1

MICRO

National Library
of Canada

Bibliothèque nationale
du Canada

Canadian Theses Service

Service des thèses canadiennes

Ottawa, Canada
K1A 0N4

NOTICE

The quality of this microform is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us an inferior photocopy.

Previously copyrighted materials (journal articles, published tests, etc.) are not filmed.

Reproduction in full or in part of this microform is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30.

AVIS

La qualité de cette microforme dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de qualité inférieure.

Les documents qui font déjà l'objet d'un droit d'auteur (articles de revue, tests publiés, etc.) ne sont pas microfilmés.

La reproduction, même partielle, de cette microforme est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30.

Cloning And Expression Of The Aspergillus oryzae
 α -Amylase Gene

by

© María Jesús Ginés B.Sc., M.Sc.

A thesis submitted to the Faculty of Graduate Studies
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy

Biology Department
Carleton University
Ottawa, Ontario
February, 1987

Permission has been granted to the National Library of Canada to microfilm this thesis and to lend or sell copies of the film.

The author (copyright owner) has reserved other publication rights, and neither the thesis nor extensive extracts from it may be printed or otherwise reproduced without his/her written permission.

L'autorisation a été accordée à la Bibliothèque nationale du Canada de microfilmer cette thèse et de prêter ou de vendre des exemplaires du film.

L'auteur (titulaire du droit d'auteur) se réserve les autres droits de publication; ni la thèse ni de longs extraits de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation écrite.

ISBN 0-315-46236-1

• A mis padres

A Danah y Aitana

The undersigned hereby recommend to the Faculty of
Graduate Studies acceptance of this thesis, submitted by
María Jesús Ginés, B.Sc., M.Sc., in partial fulfillment of
the requirements for the degree of Doctor of Philosophy

Chairman, Department of Biology

Supervisor

Co-supervisor

External Examiner

ABSTRACT

The filamentous fungus Aspergillus oryzae secretes large amounts of α -amylase upon growth in starch medium. Highest levels of α -amylase specific mRNA measured by in vitro translation (IVT) were found between 6 to 15h of growth on starch, essentially verifying earlier work (Erratt et al. 1984).

Poly(A)⁺ RNA isolated from starch cultures at 10h was used as template for cDNA synthesis and construction of a cDNA library using a M13 phage vector and Escherichia coli JM103 as host. Three libraries of 8,000 recombinant clones each were screened for α -amylase sequence using a mixture of synthetic 20-mer oligodeoxyribonucleotides directed to the carboxy terminal coding region of the published Taka-amylase (TAA) sequence (Toda et al. 1982). Of the 5 positive recombinant clones initially detected by hybridization with the probe mixture, two of these clones, M71 and M91, were demonstrated by sequencing to be α -amylase specific. Clone M71, which contained 115 codons of the carboxyl-end and 154 nucleotides of 3'untranslated region (UTR), was used as a probe for further identification of chromosomal sequences.

Two genes instead of one were found, corresponding to EcoRI fragments of 3.35 kb (designated amylase II) and 7.6 kb (amylase I), by genomic Southern blots and subsequent recovery from a Lambda Charon 4A-A.oryzae genomic bank. Restriction endonuclease and hybridization analysis of these fragments subcloned into pUC8 revealed a high degree of homology between them, lying within a common 2.64 kb EcoRI/BglII fragment. The nucleotide sequence of a 2.85 kb EcoRI/HindIII region was obtained and the predicted protein sequence compared with the TAA. Eight intervening sequences (IVS)

within the coding region were found, ranging from 55 to 86bp in length. Three of the introns (IVS-4, IVS-5 and IVS-6) divide the putative catalytic and substrate binding site domains predicted from the 3-D structure of TAA by Matsuura et al. (1984). The precursor amylase has 499 amino acids, 21 residues larger than the published sequence of TAA. Only 9 differences in amino acids were found. Little or no homology was found when the signal peptide and mature protein sequence as well as the nucleotide sequence were compared to those available of other genes, such as the glucamylases of the yeast Saccharomyces diastaticus and the fungi Aspergillus awamori and Aspergillus niger.

Nucleotide sequences of the 5' and 3' untranslated regions of amylase I were also obtained and compared to those of amylase II and to several of the additional cDNAs recovered by rescreening the M13-cDNA banks with amylase II as probe. These studies clearly revealed that the 5' regions were identical and that both genes, which differ in their 3' UTRs are transcribed. Some heterogeneity in polyadenylation signals and putative sites for poly(A) addition are apparent. The major fraction of amylase mRNA was measured at about 1,800 nucleotides in length, which is close to the expected size (~ 1,700-1,750 bases). The expression of mRNA during growth on starch is consistent with the IVT estimates. The Northern analysis further indicates that the amylase mRNA undergoes rapid degradation after 15 hours of growth on starch. Evidence for the detection of partially spliced mRNA is also presented.

ACKNOWLEDGMENTS

My special thanks to Dr. Vern Seligy for his long-term commitment in teaching, encouraging, and believing in my capabilities. His constant help and dedication have been key elements in the completion of this project.

I am indebted to Michael Dove for his skill, patience and help with lots of the experimental procedures and sequencing of the c-amylase genes.

Special thanks to Dr. H. Yamazaki and Dr. V.N. Iyer for their advice.

I thank all the other members of the group for their support and encouragement throughout the course of this work. Specifically to Bob Rutledge for his help in the construction of cDNA banks and to Shah and Steve for the late-night discussions.

I am also grateful to Lise Brammall for her help with the computer analysis of the sequencing data.

The receipt of a Government of Canada Award to Foreign Nationals as well as support provided by NRCC and NSERC to V.L.S. is greatly appreciated.

Finally, I would like to thank my friends and family for putting up with me through the tough periods and sharing with me the happier ones.

TABLE OF CONTENTS

Title	i
Abstract	iii
Acknowledgements	v
Table of Contents	vi
List of Tables	ix
List of Figures	x
List of Abbreviations	xii
 I. Introduction	 1
A) Starch As A Substrate	1
B) Enzymatic Degradation Of Starch	3
C) The α -Amylase Family	
a) General Properties	3
b) Amylases Of Bacteria	6
c) Amylases Of Eukaryotes	6
D) Amylases Of <u>A.oryzae</u>	
a) Role of Calcium and Enzyme Sulphydral Groups	7
b) Structure of Taka-amylase (TAA)	8
c) Primary Sequence Homologies	9
d) Comparison of TAA with other amylolytic enzymes	10
e) Expression and Regulation	11
E) Research Objective	12
 II. Materials and Methods	
A) Growth of <u>A.oryzae</u>	14
B) Amylase Activity, Purification and Antibody Production	14
C) Isolation of RNA	15

D) Analysis of Amylase RNA Expression By DNA-RNA Hybridization ..	16
E) Cell Free Protein Synthesis	18
F) Synthesis of cDNA	19
G) Construction of cDNA Banks	20
H) Construction of Fungal DNA/Genomic Libraries in Lambda	21
I) Isolation and Purification of DNA From Fungi and Bacteria	22
a) Preparation of <u>Aspergillus</u> DNA	22
b) Isolation of single stranded M13 DNA	23
c) Plasmid and M13 isolation	23
J) DNA Analysis by Restriction Endonucleases and Electrophoresis..	24
K) Initial Screening of Clone Banks For Amylase DNA	25
L) Initial Isolation and Characterization of Amylase cDNA Clones	27
a) Insert sizing	28
M) Screening Genomic Clones Using Amylase cDNA Probe	29
N) Subcloning of Amylase DNA Sequences	
a) Subcloning into pUC8	30
b) Subcloning into M13 mp18/19	31
O) Establishing 5'-3' Orientation of Cloned Inserts	32
P) Rescreening The cDNA Banks With Genomic Amylase DNA	33
Q) DNA Sequencing	34

III. Results

(A) Cloning of Amylase Gene Sequences	
a) cDNA and mRNA Production	35
b) Construction and screening of cDNA clones	37
c) Analysis of cDNA clones by nucleotide sequencing	42
d) Identification of genomic sequences with cDNA probes	54
e) Cloning of genomic sequences	58

• B) Comparison of Amylase Genes

- a) Subcloning of genomic amylase DNA in pUC and M13 vectors 62
- b) REN mapping and chromosomal linkage of Amylase I and II genes. 62
- c) Sequencing and organization of Amylase II 73
 - i) introns 73
 - ii) coding region 78
 - iii) nucleotide composition and codon usage 79
- d) Comparison of 5' and 3' untranslated regions of Amylase genes
I and II 83

C) Expression of Amylase I and II

- a) Identification of new cDNA clones related to amylase genomic
sequence 87
- b) mRNA levels in starch and glucose grown mycelia 91

D) Comparison of Amylase Genes to Genes From Other Organisms

- a) Genomic DNA hybridization studies 95
- b) Sequence comparisons 97

IV. Discussion

- A) Cloning Strategy, Comparison to Other Glycosidase Genes Cloned 106
- B) The Coding Region of Amylase II Gene 110
- C) Comparison of Amylase II Nucleotide Sequence With Other Genes.
 - a) Codon usage 113
 - b) 5' untranslated region 113
 - c) Intron structures 115
 - d) 3' Untranslated region 117
- D) Significance of the Amylase Cloning Study..... 119

V. References 123

List of Tables

Table I.	Classes of Amylolytic Enzymes	4
Table II.	Properties of M13mp9-cDNA recombinant phage banks made from double stranded cDNA synthesized off 10 hour polyA+ RNA ...	39
Table III.	Status of λ -Charon 4A.Ao amylase genomic clones	59
Table IV.	Summary of subclones containing EcoRI fragment I	63
Table V.	Summary of subclones containing EcoRI fragment II	64
Table VI.	Comparison of intervening sequences from <u>A.oryzae</u> amylase II	77
Table VII.	Differences between the predicted amino acid sequence of amylase II and the TAA	80
Table VIII.	Location of catalytic sites for TAA and amylase II based on exon distribution	81
Table IX.	Base composition of <u>A. oryzae</u> DNA	82
Table X.	<u>A. oryzae</u> amylase II codon usage	84
Table XI.	Amylase related REN fragment size of genomic DNA from <u>A.oryzae</u> and <u>A.niger</u>	97

List of Figures

Figure 1.	Mechanism of starch hydrolysis by amylolytic enzymes	2
Figure 2.	Production of RNA and amylase-specific mRNA during growth of <u>A.oryzae</u> mycelia on starch containing medium	36
Figure 3.	Estimate of double stranded cDNA synthesized from polyA+ RNA of <u>A.oryzae</u> grown for 10 hours on starch medium	38
Figure 4.	Plaque hybridization and purification of amylase cDNA clone M71 using oligonucleotide probe 11	41
Figure 5.	Sizing of cDNA of M71 by use of primer extension	43
Figure 6.	C-test of M71 cDNA insert cloned in opposite orientations ..	44
Figure 7.	Nucleotide sequence characterization of the M71 cDNA insert.	
	A. Nucleotide sequence of the 500 bp insert determined by dideoxy method of Sanger and co-workers	47
	B. Conversion of nucleotide sequence to amino acid sequence.	48
	C. Sequence alignment with Taka-amylase amino acid sequence of <u>A.oryzae</u> provided by Toda <u>et al</u> , 1982	49
Figure 8.	Nucleotide sequence characterization of the M91 cDNA insert.	
	A. Nucleotide sequence of the 386 bp insert determined by dideoxy method of Sanger and co-workers	51
	B. Conversion of nucleotide sequence to amino acid sequence.	52
	C. Sequence alignment with Taka-amylase amino acid sequence of <u>A.oryzae</u> provided by Toda <u>et al</u> , 1982	53
Figure 9.	REN mapping of cDNA insert of M71 clone	55
Figure 10.	Mapping of 3' amylase genomic DNA homology using M71 cDNA ..	57
Figure 11.	Screening of Charon 4A.Ao genomic library with M71 probe and purification of clone 11G1	60
Figure 12.	EcoRI cleavage and hybridization pattern of DNA prepared from plaque pure recombinant phage containing putative amylase genes	61
Figure 13.	REN analysis and map of pUC8-76 containing the 7.6kbp EcoRI fragment with the amylase gene I sequence	65

Figure 14. REN analysis and map of pUC8-76 containing the 3.35 kbp EcoRI fragment with the amylase gene II sequence	67
Figure 15. REN mapping of amylase I and II chromosomal genes and flanking regions	70
Figure 16. Complete nucleotide sequence of the 2.8 kbp EcoRI-HindIII fragment containing the amylase II gene.	74
Figure 17. Nucleotide sequence of the promoter region for <u>A.oryzae</u> amylase II gene	76
Figure 18. General map showing the structural features and organization of the amylase II gene including position of various cloned cDNA derivatives	88
Figure 19. Sequence alignment of the 3'-UTR of <u>A.oryzae</u> amylase genes I and II	86
Figure 20. Time course of appearance of α -amylase mRNA after induction with starch	93
Figure 21. Mapping of c-amylase genomic DNA homology of <u>A.oryzae</u> and <u>A.niger</u> using M71 cDNA probe	96
Figure 22. DNA sequence alignment of the 5' promoter and signal peptide regions of <u>A.oryzae</u> α -amylase II gene and fungal glucoamylases	98
Figure 23. Comparison of signal peptides of <u>A.oryzae</u> amylase II gene versus fungal glucoamylases	101
Figure 24. DNA sequence alignment of the 3'-untranslated regions of <u>A.oryzae</u> amylase II gene against fungal glucoamylases	102
Figure 25. Comparison of signal peptides of various α -amylases	104

LIST OF ABBREVIATIONS

$[\alpha^{32}\text{P}]\text{dATP}$	$[\alpha^{32}\text{P}]$ deoxyadenosine-5'-triphosphate
APS	ammonium persulfate $(\text{NH}_4)_2\text{S}_2\text{O}_8$
BSA	bovine serum albumin
Bis acrylamide	N,N'-methylenebisacrylamide
b.p	base pairs
BPB	bromo-phenol-blue
Buffer C	60% formamide, 2.5mM EDTA pH 8.0, 0.3% SDS, 0.1% BPB
Buffer I	50 mM glucose, 25mM Tris HCl pH 8.0, 10mM EDTA
cDNA	complementary DNA
Ci	curie
cpm	counts per minute
CR	coding region
ddH ₂ O	deionized distilled water
DEAE	diethylaminoethyl
DEP	diethyl oxydiformalate
DMSO	dimethyl sulfoxide
DNA	deoxynucleic acid
dNTPs	deoxynucleotide triphosphates
DTT	dithiothreitol
EDTA	ethylenediaminetetracetic acid
EtBr	ethidium bromide
HACoCl ₃	hexamine cobalt(III) chloride
H-buffer	10mM Tris HCl pH 7.5, 5mM MgCl ₂
H.S.buffer	high salt buffer (50mM Tris pH 7.5, 100mM NaCl, 10mM MgCl ₂ , 1mM DTT)

IPTG	isopropylthiogalactoside
L.S.buffer.	low salt buffer (10mM Tris pH 7.5, 10mM MgCl ₂ , 1mM DTT)
MES	morpholinoethane sulfonic acid
mM	millimolar
mRNA	messenger RNA
M.S.buffer	medium salt buffer (10mM Tris HCl pH 7.5, 50mM NaCl, 10mM MgCl ₂ , 1mMDTT)
MW	molecular weight
NCF	nitrocellulose filters
NZY	1% casamino acids, 0.5% yeast extract, 0.5% NaCl, 0.2% MgCl ₂
OD	optical density
PAGE	polyacrylamide gel electrophoresis
PMSF	phenylmethylsulfonylfluoride
RNA	ribonucleic acid
[³⁵ S]dATPαS	deoxyadenosine 5'-(α ³⁵ S)thio)triphosphate
SDS	sodium dodecyl sulfate
SSC	sodium saline citrate buffer (0.15M NaCl, 0.15M NaCitrate, pH 7.0)
STE	10 mM Tris HCl pH 8.0, 100mM NaCl, 1mM EDTA pH 8.0
Stop-buffer	100mM EDTA, 0.5% SDS, 25% glycerol, 0.1% BPB
TAE	Tris-acetate EDTA buffer (0.04M Tris-acetate, 0.002M EDTA pH 8.0)
TBE	Tris-Borate-EDTA buffer (89mM Tris-base, 89mM boric acid, 2mM Na ₂ EDTA pH 8.0)
TCA	trichloroacetic acid
TE	10mM Tris HCl pH 7.6, 1mM EDTA
TEMED	N,N,N',N'-tetramethylethylenediamine

TFB 10mM-K-MES, pH 6.2, 10mM RbCl, 45mM, $\text{MnCl}_2 \cdot 4\text{H}_2\text{O}$,
4mM $\text{CaCl}_2 \cdot 2\text{H}_2\text{O}$, 3mM HACoCl_3
Tris tris(hydroxymethyl)aminomethane
tRNA transfer RNA
2YT 1.6% tryptone, 1.0% yeast extract, 0.5% NaCl
UTR untranslated region
U.V. ultra violet
X-gal 5-bromo-4-chloro-3-indolyl-B-D-galactoside

I. INTRODUCTION

A. Starch As A Substrate

Starch is a polymer of glucose which occurs in two main structural forms: 1) amylose, a linear chain containing up to 6000 glucose units connected by α -1,4 linkages, and 2) amylopectin, a highly branched structure consisting of short-amylose chains of about 10 to 60 glucose units (average 20), connected by α -1,6 linkages or branch points that involve about 5% of the total glucose molecules. A diagrammatic representation of these structural features is shown in figure 1. Amylose makes up about 15 to 30% of the common starches (Swinkels, 1985). The ratio of amylose to amylopectin is fairly constant for a given species of starch (Roehrig, 1984).

Starch occurs mainly in a complex, granular form in various organs of plants such as, seeds, tubers, roots and stem piths. Starch is also found in a number of plant-related lower eukaryotes and probably is the most abundant, renewable carbohydrate in nature after cellulose (Swinkels, 1985). The methodology for industrial processing of starch granules is well established. Total global production of starch is about 18 million tons, obtained mainly from maize, potatoes and cereal grains, (Guilbot and Mercier, 1985). At present, only about 1% of the total agriculturally produced starch is further processed (Van Beynam and Roels, 1985), used mainly to improve the functional properties of food products and to make high fructose corn syrups and ethanol. Starch is also used in glue, coating, sizing and flocculating agents, chemicals and building materials. Like cellulose, the application potential for starch is still largely untapped; in the event that oil-based products become

Figure 1. Mechanism of starch hydrolysis by amylolytic enzymes.

Starch consists of various lengths of linear polymers of glucose (A) linked by α -1,4 and also by α -1,6 glycosidic bonds to form branch points (B). Symbols $\bullet$, $\blacksquare$, $\rightarrow$ and $\triangleright$ denote the respective attack points by α -amylase, glucoamylase, β -amylase and pullulanases and isoamylases, respectively. $\bullet$ the reducing end.

scarce, these carbohydrate polymers might provide the materials for the "biological plastics" industry (Guilbot and Mercier, 1985).

The commercialization of starch has been enormously enhanced through the application of amylolytic enzymes, glucose isomerase and fermentative microbes. However, in spite of the large amount of research activity in these areas (see review by Fogerty and Kelly, 1979), progress has been slow; most of the enzymes, and their modes of production and regulated expression are still very poorly understood, including their optimal use in various fermentation processes. With the new advances in genetic and protein engineering technologies, considerable progress is expected to be made in the amylase field.

B. Enzymatic Degradation Of Starch

Several types of enzymes have been identified and used at some time or another for the hydrolysis of starch. These include: α -amylase, β -amylase, glucoamylase, isoamylase and pullulanase (Reilly, 1985). As indicated in Table I, these enzymes are produced by a large number of diverse organisms. Microbials such as, bacteria and fungi have received much of the focus because their enzymes are produced extracellularly, often in moderate to high amounts, and are easy to manipulate.

The mechanism of starch hydrolysis of each class of enzyme is different (Table I and Figure 1). Within each enzyme class there is also a considerable amount of diversity. The significance of this diversity is not fully understood or appreciated, but likely reflects the peculiar genetic advantages evolved by host cell environmental adaptation.

TABLE 1: Classes of Amylolytic Enzymes

Class	Function
α -amylase(1,4- α -glucan-glucanohydrolases)	generally extracellular enzymes produced by bacteria, fungi, some yeasts, plants, and animals. Hydrolysis of α 1,4-glycosidic linkages, splits the substrate internally
β -amylases (α 1,4-glucan-maltohydrolases)	usually plant origin, some microbial producers, bacteria, actinomycetes, fungi calcium not required for stabilization. act on starch by splitting maltose units from nonreducing end
glucoamylases (α 1,4-glucan-glucohydrolases)	act on starch by splitting glucose units from nonreducing end, breaks down maltose slowly
pullulanases, isoamylases (1,6-glycosidases)	splits side chains of amylopectin by hydrolysis of α 1,6-bonds pullulanases and isoamylases directly affect amylopectin

C. The α -Amylase Family

(a) General Properties

Amylase-producing organisms are widely found in nature. However, the majority of these enzymes have been isolated and characterized from Bacillus bacteria and species of Aspergillus fungi, which are most often used in industry (Dasinger *et al.* 1985) because they are "generally regarded as safe" (GRAS) in foods by Food and Drug Administrations world wide. This enzyme class cleaves the amylose from within the glucose chain, retaining the anomeric configuration of the bond and releasing mainly maltobiose, maltotriose and some free glucose. After an apparent initial random cleavage most enzymes exhibit preferential hydrolysis of one of the fragments. Their action is not inhibited by α -1,6-glycosidic bonds.

The physical properties of the substrate as well as its concentration can influence the performance of individual enzymes. With amylopectin and glycogen, the hydrolysis of α -1,4-glycosidic chains is modified by the degree of branching. This is because the branch points are resistant to hydrolysis and confer resistance on neighbouring α -1,4-linkages. According to Matheson and McCleary (1985), the smallest branched fragments produced by A.oryzae amylase are:

Amylases differ from each other mainly in their relative starch-liquifying and/or saccharogenic effect and their physical properties

such as pH optimum, temperature range and stability. Starch-liquifying amylases such as from B.amyloliquifaciens and B.licheniformis, break down starch polymers without producing free sugars as do saccharogenic amylases.

