from pMC1871, containing the *lacZ* gene of *E. coli* but lacking the information for the first seven amino acids, was cloned into the *Bam*HI site of the *A. nidulans trpC* gene on plasmid pHY101. The 7.2-kb *Xho*I fragment from the resulting plasmid, designated pAN92-2 was cloned in either *Eco*RI ribosomal repeat fragment (Borsuk et al., 1982)] to produce plasmids pAN924-21 and pAN924-22. In a second step, the transcription and translation start sequences of the fusion genes were deleted by removing the 1.3-kb *Xho*I-*Sal*I fragment (Mullaney et al., 1985) as shown in Fig. 1. The remaining 6.0-kb *Sal*I-*Xho*I fragment, containing the truncated fusion gene, was inserted in either orientation into the *Sal*I site of plasmid pGW315, to yield the plasmids pAN924-41 and pAN924-42.

(c) Expression of the *trpC-lacZ* fusion gene in *A. nidulans*

The *A. nidulans amdS* deletion strain MH1277 was transformed with plasmids pAN924-21, -22, -41 and -42. AmdS$^+$ transformants were tested for βGal expression on medium containing glucose and XGal. Most (85%) of the transformants obtained with pAN924-21 and -22 (containing the intact fusion gene) produced blue colonies on glucose + XGal medium, whereas transformants obtained with plasmids p3SR2 or pGW315 produced only colourless colonies. These results indicated that introduction of the *trpC-lacZ* fusion gene into *A. nidulans* caused production of βGal activity in a majority of the transformants.

All AmdS$^+$ transformants obtained with the plasmids pAN924-41 and -42 (containing the truncated fusion gene) produced colourless colonies on glucose + XGal medium, showing that deletion of the *Aspergillus* transcription and translation start signals from the fusion gene abolished the production of βGal.

structed by cloning the 6.0-kb *Sal*I-*Xho*I fragment from pAN92-2, containing the *trpC-lacZ* fusion gene without its transcription and translation start sites (Mullaney et al., 1985), in either orientation into the *Sal*I site of pGW315. ▭, *A. nidulans amdS* fragment; ▭, *A. nidulans trpC* gene; —, *A. nidulans trpC* fragment; ▤, *A. nidulans* ribosomal RNA fragment; ■, *E. coli*

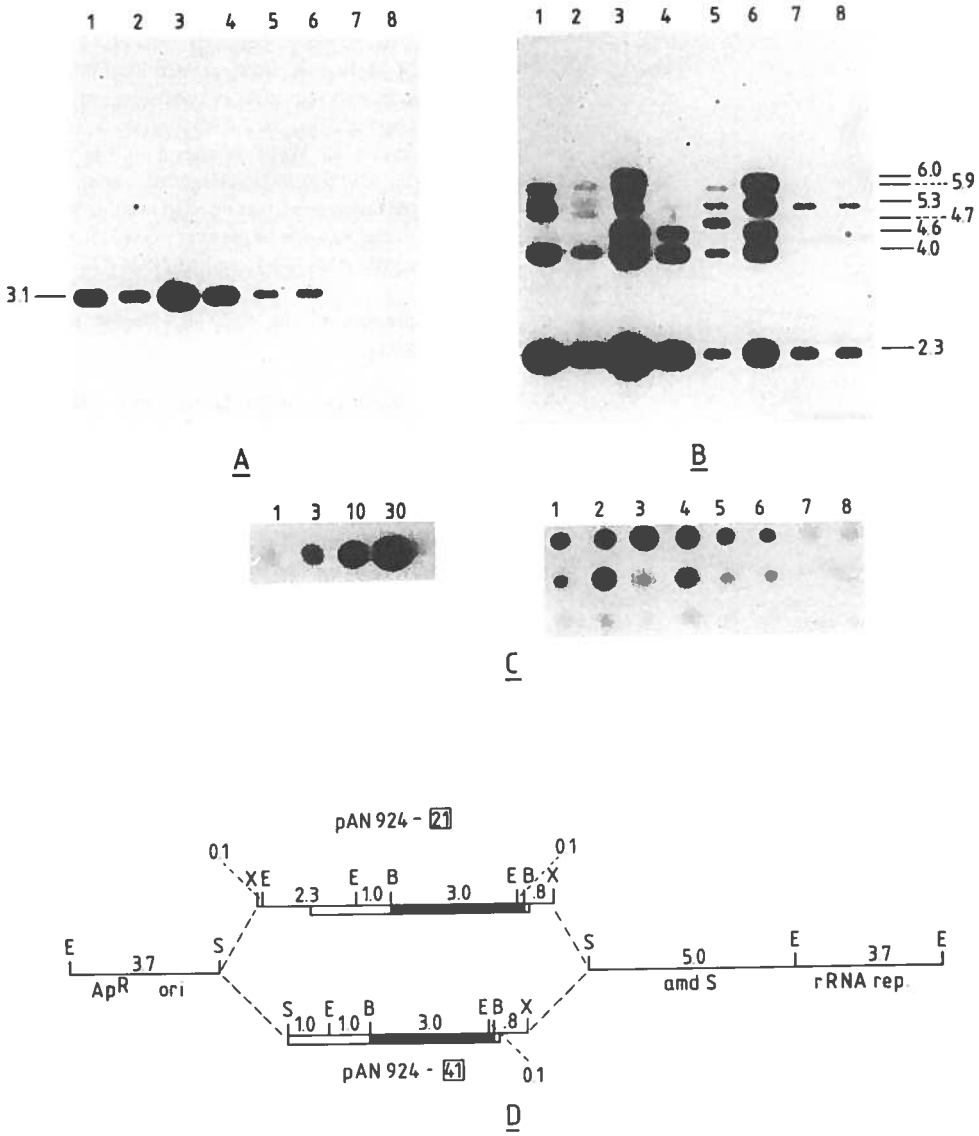


Fig. 2. Southern blot analysis of chromosomal DNA from an *A. nidulans* wild-type strain and transformants. The blots were hybridized with a ^{32}P-labelled *E. coli lacZ* probe (the 3.1-kb *Bam*HI fragment of pMC1871; Fig. 1) (panel A) or an *A. nidulans trpC* probe (a mix of the 2.3-kb *Eco*RI fragment and the 1.8-kb *Eco*RI-*Xho*I fragment of pHY101; Fig. 1) (panels B and C). Panel A shows *Bam*HI-digested chromosomal DNA fractionated on a 0.7% agarose gel. Panel B shows *Eco*RI-digested chromosomal DNA fractionated on a 0.7% agarose gel. The fragment lengths are given in kb. Panel C shows at the left dilutions of plasmid pAN92-2, equivalent to 1, 3, 10 and 30 copies and at the right a dot blot with top to bottom 0.6, 0.18 and 0.06 μg chromosomal DNA. Lanes from left to right are strains: 1, 924-21-I; 2, 924-21-II; 3, 924-22-I; 4, 924-22-II; 5, 924-41-II; 6, 924-42-II; 7, p3SR2-I; 8, wild type. Panel D shows a restriction map of the fusion genes; the closed bar depicts the *lacZ* part of the fusion gene, the open bar depicts the *trpC* part of the fusion gene. B, *Bam*HI; E, *Eco*RI; S, *Sal*I; X, *Xho*I.

(d) Southern blot analysis of the transformants

To verify the presence of the fusion gene in the AmdS$^+$ transformants, Southern blots containing digested chromosomal DNA from the *A. nidulans* wild type and from seven transformants, were analysed. The results of these experiments are shown in Fig. 2, A and B. Hybridization of *Bam*HI-digested chromosomal DNA with the *E. coli lacZ* probe (Fig. 2A), showed that the *lacZ* part of the fusion gene is present in the DNA of the transformants obtained with pAN924-21 and -22, as well as pAN924-41 and -42, because the expected 3.1-kb *Bam*HI fragment (Fig. 2D) hybridized with the probe (Fig. 2A, lanes 1 to 6). No hybridization signal was observed with chromosomal DNA from *A. nidulans* transformed with p3SR2 (Fig. 2A, lane 7) or from the *A. nidulans* wild type (Fig. 2A, lane 8).

Hybridization with the *A. nidulans trpC* probe (Fig. 2B) produced a more complex hybridization pattern. *Eco*RI-digested chromosomal DNA from the transformants obtained with pAN924-21 contained fragments of 2.3 kb, 4.0 kb and 5.9 kb, as expected (Fig. 2D). Similarly, *Eco*RI-digested chromosomal DNA of cells transformed with pAN924-22 contained the expected 2.3-kb, 4.0-kb and 4.6-kb fragments (Fig. 2B, lanes 1 to 4). These results, with the results obtained with the *lacZ* hybridization probe, indicated that the complete fusion gene was present in the transformants obtained with pAN924-21 and -22. The other hybridizing fragments were either fragments that contain the chromosomal *trpC* gene (2.3-kb and 5.3-kb *Eco*RI fragments; Fig. 2B, lanes 7 and 8; Yelton et al., 1983) or fragments that probably arose as a result of rearrangement of plasmid DNA sequences. The hybridization patterns of chromosomal DNA from transformants obtained with pAN924-41 and -42 (Fig. 2B, lanes 5 and 6) show in addition to the 2.3-kb and 5.3-kb bands originating from the chromosomal *trpC* gene (Fig. 2B, lanes 7 and 8) all of the expected bands for pAN924-41 and -42, namely the 4.0-kb, 4.7-kb and 5.9-kb bands for pAN924-41 and the 4.0-kb, 4.6-kb and 6.0-kb bands for pAN924-42 (see Fig. 2D). These results indicated that the absence of βGal activity in transformants obtained with pAN924-41 and -42 on glucose + XGal medium was not caused

The intensities of the *Eco*RI bands in DNA from cells transformed with pAN924-21 and -22 in comparison to the intensity of the 5.3-kb chromosomal *trpC* band (Fig. 2B) indicated that these transformants contain more than one copy of the fusion gene

TABLE I

Copy number analysis of the *trpC-lacZ* fusion gene in the chromosomal DNA of *A. nidulans* transformants and βGal enzyme activity measured in extracts of different *A. nidulans* transformants

A. nidulans transformant	Intact fusion-gene	Copy number[a] fusion-gene	βGal activity[b] (units/mg protein)
924-21-I	+	3.4	342
924-21-II	+	8.0	476
924-22-I	+	7.3	642
924-22-II	+	6.7	378
924-41-II	–	1.9	0.6
924-42-II	–	2.4	0.7
p3SR2-I	–	0	0.3

[a] Chromosomal DNA (0.06, 0.18 or 0.6 μg) from the transformants was spotted in a fixed volume (2 μl) onto nitrocellulose paper. Several dilutions of plasmid pAN92-2, equivalent to 1, 3, 10 and 30 copies of the gene per μg chromosomal DNA were spotted in parallel. The nitrocellulose filters were baked for 2 h at 80°C. Duplicate filters were hybridized with ^{32}P-labelled DNA fragments containing sequences of the *argB* or *trpC* genes of *A. nidulans*. The filters were used to expose X-ray film and spots were cut out and counted in a liquid scintillation counter. The results obtained with the *argB* gene served as a standard to determine the exact amount of DNA present in the spots. Results are expressed relative to wt *A. nidulans trpC* gene having one copy.

[b] Cultures (100 ml) were inoculated with 2×10^8 conidia and grown at 37°C for 18 h. The mycelium was harvested by filtration through Mira cloth, washed with 0.7% NaCl, squeezed, frozen and powdered in liquid nitrogen. The powdered mycelium was resuspended in 3 ml extraction buffer (50 mM sodium phosphate buffer, pH 7.0, 1 mM EDTA and 20 μM PMSF) by vortex mixing and kept for at least 15 min on ice. After this incubation, the extracts were mixed by vortexing several times and centrifuged for 15 min at 15600 × *g*. The supernatant solution was aspirated and stored on ice. Samples of the extracts were removed, frozen in liquid nitrogen and stored at −80°C for Western blotting experiments. This procedure was found to be optimal. Sonication or treatment of the mycelium with chloroform yielded variable results. We also observed that storage of the mycelium at −80°C after freezing in liquid nitrogen caused decreases in the βGal activity of at least 50%. The assay for βGal activity was carried out by using 5–200-μl samples of mycelial extract and ONPG as substrate as described by Miller (1972).

per haploid genome. To determine the number of fusion genes present in the transformants, a copy number analysis of the fusion gene was carried out as described in Table I; the autoradiogram of the dot blot is shown in Fig. 2C.

(e) βGal activity measurements in extracts of transformants

To quantify the amount of βGal present in the transformed *A. nidulans* cells, βGal activity was measured in extracts as described in Table I. βGal activity is seen in cells transformed with plasmids carrying the fusion gene and giving blue colonies on XGal plates. Almost no activity was found in cells transformed with the control plasmid (p3SR2) or with plasmids with the truncated fusion gene. Enzyme-specific activities were not proportional to the number of integrated copies of the fusion gene.

(f) Western blot analysis of the transformants

To demonstrate that the βGal activity found in extracts from the *A. nidulans* transformants was correlated with the presence of *E. coli* βGal in a fusion protein, a Western blotting experiment was carried out by using antiserum against *E. coli* βGal as probe. Fig. 3 shows that a protein of about the expected size (190 kDal) was present in the extracts of cells transformed with a plasmid containing the intact fusion gene (pAN924-21). Extracts of the *A. nidulans* wild type, grown on the same medium (containing glucose as carbon source) as well as wild type, grown on a medium containing galactose as carbon source, did not contain proteins that reacted with the *E. coli* βGal antiserum. No bands were visible with the extracts from cells transformed with the control plasmid p3SR2 or the plasmids carrying the truncated fusion gene (pAN924-41 and -42; results not shown).

(g) Conclusions

We have demonstrated that the *E. coli lacZ* gene, encoding the enzyme βGal, is expressed in *A. nidulans* when fused to the *A. nidulans trpC* gene. Expression is dependent on transcriptional and translational start signals of the *trpC* gene; removal of these elements abolishes βGal fusion–protein formation.

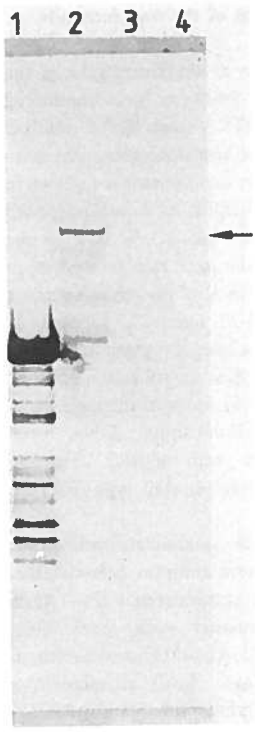


Fig. 3. Western blot analysis of extracts of *A. nidulans* wild type and a *lacZ*-positive transformant. The extracts, prepared as described in the legend of Table I, were fractionated on a 6% polyacrylamide gel. The gel was blotted and the blot was hybridized with anti-*E. coli* βGal serum. A 20% pure βGal preparation was used as a marker. Lanes from left to right: 1, marker; 2, *A. nidulans* MH1277 transformed with pAN924-21 and grown in minimal medium with glucose; 3, *A. nidulans* wild type grown in minimal medium with glucose; 4, *A. nidulans* wild type grown in minimal medium with galactose. The arrow indicates the position predicted for the *trpC-lacZ* fusion gene protein.

Expression of the fusion gene can be detected in vivo by the formation of blue colonies on minimal medium in the presence of the chromogenic substrate XGal. Suppression of the endogenous βGal activity of *A. nidulans* by addition of glucose to the medium allows monitoring of the expression of the fusion gene, even in the presence of a wild-type *A. nidulans* βGal gene. Expression of the fusion gene can also be determined quantitatively by assaying the βGal activity in extracts from transformed cells.

The transformation system of *A. nidulans* has been shown to be highly versatile. Integration events occur at homologous chromosomal positions (Tilburn et al., 1983; Yelton et al., 1984) but may also occur at heterologous sites (Ballance et al., 1983; Wernars et al., 1985). Genes may be replaced by using either one-step or two-step procedures (Miller et al., 1985) allowing for the precise manipulation of the genome. Co-transformation frequencies have been demonstrated to be high (unpublished results). A cosmid vector for cloning genes by complementation of *Aspergillus* mutants has been constructed (Yelton et al., 1985). *lacZ* fusions will provide an additional tool for the isolation and analysis of regulatory signals in this organism.

ACKNOWLEDGEMENTS

We thank Jenny Kröse and Anneke Kuijvenhoven for technical assistance, Karel Wernars for performing the transformation experiments and Dr. O. Brocades Zaalberg for preparing the *E. coli* βGal antiserum.

This work was supported in part by the Dutch Ministry of Economic Affairs, the Netherlands Organization for the Advancement of Pure Research (ZWO) and the United States Department of Agriculture.

REFERENCES

Arst Jr., H.N.: Aspects of the control of gene expression in fungi, in Glover, S.W. and Hopwood, D.A. (Eds.), Genetics as a Tool in Microbiology. Society for General Microbiology Symposium 31. Cambridge University Press, Cambridge, 1981, pp. 131–160.

Ballance, D.J., Buxton, F.P. and Turner, G.: Transformation of *Aspergillus nidulans* by the orotidine-5'-phosphate decarboxylase gene of *Neurospora crassa*. Biochem. Biophys. Res. Commun. 112 (1983) 284–289.

Borsuk, P.A., Nagięć, M.M., Stepień, P.P. and Bartnik, E.: Organization of the ribosomal RNA gene cluster in *Aspergillus nidulans*. Gene 17 (1982) 147–152.

Casadaban, M.J., Martinez-Arias, A. Shapira, S.K. and Chou, J.: β-Galactosidase gene fusions for analyzing gene expression in *E. coli* and yeast. Methods Enzymol. 100 (1983) 293–308.

Clarke, L. and Carbon, J.: Functional expression of cloned yeast DNA in *E. coli*: specific complementation of arginino-succinate lyase (argH) mutations. J. Mol. Biol. 120 (1978) 517–532.

Cove, D.J.: Genetic studies of nitrate assimilation in *Aspergillus nidulans*. Biol. Rev. 54 (1979) 291–327.

Fantes, P.A. and Roberts, C.F.: β-Galactosidase activity and lactose utilization in *Aspergillus nidulans*. J. Gen. Microbiol. 77 (1973) 471–486.

Hall, C.V., Jacob, P.E., Ringold, G.M. and Lee, F.: Expression and regulation of *Escherichia coli lacZ* gene fusions in mammalian cells. J. Mol. Appl. Genet. 2 (1983) 101–109.

Hynes, M.J.: Fine structure mapping of the acetamidase structural gene and its controlling region in *Aspergillus nidulans*. Genetics 91 (1979) 381–392.

Hynes, M.J., Corrick, C.M. and King, J.A.: Isolation of genomic clones containing the *amdS* gene of *Aspergillus nidulans* and their use in the analysis of structural and regulatory mutations. Mol. Cell. Biol. 3 (1983) 1430–1439.

Kalnins, A., Otto, K., Rüther, U. and Müller-Hill, B.: Sequence of the *lacZ* gene of *Escherichia coli*. EMBO J. 2 (1983) 593–597.

Lis, J.T., Simon, J.A. and Sutton, C.A.: New heat shock puffs and β-galactosidase activity resulting from transformation of *Drosophila* with an *hsp70-lacZ* hybrid gene. Cell 35 (1983) 403–410.

Maniatis, T., Fritsch, E.F. and Sambrook, J.: Molecular Cloning. A Laboratory Manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1982.

Miller, J.H.: Experiments in Molecular Genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1972, pp. 352–355.

Miller, B.L., Miller, K.M. and Timberlake, W.E.: Direct and indirect gene replacements in *Aspergillus nidulans*. Mol. Cell. Biol. 5 (1985) 1714–1721.

Mullaney, E.J., Hamer, J.E., Yelton, M.M. and Timberlake, W.E.: Primary structure of the *Aspergillus nidulans trpC* gene. Mol. Gen. Genet. 199 (1985) 37–45.

Nielsen, D.A., Chou, J., MacKrell, A.J., Casadaban, M.J. and Steiner, D.F.: Expression of a preproinsulin-β-galactosidase gene fusion in mammalian cells. Proc. Natl. Acad. Sci. USA 80 (1983) 5198–5202.

Pontecorvo, G., Roper, J.A., Hemmons, L.M., MacDonald, K.D. and Bufton, A.W.J.: The genetics of *Aspergillus nidulans*. Adv. Genet. 5 (1953) 141–238.

Sharma, K.K. and Arst Jr., H.N.: The product of the regulatory gene of the proline catabolism gene cluster of *Aspergillus nidulans* is a positive-acting protein. Curr. Genet. 9 (1985) 299–304.

Tilburn, J., Scazzocchio, C., Taylor, G.G., Zabicky-Zissman, J.H., Lockington, R.A. and Davies, R.W.: Transformation by integration in *Aspergillus nidulans*. Gene 26 (1983) 205–221.

Wernars, K., Goosen, T., Wennekes, L.M.J., Visser, J., Bos, C.J., Van den Broek, H.W.J., Van Gorcom, R.F.M., Van den Hondel, C.A.M.J.J. and Pouwels, P.H.: Gene amplification in

Aspergillus nidulans by transformation with vectors containing the *amdS* gene. Curr. Genet. 9 (1985) 361–368.

Yelton, M.M., Hamer, J.E., De Souza, E.R., Mullaney, E.J. and Timberlake, W.E.: Developmental regulation of the *Aspergillus nidulans trpC* gene. Proc. Natl. Acad. Sci. USA 80 (1983) 7567–7580.

Yelton, M.M., Hamer, J.E. and Timberlake, W.E.: Transformation of *Aspergillus nidulans* using a *trpC* plasmid. Proc. Natl. Acad. Sci. USA 81 (1984) 1470–1474.

Yelton, M.M., Timberlake, W.E. and Van den Hondel, C.A.M.J.J.: A cosmid for selecting genes by complementation in *Aspergillus nidulans*: selection of the developmentally regulated *yA* locus. Proc. Natl. Acad. Sci. USA 82 (1985) 834–838.

Communicated by H. van Ormondt.

CHAPTER 3

A system for the analysis of expression signals in *Aspergillus*

Robert F.M. van Gorcom, Peter J. Punt, Peter H. Pouwels and Cees A.M.J.J. van den Hondel

Gene (1986) **48**, 211-217

GEN 01814

A system for the analysis of expression signals in *Aspergillus*

(Recombinant DNA; translational *lacZ* gene fusion; promoter probe vector; *A. nidulans* genes *argB*, *gpd* and *trpC*; *A. niger*)

Robert F.M. van Gorcom, Peter J. Punt, Peter H. Pouwels and Cees A.M.J.J. van den Hondel*

Medical Biological Laboratory TNO, P.O. Box 45, 2280 AA Rijswijk (The Netherlands) Tel. (015)138777

(Received June 10th, 1986)
(Revision received August 15th, 1986)
(Accepted August 26th, 1986)

SUMMARY

To analyse gene expression signals in *Aspergillus* we have constructed a set of integration vectors each of which contains in front of the *Escherichia coli* '*lacZ* gene a unique *Bam*HI site in one of the three possible translational reading frames and the *A. nidulans argB* gene as a selection marker. The vectors allow in-phase translational fusion of any gene to '*lacZ*. After transformation of an *A. nidulans argB* strain, the vectors integrate with a high percentage at the *argB* locus of the genome, as a single copy. The insertion of the fusion genes at the *argB* locus assures the constancy of influences of the chromosomal environment on gene expression.

INTRODUCTION

With the aid of the *E. coli lacZ* gene a number of expression signals have been analysed in several organisms (for a review see Silhavy and Beckwith, 1985). To perform such analyses either the intact *lacZ* gene has been fused to the promoter region of a particular gene ('transcriptional fusion') or '*lacZ* (the protein-coding region of the *lacZ* gene lacking the first eight codons) has been fused in phase to the protein-coding region of the gene of interest ('translational fusion'). For the analysis of expression signals located downstream from the promoter region the latter fusion must be used.

Recently we described that the *E. coli lacZ* gene is expressed in the filamentous fungus *A. nidulans* after fusion to the *A. nidulans trpC* gene, resulting in the production of an active βGal fusion protein (Van Gorcom et al., 1985). This result indicates that *lacZ* fusions can also be used in *A. nidulans* for the analysis of expression signals. However, for such an analysis in *Aspergillus* only integration vectors are available. Some of these integrate at random places in the genome, sometimes in more than one copy (Ballance et al., 1983; Wernars et al., 1985; Kelly and Hynes, 1985; Buxton et al., 1985). Because of the way these vectors integrate into the host chromosome, the βGal activity may not truly reflect the activity of the expression signals (Van Gorcom et al.,

* To whom correspondence and reprint requests should be addressed.

Abbreviations: Ap, ampicillin; bp, base pair(s); βGal, β-galactosidase; kb, 1000 bp; '*lacZ*, *lacZ* gene without the translation start and the first eight codons; Km, kanamycin; ^{R}, resistant; ^{S}, sensitive; Tc, tetracycline; u, units; XGal, 5-bromo-4-chloro-3-

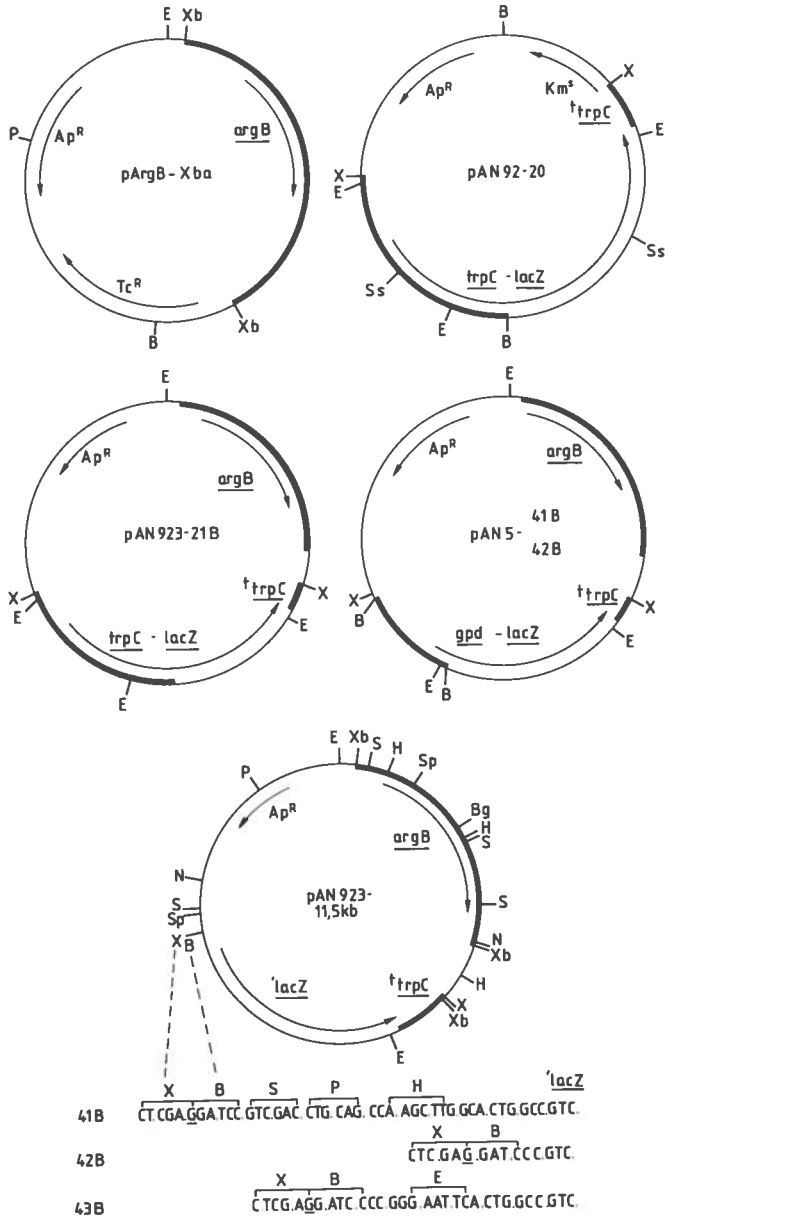


Fig. 1. Map of relevant plasmids used. Plasmids pAN923-41A and B, and pAN923-43A and B were constructed as follows. A 1.9-kb *Bam*HI-*Sst*I fragment from pSKS105 or pSKS106 (Casadaban et al., 1983) containing the N-terminal part of the *lacZ* gene of *E. coli* except for the first eight codons was fused via their *Sst*I ends to a 1.9-kb *Sst*I-*Xho*I fragment of pAN92-20 containing the C-terminal part of the *lacZ* gene and the C-terminal part of the *A. nidulans trpC* gene. The resulting *Bam*HI-*Xho*I fragments were cloned in pAN3-2A and B after digestion of the plasmids with *Bam*HI and *Sal*I, yielding pAN923-41A and B and pAN923-43A and B, respectively. Plasmids

which preferentially integrates at a unique place in the genome in one copy. In addition, it would be advantageous if the vector contained unique restriction sites in different reading frames in front of *'lacZ*, to facilitate in-phase translational fusion of the two coding regions.

In this paper we describe the construction of a set of three integration vectors. Each of these vectors contains a unique *Bam*HI site in one of the three different reading frames in front of *'lacZ*. The vectors also contain the *A. nidulans argB* gene as a selection marker (Berse et al., 1983; Miller et al., 1985).

With the aid of these vectors we have compared the expression of *'lacZ* under control of the expression signals of the *A. nidulans trpC* gene (Yelton et al., 1983; Mullaney et al., 1985) with that of *'lacZ* under control of the expression signals of the *A. nidulans* glyceraldehyde-3-phosphate-dehydrogenase (*gpd*) gene (P.J.P. and C.A.M.J.J.v.d.H., to be published). The results of this study reveal that in *A. nidulans* quantitative analysis of expression signals is feasible with the vectors constructed.

MATERIALS AND METHODS

(a) Strains and plasmids

The *A. nidulans argB* strain (*argB*, *biA*1, *methG*1) used as recipient in transformation experiments was obtained from Dr. W.E. Timberlake (Univ. of Georgia, Athens, GA, U.S.A.). Plasmids were constructed and propagated in the *E. coli* K-12 strain JA221 (Clarke and Carbon, 1978) or in strain AMA1004 (Casadaban et al., 1983). Plasmid pArgB-Xba, containing the *A. nidulans argB* gene within a 3.5-kb *Xba*I fragment cloned into the *Pvu*II site of pBR329 was obtained from Dr. W.E. Timberlake (see Miller et al., 1985). Plasmids pSKS105 and pSKS106 were described by Casadaban et al. (1983). Plasmid pAN92-20 was derived from pAN92-2, which contains the *A. nidulans trpC-E. coli lacZ* fusion (Van Gorcom et al., 1985), by removing the *Bam*HI site downstream from *'lacZ*. Plasmid pAN92-21 is a derivative of pAN92-20 in which the *Bam*HI site in front of *'lacZ* was removed by inserting a linker (GATCTCGATCGA). Plasmids pAN3-2A and B are derivatives of pArgB-Xba (= pBR329 containing the 3.5-kb *Xba*I fragment of the *A. nidulans* genome bearing the *argB* gene (Miller et al., 1985) cloned into the *Pvu*II site) in the *Bam*HI site of which a linker was cloned in two orientations with sites for *Xho*I, *Bam*HI, *Sal*I and *Xho*I.

(b) *Aspergillus nidulans* transformations, growth conditions and other procedures

Transformation of *A. nidulans* was carried out as described by Yelton et al. (1984). Unless mentioned all other procedures were carried out as described by Van Gorcom et al. (1985).

RESULTS AND DISCUSSION

(a) Construction of the *lacZ* gene fusion vectors

To analyse gene expression signals we have constructed a set of vectors (pAN923-41A and B, -42A and B, and -43A and B) which allow fusion of the 5′ part of the protein-coding region of a gene of interest together with its upstream expression signals to the protein-coding region of the *lacZ* gene of *E. coli*. To generate in-phase fusions a unique *Bam*HI site was

pAN923-42A and B were constructed by cloning a 3.8-kb *Bam*HI-*Xho*I fragment of pAN92-20 containing the *lacZ* gene except for the first eight codons and the C-terminal part of the *A. nidulans trpC* gene in pAN3-2A and B linearised with *Bam*HI and *Sal*I. The DNA sequence around the *Bam*HI cloning site in the vectors pAN923-41A and B, -42A and B, and -43A and B is shown. Plasmids pAN923-21A and B were constructed by inserting a 7.1-kb *Xho*I fragment of pAN92-21 containing the entire *trpC-lacZ* fusion in two orientations into pAN3-2A digested with *Xho*I. Plasmids pAN5-41A and B and pAN5-42A and B were constructed as follows: a 2.0-kb *Stu*I fragment, which contains the N-terminal part of the protein coding sequence of the *A. nidulans gpd* gene together with 1.6-kb 5′ non-coding sequences, was converted into a *Bam*HI fragment by adding *Bam*HI adapters. This fragment was cloned in the *Bam*HI site of pAN923-41A and B, and pAN923-42A and B, respectively. In the A plasmids the transcription directions of the fusion gene and the *areB* gene are the same. In the B plasmids the transcription units have an opposite orientation. Thick lines represent *A. nidulans* DNA;

introduced in the three different reading frames in front of *'lacZ*. Downstream of *'lacZ* a fragment was cloned which contains the 3' end of the *A. nidulans trpC* gene. This fragment presumably harbours the transcription termination signals of the *trpC* gene. As a selection marker for *Aspergillus* transformation the *A. nidulans argB* gene was inserted into the plasmids. In the A version of the plasmids the directions of transcription of the *argB* gene and *'lacZ* are the same, whereas in the B plasmids the transcription units have an opposite orientation. The construction strategy and a map with the translational phasing of the three plasmids are presented in Fig. 1.

To test whether these vectors are suitable for gene expression studies we have carried out a comparative analysis of the expression signals of the *A. nidulans gpd* and *trpC* genes.

The *gpd-lacZ* fusion plasmids were constructed by cloning a fragment containing the expression signals and the N-terminal part of the *A. nidulans gpd* gene (P.J.P. and C.A.M.J.J.v.d.H., to be published) into pAN923-41A and B and into pAN923-42A and B, resulting in the plasmids pAN5-41A and B, and -42A and B, respectively. From the nucleotide sequence of the *gpd* gene around the fusion site we could predict that after fusion the protein-coding reading frames in plasmids pAN5-41 were in phase and in plasmids pAN5-42 out of phase (Fig. 1).

The *trpC-lacZ* fusion plasmid was obtained by cloning the *trpC-lacZ* fusion gene present on plasmid pAN92-20 into the *argB* integration vectors pAN3-2A and B, resulting in plasmids pAN923-21A and pAN923-21B, respectively (Fig. 1).

(b) Analysis of *Aspergillus nidulans* transformants containing the different fusion genes

The *A. nidulans argB* strain was transformed with plasmids pAN923-21, pAN923-41, -42 and -43, and pAN5-41 and -42. ArgB$^+$ transformants were tested for βGal expression on agar plates containing XGal medium. For the transformants obtained with pAN923-21 (*trpC-lacZ* fusion) and pAN5-41 (in-phase *gpd-lacZ* fusion) 60–90% of the colonies were blue, whereas all the transformants obtained with the vectors pAN923-41, -42 and -43 (*'lacZ* without *A. nidulans* expression signals) and with pAN5-42 (out-of-phase *gpd-lacZ* fusion) gave colourless colonies on XGal plates.

Comparison of the colonies obtained after transformation with plasmids pAN923-21 and pAN5-41 showed that the plasmids containing the *gpd* expression signals resulted in a much more intense blue colour than those containing the *trpC* expression signals.

(c) Analysis of the integration patterns of the *lacZ*-fusion plasmids

To determine the site of integration of the vectors containing the different fusion genes and the number

TABLE I

Results of Southern-blot analysis of integration patterns of the *lacZ*-fusion plasmids in chromosomal DNA of *A. nidulans* ArgB$^+$ transformants which produced blue colonies on XGal plates[a]

Plasmid	Number analysed	One copy integrated at the *argB* locus	Others[b]
pAN923-21A	4	3	1
pAN923-21B	9	3	6
pAN5-41A	8	3	5
pAN5-41B	9	4	5
Total	30	13	17

[a] The *A. nidulans argB* strain was transformed with the plasmids pAN923-21A, pAN923-21B, pAN5-41A and pAN5-41B. *Eco*RI or *Bam*HI-digested chromosomal DNA of ArgB$^+$ transformants was analysed by Southern blotting analysis. The blots were hybridized with a ^{32}P-labelled *A. nidulans argB* probe (the 0.95-kb *Sph*I-*Sal*I fragment of pArgB-Xba; Fig. 1) or a ^{32}P-labelled *A. nidulans gpd* probe (a 1.8-kb *Sst*I fragment containing almost the complete *gpd* gene; unpublished results). From the resulting autoradiographs the place of integration and the number of integrated copies of the transforming plasmids was deduced.

[b] Others: integrated either at the *argB* locus in more than one copy or in one or more copies elsewhere in the genome.

of copies integrated, chromosomal DNA was isolated from a number of ArgB$^+$ transformants. Southern-blot analysis with different probes of *Eco*RI- or *Bam*HI-digested chromosomal DNA revealed (Table I) that in about 40% of the transformants which gave rise to blue colonies (amounting to 20–30% of the total number of transformants) the vector is integrated at the homologous *argB* locus in one copy per genome. This result indicates that only a limited number of transformants has to be screened to find transformants that have the *lacZ*-fusion plasmid integrated in one copy at the *argB* locus.

Southern-blot analysis also revealed that in some transformants integration had taken place at the *trpC* or *gpd* locus. In addition multiple integrations, gene conversions and complex recombinations of the *lacZ*-fusion plasmids were observed.

(d) Determination of βGal activity

To quantify the expression of the *gpd-lacZ* and *trpC-lacZ* fusion genes, the βGal activity was measured in extracts of transformants containing the plasmid pAN5-41A or B, or pAN923-21A or B, integrated in one copy at the *argB* locus.

As shown in Table II, the βGal activity in the transformants containing the *gpd-lacZ* fusion gene was much higher (approx. 10 times) than that in the transformants containing the *trpC-lacZ* fusion gene (lines 1–4 vs. lines 5–10). Hardly any activity was found in the untransformed *A. nidulans argB* strain (line 14; the endogenous βGal activity was repressed in the minimal medium containing glucose; Fantes and Roberts, 1973; Van Gorcom et al., 1985). These results indicate that expression of the *lacZ*-fusion gene is much more efficient under control of the *gpd* regulatory elements than under control of the *trpC* signals.

We also checked whether the orientation of the fusion genes with respect to the transcriptional direction of the *argB* gene (indicated as A and B, respectively) influences the expression of the fusion gene. As shown in Table II no difference in βGal activity could be detected when the orientation of the fusion gene is reversed, indicating that the orientation of the fusion gene has no significant effect on its expression (lines 1–2 vs. lines 3–4 and lines 5–7 vs. lines 8–10).

Integration of the *gpd-lacZ* fusion gene at either the *gpd* locus or the *argB* locus results in similar levels of expression (line 11 vs. lines 5–10).

TABLE II

Results of the βGal activity measurements in extracts of the ArgB$^+$ transformants[a]

Line	Transformant	Plasmid	Promoter	Phase[b]	Place of integration	βGal act. u/mg
1	21A1	pAN923-21A	*trpC*	+	*argB*	448
2	21A3	pAN923-21A	*trpC*	+	*argB*	615
3	21B1	pAN923-21B	*trpC*	+	*argB*	482
4	21B10	pAN923-21B	*trpC*	+	*argB*	418
5	41A5	pAN5-41A	*gpd*	+	*argB*	4725
6	41A6	pAN5-41A	*gpd*	+	*argB*	5598
7	41A9	pAN5-41A	*gpd*	+	*argB*	5242
8	41B5	pAN5-41B	*gpd*	+	*argB*	5511
9	41B7	pAN5-41B	*gpd*	+	*argB*	5152
10	41B12	pAN5-41B	*gpd*	+	*argB*	5236
11	41B15	pAN5-41B	*gpd*	+	*gpd*	4237
12	42A6	pAN5-42A	*gpd*	−	*argB*	7.7
13	42B2	pAN5-42B	*gpd*	−	*argB*	8.2
14	*argB*	—	—			0.3

[a] ArgB$^+$ transformants, obtained by transformation of the *A. nidulans argB* strain with the plasmids pAN923-21A, pAN923-21B, pAN5-41A, pAN5-41B, pAN5-42A and pAN5-42B, were analysed. The place of integration in the chromosomal DNA and the number of integrated copies of the transforming plasmid was determined by Southern blotting analysis as described in Table I. From 13 individual transformants containing one copy of the transforming plasmid integrated at the *argB* locus (lines 1–10, 12 and 13) or at the *gpd* locus

A very low but significant βGal activity is present in extracts of transformants obtained with the out-of-phase fusion plasmids pAN5-42A and B (lines 12 and 13). This might be caused by translation aberrantly started at an internal AUG codon which is present in phase before *'lacZ*.

In all the experiments described the *A. nidulans* βGal gene (*bgaA*) was not induced (minimal medium + glucose; Fantes and Roberts, 1973; Van Gorcom et al., 1985). However, to perform expression studies with the vectors described, under conditions that the *A. nidulans bgaA* gene is induced, a *bgaA* mutant of the *A. nidulans argB* strain was constructed (C.J. Bos and I.E. Mattern, unpublished results).

(e) Conclusions

(1) In this paper a set of vectors for the analysis of gene expression is described that allow in-phase translational fusion of an *A. nidulans* gene together with its upstream expression signals to *E. coli 'lacZ*. These vectors integrate with a high percentage at a specific location on the chromosome (the *argB* locus) of the recipient. Since integration at a specific place minimizes the influence that differences in chromosomal environment and chromatin structure might have on gene expression, a detailed quantitative analysis of the expression and regulation signals can now be carried out. With the aid of these vectors the expression efficiency of promoter regions of different transcription units can also be compared.

Integration occurs not only at the *argB* locus, but also at the locus of the gene of interest (in this paper *gpd* or *trpC* locus). Thus, the expression and regulation elements can also be analysed at their original chromosomal locus. In addition, the efficiency of expression and modes of regulation of expression of the fusion gene present in different chromosomal environments can be compared.

(2) As an example of such an analysis we have compared the efficiency of expression of *'lacZ* placed under the control of the *gpd* or *trpC* expression signals. The results of this comparison revealed that the level of expression of *'lacZ* under control of the expression signals of the *gpd* gene is approximately tenfold higher than the level of expression under control of the *trpC* signals. In addition, it was found that expression of the fusion gene under control of the *gpd* or *trpC* expression elements is not affected by its orientation in the genome. Finally, a similar level of expression was observed for the *gpd-lacZ* fusion gene integrated either at the *gpd* locus or at the *argB* locus.

(3) To test whether the *lacZ*-fusion system can also be used in other filamentous fungi, *A. niger* was transformed with the plasmids pAN5-41B and pAN923-21B. Analysis of transformants on agar plates with XGal medium showed similar differences in the blue colouring as found for *A. nidulans*, indicating that the constructed *lacZ*-fusion vectors can also be used in *A. niger*. However, to make quantitative analysis possible the presence of an homologous *A. niger* selection marker on the expression analysis vectors is required to direct integration of one copy at a specific place on the chromosome. Recently an *A. niger* orotidine-5-phosphate-decarboxylase mutant (*pyrG*) and the corresponding *A. niger pyrG* gene were isolated in our laboratory (Van Hartingsveldt et al., 1987), which will be used to adapt the expression analysis vectors for *A. niger*.

ACKNOWLEDGEMENTS

We thank M. Dingemanse, W. van Hartingsveldt, B. Jacobs, J. Kröse and C. van Zeijl for assistance with some of the experiments described in this paper and F. Bleichrodt for the critical reading of the manuscript.

REFERENCES

Ballance, D.J., Buxton, F.P. and Turner, G.: Transformation of *Aspergillus nidulans* by the orotidine-5'-phosphate decarboxylase gene of *Neurospora crassa*. Biochem. Biophys. Res. Commun. 112 (1983) 284–289.

Berse, B., Dmochowska, A., Skrzypek, M., Węgleński, P., Bates, M.A. and Weiss, R.L.: Cloning and characterization of the ornithine carbamoyl-transferase gene from *Aspergillus nidulans*. Gene 25 (1983) 109–117.

Buxton, F.P., Gwynne, D.I. and Davies, R.W.: Transformation of *Aspergillus niger* using the *argB* gene of *Aspergillus nidulans*. Gene 37 (1985) 207–214.

Casadaban, M.J., Martinez-Arias, A., Shapira, S.K. and Chou, J.: β-Galactosidase gene fusions for analyzing gene expression in *E. coli* and yeast. Methods. Enzymol. 100 (1983) 293–308.

Clarke, L. and Carbon, J.: Functional expression of cloned yeast DNA in *E. coli*: specific complementation of arginino-succinate lyase (*argH*) mutations. J. Mol. Biol. 120 (1978) 517–532.

Fantes, P.A. and Roberts, C.F.: β-Galactosidase activity and lactose utilization in *Aspergillus nidulans*. J. Gen. Microbiol. 77 (1973) 471–486.

Kelly, J.M. and Hynes, M.J.: Transformation of *Aspergillus niger* by the *amdS* gene of *Aspergillus nidulans*. EMBO J. 4 (1985) 475–479.

Miller, B.L., Miller, K.Y. and Timberlake, W.E.: Direct and indirect gene replacement in *Aspergillus nidulans*. Mol. Cell. Biol. 5 (1985) 1714–1721.

Mullaney, E.J., Hamer, J.E., Yelton, M.M. and Timberlake, W.E.: Primary structure of the *Aspergillus nidulans trpC* gene. Mol. Gen. Genet. 199 (1985) 37–45.

Silhavy, T.J. and Beckwith, J.R.: Uses of *lac* fusions for the study of biological problems. Microbiol. Rev. 49 (1985) 398–418.

Van Gorcom, R.F.M., Pouwels, P.H., Goosen, T., Visser, J., Van den Broek, H.W.J., Hamer, J.E., Timberlake, W.E. and Van den Hondel, C.A.M.J.J.: Expression of an *Escherichia coli* β-galactosidase fusion gene in *Aspergillus nidulans*. Gene 40 (1985) 99–106.

Van Hartingsveldt, W., Mattern, I.E., Van Zeijl, C.J.M., Pouwels, P.H. and Van den Hondel, C.A.M.J.J.: Mol. Gen. Genet. (1987) in press.

Wernars, K., Goosen, T., Wennekes, L.M.J., Visser, J., Bos, C.J., Van den Broek, H.W.J., Van Gorcom, R.F.M., Van den Hondel, C.A.M.J.J. and Pouwels, P.H.: Gene amplification in *Aspergillus nidulans* by transformation with vectors containing the *amdS* gene. Curr. Genet. 9 (1985) 361–368.

Yelton, M.M., Hamer, J.E., De Souza, E.R., Mullaney, E.M. and Timberlake, W.E.: Developmental regulation of the *Aspergillus nidulans trpC* gene. Proc. Natl. Acad. Sci. USA 80 (1983) 7576–7580.

Yelton, M.M., Hamer, J.E. and Timberlake, W.E.: Transformation of *Aspergillus nidulans* using a *trpC* plasmid. Proc. Natl. Acad. Sci. USA 81 (1984) 1470–1474.

Communicated by H. van Ormondt.

CHAPTER 4

Expression analysis vectors for *Aspergillus niger*

Robert F.M. van Gorcom and Cees A.M.J.J. van den Hondel

Nucleic Acid Research (1988) **16**, 9052

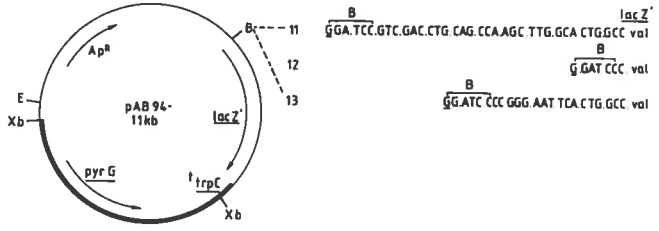
Volume 16 Number 18 1988 Nucleic Acids Research

Expression analysis vectors for *Aspergillus niger*

Robert F.M.van Gorcom and Cees A.M.J.J.van den Hondel

Medical Biological Laboratory TNO, PO Box 45, 2280 AA Rijswijk, The Netherlands
Submitted July 20, 1988

For the analysis of expression signals in Aspergillus nidulans a set of integration vectors (pAN923-41B, -42B and -43B) has been described recently (1). The fusion of transcription and translation start regions to the β-galactosidase gene of E. coli in these vectors facilitates the quantitative functional analysis of these sequences at a defined site in the chromosome. To carry out such gene expression studies in Aspergillus niger similar vectors were constructed, by replacing the A. nidulans argB gene present on pAN923-41B, -42B and -43B by a mutant allele of the A. niger pyrG gene (2). A diagram of the resulting plasmids pAB94-11, -12 and -13 is shown below. To facilitate the selection of transformants in which the vector was integrated at the homologous locus, a mutant pyrG gene was used (3). The pyrG gene was mutated by filling in the BamHI restriction site that is present in the protein-coding region of the gene.
The advantage of a mutated pyrG gene to direct integration of a single copy to the chromosomal pyrG locus was demonstrated by the results of a transformation experiment wherein two plasmids were used only differing in the mutation in the pyrG selection marker. Of 48 transformants that were made using the vector with the intact pyrG gene none had a single copy of the plasmid integrated at the pyrG locus. Out of 32 transformants that were made with the mutated pyrG gene 14 had a single copy of the plasmid integrated at the pyrG locus. Transformation frequencies of 10 per μg DNA were obtained with plasmids pAB94-11, -12 and -13.
At present we are using these expression analysis vectors to study the promoters of the gpdA genes of A. nidulans and A. niger in A. niger.



REFERENCES
(1) van Gorcom R.F.M. et al. (1986) Gene 48, 211-217.
(2) van Hartingsveldt W. et al. (1987) Mol.Gen.Genet. 206, 71-75.
(3) Hamer J.E. et al. (1987) Mol.Cell.Biol. 7, 2352-2359.

CHAPTER 5

Isolation and molecular characterization of the benzoate-para-hydroxylase gene (*bphA*) of *A. niger*; a member of a new gene-family of the cytochrome P450 superfamily

Robert F.M. van Gorcom, Johan G. Boschloo, Anneke Kuijvenhoven, Jacqueline Lange, Arjen J. van Vark, Cees J. Bos, Johannes A.M. van Balken, Peter H. Pouwels and Cees A.M.J.J. van den Hondel

Molecular and General Genetics (1990) **223**, 192-197

Isolation and molecular characterisation of the benzoate-para-hydroxylase gene (*bphA*) of *Aspergillus niger*: A member of a new gene family of the cytochrome P450 superfamily

Robert F.M. van Gorcom[1], Johan G. Boschloo[2], Anneke Kuijvenhoven[1], Jacqueline Lange[1], Arjen J. van Vark[2], Cees J. Bos[2], Johannes A.M. van Balken[3], Peter H. Pouwels[1], and Cees A.M.J.J. van den Hondel[1]

[1] Medical Biological Laboratory TNO, P.O. Box 45, 2280 AA Rijswijk, The Netherlands
[2] Department of Genetics, Agricultural University, De Dreijenlaan 2, 6703 HA Wageningen, The Netherlands
[3] DSM Research, CP-BOC, P.O. Box 18, 6160 MD Geleen, The Netherlands

Received April 17, 1990

Summary. The gene coding for benzoate-para-hydroxylase (*bphA*) of *Aspergillus niger* was cloned using differential hybridisation techniques and complementation of mutants deficient in this enzyme activity. The nucleotide sequence of the gene was determined, the presence of two introns was shown and the transcription start and termination sites were determined. The structure of the mRNA upstream from the long open reading frame (ORF) is unusual. It contains two small, overlapping ORFs whose function is unknown. Comparison of the deduced amino acid sequence of the protein with the sequences present in the databanks, indicated a significant similarity of BPH to the superfamily of cytochrome P450 enzymes. Further analysis revealed that this protein is a member of a new P450 gene family designated P450LIII. The gene is designated *CYP53*. To increase the BPH activity of *A. niger*, multiple copies of the *bphA* gene were introduced into the genome of a recipient strain by transformation. Although increased intracellular levels of the BPH protein could be detected, the BPH enzyme activity was decreased, suggesting titration of another essential component.