(b) Amylases from Bacteria

The most intensely studied and used bacterial amylases are from the genus Bacillus. (MacKay et al., 1985; Ihara et al., 1985). Several of the corresponding genes have now been cloned and characterized and the primary (amino acid sequence) structures and mechanism of secretion of the enzymes have been elucidated (MacKay et al., 1985; Ihara et al., 1985). Detailed studies are still lacking on the gene expression and the substrate induction-repression properties of bacterial amylases, albeit commercial applications of genetically modified strains have been described.

(c) Amylases from Eukaryotes

Amylases have been isolated from both lower eukaryotes such as fungi and yeast and higher eukaryotes such as animals, insects and plants. In some cases the amino acid sequence is known as well as that of the gene and its expression (see reviews by MacKay et al., 1985; Rogers, 1985; Nakajima et al., 1986). Surprising, work on the α -amylases of fungi and yeast, until very recently, has not been very extensive except, in the case of α -amylase of A.oryzae (Toda et al., 1982; Matsuura et al., 1984; Erratt et al., 1984; MacKay et al., 1985). The level of genetic characterization has lagged considerably behind that of bacterial work mainly because the genes coding for these enzymes are not readily cloneable by heterologous expression. Much of the recent effort has gone into the genetic characterization of glucoamylases which show very little

primary sequence homology with amylases (MacKay et al. 1985).

Some discussion is in order on the famous "Taka-diastase" or "Taka-amylase" of A.oryzae which was developed by Dr. Jokichi Takamine, the "Japanese Pasteur", as the first patented microbial industrial enzyme (Takamine, 1894; see Bennett, 1985).

D. A.oryzae α -amylase

Most of the earlier studies on this amylase were done using a commercial source, "Taka-diastase" (from Sankyo Co., Tokyo, Japan) which was generally produced in batches and therefore subject to variation due to strain and growth conditions.

a) Role of Calcium and Enzyme Sulphydral Groups

The Taka-amylase A (TAA) consists of four disulphide (-S-S-) bridges and one free sulphydral group (Narita et al., 1966). One mole of calcium (Ca^{++}) per mole of enzyme is required for enzymatic activity and structural stability (Oikawa and Maeda, 1957; Oikawa, 1959). However, nine additional moles of Ca^{++} per mole of enzyme can be bound and removed without any affect on either the physical or chemical properties of the enzyme. Takagi and Isemura (1965) found that reoxidation of reduced TAA requires the presence of Ca^{++} for recovery of enzyme activity. The enzyme can be denatured by low pH (Tagaki and Toda, 1962), urea (Tagaki and Isemura, 1962; Isemura et al. 1963) and EDTA (Friedmann and Epstein, 1967). Neutralizing, diluting or removal of the offending agent or addition of Ca^{++} promotes renaturation. Friedman and Epstein (1968) indicated that reoxidation involved conformational transitions of large domains of the protein induced by the Ca^{++} -dependent closure of one critical -S-S-. The free sulphydral group is shielded by the essential

Ca^{++} (Toda et al. 1968) and apparently maintains the active configuration of the enzyme by chelating with the essential calcium atom.

b). Structure of Taka-amylase

The amino (N) and carboxy (C) terminal residues are alanine (Akabori and Ikenaka, 1955) and serine (Narita et al. 1966). Large portions of amino acid sequence of both the N and C termini were obtained by Narita (1975) and the complete amino acid sequence of TAA was reported by Toda et al. (1982). The mature enzyme consists of 478 amino acids and carbohydrate is attached to asparagine-197 (see next section for further details).

The three-dimensional structure of TAA was investigated by X-ray structure at the 3 Angstrom resolution level using the multiple isomorphous method (Matsuura et al. 1979; Matsuura et al. 1980). A complete molecular model was constructed by Matsuura et al. (1984) by fitting skeletal models based on the full amino acid sequence to the electron density results. The 478 amino acid residues are folded into two large domains; the first, Domain A, is the largest, consisting of 380 residues from the N-terminus. Domain A contains a (β, α)₈ supersecondary structure. A large cleft is located at the carboxyl end of the parallel β -barrel in domain A. Domain B is an eight-stranded antiparallel β -sandwich (pleated sheet) structure. The two domains are linked by a single polypeptide chain with an extensive interface which is mainly composed of hydrophobic residues.

A 3-dimensional model of α -amylase fitted with an amylose chain in the catalytic site of the enzyme showed a possible productive binding mode between substrate and enzyme. Glutamic acid (Glu 230) and aspartic acid (Asp 297), which are located at the bottom of the cleft in domain A,

were concluded to be the catalytic residues, serving as the general acid and base, respectively. The configuration and active site residues are strikingly similar to those of lysozyme (Weaver et al, 1985) and other glycosidases such as xylanase and cellulases (Paice et al, 1978; Yaguchi et al, 1983; Paice et al, 1984 ; MacKay et al, 1985 ; Weaver et al, 1985), suggesting a common mechanism of hydrolysis in spite of the lack of sequence homology between the classes of enzymes (MacKay et al, 1985).

Site specific mutagenesis and protein engineering similar to that now being applied to proteases (Craik et al, 1985; Thomas et al, 1985) would help to clarify the role of various amino acid residues in the formation of the substrate binding and hydrolytic sites.

c) Primary Sequence Homologies

Comparative amino acid sequence analysis involving TAA and α -amylases from animal, bacteria and plant (MacKay et al, 1985; Rogers, 1985) has indicated that roughly four regions of homology exist. The larger of the conserved regions are located in the N-terminal half of each of the enzymes compared. These regions span domain A and part of the substrate-binding cleft proposed by Matsuura et al (1984), including the postulated active site amino acids, His-122, Glu-230 and Asp-297. Further comparisons involving eleven different amylases (Ihara et al, 1985; Nakajima et al, 1986) confirmed the previously described observations. The 6 animal amylases tested were found to be highly homologous (about 80 to 90%, depending on the species compared). However, in the cases where the two tissue-specific α -amylases have been sequenced, the comparisons reveal a surprising degree of dissimilarity, suggesting that these genes have diverged since duplication, about as much as expected for individual organisms (MacKay et al, 1985). The homology of

the amylases from a wider range of organisms including plants and microbes, is no more than 10%. Even within the bacterial genus, Bacillus, there was only 60% identity (e.g., B.steafothermophilus and B.amyloliquifaciens). A partial, N-terminal sequence of an α -amylase from the yeast, Schwanniomyces alluvius has been obtained and compared with that of the TAA and the homology has been found to be quite high. Since no other fungal or yeast α -amylases have been sequenced, it is not clear how much diversity exists at this taxonomic level. Little or no homology exists between the cellulases from fungi and/or bacteria (Mackay et al. 1985; Mackay et al., 1986). However, a region of homology has been found between at least one region of β -glucosidases of the basidiomyceteous fungus Schizophyllum commune and the yeast Candida pelliculosa (Moranelli et al. 1986) as well as with xylanases from S.commune and the yeast Cryptococcus albidans (Morosoli et al., 1986).

d) Comparison of TAA with other amylolytic enzymes

At present, the only enzymes that have been extensively characterized are from the glucoamylase class. The predicted amino acid sequences derived from the genes coding for glucoamylase of two ascomycetous fungi A.awamori (Nunberg et al., 1984) and A. niger (Boel et al., 1984) and the yeast, Saccharomyces diastaticus (Yamashita et al., 1984) and TAA have been compared and shown to have very little shared homology (Mackay et al., 1985). However, they reveal a single region of homology not present in other eukaryotic α -amylases which surrounds the Tyr-79 and Trp-83 residues of TAA located in the substrate-binding cleft. In A.niger glucoamylase (Clarke and Svensson, 1984), the six amino acid peptide, of which five are identical to the TAA peptide (Ala-78 to Trp-83), containing the Trp-120 (equivalent to Trp-83 of TAA) is essential for enzymatic function.

This peptide is also located at the centre of a highly conserved region shared by the glucoamylase of S.diastaticus.

The genetic relationship between α -amylase and glucoamylase is uncertain. MacKay and co-workers (1985) have suggested that the common pentapeptide Ala-78 / Trp-83 of TAA is peculiar to fungal amylase lineage, and may have arisen by a gene duplication event after divergence from other eukaryotes which generated the amylase and glucoamylase families. To pursue this matter further work should be carried out at the gene level.

c) Expression and Regulation of TAA

Like other amylolytic enzymes the production of TAA is influenced by the substrate as well as the end-product glucose (Meyrath et al, 1971; Angelova et al, 1980; Erratt et al, 1984). In comparison to the amount of biochemical characterization done on α -amylase and glucoamylase, studies on the mechanisms of induction and repression of these amylolytic enzymes have not progressed very rapidly. The mechanism of induction of the glucoamylase of A.awamori (Nunberg et al, 1984) is primarily at the transcription level. The mature 2.3 kb transcript codes for a preprotein which is processed by proteolytic cleavage and glycosylation during synthesis and secretion. Until recently (Erratt et al, 1984) very little was known about the molecular events surrounding expression of TAA in A.oryzae. In the presence of starch, the intracellular and extracellular TAA activity is induced by a factor of about 6.5 and 20 over that produced in the presence of glucose as the sole carbon source. The molecular weights of the active intracellular and extracellular enzymes appear to be the same, about 52.5kDa. However, the nascently synthesized α -amylase is likely inactive and slightly larger in size, since anti- α -amylase antibody, used in immunoprecipitation studies of in vitro

translated proteins directed by polyA⁺ RNA isolated from growing mycelia, revealed a product of about 54.0 kDa. These same studies indicated that the large increase in TAA activity in starch cultures could be accounted for by increasing the level or rate of α -amylase transcription. However, at this time it is not known whether other processing events are involved such as differential turn over of mRNA and protein glycosylation. Studies involving the cloned gene of TAA would provide further insight into the mechanisms governing this expression.

E) Research objective

From both fundamental as well as applied points of view, cloning of the α -amylase, TAA of A.oryzae would be desirable. Little is known about the regulation and expression of amylases even at the level of bacteria. The TAA is structurally, the most well characterized of this class of enzyme and one of the most highly used commercially. Cloning and characterization of the TAA gene(s) should provide insight into the nascent protein primary structure and its mechanism of regulated expression. More importantly, in the long term, beyond the scope of this thesis, cloning of the TAA gene would allow for construction of novel amylolytic, conventional industrial organisms (Sillis et al, 1984; Stewart and Russell, 1982; Innis et al, 1984; Seligy et al, 1986). Following closely the global objectives of protein engineering (Craik et al, 1985), modifications to the gene and ultimately its product would allow for the critical evaluation of the role of specific amino acid side chains in relation to the binding and active sites and open the exciting possibilities for the altered function, optimization and adaptation of TAA to heterologous physical and chemical environments.

The increased complexity of eukaryotic gene organization and

expression mechanisms in comparison with prokaryotic counterparts such as Escherichia coli reduced the possibility of directly screening for a functional α -amylase gene in E.coli as has been done for bacterial amylases (for further discussions see MacKay et al, 1985). Since initial studies by (Erratt et al, 1984) indicated that the induction of A.oryzae α -amylase by starch was due mainly to an elevated transcription of the α -amylase gene, the approach taken here was to start by cloning a cDNA copy of the α -amylase mRNA which was produced in high abundance by mycelia grown for 8 to 12 hours on starch. This cDNA copy either complete or partial could then be used to acquire the complete genomic sequence for characterization.

III. MATERIALS AND METHODS

A. Growth of *Aspergillus oryzae*

The strain of *A.oryzae* used in this work was obtained from the NRCC stock collection (NRC 401013). Further stocks were made by streaking loopfuls of conidia onto slants containing 8 ml of malt media agar (3% malt extract, 0.5% yeast extract, 0.5% glucose and 1.5% agar) (Law and Timberlake, 1980). The slant was incubated at 35C for 2-3 days until conidia appeared. For liquid cultures, 5ml of sterile 1% Tween 80 solution were added to a slant with conidia to produce a spore suspension containing $4-4 \times 10^7$ spores/ml. The spore suspension was added to 200mls of liquid malt media in a baffled flask and incubated at 35C for about 72hr on a rotary shaker at 300rpm. The mycelia were collected by filtration, washed with sterile water and reinoculated at 0.2% (wet weight/volume) into a baffled flask containing 1% yeast extract, 0.5% bactopectone and either 2% soluble starch or 2% glucose as the sole carbon source. The incubation was done at 35C with agitation. At various times (see Results Section A) the mycelia were collected by filtration through a sintered glass filter, washed with sterile water and frozen by immersion into liquid nitrogen. All samples were kept frozen at -70C.

B. Amylase activity, purification and antibody production

At various times throughout the growth period 10ml samples were removed and extracellular-amylase activity was determined from the filtered culture media using the starch iodine assay described by Smith and Roe (1949) with minor modifications. Purified amylase from *A.oryzae* was provided by Dr. Frank Moranelli as described earlier (Erratt et

al. 1984). One unit of activity equals that amount of enzyme required to produce 1mg of maltose from starch/ 3 min at pH 6.9, 20C.

The α -amylase and polyclonal antibody purifications were carried out by Dr.F.Moránelli (NRCC) as outlined in previous work (Erratt, et al.,1984; Willick et al.,1984).

C. Isolation of RNA

Two methods were used to purify RNA. Procedures for the preparation of ribonuclease-free glassware and equipment were followed (Chirgwin, et al. 1979; Maniatis et al.,1982).

Method A involved the use of phenol and SDS to deproteinize the RNA and DNA of frozen mycelia disrupted in a Braun homogenizer (Erratt et al. 1984). High molecular weight RNA was further separated from DNA and carbohydrate contaminants by precipitation in a solution containing 5mM EDTA, pH 8.0, and 2M LiCl (Zantinge et al.,1979; Willick and Seligy, 1985). After collecting the RNA precipitate at 15,000xg, 10min, 4C, the RNA was dissolved in 50mM Tris-HCl (pH 7.6) and residual carbohydrate was removed by centrifugation (15,000xg, 2min, 4C). Supernatant containing the RNA was stored frozen at -70C until further use.

Method B involved the use of guanidine isothiocyanate as a protein denaturant (Chirgwin et al. 1979). Powdered mycelial were prepared (Yelton et al. 1984) and stored at -70C. RNA was extracted by mixing the powdered mycelia with 4 volumes of guanidine isothiocyanate buffer (4M guanidine thiocyanate, 2mM sodium citrate pH 7.0, 1mMEDTA and 1% 2-mercaptoethanol), 0.5% antifoam and 0.5% sarkosyl. It was vortexed twice for 1 min before centrifugation at 15,000xg, 15 min, 4C. The clear supernatant was layered on top of a CsCl cushion containing 2ml of 5.7M CsCl and 100mM EDTA pH 7.0 (Glisin et al. 1974) and centrifuged in a

Beckman SW41.T1 rotor at 30,000 rpm for 18h at 18C. RNA pellet was resuspended in sterile water and precipitated in 2.5M ammonium acetate and two volumes of ethanol. The total RNA was kept dissolved in H₂O at -70C.

The PolyA⁺RNA fraction was obtained using the method of Aviv and Leder (1972) with modifications adapted for filamentous fungi (Erratt *et al.* 1984; Willick and Seligy, 1985). An aliquot of frozen RNA (200 µg) was heat denatured at 70C for 1 min and adjusted to 10mM Tris-HCl, pH 7.5 and 0.5M LiCl before rapidly chilling on ice and passage through a 3 ml oligo-dT cellulose column (BRL) equilibrated with the same buffer. RNA elution was monitored using a fraction collector with a UV-absorbance detector and chart recorder, calibrated to one O.D. (254nm) = 40 µg RNA. PolyA⁺RNA was eluted with 10mM Tris-HCl pH 7.5 and 0.5% SDS. The SDS was removed by precipitation with KAcetate (0.3M final) on ice and centrifugation (2,500xg, 10min). RNA was precipitated by addition of 2.5 volumes of ethanol and 0.1 volumes of 4M LiCl. RNA was recovered by centrifugation (15,000xg, 30min, 4C) and washed once in 66% ethanol before drying under vacuum and redissolving in sterile distilled water at 1mg/ml for storage at -70C.

D. Analysis of Amylase Specific RNA Expression By DNA-RNA Hybridization

Total RNA was prepared from A.oryzae mycelia grown on starch or glucose medium (see Section II.A) for 6, 8, 9, 10, 12, 15, 24 and 48 hrs using method B (see Section II.C). Aliquots containing equal amounts of RNA (1 to 5 µg) were immobilized with heat (80C, 1hr under vacuum) on nitrocellulose (Schleicher and Schuell, Bactiflex 0.45 µm) or nylon membrane (Pall Biodyne) filters either by direct adsorption (Dot blot) or by transfer from agarose gels after electrophoretic fractionation (Northern blot).

For dot blotting, total RNA in 20XSSC was added to membranes using the "Hybri-Dot" vacuum manifold system according to the manufacturer's specifications (BRL) and procedures described by Thomas (1980). RNA samples were denatured by mixing with one volume of 4XSSC and heating at 60C for 10min. Samples were cooled rapidly on ice and applied directly to the filter (Amorese, 1986).

Electrophoretic fractionation of RNA according to size was carried out after glyoxal treatment (Gerard and Miller, 1986). Total RNA (3-5 μ g) was incubated in 7 μ l of 10% glyoxal, 73% DMSO and 17.6 mM NaPO₄ pH 7.0 at 50C for 1hr. Samples were cooled to 23C and 1 μ l of 10mM NaPO₄, 0.5% xylene cyanol, 0.5% bromophenol blue was added and electrophoresed on a horizontal, 1% agarose nondenaturing gel in 10mM NaPO₄ buffer. When the bromophenol blue marker passed the 9 cm mark or the bottom of the gel, the gel was stained with ethidium bromide (0.5 μ g/ml), photographed (Polaroid MP, Type 52 film with red filter) under UV-light (310 nm). Transfer of RNA was carried out for 16 hrs as described by Thomas (1980) using 20xSSC.

DNA-RNA hybridizations were carried out either at 65C in 1M NaCl and 1% SDS after prehybridization in the same solution and denatured salmon sperm DNA (200 μ g/ml) or at 45C according to the protocol of the manufacturer (Pall) using 50% formamide. The DNA probe used for the annealing studies was a BglII/SalI 1.9kbp fragment of α -amylase gene II. DNA (750ng) was nick translated (see conditions described in Section II.P) to a level of $1-2 \times 10^7$ cpm and hybridization was carried out for 48 hr. Filters were washed and prepared for autoradiography as described by Maniatis et al. (1982).

E. Cell Free Protein Synthesis

In vitro, cell-free, translations of various RNA fractions was carried out using commercially prepared rabbit reticulocyte system according to recommendations of the supplier (BRL, Bethesda MD). The translations (30 μ l final) contained: 25-50 μ Ci (1026 Ci/mMole) L-(35 S)-methionine (New England Nuclear Corp.), 5 μ g of RNA (10 μ l), 3 μ l reaction mixture, 3.2 μ l water, 1.3 μ l 2M Kacetate and 10 μ l rabbit reticulocyte. The reaction was carried out for 60 to 90 min at 30C, after which 9 μ l of RNase A (10mg/ml) were added and incubated further for 20 min at 23C. Before further analysis 5 μ l of 50mM L-methionine was added and samples were processed directly or stored at -70C before further use.

For immunoprecipitation studies, the translation products were first mixed with an equal volume of Buffer T (50mM Tris-HCl pH 7.0, 150mM NaCl, 5mM EDTA, 0.005% Triton X-100) and then clarified by microfuge (13,000 rpm, 1 min). Supernatants were divided into two equal aliquots; one aliquot was mixed with 10 μ l of anti-~~amylase~~ antibody and the other with water. Both samples were incubated for 2hrs at 4C by rocking (Labquaker) after which 100 μ l of a 10% (v/v) solution of protein A-Sepharose 4B (Pharmacia) in Buffer T were added and further mixed at 23C for 30 min. The complexes were collected by centrifugation (8,000 rpm, 10min) and the pellets washed thrice with one ml Buffer T by vortexing. The final pellet was suspended in Buffer T without Triton X-100, collected (8,000 rpm, 2min) and dissolved in 50 μ l of 2X polyacrylamide gel (PAG), Laemmli (1970) loading buffer (125mM Tris-HCl pH 6.8, 0.4% SDS, 10% glycerol, 20% 2-mercaptoethanol, 6M urea), boiled for 5min and clarified (13,000 rpm/5min) before subjecting to electrophoresis in a 10% PAG. Electrophoresis was carried out overnight at 45 volts using 0.25M Tris-HCl

pH 8.3, 1% SDS and 1.92 M glycine as the running buffer (Laemmli, 1970).

For fluorography, proteins within gels were fixed by agitating for 60 min in 10% TCA, 10% glacial acetic acid and 30% methanol. Gels were then soaked in fluor (En³Hance, New England Nuclear) for 30 to 60 min, followed by washes for 30 min in cold water to precipitate the fluor and drying under vacuum and heat (60-70°C) for 2 hrs on Whatmann 3MM paper as backing. The dried gels were wrapped in plastic wrap and exposed to Kodak X-Omat AR-5 X-Ray film for various times at -70°C.

F. Synthesis of cDNA

cDNA synthesis from RNA was carried out as described by Gubler and Hoffman (1983) with modifications. 10 µg of polyA⁺RNA from 10 hr mycelia grown on starch was heat denatured in 10mM CH₃ HgOH for 10 min at 23°C followed by inactivation of mercury with 2 µl of 700mM 2-mercaptoethanol (Maniatis *et al.* 1982) and additions of: 10ng oligo-dT, cDNA buffer (100mM Tris-HCl 8.2, 10mM MgCl₂ and 250 mM KCl, final), 1mM of each of dATP, dCTP, dGTP and dTTP, 11 µCi (α³²P) dATP (~ 3000 Ci/mMol) and 50 units of AMV reverse transcriptase (Life Sciences Inc.) in a final volume of 50 µl. Synthesis of first strand cDNA was carried out at 37°C for 30 min and monitored by Cerenkov counting of TCA-precipitable radioactivity (Berger, 1984).

For the second strand synthesis, the reaction was simply adjusted to 100 µl by addition of one volume of nick translation buffer (50mM Tris-HCl pH 7.5, 5mM MgCl₂, 20mM KCl and 5mM DTT), 1-unit RNase H (BRL) and 25 units of *E. coli* Pol I (BRL) as recently described by Rutledge, 1987. The use of *E. coli* ligase and a two second step reaction (Gubler and Hoffman, 1983) were omitted. After incubation at 15°C for 2 hours, the double stranded cDNA was extracted twice with phenol-chloroform-isoamylalcohol (50:48:2),

passed through a Sephadex G-50 "spin column" (Maniatis et al., 1982) to remove small incomplete reaction products, and recovered by precipitation with 2.5 volumes of ethanol. The final product was dissolved in water (10ng/ μ l) and analyzed by TCA precipitation and counting in a liquid scintillation counter and by autoradiography after electrophoresis in a 1.4% agarose gel (see Section II.J).

G. Construction of cDNA Banks

Cloning of flush-ended double stranded cDNA involved the use of RF form M-13 mp9 vector cut at the single HincII site (Messing, 1983). Ten μ g of vector was digested with 50 units of HincII (BRL) in 50 μ l of 10mM Tris-HCl 7.5, 50mM NaCl, 10mM MgCl₂ and 1mM DTT at 37C for 2 hrs. The blunt ends were dephosphorylated with calf intestinal alkaline phosphatase (CIP) for 1 hr and the vector was purified by addition of SDS (0.5% final) and heat (65C, 10 min) followed by thrice extraction with phenol-chloroform and precipitation with 2 volumes of ethanol in 0.1M NaCl. The vector was dissolved in 10mM Tris-HCl pH 7.5 - 1mM EDTA to 50ng/ μ l. For ligation, 2 units of T4 ligase (BRL) was mixed with 100ng vector and 20ng cDNA in a 20 μ l final volume containing 7mM Tris-HCl pH 7.5, 10 mM DTT, 7mM MgCl₂ and 1 μ M ATP at 4C for 16hrs.

E.coli strain JM103 (Messing, 1983) was used for transfections as well as infections of recombinant M13 vector derivatives mp8 and mp9, while strain JM109 was used for vectors mpl8 and mpl9. E.coli strain HB101 and RR1 were also used to transform chimeras of plasmids pUC8 and pUC9. Competent cells were prepared by the method of Hanahan (1983) with modifications. Cells (4 μ l) from an M9 glucose minimal plate (Miller 1972) were grown over night in a 50 ml erlenmeyer flask containing 10 ml of 2YT (1.6% Bactotryptone; 1.0% Bacto yeast extract and 0.5% NaCl) at 37C with

rotary mixing (300 rpm), diluted in 2YT with 20mM MgSO_4 (25 μl / 25 ml) for further growth to a density of 0.6 to 0.7 O.D. 660nm and collected by centrifugation (8,000 x g 5min, 4C). Cell pellets were suspended (~3X concentration) in TFB (10mM K-Mes. pH 6.2, 100mM RbCl , 45 mM $\text{MnCl}_2 \cdot 4\text{H}_2\text{O}$, 10mM $\text{CaCl}_2 \cdot \text{H}_2\text{O}$, 3mM HA CoCl_3) by gentle vortexing and placed on ice for 10-15 min before pelleting and resuspension in TFB at 12.5 X the original volume of cells. Fresh DMSO was added to 3.5%, cells were swirled and left on ice for 10 min. DTT was added to 75mM (7 μl of 2.25M stock/200 μl of cells), mixed by swirling and left on ice for 10min before addition of an equal volume of DMSO and incubation at 4C for 5 min. Approx. 200 μl of treated cells were mixed with 4 μl (24 ng) of DNA ligation mixture in a 15 ml polypropylene tube (Falcon 2059) and incubated on ice for 30 min before shifting without agitation to 42C for 90 sec and back to ice for 1-2 min.

For M13 plaque production, the competent cells were mixed with 200 μl of exponentially growing JM103 cells, 10 μl of 100mM IPTG, 50 μl of Xgal (in 2% dimethylformamide) and 3 ml of top agarose containing 0.6% (w/v) agarose in B broth (1% bactotryptone, 0.8% NaCl, 0.001% Vitamin B1). The mixture was vortexed for 2 sec and plated on B broth plates containing 1.5% (w/v) agar. Plates were incubated at 37C overnight. Transformations with plasmid pUC derivatives was carried out as above without cell embedding in top agar or addition of exponential cells and with ampicillin (30 $\mu\text{g}/\text{ml}$) in the bottom agar for selection. (Also see section II.N.a).

H. Construction of Fungal DNA/Genomic Libraries in Lambda

A collection of approx. 219,000 individual λ Charon 4A phage containing A.oryzae DNA of average size of 15.6 kbp was made by V.L.Seligy and M.J.Dove based on previously published methods. Fungal DNA was

prepared from crude chromatin essentially as described by Seligy et al. (1980), see also Section II.I.a The EK2 vector λ Charon 4A was chosen because of its capacity for carrying large DNA fragments (8 to 22 kbp). Charon 4A "arms" which contain all the genes essential for phage propagation were prepared from pre-ligated phage DNA digested to completion with EcoRI to release the Bio and Lac containing "stuffer" fragments (Slightom et al. 1980), using 5 ml, 5 to 20% NaCl linear gradients (Beckman SW-50.1, 45,000 rpm, 2.8 hrs, 20C). Approximately 10 μ g of "arms" were ligated to an equivalent concentration of fungal DNA (12 to 20 kbp range, see Section II.I.a) in 50 μ l ligation buffer containing 1 unit T4 DNA ligase incubated at 4C for 18 hrs. The λ Ch4A-fungal DNA recombinants were packaged in vitro following the method of Hohn (1979) and plated and propagated on mini plates (9cm x 9cm) or megaplates (27cm x 90 cm enamel trays) with modifications (Davis et al., 1980; Slightom et al. 1980) using E.coli strain K802 as host.

I. Isolation and Purification of DNA from Fungi and Bacteria

(a) Preparation of Aspergillus DNA

DNA from the fungi, A.niger and A.oryzae was initially prepared as by a method used for yeast DNA and plasmids which basically selected for nuclear DNA (Seligy et al., 1980). However, most of this work used the method developed by Yelton et al. (1984) for A.nidulans. Powdered mycelia (~ 2g, see Section II.A) was rapidly suspended in 10 ml of 50mM NaEDTA pH 8.5 ,0.2% SDS with 10 μ l DEP with shaking for 1 min at 23C before heating at 68C for 15 min followed by cooling to 23C and centrifugation for 10 min at 15,000 x g. The supernatant (~ 8 mls) was transferred to a fresh tube, mixed with 1 ml 8.0M K acetate (pH 4.2) and chilled on ice for 1hr before clarifying at 12,000xg for 30 min at 4C. The supernatant was

carefully removed and extracted once with an equal volume of phenol-chloroform (1:1) followed by extraction with chloroform, removing each time the aqueous phase. Nucleic acid was precipitated from the final aqueous phase by addition of one volume of 2-propanol overnight before collection by centrifugation (8,000xg 10 min, 4C). The precipitate was dissolved in TE and DNAase-free RNAse A (treated at 90C for 10 min) added at 10µg/ml for 1 hr at 23C before a double precipitation step with 2-propanol. The final pellet was dissolved in 300 µl of TE or approximately 1mg/ml and stored at 4C.

(b) Isolation of single stranded M13 DNA

Single stranded ("+" DNA) M13 DNA (also referred to as template) was purified as described by Messing (1983). Single plaques were transferred and incubated for 8 hrs in a 50 ml erlenmeyer containing 10ml of 2YT, 50 to 100 µl of fresh, exponentially growing JM103 cells (see Section II.G). Cells were pelleted at 6,000xg for 10 min and then the phage progeny in the supernatant were removed and precipitated by a 1/10th volume of 20% PEG, 2.5M NaCl mixed at 4C for 15 min. The phage were collected (6,000xg, 15 min, 4C) and the supernatant was thoroughly removed before dissolving the pellet in 400 to 500 µl of STE (100M NaCl in TE). Phage DNA was initially extracted with phenol (v/v) by shaking for 5 min and recovered (microfuge, 13,000rpm, 4min) for further purification (see Fungal DNA). Final DNA was precipitated with two volumes of ethanol and redissolved in sterile water at 200ng/µl.

(c) Plasmid and M13(RF) DNA isolation

Plasmids and RF-form M13 phage DNA were prepared essentially by the rapid alkaline mini-plasmid procedure of Birnboim and Doley (1979) which involves an initial weakening of the bacterial cell wall by treatment with

lysozyme, SDS and NaOH. For extraction of total E.coli DNA without plasmid or phage, the alkaline addition step was omitted. M13 RF-form DNA was extracted from pellets of JM103 or JM109 cells grown for 8 hrs from an initial infection of 100 μ l of fresh cells (see Section G) with 50 μ l of plaque pure M13 stock (10^6 pfu/ml). Plasmid DNA was from various HB101, RRI or JM103 transformants grown (10 ml to 100 ml cultures) in the presence of ampicillin. Ultra-pure plasmid or RF-form DNA was obtained by further fractionation using cesium chloride ~~ultra~~-centrifugation with the intercalative dye ethidium bromide (Maniatis et al., 1982).

(d) DNA isolation from λ Charon 4A recombinant clones

DNA from plaque pure lambda phage chimeras propagated by the confluent plate method (Slightom et al., 1980) was extracted directly from PEG precipitated phage or from phage purified through a cesium chloride step gradient (Davis et al., 1979).

J. DNA Analysis by Restriction Endonucleases and Electrophoresis.

DNA from various sources were routinely subjected to limit digestion with different restriction endonucleases using either the buffers specified by the commercial supplier (BRL, New England Biologicals, Pharmacia) or the high, medium and low salt buffers described by Maniatis et al. (1982). Digests were routinely carried out for 1-2 hours at the recommended temperatures with enzyme units per DNA (μ g) ranging from 0.5 to 10 depending on whether the DNA was total genomic or cloned. Digests were inactivated by addition of 1/10th volume of stop buffer (100mM EDTA, 0.5% SDS, 25% glycerol, 0.1% BPB) by heating at 55C for 3-5 min. Samples were then stored frozen or at 4C or immediately used for electrophoresis.

Electrophoresis of DNA (up to 10 μ g/slot or track) was carried out using slabs of variable sizes of agarose (Seakem, FMC) as the supporting matrix (0.7 to 1.4% w/v) in either TAE (25mM Tris, 25mM acetic acid, 10mM EDTA, pH 7.5) or TBE (10mM Tris, 10mM Boric acid, 10mM EDTA, pH 8.0) (Maniatis et al. 1982) and with or without the dye ethidium bromide (Sigma Chemicals) at 0.1 to 0.5 μ g/ml. Electrophoresis was carried out in the horizontal position at approx 5 volts/cm for various times. Position of migration of DNA fragments within the gels was recorded by photographing using a camera equipped with a red filter and clear polaroid screen to detect the ethidium bromide fluorescence by ultra-violet light (310 nm).

> Molecular weight markers were : lambda DNA wild type digested with EcoRI, HindIII separately or together and pBR322 digested with TaqI and HinfI. In some cases these marker DNAs were radiolabelled by a 3' fill in reaction using PolI Krenow or T4 DNA polymerase and α -(³²P)-dATP (Maniatis et al. 1982). Labelled DNA was detected by autoradiography after fixing the agarose gels in 5% TCA, drying on Whatmann 3MM paper as a support with a slab dryer and exposing to X-Ray film (see Section II.E).

K. Initial Screening of Clone Banks For Amylase DNA

The M13-mp9.cDNA bank was initially screened for α -amylase recombinant clones by using the polyclonal anti-amylase antibody following the dot-immunobinding assay of Hawkes et al. (1982) that was successfully used to identify two cDNA clones coding for lysine-rich histone H5 of Xenopus laevis (Rutledge, 1987). However, since no clones were detected by this method this procedure was abandoned and a second approach using a synthetic oligonucleotide probe was used. For these studies, two oligonucleotides were designed by Dr.Seligy and synthesized in the laboratory of Dr. Saran Narang. As can be seen below,

Probe I

5' GA(Y)CG(N)TT(Y)GC(N)CG(Y)AC(N)GA(Y)GG(N)TC 3'

Probe II

5' CC(Z)CGCAT(Z)GG(Z)AC(Z)GG(Z)AC 3'

(N = A,C,G,T ; Y = T or C ; Z = A or C or G)

these oligomers are a mixture of several related sequences made to optimize complementation to putative sense codons of regions 17 to 25 of the N-terminal amino acid sequence and 450 to 456 of the C-terminal sequence respectively, derived from the sequence published by Toda et al. (1982).

Labelling of synthetic oligonucleotide probes was done with T4 polynucleotide kinase (Pharmacia) using the following conditions: 400ng ~~of~~ probe in 50µl buffer containing 50mM Tris.HCl pH 7.6, 10mM MgCl₂, 5mM spermidine (Sigma) , 0.1mM EDTA , 100 µCi γ-(³²P)-ATP (3000Ci/mM) Amersham) and 10 units of enzyme at 37C for 1 hr. Unincorporated isotope was removed by passage through a Sephadex G-25 spin column.

The procedure for oligonucleotide hybridization to E.coli clones was similar to that described by Woods (1984). Original M13-cDNA library plates containing 1000 to 2000 or so plaques were replica transferred to 80mm nitrocellulose filters (previously marked for clone location) by soaking for 1-2 min and then transferred to B agar plates (see Section II.G) cell-side up for 6-8 hrs incubation at 37C.

Filters were gently lifted from the plates and bacterial colonies including M13 phage were lysed using the alkali method described by Benton and Davis (1977). After neutralization with 1 M Tris-HCl pH 7.0 in 1.5M NaCl for 2-3 min, the filters were briefly immersed in 3X SSC (15-20sec) followed by adsorption of excess salt on Whatmann 3MM paper, air drying

and baking at 80C under vacuum for 2hr. Filters were washed in 3XSSC, 0.1% SDS at 65C for 16 to 20 hrs using larger volumes of buffer with several changes.

For prehybridization, filters were wetted in a 3XSSC filled plastic box or bag before exposure to 6XSSC, 1XDenhardt's, 0.5% SDS and 100 µg/ml sheared, heat denatured salmon sperm DNA for 2 hrs at 45C. Prehybridization buffer was drained away and ^{32}P -5' end labelled probe ($10^6/\text{ml}$) added, approximately 1-2ml of 6XSSC, 1XDenhardt's, 20 µg/ml yeast tRNA (Sigma Chemical), per filter. Incubation was at 45C for 16 to 20 hrs. Post-hybridization washing was done in several changes of 6XSSC with 0.5% sodium pyrophosphate (SSC-PP) using the following schedule: 1hr at 45C and 5 min at 50C with shaking. Filters were blotted dry and autoradiographed at -80C overnight.

1. Initial Isolation and Characterization of Amylase cDNA clones

Positively hybridizing plaques detected by the exposure on X-ray film were matched against the clones on the master plates. Each of the clones or plugs was picked up on the tip of a sterile pasteur pipette and placed into 1.5ml microfuge tubes containing 1 ml of 2YT broth. After mixing, the cells were pelleted ($13,000\times g$ 4 min) and a series of transfections at different phage dilutions was made and plated as before. The plates with the appropriate dilution 100 to 500 were screened by hybridization as before. Positive clones from this secondary screening were generally rescreened once more at a 10 to 100 plaque concentration. At this tertiary stage of screening all plaques should be positive and deemed therefore plaque pure. Results for the purification of α -amylase cDNA clone M13.mp9.M71 is shown in Figure 4 as an example.

Single strand phage DNA (template) from each of the plaque pure clones was verified for the presence of oligo-probe sequence and amylase DNA sequence by the Sanger dideoxy sequencing procedure as described by Messing (1983) (see Section Q.). Different templates were distinguished by employing the technique of "T tracking", which uses only the dideoxy TTP as the primer extension inhibitor (Messing, 1983). Determination of the cloned cDNA insert size (in kbp) was done using restriction endonuclease cleavage of double stranded cDNA produced by either the primer extension method (Chandler, 1982), which makes localized double stranded DNA or, by use of isolated RF-form DNA.

(a) Insert sizing

For primer extension, one μ g of template was annealed to 2 μ l (4 ng) of 17-mer universal primer (Pharmacia) in 12 μ l final of H-buffer (10mM Tris-HCl pH 7.5, 5mM $MgCl_2$) using a temperature gradient starting at 85C for 5 min and cooling to 23C, about 45 to 60 min. Extension reactions without the use of radioactive dATP, was started by adding 1 μ l each of 200mM DTT, 10mM dNTP (dATP, dCTP, dGTP, dTTP) and E.coli DNA polymerase I Klenow fragment (Pharmacia) (1.5 units) and incubated for 1hr at 37C. For radiolabelled primer extensions, 10 μ Ci of α - ^{32}P -dATP (3000Ci/mM) was added to the reaction. Unincorporated isotope was removed by passage of the mixture through a Sephadex G-50 "spin column" (Maniatis et al., 1982). All reactions were heat inactivated at 65C (10min) before adjusting the volume to 20 μ l with buffer necessary for digestion (1-2hr) with the appropriate restriction enzymes (BamHI, EcoRI, HindIII, PstI). Digests were terminated by use of stop buffer and fragment lengths were analyzed by electrophoresis (see Section II.J).

M. Screening X-4A.A.oryzae Genomic Clones Using Amylase cDNA Probe

Twenty plates each with approximately 2,000 pfu were prepared using a titer of the amplified recombinant library ($> 7 \times 10^8$ pfu/ml) prepared as described in Section II.H. Nitrocellulose filter lifts were prepared according to the Benton and Davis (1977) procedure using K802 cells (see also Section K). After baking, the filters were washed at 65C for 3 hrs or more in three changes of 6XSSC, 0.1% SDS and prehybridized for 90 min in 6XSSC, 0.1% SDS, 1X Denhardt's solution, 250 μ g/ml yeast tRNA and 250 μ g/ml denatured, sheared salmon sperm DNA (Sigma Chemical). For hybridization, about 10^6 cpm/ml of radiolabelled M13mp9.M71 cDNA insert was used in the same solution for 16 hrs of incubation. Filters were washed at 65C sequentially for 30 min each in 6XSSC, 0.1% SDS, 1X Denhardt's solution (three times) and 6XSSC, 0.1% SDS (twice), 15 min each in 2XSSC, 0.1% SDS (once) and 2XSSC (once). Filters were air dried mounted on card board backing and exposed to X-ray film at -70C for one day or more. Positively hybridizing plaques were located and purified to homogeneity as described in Section G. Cloned DNA was analyzed by restriction endonuclease digestion with EcoRI (2.5 units / μ g DNA), electrophoresis, transfer to nylon membrane (Taylor, 1985) and hybridization to the M71 probe as described above. Examples of this type of analysis is shown in Figure 12 for clones λ -4A.Ao.3C1 and λ -4A.Ao.2B1 which contain full length genes homologous to the probe located on EcoRI fragments 3.35 kbp and 7.6 kbp, respectively.

The labelling of single strand specific M13mp9.M17 hybridization probe was done essentially as described (see Section II.L.a. and Hu and Messing, 1982). The reaction contained: 1 μ g M71 template, 2 μ l sequencing primer (BRL), 1.1 μ l 10x H-buffer incubated at 70C for 10 min with dried

α -(32 P) dATP, dCTP and dTTP before addition of 1 μ l 0.2M DTT, 1 μ l of 100 μ M dATP, dCTP, dGTP and dTTP and 1 μ l (1.5 units) PolI Klenow enzyme. Synthesis proceeded for 1 hr at 37C before termination (see Section II.L.a.).

N. Subcloning of Amylase DNA sequences

The two EcoRI fragments 3.35 kbp and 7.6 kbp from λ -4A.Ao.3C1 and λ -4A.Ao.2B1 respectively as well as others from pUC plasmid cloning were separated by electrophoresis and recovered from the agarose by electroelution onto DEAE membranes (NA-45, Schleicher and Schuell) following the method of Dretzen *et al.* (1981). The membranes were cut to various sizes to match different agarose comb/slot widths. Prior to use, the membranes were soaked for several hrs in 2.5M NaCl before rinsing several times in water and storage in 1 mM EDTA pH 8.0 at 4C. DNA was electrophoresed (see Section J) using TAE buffer with ethidium bromide and located after by exposure to long wave-length UV to position the membranes. DNA was eluted off the membranes after further electrophoresis by soaking in 500 to 700 μ l of salt buffer (20mM Tris-HCl, pH 7.5, 1 mM EDTA, 1.5 M NaCl) for 1 hr at 68C. The eluted DNA was extracted successively with 3 volumes water saturated n-butanol to remove ethidium, and phenol-chloroform (1:1) to remove traces of agarose before recovery by precipitation in two volumes of ethanol. DNA precipitate was recovered by centrifugation (13,000xg 15 min) and dissolved in water at 10 to 20 ng/ μ l.

(a) Subcloning into pUC8

Ligation of fragments to pUC 8 vector (Viera and Messing, 1982), previously treated with CIP (section II.G), involved mixing of DNA (20 ng per ml final) at a vector to insert molar ratio of 1:2, in 10 μ l final

volume containing .1 units T4 DNA ligase, 5µg BSA, 25mM Tris-HCl pH 7.5, 10 mM MgCl₂, 10 mM DTT and 1µM ATP. Incubation was done at 12C for 6 or more hours.