Key words: Monooxygenase – Recombinant DNA – Overexpression – Upstream reading frames – Complementation

Introduction

Monooxygenases, a very interesting class of enzymes, generally fall into one of the following groups: pteridine-dependent monooxygenases, cytochrome P450 monooxygenases and flavin-linked monooxygenases (Walsh 1978). These enzymes are involved in both biosynthetic and catabolic pathways in virtually all organisms. They are often also involved in detoxification reactions. For example, in the mammalian liver a number of these enzymes are induced by toxic compounds.

Not much is known about monooxygenases in filamentous fungi. It has been reported that cytochrome P450 activities in *Aspergillus flavus* are involved in aflatoxin biosynthesis (Hamid and Smith 1987). The best characterised enzyme which carries out an aromatic para-hydroxylation is benzoate-para-hydroxylase (BPH; EC 1.14.13.12) of *A. niger* (Reddy and Vaidyanathan 1975). These authors described the purification and induction of this enzyme and reported that BPH contains a tetrahydropteridine prosthetic group. To study this type of enzyme in more detail and to analyse its potential for industrial applications, we have carried out studies on the para-hydroxylation of benzoate by *A. niger*, employing whole cell technology.

Benzoate is metabolised by *A. niger* in a series of steps, the first of which is the para-hydroxylation of the aromatic ring, carried out by BPH. The enzyme is present only after induction by benzoate. Mutants disturbed in this enzymatic activity have been isolated and most of them probably involve the structural gene designated *bphA* (Boschloo et al. in preparation). The BPH activity of benzoate-induced *A. niger* cells is low. In order to increase this activity and obtain more information about the protein and the regulation of its biosynthesis, we decided to clone the gene coding for BPH and overexpress it in *A. niger*.

This report describes the isolation of the *bphA* gene from a cosmid library using differential hybridisation techniques and subsequent complementation of *bphA* mutants with the selected cosmids. The molecular structure of the *bphA* gene was determined. From the deduced amino acid sequence of the gene it was concluded that BPH is a cytochrome P450 monooxygenase which is a member of a new gene family in the cytochrome P450 superfamily (Nebert et al. 1989).

Offprint requests to: R.F.M. van Gorcom

Materials and methods

Strains and plasmids. *Escherichia coli* strains JA221 (Clarke and Carbon 1978) and JM109 (Yanisch-Perron et al. 1985) were used for the construction and propagation of vector molecules. *A. niger* NV DSM 2061 (ATCC 1015) and its derivatives N204 (*csp21, met21*), N271 (*csp21, fwn21, pdx21*) and N282 (*csp21, bphA21, pyr25, pab21*) were used in this study for transformation experiments. In this paper the nomenclature of Van Hartingsveldt et al. (1987) is used for the *A. niger* orotidine-5'-phosphate decarboxylase gene (*pyrG*). Goosen et al. (1987), who isolated the same gene, used the name *pyrA*. Poly(A)$^+$ RNA from *A. niger* strain T16 (an *A. niger* N271 transformant containing several copies of the *bphA* gene) was used as a template for mRNA sequence determination. Subcloning experiments were carried out using plasmid pUC19 (Yanisch-Perron et al. 1985). For selection after transformation of *A. niger*, plasmids p3SR2 (or part of it) containing the *A. nidulans amdS* gene (Hynes et al. 1983), pAB4-1 (or part of it) containing the *A. niger pyrG* gene (Van Hartingsveldt et al. 1987) and pAN7-1 containing a hygromycin B resistance unit, functional in *Aspergilli*, (Punt et al. 1987) were used.

Gene library construction. A cosmid library of *A. niger* NV DSM 2061 was constructed in the cosmid vector pKBY2 (Yelton et al. 1985). Chromosomal DNA isolated from frozen mycelium (Yelton et al. 1984) was partially digested with *Sau*3AI and size fractionated on a 1.25–5 M NaCl gradient. Fractions ranging in size from 25 to 40 kb were ligated to the vector linearised with *Bgl*II. Packaging was carried out using commercial packaging extracts.

Recombinant DNA procedures; DNA and RNA sequencing. All methods used in this work have been published previously (Van Gorcom et al. 1986; Punt et al. 1988) or were carried out as described by Maniatis et al. (1982). Computer analysis of the data presented in this paper was performed using software of the GCG group of the University of Wisconsin (Devereux et al. 1984).

Enzyme activity determination and immunological procedures. Mycelium was grown in liquid CM (Pontecorvo et al. 1953) and after 24 h at 30° C transferred to 50 mM HEPES (pH 6.4) containing 0.1% benzoate. After 5 h the cells were harvested at 4° C by filtration, and the mycelium was frozen in liquid nitrogen and stored at −80° C. For the preparation of a crude extract, mycelium was disrupted by means of a microdismembrator (Braun). The macerated mycelium was resuspended in 50 mM TRIS-HCl (pH 7.8) and BPH activity was measured spectrophotometrically by following the benzoate-dependent rate of oxidation of NADPH at 340 nm. One unit of enzyme activity is defined as the amount required for the oxidation of 1 µmol of NADPH per minute (Cordewener et al. 1987). Western blotting was carried out using the methods of Towbin et al. (1979) and an antiserum (α-SP32) specific for BPH, directed against a synthetic peptide derived from the deduced amino acid sequence (residues 199–215) (Gerritse et al. 1990).

Results

Cloning of the bphA *gene of* A. niger

Attempts to isolate the benzoate-para-hydroxylase (*bphA*) gene by transformation of a *bphA* mutant strain with a cosmid library and direct selection for growth on benzoate as sole carbon source were not successful. In order to preselect cosmids containing genes that are induced by benzoate (BPH activity is only present after induction with benzoate; Cordewener et al. in preparation), the cosmid library was hybridised with radioactive cDNA synthesised on poly(A)$^+$ RNA isolated from benzoate-induced (+ probe) or para-hydroxybenzoate-induced (− probe) mycelium as a template. Cosmids that gave a much stronger hybridisation signal with the + probe than the − probe were isolated. After rescreening, 25 different cosmids were isolated containing genes that were inducible by benzoate. These cosmids were tested for the presence of the *bphA* gene by cotransformation of *A. niger* N282 (*csp21, bphA21, pyr25, pab21*) with individual cosmids and plasmid pAB4-1. One of the cosmids tested (called pAB8-1) was able to complement the *bphA21* mutation of this strain.

Cosmid pAB8-1 also complemented eight other, independently isolated *bphA* mutants, strongly suggesting the presence of the *bphA* gene on pAB8-1. In view of the different genetic backgrounds of the various *bphA* mutant in the latter experiments, we used besides pAB4-1 two other selectable plasmids p3SR2 and pAN7-1, for cotransformation. The benzoate-induced gene could be localised by hybridisation experiments, on a 9.8 kb *Eco*RI fragment and a 6.2 kb *Eco*RI-*Pvu*II subfragment of pAB8-1 (results not shown). These fragments were cloned in pUC19, resulting in plasmids pAB8-2 and pAB8-22, respectively.

Characterisation of the cloned bphA *gene*

To demonstrate the presence of a functional copy of the *bphA* gene on the cloned 6.2 kb *Eco*RI-*Pvu*II fragment, a plasmid (pAB8-25) containing this fragment was used to transform strain N282. On this plasmid a mutant allele of the *A. niger pyrG* gene is present. Transformation with this plasmid of a strain carrying a different mutant allele results in Pyr$^+$ transformants in which the vector has been preferentially integrated at the homologous *pyr* locus (Van Gorcom and Van den Hondel 1988). Of 32 Pyr$^+$ transformants of strain N282 tested, 14 showed a restored ability to grow on benzoate. Southern analysis of these 32 transformants revealed that these 14 transformants had a single copy of the plasmid integrated at the *pyr* locus. The other Pyr$^+$ transformants did not contain plasmid sequences. The ability to restore growth on benzoate to the *bphA* mutant strain N282 with one copy of pAB8-25 integrated at the *pyr* locus demonstrates the presence of a functional copy of the *bphA* gene on this plasmid. Genetic analysis of strain N282 (Boschloo et al. to be published) showed that the

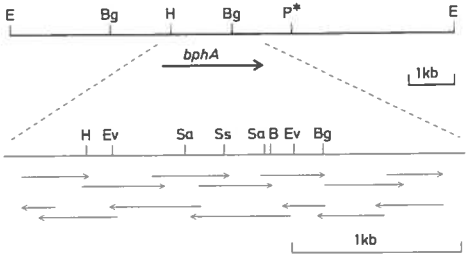


Fig. 1. Restriction map of a 9.8 kb *Eco*RI fragment of the *Aspergillus niger* chromosome that contains the *bphA* gene. The location of the gene and the transcription direction is shown (*thick arrow*). The sequencing strategy used is shown by the *thin arrows*. B, *Bam*HI; Bg, *Bgl*II; E, *Eco*RI; Ev, *Eco*RV; H, *Hpa*I; P*, *Pvu*II (not unique); Sa, *Sac*I; Ss, *Sst*II. There are no sites for *Hin*dIII, *Kpn*I and *Xba*I

bphA gene is located at chromosome I, whereas the *pyr* locus is on chromosome III (Bos et al. 1989).

Molecular characterisation of the bphA *gene*

A restriction map of the cloned 9.8 kb *Eco*RI fragment was made (Fig. 1) and the *bphA* gene was further localised on this fragment by Northern blot experiments (results not shown). The DNA sequence of the *bphA* gene was determined, using the strategy shown in Fig. 1, and the results are shown in Fig. 2 (EMBL accession number X52521). To confirm the presence of the two putative introns, deduced from the DNA sequence, the sequence of the mRNA in these regions was also determined.

Primer extension experiments and nuclease S1 protection experiments (results not shown) were carried out to determine the start site for transcription of the *bphA* gene (Fig. 2). The 3' end of the mRNA was also determined by nuclease S1 experiments (Fig. 2). Two major stops were found which would result in *bphA* mRNAs of 2130 and 2138 nucleotides, corresponding well with the size deduced from Northern blot analysis (2.1 kb; result not shown).

From the combination of DNA and mRNA sequence data the amino acid sequence of BPH was deduced. Three open reading frames (ORFs) were found. The longest ORF, which presumably codes for BPH, is preceded by two very small, partially overlapping reading frames. The first ORF, which starts only 6 bases downstream from the main transcription initiation site, is 18 amino acids in length. The second ORF, which starts near the end of the first ORF is only 5 amino acids long. Both ORF's are followed in phase by more than one translation stop codon.

BPH is a member of the cytochrome P450 superfamily

The long ORF present on the mRNA of the *bphA* gene codes for a protein of 517 amino acids with a theoretical molecular weight of 58 kDa. The N-terminus of the protein is very hydrophobic, but this sequence is not followed by a signal peptidase-processing site (according to the predictions of Von Heijne 1986). A computer search for membrane-spanning fragments in BPH reveals one such fragment, the N-terminus (Eisenberg et al. 1984; Rao and Argos 1986; Klein et al. 1985).

Database searches indicate a possible relationship of BPH to monooxygenases of the cytochrome P450 superfamily. A cysteinyl-containing peptide involved in the haem-binding site (Gotoh et al. 1983) was found in the C-terminal part of the protein (stippled in Fig. 2). Furthermore, eight out of ten amino acids found in all cytochrome P450 proteins characterised so far (Nelson and Strobel 1988) are present (underlined in Fig. 2) at approximately the same positions in BPH. Also the presence of a hydrophobic N-terminus is a common feature of all eukaryotic cytochrome P450 proteins (Nelson and Strobel 1988, 1989). Comparison of the BPH sequence with the sequences of other cytochrome P450 proteins from different gene families indicated that BPH is a member of a new gene family. According to the nomenclature proposed by the authors of the P450 superfamily review (Nebert et al. 1989) this gene family is called P450LIII and the gene is designated *CYP53* (D.W. Nebert, personal communication).

Table 1. Results from the overexpression experiments of the *bphA* gene

Strain	Copy number *bphA*	Growth on benzoate	In vitro enzyme-activity	Relative BPH protein production
N204	1	+	3.1	1
N271(T16)	12	−	0.4	11
N271(T18)	10	−	0.5	8
N271(T21)	3	+/−	2.0	6

N204 (*csp21, met21*) is a *bphA* wild-type strain used as a reference; N271 (*csp21, fwn21, pdx21*) is another *bphA* wild-type strain that was used to obtain *bphA* multiple copy transformants. The copy number of the *bphA* gene was determined with spot blots using *pyrG* as a reference. Growth on benzoate was scored on minimal medium plates with 25 mM benzoate as carbon source (Boschloo et al. in preparation). Wild type growth is denoted +, intermediate growth +/−, and no growth −. The in vitro enzyme activities were measured as described in the Materials and methods and are given as units per gram dry weight. The relative BPH protein production was estimated using Western blots (Fig. 3) developed with the α-SP32 antiserum (Gerritse et al. 1990) using a laser densitometer and total protein concentration assays (BCA, Pierce). The BPH protein production in the *bphA* wild-type strain N204 is set to 1

Overexpression of the bphA *gene*

In order to increase the BPH activity, extra copies of the *bphA* gene were introduced into *A. niger* using the *amdS* transformation system (Kelly and Hynes 1985). The results of these experiments are given in Table 1. In multicopy transformants increased levels of *bphA*-specific mRNA were detected (results not shown). Using an antiserum (α-SP32) raised against a synthetic peptide derived from the amino acid sequence of BPH (Gerritse

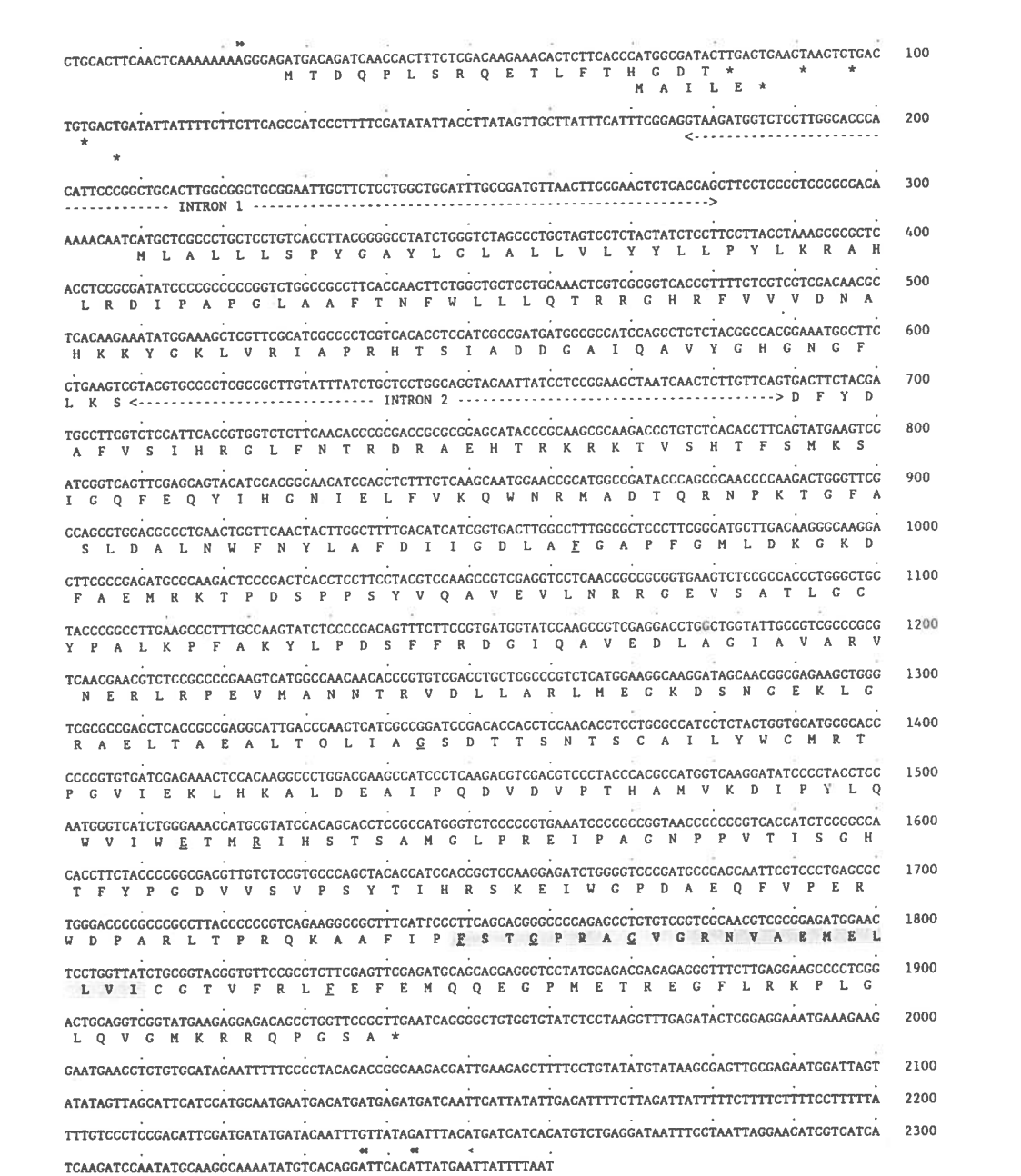


Fig. 2. DNA sequence of the mRNA coding region of the *A. niger bphA* gene. ⩾, main transcription start point; ⩽, main transcription stops; <, minor transcription stop. The introns are indicated by a *dashed line*. The three ORFs are translated using the one letter code. The eight amino acids that are identical with all other cytochrome P450 proteins are *underlined*. The cysteinyl-containing peptide involved in the haem-binding site is *stippled*

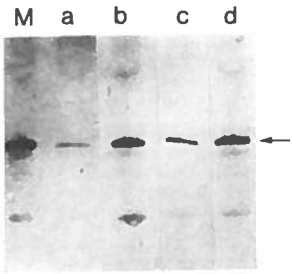


Fig. 3. Western blot of extracts (1 µl) from benzoate-induced mycelium of wild type and *bphA* multicopy transformants. The gels were run and blotted using the Phastsystem (Pharmacia LKB) and the blots were developed using the α-SP32 antiserum (Gerritse et al. 1990). Lane M, BPH marker; lane a, N204; lane b, T16; lane c, T18; lane d, T21. The *arrow* indicates the position of intact BPH. Protein concentrations were (µg/µl): N204, 1.9; T16, 3.5; T18, 1.7; T21, 4.8

et al. 1990) it could also be shown that the BPH production was increased in these transformants (Fig. 3). However, transformants with an increased copy number of the *bphA* gene invariably showed a decreased ability to grow on benzoate and also the in vitro BPH activity was strongly decreased (Table 1). Transformants with more than about 5 copies of the gene were not able to grow on benzoate at all. These results suggest that the *bphA* gene product is not the only factor necessary for BPH enzyme activity.

Discussion

Aspergillus niger, a filamentous fungus widely used in the fermentation industry, is able to grow on benzoate. The first step in the degradation of this compound is para-hydroxylation of the aromatic ring, carried out by an enzyme called benzoate-para-hydroxylase. To increase the activity of the enzyme, we have cloned the gene (*bphA*) that codes for this enzyme. In this report we describe the isolation and molecular characterisation of the gene, and the characterisation and overexpression of the encoded protein.

The DNA sequence and part of the mRNA sequence of the gene were determined. From the results it can be concluded that the gene contains two introns and codes for a protein of 517 amino acids. The structure of the 5′ region of the mRNA shows some unexpected characteristics. Upstream of the long ORF two smaller ORFs are present, both followed by several stop codons. Since the *bphA* gene is tightly regulated, these sequences might be involved in the regulation of the expression of the gene. The presence of short ORFs before the main ORF has been found in regulatory genes like the *GCN4* gene of the yeast *Saccharomyces cerevisiae* (Hinnebusch 1988). In the sequence of the acid phosphatase gene (*pacA*) of *A. niger* (MacRae et al. 1988) a structure similar to that found in the *bphA* gene has been described. The data presented by these authors reveal the presence of two ORFs, 33 and 11 amino acids long, ending just before the beginning of the long ORF. The TGA stop codon of the smaller one even overlaps the ATG start codon of the main ORF. It is not known whether the ORFs of the *bphA* gene or those of the *pacA* gene are actually translated.

Introns are also found in the 5′ untranslated regions of other genes in filamentous fungi. Both the *gpdA* genes of *A. nidulans* (Punt et al. 1988) and *A. niger* (Van Gorcom et al. in preparation) and the *egIII* gene of *Trichoderma reesei* (Saloheimo et al. 1988) have introns in this part of the gene.

Analysis of the BPH amino acid sequence shows a similarity to that of the cytochrome P450 superfamily. All characteristics of this class of proteins are present in BPH (conserved amino acids, length, membrane-spanning domain, haem-binding site). Furthermore it has been shown that in vitro BPH activity can be inhibited by CO and metyrapone (W.J.J. van den Tweel, personal communication). One striking difference from all other cytochrome P450 amino acid sequences is the absence of one or two of the ten conserved amino acids (Nelson and Strobel 1988). The first normally conserved amino acid (a glycine at position 96) might still be present in BPH. In that region of the BPH amino acid sequence several glycines are present but since the region is not very homologous to other cytochrome P450 proteins, it is difficult to assign the conserved glycine. The sixth conserved amino acid (a proline at position 488) is not present in BPH, but the surrounding region contains several amino acids also present in other cytochrome P450 proteins. Since BPH is not a member of one of the presently known gene families, it is classified as the first representative of a new gene family designated P450LIII (D.W. Nebert, personal communication).

Mutagenesis experiments in *A. niger* NV DSM 2061 indicated the presence of a single gene coding for a benzoate-para-hydroxylase (Boschloo et al. to be published). Classification of BPH as a cytochrome P450 monooxygenase is in conflict with previously published data indicating that BPH is a tetrahydropteridine containing enzyme (Reddy and Vaidyanathan 1975). The experiments presented by these authors do not exclude the possibility that BPH is a cytochrome P450 monooxygenase (absence of inhibition by CO or metyrapone) nor did the authors present a spectrum of the purified enzyme preparation. But we cannot, however, exclude the possibility that one strain of *A. niger* contains a tetrahydropteridine containing BPH whereas another contains a cytochrome P450 class of enzyme.

Overexpression of the gene in *A. niger* was achieved by increasing the copy number of the *bphA* gene. Although the production of mRNA and the BPH protein were increased, the multicopy transformants showed a decreased level of enzyme activity. Multicopy transformants also lost their ability to grow on benzoate. If the properties of cytochrome P450 systems in other eukaryotic organisms can be extrapolated to the *Aspergillus* enzyme, then the missing component could be the NADPH cytochrome P450 reductase. This membrane-bound enzyme must form a complex with cytochrome

P450 in order to transport an electron from NADPH to cytochrome P450 (for a review see Nebert and Gonzalez 1987). In mammalian cells, the NADPH cytochrome P450 reductase is present in limiting supply (Nebert and Gonzalez 1987). The negative effect of overexpression of the *bphA* gene on growth on benzoate might be caused by the titration of the NADPH cytochrome P450 reductase which is an essential component of several biosynthetic enzyme systems. So, to increase the ability of *A. niger* to convert benzoate enzymatically to para-hydroxybenzoate, it will probably be necessary to improve also the expression of the NADPH cytochrome P450 reductase in the cell.

Acknowledgements. We thank Drs. Y.M.J.T. de Ruiter-Jacobs and J. Cordewener for excellent help in some of the experiments described and Drs. F. Hermes and W.J.J. van den Tweel for critical reading of the manuscript.