For transformation, JM103 competent cells were used (see Section II.G). In this case, SOB (2% Bactotryptone, 0.5% yeast extract, 10mM NaCl, 2.5mM KCl, 10mM MgSO₄ and 10mM MgCl₂) was used for growing the cells instead of 2YT. (see Section II.G). After the heat pulse, 800ul of SOC (SOB and 2mM glucose) were added to the cells, and incubated at 37C for 1h. at 275rpm. 75µl of 100mM IPTG were added to the cells previously to plating. The cell suspension was divided into 100µl aliquots and plated on LAM (1% Bactotryptone, 0.5% yeast extract, 10mM NaCl, 10mM MgSO₄.7H₂O, 35µg/ml ampicillin and 1.5% Bacto agar) plates, previously streaked with 50µl of 2% X-gal, containing 100µl of SOB. The cells were streaked out onto a plate, allowed to dry for 20 min, and incubated upside-down at 37C overnight.

Identification of recombinant plasmid containing colonies was made using replica plating and colony hybridization according to Grunstein and Wallis (1979) with appropriate hybridization probe. Of the positively hybridizing colonies, generally 6 putative clones were picked at random and tested for presence of insert by the mini-plasmid procedure (see Section II.L). Clones containing the 7.6 and 3.35 kbp EcoRI fragments respectively in pUC8-76 and pUC-33 were further scaled up to one litre growths for bulk isolation of plasmid by cesium chloride density centrifugation (Maniatis et al., 1982).

(b) Subcloning into M13 mp18/19.

Further subcloning for sequencing was done using M13 mp18/19 vectors, (Yanisch-Perron et al. 1985). Fragments and vectors were cut by

combinations of 2 different enzymes to render different termini used for forced cloning. (see next Section).

Ligation conditions of vector to insert ratio of 1pmol ends/2pmol ends were chosen. Ligation mixture contained 20µg/ml of DNA, BSA (5mg/ml) and 0.1 units of T4 DNA ligase in 10µl of ligation buffer (25mM Tris-HCl pH 7.5 10mM MgCl₂, 10mM DTT and 4mM ATP). Ligation was done at 4-10°C for 4-14h followed by incubation for 2h at 23°C.

Preparation of E.coli K12 (JM109) (Yanisch-Perron et al. 1985) competent cells and transformation were done as described in Section II.G. Confirmation of positive recombinant clones (white plaques) was done using replica plating on nylon membranes and hybridization against the original fragment. Recombinant clones were picked for preparation of single stranded template DNA (Section II.I.B.).

0. Establishing 5'-3' Orientation of Cloned Inserts

The template strand of M13mp9M71 was shown to be the same as for the α-amylase mRNA by nucleotide sequencing (see Figure 7). This fact and the knowledge that a single BglIII site, located five codons upstream from the TGA stop codon, permitted the construction of 3' and 5' oriented DNA hybridization probes for mapping cloned genomic sequences.

The procedure of "force cloning" (Messing, 1983) which involves the use of reversed oriented, homologous restriction sites on M13 vectors mp8 and mp9, and mpl8 and mpl9, respectively, provided a means to obtain strand specific templates for hybridization and/or nucleotide sequencing. The cDNA insert of the RF-DNA of M13-mp9-M71, for example, was reverse oriented by cleaving the former with BamHI and HindIII, located at the two extremities, purifying by electroelution and ligating with RF-DNA M13-mp8 digested with the same enzymes. In a similar manner, two recombinant

phage DNAs can be tested for complementarity by use of the "C-test" (Messing, 1983). This procedure can be used to also excise these sequences by use of either restriction enzymes or single strand nucleases to eliminate the M13 DNA (Seligy *et al.*, 1986). Template DNA (250ng) from M13mp9M71 was annealed to 250 ng of recombinant M13mp8M71 (A to F) in 50 μ l final of 0.3M NaCl and buffer C (60% formamide, 2.5mM EDTA pH 8.0, 0.3% SDS and 0.1% bromophenol blue) at 65C for 1 hr. The results were analyzed by agarose gel electrophoresis using individual DNAs as controls (see Section J).

P. Rescreening The M13mp9.cDNA Bank With Genomic Amylase Sequence

The three independently constructed M13mp9.cDNA banks (see Section II.G.) were pooled by scraping the top agar from each plate into 15 ml Falcon tubes (Gibco) containing 2-4 ml of 2YT. The tubes were vortexed several times at 23C to release phage and cells from the soft top agar. The suspension was cleared of cells and agar by centrifugation (6,000 xg, 10 min, 4C) before removing the supernatant and recentrifuged. The final phage stock was plated out to approximately 2000 pfu/ plate and subjected to the DNA-hybridization screening procedure (Sections II.G. and II.K.) using the 3.35 kbp EcoRI genomic fragment radiolabelled with 32 P by nick translation (Maniatis *et al.*, 1982). The reaction mixture contained: 250 μ g DNA, 50 μ Ci each of α -(32 P)-dCTP and dTTP, 100 pM of dCTP and dTTP and 1nM each of dATP and dGTP, 5 μ l 10X NT buffer (50mM Tris-HCl pH 7.5, 10mM MgCl₂) and 4 units *E.coli* DNA polymerase (Pharmacia), 100pg DNAase I (BRL) at 16C for 1 hr. Unincorporated isotope was removed by "spin column" and DNA was denatured by adjusting to 0.3N NaOH followed by neutralization prior to addition to the hybridization buffer. Specific activity was based

on Cerenkoff radiation counted in a liquid scintillation counter (Beckman LS 250).

Three different classes of positive hybridization intensities (signals) were observed, weak, medium and strong. A total of 40 clones, 20 corresponding to strong signals (the C-series) and 20 giving medium signals (T-series) were rescreened twice more. A total of 12 plaque pure clones were selected for template DNA preparation and further analysis.

Q. DNA sequencing

DNA sequencing was carried out following the Sanger dideoxy nucleotide sequencing procedure (Sanger et al., 1977), involving single strand phage M13 DNA (template), the 17-mer universal primer with Pol I Klenow enzyme (Messing 1983), and [^{35}S]dATP α S (Amersham) and gradient polyacrylamide gels to detect primer extension products (Biggin et al. 1983). The strategies for sequencing the cDNA and genomic DNA are shown in Figures 7 and 14. Electrophoresis involved the use of either S1 or S0 equipment manufactured by BRL. After electrophoresis was completed, the front glass plate was removed and Whatmann 3M paper was used to back the gels for support during drying in a slab gel dryer (BioRad 483) at 80°C and autoradiography at -70°C. An overnight exposure was generally sufficient to read most X-ray films.

III. RESULTS

A. Cloning of Amylase Gene Sequences

(a) mRNA and cDNA Production

The work of Erratt et al. (1984) indicated that polyA⁺ RNA isolated from A.oryzae mycelia after 10 to 12 hours of growth on starch could provide an enriched source of specific mRNA for the synthesis of amylase cDNA. As shown in Figure 2 in the present work, production of total RNA occurred at approximately the same time, coinciding with the bulk of mycelial growth. The yield of polyA⁺RNA obtained by oligo-dT cellulose column chromatography was approximately the same from each RNA sample, about 2%. This RNA was qualitatively analyzed for size distribution by agarose gel electrophoresis and template activity by use of a rabbit reticulocyte in vitro translation system. The relative abundance of amylase mRNA was further checked by immunoprecipitation of the 54-56 kDa amylase-polypeptide from the cell-free translation products using the polyclonal anti-amylase antibody and complexed with protein A-Sepharose. These results, which are also summarized in Figure 2, were about the same as reported earlier by Erratt et al. (1984).

Based on the preceding analyses, the 10 hr polyA⁺RNA was chosen for use as the template for synthesis of double stranded cDNA using avian melanoblast virus (AMV) reverse transcriptase synthesis. The method for this synthesis was basically that described by Gubler and Hoffman (1983) but without the use of a separate step for second strand synthesis and the use of E.coli ligase as recently reported (Gillen et al., 1986; Rutledge, 1987). From approximately 10 ug of template, 494 ng of blunt-ended, double-stranded cDNA was recovered. Analysis of the size of this material by agarose gel electrophoresis in Figure 3, indicated that a fairly

Figure 2. Production of RNA and amylase-specific mRNA during growth of *A.oryzae* mycelia on starch containing medium.

Details for the growth (□—□), isolation and determination of total RNA (●—●) and estimation of α-amylase mRNA (▲—▲) are given in II.A, C and E, respectively. The relative amount of amylase mRNA was determined from the α-amylase specific immunoprecipitation of the *in vitro* translation products. Concentration of translatable α-amylase mRNA was made relative to the amount of α-amylase synthesized by 10h polyA⁺ RNA.

broad range of sizes were made, the bulk of which was about 1000 bp in size. This material and that of two other preparations were used to form the stock for the construction of double stranded cDNA libraries.

(b) Construction and screening of cDNA clones

The vector used for the cloning of the cDNA was the mp9 derivative of the phage M13 developed by Messing (1983). These phage vectors, which do not lyse their host cells, allow packaging of extraneous DNA up to 50% or more of their unit length. The vectors contain a MCS (multiple cloning site) with sites for several restriction enzymes, including the *HincII* that was used here for the blunt end ligation of cDNA into the *lac Z* gene. The double stranded, RF-form of M13mp9 is required for in vitro recombinant DNA production. The mp9 vector has recently been successfully used in the same manner to clone cDNA specific for mRNA coding for the Xenopus laevis H5 (Rutledge, 1987) and the rat oncomodulin (Gillen et al. 1986).

The E.coli JM103 host used for transfection of mp9 DNA (phage or RF form) and infection with phage is K12 restriction minus, but modification plus. The lac-pro region of the chromosome has been deleted. Complementation of the proline deficiency is facilitated by the F episome which is required to propagate the male specific M13 phage. The non-infected host is lac⁻ and produces a defective β -galactosidase. The polypeptide products of the lac region of the episome and the phage form a complex which restores activity and infected cells become lac⁺. The latter provides the basis for the blue/white plaque assay of JM strains infected with M13mp phages that express functional or nonfunctional *lac* DNA sequence when induced with IPTG in the presence of the colourless substrate X-gal which produces a deep blue colour if cleaved by active

Figure 3. Estimate of double stranded cDNA synthesized from polyA⁺RNA of A.oryzae grown for 10 hours on starch medium.

Approximately 10 ng of ds cDNA synthesized by AMV reverse transcriptase (track b) was electrophoresed for 3 hours in 1.4% agarose along with TaqI + HinfI (track a) and TaqI (track c) digested pBR322 DNA and ³²P-end labelled to serve as markers (see II.D and E)

B-galactosidase.

A summary of the three cDNA banks constructed is given in Table II. The total number of recombinants, based on the production of white plaques by IPTG-Xgal assay was approximately 25,000.

Table II. Properties of M13mp9-cDNA recombinant phage banks made from double stranded cDNA synthesized off 10 hour polyA⁺RNA

Vector Stock	Bank Number	Transformants per ng vector	% Blue Plaques	Total Recombinants	Plate Numbers
V-1	I	2,100	25	8,500	1,2,3
NV-1	II	20,000	17	8,700	21,22,41,42,43,44,45
NV-3	III	13,000	14	8,000	31,32,33,34,35,36,37,38,39,40

As previously mentioned in the methods (II.K), considerable effort was made to initially screen the three cDNA banks for production of amylase antigen using the polyclonal anti-amylase antibody. Several clones were selected by this method but none proved positive after the second round of screening. Since the amino acid sequence of α -amylase was available (Toda et al, 1982), it was felt that identification of an amylase cDNA clone could be better achieved by use of oligonucleotide hybridization probes whose sequence was complementary to 7 or 8 codons of the mRNA.

As described in the Methods (II.K) two probes were designed and synthesized: probe I was directed to cDNA sequences corresponding to the N-terminus (positions 17 through 25) and probe II was directed to the

C-terminal region (positions 450 through 456). Probe II proved to be a very effective probe probably because even as a mixture of 15-different 20-mers, it has a high G+C content (> 55%) which is useful for carrying out stringent annealing conditions. Also, the location of the probe sequence would pick up even relatively short cDNA sequences initiated by oligo-dT primer binding to the polyA tail of the mRNA.

Approximately the entire 25,000 recombinant clones were screened using a total of 20 filter lifts. Replication of clones onto NCF and lysis of the bacteria was done as described in Methods (II, G and K). The conditions for hybridization with the probes were those established by Woods (1984). The T_m maximum = 74°C and minimum = 64°C of probe II was calculated using the formula:

$$T_m = 2 + (T+A)_n \times 2 + (G+C)_i \times 4$$

Since the recommended hybridizations should be carried out at 10 to 15°C, at least, below the T_m and the washes at temperatures up to 5°C below the T_m , the conditions used were 45°C for annealing and 50°C for washing. Based upon the actual data derived from sequencing the cDNA (CC-TGC-CAT-AGG-AAC-AGG-CAC, see also section III, A, c) the T_m was calculated to be 66°C.

The hybridizations were carried out using probe labelled by T4 polynucleotide kinase and γ -(^{32}P)-ATP (1.6×10^8 cpm/ μg). The hybridization resulted in filters with some non-specific binding of probe to bacterial debris as well as, genuine positive signals. Only those which gave distinct, strong signals were picked and replated for further purification as shown in Figure 4. Bank I was negative. Bank II and III gave 10 and 3 putative positives, respectively, representing about 0.07%

Figure 4. Plaque hybridization and purification of amylase cDNA clone M71 using probe II.

A. initial filter with approx. 2,000 plaques, arrow shows location of positive M71 (A) clone. B,C. plaque purifications by second (B) and third (C) rounds of hybridizations with probe II. Exposure of autoradiographs was for 24 hours at -80°C .

of the total clones screened. Plaques of JM103/M13mp9 alone were negative. In the final purification, only five clones proved to be positive. One plaque from each master plate was kept as stock for further manipulation.

(c) Analysis of cDNA clones by nucleotide sequencing

The five plaque pure clones were characterized with respect to size of the cDNA insert, complementarity and nucleotide sequence content. These experiments were carried out using purified RF-form from the host cell or the single stranded phage DNA obtained from the culture medium.

Direct gel electrophoresis allowed the relative comparison of recombinant phage DNA sizes with M13mp9 phage DNA. According to this procedure all the clones had higher mobilities than the phage control and M91 and M71 appeared to contain the largest inserts. Direct measurement of insert sizes was obtained by use of the primer extension method (see II, L) or isolation of RF-form DNA prior to digestion with EcoRI or BamHI upstream and HindIII or PstI downstream of the HincII site to release the insert. An example of M71 insert sizing by the primer extension method is given in Figure 5. The insert was estimated to be about 510 ± 20 bp which is relatively close to the actual size by sequence analysis (500 bp plus the 5 or so M13 bases on either side).

To obtain accurate sequence information, the cDNA inserts were subcloned in the opposite orientation using the vector M13mp8 which has the MCS in reverse orientation to that of mp9 (Messing 1983). For forced cloning, both the mp8 and the recombinant RF were cleaved with HindIII and BamHI. The BamHI/HindIII cDNA fragments were purified by agarose gel electrophoresis and binding to DEAE membrane (See II, N). Two plaques from each transfection plate were picked and the phage DNA (+ strand) was tested for presence of insert by annealing the DNA to the original mp9.

Figure 5. Sizing of cDNA of M13mp9-M71 by use of primer extension.

The M71 phage DNA template was primed with the 17-mer sequencing primer in the presence of α - ^{32}P -dATP to form a double stranded DNA product that was cut from the vector with various restriction nucleases. The digests were analyzed by agarose gel (1.5%) electrophoresis, dried and autoradiographed. Tracks a and m contain primer extended undigested M71 and d, g, j the HinfI pBR322 marker fragments. Tracks b, c, correspond to primer extended M71 cut with EcoRI , $\text{EcoRI}/\text{HindIII}$; e, f, HindIII , $\text{HindIII}/\text{BamHI}$; h, i, BamHI , BamHI/PstI ; k, l, PstI , PstI/EcoRI .

Figure 6. C-test of M71 cDNA insert cloned in opposite orientations. M13mp9-M71 phage DNA containing the 500 bases cDNA insert was annealed for 1h at 65C with putative M13mp8 clones containing the complementary cDNA sequence (see II,K for details). The samples were analyzed by gel electrophoresis in 0.8% agarose containing ethidium bromide. The photograph shows: tracks a to g, 500ng of mp9-M71, 250ng of mp9-M71 with mp8-M71A, B, C, D, E, F. Track h contains λ -HindIII DNA fragments for size markers.

derived clone. This C-test procedure is illustrated in Figure 6 for the M13mp8-M71 A to F derivatives.

Sequencing of the cDNA clones was carried out using the primer extension method, the universal 17-mer primer and dideoxy inhibitors (Sanger *et al* 1977 and Messing *et al*, 1981, see also II,Q).

The full nucleotide sequences of the coding strands of M71 and M91 are given in Figures 7.A and 8.A, respectively, along with the predicted amino acid sequence (Figure 7.B and 8.B) and its comparison with the amino acid sequence of taka-diastase amylase by Toda *et al*, 1982. As shown, the 500 base cDNA of M71 contains 114 codons which includes probe II sequence, terminating in the stop codon TGA, followed by 154 nucleotides of 3' untranslated region containing two additional stop codons, and terminating in a 9 base A-track. The insert of M91 is similar to M71 but smaller; it contains only 384 bp with the same 3' UTR sequence and 75 codons. Comparison with the published amino acid sequence reveals a high degree of sequence alignment from amino acid residues beginning at 365 and ending at 478. Some sequence differences are apparent as listed below.

- 1) An extra Trp is inserted between Asn and Pro at positions # 384 and 385. This difference was also predicted from the genomic sequencing (see Figures 7-G and 16).
- 2) The TYR and ILE originally reported at positions 386 and 387, respectively, are reversed. These differences were also predicted in the genomic sequence (see Figures 7.C and 16) and can not be explained by simple base transitions.
- 3) A Gly is found at # 426 instead of Ser. This can be accounted for by a simple base transition.
- 4) An Asp at position # 476 is missing, leaving three Serines in a row.

The M91 sequence also confirmed the latter three differences.

Figure 7. Nucleotide sequence characterization of the M71 cDNA insert.

A. Nucleotide sequence of the 500 base insert determined by dideoxy method of Sanger and coworkers, (1977). Long heavy line indicates position of nucleotides 260-279 corresponding to 20-mer oligonucleotide probe II used in original screening. Short line shows position of BglII restriction site used to create 5' and 3' probes. Dashed line arrows indicate direct repeats. Arrows indicate positions of inverted repeats.

***-signify position of stop codons.

B. Conversion of nucleotide sequence to amino acid sequence.

The coding region is 342 nucleotides long corresponding to 114 amino acids. The 3' UTR is 157 nucleotides, including the TGA stop codon, and poly (A) tail.

C. Sequence alignment with the Taka-amylase amino acid sequence of A. oryzae provided by Toda et al. (1982). Sequence alignment was carried out by the MATCH program provided by NRCC Biomathematics section.