References

Bos CJ, Debets AJM, Slakhorst SM, Swart K (1989) Adenine and pyrimidine genes of *Aspergillus niger* and evidence for a seventh linkage group. Curr Genet 16:307–310

Clarke L, Carbon J (1978) Functional expression of cloned yeast DNA in *E. coli*: specific complementation of arginino-succinate lyase (*argH*) mutations. J Mol Biol 120:517–532

Cordewener JHG, Busink R, Boschloo J, Visser J, Bos CJ (1987) Development of biocatalysts with monooxygenase activity: Pathway and characterisation of enzymes of benzoate degradation in *Aspergillus niger*. In: Neijssel OM, Van der Meer RR, Luyben KChAM (eds) Proceedings 4th European Congress on Biotechnology, vol 3. Elsevier, Amsterdam, pp 366

Devereux J, Haeberli P, Smithies O (1984) A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res 12:387–395

Eisenberg D, Schwarz E, Komaromy M, Wall R (1984) Analysis of membrane and surface protein sequences with the hydrophobic moment plot. J Mol Biol 179:125–142

Gerritse K, Van Gorcom RFM, Fasbender M, Lange J, Schellekens MM, Zegers ND, Claassen E, Boersma W (1990) Specific immuno detection of benzoate-para-hydroxylase with antibodies raised against specific peptides. Biochem Biophys Research Commun 167:33–39

Goosen T, Bloemheuvel G, Gysler C, De Bie DA, Van den Broek HWJ, Swart K (1987) Transformation of *Aspergillus niger* using the homologous orotidine-5'-phosphate-decarboxylase gene. Curr Genet 11:499–503

Gotoh O, Tagashira Y, Iizuka T, Fujii-Kuriyama Y (1983) Structural characteristics of cytochrome P450: Possible location of the heme-binding cysteine in determined amino acid sequences. J Biochem 93:807–817

Hamid AB, Smith JE (1987) The involvement of cytochrome P-450 monooxygenase system in aflatoxin biosynthesis by *Aspergillus flavus*. J Ind Microbiol 2:137–142

Hinnebusch AG (1988) Novel mechanisms of translational control in *Saccharomyces cerevisiae*. Trends Genet 4:169–174

Hynes MJ, Corrick CM, King JA (1983) Isolation of genomic clones containing the *amdS* gene of *Aspergillus nidulans* and their use in the analysis of structural and regulatory mutations. Mol Cell Biol 3:1430–1439

Kelly JM, Hynes MJ (1985) Transformation of *Aspergillus niger* by the *amdS* gene of *Aspergillus nidulans*. EMBO J 4:475–479

Klein P, Kanehisa M, DeLisi C (1985) The detection and classification of membrane-spanning proteins. Biochim Biophys Acta 815:468–476

MacRae WD, Buxton FP, Sibley S, Garven S, Gwynne DI, Davies RW, Arst Jr HN (1988) A phosphate-repressible acid phosphatase gene from *Aspergillus niger*: its cloning, sequencing and transcriptional analysis. Gene 71:339–348

Maniatis T, Fritsch EF, Sambrook J (1982) Molecular cloning. A laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY

Nebert DW, Gonzalez FJ (1987) P450 genes: Structure, evolution, and regulation. Annu Rev Biochem 56:945–993

Nebert DW, Nelson DR, Adesnik M, Coon MJ, Estabrook RW, Gonzalez FJ, Guengerich FP, Gunsalus IC, Johnson EF, Kemper B, Levin W, Phillips IR, Sato R, Waterman MR (1989) The P450 superfamily: Updated listing of all genes and recommended nomenclature for the chromosomal loci. DNA 8:1–13

Nelson DR, Strobel HW (1988) On the membrane topology of vertebrate cytochrome P450 proteins. J Biol Chem 263:6038–6050

Nelson DR, Strobel HW (1989) Secondary structure prediction of 52 membrane-bound cytochromes P450 shows a strong structural similarity to $P450_{cam}$. Biochemistry 28:656–660

Pontecorvo G, Roper JA, Hemmons LJ, MacDonald KD, Bufton AWJ (1953) The genetics of *Aspergillus nidulans*. Adv Genet 5:141–238

Punt PJ, Oliver RP, Dingemanse MA, Pouwels PH, Van den Hondel CAMJJ (1987) Transformation of *Aspergillus* based on the hygromycin B resistance marker from *Escherichia coli*. Gene 56:117–124

Punt PJ, Dingemanse MA, Jacobs-Meijsing BJM, Pouwels PH, Van den Hondel CAMJJ (1988) Isolation and characterization of the glyceraldehyde-3-phosphate dehydrogenase gene of *Aspergillus nidulans*. Gene 69:49–57

Rao JKM, Argos P (1986) A conformational preference parameter to predict helices in integral membrane proteins. Biochim Biophys Acta 869:197–214

Reddy CC, Vaidyanathan CS (1975) Purification, properties and induction of a specific benzoate-para-hydroxylase from *Aspergillus niger*. Biochim Biophys Acta 384:46–57

Saloheimo M, Lehtovaara P, Penttilä M, Teeri TT, Ståhlberg J, Johansson G, Claeyssens M, Tomme P, Knowles JKC (1988) EGIII, a new endoglucanase from *Trichoderma reesei*: the characterisation of both gene and enzyme. Gene 63:11–21

Towbin H, Staehelin T, Goordon J (1979) Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. Proc Natl Acad Sci USA 76:4350–4354

Van Gorcom RFM, Punt PJ, Pouwels PH, Van den Hondel CAMJJ (1986) A system for the analysis of expression signals in *Aspergillus*. Gene 48:211–217

Van Gorcom RFM, Van den Hondel CAMJJ (1988) Expression analysis vectors for *Aspergillus niger*. Nucleic Acids Res 16:9052

Van Hartingsveldt W, Mattern IE, Van Zeijl CMJ, Pouwels PH, Van den Hondel CAMJJ (1987) Development of an homologous transformation system for *Aspergillus niger* based on the *pyrG* gene. Mol Gen Genet 206:71–75

Von Heijne G (1986) A new method for predicting signal sequence cleavage sites. Nucleic Acids Res 14:4683–4690

Walsh C (1978) Chemical approaches to the study of enzymes catalyzing redox transformations. Annu Rev Biochem 47:881–931

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

Yelton MM, Hamer JE, Timberlake WE (1984) Transformation of *Aspergillus nidulans* using a *trpC* plasmid. Proc Natl Acad Sci USA 81:1470–1474

Yelton MM, Timberlake WE, Van den Hondel CAMJJ (1985) A cosmid for selecting genes by complementation in *Aspergillus nidulans*: selection of the developmentally regulated *yA* locus. Proc Natl Acad Sci USA 82:834–838

Communicated by W. Gajewski

CHAPTER 6

Analysis of the effect of upstream open reading frames on the expression of the *Aspergillus niger bphA* gene

Robert F.M. van Gorcom, Peter J. Punt and Cees A.M.J.J. van den Hondel

ABSTRACT

The *Aspergillus niger bphA* gene encoding the strongly inducible, highly expressed cytochrome P450 benzoate para-hydroxylase contains two small overlapping upstream open reading frames (uORFs) just downstream of the 5' end of the mRNA. By fusion of the start codons of these uORFs to *lacZ* and measurement of β-galactosidase activities of benzoate-induced and non-induced transformants, it is shown that the first AUG codon (located 6 bp downstream of the transcription start point) is hardly used as a start point for translation. The AUG codon of the second uORF appears to be used with about the same efficiency as the AUG of the main ORF. In line with the absence of substantial β-galactosidase expression of the *lacZ* fusion to the first upstream start codon, we found that mutation of this AUG codon in a vector where *lacZ* is fused to the AUG codon of the main open reading frame, did not result in significantly different expression levels. Mutation of the AUG codon of the second uORF and combination of both mutations resulted only in a small, and in most cases non-significant, increase of *lacZ* expression levels both in a wildtype strain and in a strain in which the chromosomal region encoding the *bphA* uORFs was deleted.

Under the growth and induction conditions used in our experiments we can not find a clear effect of translation of the uORFs (*cis* effect), nor of the peptides encoded by the two uORFs (*trans* effect) on the regulation of expression of the *bphA* gene.

INTRODUCTION

Fungi, one of the most important classes of biodegrading organisms in nature, are able to modify and metabolize a vast array of substances. This ability is exemplified by the production by fungi of a large range of biopolymer-degrading enzymes, like cellulases (and other sugar polymer degrading enzymes), ligninases and proteinases. Furthermore, fungi produce a large diversity of primary and secondary metabolites, some of which are used to increase the bioavailability of the substrates on which they feed. Other metabolites are involved in strategies to invade and/or kill their future nutrient source and/or compete with other microorganisms that can make use of the same nutrient sources. Since fungi encounter a large range of threats in nature, they also developed a large spectrum of enzymes that are able to detoxify many kinds of substances.
Man started to use some of their capacities already a few thousand years ago in food fermentation processes. More recently we have learned that many of these abilities can be commercially exploited. As examples can be mentioned the production of antibiotics, citric acid or other primary and secondary metabolites, the production of polymer degrading enzymes and the hydroxylation of steroids and other biotransformation processes.
We are interested in the role cytochrome P450 enzyme systems play in biosynthetic and biotransformation processes and in the use of this class of enzymes for industrial and medical purposes. Previously we have cloned a gene from the fungus *Aspergillus niger* encoding benzoate para-hydroxylase (*bphA*), which is a member of the cytochrome P450 superfamily (*cyp53A1*; van Gorcom et al. 1990). We use this gene to study the possibilities for improving cytochrome P450 mediated biocatalytic reactions by fungal cultures (van den Brink et al. 1996a, 1996b).
The fungus *A. niger* possesses many enzyme systems involved in the degradation or detoxification of a large spectrum of compounds. Therefore a strict regulation of their synthesis and/or activity is to be expected. The coordinate regulation at the transcriptional level of the expression of the two genes encoding the components of the enzymatic complex that performs the para-hydroxylation of benzoate, the *bphA* gene and the *cprA* gene, encoding NADPH cytochrome P450 oxidoreductase (van den Brink et al. 1995), is currently under study in our laboratory (van den Brink et al. 1997). The mRNA of both genes (*bphA* both the benzoate-induced and non-induced mRNA, *cprA* only the benzoate-induced mRNA) contains one (*cprA*) or two (*bphA*) small open reading frames in its 5' region indicating that regulation of expression at the translational level might also be important.

Small open reading frames present within mRNA molecules before the start of the main protein encoding region(s) are relatively rare, but -if present- often found in mRNA encoded by genes of which regulation is complex, such as genes involved in amino acid biosynthesis or cell differentiation. They are, however, also found in mRNAs encoded by genes whose regulation seems less critical for the cell. In prokaryotes these upstream open reading frames (uORFs), whose encoded peptides are usually referred to as 'leader peptides', often have a role in gene regulation at the transcriptional level. In eukaryotes some of these uORFs have been shown to

be involved in regulation of gene expression at the translational level (for a recent overview see Geballe and Morris 1994).

Research on eukaryotic microorganisms has, for a few uORF containing genes, resulted in the (partial) elucidation of the way these uORFs are involved in the regulation of translation of the protein encoding part of the gene. The best studied example in lower eukaryotes is the gene *GCN4* of the yeast *Saccharomyces cerevisiae*, encoding a positive regulator for amino acid biosynthesis (for a review see Hinnebusch 1990). On the mRNA of this gene 4 uORFs, each only a few amino acid codons long, are present. Regulation of expression of *GCN4* mainly takes place at the post-transcriptional level and the (four) uORFs play a crucial role in this regulation. Deletion or mutation of the translation start codons of these uORFs is sufficient for distortion of the regulation of *GCN4* expression (Mueller and Hinnebusch 1986). The putative peptides encoded by these uORFs do not seem to be involved in this regulation as their sequence can be replaced by other sequences without changing the pattern of *GCN4* regulation (Mueller et al. 1988; Williams et al. 1988). Reinitiation of ribosomes on *GCN4* mRNA, mediated by sequences around the stop-codon of the most 5' uORF that prevent dissociation of the 40S ribosomal subunit from the mRNA, is crucial for translation of the protein encoding part of the *GCN4* mRNA (no reinitiation --> no *GCN4* expression; Abastado et al. 1991; Grant and Hinnebusch 1994). Regulation of *GCN4* expression is mediated by changes in the efficiency of this reinitiation process caused by differences in the availability of the translation initiation factor eIF-2 (Abastado et al. 1991; Grant et al. 1994).

The *CPA1* gene of *S. cerevisiae*, encoding the small subunit of arginine-specific carbamoyl phosphate synthetase, illustrates that also the peptides encoded by uORFs can play a role in regulation of gene expression (Werner et al. 1985). As a response to the arginine concentration in the growth medium, *CPA1* gene expression is mainly regulated at the post-transcriptional level. Mutation of the AUG codon of the uORF, present in the 5' region of the mRNA encoded by this gene, leads to constitutive *CPA1* gene expression. In contrast to *GCN4*, the peptide encoded by the uORF plays an important role, since modification (even of a single amino acid) of the uORF peptide has the same effect (constitutive expression) as abolishing uORF expression by mutation of its AUG codon (Werner et al. 1987). Furthermore, it was shown that this uORF confers its regulatory properties to another gene when placed in the 5' upstream mRNA encoding region of that gene (Delbecq et al. 1994).

Upstream ORFs are also found in the *GCN4* and *CPA1* homologues in the filamentous fungi *Neurospora crassa* and *Aspergillus niger* and are thought to play a similar role in the regulation of expression of these genes (*N. crassa cpc-1* (*GCN4*), Paluh et al. 1988; *A. niger cpcA* (*GCN4*), Wanke et al. 1997; *N. crassa arg-2* (*CPA1*), Orbach et al. 1990). The peptide encoded by the uORF of *arg-2* is 63% similar to the uORF encoded peptide of *CPA1*, and its role in mediating arginine-specific translational control, has recently been shown (Luo et al. 1995, Luo and Sachs 1996).

The role of an uORF in the regulation of expression of the *Aspergillus nidulans brlA* gene, a regulatory gene involved in cell differentiation (conidiophore development), has been studied by Han et al. (1993). It has been shown that mutation of the AUG codon of the uORF, that is present

on only one of the two different transcripts of the *brlA* gene, results in strongly increase translation of the main ORF from that particular mRNA, yielding strains that show prematur development of conidiophores (Han et al. 1993).
In several other genes in filamentous fungi, uORFs are present but in most cases it is not know whether, and if so in which way, these uORFs are involved in regulation of gene expression (e.g *Aspergillus niger pacA*, MacRae et al. 1988; *Trichoderma viride pgk*, Goldman et al. 1992 *Aspergillus nidulans stuA*, Miller et al. 1991; *Ustilago violacea γ-tub*, Luo and Perlin 1993; see also chapter 7).

The *bphA* gene of *A. niger* contains two uORFs located close to the 5' end of the mRNA, ± 12 bp upstream of the start of the large ORF encoding BPH (van Gorcom et al. 1990). To analyz whether these uORFs are translated *in vivo*, we have fused a reporter gene (*lacZ*) to th translation start codons of both uORF encoding sequences and compared *lacZ* expression of th resulting reporter strains to that of a strain in which the *lacZ* gene was fused to the translatio start codon of the BPH encoding reading frame. Furthermore, we have analyzed the effects c mutation of the translation start codons of the uORFs on the expression of the reporter gen fused to the translation start codon of the main ORF both in a wildtype and in a *bphA* deletio strain.

IATERIALS AND METHODS

trains and plasmids

. *niger* strain N245 (ATCC 1015, *csp21, metB21, pyrG*) used as a recipient in transformation xperiments was obtained from Drs. J.G. Boschloo and C.J. Bos (Agricultural University Vageningen, NL). Vectors were constructed and propagated in *E. coli* K-12 JM109 (Yanisch-'erron et al. 1985) and DH5α (GibcoBRL, Life Technologies, Gaithersburg MA, USA). Uracil-ontaining DNA templates were prepared in *E. coli* K-12 BW313 (Kunkel 1985). Bacteriophage 113mp11 has been described by Messing (1983), plasmids pAB8-2 by van Gorcom et al. (1990), AB94-12 and -13 by van Gorcom and van den Hondel (1988), pAB94-53 by van Gorcom et al. 1993), pAN5-d1 by Punt et al. (1990) and pAN7-1 by Punt et al. (1987).

ransformations, growth and induction conditions

. *coli* transformations were carried out according to Hanahan (1983) and M13 transfections ccording to Kunkel (1985). *A. niger* transformation was carried out according to van lartingsveldt et al. (1987). Complete medium (CM) was as described by Pontecorvo et al. 1953). Minimal medium (MM) contained 5.24 g KH_2PO_4, 5.24 g K_2HPO_4, 1.04 g NaCl, 0.25 g $MgSO_4$, 0.22 g urea, 0.1 g methionine and 2.44 g uridine per litre. For the induction of *bphA* xpression, 200 ml CM cultures were inoculated with 10^5 spores/ml and cultivated overnight at 0°C at 300 rpm. Mycelium was filtered through Miracloth (Calbiochem) and washed with MM. .0 g mycelium (w.w.) was transferred to a flask with 125 ml induction medium (MM + inducer t the given concentration) and incubated for 4 hours at 30°C at 300 rpm. The resulting mycelium vas used for the analysis of β-galactosidase activity.

onstruction of a *bphA* deletion strain

o be able to distinguish between *cis* and *trans* effects of the uORFs, a strain was constructed which a large part of the transcription control region and the 5' part of the *bphA* gene had been leleted. This deletion was made by replacement of a 2.7 kb *Bgl*II fragment containing 1 kb of the *phA* 5' flanking region and 1.7 kb of the *bphA* gene (van Gorcom et al. 1990) in the genome of *A. niger* N245 (Fig. 1) by a 4.0 kb fragment from plasmid pAN7-1 (Punt et al. 1987), containing he *E. coli hph* gene flanked by *Aspergillus* expression signals (hygromycin resistance cassette). The expected genotype of the resulting strain, N245Δ*bphA*, was verified both by PCR analysis nd Southern blotting. Growth tests showed that this strain had lost its ability to grow in enzoate.

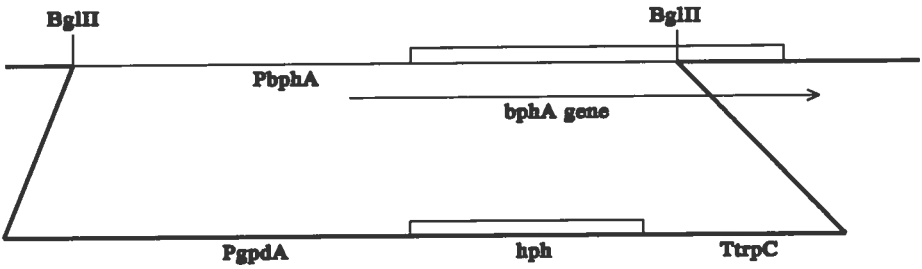


Figure 1: Schematic representation of the replacement (by double cross-over) of a 2.7 kb *Bgl*II fragment, containing a large part of the *Aspergillus niger bphA* gene, by the *gpdA-hph* fusion gene (hygromycin resistance cassette) present on an about 4 kb *Bgl*II-*Hin*dIII (*Hin*dIII end modified to *Bgl*II end by a *Hin*dIII-*Bgl*II adapter) fragment isolated from pAN7-1 (Punt et al. 1987) in strain N245 resulting in strain N245Δ*bphA*. The 4 kb hygromycin resistance cassette was cloned in pAB8-2, from which at the same time the 2.7 kb *Bgl*II fragment was deleted, in order to provide appropriate flankings for the double cross-over. Upper line: situation as present in wild-type *A. niger* genome. Lower line: schematic presentation of the *gpdA-hph* fusion gene. Thick line: situation as present in the disrupted strain N245Δ*bphA*.

β-galactosidase activity determination

β-Galactosidase activities were determined using mycelium which, after a wash step, was immediately frozen in liquid nitrogen. The β-galactosidase activity was determined in extracts prepared by grinding of mycelium in liquid nitrogen, as described before (van Gorcom et al. 1985). Protein concentrations of the extracts were determined using the BIORAD assay with BSA as a standard. All activities are calculated as units activity per mg protein.

Molecular methods

Fungal DNA isolations were performed as described by Kolar et al. (1988). Unless indicated, all other methods were essentially as described by Sambrook et al. (1989).

DNA sequence analysis of the *A. niger bphA* transcription control region

The DNA sequence of about 1.6 kb upstream of the transcription start point of the *A. niger bphA* gene was determined and is presented in Fig. 2A. Analysis of this region identified an open reading frame (bases 86 - 1123) that showed clear homology with 3' parts of the *Schizosaccharomyces pombe C56F8.17C* gene with an unknown function (50% identity), and the *S. cerevisiae SNM1* gene, encoding a DNA cross-link repair protein (33% identity; the *SNM1* gene and the *C56F8.17C* gene are 30 % identical). This open reading frame is clearly overlapping with the transcription control region of the *bphA* gene (van den Brink et al. 1997). In the DNA sequence presented in Fig. 2A no strong secondary structures or large direct repeats are found. The sequence has been deposited in the EMBL DNA data library and can be found with accession number A22974.

Induction of the *bphA* transcription control region

Northern blots, containing mRNA isolated from *A. niger* mycelium which was grown in the presence of 0.1% benzoate as sole carbon source, probed with a fragment containing the *bphA* gene, revealed a strongly expressed mRNA of about 2.1 kb (not shown). The intensity of the band was comparable to that of mRNA of the highly expressed *A. niger gpdA* gene. The *bphA* mRNA could not be detected in mycelium grown in 1% glucose containing medium, indicating an efficient regulation of *bphA* gene expression at the transcriptional level. In mycelium grown in medium with both 0.1% benzoate and 1% glucose hardly any *bphA* mRNA could be detected, suggesting that the expression of the *bphA* gene is subject to carbon catabolite repression (not shown).

Subsequent experiments revealed that para-amino-benzoate (PABA) is also a good inducer of the *A. niger bphA* gene. Optimal induction was obtained at concentrations of 0.1% benzoate or 0.3% PABA. At a higher benzoate concentration benzoate inhibits the growth of the *A. niger* strain used in these experiments.

Construction of reporter plasmids

As described before (van Gorcom et al. 1990; see Fig. 2A), the 5' untranslated region of the *A. niger bphA* gene contains two small upstream open reading frames (uORFs). To analyze whether the translation start codons of the uORFs of the *bphA* gene are used in vivo, three reporter vectors were made in which the AUG codons of the two uORFs (AUG_1 and AUG_2) and that of the benzoate para-hydroxylase (BPH) encoding ORF (AUG_3) were fused to a reporter gene (*lacZ*). The DNA sequence around the transcription and translation start points is shown in Fig. 2A. In the vectors about 1.6 kb of genomic DNA 5' to the transcription start point is present. A map of the resulting plasmids (pAB94-84 (fusion at AUG_1), pAB94-85 (fusion at AUG_2) and pAB94-83 (fusion at AUG_3)) is shown in Fig. 3 and the DNA sequence at the fusion points is shown in Fig. 2B.

A

```
GTCGACCGGAAGTGGCTTGTACCACTTGAGTTCGAGCGGAAGACGGAAATCCCGGATACAGGAGGAGCGC    70
AGGTGACTTTGATCGATGGCTAATCATTGTCCTGGGAGCGCCATCTTTCTCTTCGAGAAATCAATGGGAT   140
CGGGTCCCTCGCAGAGAACACATCGTGTCCTCCACTGTGGTGACTTTCGCGCCTCGCCGCTTCATGTGCA   210
ACATGCCCTTCCCCGCCCGGAGATTGCTGACCCCGCAACCGGCAAGGCTCGCCAGCAACGAATCGATGCC   280
TGCTATCTGGACACTACATATTTGAGCCCCAAGTATGCATTCCCTGGCCAGGAAGATGTCATACAAGCCT   350
GCGCAGAACTTTGCGTTGAGCTCGATGGGGACGCCAACGACACCAATGGACGAGCATTTGGACGACCAGT   420
CAATGGAAAAAGCGGAATGCTGAGCAAGTTTGTTACGGCTGTGACTGGATCCCGCCCGTCTCCGACGCAA   490
GACAGCCGCCCCCCTGGCCGGCTATTGGTAGTAATAGGGACGTACAGCATCGGCAAAGAACGCATCTGTC   560
TGGGGATCGCACGGGCATTGAAGAGCAAGATCTACGCGACGCCAGCTAAGCAGCGCGTCTGTGCGTGCCT   630
CGAGGATGCTGAGCTGTCATCGCTGCTGACAGACGATCCCACGGAGGCGCAGGTGCATATGCAAACGCTA   700
TTCGAGATCCGGGCGGAAACGCTGGCGGATTACCTGGACTCGATGAAGCCGCACTTCACGCGGGTGGTGG   770
GATTTCGACCAACCGGGTGGACGTATCGCCCGCCAGCTGGCCGAATGCTGGACAACCCACCGGTGTCGGT   840
GGTGCTCAATTCGGCACATTGGAAGACGCCCTTTTCTGCGAAAGACCTGGTGCCACAGCGAGGGAGTACG   910
CGGGAAAGCGCATGCTTTGGAGTGCCGTACAGTGAGCACAGCTCATTTCGGGAGTTGAGCATGTTCTGCT   980
GCGCACTCCGGATCGGACGGGTGATCCCGACAGTGAACGTAGGTAGCCGGAAAAGTCGGGAGCGCATGAA  1050
GGCGTGGATTGAGCGATGGGAGGCGGAGAAGCGGAAGAATGGGTTGTACCGCGTGGAGGGGAATAGCTGG  1120
TAGGGAAGGGAATAGATGGCTCTACCAATGTCCAAAGTACTGGTGGAACAGAAGGATCAGAAGGATTGCG  1190
AAAGACGGGTCGGAACATGATGCCTAATAGAGTAAGTAAGGAGTTGGTGCTGTAACTAGTCACAAGTTAC  1260
GAGTTGTGTACATAACATCATTAGTCATGAAGATCAATTGCCTTTATGCTTCCGTAACTCTCGCCTCCCC  1330
GGAGTCACGAGATCAATAGAAACCACCGCCGTTGACCATTCGCGATGCTCTCACTGGCTGTATGCTGTCG  1400
ATAGCCATGGAGCCGTTCAAAGTATGGACCTTTGGGTGAGGATCTCCCTCCAACCCCACGGGACGTACAC  1470
GAACAACCGAGCAGAGGCGGGGGAGGGCAAAGAGGCCGGCGCTGCAAATCGGCTGGCAGATCAGTCGCGG  1540
CTCAGCAGAGACTCCCGATTTTCCCTTCCGTTGCCTGGCTTTGCCTCGGGGTTCGAGAGGAGCCCGTCTG  1610
                                    #
CCATAAATAAGCCTGCACTTCAACTCAAAAAAAAGGGAGATGACAGATCAACCACTTTCTCGACAAGAAA  1680
                                        M1 T  D  Q  P  L  S  R  Q  E  T

CACTCTTCACCCATGGCGATACTTGAGTGAAGTAAGTGTGACTGTGACTGATATTATTTTCTTCTTCAGC  1750
  L  F  T  H  G  D  T  *
           M2 A  I  L  E  *

CATCCCTTTTCGATATATTACCTTATAGTTGCTTATTTCATTTCGGAGgtaagatggtctccttggcacc  1820
                                                   <----------------------

cacattcccggctgcacttggcggctgcggaattgctttctcctggctgcatttgccgatgttaactttccg  1890
---------- intron ------------------------------------------------------

aactctcaccagCTTCCTCCCCTCCCCCCACAAAAACAATCATGCTCGCCCTGCTCCTGTCACCTTACGG  1960
----------->                                M3 L  A  L  L  L  S  P  Y  G

GGCCTATCTGGGTCTAGCCCTGCTAGTCCTCTACTATCTCCTTCCTTACCTAAAGCGCGCTCACCTCCGC  2030
 A  Y  L  G  L  A  L  L  V  L  Y  Y  L  L  P  Y  L  K  R  A  H  L  R
```

`#:` transcription startpoint
M_1: ATG_1 = translation startcodon of the first uORF
M_2: ATG_2 = translation startcodon of the second uORF
M_3: ATG_3 = translation startcodon of the main ORF encoding BPH

B

```
ccccacaaaaacaatcatg1TCGACGAATTCGGATCC              pAB94-83

ctcaaaaaaaagggagatg1TCGACGAATTCGGATCC              pAB94-84

gaaacactcttcacccatg2TCGACGAATTCGGATCC              pAB94-85
```

C

```
                                    #
GTCTGCCATAAATAAGCCTGCACTTCAACTCAAAAAAAAtcttcagtatattcatcttcccatccaagaa   70
cctttatttcccctaagtaagtactttgctacatccatactccatccttcccatcccttattcctttgaa  140
cctttcagttcgagctttcccacttcatcgcagcttgactaacagctaccccgcttgagcagacatcacc  210
ATGgatcctgtcgacgaattcggatccgtcgacctgcag ---> rest of lacZ
```

D

```
#                                                    NcoI
GGGAGATGACAGATCAACCACTTTCTCGACAAGAAACACTCTTCACCCATGGCGATACTTGAGTGA   pAB94-83
     M1 T  D  Q  P  L  S  R  Q  E  T  L  F  T  H  G  D  T  *
                                                      M2 A  I  L  E  *

#   AccI                                             NcoI
GGGAGtaGACAGATCAACCACTTTCTCGACAAGAAACACTCTTCACCCATGGCGATACTTGAGTGA   pAB94-91
                                                      M2 A  I  L  E  *
#
GGGAGATGACAGATCAACCACTTTCTCGACAAGAAACACTCTTCACCCAcGGCGATACTTGAGTGA   pAB94-92
     M1 T  D  Q  P  L  S  R  Q  E  T  L  F  T  H  G  D  T  *

#   AccI
GGGAGtaGACAGATCAACCACTTTCTCGACAAGAAACACTCTTCACCCAcGGCGATACTTGAGTGA   pAB94-93
```

Figure 2: (A) DNA sequence of the region 5' of the *Aspergillus niger bphA* gene and the 5' region of the *bphA* gene, including the position of the first intron (lower case), the position of the main start point of transcription (#) and the positions of the two uORFs and the start of the main ORF; EMBL data library accession number: A22974; (B) DNA sequence of the region in plasmids pAB94-83, -84, and -85 where the fusions at the AUG codon (83: AUG_3, 84: AUG_1, 85: AUG_2) between the *bphA* 5' region and *lacZ* were made; (C) partial DNA sequence of pAB94-86 (transcription-start fusion *bphA-[gpdA-lacZ]*) around the fusion site: the position of the presumed start point of transcription (#) and the start point of translation (underlined) are indicated, the 5' untranslated region of the *A. nidulans gpdA* gene is presented in lower case; (D) DNA sequence of the region in plasmids pAB94-83, -91, -92 and -93 in which the mutations (mutant bases in lower case) resulting in the deletion of the AUG's of the uORFs have been made.