A

M13-mp9-M71

10 20 30 40 50
CGCGAACGCA ATCCGGAAGT ATGCCATTAG CAAAGATACA GGATTCTGTGA

60 70 80 90 100
CCTACAAGAA CTGGCCCATC TACAAAGACG ACACAACGAT CGCCATGCCG

110 120 130 140 150
AAGGGCACAG ATGGGTGCGA GATCGTGAAT ATCTTGTCCA ACAAGGGTGC

160 170 180 190 200
TTCGGGTGAT TCGTATACCC TCTCCTTGAG TGGTGCGGGT TACACAGCCG

210 220 230 240 250
GCCAGCAATT GACGGAGGTC ATTGGCTGCA CGACCGTGAC GGTTCGTTCC

260 270 280 290 300
GATGGAAATG TGCCTGTTCC TATGGCAGGT GGGCTACCTA GGGTATTGTA

310 320 330 340 350
TCCGACTGAG AAGTTGGCAG GTAGCAAGAT CTGTAGTAGC TCGTGAAGGG

360 370 380 390 400
TGGAGAGTAT ATGATGGTAC TGCTATTCAA TCTGGCATTG GACAGTGAGT

410 420 430 440 450
TTGAGTTTGA TGTAAGTTGT CTATTCTATG ATGTATGGTC TTTTGTCT

460 470 480 490 500
ATAGTTGGAA ATCGGAATGA TCTCAAATCT TGAATAAATA TAAAAAATAA

B

mp9-M71

														5															10															15
GCG	AAC	GCA	ATC	CGG	AAC	TAT	GCC	ATT	AGC	AAA	GAT	ACA	GGA	TTC	ALA	ASN	ALA	ILE	ARG	ASN	TYR	ALA	ILE	SER	LYS	ASP	THR	GLY	PHE															
														20															25															30
GTG	ACC	TAC	AAG	AAC	TGG	CCC	ATC	TAC	AAA	GAC	GAC	ACA	ACG	ATC	VAL	THR	TYR	LYS	ASN	TRP	PRO	ILE	TYR	LYS	ASP	ASP	THR	THR	ILE															
														35															40															45
GCC	ATG	CGC	AAG	GGC	ACA	GAT	GGG	TCG	CAG	ATC	GTG	ACT	ATC	TTG	ALA	met	ARG	LYS	GLY	THR	ASP	GLY	SER	GLN	ILE	VAL	THR	ILE	LEU															
														50															55															60
TCC	AAC	AAG	GGT	GCT	TCG	GGT	GAT	TCG	TAT	ACC	CTC	TCC	TTG	AGT	SER	ASN	LYS	GLY	ALA	SER	GLY	ASP	SER	TYR	THR	LEU	SER	LEU	SER															
														65															70															75
GGT	GCG	GGT	TAC	ACA	GCC	GGC	CAG	CAA	TTG	ACG	GAG	GTC	ATT	GGC	GLY	ALA	GLY	TYR	THR	ALA	GLY	GLN	GLN	LEU	THR	GLU	VAL	ILE	GLY															
														80															85															90
TGC	ACG	ACC	GTG	ACG	GTT	GGT	TCG	GAT	GGA	AAT	GTG	CCT	GTT	CCT	CYS	THR	THR	VAL	THR	VAL	GLY	SER	ASP	GLY	ASN	VAL	PRO	VAL	PRO															
														95															100															105
ATG	GCA	GGT	GGG	CTA	CCT	AGG	GTA	TTG	TAT	CCG	ACT	GAG	AAG	TTG	met	ALA	GLY	GLY	LEU	PRO	ARG	VAL	LEU	TYR	PRO	THR	GLU	LYS	LEU															
														110															115															120
GCA	GGT	AGC	AAG	ATC	TGT	AGT	AGC	TCG	TGA	AGG	GTG	GAG	AGT	ATA	ALA	GLY	SER	LYS	ILE	CYS	SER	SER	SER	***																				
														125															130															135
TGA	TGG	TAC	TGC	TAT	TCA	ATC	TGG	CAT	TGG	ACA	GTG	AGT	TTG	AGT																														
														140															145															150
TTG	ATG	TAA	CTT	GTC	TAT	TCT	ATG	ATG	TAT	GGT	CTT	TTT	GTT	CTA																														
														155															160															165
TAG	TTG	GAA	ATC	GGA	ATG	ATC	TCA	AAT	GTT	GAA	TAA	ATA	TAA	AAA																														
														170															175															180
														175															180															185
														180															185															190
														185															190															195
														190															195															200
														195															200															205
														200															205															210
														205															210															215
														210															215															220
														215															220															225
														220															225															230
														225															230															235
														230															235															240
														235															240															245
														240															245															250
														245															250															255
														250															255															260
														255															260															265
														260															265															270
														265															270															275
														270															275															280
														275															280															285

SEQUENCE ALIGNMENT

MP9-M71 cDNA

vs

A.oryzae Taka-amylase

(elements 1-115 of 166)

(elements 360-478 of 478)

```

----- ALA ASN ALA ILE ARG ASN TYR ALA ILE SER 10
      *** **
LYS LEU ILE ALA SER ALA ASN ALA ILE ARG ASN TYR ALA ILE SER 374

LYS ASP THR GLY PHE VAL THR TYR LYS ASN TRP PRO ILE TYR LYS 25
*** **
LYS ASP THR GLY PHE VAL THR TYR LYS ASN --- PRO TYR ILE LYS 388

ASP ASP THR THR ILE ALA MET ARG LYS GLY THR ASP GLY SER GLN 40
*** **
ASP ASP THR THR ILE ALA MET ARG LYS GLY THR ASP GLY SER GLN 403

ILE VAL THR ILE LEU SER ASN LYS GLY ALA SER GLY ASP SER TYR 55
*** **
ILE VAL THR ILE LEU SER ASN LYS GLY ALA SER GLY ASP SER TYR 418

THR LEU SER LEU SER GLY ALA GLY TYR THR ALA GLY GLN GLN LEU 70
*** **
THR LEU SER LEU SER GLY ALA SER TYR THR ALA GLY GLN GLN LEU 433

THR GLU VAL ILE GLY CYS THR THR VAL THR VAL GLY SER ASP GLY 85
*** **
THR GLU VAL ILE GLY CYS THR THR VAL THR VAL GLY SER ASP GLY 448

ASN VAL PRO VAL PRO MET ALA GLY GLY LEU PRO ARG VAL LEU TYR 100
*** **
ASN VAL PRO VAL PRO MET ALA GLY GLY LEU PRO ARG VAL LEU TYR 463

PRO THR GLU LYS LEU ALA GLY SER LYS ILE CYS SER --- SER SER 115
*** **
PRO THR GLU LYS LEU ALA GLY SER LYS ILE CYS SER ASP SER SER 478

```

Figure 8. Nucleotide sequence characterization of the M91 cDNA insert.

A. Nucleotide sequence of the 384 base insert.

Long heavy line indicates position of nucleotides 144-163 corresponding to 20-mer oligonucleotide probe II used in original screening. Short line shows position of BglII restriction site used to create 5' and 3' probes. Dashed line arrows indicate direct repeats. Arrows indicate positions of inverted repeats. ***- signify positions of stop codons.

B. Conversion of nucleotide sequence to amino acid sequence.

The coding region is 227 nucleotides long corresponding to 75 amino acids. The 3' UTR is 157 nucleotides, including the TGA stop codon and poly (A) tail.

C. Sequence alignment with the Taka-amylase amino acid sequence of A.oryzae provided by Toda et al. (1982).

A

M13-mp9-M91

10 20 30 40 50
 CGCAGATCGT GACTATCTTG TCCAACAAGG GTGCTTCGGG TGATTCGTAT
 1

60 70 80 90 100
ACCCTCTCCT TGAGTGGTGC GGGTTACACA GCCGCCCAGC AATTGACGGA
 2

110 120 130 140 150
 GGTCATTGGC TGCACGACCG TGACGGTTGG TTCGGATGGA AATGTGCCTG

160 170 180 190 200
TTCCTATGGC AGGTGGGCTA CCTAGGGTAT TGTATCCGAC TGAGAAGTTG
 3 2

210 220 230 240 250
GCAGGTAGCA AGATCTGTAG TAGCTCGTGA AGGGTGGAGA GTATATGATG
 3 Bgl II ***

260 270 280 290 300
 GTACTGCTAT TCAATCTGGC ATTGGACAGT GAGTTTGAGT TTGATGTAAC
 4 1

310 320 330 340 350
 TTGTCTATTG TATGATGTAT GGTCTTTTTG TTCTATAGTT GGAAATCGGA

360 370 380
 ATGATCTCAA ATCTTGAATA AATATAAAAA AAAA
 4

B

mp9-M91

	5	10	15
CAG ATC GTG ACT ATC TTG TCC AAC AAG GGT GCT TCG GGT GAT TCG			
GLN ILE VAL THR ILE LEU SER ASN LYS GLY ALA SER GLY ASP SER			
	20	25	30
TAT ACC CTC TCC TTG AGT GGT GCG GGT TAC ACA GCC GGC CAG CAA			
TYR THR LEU SER LEU SER GLY ALA GLY TYR THR ALA GLY GLN GLN			
	35	40	45
TTG ACC GAG GTC ATT GGC TCC ACG ACC GTG ACG GTT GGT TCG GAT			
LEU THR GLU VAL ILE GLY CYS THR THR VAL THR VAL GLY SER ASP			
	50	55	60
GGA AAT GTG CCT GTT CCT ATG GCA GGT GGG CTA CCT AGG GTA TTG			
GLY ASN VAL PRO VAL PRO met ALA GLY GLY LEU PRO ARG VAL LEU			
	65	70	75
TAT CCG ACT GAG AAG TTG GCA GGT AGC AAG ATC TGT AGT AGC TCG			
TYR PRO THR GLU LYS LEU ALA GLY SER LYS ILE CYS SER SER SER			
	80	85	90
TGA AGG GTG GAG AGT ATA TGA TGG TAC TGC TAT TCA ATC TGG CAT			

	95	100	105
TGG ACA GTG AGT TTG AGT TTG ATG TAA CTT GTC TAT TCT ATG ATG			
	110	115	120
TAT GGT CTT TTT GTT CTA TAG TTG GAA ATC GGA ATG ATC TCA AAT			
	125		
CTT GAA TAA ATA TAA AAA AAA A			

SEQUENCE ALIGNMENT

MP9-M91 cDNA vs A.oryzae Taka-amylase
 (elements 1-75 of 76) (elements 403-478 of 478).

```

GLN ILE VAL THR ILE LEU SER ASN LYS GLY ALA SER GLY ASP SER 15
*** *** *** *** *** *** *** *** *** *** *** *** ***
GLN ILE VAL THR ILE LEU SER ASN LYS GLY ALA SER GLY ASP SER 417

TYR THR LEU SER LEU SER GLY ALA GLY TYR THR ALA GLY GLN GLN 30
*** *** *** *** *** *** *** *** *** *** *** *** ***
TYR THR LEU SER LEU SER GLY ALA SER TYR THR ALA GLY GLN GLN 432

LEU THR GLU VAL ILE GLY CYS THR THR VAL THR VAL GLY SER ASP 45
*** *** *** *** *** *** *** *** *** *** *** *** ***
LEU THR GLU VAL ILE GLY CYS THR THR VAL THR VAL GLY SER ASP 447

GLY ASN VAL PRO VAL PRO MET ALA GLY GLY LEU PRO ARG VAL LEU 60
*** *** *** *** *** *** *** *** *** *** *** *** ***
GLY ASN VAL PRO VAL PRO MET ALA GLY GLY LEU PRO ARG VAL LEU 462

TYR PRO THR GLU LYS LEU ALA GLY SER LYS ILE CYS SER --- SER 74
*** *** *** *** *** *** *** *** *** *** *** *** ***
TYR PRO THR GLU LYS LEU ALA GLY SER LYS ILE CYS SER ASP SER 477

SER
***
SER

```

75
478

For further use of the cDNA sequence, REN mapping was done using the RMAP program of the computer. These results are summarized in Figure 9. The presence of a BglII site and the absence of BamHI, EcoRI, HindIII and Pst I sites were confirmed by actual experiments (see Figure 5).

(d) Identification of genomic sequences with cDNA probes

The sequence results unequivocally establish that the 3' portion of the mRNA coding for α -amylase mRNA sequence has been cloned. Because of the substantial length of the sequences in M71 and M91, they were used as specific amylase gene probes for the identification of the chromosomal genes. The presence of the single BglII site six codons upstream from the stop codon conveniently divides the cDNA into coding and non-coding regions (see Figure 9). This feature was used to establish 5' and 3' orientation of the genomic sequence as well as the 3' extremity of the gene.

The first step in the procedure to acquire the genomic copy of the amylase gene was to carry out a DNA-DNA hybridization analysis of purified A.oryzae total DNA with the M71 probe made radioactive by nick translation of the RF-DNA copy. Total DNA was digested to limit with the restriction enzymes EcoRI, BamHI and HindIII. The digests were electrophoresed in agarose and transferred to a membrane for further analysis by the Southern (1975) procedure. These results shown in Figure 10 and others (see Figure 15) indicated that two fragments are released by each enzyme. The sizes of these fragments are listed below:

(in Kilobase pairs)

EcoRI	BamHI	HindIII
7.6	8.6	8.7
3.35	7.2	6.6

A

SIZE OF SITE	ENZYME	# OF FRAGMENTS	SIZE OF FRAGMENTS				
			1	2	3	4	5
6	AccI	2	335	165			
4	AluI	2	340	160			
5	BbvI	2	287	213			
6	BglII	2	328	172			
6	BspI286	2	392	108			
6	CfrI	2	300	200			
5	DdeI	2	307	193			
4	DpnI	5	207	141	90	33	29
4	FnuclI	2	497	3			
5	Fnu4hI	2	273	227			
5	FokI	2	264	236			
6	FspI	2	401	99			
4	HaeIII	3	298	137	65		
4	HhaI	2	400	100			
5	HinfI	3	341	116	43		
4	HpaII	3	301	185	14		
5	HphI	2	332	168			
4	MnlI	3	292	180	28		
6	NaeI	2	300	200			
4	NlaIII	2	402	98			
6	PvuI	2	409	91			
4	RsaI	2	369	131			
6	StyI	2	288	212			
6	XhoII	2	328	172			

Figure 9. REN mapping of cDNA insert of M13mp-M71 clone.

A. Catalogue of 6, 5 and 4 base recognition restriction enzyme cutting sites giving the number, position and size in base pairs of the fragments.

B. Scale map showing positions.

mp9-M71 cDNA RESTRICTION MAP 27

	10	20	30	40	50	
1	CGCGAAGCA GCGCTTGGT **** FnuDI	ATCCGGAAGT TAGGCCTTGA **** HpaII	ATCCCATJAG TACGGTAATC	CAAAGATACA GTTTCTATGT	GGATTGCTGA CCTAAGCACT ***** HinfI	50
	60	70	80	90	100	
51	CGTACAAGAA GGATGTICTT	CTGGGCCAATC GACCGGCTAG **** HaeIII	TACAAAGACG ATGTTTCTTC	ACACAACGAT TGTGTTCCTA **** PvuI *** DpnI	CGGCATGXXC GCGGTACGCG ** **** FspI * **** NlaIII **** HhaI	100
	110	120	130	140	150	
101	AAGGGCAGAG TTCCCGTGTC *	ATGGGTGCGA TACCCAGCGT	GATCGTGACT CTAGCACTGA **** DpnI	AICTTGTGCA TACAACAGGT	ACAAGGGTGC TGTTCGACG	150
	***** BspI286					
	160	170	180	190	200	
151	TTCCGGTGAT AAGCCCACTA ***** HphI *** Hin	TGGTATACCC AGCATATGGG ***** AccI ** FI	TCTCTTTGAG AGAGCAACTC	TGGTGGCGGT ACCAAGGCGA	TACACAGCGG ATGTCTGCGC **** NaeI *** Hpu ** CI * H	200
	210	220	230	240	250	
201	GCCAGCAATT CGCTCGTTAA ** * II **** rI *** nelII	GACGGAGGTC CTGCCCTCCAG **** MnII	ATTCGGTGGCA TAACCGAGGT ***** BbvI ***** Fnu4hI	CGACCGTGAC GCTGGCACTG	GCTTGTTCG CCAAGCAAGC * F	250
	260	270	280	290	300	
251	GATGGAATG CTACCTTTAC **** okI	TGGCTGTTCG ACGGACAAGG	TATGGCAAGT ATACCGTCCA	GGCTACCTTA CCCGATGGAT **** StyI	GGGTATTGTA CCCATACAT **	300
	310	320	330	340	350	
301	TCGACTGAG AGGCTGACTC ***** DdeI	AAGTTGGCAG TTCAACCGTC	GTAGCAAGAT CATGGTTCTA **** BglI **** XhoI *** DpnI	GTGTAGTAGC GACATCATCG ** *** Alu I	TGTGAACGG AGCACTTCCG * I	350
	360	370	380	390	400	
351	TGGAGACTAT ACCTCTCATA	ATGATGTGAC TACTACCATG **** RsaI	TGCTATTCAA ACGATAAGTT	TCTGGCATTG AGACCGTAAC	GACAGTGAGT CTGTCACTCA	400
	410	420	430	440	450	
401	TTGAGTTTGA AACTCAAAC	TGTAACCTGT ACATTGAACA	CTATTCTATG GATAAGATAC	ATGTATGGTC TACATACCAG	TTTTTGTCT AAAAACAAGA	450
	460	470	480	490	500	
451	ATAGTTGGAA TATCAACCTT	ATCGGAATGA TAGCCTTACT ** DpnI	TCTCAAATCT AGAGTTTACA ** nI	TGAATAAATA ACTTATTAT	TAAAAAATA ATTTTTTTT	500

Figure 10. Mapping of 3' α -amylase genomic DNA homology using M71 cDNA.

A. Agarose (0.8%) gel electrophoresis of REN digests stained with ethidium bromide. Tracks a-c, EcoRI, BamHI and HindIII digested A.oryzae DNA. Track d contains λ -HindIII digested DNA as size markers.

B. 4 day autoradiograph showing positively hybridizing fragments detected by use of 5×10^7 cpm of nick translated RF-form M71 DNA.

Further probing of similar DNA filters with the two halves of the M71 cDNA insert did not distinguish between the two fragments released by each enzyme. These results established that the M71 is an effective probe that can be used for screening the genomic libraries and also indicated that more than one α -amylase gene sequence is present in the genome of A.oryzae.

(e) Cloning of genomic sequences

Approximately 40,000 clones, 2,000 plaque forming units per plate of the λ -Charon 4A-Ao genomic bank (see II, M for procedure) were screened using M71 cDNA insert radiolabelled by the primer extension method (Hu and Messing, 1982) as probe. From the primary screening, 16 strong positives were obtained, about 0.04% of the total. These positively hybridizing plaques were picked and subjected to two more rounds of hybridization to obtain plaque purity as shown in Figure 11 for clone 11G1. Only 12 of the picks were carried through to purity. One plaque from each plate was picked and kept in 500 μ l of plaque storage buffer, giving a final stock of 2×10^6 phage/ml.

An analysis of the inserts in the 12 plaque pure clones was carried out by the Southern method using a mini-scale up of phage DNA involving production of plaques confluent on lawns of K802 plates (Davis et al. 1981). These results are summarized in Figure 12 and Table III. Only one of the clones failed to hybridize with the M71 probe. Most of the clones contained either one or the other of the two classes of DNA fragments originally detected in the EcoRI genomic digests by the M71 probe (see Figures 10 and 12). Two clones (2B1, 11G1) containing the 7.6kbp fragment also carried a 4 kbp fragment that positively hybridized to the M71 probe. The origin of this fragment is uncertain. Another fragment

Table III. Status of Charon4A.Ao Amylase Genomic clones

Clone Number			Type of Insert
Primary	Secondary	Tertiary	Hybridized to M71
2A	0/50	-	
2B	19/90	2B1 +	I + 4 kbp *
3C	6/70	3C1 +	II *
4CH	30/120	4CH1 -	
4D	5/150	4D1 +	I *
5E	2/70	5E1 +	I
10F	5/180	10F1 +	I
11G	7/37	11G1 +	I + 4 kbp
11H	3/43	11H1 +	II
13I	4/140	13I1 +	II
17J	0/71	-	
17K	20/160	17K1 +	II *
18L	0/100	-	
19M	0/240	-	
19N	12/120	19N1 +	II
19P	18/100	19P1 +	II

I = 7.6 Kbp insert; II = 3.35 Kbp insert; control Charon 4A was -ve.
 * indicate clones containing a 5.8 Kbp insert.

Figure 11. Screening of λ -Charon 4A-Ao genomic library with M71 probe and purification of clone 11G1. A-C, autoradiograms of nitrocellulose membrane lifts from plates containing: A, about 2,000 plaques from original genomic bank; B and C represent secondary and tertiary rounds of screening. All signals on C correspond to 100% of the plaques. Arrows in A and C indicate plaques picked for starting/primary and final purified stocks.

Figure 12. EcoRI cleavage and hybridization pattern of phage DNA prepared from plaque pure recombinant phage containing putative amylase genes.

DNA was prepared by the mini-lysate and subjected to Southern analysis using 10^7 cpm of primer extended M71 probe. Autoradiogram of nylon membrane filter blotted with EcoRI digested phage DNA electrophoresed in 0.8% agarose (see II,M). Tracks, a and n contained HindIII digested lambda DNA used as marker. Tracks, b-m correspond to clones 2B1, 3C1, 4CH1, 4D1, 5E1, 10F1, 11G1, 11H1, 13I1, 17K1, 19N1 and 19P1.

the 5.8kbp found in four clones (Table III) was studied further (see III,B,a).

These cloning results indicate that the genomic amylase sequences have been successfully cloned. The 26-mer oligonucleotide probe directed to codons at the 5' end of the amylase coding region was also used to hybridize a filter similar to that prepared in Figure 12. These results further indicated that only the 7.6 and 3.35 kbp EcoRI fragments carried homologous sequence albeit some hybridization was also detected in the Charon 4A right arm (10.6kbp) (data not shown).

B. Comparison of Amylase Genes

(a) Subcloning of genomic amylase DNA in pUC8 and M13 phage vectors

To further characterize the amylase genomic sequences located on EcoRI fragments I (7.6 kbp) and II (3.35 kbp) it was necessary to excise these sequences from clones 11G1 and 3C1, respectively. The procedure involved electrophoretic fractionation and elution of the fragments onto DEAE paper. Vector pUC8 was cut at the single EcoRI site located in the MCS and dephosphorylated with calf thymus intestinal phosphatase before ligation and transformation using the host JM103. Recombinants were detected by hybridization of the colonies using the respective fragments radiolabelled by nick translation as probes. Detailed mapping of these subclones is presented in the ensuing sections (Table IV and V).

b) REN mapping and chromosomal linkage of Amylase I and II genes

A physical map of the common restriction sites, including the orientation of the 5' and 3' ends of the amylase genes, was carried out on the inserts of pUC76 and pUC33. These results are shown in Figures 13 and 14. The REN analysis involved the use of one or two enzymes during

Table IV. Summary of subclones containing EcoR1 fragment I.

(Refer to Figure 13.B for map location of each fragment)

Clone Nomenclature	Vector	Restriction Sites	Insert size(kbp)
pUC8-76	pUC8	EcoR1/EcoR1	7.6
5C	M13mp18	BamH1/EcoR1	6.0
6I		SmaI/EcoR1	1.66
* 7I		SmaI'/EcoR1'	0.65
* 8L		SalI'/SmaI	2.35
* 9L		SalI'/SmaI'	2.9
10B		Bgl II/SmaI	0.9
13R		PstI'/SmaI	0.5
14Q		PstI'/SalI	0.43
18Q		PstI/ SalI	0.35
19T		PstI/ EcoR1	0.42
5G	M13mp19	EcoR1/BamH1	6.0
6K		EcoR1/SmaI	1.66
* 7K		EcoR1'/SmaI'	0.65
8M		SmaI/SalI'	2.35
* 9M		SmaI'/SalI'	2.9
10FU		SmaI/BglIII	0.98
14U		SalI/PstI'	0.43
20Z		BglIII/SalI'	1.43
31U2		AvaI/BglIII	0.45
60U4		AvaI/SmaI	0.53

* not used in any experiments; /' denotes the second restriction site

Table V. Summary of subclones containing EcoRI fragment 11.

(Refer to Figure 14B for map location of each fragment)

Clone Nomenclature	Vector	Restriction Sites	Insert size(kbp)
pUC8-33	pUC8	EcoRI/EcoRI	3.35
pH-29	pUC8	EcoRI/HindIII	2.882*
pH-3A		EcoRI/HindIII	2.882*
pUC-II	pUC9	EcoRI/HindIII	2.882*
YRP-II	YRp(Adel)3	EcoRI/HindII blunt end	2.882*
1A	M13mp18	SalI/EcoRI	0.75
2B		BamHI/SmaI	1.55
3C		BamHI/EcoRI	2.82
4D		SalI/SmaI	0.913
11I		SmaI/EcoRI	1.665
12Q		PstI'/SalI'	0.425
12P		PstI'/SmaI	0.488
16Q		PstI/SalI	0.335
17T		PstI/EcoRI	0.417
21C		PstI/EcoRI'	0.530
27XD		AvaI/SmaI	0.528
1E	M13mp19	EcoRI/SalI	0.75
2F		SmaI/BamHI	1.155
3G		EcoRI/BamHI	2.820
4H		SmaI/SalI	0.913
11K		EcoRI/SmaI	1.665
28NB		AvaI/BglII	0.450

* not used in any experiments; ' denotes the second restriction site

Figure 13. REN analysis and map of pUC8-76 containing the 7.6 EcoRI fragment with amylase gene 1 sequence. A, 0.8% agarose gel analysis of REN digests of pUC8-76. Tracks a and o contain lambda DNA digested with EcoRI/HindIII and HindIII respectively. Tracks b-n contain plasmid digested with BamHI, BglII, HindIII, PstI, SalI, SmaI, EcoRI, and double digests EcoRI with BamHI, BglII, HindIII, PstI, SalI, and SmaI. B, map showing restriction sites and subclones (see Table IV).

A

B

Figure 14. REN analysis and map of pUC8-33 containing the 3.35 kbp EcoRI fragment with amylase gene II sequence. A, 0.8% agarose gel analysis of REN digests of pUC8-33. Tracks a and o contain lambda DNA digested with HindIII and HindIII/EcoRI, respectively. Tracks b-n contain plasmid digested with BamHI, BglII, HindIII, PstI, SalI, SmaI, EcoRI, and double digests EcoRI with BamHI, BglII, HindIII, PstI, SalI, and SmaI. B, map showing restriction sites and subclones used in DNA sequencing (see Table V). ----> and <---- indicate clones sequenced using two synthetic primers made to 303-308 and 411-416 codons positions respectively. Arrows indicate (5'-3') direction of DNA sequence determination.

A

a b c d e f g h i j k l m n o

B

Table VII. Differences between the predicted amino acid sequence of
Amylase II and the TAA

Numbering of amino acid residues follows that reported for TAA by Toda
et al. 1982.

Residue Number	TAA	Amylase II
72-73	ASP-CYS (GAPy-TGPy)	THR-THR (ACC-ACC)
80	THR (ACX)	HIS (CAT)
85	THR (ACX)	GLN (CAG)
384-385	ASN-PRO (AAPu-CCX)	ASN-TRP-PRO (AAC-TGG-CCC)
386-387	TYR-ILE (TAPu-ATPy/A)	ILE-TYR (ATC-TAC)
426	SER (TCX/AGPy)	GLY (GGPy)
476	ASP (GAPy)	missing
Total residues each case is 478		

Figure 15 REN mapping of amylase I and II chromosomal genes and flanking regions. A,B- Digests of total DNA of A.oryzae electrophoresed in 0.8% agarose stained with ethidium (A) and hybridized with M71 cDNA (B) or 3C1 5.8kbp fragment (see C) after transfer to membrane. Tracks a and k contain lambda DNA digested with HindIII. Tracks b-j, correspond to EcoRI, EcoRI/BamHI, BamHI, BamHI/HindIII, HindIII, EcoRI/HindIII, EcoRI/PstI and Pst I, respectively. C- shows approximate physical map of chromosomal DNA flanking amylase I and II genes. Dashed line indicates extent of hybridization with 3C1 5.8kbp fragment.

fragment and the class I amylase gene. These results indicated that the 3C1 5.8kbp probe hybridized only to itself and not to the 5.8kbp of 2B1. Further more, the hybridization pattern obtained with the 3C1 5.8kbp probe against the same genomic blot probed with M71 (see Figure 15 C, dashed lines) indicated this fragment originates from the 3' side of the amylase II gene. This is based on the fact that the chromosomal DNA 3.7 kbp PstI-PstI fragment hybridized to both the M71 and the 3C1 5.8kbp fragment. The sequence on the 7.6kbp EcoRI fragment flanking the 3' end of the amylase I gene should have hybridized to the 5.8kbp fragment if there was any homology there (see Figure 15C).

These results suggest that the two amylase genes are not directly linked to each other; a very conservative estimate would place the two genes at least 6 to 10 kbp apart.

(c) Sequence and organization of Amylase II

The complete nucleotide sequence of the 2,882 base pair long EcoRI-HindIII DNA fragment containing the structural gene of Amylase II was determined and the results are shown in Figure 16 and 17. The Amylase II gene was chosen because of convenient endonuclease restriction sites such as HindIII and BamHI that were located just downstream from the 3' end of the stop codon. Some points of interest are noted here.

i) Introns.- The coding region of the gene is longer than expected because it contains 8 intervening sequences. The introns are all relatively short in length, between 55 and 86 nucleotides (Table VI). Sequences at the intervening junctions were compared to the consensus junction sequences of the yeast Saccharomyces cerevisiae (Langford and Gallwitz, 1983) and eukaryotic genes in general (Mount, 1982). The sequences for the donor splice site and acceptor site given below are found in all introns.

(A/C A G / G T A/G A G T.....(T/C)_n N C/T A G / G)

The first two GT and the last two AG are strictly conserved. The C- and T-rich regions found at the 3' end of the introns in other eukaryotes is less pronounced in this gene, albeit 7 out of 8 introns end in T/CAG. In general, the 5' end of the splice junction is less conserved than in the case of the A.niger glucoamylase gene (Boel et al., 1984) which contains 4 introns, and other fungal genes reviewed by Dons et al. (1984).

Further analysis was carried out for the consensus sequence,

5' - Pu C T Pu A C - 3'

found near the 3' terminus of the intervening sequences of other filamentous ascomycetes such as A.awamori (Nunberg et al. 1984), Neurospora crassa (Kinnaird et al., 1982) and Trichoderma reesei (Shoemaker et al. 1983). This sequence is related to the " internal conserved

Figure 16. Complete nucleotide sequence of the 2.8kbp EcoRI-HindIII fragment containing the amylase II gene.

Sequencing was carried out using various restriction fragments (see Tables IV and V, Figures 13 and 14) as described in II,Q.

Symbols used:

- - position of start codon and beginning of precursor protein
- - hydrophobic amino acids located within the putative signal peptide
- A to I - signify Exons
- - cysteine, potential -S-S- bonds
- - putative catalytic residues
- ⌒ - residues located at the active site
- ↓ - beginning of mature protein and site of proteolytic processing of signal peptide
- ◻ - putative glycosylation site / carbohydrate attachment
- ∇ - differences with respect to published amino acid sequence (Toda et al. 1982).
- *** - STOP codon

[illegible]

-200
 CTGCAGAATG CAATTTAAAC TCTTCTGGGA ATCGCTTGA TCCCCGCCC
 Pst I

 -150
 CTGGCCGTAG AGCTTAAAGT ATGTCCCTTG TCGATGCGAT GTATCACAAC

 -100
ATATAAATAC TAGCAAGGGA TGCCATGCTT GGAGGATAGC AACCGACAAC

 -50
 ATCACATCAA GCTCTCCCTT CTCTGAACAA TAAACCCAC AGAAGGCATT

 1
 TATGATGGTC GCGTGGTGGT CTCTATTCT GTACGGCCTT CAGGTGCGCG
 Met

 50
 CACCTGCTTT GGCTGCAACG CCTGCGGACT GCGGATCGCA ATCCATTAT
 ALA
 +1

 100
 TTCCTTCTCA CGGATCGATT TGCAAGGACC GATGGGTCGA C
 Sal I

Figure 17. Nucleotide sequence of the promoter region for *A. oryzae* α -amylase II gene.

Translation initiation site is indicated by an arrow. Nucleotides upstream of the START codon are assigned negative numbers. CAAT and TATA signals are boxed. Dashed lines indicate position of pyrimidine track analogous to CTTCCTCCCTTT sequence associated with position of transcription initiation in TrpC (see text for details).

sequence", the "ICS", - TACTAAC -, found within intervening sequences near the 3' junction in *S.cerevisiae* genes (Langford and Gallwitz, 1983). As shown in Figure 16 and Table VI, all 8 introns of amylase II contain sequences resembling the yeast ICS. Some variability of the consensus sequence is observed for the "Pu" (purine) in introns 2, 6 and 7, where a C (pyrimidine) is found instead. In the ICS of intron 5 there is a purine to pyrimidine (T) transversion at position # 4. In the ICS of intron 8 there is a C to T change at position #6. Both intron 4 and 8 contain an additional ICS, but further upstream from the 3' junction. It may be assumed that the one closer to the 3' end is the preferred sequence used in splicing (Cellini et al. 1986; see also Discussion).

ii) Coding Region.- The N-terminal region of amylase II is located in the PstI-SalI fragment. The entire coding region is divided into 9 segments due to the presence of the 8 IVS elements. Comparison with the published TAA amino acid sequence (Toda et al., 1982) and exclusion of the IVS sequences gave an open reading frame coding for 478 + 21 residues, the same as for TAA. The first open reading frame begins at Met (-21) and extends until it reaches Gln-35. Near complete amino acid sequence homology with the previously published sequence of the mature TAA begins at Ala-1, exactly 21 amino acids down stream from Met (position -21) and 681 base pairs from the EcoRI Site. Two putative start codons are present (ATG) 63 base pairs upstream. It is assumed that the first ATG is used. Other putative start codons are found at 140 and 145 base pairs upstream from Ala-1, but they are not in the same reading frame. Some sequences characteristic of 5' transcription start are apparent in this region (see III, B, d), suggesting that the precursor amylase II begins at Met-21. Fifteen of the first 21 residues are hydrophobic, consistent with the

properties of a secretion signal peptide, generally regarded as important for secretion of such enzymes (Innes et al., 1985).

Several of the introns interrupt codons, Ile-87 (IVS-3), Ile-200 (IVS 5), Ile-254 (IVS-6) and Ser-303 (IVS-7). Other intron junctions are located between codons such as IVS-4 between Met-123 and Gly-124, and IVS-8 located between Lys-383 and Asn-384.

Some differences in amino acid sequence exists between the TAA and amylase II which are summarized in Table VII. All the differences from the published sequence are found in exons C and I. These differences do not affect the sequences postulated to be part of the functional, catalytic domain (Matsuura et al. 1984). All of the amino acids purported to be involved with the substrate binding site are conserved. However, their locations are in different domains, distributed over four different exons, D to G (see Table VIII for further details).

Nine Cys residues were found, consistent with an earlier report (Narita et al., 1966). Four disulphide bonds defined by electron density maps have been previously assigned (Matsuura et al., 1984), linking residues 30-38, 150-164, 240-283, and 439-474. The positions for the fourth bridge would be different due to the shift in the extra Trp-385 found in both cDNA and genomic sequences.

iii) Nucleotide composition and codon usage.- The nucleotide sequence of the coding and extragenic DNA of Amylase II was analyzed and is summarized in Table IX. These data were compared to available published sequence of Aspergilli. Adenine is slightly higher than cytosine and thymine with the less abundant being guanine. The lower frequency of G is also reflected in the G+C content (48.7%) and distribution in dinucleotides, AG, GC, and GT which are the less frequent combinations. The G-C content drops further

Table VII. Differences between the predicted amino acid sequence of
Amylase II and the TAA

Numbering of amino acid residues follows that reported for TAA by Toda
et al. 1982.

Residue Number	TAA	Amylase II
72-73	ASP-CYS (GAPy-TGPy)	THR-THR (ACC-ACC)
80	THR (ACX)	HIS (CAT)
85	THR (ACX)	GLN (CAG)
384-385	ASN-PRO (AAPu-CCX)	ASN-TRP-PRO (AAC-TGG-CCC)
386-387	TYR-ILE (TAPu-ATPy/A)	ILE-TYR (ATC-TAC)
426	SER (TCX/AGPy)	GLY (GGPy)
476	ASP (GAPy)	missing
Total residues each case is 478		

2 of /de 2

PRINTED IN THE UNITED STATES OF AMERICA
Metric
THE LIGHTS TO GOVERN THE WORLD

Table VIII. Location of catalytic sites for TAA and Amylase II based on exon distribution

Residue Number	Exon	Comments

Catalytic Residues		
GLU-230	F	
ASP-297	G	
Unpaired		
CYS-227	F	
Disulphide		
Bonds		
30-38	A-B	
150-164	E-E	
240-283	F-G	
439-474	I-I	shift in position to 440-475
Carbohydrate		
Attachment		
ASN-197	E	
Substrate		
Binding Sites		
HIS-122	D	
TYR-155	E	
ASP-206	F	
LYS-209	F	
HIS-210	F	
VAL-231	F	
HIS-296	G	

Table IX. Base composition of A.oryzae DNA.(a) Data obtained from A.oryzae amylase II nucleotide sequence.(This study)

(b) Data reported by Kogane and Yanagita, 1964.

DNA sample	Base composition				Molar ratios			
	Ade	Gua	Cyt	Thy	$\frac{Pu}{Py}$	$\frac{A}{T}$	$\frac{G}{C}$	$\frac{A + T}{G + C}$
^a AMYLASE II PstI/HindIII. (2471 bp)	25.8	23.2	25.5	25.1	0.97	1.02	0.90	1.05
^a AMYLASE II EcoRI/HindIII (2887)	27.5	22.3	25.2	25.5	0.98	1.07	0.88	1.11
^b <u>A.oryzae</u> DNA germinating conidia	27.2	22.0	25.5	25.4	0.97	1.07	0.86	1.11
^b <u>A.oryzae</u> DNA dormant conidia	28.2	20.5	24.7	26.5	0.95	1.06	0.83	1.21

The results obtained from the hybridization of the α -amylase cDNA clone M71 against total RNA from 10h starch or glucose cultures indicate that there is a fair amount of α -amylase messenger being produced even by the glucose cultures. This result might be related to the fact that at 10h very little glucose is left in the media, which might stimulate the production of the amylase messenger if the induction is exerted at the level of glucose depletion rather than starch presence. A different possibility is that both situations could have an effect on the induction of amylase mRNA, in which case the results obtained from the time course experiment could be explained as a first wave of induction exerted by the lack of glucose in the media at around 6h, followed by some turnover of the amylase messenger at 9h, when it has already been translated into the protein. The second peak of induction would be exerted by some of the products resulting from the starch breakdown.

D. Comparison of amylase genes to genes from other organisms.

(a) Genomic DNA hybridization studies

Total DNA from A.oryzae and A.niger were digested with EcoRI, BamHI and HindIII. The samples were fractionated by electrophoresis and transfer to a Biodyne filter. Hybridization analysis were done using an amylase cDNA clone, M71. The probe was labelled by a primer-extension reaction, using ssDNA as a template and α -(32 P) deoxynucleotides. Radioisotope incorporation was 54% and total amount used was 2×10^7 cpm. Hybridization was done as described previously.

Two different bands were seen in the lanes corresponding to A.oryzae DNA for each of the different enzymes, the size and intensity being very similar to the results obtained using nick-translated probe (Fig. 21).

Table X. *A. oryzae* amylase II codon usage (500 codons).

Percentage utilization in each codon family is shown.

AMINO ACID	SUM	1	2	3	4	5	6
ALA	8.4	GCA 28.6	GCC 35.7	GCG 23.8	GCT 11.9		
ARG	2.0	AGA 0.0	AGG 30.0	CGA 20.0	CGC 20.0	CGG 20.0	CGT 10.0
ASN	5.2	AAC 88.5	AAT 11.5				
ASP	8.4	GAC 52.4	GAT 47.6				
CYS	1.8	TGC 22.2	TGT 77.8				
GLN	4.0	CAA 25.0	CAG 75.0				
GLU	2.4	GAA 41.7	GAG 58.3				
GLY	8.6	GGA 23.3	GGC 41.9	GGG 11.6	GGT 23.3		
HIS	1.4	CAC 57.1	CAT 42.9				
ILE	5.6	ATA 7.1	ATC 67.9	ATT 25.0			
LEU	7.4	CTA 8.1	CTC 35.1	CTG 16.2	CTT 10.8	TTA 2.7	TTG 27.0
LYS	4.0	AAA 35.0	AAG 65.0				
MET	2.2	ATG 100.0					
PHE	3.0	TTC 80.0	TTT 20.0				
PRO	4.4	CCA 13.6	CCC 36.4	CCG 22.7	CCT 27.3		
SER	7.4	AGC 16.2	AGT 10.8	TCA 10.8	TCC 24.3	TCG 27.0	TCT 10.8
THR	8.0	ACA 25.0	ACC 32.5	ACG 20.0	ACT 22.5		
TRP	2.4	TGG 100.0					
TYR	7.0	TAC 68.6	TAT 31.4				
VAL	6.2	GTA 12.9	GTC 38.7	GTG 32.3	GTT 16.1		
TER	0.2	TAA 0.0	TAG 0.0	TGA 100.0			

Upstream are located features noted for typical eukaryotic promoters by Breathnach and Chambon (1981): a CAAT box at -191 base pairs and a prominent TATAA box (Gannon *et al.*, 1979) at -100 base pairs from the MET codon. Further upstream within the EcoRI-PstI fragment which is A-rich, are a number of repeat sequences which may be part of a cis-acting enhancer required for regulation of the Amylase gene expression (see III, C, b).

The stop codon for the Amylase II gene is TGA, located immediately after the SER codon reported by Toda and colleagues (1982). There is no AAUAAA signal sequence for polyadenylation (Proudfoot and Brownlee, 1976). However, some information can be obtained from the analysis of the cDNA clones which were made by using oligo.dT as primer for initiation of transcription at the 3' polyA tails. The 3' ends of two cDNA clones (C15-ON2 and C4) which are completely homologous to the Amylase II gene indicates a 3' trailer sequence with a polyA tail located 179 nucleotides downstream from the stop codon. However, some heterogeneity probably exists because a third clone (C17) was found to have a 240 nucleotide long trailer.

The 3' UTRs of Amylase I and II were compared and found to be different (Fig. 19). The homology extends 66 nucleotides downstream from the stop codon at which point most of the homology is lost. Amylase I does contain a polyadenylation signal AATAAA. The cDNA clones M71 and C16 which match with the Amylase I gene establish further that M71 contains the AATAAA sequence and a polyA tail only 3 bases further away rather than the usual 11-19 for most genes (Nevins, 1983). In clone C-16 the polyA addition was found upstream from the AATAAA signal. The presence of TACT, TATGT signals recognized in yeast are apparent in both types of genes (see III, C,).


```

      Bgl II          STOP
AGATCTGTAG TAGCTCGTGA AGGGTGGAGA GTATATGATG GTA CTGCTAT TCAATCTGGC 60
*****
AGATCTGTAG TAGCTCGTGA AGGGTGGAGA GTATATGATG GTA CTGCTAT TCAATCTGGC 60

ATTGGACAGT GAGTTTGAGT TTGATGTACA GTTGGAGTCG TTA CTGCTGT CATCCCCTTA 120
*****
ATTGGACAGT GAGTTTGAGT TTGATGT--A ACTTGTCTAT TCTATGATGT -ATGGTCTTT 117

TACTCTTCGA TTGTTTTCG AACCTAATG CCAAGCACGC TATCTATTA TAGGAAAGGA 180
* * * * *
T--TGTTCTA TAG--TTGGA AATCGGAATG ATCTCAAATC TTGAATAAAT ATAAAAAGGA 173
                                     ↑
HI
TCGGATTAA TATGTTTCA TAACGCGGTA CTGTATGGTA CTTCTGTATT ATAT--CACC 238
* * * * *
T----AATAC TCACATCCAT CACAACCTTA C--AAGGTTA ATTCCGAGCT ATATTCCACC 227

Hind III
GAAGCTTAT- ----GTATCT TACATATATA TATT----AT A-- 270
** * **
GACACAAATA GGCAGTATCT TCTCTCGCCA GAATCGCGAT ATT 270

```

Figure 19. DNA sequence alignment of the 3'-untranslated regions of *A. oryzae* α-amylase genes I (bottom) and II (top).

The analysis was done using a MATDNA program, which finds the best alignment between two nucleotide sequences, for the regions comprising the Bgl-II site upstream of the STOP codon to 270bp downstream, and the polyadenylation site. Restriction sites are indicated. The putative polyadenylation signals for yeast (---) or eukaryotes (—) are underlined. Putative polyadenylation sites predicted from the cDNA clones are indicated by an arrow.

C) Expression of Amylase Genes I and II

(a) Identification of new cDNA clones related to Amylase genomic sequence

The cDNA banks were hybridized with the EcoRI 3.35 kbp genomic fragment which contains the entire sequence of Amylase II. This fragment also carries 420 and 470 base pairs of 5' and 3' flanking sequence that may also code for structural information. Approximately 8,000 clones from the three banks replica plated onto 20 filters were screened. A large number of positives were detected with hybridization signals of "strong", "medium" and "weak" intensities. From these a total of 40 clones, 20 "strong" (C1 to C20) and 20 "medium" (T1 to T20) were carried through. In the second round of screening only C-6 and T-12 failed to give good hybridizations. Out of the 38 positives, 20 were chosen for rescreening and only 15 gave any indication of insert after sizing.

The range of inserts was between 50 to 1000 bases in length; however, 13 clones had inserts of 600 bases or greater. None of these would be the size of the putative Amylase transcript which would be about 1800 bases in length. A "T-track" sequencing gel was run with all 15 templates in an attempt to finger print and compare the sequences. T-3 was discarded because it had a small insert. The initial ladder of T in 6 clones suggested that these clones were derived from a polyA mRNA template during cDNA synthesis. However, none of these resembled the sequences for the 3' end of either Amylase I or II. Several of the remaining clones were partially sequenced. With the exception of C-2 all other showed some homology with the 3' end of the amylase mRNA.