To be able to analyze transcriptional control of the *bphA* gene without interference of elements present on the *bphA* mRNA (and/or the DNA encoding the *bphA* mRNA) also a fusion at the transcription start point (TSP) with a *gpdA-lacZ* fusion gene was made, resulting in plasmid pAB94-86. This vector contains the same *bphA* transcription control region as present in the other vectors but the DNA sequence from the start point of transcription to the start point of translation of the *bphA* gene has been replaced by the similar region of the *A. nidulans gpdA* gene. The DNA sequence of pAB94-86 in this region is shown in Fig. 2C.

Analysis of *lacZ* expression in reporter-plasmid transformants

To determine the *lacZ* expression by the reporter genes described above, *A. niger* N245 was transformed with DNA of plasmids pAB94-83, -84, -85 and -86. Transformants with one copy of the plasmids integrated at the *pyrG* locus of the host were identified by Southern blot analysis and used for further studies. β-Galactosidase expression was determined in these transformants after growth on complete medium and transfer to, and subsequent incubation for 4 hours in, minimal medium containing various inducers (para-hydroxy-benzoate, para-amino-benzoate or benzoate) or, as a non-induced control, again in CM. The results of these experiments are shown in Table 1.

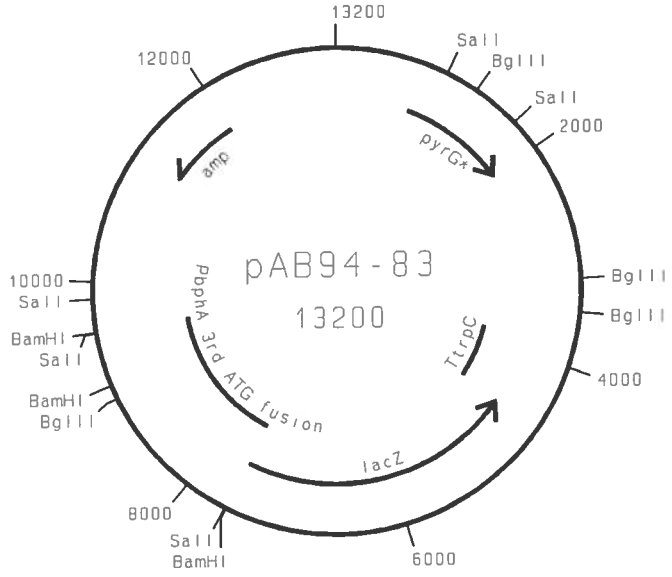


Figure 3: Example of the reporter plasmids used in this study. The construction of the plasmids used in this study is described below. A 2.1 kb *Sal*I fragment isolated from pAB8-2 (van Gorcom et al. 1990) containing the transcription control region and 5' part of the *A. niger bphA* gene was cloned in the *Sal*I site of M13mp11 (Messing 1983) resulting in vector mAB8-81. In this vector the AUG codons of the three ORFs present in the *bphA* gene were placed in front, in phase, of *lacZ*α present in this vector using *in vitro* mutagenesis (Kunkel 1985). This resulted for AUG_1 in vector mAB8-84 (oligo: GGATCCTCTAGAGTCGACATCTCCC-TTTTTTTTGAG), for AUG_2 in vector mAB8-85 (oligo: GGATCCTCTAGAGTCGACATGGGTGAAGAGTG-TTTC) and for AUG_3 in vector mAB8-83 (oligo: GGATCCTCTAGAGTCGACATGATTGTTTTTGTGGGG, all 5' ---> 3', complementary to + strand). The *Sal*I fragment, containing the transcription control region of the *bphA* gene and the first few codons of *lacZ*, of mAB8-83 was cloned in the *Bam*HI site of pAB94-12 (van Gorcom and van den Hondel 1988) using a *Bam*HI-*Sal*I adapter (GGATCCGAATTCGTCGAC). The resulting plasmid was called pAB94-83 (AUG_3-fusion). The *Bam*HI fragments from mAB8-84 and mAB8-85 containing most of the *bphA* transcription control region and the fusion of AUG_1 resp. AUG_2 to *lacZ* were cloned in pAB94-83 to replace the similar fragment that was removed by partial *Bam*HI digestion. This resulted in plasmids pAB94-84 (AUG_1 fusion) and pAB94-85 (AUG_2 fusion). For the construction of the transcription-start fusion a 210 bp *Pvu*II-*Bam*HI fragment from pAN5-d1 (Punt et al. 1990), containing the 5' region of an *A. nidulans gpdA-E. coli lacZ* fusion (including some bp from the *gpdA* promoter and a few codons of the fusion gene), was cloned in mAB8-81 digested with *Sma*I and *Eco*RI using a *Bam*HI-*Eco*RI adapter (GGATCCTGTCGACGAATTC) resulting in vector mAB8-82. Using *in vitro* mutagenesis a correct fusion at the transcription start point of the *bphA* gene and the transcription start point of the *gpdA-lacZ* fusion was made (oligo: GATGAATATACTGAAGATTTTTTTTGAGTTGAAGTGC, 5' ---> 3', complementary to + strand) resulting in vector mAB8-86. The *Sal*I fragment of mAB8-86 containing the *bphA* transcription control region and the 5' part of the fusion gene, including a few codons of *lacZ*, was cloned in the *Bam*HI site of pAB94-13 (van Gorcom and van den Hondel, 1988) using a *Bam*HI-*Sal*I adapter (GGATCCGAATTCGTCGAC). The resulting plasmid was called pAB94-86. The DNA sequence of essential regions of all plasmids used in this study has been verified in the final vectors.

Growth medium Strain: N245	plasmid used	fusion type	relative activity
Complete medium	-	none	<0.1
	pAB94-83	3rd AUG	2.2
	pAB94-85	2nd AUG	1.0
	pAB94-84	1st AUG	1.1
	pAB94-86	transcr.f.	2.3
MM + para-hydroxy-benzoate (0.1%)	-	none	<0.1
	pAB94-83	3rd AUG	6.4
	pAB94-85	2nd AUG	13.7
	pAB94-84	1st AUG	1.5
	pAB94-86	transcr.f.	20.1
MM + para-amino-benzoate (0.3%)	-	none	<0.1
	pAB94-83	3rd AUG	74
	pAB94-85	2nd AUG	89
	pAB94-84	1st AUG	2.2
	pAB94-86	transcr.f.	83
MM + benzoate (0.1%)	-	none	<0.1
	pAB94-83	3rd AUG	**100**
	pAB94-85	2nd AUG	139
	pAB94-84	1st AUG	2.6
	pAB94-86	transcr.f.	118

Table 1: Relative β-galactosidase activities determined in extracts of *A. niger* N245 transformants containing a *lacZ* reporter construct. Activity was determined as described in the Materials and Methods section. As 100% value the activity of *A. niger* N245::pAB94-83 grown in MM with 0.1% benzoate is used. This transformant gives an activity of about 28,000 units/mg protein. The σ^n of all values in this table is lower than 20%, except for values below 5.

All transformants, except those containing pAB94-84, showed considerable levels of β-galactosidase activity under inducing conditions, indicating that apart from the AUG start codon of BPH (AUG_3) also the AUG codon of the second uORF (AUG_2) can initiate translation very efficiently in vivo. The best inducer for *bphA* expression, as far as tested, is benzoate. Also para-amino-benzoate is a good inducer. Induction by para-hydroxy-benzoate, the product of the reaction catalyzed by BPH, is clearly less efficient. In both benzoate and para-amino-benzoate induced cultures the β-galactosidase levels are similar for all reporter strains, except pAB94-84 transformants. When para-hydroxy-benzoate is used as inducer significant differences between pAB94-83, -85 and -86 transformants were observed. Each value presented in Table 1 is the result of at least two independent experiments. Furthermore, for each plasmid two one-copy transformants have been tested which all gave similar results.

Mutation of the uORF AUG codons in pAB94-83

To analyze the effects of translation from the upstream AUG codons on the expression of th main open reading frame, we have mutated these AUG codons (AUG_1 and AUG_2) in the reporte plasmid in which *lacZ* was fused to the AUG codon (AUG_3) of the benzoate para-hydroxylas encoding ORF (pAB94-83). This resulted in plasmids pAB94-91 (mutation of AUG_1), pAB94-9 (mutation of AUG_2) and pAB94-93 (mutation of AUG_1 and AUG_2). The mutation of AUG_2 has bee carried out in such a way that the amino acid sequence encoded by the frame starting fror AUG_1, which overlapped with AUG_2, was unchanged. Furthermore, we introduced or remove specific restriction sites around these AUG codons enabling an easy identification of the correc mutations (Fig.2D).

Growth medium	plasmid used	AUG mutated	relative activity N245	relative activity N245Δ*bphA*
Complete medium	pAB94-83	none	1.6	1.6
	pAB94-91	1st AUG	1.4	1.7
	pAB94-92	2nd AUG	1.5	1.7
	pAB94-93	1st+2nd AUG	1.8	1.6
MM + para-hydroxy-benzoate (0.1%)	pAB94-83	none	6.5	4.9
	pAB94-91	1st AUG	6.3	6.3
	pAB94-92	2nd AUG	8.3*	7.4
	pAB94-93	1st+2nd AUG	8.5	8.0
MM + para-amino-benzoate (0.3%)	pAB94-83	none	74	77
	pAB94-91	1st AUG	67	85
	pAB94-92	2nd AUG	124	108*
	pAB94-93	1st+2nd AUG	118	124*
MM + benzoate (0.1%)	pAB94-83	none	**100**	97
	pAB94-91	1st AUG	60	110
	pAB94-92	2nd AUG	121*	162
	pAB94-93	1st+2nd AUG	134	165

Table 2: Relative β-galactosidase activities determined in extracts of *A. niger* N245 and N245Δ*bphA* transformants containing a *lacZ* fusion at the AUG codon of the main ORF and in which the AUG codons of the uORFs have been mutated. Activity has been determined as described in the Materials and Methods section. As 100% value the activity of *A. niger* N245::pAB94-83 grown in MM with 0.1% benzoate is used. This transformant gives ar activity of about 28,000 units/mg protein. The σ^n of all values in this table is lower than 20%, except for values below 5 u/mg and for N245::pAB94-92/0.1% benzoate (23%), N245Δ*bphA*::pAB94-92/0.3% para-amino-benzoate (34%), N245Δ*bphA*::pAB94-93/0.3% para-amino-benzoate (38%), N245:: pAB94-92/0.1% para-hydroxy-benzoate (28%). These values are marked with an asterix.

A. niger N245 and N245Δ*bphA* were transformed with plasmids pAB94-83, -91, -92 and -93. N245Δ*bphA* was used to analyze the effect of the mutation of the AUG codons of the uORFs ir

ne copy of plasmids integrated at the *pyrG* locus were identified by Southern blotting, making se of the changes in restriction pattern which were created by modification of the AUG_1 and AUG_2 codons. In these transformants *lacZ* expression was determined after cultivation in the resence of different inducers (Table 2).

Transformants of both strains with the same plasmid gave almost identical β-galactosidase ctivity levels. Mutation of AUG_1 did not have a clear effect on the β-galactosidase expression evel. In the wildtype strain (N245) we found a small decrease of activity under all conditions when the AUG_1 was mutated, whereas in the strain which did not contain the 5' region of the chromosomal *bphA* gene (N245Δ*bphA*) we found a small increase of activity under all conditions with the same construct. Removal of the translation start codon of the second uORF (AUG_2) esulted, either alone or in combination with mutation of AUG_1, under all induction conditions in a small (but in most conditions non-significant) increase in the β-galactosidase expression nitiated at the start codon of benzoate para-hydroxylase (AUG_3).

DISCUSSION

To investigate the role of the upstream open reading frames present in the 5' part of the *Aspergillus niger bphA* mRNA, we have determined *in vivo* the translation initiated at the start codons of these uORFs by fusion of a reporter gene to these AUG codons. The results of these experiments (Table 1) clearly demonstrate that the AUG of the first uORF, which is located only 6 bases downstream of the 5' end of the mRNA, does not seem to be used efficiently as a translation initiation codon. A low efficiency of translation from AUG codons present within the first bases of the mRNA has been described before by Sedman et al. (1990) and Kozak (1991a), although the effects in those studies were not as pronounced as observed here. Fusion of the reporter gene to the start codon of the second uORF results in levels of expression of β-galactosidase that are comparable to (or even slightly higher than) the expression of the reporter gene fused to the translation start codon of the main open reading frame. These results strongly suggest that uORF2 is translated *in vivo*.

Expression of the reporter gene in a vector in which the transcription control region of the *bphA* gene was fused at the transcription initiation site to an *A. nidulans gpdA-lacZ* fusion gene (pAB94-86) is comparable to that of the reporter gene fused to the translation start codon of the main open reading frame, indicating that the 5' region of the *bphA* mRNA contains no elements which have a strong negative or positive effect on translation.

After induction the amount of *bphA* mRNA found in mycelium is considerable, indicating that the expression control region of this gene is rather efficient. To determine its expression level more precisely, we have compared its expression level to that of a well known efficiently expressed promoter. These experiments showed that the expression level of the benzoate induced *bphA* promoter (using a pAB94-83 transformant) is about 150% of the level directed by the promoter of the highly expressed glyceraldehyde-3-phosphate dehydrogenase gene (*gpdA*) of *A. niger* present on an identical expression analysis vector (pAB94-53; van Gorcom et al. 1993).

The context of both AUG_2 and AUG_3 (main ORF) is not optimal for initiation of translation based on the rules presented by Kozak (GCCA/GCC**AUG**G, positions -3 and +4 have highest importance; 1991b). AUG_2 is missing a purine in position -3 and AUG_3 is missing the G in position +4. However, from the experiments presented in this paper it can be concluded that both AUG regions are able to efficiently initiate translation.

Only small non-significant differences (max. 39%) of the *lacZ* expression between the different transformants (obtained with plasmids pAB94-83, -85 and -86) were found after induction by benzoate and PABA. However, a larger relative effect of the different fusion positions (AUG_2, AUG_3 or TSP) on the expression of *lacZ* was observed when para-hydroxy-benzoate is used as inducer. The *lacZ* expression in transformants in which the transcriptional fusion (pAB94-86) is present is about three times higher compared to the expression level in pAB94-83 transformants (fusion at the AUG of the main ORF). The *lacZ* expression of pAB94-85 transformants (fusion at AUG_2) is about twice as high as those in pAB94-83 transformants. So using para-hydroxy-

translational control. Alternatively differences in mRNA stability occur, caused by differences in the 5' regions of the mRNAs formed in the different transformants.

β-Galactosidase activity levels determined in transformants containing the reporter constructs where *lacZ* is fused to AUG_3 in which mutations were made in the AUG codons of the uORFs one by one and both together (shown in Table 2), show that the absence of AUG_1 has no significant positive effect on the translation of the main open reading frame. In one case (N245::pAB94-91, benzoate induction) we even find a decrease of β-galactosidase activity. Mutation of AUG_2 results in a small, but in most cases non-significant, increase of β-galactosidase production starting at AUG_3 both in wildtype cells and in the Δ*bphA* strain with all inducers tested. The effects of the mutation of the AUG codons of both uORFs are comparable to the effects found with the mutation of only AUG_2. The effect of a mutant AUG_2 on β-galactosidase expression is much lower than we expected, based on the scanning model for translation initiation postulated by Kozak (1989, 1992) and the efficient β-galactosidase expression found in transformants with a fusion of *lacZ* to AUG_2. The presence of an active upstream AUG codon on the mRNA before the start codon of the main open reading frame has been shown to have a strong negative effect on translation initiation at the next AUG codon (Kozak 1995).

There are several possible explanations for our results. The region between AUG_2 and AUG_3 might contain an IRES (internal ribosome entry site). Translation can be resumed at such sites after termination of an upstream ORF. These IRES sites, however, have sofar only been found in viruses (Macejak and Sarnow 1991) and in a gene from *Drosophila* (Soo-Kyung et al. 1992). The distance between uORF2 and the main ORF is 117 bases. This distance might be long enough for the ribosomes which resume scanning after translation of uORF2 to reassemble with the necessary initiation factors in order to be able to (re)initiate at AUG_3 (Grant et al. 1994). However, the sequence context at the translation termination codon of uORF2 does not particularly fit to the rules that have been worked out for optimal resumption of scanning by Grant and Hinnebusch (1994).

Analysis of the 5' region of the *bphA* gene with respect to the presence of possible secondary structure elements did not reveal potential stems of more then 6 bp in length (6 bp only if one G-T pair is allowed), indicating that it is unlikely that secondary structure plays a major role in the translational regulation of *bphA* expression.

Implicit in our interpretation of the data is the assumption that the transcription start points are the same under all conditions. So far we have only precisely determined the transcription initiation point with benzoate as inducer. The high level expression of *lacZ* from $uAUG_2$ under most conditions tested indicates that this uAUG seems to be present on the mRNAs produced under these conditions. Recent preliminary results obtained in our laboratory indicate that *bphA* mRNA, synthesized at a very low level during growth in high glucose concentrations (1%), might have a larger 5' untranslated region compared to the mRNA formed under benzoate induced conditions (van den Brink and van Gorcom, unpublished). Such benzoate-dependent differential promoter use has also been found in the *cprA* gene encoding the second component of the BPH

enzyme complex (van den Brink et al. 1997). Nevertheless, it seems unlikely that differential promoter use can be an explanation for the results found in this study.
If the results obtained with the reporter plasmids in the wildtype and the Δ*bphA* strain (Table 2) are compared, no differences are found indicating the absence of a clear *trans* effect of the putative peptides encoded by the uORFs on the expression of the *bphA* gene.
In conclusion, we have not been able to demonstrate, under the conditions used, any of the effects of uORFs which have been found in the regulation of expression of other lower eukaryotic, uORF containing genes.
The real biological function of the *bphA* gene of *A. niger* is not (yet) known. The *Aspergillus niger* strain, used in this work, has been selected for its ability to hydroxylate one-ring aromatic compounds at the para-position (de Jonge et al. 1990). The *bphA* gene has been isolated using a differential hybridization screening making use of the inducibility of the expression of the gene by benzoate (van Gorcom et al. 1990). Mutants made in this gene both by classical means (Boschloo et al. 1990 and 1991) and by directed gene disruption (this study) have no apparent phenotype, apart from their loss of the ability to para-hydroxylate benzoic acid and related molecules. Verification of the genuine substrate of BPH may be required to get more insight in the possible role(s) of the uORFs in expression of the *bphA* gene.

ACKNOWLEDGEMENT
The authors would like to thank Jacqueline Busscher-Lange, Sandra van den Broek, Theo Sonke, Marjolein van Helmond and Inge Frantzen for excellent technical assistance. Peter Pouwels is acknowledged for critical reading of the manuscript and helpful discussions on the conclusions. Part of the work has been sponsored by DSM Fine Chemicals.

REFERENCES

Abastado, J.-P., P.F. Miller, B.M. Jackson, and A.G. Hinnebusch (1991). Suppression of ribosomal reinitiation at upstream open reading frames in amino acid-starved cells forms the basis for *GCN4* translational control. *Mol. Cell. Biol.* **11**, 486-496.

Boschloo, J.G., A. Paffen, T. Koot, W.J.J. van den Tweel, R.F.M. van Gorcom, J.H.G. Cordewener & C.J. Bos (1990). Genetic analysis of benzoate metabolism in *Aspergillus niger*. *Appl. Microbiol. Biotech.* **34**, 225-228.

Boschloo, J.G., E. Moonen, R.F.M. van Gorcom, H.F.M. Hermes and C.J. Bos (1991). Genetic analysis of *Aspergillus niger* mutants defective in benzoate-4-hydroxylase function. *Curr. Genet.* **19**, 261-264.

De Jonge, P.H., Th.J.G.M. van Dooren, G.J. Schmitz, J.A.M. van Balken, H.F.M. Hermes, E.M. Meyer and W.J.J. van den Tweel (1990) Formation of para-hydroxybenzoate from benzoate by immobilized *Aspergillus niger*. In: Physiology of Immobilized Cells, (Eds. De Bont, J.A.M., J. Visser, B. Mattiasson and J. Tramper) Elsevier Science Publishers B.V., Amsterdam, The Netherlands, 317.

Delbecq, P., M. Werner, A. Feller, R.K. Filipkowski, F. Messenguy and A. Piérard (1994). A segment of mRNA encoding the leader peptide of the *CPA1* gene confers repression by arginine on a heterologous yeast gene transcript. *Mol. Cell. Biol.* **14**, 2378-2390.

Geballe, A.P. and D.R. Morris (1994). Initiation codons within 5'-leaders of mRNAs as regulators of translation. *Trends Biochem. Sci.* **19**, 159-164.

Goldman, G.H., R.A. Geremia, A.B. Caplan, S.B. Vila, R. Villarroel, M. van Montagu and A. Herrera-Estrella (1992). Molecular characterization and regulation of the phosphoglycerate kinase gene from *Trichoderma viride*. *Mol. Microbiol.* **6**, 1231-1242.

Grant, C.M. and A.G. Hinnebusch (1994). Effect of sequence context at stop codons on efficiency of reinitiation in *GCN4* translational control. *Mol. Cell. Biol.* **14**, 606-618.

Grant, C.M., P.F. Miller and A.G. Hinnebusch (1994). Requirements for intercistronic distance and level of eukaryotic initiation factor 2 activity in reinitiation on *GCN4* mRNA vary with the downstream cistron. *Mol. Cell. Biol.* **14**, 2616-2628.

Han, S., J. Navarro, R.A. Greve and T.H. Adams (1993). Translational repression of *brlA* expression prevents premature development in *Aspergillus*. *EMBO J.* **12**, 2449-2457.

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

Hinnebusch, A.G. (1990). Involvement of an initiation factor and protein phosphorylation in translational control of *GCN4* mRNA. *Trends Biochem. Sci.* **15**, 148-152.

Kolar, M., P.J. Punt, C.A.M.J.J. van den Hondel, H. Schwab (1988). Transformation of *Penicillium chrysogenum* using dominant selection markers and expression of an *Escherichia coli lacZ* fusion gene. *Gene* **62**, 127-134.

Kozak, M. (1989). The scanning model for translation: an update. *J. Cell. Biol.* **108**, 229-241.

Kozak, M. (1991a) A short leader sequence impairs the fidelity of initiation by eukaryotic ribosomes. *Gene Expression* **1**, 111-115.

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

Kozak, M. (1992) Regulation of translation in eukaryotic systems. *Annu. Rev. Cell. Biol.* **8**, 197-225.

Kozak, M. (1995) Adherence to the first-AUG rule when a second AUG codon follows closely upon the first. *Proc. Natl. Acad. Sci. USA* **92**, 2662-2666 and 7134 (erratum).

Kunkel, T.A. (1985). Rapid and efficient site-specific mutagenisis without phenotypic selection. *Proc. Natl. Acad. Sci. USA* **82**, 488-492.

Luo, H. and M.H. Perlin (1993). The γ-tubilin-encoding gene from the basidiomycete fungus *Ustilago violacea*, has a lon 5'-untranslated region. *Gene* **137**, 187-194.

Luo, Z., M. Freitag and M.S. Sachs (1995) Translational regulation in response to changes in amino acid availability in *Neurospora crassa*. *Mol. Cell. Biol.* **15**, 5235-5245.

Luo, Z. and M.S. Sachs (1996) Role of an upstream open reading frame in mediating arginine-specific translational control in *Neurospora crassa*. *J. Bacteriol.* **178**, 2172-2177.

Macejak, D.G. and P. Sarnow (1991). Internal initiation of translation mediated by the 5' leader of a cellular mRNA. *Nature* **353**, 90-94.

MacRae, W.D., F.P. Buxton, S. Sibley, S. Garven, D.I. Gwynne, R.W. Davies and H.N. Arst (1988). A phosphate-repressible acid phosphatase gene from *Aspergillus niger*: its cloning, sequencing and transcriptional analysis. *Gene* **71**, 339-348.

Messing, J. (1983). New M13 vectors for cloning. In: Methods in Enzymology **101** (Eds. Wu, R., L. Grossman and K. Moldave), Academic Press, New York, USA, pp-20-78.

Miller, K.Y., T.M. Toennis, T.H. Adams and B.L. Miller (1991). Isolation and transcriptional charaterization of a morphological modifier: the *Aspergillus nidulans* stunted (*stuA*) gene. *Mol. Gen. Genet.* **227**, 285-292.

Mueller, P.P. and A.G. Hinnebusch (1986). Multiple upstream AUG codons mediate translational control of *GCN4*. *Cell* **45**, 201-207.

Mueller, P.P., B.M. Jackson, P.F. Miller and A.G. Hinnebusch (1988). The first and fourth upstream open reading frames in *GCN4* mRNA have similar initiation efficiencies but respond differently in translational control to changes in length and sequence. *Mol. Cell. Biol.* **8**, 5439-5447.

Orbach, M.J., M.S. Sachs and C. Yanofsky (1990). The *Neurospora crassa arg-2* locus. *J. Biol. Chem.* **265**, 10981-10987.

Paluh, J.L., M.J. Orbach, T.L. Legerton and C. Yanofsky (1988). The cross-pathway control gene of *Neurospora crassa, cpc-1*, encodes a protein similar to *GCN4* of yeast and the DNA binding domain of the oncogene v-*jun*-encoded protein. *Proc. Natl. Acad. Sci. USA* **85**, 3728-3732.

Pontecorvo, G., J.A. Roper, L.J. Hemmons, K.D. MacDonald and A.W.J. Bufton (1953) The genetics of *Aspergillus nidulans*. *Adv. Genet.* **5**, 141-238.

Punt, P.J., R.P. Oliver, M.A. Dingemanse, P.H. Pouwels, C.A.M.J.J. van den Hondel (1987) Transformation of *Aspergillus* based on the hygromycin B resistance marker from *Escherichia coli*. *Gene* **56** 117-124.

Punt P.J., Dingemanse, M.A., Kuyvenhoven A., Soede, R.D.M., Pouwels, P.H. & C.A.M.J.J. van den Hondel (1990). Functional elements in the promoter region of the *Aspergillus nidulans gpdA* gene encoding glyceraldehyde-3-phosphate dehydrogenase. *Gene* **93**, 101-109.

Sambrook, J., E.F. Fritsch and T. Maniatis (1989). Molecular cloning. A laboratory manual, 2nd ed.. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, USA.

Sedman, S.A., G.W. Gelembiuk and J.E. Mertz (1990). Translation at a downstream AUG occurs with increased efficiency when the upstream AUG is located very close to the 5' cap. *J. Virol.* **64**, 453-457.

Soo-kyung, O.H., M.P. Scott and P. Sarnow (1992). Homeotic gene *Antennapedia* mRNA contains 5'-noncoding sequences that confer translational initiation by internal ribosome binding. *Genes & Dev.* **6**, 1643-1653.

van den Brink, J.M., C. van Zeijl, C.A.M.J.J. van den Hondel and R.F.M. van Gorcom (1995) Cloning and characterization of the NAPDH cytochrome P450 oxidoreductase gene from the filamentous fungus *Aspergillus niger*. *DNA Cell Biol* **14**, 719-729.

van den Brink, J.M., C.A.M.J.J. van den Hondel and R.F.M. van Gorcom (1996a) Optimization of the benzoate-inducible benzoate *p*-hydroxylase cytochrome P450 enzyme system in *Aspergillus niger*. *Appl. Microbiol. Biotechnol.* **46** in press.

van den Brink, J.M., J.G.M. van Nistelrooij, M.A. de Waard, C.A.M.J.J. van den Hondel, R.F.M. van Gorcom (1996b) Increased resistance to 14α-demethylase inhibitors (DMI's) in *Aspergillus niger* by coexpression of the *Penicillium italicum* sterol-14α-demethylase (*cyp51*) and the *A.niger* cytochrome P450 reductase (*cprA*) genes. *J.of Biotechnol.* **49**, 13-18.

van den Brink, J.M. (1997) Thesis. University of Leiden, in press.

van Gorcom, R.F.M., Pouwels, P.H., Goosen, T., Visser, J., van den Broek, H.W.J., Hamer, J. E., Timberlake, W.E. & van den Hondel, C.A.M.J.J.(1985). Expression of an *Escherichia coli* ß-galactosidase fusion gene in *Aspergillus nidulans*. *Gene*, **40**, 99-106.

van Gorcom, R. F. M. & van den Hondel, C. A. M. J. J. (1988). Expression analysis vectors for *Aspergillus niger*. *Nucleic Acid Research*, **16**, 9052

van Gorcom, R.F.M., Boschloo, J.G., Kuyvenhoven, A., Lange, J., van Vark, A.J., Bos, C.J., van Balken, J.A.M., Pouwels, P.H. & C.A.M.J.J. van den Hondel (1990). Isolation and molecular characterization of the benzoate-para-hydroxylase gene (*bphA*) of *A. niger*, a member of a new gene-family of the cytochrome P450 superfamily. *Mol. Gen. Genet.* **223** 192-197.

van Gorcom, R.F.M., Th. Sonke and C.A.M.J.J. van den Hondel (1993) Expression cassette for heterologous proteins in filamentous fungi. PCT International Patent Application, WO 93/07277.

van Hartingsveldt, W., Mattern, I. E., van Zeijl, C. M. J., Pouwels, P. H. & van den Hondel, C. A. M. J. J. (1987). Development of an homologous transformation system for *Aspergillus niger* based on the *pyrG* gene. *Mol. Gen. Genet.*, **206**, 71-75.

Wanke, C., S. Eckert, G. Albrecht, W. van Hartingsveldt, P.J. Punt, C.A.M.J.J. van den Hondel and G.H. Braus (1997) The *Aspergillus niger GCN4* homologue *cpcA* is transcriptionally regulated and encodes an unusual leucine zipper. *Mol. Microbiol.*, submitted.

Werner, M., A. Feller and A. Piérard (1985). Nucleotide sequence of yeast gene *CPA1* encoding the small subunit of arginine-pathway carbamoylphosphate synthetase: homology of the deduced amino acid sequence to other glutamine amidotransferases. *Eur. J. Biochem.* **146**, 371-381.

Werner, M., A. Feller, F. Messenguy and A. Piérard (1987). The leader peptide of yeast gene *CPA1* is essential for the translational repression of its expression. *Cell* **49**, 805-813.

Williams, N.P., P.P. Mueller and A.G. Hinnebusch (1988). The positive regulatory function of the 5'-proximal open reading frames in *GCN4* mRNA can be mimicked by heterologous, short coding sequences. *Mol. Cell. Biol.* **8**, 3827-3836.

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

CHAPTER 7

THE PRESENCE OF UPSTREAM OPEN READING FRAMES PRESENT IN THE 5' REGIONS OF mRNA MOLECULES OF *Aspergillus* GENES: A REVIEW

Robert F.M. van Gorcom

ABSTRACT

A systematic analysis of gene sequences of the genus *Aspergillus* has been performed aiming at the identification of upstream open reading frames (uORFs) present on the encoded mRNAs before the main open reading frames. From 942 entries which were selected using the keywords 'Aspergil*' and 'Emericel*', only 162 contained enough information for this analysis. Thirty five genes (22%) were shown to contain one or more uORFs. Analysis of the sequence context of the AUG codons of the uORFs showed, in contrast to the sequence context of AUG codons of the main ORFs, an absence of biased use of specific bases.

In most cases where research has been performed on the function of uORFs, they were shown to play a role in the regulation of gene expression at the translational level. *Aspergillus* genes involved in the regulation of important cellular processes seem to contain such uORFs with a much higher frequency compared to other genes, indicating that translational control of gene expression might be a key mechanism in the way cells control particular processes.

INTRODUCTION

In eukaryotes, translation usually starts on the first AUG codon present on the mRNA (Kozak, 1987a, 1995). For some genes deviations from this rule have been observed. For example if the first AUG codon is located very close to the 5' end of the mRNA molecule it is often not or rarely used as a translation start codon (Kozak, 1991a; Slusher et al., 1991; Schutze et al., 1994). For some other genes, usually small open reading frames (ORFs), whose AUG codon is at a position that should allow proper initiation of translation, are found, preceding the main protein encoding region on the messenger RNA. For a few of these genes it has been shown that these upstream open reading frames (uORFs) are involved in the regulation of translation of the main open reading frame. In chapter 6 of this thesis a short description has been given of two types of translational control by uORFs in genes of the yeast *Saccharomyces cerevisiae.* In the same chapter research is described aiming at the analysis of the function of two small uORFs present on the mRNA of the benzoate para-hydroxylase gene of the filamentous fungus *Aspergillus niger.* In order to analyze the occurrence and possible function of uORFs on mRNAs of genes from fungi of the genus *Aspergillus* a detailed screening of all sequences of *Aspergillus* genes available has been performed. In this screening we have analyzed the 5' mRNA encoding region for the presence of AUG codons before the start of the main protein encoding open reading frame. In this study an upstream open reading frame is defined as a sequence present on the mRNA, which starts at an AUG codon located before the start of the main open reading frame, and ends with one of the three translation stop codons. The amount of amino acid codons in between start and stop codon may vary from 0 to a large number. In this chapter the results of the screening will be presented and discussed.

Analysis of the DNA/RNA sequence data libraries

To analyze the occurrence of uORFs in *Aspergillus* genes a computer search has been carried out on all gene sequences of the genus *Aspergillus* (or *Emericella*[1]) that were available in the EMBL/Genbank/GENESEQ libraries on September 25, 1996. Nine hundred forty two sequences were selected with the keywords 'Aspergil*' and 'Emericel*'. Within this group 40 entries did not contain sequences of species of the genus *Aspergillus*. All sequences which encoded (parts of) ribosomal RNA molecules, originated from the mitochondrial genome, contained vector or primer sequences, contained only promoter or terminator sequences or for which no features were indicated, were excluded from this analysis. This resulted in a list of entries which contained nuclear, protein encoding sequences. In this list a number of multiple entries of the same gene were found (e.g. several partial entries of one gene or both cDNA and genomic sequence data or data deposited by different authors) which further reduced the number of entries to be analyzed. A further reduction had to be made by removing all entries which lacked sequences 5' of the main ORF. All these restrictions reduced the set of 942 entries to 287 genes that could be used for this study.

These 287 genes have been analyzed for the availability of data on the start point of transcription (TSP) or cDNA data containing (part of) the 5' region of the mRNA. The 162 genes for which this information was available, have been analyzed for the presence of AUG codons upstream of the AUG codon of the main protein encoding open reading frame. The quantitative results of this analysis are shown in Table 1. Upstream open reading frames were found in 35 of the 162 genes analyzed.

[1] Throughout this chapter the name *Aspergillus* will also be used for species for which the name of the teleomorph (*Emericella*) should be used.

Species	genes used in this study	genes with known TSP	genes with known cDNA	genes with uORF
Aspergillus aculeatus	15	0	13	0
Aspergillus awamori	6	3	0	0
Aspergillus clavatus	1	0	1	1
Aspergillus ficuum	4	0	0	0
Aspergillus flavus	9	1	3	0
Aspergillus fumigatus	18	3	2	0
Aspergillus giganteus	3	1	1	0
Aspergillus kawachii	5	0	1	0
Aspergillus nidulans	126	53	30	28
Aspergillus niger	54	18	8	4
Aspergillus oryzae	22	6	4	1
Aspergillus parasiticus	12	2	5	0
Aspergillus restrictus	1	0	1	0
Aspergillus saitoi	2	0	1	0
Aspergillus shirousamii	2	1	1	0
Aspergillus terreus	5	1	1	1
Aspergillus tubigensis	2	1	0	0
Total	287	90	72	35

Table 1: Quantitative summary of the results of the search for *Aspergillus* genes present in the DNA data library (EMBL/Genbank/GENESEQ) on September 25, 1996. For each species are given the number of genes that have been analyzed, the number of genes of which the start point of transcription has been determined, the number of genes of which 5' cDNA data are available (only scored if the TSP is not known) and the number of genes in which one or more uORFs have been found. For the species *A. amstelodami*, *A. anthodesmis*, *A. avenaceus*, *A. carbonarius*, *A. dimorphicus*, *A. flaschentraegeri*, *A. glaucus*, *A. gorakhpurensis*, *A. itaconicus*, *A. japonicus*, *A. ochraceus*, *A. sojae*, *A. sparsus*, *A. tamarii*, *A. terricola*, *A. thomii*, *A. unguis*, *A. usamii*, *A. ustus* and *A. wentii* only entries have been found in the library that could not be used in this study.

For the 35 genes in which upstream AUG codons occurred, the position and size of the putative uORF(s) were calculated and these results are shown in Table 2. In Table 2 it is also indicated whether or not the uORFs overlap with the main protein encoding open reading frame. The number of uORFs per gene varies from 1 to 10. Sixteen of the 35 genes (46%) contain more than one uORF on their mRNA.

Table 2: Number, length and position of the upstream open reading frames present in *Aspergillus* genes. When the uORF overlaps with the main protein encoding ORF this is indicated in the last column.

LIBRARY CODE	GENE	cDNA or genomic[1]	number of uAUGs	relative position to TSP=1[2]	relative position to main AUG=1	length uORF bp[3]	overlapping with main ORF[4]
Ac19383	A. clavatus cla	c	4	18	-262	9	-
				89	-191	165	-
				141	-139	66	-
				173 [5]	-107	81	-
Anaciag	A. nidulans aciA	g	3 [6]	113	-167	282	+
				177	-103	72	-
				203 [7]	-77	192	+
Anamdr	A. nidulans amdR	g	4	51	-108	9	-
				56	-103	81	-
				110 [8]	-49	27	-
				148	-11	117	+
Anapsa	A. nidulans apsA	g	10	9	-1145	177	-
				48 [9]	-1106	138	-
				68	-1086	33	-
				107 [10]	-1047	15	-
				141 [11]	-1013	45	-
				147 [12]	-1007	39	-
				507	-647	6	-
				1028	-126	42	-
				1038	-116	12	-
				1052 [13]	-102	15	-

1 Only genomic sequences for which the TSP is determined are included
2 Major transcription startpoint or to 5' end of cDNA
3 Including stop codon
4 -: not overlapping; +/-: stop codon uORF and start codon main ORF overlap; +: uORF and main ORF overlap in different reading frames
5 This uATG is located within the uORF starting at position 89
6 The *aciA* gene has two TSPs. uATGs are only present in the larger transcript
7 This uATG is located within the uORF starting at position 113
8 This uATG is located within the uORF starting at position 62
9 This uATG is located within the uORF starting at position 9
10 This uATG follows the stop codon of the uORF started at 68 (TAAATG)
11 This uATG is located within the uORF starting at position 9
12 This uATG is located within the uORF starting at position 9
13 This uATG is located within the uORF starting at position 1028

BRARY ODE	GENE	cDNA or genomic	number of uAUGs	relative position to TSP=1	relative position to main AUG=1	length uORF bp	overlapping with main ORF
aromg	A. nidulans arom	g	1 [14]	90/53	-51	9	-
bima	A. nidulans bimA	c	1	250	-109	18	-
bime	A. nidulans bimE	c	1	126	-37	42	+/- [15]
nima	A. nidulans nimA	c	8	18	-422	39	-
				67	-373	57	-
				85 [16]	-355	39	-
				123	-317	54	-
				126 [17]	-314	51	-
				167	-273	144	-
				281 [18]	-159	30	-
				346	-94	63	-
nimtmr	A. nidulans nimT	c	1	82	-81	12	-
npaccgen	A. nidulans pacC	c	1	4	-169	9	-
npalbgn	A. nidulans palB	c	2	129	-174	6	-
				219	-84	39	-
nqutr	A. nidulans qutR	g	4	110	-86	72	-
				117	-79	24	-
				142	-54	12	-
				155 [19]	-41	27	-
nrnasepb	A. nidulans sepB	c	4	20	-83	60	-
				29 [20]	-74	51	-
				62 [21]	-41	18	-
				89	-14	165	+
nrps16a	A. nidulans rps16	g	3	14	-280	33	-
				69	-225	42	-
				75 [22]	-219	36	-
nsagamr	A. nidulans sagA	c	1	12	-185	12	-
nuapc	A. nidulans uapC	g	1	30	-34	18	-
n12427	A. nidulans myoA	c	1	99	-123	60	-

The *arom* gene has two TSPs
Stop codon of uORF and start codon of main ORF overlap (AUGA)
This uATG is located within the uORF starting at position 67
This uATG is located within the uORF starting at position 123
This uATG is located within the uORF starting at position 167
This uATG is located within the uORF starting at position 110
This uATG is located within the uORF starting at position 20
This uATG is located within the uORF starting at positions 20 and 29
This uATG is located within the uORF starting at position 69

LIBRARY CODE	GENE	cDNA or genomic	number of uAUGs	relative position to TSP=1	relative position to main AUG=1	length uORF bp	overlapping with main ORF
En13919	A. nidulans cnaA	c	2	79 [23]	-72	72 [23]	+ [23]
				146	-5	36	+
En22009	A. nidulans nudF	c	4	13	-151	39	-
				24	-140	9	-
				32	-132	81	-
				54	-110	33	-
En25693	A. nidulans ankA	g	1	29	-286	78	-
En40146	A. nidulans sudA	c	1	118	-45	9	-
Enacue	A. nidulans acuE	g	1	55..34 [24]	-23	12	-
Enbimb	A. nidulans bimB	c	1	16	-15	9	-
Enbimd	A. nidulans bimD	c	1	81	-52	12	-
Enbimg	A. nidulans bimG	c	1	32	-854	105	-
Enbrlaab	A. nidulans brlA	g	1 [25]	68	-677	126	-
Encytc	A. nidulans cytC	g	6 [26]	75	-278	6	-
				95	-258	21	-
				98 [27]	-255	18	-
				233/118	-120	33	-
				262/147	-91	135	+
				304/189 [28]	-49	93	+
Enlam	A. nidulans lamA	g	1	53	-75	27	-
Enstua	A. nidulans stuA	g	5	27	-1062	75	-
				63 [29]	-1026	39	-
				286	-803	27	-
				354	-735	6	-
				589	-500	12	-

[23] In frame with main ORF

[24] The *acuE* gene has seven closely spaced TSPs

[25] The *brlA* gene encodes two transcripts. The longest transcript contains an uATG

[26] The *cytC* gene has two TSPs. A minor TSP is located 115 upstream from the major TSP. uATGs 1, 2 and 3 are only present on the mRNA started at the minor TSP

[27] This uATG is located within the uORF starting at position 95

[28] This uATG is located within the uORF starting at position 262/147

[29] This uATG is located within the uORF starting at position 27

IBRARY ODE	GENE	cDNA or genomic	number of uAUGs	relative position to TSP=1	relative position to main AUG=1	length uORF bp	overlapping with main ORF
nbpha	A. niger bphA	g	2	6	-281 [30]	57	-
				49	-238 [31]	18	-
ncpra	A. niger cprA	g	1 [32]	472	-382	147	-
nniado	A. niger niaD	g	1	13	-48	30	-
npaca	A. niger pacA	g	2	32	-114	102	-
				114	-32	36	+/- [33]
orntb	A. oryzae rntB	g	1	22/30	-26	60	+
tdhgo	A. terreus dhgo	g	6	6	-475	81	-
				11	-470	63	-
				204	-277	6	-
				303	-178	132	-
				373	-108	54	-
				415 [34]	-66	12	-

in DNA; in mRNA -177 due to the presence of an intron in this region
in DNA; in mRNA -134 due to the presence of an intron in this region
The *cprA* gene has two TSPs. The uORF is only present on the longest mRNA which is formed after induction with benzoate
Stop codon of uORF and start codon of main ORF overlap (AUGA)
This uATG is located within the uORF starting at position 373

Analysis of the sequence context around the start codons of the uORFs and the main ORF of the uORF containing *Aspergillus* genes

Around the translational start codons of open reading frames usually a clear bias in the use of certain bases can be found. This bias is thought to be more pronounced when the gene is highly expressed. Kozak has made many analyses of the sequences in these regions in different organisms (Kozak, 1986, 1987abc, 1988, 1989ab, 1990, 1991abc, 1992ab, 1995) and several years ago such a compilation has also been made for fungi (Gurr et al., 1988). More recently a database (called TransTerm) has been created which contains data about the sequence context around start codons for more than 40.000 genes of 150 different organisms (Dalphin et al., 1996).

The typical sequence context of all AUG codons of the uORFs found in this study and of the AUG codons of the main ORFs of the same genes has been calculated and the results are summarized in Table 3. In Table 3 also the data present in the TransTerm database for all *A. nidulans* and *A. niger* genes are shown. As can be seen in this Table, clear differences in the use of specific bases around the uAUG codons compared to the usage around the main AUG codons (both in the data resulting from this study and in the TransTerm data) are found at positions -5, -4, -3, -2, -1 and +4. The differences between the TransTerm data and the sequence context calculated from the AUG of the main ORF of the uORF containing genes listed in this study are much smaller. Most uORF percentages found at the different positions, are close to the mean. The differences in the sequence context of the uORF AUG codons and the AUG codons of the main ORFs seem to indicate an absence of evolutionary pressure on the base context around the uORF AUG codons.

Table 3: Comparison of the base composition of the sequences around the translation start codons compiled in this study and as present in the TransTerm database. (Position A of initiation codon = +1)

		-10	-9	-8	-7	-6	-5	-4	-3	-2	-1	+1	+2	+3	+4	+5	+6
Transterm *A. niger*	%G	20	11	4	27	31	11	7	18	9	4	0	0	100	29	9	38
Transterm *A. nidulans*		23	21	9	18	24	14	7	19	10	14	0	0	100	45	18	21
Main ORFs this study		29	17	14	37	23	9	6	23	11	20	0	0	100	51	17	26
uORFs this study		**15**	22	17	26	**15**	**27**	**23**	16	**26**	**34**	0	0	100	**15**	**23**	28
Transterm *A. niger*	%A	31	40	40	20	11	11	22	73	27	20	100	0	0	24	31	4
Transterm *A. nidulans*		26	26	26	29	25	13	20	67	29	33	100	0	0	12	21	13
Main ORFs this study		31	34	40	26	34	20	29	57	23	20	100	0	0	14	20	17
uORFs this study		30	24	**20**	25	20	23	24	**30**	19	25	100	0	0	28	19	15
Transterm *A. niger*	%U	13	33	20	13	36	36	7	4	33	9	0	100	0	18	27	18
Transterm *A. nidulans*		21	24	33	22	29	37	17	6	23	14	0	100	0	25	15	39
Main ORFs this study		9	26	14	20	20	26	26	11	23	23	0	100	0	20	17	26
uORFs this study		19	32	27	20	34	28	28	**31**	31	15	0	100	0	24	28	34
Transterm *A. niger*	%C	36	16	36	40	22	42	64	4	31	67	0	0	0	29	33	40
Transterm *A. nidulans*		29	29	32	31	21	36	56	8	38	39	0	0	0	18	46	28
Main ORFs this study		31	23	31	17	23	46	40	9	43	37	0	0	0	14	46	31
uORFs this study		36	23	35	28	**31**	**22**	**25**	**24**	**24**	**26**	0	0	0	33	30	23
Positions around the start codons of uORFs with major deviations from nucleotide use around main ORFs							↑	↑	↑	↑	↑				↑		
Consensus main ORFs (TransTerm + main ORF)							u/c	C	A	c	C	A	U	G	g	c	
Consensus uORFs							x	x	x	x	x	A	U	G	x	x	

X: indicates percentage usage of nucleotide in uAUG context clearly different (≥20%) from main ORF and Transterm percentages
X: indicates percentage usage of nucleotide in uAUG context strongly different (≥40%) from main ORF and Transterm percentages

DISCUSSION

We have analyzed all *Aspergillus* genes present in the DNA data libraries for the presence of upstream open reading frames (uORFs) in the mRNA encoding region. In the 942 entries which were identified with the keywords 'Aspergil*' and 'Emericel*' only 162 genes contained enough information to allow such an analysis. The analysis presented in this chapter is severely hampered by the lack of sequence data of the 5' regions of the genes present in the data library and the lack of information about start points of transcription for a large number of genes. In 44% of the entries which contain sequence data 5' of the translational start codon (125 genes), information about the start point of transcription or 5' end(s) of cDNA clones is absent (see Table 1). Remarkably, this lack of data is not restricted to the more industrially used species of the genus *Aspergillus*. Even for the 'academic' species *Aspergillus nidulans* in 34% of all entries and the accompanying scientific publications no data about the transcription start point can be found. For almost all *A. nidulans* and *A. niger* genes of which no cDNA data or data about the TSP were available, the existence of additional data has been checked with the authors but in most cases such information was not available.

Of the 162 genes for which enough information was available, 35 genes (22%) were identified that contained one or more uORFs. The actual number of uORF containing genes within this group of 162 genes might even be higher, since for 72 genes only cDNA data were available and the percentage of completely full-size cDNAs among these entries is unknown.