Clone C-15 was sequenced because it showed homology with the 920 base pair clone 4H (SaliI/SmaI) containing part of the 5' end of Amylase II. The 800 base pair C-15 starts at codon SER-89 of TAA and extends to

Figure 18. General map showing the structural features and organization of the amylase II gene including position of various cloned cDNA derivatives.

α -AMYLASE I

α -AMYLASE II

Ser-141 (160 nucleotides) after which no homology was detected for 220 bases. The Ser-89 to Ser-141 begins immediately down stream of IVS-3 and by passes IVS-4. Sequencing the subclone of C-15 in the opposite direction (C15-ON2) revealed a 10-mer dT tail, a 189 base 3' UTR homologous to Amylase II, including the BamHI sequence 16 nucleotides upstream from the polyA tail, and coding sequence up to Ser-350. Beyond this codon the sequence again becomes non-homologous. Using the 3' end as the point of reference, the 610 remaining nucleotides should have coded for 203 amino acids, or from positions 274 to 478 and not to 89. Several other clones have been found which begin or end at or near introns.

T-17 contains 317 base pairs and is inserted in the 3'-->5' orientation for sequencing. However, no 3'UTR is present and sequence begins at Asp-390 and extends upstream to Ser-286, by passing IVS-8 and IVS-7 by 5 and 17 codons, respectively. The rest of the clones are inserted in the 5'--> 3' orientation. Partial sequence of these clones could not give any information about which gene the transcript belong to. However, none appear to be large enough to include everything and the complete 3'UTR. C-9 begins at Tyr- 304 and extends beyond Thr-429, starting just after IVS-7 and bypassing IVS-8. T-1 begins at Pro-299, four codons before IVS-7 and extends beyond Lys-389, by passing IVS-7 and 8. T-14 begins at Pro-284, 19 codons before IVS-7 and continues by passing IVS-7 up to and beyond Ser-364. C-14 for 28 bases has no homology and then starts at Tyr-255, immediately downstream of IVS-6; it by passes IVS-7 and extends to at least Gly-337.

Of the clones that seemed to have a polyA tail (C-17, C-16, C-4, C-18 and T-13), three were analyzed further. The template strands for C-4, C-16 and C-17 all correspond to the mRNA. The 414 nucleotides of C-4 has a 181 base long 3'UTR with a 15-mer polyA tail 16 nucleotides after the BamHI

site, and extends upstream to at least Val-406. C-17 has a 240 base 3'UTR with a 13-mer polyA, homologous to Amylase II but includes 63 bases more than the 3'UTRs of C-4 or C-15-ON2. The coding region extends from 478 up to at least Thr-429. C-16, in contrast, has a 128 base 3'UTR and a 8-mer polyA tail and matches with M-71, a transcript of Amylase I, but C-16 is 17 bases shorter than M-71 in this region. C-16 has amylase coding sequence, at least up to Gly-427. Positioned on the Amylase I genomic sequence, the polyA tail of C-16 starts at a 3-mer A track before the AATAAA polyadenylation site. This is in contrast to M-71 which is 17 bases longer and whose polyA tail begins at a 5-mer A cluster down stream from AATAAATAT-AAAAAA.

Partial sequences have been obtained for two clones which appear to be fusions between amylase sequence and other unknown sequences. C-12 contains a 150 base pair homology with the 3'UTR of Amylase II, starting one nucleotide down from the stop codon, continuing up to the site of polyadenylation normally for Amylase-II. However, no polyA tract was found and the remainder of the sequence does not match amylase sequence. T-9 also has two different fragments of unrelated sequence. The first consists of a 127 base region of non-homology fused with an 11-mer A track and sequence matching the 3'UTR of Amylase II. This polyA tract is in a new location, starting at a 3-mer A track immediately upstream of the BamHI sequence and extends to at least 24 bases downstream of the stop codon.

b) mRNA levels in starch and glucose grown mycelia

Erratt et al. (1984) were the first to demonstrate the increase in amylase mRNA during the early stages of A.oryzae growth on starch. These studies were essentially verified in this work. It was shown that the level of translatable amylase mRNA was maximal between 6 to 15 h (Fig.20).

To assess the transcript size and abundance further, the coding regions of amylase I and II genes were used as probes against filter blots of total RNA separated by electrophoresis or dot blots. As shown in Fig. 20 A, a major band approximately 1.8kbp in size is detected with the highest intensity of hybridization appearing at 12 to 15h. Some hybridization was detected at 24h only after 3 days of exposure of the film. At 48h no hybridization was detected. The 6 h sample also hybridized strongly, however, the nine hours one presented relatively weaker hybridization. These results were in duplicated experiments and when filters were hybridized with the c-amylase II. CR (data not shown).

The quantification of these results was done by using the same probe against a dot-blot of the same total RNA samples at different concentrations. As shown in Fig. 20 C, there is a linear increase in the intensity of hybridization from 2 μ g to 15 μ g.

These results are in general agreement with the ones obtained by Erratt et al. (1984) and in the present study using IVT, where it was shown that maximum levels of c-amylase mRNA were present at 10-12h starch cultures, after which the RNA is quickly degraded.

The result for the 9h sample is interesting, since it was expected to give a level of hybridization between 6h and 12h samples. The fact that the same results were obtained in more than one separate experiment eliminates the possibility of spurious sample degradation or hybridization. The EtBr picture above the dot blot (Fig. 20B) indicates the intactness of the samples. It could be argued that the α -amylase specific mRNA is being differentially degraded reflecting a turnover of message. In which case there would be two peaks in the induction and expression of the α -amylase messenger.

Figure 20 . Time course of appearance of α -amylase mRNA after induction with starch.

A. Analysis of total RNA extracted from mycelia cultures grown on starch at various times. 4 μ g of RNA from each sample were denatured with glyoxal and electrophoresed on a 10mM sodium phosphate (pH 7.0) , 1% agarose gel, before transfer onto nylon membrane and hybridization against nick translated 32P-1.9kb Sal/BglIII fragment containing the coding region for α -amylase I gene. Lanes a to f: samples from 6, 9, 12, 15, 24 and 48h respectively. Molecular weight marker was the BRL RNA ladder.

B. 2 μ g of total RNA from same samples than A) were electrophoresed on a 0.025M sodium citrate, 1.5% agarose gel and stained with EtBr.

C. Dot blot of total RNA samples, denatured in 2XSSC, onto nylon membrane and hybridized against same probe that in A) 2, 5, 10, and 15 indicate the amount of RNA loaded in μ g.

D. Total RNA samples from A.oryzae mycelia cultures grown for 10h on glucose (g) or starch (h) media. Electrophoresis was done as in A). Transfer was done on NCF. Hybridization was against primer extended (32P)-cDNA clone M71.

← 1.8kb

The results obtained from the hybridization of the α -amylase cDNA clone M71 against total RNA from 10h starch or glucose cultures indicate that there is a fair amount of α -amylase messenger being produced even by the glucose cultures. This result might be related to the fact that at 10h very little glucose is left in the media, which might stimulate the production of the amylase messenger if the induction is exerted at the level of glucose depletion rather than starch presence. A different possibility is that both situations could have an effect on the induction of amylase mRNA, in which case the results obtained from the time course experiment could be explained as a first wave of induction exerted by the lack of glucose in the media at around 6h, followed by some turnover of the amylase messenger at 9h, when it has already been translated into the protein. The second peak of induction would be exerted by some of the products resulting from the starch breakdown.

D. Comparison of amylase genes to genes from other organisms.

(a) Genomic DNA hybridization studies

Total DNA from A.oryzae and A.niger were digested with EcoRI, BamHI and HindIII. The samples were fractionated by electrophoresis and transfer to a Biodyne filter. Hybridization analysis were done using an amylase cDNA clone, M71. The probe was labelled by a primer-extension reaction, using ssDNA as a template and α -(32 P) deoxynucleotides. Radioisotope incorporation was 54% and total amount used was 2×10^7 cpm. Hybridization was done as described previously.

Two different bands were seen in the lanes corresponding to A.oryzae DNA for each of the different enzymes, the size and intensity being very similar to the results obtained using nick-translated probe (Fig. 21).

Figure 21 . Mapping of 3'-amylase genomic DNA homology for A.oryzae and A.niger using M71 cDNA probe.

A. Agarose (0.8%) gel electrophoresis of REN digests stained with EtBr. Tracks b, c and d, EcoRI, BamHI and HindIII digested A.oryzae DNA. Tracks e, f and g, A.niger DNA. Tracks a and h contain λ -HindIII digested DNA as size markers.

B. 10h autoradiograph showing positively hybridizing fragments detected by use of 2×10^7 cpm. of primer-extended radiolabelled M71 cDNA.

A.niger DNA hybridized differently to the probe. There was only one band lighting up in every case. The restriction fragments sizes can be seen in table XI. These results clearly indicate that while there are two α -amylase related genes in A.oryzae, only one is detected by A.oryzae amylase cDNA 3' probe in the A.niger genome.

Table XI . Amylase related REN fragment size of genomic DNA from A.oryzae and A.niger.

restriction endonuclease	Fragment size (kbp)		
	<u>A.oryzae</u>		<u>A.niger</u>
EcoRI	7.6	3.35	2.85
BamHI	8.6	7.2	9.0
HindIII	8.7	6.6	8.9

(b) Sequence comparisons.

An extensive study on the amino acid homologies between α -amylases from several sources has been published by Mackay et al. (1985) (see also reviews by Nakajima et al. 1986 and Ihara et al. 1985). However, homologies at the 5' and 3'-UTR were not possible given the lack of information on genomic sequences. With the availability of α -amylase II gene sequence, a search was made for regions of homology at the 5' and 3'-UTR between this fungal α -amylase and other amylases and fungal glucoamylases. Fig. 22 illustrates the results of comparing α -amylase II 5'-UTR and signal peptide nucleotide sequence against S.diastaticus glucoamylase (Yamashita et al. 1985)(A), or A.niger glucoamylase (Boel et

Figure 22. DNA sequence alignment of the 5' promoter and signal peptide regions of A. oryzae α -amylase II gene and fungal glucoamylases.

A). A. oryzae α -amylase II (top) versus S. diastaticus glucoamylase (bottom).

B). A. oryzae α -amylase II (top) versus A. niger glucoamylase (bottom).

CAAT and TATA signals are boxed. Translation initiation sites are indicated by an arrow. Asterisks emphasize identical positions in this alignment. Dashed lines indicate where it is necessary to introduce gaps in the sequence for best alignment.

A

CTGCAGAATG CAATTTAAAC TCTTCTGCGA ATCGCTTGGG TTCCCCGCC CTGGCCGTAG 60
 * *** ** * * * * * * * * * *
 -----G ATCTTTTGCT TCCTAAACTA AACTATAAA AAGC---ACC CTATTATCA 48
 AGCTTAAAGT ATGTCCCTTG TCGATGCGAT GTATCACAAC ATATAAATAC TAGCAAGGGA 120
 *** ** **** ** * * * * * * * * * *
 GTTATAA--- ---TCTCTTG TC-ATGTTGT GGTCTAATT GAAAATATAC T----- 92
 TGCCATGCTT GGAGGATAGC AACCGACAAC ATCACAACAA GCTCTCCCTT CTCTGAACAA 180
 *** * ** ** * * * * * * * * * *
 ---ATGGTA GG----- --CCTCAAAA ATCCATATAC GCACACTATG CAAAGACCAT 138
 TAAACCCAC AGAAGGCATT TATGATGGTC GCGTGGTGGT CTCTATTCT GTACGGCCTT 240
 * * * * * * * * * * * * * * * *
 TTCTACTCGC -----T TAT-TTGGTC ---CTTTCGC TTCTATT-- -----AACT 277
 CAGGTCGCGG CACCTGCTTT GGCTGCAACG 270
 *** *
 CAGCT----- 182

B

CTGCAGAATG CAATTTAAAC TCTTCTGCGA ATCGCTTGG- ATTCCCCGCC CCTGGCCGTA 59
 **** ** * * * * * * * * * *
 -----GAATT CAAGCTAGA- TGCTAAGCGT ATTGCATGCC AATATGTGTT GATGCATGTG 54
 GAGCTTAAAG TA--TGTCCC TTGTGATGC GATGTATCAC AACTAT--A AATACTAGCA 115
 *** * * * * * * * * * *
 CTTCTTCCTT CAGCTTCCCC TCGTGGGAGT GAGGTTTGGC TATAAATTGA AGTGGTTGGT 114
 AGGGATGCCA TGCTTGGAGG ATAGCAACCG ACAACATCAC ATCAAGCTCT CCCTTCTCTG 175
 *** * * * * * * * * * *
 CGGGGTTCCTG TGAGGGGETG A----- AGTGGTTCCT TTTTAG---G CGCAACTGAG 162
 AACATAAAC CCCACAGAAG GCATTTATGA TGGTC-GCGT GGTGGTCTCT ATTTCTGTAG 234
 * * * * * * * * * * * * * * * *
 AGC-CTGAGC TTCATCGGCA GCATCATTAC ACCTCAGCAA TGTGTTCCG ATCTCTACTC 221
 GGCCTTCAGG TCGGGGCACC TGCTTTGGCT --GCAACG 270
 * * * * * * * * * * * * * * * *
 GCCCT--GAG CGGCTCGTC TGCACAGGT TGGCA--- 254

al. 1984) (B). The yeast sequence does not have a CAAT box, and although some similarities are found, these do not correspond to functionally equivalent areas. The homologies at the sequence level are not very high between the Aspergillus sequences, however the alignments reflect some similarities in terms of location of the TATA and CAAT boxes with respect to the origin of translation (or start codon). Vertical arrows correspond to the 2 different initiation points for the glucoamylase transcripts as determined by Boel et al. (1984).

Amino acid sequences of the signal peptides for the same 3 genes were compared for maximum homology (Figure 23). No significant homologies were found. The 3 peptides differ in size as well as sequence, but the 2 Aspergillus ones are more alike.

Comparison of the 3'-UTR of A.oryzae amylase II gene against the two fungal glucoamylases was also done (Figure 24). And again, no significant homologies could be found. Both A.oryzae amylase II and A.niger glucoamylase present sequences that resemble the yeast sequences postulated to be part of the polyadenylation signal (Zaret and Sherman, 1982). However, their relative distances to the polyadenylation site are different. S.diastaticus contains the sequence 5'-AATAAA-3' and analogous sequence in the 3' flanking region of the glucoamylase gene (Yamashita et al. 1986) some of which are underlined in figure 24. This region also contains yeast-related polyadenylation signal sequences, however the 3' polyadenylation site is not being determined.

For comparison of secretion signal peptides, amylases from different sources were aligned versus the A.oryzae amylase II sequence for the putative signal peptide (Fig. 25). Bacterial, animal and plant amylase signal peptides all have an equivalent degree of similarity to the fungal one, although the animal sequences resemble each other to a higher extent

<u>S. diastaticus</u> , gluco	M V G L K N P Y T H T M Q R	
	R P F L L A Y L V L S L L F N S A	30
<u>A. oryzae</u> , amylase	M M V A W W S L F L Y G L Q V A A P A L A	21
<u>A. niger</u> , gluco	M S F R S L L A L S G L V C T G L A	18

Figure 23. Comparison of signal peptides of A. oryzae α -amylase II gene versus fungal glucoamylases.

Amino acid residues are shown by single letters. Sequences have been aligned for maximum homology. Asterisks indicate identical positions with respect to the A. oryzae α -amylase sequence. The number of amino acids is indicated to the right of the sequence. Sequence sources:

A. niger, Boel et al. 1984; S. diastaticus, Yamashita et al. 1985;
A. oryzae, original results.

Figure 24. DNA sequence alignment of the 3'-untranslated regions of A.oryzae α -amylase II gene against fungal glucoamylases.

A). A.oryzae α -amylase (top) versus S.diastaticus glucoamylase (bottom).

B). A.oryzae α -amylase (top) versus A.awamori glucoamylase (bottom).

The analysis was done using a MATDNA program, which finds the best alignment between two nucleotides sequences. Stop codons are indicated.

The putative polyadenylation signals for yeast (----) or eukaryotes (—) are specified. Polyadenylation sites are indicated by an arrow.

B

<u>BglII</u>		STOP					
AGATCTGTAG	TAGCTCGTGA	AGGGTGGAGA	GTATATGATG	GTACTGCTAT	TCAATCTGGC		60
	*** * * *		* * * *	* * * *	*** **		
-----	<u>TAGACAATCA</u>	A-----	-TCCATTTCG	CTATAGTTAA	AGGATGGGGA		40
	STOP						
ATTGGACAGT	GAGTTTGAGT	TTGATGTACA	GTTGGAGTCG	TTACTGCTGT	CATCCCCCTTA		120
*** **	*** *	* * * * *	* * * * *	*	* * *		
TGAGGGCAAT	TGGTTATATG	ATCATCTA--	--TGTAGTGG	GT-----	--GTGCATAA		86
TACTCTTCGA	TTGTTTTTCG	AACCCTAATG	CCAAGCACGC	TAGTCTATTA	TAGGAAAGGA		180
*** * *	** *	* * * * *	* * *	* * * * *	* * * * *		
TAGTAGTGAA	ATG-----G	AAGCCAAGT-	-----CATGT	GATTGTAATC	GACCGACGGA		134
<u>HI</u>	-----	↓		↑			
--TCCGGATT	AATATGTTTT	CATA-ACGCG	GTACTGTATG	GTACTTC---	--TGTATTAT		232
* * * * *	* * * * *	* * * * *	* * * * *	* * * * *	* * * * *		
ATTGAGGATA	TCCGGAAATA	CAGACACCGT	GAAAGCCATG	GTCTTTCCCTT	CGTGTAGAAG		194
	<u>HindIII</u>	---	↓				
ATC-----A	CCGAAGCTTA	TGTATCTTAC	ATATATATAT	TATA--			270
* * *	* * * * *	* * * * *	* * * *	*			
ACCAGACAGA	CAGTCCCTGA	TTTACCCTGC	ACAAAGCACT	AGAAAA			240

<u>B. stearothermophilus</u>	MLTFHRIIRKGMELLAFLLTALLFCPTGQPAKA	34
<u>B. amyloliquefaciens</u>	MIQKRKRTVSFRLVLMCTLLFVSLPITKTS*	31
<u>B. subtilis</u>	MFAKRFKTSLLPLFAGFLLLFHLVLAGPAAASA	33
<u>A. oryzae</u>	MMVAWWSLFLYGLQVAAPALA	21
Rat	MKFVLLLSLIGFCWA	15
Mouse, s	MKFFLLLSLIGFCWA	15
Mouse, p	MKFVLLLSLIGFCWA	15
Human, s	MKLFWLLFTIGF	12
Human, p	MKFFLLLSLFTIGF	12
Barley	MGKNGSLCCFSLLLLLLLLAGLAS	24
Wheat	MANKHLSSLFLVLLGLSASLAS	23

Figure 25. Comparison of signal peptides of various α -amylases.

Amino acid residues are shown by single letters. Sequences have been aligned for maximum homology. Asterisks indicate identical positions with respect to the A. oryzae α -amylase sequence. The number of amino-acids is indicated to the right of the sequence. Sequence sources:

B. stearothermophilus, Nakajima *et al.* 1985; B. amyloliquefaciens, Takkinen *et al.* 1983; B. subtilis, Yang *et al.* 1983; Rat, Macdonald *et al.* 1982; Mouse, pancreatic and salivary, Hagenbuchle *et al.* 1980; Human pancreatic and salivary, Nakamura *et al.* 1984; Wheat, Rothstein *et al.* 1984; A. oryzae, original results.

(also see Ihara et al. 1985). The homologies found seem more coincidental than specific, considering that the signal peptide serves an identical function in all cases, which already is defining constraints for what the sequence can be.

IV. DISCUSSION

A. Cloning Strategy, Comparison With Other Systems

A large number of structural genes have now been cloned from a variety of organisms, from bacteria, lower eukaryotes and higher eukaryotes. The majority of these genes were cloned by using various strains of Escherichia coli K12 as a host and derivatives of its naturally occurring plasmids and phage such as lambda and M13 as the cloning vectors. Most of the bacterial genes and only a very few of the eukaryotic genes have been identified by direct expression of either the protein product or genetic complementation. Even the use of alternate host expression systems such as the well characterized yeast, Saccharomyces cerevisiae, has not proven particularly effective for cloning eukaryotic genes, even from other yeast and ascomycetous fungi (Seligy et al., 1986). The major problem appears to be the inability of such hosts to recognize heterologous transcription and intron splicing signals. Additional barriers also exist at the levels of translation and particularly post-translational modification, involving glycosylation, proteolytic processing and secretion (Schekman and Novick, 1982). As a consequence a large number of different strategies for cloning eukaryotic genes have been proposed and used.

The cloning of complementary DNA (cDNA) copies of eukaryotic gene transcripts has been an important innovation, making use of the generally higher copy of specific mRNA, in comparison to its single copy genomic sequence, the absence of probable intervening sequence(s) found in the gene and occasionally 5' untranslated sequences, which may allow for fusions with E.coli or S.cerevisiae sequences needed for expression.

Prior to this cloning project, preliminary work in this laboratory on attempts to clone the TAA gene by shot-gun cloning and screening for expression of an active enzyme product, and work on other glycosidase genes from filamentous fungi (Seligy et al., 1983; Shoemaker et al. 1983; Boel et al. 1984; Nunberg et al., 1984; Innis et al., 1985), indicated the high probability that the gene coding for the TAA may contain introns. However, the extensive structural characterization of the TAA (Toda et al., 1982; Matsuura et al., 1984) and its regulative expression in A.oryzae (Erratt et al., 1984), suggested that a cDNA approach to cloning the TAA gene was feasible. The key observation for the cDNA approach was the temporal correlation of TAA production, induced during A.oryzae growth on starch medium, and its mRNA abundance, assayed by in vitro translation and immunoprecipitation by use of a polyclonal anti-amylase antibody (Erratt et al. 1984). The optimization of these two parameters can vary considerably from one system to another as has been demonstrated for the induction of galactosidase of the yeast, S.cerevisiae (St.John's and Davis, 1981), and the cellulases of the fungus Schizophyllum commune (Willick and Seligy, 1985). In the present work, the optimization of amylase mRNA (polyA⁺ RNA) was confirmed to be between 6 to 15 hours of growth on starch, and RNA from 10 hour cultures was used as the source for template to synthesize cDNA from which a shot-gun cDNA clone bank was made.