The percentage of all genes that contain one or more uORF(s) (22%) is much higher than we had anticipated, as only in few reports the existence of these uORFs was noted. A reason for the frequent occurrence of uORFs might be a bias towards a specific type of genes that have been isolated so far. *Aspergillus nidulans* is extensively used for the study of (regulation) of differentiation, (regulation) of mitosis and nuclear movement, of pH control, and of general and pathway-specific carbon and nitrogen source dependent regulation of gene expression. This organism clearly shows the highest percentage of genes containing one or more uORFs (34%). Of the genes of the other Aspergilli together only 9% contain one or more uORFs.

Analysis of the list of uORF containing *Aspergillus* genes shows that uORFs are present in several types of genes. They are found in genes encoding proteins which have a(n) (important) regulatory function (e.g. in differentiation, nuclear distribution and movement, mitosis or pH dependent gene regulation) but they are also found in genes that are thought to have a function in the metabolism of common and less common nutrients/chemicals which the organism may encounter in its environment, e.g. acetate, nitrate, poly-galacturonic acid and benzoate. The rather frequent occurrence of uORFs can be interpreted in two ways: (1) most uAUG codons for some reason do not initiate translation and therefore have no role in the (regulation of) translation of the main ORF or (2) most uAUG codons and/or the peptides that are translated from the uORFs are part of mechanisms that controls the amount of protein translated from the main ORFs. To analyze which of these two interpretations is most likely, we have further analyzed the uORFs.

An important question with regard to the role of uORFs is whether these uORFs are actually translated. In many cases this question can only be answered indirectly since the short length of the uORF does not allow the (easy) detection of its encoded peptide. Therefore an indirect method is used in most studies by fusing a reporter gene in frame to the uORF. All uAUG codons that are not positioned at the very 5' end of the mRNA were shown to be able to initiate translation in the few studies done so far in *Aspergillus* (Miller et al., 1992; Han et al., 1993; chapter 6). Kozak (1991a) and others have described a few examples of genes containing uAUG codons which are located relatively close to the 5' end of the mRNA (<= 20 bases) and they showed that, because of the position of these uAUG codons, leaky scanning occurs resulting in initiation of translation at the second AUG codon. Of the 88 uAUG codons identified in this study (Table 2), 13 are located in the first 20 nucleotides of the mRNA or cDNA (12 genes; one gene (*A. terreus dhgo*) contains two 'early' AUG codons). For 7 of these uAUG codons the calculation of their distance to the 5' end is based on cDNA data, indicating that their actual position on the (full length) mRNA is unknown. Nine of the 13 'early' uAUG codons are present in mRNAs which have at least one other uAUG codon further downstream. These data indicate that only a small minority of the uORFs described in this study will have no function just because of their *position* in the very 5' region of the mRNA. That uAUG codons in fungi near the 5' end of mRNA are not influencing the translation initiation at a downstream AUG codon can be concluded from the results presented in chapter 6. The presence or absence of the first uAUG codon of the *A. niger bphA* gene does not have a significant effect on translation initiated at the second uAUG or the AUG of the main ORF.

The longer the 5' region of an mRNA the higher the chance for the (random) occurrence of an AUG triplet. Statistically one would expect the occurrence of an AUG codon in a random sequence once every 64 nucleotides. The average length of the 5' region of the 35 uAUG containing *Aspergillus* genes is 309 nucleotides (ranging from 31 to 1154 nucleotides). The occurrence of an uAUG codon in this region is once every 123 nucleotides. The average length of the 5' region of the 128 other genes (containing no uORF) is only 97 nucleotides (ranging from 3 to 834 nucleotides). It is difficult to conclude whether the larger average size of the 5' region of the 35 uORF containing genes is the (only) reason for the presence of uORFs or whether the 5' regions are longer because of the presence of uORFs.

The longest uORF found in this study is 282 nucleotides long and encodes a putative polypeptide of 93 amino acids. The smallest uORF consists only of a start and a stop codon (6 nucleotides). The size distribution of the 88 uORFs found in this study is: 6-33 nucleotides (1-10 aa codons): 39; 36-78 nucleotides (11-25 aa codons): 30; 81-282 nucleotides (26-93 aa codons): 19. The average length of these 88 uORFs is 54 nucleotides (18 codons). This average is to be expected for a random use of codons (statistical occurrence of a stop codon is once every 21 codons). These results suggest that it seems likely that the presence of the uAUG start codon is in most cases more important than the encoded putative peptide or length of the peptide. However, it has been shown that in some cases the uORF encoded peptide is essential in the control of gene expression of the gene (see later).

Differences in translation initiation efficiency are often reflected in differences in the sequence context of AUG codons. Therefore, we have also analyzed the sequence context around the AUG codons of the uORFs and of the main ORFs. At most positions where differences in the use of specific bases are found (Table 3) the sequence context of the AUG codons of the uORFs shows hardly any bias, whereas the sequence context of the AUG codons of the main ORFs clearly shows a specific bias (e.g. U/C at -5, C at -4, A at -3, G at +4) . This bias might be the result of an evolutionary pressure for the use of preferred bases at certain positions to achieve optimal conditions for the start of translation as is found in many organisms. These results clearly indicate that the AUG start codons of the uORFs must have a different role (and are under a different evolutionary pressure) compared to the AUG start codons of the main open reading frames.
Of the 88 uORFs found in this study only 9 overlap with the main ORF and of two uORFs the stop codons overlap with the start codon of the main ORF. Therefore it can be concluded that direct interference of translation of the uORF with the initiation of translation at the AUG of the main ORF does not seem to be the mechanism mediating translation control of the translation of the majority of these genes.

Only for a very limited number of uORF containing eukaryotic genes the possible function of their uORF(s) has been analyzed. In these studies it was shown that the uORFs have a very important role in the regulation of gene expression by controlling the level of translation initiation at the AUG of the main protein encoding ORF. The mechanisms used for this control are different for different genes. The translation of the *Saccharomyces cerevisiae GCN4* gene (encoding a regulatory protein involved in amino acid biosynthesis) is controlled by a complex system involving reinitiation of translation (Hinnebusch, 1990; Abastado et al., 1991; Grant and Hinnebusch, 1994). This type of regulation also seems to be present in the filamentous fungal homologues of the GCN4 gene (*Neurospora crassa cpc-1* (Paluh et al., 1988), *Aspergillus niger cpcA*, Wanke et al., 1997). The mode of action of the translational control of the *GCN4* gene is not dependent on the codons used in the uORFs of the gene (Mueller et al., 1988; Williams et al., 1988).
A completely different mechanism of control of gene expression by a uORF is found in the *CPA1* gene of *S. cerevisiae* (Werner et al., 1985) and its fungal homologue the *N. crassa arg-2* gene (Orbach et al., 1990; Luo et al., 1995). These genes contain a uORF encoding a 25 (*CPA1*) resp. 24 (*arg-2*) amino acid long peptide, whose amino acid sequences are similar. It has been shown for both genes that these uORFs confer control of translation initiation of the main ORF present on the mRNA irrespective of the sequence of the surrounding 5' region (the *CPA1* uORF is also functional if placed in the 5' region of another gene; Delbecq et al., 1994). Even a single amino acid substitution in this uORF completely abolishes its role in control of gene expression (Werner et al., 1987; Luo and Sachs, 1996).

The function of the uORF(s) in most of the genes identified in this study is unknown. Only in very few cases (some) research has been carried out on their role in the control of gene expression.

The regulation of expression of two *A. nidulans* genes (*brlA* and *stuA*) and the role of uORFs in this regulation has been studied. These two genes are encoding proteins involved in the regulation of differentiation of mycelial cells to conidiophores (asexual spore forming multicellular structures) and conidiospores (asexual spores). The expression of the *brlA* gene is controlled by a complex mixture of regulatory mechanisms (regulation at the level of transcription induction, differential promoter use, differential splicing resulting in mRNAs encoding slightly different proteins (at their N-terminal end) and down regulation of translation by the presence of a uORF). Under normal cultivation conditions the *brlA* gene is transcribed at rather low levels and the transcript formed (*brlAβ*) contains a uORF. The presence of this uORF prevents the translation of the main ORF. Removal of the initiation codon of this uORF leads to deregulated *brlA* expression, resulting in an inappropriate activation of development (Han et al., 1993). After induction a different mRNA species (*brlAα*) is predominantly formed which lacks the uORF.
The expression of *stuA* is, apart from transcriptional regulation, also controlled at the translational level, which is probably also mediated by the presence of uORFs. The *stuA* gene is, like the *brlA* gene, also transcribed from two promoters. Both transcripts have very long 5' regions which contain 3 very small uORFs. Their role in translational control has not yet been studied. On the *stuAα* transcript an additional uORF, encoding a putative peptide of 24 amino acids, is present. Miller et al. (1992) have shown -by in frame fusion of *lacZ* with the AUG codon of this uORF- that the AUG codon of this most 5' uORF can be used as start codon for translation. Their results also indicate that the integrity of this uORF is required for the induction-dependent increase in StuA protein production and therefore the authors propose that translation of this uORF has a regulatory function.
Apart from the two examples mentioned above and the study presented in chapter 6 of this thesis the possible function of uORFs has not been studied for any other *Aspergillus* gene containing uORFs.

Inspection of the list of uORF containing genes presented in Table 2, reveals the remarkable feature that a high percentage of genes, whose gene products are involved in the process of cell differentiation and the cell cycle and its regulation, contain uORFs. For example, all genes, except *nimE*, identified so far, whose gene products are involved in mitosis and nuclear movement and for which data on the 5' region of their mRNA are known, contain at least one uORF. These results strongly suggest that apart from regulation at the transcriptional level (which is found for some of these genes) the expression of these genes and, consequently, the regulation of mitosis in *Aspergillus nidulans* might also be controlled at the translational level. However, so far the involvement of uORFs in the regulation of expression of these genes has not been studied.

Taken together the results obtained till now about the role of uORFs in the control of gene expression in several lower eukaryotic genes, interpretation of the results obtained in this study and the rigidity of the 'Kozak rules' as described again recently by Kozak (1995) it is tempting to

in a position that allows translation (= not too close to the 5' end), play a role in translational control of expression of that gene.

In conclusion, we have found that an unexpectedly high percentage (22%) of the genes of the genus *Aspergillus* contain small open reading frames in the 5' region of their mRNA. Considering the relatively limited number of genes which have been characterized so far it is difficult to draw very firm conclusions. In some classes of regulatory genes the percentage of uORF containing genes is much higher than average. For example, of the genes that are involved in mitosis 85% (7 from 8) contain at least one uORF, suggesting that translational control of expression of these genes may be very important. In the few cases where the role of uORFs in regulation of gene expression has been studied, the results show, with the exception of *bphA*, that translational control mediated by the presence of upstream starts of translation is a important element in the regulation of expression of those genes.
Given these results and the high frequency of occurrence of uORFs particularly in genes that are part of or control important cellular processes such as differentiation, mitosis and nuclear movement, it will be important to start (or expand) research on the analysis of the role in translational control of these uORFs. This will allow us to get a more complete picture of the way the expression of these genes and thereby the processes they perform or control, is regulated.

REFERENCES

Abastado, J.-P., P.F. Miller, B.M. Jackson, and A.G. Hinnebusch (1991). Suppression of ribosomal reinitiation at upstream open reading frames in amino acid-starved cells forms the basis for *GCN4* translational control. *Mol. Cell. Biol.* **11**, 486-496.

Dalphin, M.E., C.M. Brown, P.A. Stockwell and W.P. Tate (1996) TransTerm: a database of translational signals. *Nucleic Acid Res.* **24**, 216-218.

Delbecq, P., M. Werner, A. Feller, R.K. Filipkowski, F. Messenguy and A. Piérard (1994). A segment of mRNA encoding the leader peptide of the *CPA1* gene confers repression by arginine on a heterologous yeast gene transcript. *Mol. Cell. Biol.* **14**, 2378-2390.

Grant, C.M. and A.G. Hinnebusch (1994). Effect of sequence context at stop codons on efficiency of reinitiation in *GCN4* translational control. *Mol. Cell. Biol.* **14**, 606-618.

Gurr, S.J., S.E. Unkles and J.R. Kinghorn (1988) The structure and organization of nuclear genes of filamentous fungi. *In:* Gene structure in Eukaryotic Microbes (Ed. Kinghorn, J.R.), IRL Press, SGM Spec. Publ. Vol. 23, 93-139.

Han, S., J. Navarro, R.A. Greve and T.H. Adams (1993) Translational repression of *brlA* expression prevents premature development in *Aspergillus. EMBO J.* **12**, 2449-2457.

Hinnebusch, A.G. (1990). Involvement of an initiation factor and protein phosphorylation in translational control of *GCN4* mRNA. *Trends Biochem. Sci.* **15**, 148-152.

Kozak, M. (1986) Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. *Cell* **44**, 283-292.

Kozak, M. (1987a) An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs. *Nucleic Acids Res.* **15**, 8125-8148.

Kozak, M. (1987b) Effects of intercistronic length on the efficiency of reinitiation by eukaryotic ribosomes. *Mol. Cell. Biol.* **7**, 3438-3445.

Kozak, M. (1987c) At least six nucleotides preceding the AUG initiator codon enhance translation in mammalian cells. *J. Mol. Biol.* **196**, 947-950.

Kozak, M. (1988) Leader length and secondary structure modulate mRNA function under conditions of stress. *Mol. Cell. Biol.* **8**, 2737-2744.

Kozak, M. (1989a) The scanning model for translation: an update. *J. Cell. Biol.* **108**, 229-241.

Kozak, M. (1989b) Circumstances and mechanisms of inhibition of translation by secondary structure in eukaryotic mRNAs. *Mol. Cell. Biol.* **9**, 5134-5142.

Kozak, M. (1990) Downstream secondary structure facilitates recognition of initiator codons by eukaryotic ribosomes. *Proc. Natl. Acad. Sci. USA* **87**, 8301-8305.

Kozak, M. (1991a) A short leader sequence impairs the fidelity of initiation by eukaryotic ribosomes. *Gene Expression* **1**, 111-115.

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

Kozak, M. (1991c) An analysis of vertebrate mRNA sequences: intimations of translational control. *J. Cell. Biol.* **115**, 887-903.

Kozak, M. (1992a) A consideration of alternative models for the initiation of translation in eukaryotes. *Crit. Rev. Biochem. Mol. Biol.* **27**, 385-402.

Kozak, M. (1992b) Regulation of translation in eukaryotic systems. *Annu. Rev. Cell. Biol.* **8**, 197-225.

Kozak, M. (1995) Adherence to the first-AUG rule when a second AUG codon follows closely upon the first. *Proc. Natl. Acad. Sci. USA* **92**, 2662-2666 and 7134 (erratum).

Luo, Z., M. Freitag and M.S. Sachs (1995) Translational regulation in response to changes in amino acid availability in *Neurospora crassa. Mol. Cell. Biol.* **15**, 5235-5245.

Miller, K.Y., J. Wu and B.L. Miller (1992) StuA is required for cell pattern formation in *Aspergillus*. *Genes Dev.* **6**, 1770-1782.

Mueller, P.P., B.M. Jackson, P.F. Miller and A.G. Hinnebusch (1988). The first and fourth upstream open reading frames in *GCN4* mRNA have similar initiation efficiencies but respond differently in translational control to changes in length and sequence. *Mol. Cell. Biol.* **8**, 5439-5447.

Orbach, M.J., M.S. Sachs and C. Yanofsky (1990). The *Neurospora crassa arg-2* locus. *J. Biol. Chem.* **265**, 10981-10987.

Paluh, J.L., M.J. Orbach, T.L. Legerton and C. Yanofsky (1988). The cross-pathway control gene of *Neurospora crassa, cpc-1*, encodes a protein similar to *GCN4* of yeast and the DNA binding domain of the oncogene v-*jun*-encoded protein. *Proc. Natl. Acad. Sci. USA* **85**, 3728-3732.

Schutze, M.-P., P.A. Peterson and M.R. Jackson (1994) An N-terminal double-arginine motif maintains type II membrane proteins in the endoplasmic reticulum. *EMBO J.*, **13**, 1495-1501.

Slusher, L.B., E.C. Gillman, N.C. Martin and A.K. Hopper (1991) mRNA leader length and initiation codon context determine alternative AUG selection for the yeast gene *MOD5*. *Proc. Natl. Acad. Sci. USA*, **88**, 9789-9793.

Wanke, C., S. Eckert, G. Albrecht, W. van Hartingsveldt, P.J. Punt, C.A.M.J.J. van den Hondel and G.H. Braus (1997) The *Aspergillus niger GCN4* homologue *cpcA* is transcriptionally regulated and encodes an unusual leucine zipper. *Mol. Microbiol.*, submitted.

Werner, M., A. Feller and A. Piérard (1985). Nucleotide sequence of yeast gene *CPA1* encoding the small subunit of arginine-pathway carbamoylphosphate synthetase: homology of the deduced amino acid sequence to other glutamine amidotransferases. *Eur. J. Biochem.* **146**, 371-381.

Werner, M., A. Feller, F. Messenguy and A. Piérard (1987). The leader peptide of yeast gene *CPA1* is essential for the translational repression of its expression. *Cell* **49**, 805-813.

Williams, N.P., P.P. Mueller and A.G. Hinnebusch (1988). The positive regulatory function of the 5'-proximal open reading frames in *GCN4* mRNA can be mimicked by heterologous, short coding sequences. *Mol. Cell. Biol.* **8**, 3827-3836.

Summary

/

Samenvatting

Summary

The goal of the research described in this thesis was the development and application of a system for the analysis of gene expression in filamentous fungi. The research presented in this thesis has been performed over a period of almost 14 years. In general terms the research can be divided into four parts. In the first part the tools to quantify gene expression in the filamentous fungi *Aspergillus nidulans* and *Aspergillus niger* have been worked out. In the second part of this thesis an *A. niger* gene has been cloned encoding a specific cytochrome P450 enzyme. In the third part the tools developed in the first part have been used to analyze a specific topic of the regulation of gene expression of the gene isolated in the second part. This specific topic relates to the role of small open reading frames present in the 5' region of the mRNA of this gene. In the fourth part a theoretical analysis is reported in which all *Aspergillus* genes, whose sequence is known, are analyzed for the presence of such small upstream open reading frames.

Chapters 2, 3 and 4 describe the set up of tools for systematic analysis of gene expression in *Aspergillus*. When we started this research it was not yet known whether bacterial reporter genes (genes which can be used to easily monitor gene expression) could be expressed in species of the genus *Aspergillus*. Therefore we first tried to express the reporter gene of choice, the *Escherichia coli lacZ* gene (encoding β-galactosidase), as a fusion protein with one of the few enzymes whose gene was already cloned at that moment. We found that the *A. nidulans trpC-E. coli lacZ* fusion gene we constructed, could indeed be expressed at significant levels. But we also found that the expression levels determined in different transformants obtained with the same fusion gene construct were clearly different. We could show that this might be caused by different copy numbers in which the fusion gene was present in the transformants, but also the position of integration of the fusion gene containing plasmid into the genome of the transformant could be influencing the activity which was obtained. Since we wanted to obtain transformants with different fusion genes the activity of which could be compared, we designed a strategy to eliminate the differences found in the first set of transformants. Therefore, we used transformation procedures which resulted in transformants, which had only one copy of the vector integrated at a specific site of the genome. The vectors to be used for gene expression studies, were also modified to accommodate in frame fusion of gene expression control regions including the start of translation. For *A. niger* the selection system for one copy, single locus transformants was even further refined by using a mutated selection gene on the expression analysis plasmids. The tools developed in this part of the research indeed allowed the quantitative comparison of different expression control regions (either mutated versions of the same expression control region or expression control regions of different genes).
The research described in chapter 5 aimed at the improvement of a biotransformation process performed by *A. niger*. For growth on benzoate (or for detoxification of benzoate or related compounds) *A. niger* makes an enzyme system which, as the first step in the breakdown process, hydroxylates benzoate at the para-position. In order to improve this activity we have set out experiments to analyze the gene encoding benzoate para-hydroxylase. This gene, *bphA*, was

cloned using heterologous hybridization techniques. Sequencing of the gene and comparison of the deduced protein sequence with those of all genes of which sequences were available revealed that the *bphA* gene encoded a monooxygenase which was a new member of the cytochrome P450 superfamily (*cyp53A1*). The experiments we performed in the meantime, aiming at overexpression of the *bphA* gene, were successful but did not result in increased enzyme activity, probably since the second component of each microsomal cytochrome P450 enzyme system, the NADPH cytochrome P450 reductase (CPR) was limiting. Improvement of the activity of *A. niger* using the gene encoding CPR will be described in the thesis of Hans van den Brink which will soon be published.

The *bphA* gene contains in its 5' upstream region two small open reading frames encoding 18 resp. 5 amino acids. Such upstream open reading frames (uORFs) are not very often found in eukaryotic genes. In the few instances where their role has been studied, they are involved in the regulation of gene expression at the translational level. Therefore, we have performed research using the tools described in the first part of this thesis to analyze whether the start codons of the uORFs can be used to initiate translation. Furthermore we have analyzed whether mutation of these start codons influences the amount of protein produced starting at the start codon of the main open reading frame. The results of this study are described in chapter 6. Unexpectedly we did not find a clear effect of mutation of the upstream translational start codons, although we could show that the second start codon is very efficiently used to initiate translation of a reporter gene when it is coupled to this upstream start codon. One explanation for our results can be that the upstream open reading frames of the *bphA* gene have no particular function in control of gene expression but it might also be quite possible that we are not (yet) aware of the physiological conditions in which the uORFs play their role.

Since we noticed that in several publications the presence of uORFs in genes which were described were not mentioned, we have carried out a survey of all gene sequences of *Aspergillus* species present in the EMBL/Genbank/GENESEQ data library for the presence of such uORFs. To our surprise we found that of the 162 entries which contained enough information to allow this analysis, 35 (22%) contained at least one uORF. The list of uORF containing genes comprises many genes involved in (the regulation of) important cellular processes such as mitosis, nuclear distribution and movement, differentiation and pH dependent gene regulation. From this study it can be concluded that it will be important to expand research on the regulation of gene expression for the cellular processes mentioned above also to the regulation at the translational level.

Samenvatting

De basenpaarvolgorde en structuur van DNA, de drager van alle erfelijke informatie, bepalen de eigenschappen van levende wezens. De moleculaire genetica is het vakgebied dat gericht is op de bestudering van de structuur en eigenschappen van deze erfelijke informatie. Binnen dit vakgebied is een van de belangrijkste onderwerpen de vraag hoe de genetische informatie van een organisme tot expressie wordt gebracht. Met name de vraag hoe deze genexpressie wordt gereguleerd, is reeds vele jaren onderwerp van studie. Hierbij maakt men gebruik van zeer veel verschillende organismen, waarbij zowel fundamenteel wetenschappelijke belangstelling als ook mogelijke toepassingen bepalend zijn voor de keuze van het organisme. In de afdeling Moleculaire Genetica en Gentechnologie van TNO Voeding (voorheen de afdeling Biochemie van het Medisch Biologisch Laboratorium TNO) wordt reeds een groot aantal jaren moleculair genetisch onderzoek verricht aan diverse soorten organismen. Gedurende de laatste vijftien jaar wordt dit onderzoek geconcentreerd op microorganismen (bacteriën, gisten en schimmels) die worden toegepast in de industrie.

Het moleculair-genetisch onderzoek aan schimmels werd door TNO in 1981 gestart. Schimmels worden in de industrie op zeer grote schaal toegepast voor de produktie van m.n. enzymen, organische zuren en antibiotica. In die tijd waren schimmels een van de weinige soorten microorganismen waarvoor nog geen methode was ontwikkeld voor gerichte genetische modificatie.

In dit proefschrift wordt de ontwikkeling en toepassing beschreven van een systeem dat het mogelijk maakt de regulatie van expressie van erfelijke eigenschappen in schimmels te bestuderen. Omdat het vaak niet eenvoudig is om rechtstreeks te bepalen hoe bepaalde genen tot expressie komen, is gebruik gemaakt van een systeem waarbij de expressie wordt bestudeerd van een zogenaamd reporter gen. Als reporter gen wordt meestal een gen gebruikt dat codeert voor een eenvoudig meetbaar produkt. Door het reporter gen de plaats te laten innemen van het te onderzoeken gen, wordt aan de schimmel een eigenschap toegevoegd (te meten als de activiteit of hoeveelheid van het produkt van het reporter gen) waarvan de vorming op dezelfde wijze gereguleerd wordt als die van het te onderzoeken schimmeleiwit. Op deze wijze kan een, meestal zeer nauwkeurig, beeld worden verkregen van de wijze waarop genen in het betreffende organisme worden gereguleerd.

Na een korte inleiding in hoofdstuk 1, wordt in de hoofdstukken 2, 3 en 4 de ontwikkeling beschreven van het analyse systeem. Hierbij werd als reporter gen gebruik gemaakt van het gen *lacZ* van de bacterie *Escherichia coli* dat codeert voor ß-galactosidase, een enzym waarvan de activiteit goed te meten is. De aanwezigheid van deze enzymatische activiteit kan zowel in schimmels als in extracten worden aangetoond. In de schimmel zelf kan alleen een kwalitatief beeld van de hoeveelheid ß-galactosidase worden verkregen. Een kleuromslag van kleurloos naar blauw bij groei van de schimmel in aanwezigheid van de stof X-gal geeft een indruk van de hoeveelheid ß-galactosidase die wordt geproduceerd. In extracten kan de hoeveelheid van dit reporter eiwit zeer nauwkeurig worden bepaald d.m.v. een enzymbepaling. In het onderzoek beschreven in de betreffende hoofdstukken, werd allereerst bepaald of de gekozen reporter in

schimmels gebruikt kon worden. Toen bleek dat dit het geval was, werden vectoren geconstrueerd waarmee zowel in *Aspergillus nidulans* als in *Aspergillus niger* de regulatie van gen-expressie kon worden bestudeerd.
In het vijfde hoofdstuk wordt onderzoek beschreven dat werd uitgevoerd in opdracht van het Nederlandse chemieconcern DSM. DSM was geïnteresseerd in de biotechnologische produktie van aromatische verbindingen met een hydroxyl groep op de para-positie van de benzeen ring. Als voorbeeld-reactie werd gekozen voor de para-hydroxylering van benzoëzuur door *Aspergillus niger*. Aangezien de natuurlijke activiteit van *A. niger* veel te laag is om een commercieel proces mogelijk te maken, kregen TNO en de Landbouw Universiteit Wageningen de opdracht om een sterke verbetering van dit proces te bewerkstelligen m.b.v. recombinant DNA methoden. Hiertoe werd bij TNO het gen, dat codeert voor het enzym dat de reactie uitvoert, het benzoëzuur para-hydroxylase, geïsoleerd en gekarakteriseerd. In hoofdstuk 5 worden ook de initiële -nog niet succesvolle- experimenten beschreven die tot doel hadden de benzoëzuur para-hydroxylase activiteit van *A. niger* te vergroten. Het voortgezette onderzoek naar de verbetering van deze enzymactiviteit zal worden beschreven in het proefschrift van Hans van den Brink dat over enkele maanden zal verschijnen.
Het benzoëzuur para-hydroxylase gen (*bphA*), bleek een vrij opvallende structuur te hebben in het gebied dat codeert voor het 5' gedeelte van het messenger RNA (het eerste stuk van het mRNA). In dit deel van het mRNA komen twee kleine open leesramen (uORFs = upstream open reading frames) voor. D.w.z. het mRNA codeert hier voor twee oligopeptiden, resp. 18 en 5 aminozuren lang. Het voorkomen van uORFs in schimmelgenen is tot nu toe slechts in enkele gevallen waargenomen. In hoofdstuk 6 wordt een onderzoek beschreven waarin geanalyseerd is welke rol deze uORFs hebben bij de regulatie van de expressie van het *bphA* gen. Hierbij werd gebruik gemaakt van de materialen en methoden waarvan de ontwikkeling is beschreven in de hoofdstukken 2, 3 en 4. Alhoewel duidelijk kon worden aangetoond dat één van de twee uORFs zeer waarschijnlijk wordt 'afgelezen', kon geen duidelijke rol voor de aanwezigheid van deze uORFs in het *bphA* gen worden gevonden.
In het laatste hoofdstuk van dit proefschrift wordt een onderzoek beschreven waarin geanalyseerd is hoe vaak de hierboven omschreven kleine leesramen in het 5' gedeelte van mRNAs van schimmelgenen voorkomen. Hiertoe zijn alle genen van de soort *Aspergillus*, waarvan de DNA sequentie en het startpunt van transcriptie bekend zijn, onderzocht op de aanwezigheid van deze uORFs. Dit onderzoek leidde tot verrassende resultaten. Deze uORFs blijken namelijk veel vaker voor te komen dan tot nu toe werd aangenomen. Met name genen die betrokken zijn bij (de regulatie van) processen die te maken hebben met de celcyclus en differentiatie lijken in veel gevallen één of meer uORFs te bevatten. Mede op grond hiervan mag worden aangenomen dat de expressie van deze genen, behalve op transcriptie niveau, zeer waarschijnlijk ook op translatie niveau wordt gereguleerd. Aangezien hieraan momenteel nog betrekkelijk weinig aandacht wordt besteed, zullen deze resultaten hopelijk er toe bijdragen dat meer onderzoek in deze richting wordt uitgevoerd.

Curriculum vitae

De auteur van dit proefschrift werd geboren op 24 juni 1956 te Rijswijk (ZH). Na een lagere school opleiding in Geleen en Stein volgde hij zijn middelbare schoolopleiding aan het Albert Schweitzer Atheneum en de Scholengemeenschap St. Michiel beiden in Geleen. Hij behaalde in 1974 het diploma Atheneum B aan de Scholengemeenschap St. Michiel. Hierna studeerde hij Moleculaire Wetenschappen (chemische oriëntatie) aan de Landbouw Hogeschool Wageningen. In de doctoraal fase koos hij als hoofdvakken Moleculaire Biologie en Fysische en Kolloïdchemie en liep een stage van 9 maanden in de Moleculaire Biologie bij de sectie recombinant DNA van de afdeling Biochemie van het Medisch Biologisch Laboratorium TNO (MBL-TNO). In januari 1981 behaalde hij het doctoraal diploma aan de LH. Op 15 januari 1981 kwam hij in dienst van dezelfde sectie Recombinant DNA van het MBL-TNO te Rijswijk, eerst als wetenschappelijk medewerker en vanaf 1985 als projectleider. Na aanvankelijk gewerkt te hebben aan de ontwikkeling van vectoren voor *Escherichia coli* stapte hij in 1982 over naar de groep die ging werken aan het ontwikkelen en toepassen van genetische modificatie technieken voor schimmels. In 1988 ging hij leiding geven aan alle industriële projecten die binnen de 'schimmel werkgroep' in uitvoering waren. Op 1 mei 1992 werd hij plaatsvervangend hoofd van de afdeling Moleculaire Genetica en Gentechnologie (voortgekomen uit de sectie Recombinant DNA) van het MBL-TNO en in op 1 december 1993 hoofd van dezelfde afdeling. Organisatorisch werd deze afdeling per 1 januari 1994 onderdeel van de divisie Biochemie en Gentechnologie van TNO Voeding. Per 1 april 1997 zal hij naast zijn huidige functie ook hoofd worden van de afdeling Plantenbiotechnologie van TNO Voeding in Leiden.

List of Publications

van Gorcom, R.F.M., J.A.M. Kröse, P.H. Pouwels, C.A.M.J.J. van den Hondel, K. Wernars, T. Goosen, J. Visser, H.W.J. van den Broek & W.E. Timberlake (1984). Development of a system for the isolation and analysis of promoters from *Aspergillus*: Expression of the *E. coli* β-galactosidase gene in *A. nidulans*. In: Third European Congress on Biotechnology, Vol. III, pp. 305-310.

Goosen, T., K. Wernars, L.M.J. Wennekes, J. Visser, C.J. Bos, H.W.J. van den Broek, **R.F.M. van Gorcom**, A. Kos, C.A.M.J.J. van den Hondel & P.H. Pouwels (1984). Multiple copy integration of transforming sequences in *Aspergillus nidulans*. In: Third European Congress on Biotechnology, Vol. III, pp. 311-316.

Kos A., M.P. Broekhuijsen, **R.F.M. van Gorcom**, J. van Rotterdam, B.E. Enger-Valk, A.J.van Putten, E. Veltkamp, C.A.M.J.J. van den Hondel and P.H. Pouwels (1984). Construction of vectors suitable for efficient expression of genes in *E. coli*, Innovation in Biotechnology (Eds. Houwink E. and Van der Meer R.R.) 415-425.

Wernars K., Goosen T., Wennekens L.M.J., Visser J., Bos C.J., Van den Broek H.W.J., **Van Gorcom R.F.M.**, Kos A., Van den Hondel C.A.M.J.J. and Pouwels P.H. (1984). Transformation of *Aspergillus nidulans* by tandem integration, Proc. Seventh European Meeting on Genetic Transformation, Paris, 151-153.

van Gorcom, R.F.M., Pouwels, P.H., Goosen, T., Visser, J., van den Broek, H.W.J., Hamer, J. E., Timberlake, W.E. & van den Hondel, C.A.M.J.J.(1985). Expression of an *Escherichia coli* ß-galactosidase fusion gene in *Aspergillus nidulans*. *Gene*, **40**, 99-106.

Wernars, K., Goosen, T., Wennekes, L.M.J., Visser, J., Bos, C. J., van den Broek, H.W.J., **van Gorcom, R.F.M.**, van den Hondel, C.A.M.J.J. & Pouwels P.H. (1985). Gene amplification in *Aspergillus nidulans* by transformation with vectors containing the *amdS* gene. *Current Genetics*, **9**, 361-368.

van den Hondel, C.A.M.J.J., **van Gorcom, R.F.M.**, Pouwels, Goosen, T., van den Broek, H.W.J., Timberlake, W.E. & P.H. Pouwels (1985). Development of a system for analysis of regulation signals in *Aspergillus*. In: Molecular Genetics of Filamentous fungi, ed. W.E.Timberlake (Alan R. Liss, inc. New York), pp 29-38.

van Gorcom, R.F.M., P.J.Punt, P.H.Pouwels, C.A.M.J.J.van den Hondel: A system for the analysis of expression signals in *Aspergillus*. *Gene* **48** (1986) 211-217.

van den Hondel, C.A.M.J.J., P.J. Punt, B.L.M. Jacobs-Meijsing, W. van Hartingsveldt, **R.F.M. van Gorcom** & P.H. Pouwels (1986). Analysis of transcription control sequences in Aspergillus. In: Biology and Molecular Biology of Plant-Pathogen Interactions, ed. J. Bailey (Springer-Verlag Berlin Heidelberg) pp. 365-369.

van Gorcom, R.F.M. & van den Hondel, C.A.M.J.J. (1988). Expression analysis vectors for *Aspergillus niger*. *Nucleic Acid Research*, **16**, 9052.

van Hartingsveldt, W., van Zeijl, C.M.J., Veenstra, A.E., van den Berg, J.A., Pouwels, P.H., **van Gorcom, R.F.M.** & C.A.M.J.J. van den Hondel (1990). Heterologous gene expression in *Aspergillus*: analysis of chymosin production in single-copy transformants of *Aspergillus*

van Gorcom, R.F.M., Boschloo, J.G., Kuyvenhoven, A., Lange, J., van Vark, A.J., Bos, C.J., van Balken, J.A.M., Pouwels, P.H. & C.A.M.J.J. van den Hondel (1990). Isolation and molecular characterization of the benzoate-para-hydroxylase gene (*bphA*) of *A. niger*, a member of a new gene-family of the cytochrome P450 superfamily. *Mol. Gen. Genet.* **223**, 192-197.

Gerritse, K., **van Gorcom, R.**, Fasbender, M., Lange J., Schellekens, M., Zegers. N., Claassen, E. & W. Boersma (1990). Specific immuno-detection of benzoate-para-hydroxylase with antibodies raised against synthetic peptides. *Bioch.Biophys.Res.Comm.* **167**, 33-39.

Boschloo, J.G., A. Paffen, T. Koot, W.J.J. van den Tweel, **R.F.M. van Gorcom**, J.H.G. Cordewener & C.J. Bos (1990). Genetic analysis of benzoate metabolism in *Aspergillus niger*. *Appl. Microbiol. Biotech.* **34**, 225-228.

Boschloo, J.G., E. Moonen, **R.F.M. van Gorcom**, H.F.M. Hermes and C.J. Bos (1991). Genetic analysis of *Aspergillus niger* mutants defective in benzoate-4-hydroxylase function. *Curr. Genet.* **19**, 261-264.

Gouka, R.J., W. van Hartingsveldt, R.A.L. Bovenberg, C.A.M.J.J. van den Hondel and **R.F.M. van Gorcom** (1991). Cloning of the nitrate-nitrite reductase gene cluster of *Penicillium chrysogenum* and use of the *niaD* gene as a homologous selection marker. *J. Biotech.* **20**, 189-200.

Van den Hondel, C.A.M.J.J., P.J. Punt and **R.F.M. van Gorcom** (1991). Heterologous gene expression in filamentous fungi. In: More gene manipulation in fungi (Eds. J.W. Bennett and L.L. Lasure), Academic press, pp. 396-428.

Van Gorcom, R.F.M., W. van Hartingsveldt, P.A. van Paridon, A.E. Veenstra, R.G.M. Luiten and G.C.M. Selten (1991). Cloning and expression of microbial phytase. European Patent Application, EP 0420358.

Cooley, R.N., **R.F.M. van Gorcom**, C.A.M.J.J. van den Hondel and C.E. Caten (1991). Isolation of a Benomyl-Resistant Allele of the β-Tubulin Gene from *Septoria nodorum* and its Use as a Dominant Selectable Marker. *J. Gen. Microbiol.* **137**, 2085-2092.

Van Gorcom, R.F.M. (1991) New enzyme treatment for pig and chicken feed may cut phosphate emissions by a third. *Applied Research* **39**, 6-7.

Van Gorcom, R.F.M., J.G.M. Hessing, J. Maat, M. Roza and J.M.A. Verbakel (1991) Xylanase production. PCT International Patent application, WO/91/19782.

Maat, J., M. Roza, J. Verbakel, H. Stam, M.J. Santos da Silva, M. Bosse, M.R. Egmond, M.L.D. Hagemans, **R.F.M. van Gorcom**, J.G.M. Hessing, C.A.M.J.J. van den Hondel and C. van Rotterdam (1992) Xylanases and their application in bakery. In: Proceedings of the International Symposium on Xylans and Xylanases, Wageningen, The Netherlands, December 8-11, 1991, 349-360.

Van den Hondel, C.A.M.J.J., P.J. Punt and **R.F.M. van Gorcom** (1992). Production of extracellular proteins by the filamentous fungus *Aspergillus*. *Anthonie van Leeuwenhoek* **61**, 153-160.

Van Hartingsveldt, W, **R.F.M. van Gorcom**, R.J. Gouka and R.A.L. Bovenberg (1992) Transformation selection marker system adapted for *Penicillium.* European Patent Application, EP 0487118.

Van Hartingsveldt, W., C.M.J. van Zeijl, G.M. Harteveld, R.J. Gouka, M.E.G. Suykerbuyk, R.G.M. Luiten, P.A. van Paridon, G.C.M. Selten, A.E. Veenstra, **R.F.M. van Gorcom** and C.A.M.J.J. van den Hondel (1993) Cloning, characterization and overexpression of the phytase-encoding gene (*phyA*) of *Aspergillus niger. Gene* **127**: 87-94.

Gouka, R.J., W.v. Hartingsveldt, R.A.L. Bovenberg, C.M.J. van Zeijl, C.A.M.J.J.van den Hondel and **R.F.M.van Gorcom** (1992) Development of a new transformant selection system for *Penicillium chrysogenum*: isolation and characterization of the *P.chrysogenum* acetyl-coenzyme A synthetase gene (*facA*) and its use as a homologous selection marker. *Appl. Microbiol Biotechnol* **38**: 514-519.

Van Gorcom, R.F.M., Th. Sonke and C.A.M.J.J. van den Hondel (1993) Expression cassette for heterologous proteins in filamentous fungi. PCT International Patent Application, WO 93/07277.

Hessing, J.G.M., C. van Rotterdam, J.M.A. Verbakel, M.Roza, J. Maat, **R.F.M. van Gorcom** and C.A.M.J.J. van den Hondel (1994) Isolation and characterization of a 1,4-ß-endoxylanase gene of A. awamori. *Curr. Genet.* **26**: 228-232.

Van den Hondel, C.A.M.J.J., P.J. Punt and **R.F.M. van Gorcom** (1994). Strategies for the over-production of fungal proteins in *Aspergillus.* IN L. Alberghina, L. Frontali and P. Sensi (Eds.), Progress in Biotechnology vol.9, Elsevier, Amsterdam, pp. 287-290.

Van Gorcom, R.F.M., P.J. Punt, C.A.M.J.J. van den Hondel (1994). Heterologous gene expression in Aspergillus. In K.A. Powell, A. Renwick and J.F. Peberdy (Eds.), The Genus Aspergillus Plenum Press, New York 241-250.

Van den Brink, J.M. and **R.F.M. van Gorcom** (1994). Oxidoreductase from filamentous fungi, DNA coding therefor and cells transformed with said DNA. PCT International Patent Application PCT/WO/94/29453.

Frenken, L.G.J., **R.F.M. van Gorcom**, J.G.M. Hessing, C.A.M.J.J. van den Hondel, W. Musters, J.M.A. Verbakel, C.Th. Verrips (1994). Process for producing fusion proteins comprising SCFV fragments by a transformed mould. PCT International Patent Application PCT/WO94/29457.

Van den Brink, J.M., C. van Zeijl, C.A.M.J.J. van den Hondel and **R.F.M. van Gorcom** (1995). Cloning and characterization of the NAPDH cytochrome P450 oxidoreductase gene from the filamentous fungus *Aspergillus niger. DNA Cell Biol* **14**: 719-729.

Selten, G.C.M., **R.F.M. van Gorcom** and B-W. Swinkels (1995). Selection marker gene free recombinant strains, a method for obtaining them and the use of these strains. European Patent Application. EP 0635574 A1.

Van den Brink, J.M., J.G.M. van Nistelrooij, M.A. de Waard, C.A.M.J.J. van den Hondel and **R.F.M. van Gorcom** (1996). Increased resistance to 14α-demethylase inhibitors (DMI's) in *Aspergillus niger* by coexpression of the *Penicillium italicum* sterol-14α-demethylase (*cyp51*) and the *A.niger* cytochrome P450 reductase (*cprA*) genes. *J.of Biotechnol.*, in

14α-demethylase (CYP51) from *Penicillium italicum. Mol Gen Genet* **250**: 725-733.

van den Brink, (Hans) J.M., Cees A.M.J.J. van den Hondel and **Robert F.M. van Gorcom** (1996). Optimization of the benzoate-inducible benzoate *p*-hydroxylase cytochrome P450 enzyme system in *Aspergillus niger. Appl Microbiol Biotechnol (1996)*, **46**:

Faber, F., M. van Giezen, **R. van Gorcom** and W. Harder. (1996). Identification of two *Echerichia coli* K12 proteins which are induced in response to pollutant stress. *Thesis F. Faber, University of Groningen,* pp. 55-72.

Faber, F., H. Hessing, **R. van Gorcom** and W. Harder. (1996). Construction of firefly luciferase reporter-plasmids containing promoters of proteins which expression was affected by pllutant stress. *Thesis F. Faber, University of Groningen,* pp. 73-100.

Faber, F., **R. van Gorcom**, R. Leer, H. Vonk and W. Harder. (1996). Characterization of the effects of pollutants on the expression of luciferase from reporter-plasmids in *Escherichia coli* K12 by *in vivo* activity measurements. *Thesis Faber, University of Groningen,* pp. 101-128.

van Gorcom, R.F.M., J.M. van den Brink and C.A.M.J.J. van den Hondel (1006). Improvement of Cytochrome P450 Mediated Biotransformations by Filamentous Fungi. *Chimia 50*, 433-434.

Dankwoord

Bij het tot stand komen van een proefschrift dat een verzameling is van stukken onderzoek uitgevoerd in een periode van 14 jaar is het niet eenvoudig je te realiseren, wie hieraan op welke wijze hebben bijgedragen. Ik zal dan ook ongetwijfeld een aantal mensen vergeten te vernoemen, maar ik hoop dat zij mij dit niet kwalijk nemen.
Bij het schrijven van dit dankwoord heb ik even stilgestaan bij mijn vorming als onderzoeker. Mijn geheugen is nogal slecht, of misschien beter gezegd selectief, maar als ik de verhalen mag geloven ging ik er graag reeds op zeer jonge leeftijd op uit om mijn omgeving en hetgeen zich daarin afspeelde te onderzoeken. Gezien de ruimte die ik daarvoor kreeg, zijn mijn eerste 18 jaar thuis een belangrijke voedingsbodem geweest voor mijn verdere vorming. Ik wil mijn ouders hiervoor dan ook van harte bedanken.
Mijn belangrijkste opleider/vormer in het wetenschappelijk onderzoek in de afgelopen 14 jaar was mijn co-promotor Kees van den Hondel. Ik vind het jammer voor de huidige AIO's dat Kees minder tijd heeft gekregen voor het dagelijks begeleiden van hun onderzoek. Ik was nooit van hem af. Elke discussie waarvan ik dacht dat hij afgerond was, kreeg een vervolg, meestal de volgende morgen om halfnegen. Karakteristiek was jouw start Kees: "Ik heb nog eens nagedacht, maar ....". Kees bekijkt een resultaat, een probleem, een hypothese, een uitspraak, etc. keer op keer, belicht het van diverse kanten en meestal vindt hij daardoor iets wat we in de eerdere gesprekken over het hoofd hadden gezien. Kees, je weet dat ik jouw kritisch opbouwende houding zeer op prijs stel en ik hoop dat we nog vele jaren in dit vak zullen samenwerken.
Mijn 'carrière' bij TNO ben ik begonnen als student bij mijn promotor, Peter Pouwels. Ik ben Peter in de afgelopen ruim 16 jaar in Rijswijk op verschillende manieren tegengekomen. Keer op keer, in welke verhouding dan ook, leer ik steeds weer van je. Elk artikel of ander stuk dat in jouw tas verdwijnt, komt er vaak na zeer korte tijd weer uit met een groot aantal opmerkingen en suggesties die hout snijden en meestal het geheel substantieel doen verbeteren. Hoofdstuk zes van dit proefschrift was voor wat betreft de 'korte tijd' een duidelijke uitzondering: Peter was bang dat zijn belangrijkste opmerking (een prima hoofdstuk, maar niet publiceerbaar) voor mij moeilijk te verteren zou zijn. Peter ik wil je van harte bedanken voor de fijne samenwerking de afgelopen jaren en met name voor de positief kritische bijdrage aan dit boekje. Tot op de laatste dagen bleef je snel en steeds uiterst scherp reageren.
Ook de tweede Peter in onze afdeling (Peter Punt) behoort in dit dankwoord vernoemd te worden. Peter heeft op twee tijdstippen in het in dit proefschrift beschreven onderzoek een belangrijke rol gespeeld. Als co-auteur van het in hoofdstuk 3 beschreven werk en als drijvende kracht in het werk uit hoofdstuk 6. Toen ik de mogelijkheid kreeg om projectleider te worden in een in 1985 te starten project voor DSM was Peter net actief in het genexpressie onderzoek. Peter heeft toen de laatste eindjes van het in hoofdstuk 3 beschreven werk aan elkaar geknoopt en heeft het onderzoek voortgezet met name in de richting van het toepassen van de ontwikkelde 'tools'. Toen ik er in 1990 achter kwam dat ergens diep in mij iets zei, dat ik toch best graag wilde promoveren en toen het onderzoek voor DSM, met name het *bphA* promotor werk, hiertoe ineens nieuwe kansen leek te gaan bieden, was Peter van harte bereid om de afrondende werkzaamheden (mede) te begeleiden. Deze klus bleek nog tamelijk pittig, maar gelukkig hebben

de diverse hoofdstukken.

In de loop van de afgelopen veertien jaar hebben diverse mensen met mij samen aan de verschillende onderzoeksprojecten gewerkt. Één iemand wil ik nog iets uitgebreider bedanken, Anneke Drint-Kuijvenhoven. Anneke was ten tijde van het onderzoek beschreven in hoofdstuk 5 de drijvende kracht. Zolang het neerkwam op plannen, analyse en rapportage kon ik de kar best trekken, maar zodra het neerkwam op planning en uitvoering was mijn rol duidelijk anders. Ik voelde mij dan het best op mijn gemak als hulpje van mijn analist. Zo ik al eens de pretentie had om mij een grotere rol aan te meten, werd mij haarfijn (gestaafd met feiten!) duidelijk gemaakt dat ik meer voor iets anders in de wieg was gelegd. Anneke, je weet dat ik altijd heel graag met je heb samengewerkt en ik hoop dat ik in die tijd, waar ik erg leunde op jouw inzet en kwaliteit, niet te veel van je heb gevraagd. Echter bovenal wil ik je bedanken voor de (misschien onbedoelde, maar erg belangrijke) lessen die ik van je gehad heb in het leidinggeven aan collega's en voor het helpen kuisen van mijn taalgebruik.

Aan het in dit boekje beschreven onderzoek hebben nog veel meer mensen meegewerkt. In alfabetische volgorde wil ik jullie, middels het noemen van jullie namen, van harte bedanken: Sandra van den Broek, Jacqueline Busscher-Lange, Inge Frantzen, Marjolein van Helmond, Yolanda de Ruiter en Theo Sonke. Ook de andere leden van de afdeling MGG (vroeger de sectie recombinant DNA) en de ondersteunende diensten van het MBL en nu TNO Voeding wil ik niet vergeten (ik ga niet iedereen noemen maar met name ook Jeanne, Michel, Wim, Lo, Marius en Floor). Zonder de unieke sfeer en collegialiteit die we gelukkig nog steeds in Rijswijk kennen is leuk werk en goed onderzoek onmogelijk. Ik ben er van overtuigd dat we dit de komende jaren, ook in een nieuwe omgeving, als drijvende kracht weten te behouden.

Naast de collega's binnen het instituut is aan dit werk ook een belangrijke bijdrage geleverd door collega's van buiten TNO. Hun namen staan meestal vermeld als auteur bij het betreffende hoofdstuk, maar ik wil een aantal mensen ook hier graag vernoemen. Het werk met schimmels is bij TNO gestart als samenwerkingsverband met de LUW vakgroep Erfelijkheidsleer. Karel, Theo, Bert, Henk, Jan, Harry, Jaap, Johan, Edu, Klaas en Kees, we hebben destijds veel van elkaar kunnen leren, hartelijk dank daar voor. Ik vind het nog steeds jammer dat de contacten wat minder zijn geworden. Ook de samenwerking met DSM Research heb ik altijd zeer gewaardeerd (niet alleen omdat ik mij enigszins Limburger voel). Zonder de anderen tekort te willen doen wil ik met name de contacten met Hans van Balken (inductie was voor de een niet altijd wat de ander er onder verstond) en Theo Sonke noemen. Ik heb in deze samenwerking geleerd hoe leuk het is om vanuit totaal verschillende disciplines en invalshoeken aan een gezamenlijk probleem te werken.

Als laatste wil ik Christina, Jort en Luuk bedanken voor het geduld dat zij hebben gehad met een echtgenoot/vader die vaak geen tijd had voor huiselijke plichten en geneugten. Gedurende een lange periode waren de weekends geblokkeerd voor leukere zaken. Een bijkomend nadeel voor de jongeheren was de blokkade van "hun" computer, omdat Robert weer zo nodig moest werken (zelfs bij een van Jort z'n vrienden kreeg ik een slechte naam). Jongens, dat is weer voorbij. Motorracen, Keen en de andere spelletjes staan weer tot jullie beschikking en vooral, we gaan er weer lekker op uit!