The approach taken for the isolation of TAA specific cDNA clones originally involved two options: 1) identification of clones producing amylase antigen by use of the available polyclonal amylase-antibody, and 2) identification of clones carrying nucleotide sequences homologous to an oligonucleotide hybridization probe whose sequence was predicted from the published TAA amino acid sequence (Toda et al., 1982). The vector of

choice was the M13 phage derivative, mp9, of Messing (1983), which contains part of the E.coli β -galactosidase expression system, potentially capable of producing fusion proteins and allowing direct identification of the coding sequence by nucleotide sequencing (Sanger et al., 1980; Messing, 1983). This type of approach has been successfully employed in this lab on another project, involving the cloning of cDNA coding for parts of the lysine-rich histone H5 of Xenopus laevis (Rutledge, 1987) and β -glucosidase of S.commune (Moranelli et al., 1986). In the latter example, the galactosidase expression system is carried in the lambda gtl1 vector instead of M13. Unfortunately, this approach failed to turn up any expression clones, albeit several false-positives were initially isolated after two rounds of screening with anti-amylase antibody according to published procedures.

The 20-mer, mixed oligonucleotide probe used to successfully identify TAA cDNA clones was derived from the amino acid sequence, VAL-450 to GLY-456 (Toda et al., 1982). This type of approach has also been used to isolate identical glucoamylase DNA sequences from A.niger (Boel et al., 1984) and A.awamori (Nunberg et al., 1984). The method used in the present work was that of Wallace et al. (1981) which makes allowances for mixed sequences. Analysis of the probe-positive clones by DNA sequence and comparison of the predicted amino acid sequence with the published TAA sequence, indicated that none of the clones contained a full length copy of the predicted 1750 to 1800 nucleotide amylase transcript. However, clone M13mp9-M71 contained enough sequence to adequately identify amylase genomic DNA and corresponding clones from a shot-gun library made from A.oryzae DNA partially digested with EcoRI and ligated into the phage vector, lambda Charon 4A (Slightom et al., 1980). The presence of a single BglII restriction site, 6 codons upstream from the stop codon,

proved exceedingly useful for mapping the chromosomal genes. Another important step was the use of the cloned genomic DNA as a probe to rescreen the cDNA banks for sequences in the 5' end of the transcript and sequences in the same orientation (non-complementary) to the original 20-mer probe.

The oligonucleotide probe with sequence for the 5' coding region of the amylase transcript (Probe I) should also have been very useful for obtaining clones with sequence in that region and transcript mapping. However, this probe cross-hybridized extensively with non-amylase sequence. Also, in retrospect, probe I probably would not have detected any complementary cDNA clones because full-length or 5' end amylase cDNA may not be present, as indicated by screening of the cDNA banks with the amylase-II genomic sequence. Failure to obtain full-length cDNA could be due to several factors: 1) quality (size) of the mRNA as a template for reverse transcriptase and 2) presence of secondary structures due to sequences within the template, which promote premature transcription termination or deletions on cloning in M13mp9. The polyA+RNA used for cDNA synthesis must have been of reasonable quality as amylase-specific peptides (50,000 to 55,000 Da) were readily detected by gel electrophoresis after cell-free translation of the RNA and immunoprecipitation with antibody and Sepharose-protein A. Moreover, the size of the cDNA synthesized (Figure 3) was also adequate. The hybridization analysis of the RNA sized by electrophoresis (Figure 20) indicates that the bulk of the amylase specific RNA corresponds to approximately the size expected, about 1,710 to 1,800 bases. The presence of 8 introns within the coding region of the amylase gene may be an important factor as evidenced by the recovery of several partial cDNAs

which have either an apparent deletion or termination before or after an intron. Judging by the number of clones of this type recovered, these results indicate that the initial RNA extraction probably caused a disruption in amylase mRNA splicing. Deletions due to instability of heterologous DNA in M13 during propagation in E.coli may not be involved, since the genomic sequences have been successfully cloned into the same vector along with several other types of sequences cloned in this lab (the B.subtilis and B.polymyxa cellulase genes, Seligy et al., 1986, S.Baird and V.Seligy, unpublished; the H5 cDNA, Rutledge, 1987 and rat oncomodulin cDNA, Gillen et al., 1986).

B. The Coding Region of Amylase-II Gene

The results from the sequence analysis of the cloned 3.35 kbp EcoRI fragment clearly indicates that it contains a complete structural gene coding for a 499 amino acid precursor amylase. Based on the reported sequence of TAA, this precursor when correctly processed would lead to production of the mature, active enzyme of nearly identical amino acid sequence and of exactly identical size (478 residues) to the TAA. Kozak (1983) has proposed that the consensus nucleotide surrounding the translation start codon ATG is either purine (usually A) in position -3 or G in position +4, or both. Neither of the two nucleotides can be found if the first ATG is considered to be used as the translation initiation point. If the second ATG is used, both A at -3 and G at +4 are found. In a later compilation of data, Kozak (1984) indicated that CCACC-ATG should be considered as the translation initiation site consensus sequence, which was not found in the amylase II in this region. Sequence of the pre-amylase protein would resolve this point. For this work the first A of the first ATG is considered as the start codon for the preprotein.

The open reading frame for the presumptive precursor, begins at one of two adjacent ATG codons. The 21 amino acids defining the precursor amino terminus, is rich in hydrophobic residues, characteristic of other signal peptides of secretory proteins (Schekman and Novick, 1982), but like other signal peptides, not particularly homologous as shown by the comparison with other signals (Fig. 25). The predicted size of this amylase precursor is also consistent with the size proposed from the cell-free translation analysis (Erratt *et al.*, 1984).

The coding region is divided up by 8 introns into 9 segments or exons. Four of the eight introns interrupt the sequence in the middle of codons. In four cases, codons (three for Ile and one for Ser) are generated by removal of an intron.

Several differences are observed between the amino acid sequence predicted from the amylase-II gene and that corresponding to TAA (see III, Table IV). The changes observed within exon region "1" have been corroborated with sequence from several cDNA clones, leaving no doubt that these changes are authentic rather than artifacts originating from the cloning and sequencing procedures. However, the amino acid residues proposed at positions 72, 73, 80 and 85, in the N-terminal region, could have originated from ambiguities introduced during subcloning, when these segments were amplified by replication in *E. coli*. Since the error rates for amplification of cloned DNA in a bacteriophage is about 10^{-8} in comparison to 10^{-4} for *E. coli* DNA polymerase I used in the sequencing reaction (Weymouth and Loeb, 1978) any error is more likely due to the latter rather than the former. However, no discrepancy in sequence was observed between repeated cycles of sequencing the same templates and the opposite strands of subclones of these templates. Also, the changes

listed in Table IV, when correlated with the respective nucleotide changes, suggest that a minimum of two nucleotides per codon are required to account for the discrepancies.

The disagreement between the predicted sequence and the sequence published by Toda et al. (1982) may be due to inaccuracies in the reported TAA amino acid sequence due to faulty analysis or use of more than one source of TAA for sequencing. The TAA was purified from Takadiastase Sankyo, a commercial source in which no strain specification was given. Initially the C-terminal TAA sequence was reported by Narita (1975) as -LYS-ILE-CYS-SER-SER-SER- which agrees with the gene sequence work rather than the -LYS-ILE-CYS-SER-ASP-SER-SER- reported by Toda et al. (1982). The finding of two genes here that are both expressed might account for some of the amino acid changes. The amylase-I gene is currently being sequenced (M.Dove, M.J.Gines and V.Seligy). However, the present sequence evidence from the 3' end (C-terminal region) of these two genes and their respective cDNAs predict the same sequence as Narita for both proteins. It is likely that the presence of -TRP- between ASN-384 and PRO-385 as predicted here probably would not have been detected by amino acid analysis (M. Yaguchi and V.Seligy, personal communication). It is interesting that all of these amino acid changes are located in regions not directly postulated to be involved in the catalytic or binding sites of the enzyme (Matsuura et al., 1984). Therefore, the minor differences between specific amino acid residues of TAA and of amylase II may be real.

C. Comparison of Amylase-II Nucleotide Sequence With Other Genes

(a) Codon Usage

Analysis of codon utilization predicted by the cDNA and genomic sequence indicated some preference for certain codons over others (Table V). However, comparison with other genes such as the glucoamylase of A. awamori (Nunberg et al., 1984), cellobiohydrolase of Trichoderma reesei (Shoemaker et al., 1983), several yeast genes (Bennetzen and Hall, 1982) and the Trp-1 of Neurospora crassa (Schechtman and Yanofsky, 1983) and TrpC of Aspergillus nidulans (Mullaney et al., 1985) did not reveal any particular striking novelty in the results, except as expected the codon bias is not anywhere the extreme as already reported for S. cerevisiae. As Nunberg and co-workers (1984) pointed out, species differences in codon utilization within well-expressed genes may reflect species differences in the distribution of isoacceptor t-RNA species. In most of the published work, the distribution of t-RNA isoacceptors is not known.

(b) 5' Untranslated Region

The promotor regions of eukaryotic genes are composed generally of at least three separate consensus blocks: the capping site for mRNA, the TATA box which fixes the site for initiation of transcription and the CAAT box which is required for efficient initiation of transcription (Nevins, 1983). Sequence comparison of the 5' flanking region (PstI to SalI) of amylase-II with sequences from other eukaryotes revealed homology at the regions of interest. At 191 bp from the putative start codon (or 254 bp upstream from the ALA-1 codon of the mature protein) a CAAT box-like sequence is found (Breathnach and Chambon, 1981). At 90 bp downstream from the CAAT box (or -100 bp from first ATG) is located a TATAAATA, Goldberg-Hogness box (Gannon et al., 1979; Kozak, 1984). TATA

boxes are usually located 30 bp or so upstream from the start of transcription and 5' untranslated leaders are 20 to 80 nucleotides in length (Nevins, 1983; Kozak, 1984). This agrees well with the distance between the TATA box found here and the position of the putative initiator codon.

There are a number of examples of highly transcribed genes of lower eukaryotes in which transcription starts at the second A in the sequence G/C A A C/G (Nevins, 1983). The first one of these sequences is found in the 5' flanking region of amylase-II 31 nucleotides downstream from the TATA box; three nucleotides downstream is a second one and eight nucleotides further is a third one. The third CAAG is positioned approximately mid-way between the TATAAAT and the start codon and is immediately upstream from the pyrimidine tract CTCTCCCTTCTCT which is similar to the dodecamer pyrimidine track CTCCTCCCTTT found in close association with the presumed positions of transcription initiation in both the TrpC and glucoamylase genes (Mullaney *et al.*, 1985). Multiple initiation sites have been reported for other Aspergillus genes (Mullaney *et al.*, 1985; Nunberg *et al.*, 1984).

The actual site(s) for transcription were not mapped in this study. In the initial characterization of the genomic clones by restriction enzyme digestion and DNA-hybridization analysis (Fig. 15), the data indicated that the 5' flanking sequences of the two genes (EcoRI to Pst I) were highly homologous if not identical. A preliminary comparison of the nucleotide sequence of amylase I with that obtained from amylase II indicates that they are identical (M.Dove, M.J.Gines and V.Seligy unpublished) which is in contrast to the results obtained for the 3' untranslated regions. Without a full-length cDNA copy of each gene

it may be impossible to assign any differences in the putative transcription initiation sites used by these two genes.

(c) Intron Structures

Amylase-II contains eight introns or intervening sequences (IVS), all located within the coding region. Correlation of the genomic sequence data with the sequences of the cDNA clones suggests that there are no introns located in the 3' untranslated regions of either gene. This is the highest number of IVS recorded in a fungal gene to date, the numbers are usually 2 to 5 (MacKay *et al.*, 1985; McKnight *et al.*, 1986). However, the actual sizes of these introns (55-86 bp) are in good agreement with those of other examples (MacKay *et al.*, 1985; McKnight *et al.*, 1986; Gilbert *et al.*, 1986).

Comparison of intron distribution with respect to exon regions of amylase genes from higher organisms was not made here but certainly will be after the DNA sequence of amylase I is obtained (M.Dove, M.J.Gines and V.Seligy in progress) and when ~~more~~ genomic sequence data of other amylases genes becomes available. It is interesting to note, however, that the human salivary gland amylase gene is about 10 kbp in length and contains 10 introns which divide the coding region into 11 exons, from 100 to 231 bp in size (Nishide *et al.*, 1986). The introns are generally much larger than those of the amylase-II gene, 94 to 2,400 bp. Eleven exons have been predicted by r-looping analysis of mouse salivary gland amylase gene (Schibler *et al.*, 1982). There is some indication that the sites for interruption of exons may be conserved between mouse and man. Several rat genomic sequences have been detected by use of a cloned amylase cDNA sequence (MacDonald *et al.*, 1980). The cDNA homology is distributed along a 9 kbp fragment and is interrupted by at least seven

introns. The sizes of the introns (300 to 1,700 bp) and exons (140 to 250 bp) are variable.

Analysis of the eight IVS sequences revealed the presence of consensus sequences for 5' and 3' intron/exon junctions (Mount, 1982; Guthrie et al. 1986) as well as internal intron sequences (ICS), resembling the PuCTPuAC consensus for filamentous Ascomycetes (Guthrie et al., 1986). Examination of 5' junctions, seems to indicate three levels of sequence conservation: three residues, the GT at the 5' junction point and the G in the fifth position of the consensus (G T Pu A G T) are invariant in all except G (position 5) in IVS-2. The third and sixth positions also show high conservation and the least is the fourth position, which is in good agreement with earlier analyses. The trinucleotide 3' junction sequence seems to be highly conserved, since all but one (IVS-7) have T/C A G.

None of the introns contain the full TACTAAC required for S.cerevisiae mRNA precursor splicing (Langford et al. 1984). Further search for the internal sequence caGCTGACTg found in both introns of the exo-cellobiohydrolase gene CBH-I from T.reesei and in three of the four introns of glucoamylase from A.niger (Shoemaker et al., 1983; Boel et al., 1984) revealed that only IVS-1, 3 and 4 contained similar sequence, ACTPuACTT. Boel and co-workers (1984) suggested that in the A.niger glucoamylase gene two different splicing mechanisms would be operative, one using the TACTAAC sequence would involve a self-splicing mechanism, the other would make use of a U₂-like snRNA. However, recent evidence indicates that yeast also have snRNA and therefore fungal introns with sequences related to TACTAAC such as PuCTPuAC might also involve snRNA (U₂-like) in splicing (Guthrie et al., 1986). The finding of three cDNA clones in which the 5' ends correspond exactly to the 3' side of the 3'

intron/exon junction of IVS-3, IVS-6 and IVS-7 indicates the possibility that RNA intermediates have been cloned. The precise ending to these RNA templates can not be readily explained by current splicing models. It is also still too early to conclude whether it is the ICS or the caGCTGACTg sequence that is facilitating the splicing.

Another interesting feature of the amylase-II gene is the presence of more than one ICS in IVS-4 and IVS-8. For IVS-4, both ICSs are identical, one is 15 bp upstream from the 3' junction and the other is 45 bp further away. The two ICSs of IVS-8 differ by one base in homology and the relative distances to the 3' junction are 14 and 27 bp respectively. If sequence difference is not a major consideration for splicing, the ICS closest to the 3' junction would be preferred. Cellini *et al.* (1986) have noted the presence of TACTAAC-like sequences in the vicinity of the TACTAAC box in yeast. When the TACTAAC box is deleted the secondary TACTAAC sequences are functional but at a lower efficiency, implying that they may be functional at all times *in vivo*. This may be also the case for the amylase II IVS-4 and 8 ICSs and might even represent some mode of regulation of expression.

(d) 3' Untranslated Region

The stop codon TGA marks the beginning of the 3'UTR. The comparative 3' UTR sequence analysis of all pertinent cDNAs as well as the amylase-I and amylase-II genomics indicate several interesting points.

First, the sequences of amylases I and II are distinct enough that they could be used to distinguish their respective cDNAs, implying that both genes were transcribed at 10 hours of growth on starch. However, the heterogeneity of the sequence at which the oligo.dT primer initiated cDNA

synthesis suggests that there is considerable heterogeneity in the actual site at which polyadenylation probably occurs.

Second, only the amylase I gene contains a typical AATAAA, a general feature of eukaryotic mRNA which is located about 20 nucleotides or more upstream of the actual polyadenylation site. In fungi and yeast several exceptions to this rule have been found (Dons *et al.*, 1984; Nunberg *et al.* 1984; Shoemaker *et al.*, 1983; Zaret and Sherman, 1982). The analysis presented on the next page indicates that two types of sites for polyA termini were found for each gene. The TAGT and TATGT which also has been postulated to have a role in transcription termination and polyadenylation in some but not all genes of yeast (Zaret and Sherman, 1982) and the filamentous fungi (Nunberg *et al.*, 1984) are both present in amylase-II but TAGT and a TATGatgTATGgt was found in amylase-I. The relative distance of these signals from each other in reference to the actual polyadenylation site is relatively constant except for type 2 of amylase-I.

3'-UTR		Nucleotides from 3' end of signal polyA site			Nucleotides between signals
		TAGT	TATGT	AATAAA	
amylase-I	type 1	36		3	27
	type 2	19	37-43		14-20
Amylase-II	type 1	35	4		26
	type 2	45	13		27

The putative polyadenylation site predicted from comparison of sequences of the cDNA primed with oligo.dT and genomic DNA may be

questionable for the type-I amylase-I results. It is possible that the 9-mer polyA tail may have arisen from the oligo.dT priming off the polyA track, 3 bases downstream from the putative AATAAA signal sequence.

A search for the CAYUG consensus sequences (CAATG, CATTG, CATTC) which are also proposed to play a role in polyadenylation (Mullaney et al. 1985) was made but these sequences were not found.

Third, the length of the 3' UTRs of the two genes albeit heterogeneous are generally 5 to 6 times larger than those reported for the amylase mRNAs from vertebrate and invertebrate sources. These lengths are 33 and 35 bases for mouse salivary and pancreatic amylase (Hagenbuchle et al. 1980), 27 and 32 bases for the human versions (Nakamura et al., 1984) and 86 for fruit fly (Boer and Hickey, 1986). It has been postulated by Scheele et al. (1980) that mRNAs for secretory proteins that are synthesized in large quantities may have some advantage in either differential transcription or translation by having short 3' trailer sequences. Comparison with other, highly expressed fungal glycosidases indicates that the 156 base amylase-I and the 263 base amylase-II 3' UTRs are not unusual (Boel et al., 1984; Shoemaker et al. 1983; Yamashita et al., 1985). Other than the high AU content, 62% for amylases I and II, 65% and 58% for S.diastaticus and A.niger glucoamylase respectively, very little homology exists between the 3' UTRs of these genes.

D. Significance of the Amylase Cloning Study

Amylases of the α 1,4 type are broadly represented in nature and extensive comparisons of both amino acid sequence and to a much lesser extent, nucleotide sequence, have been carried out (MacKay et al., 1985; Nakajima et al. 1986). This is the first report of the cloning and

complete sequence of any fungal α -amylase gene. The identification of two functional copies of the gene in A.oryzae indicates that the amylase gene duplications reported for human (Nakamura et al., 1984), mouse (Schibler et al., 1982), rat (MacDonald et al., 1980) and fruit flies (Gemmell et al. 1985; Boer and Hickey, 1986) may have occurred quite early in amylase evolution. However, the amount of sequence divergence detected between amylase-I and -II appears to be quite low but some differences may arise after the full sequence of amylase-I is completed. The orientation of the two amylase genes with respect to each other is also not known at this time. However, the genomic cloning and mapping studies suggest that these genes are at least 6 kbp apart.

The major differences in sequence, so far detected, are outside the coding region, near the sites for polyadenylation, instead of in the 5' flanking region as noted in other studies (Boer and Hickey, 1986). The high conservation of the 5' flanking regions of the two genes and the presence of clusters of minor repeats 200 bases, and more, upstream suggest that these sequences may be involved in regulation of amylase gene transcription. Since in the mouse salivary-liver system there is only one amylase gene which is expressed by two different promoters in these tissues (Schibler et al., 1983), it is possible that amylase-I and -II could be still differentially regulated at some other stage of the life cycle as it has been suggested for one of the glucoamylase genes in yeast (Erratt and Nasim, 1986). The present expression results and those obtained earlier by Erratt et al. (1984) establish that the expression of the two A.oryzae amylases are regulated largely at the level of enhanced mRNA synthesis and degradation. What role glucose repression plays on regulation of these genes is not clear. The preliminary observation that

amylase mRNA is present in RNA from 10 hour glucose cultures which gives little or no in vitro translation product (Erratt et al. 1984) is interesting. The relative sizes of amylase mRNA from both glucose and starch grown mycelia are about the same, albeit it is not known to what extent processing of the introns has taken place in these two mRNA populations. The use of various sequence components of the cloned amylase-II gene should provide insight in determining what role mRNA processing might play in expression and regulation. Harding and Rutter (1978) observed that differences in amylase levels in pancreas, parotid and liver paralleled abundance of amylase mRNA in these tissues. More recently Benkel and Hickey (1986) demonstrated glucose repression occurs at a pretranslational stage in *Drosophila*. Further studies on *A.oryzae* amylase gene transcription must await 5' mapping data and the use of 3'-specific probes for the two respective genes.

From a functional point of view, only one gene was expected, since one amylase sequence was ever obtained from taka-diastase preparations (Toda et al. 1982). However, two different amylase enzymes were reported for *A.niger* (Minoda et al. 1969), one was acid stable and the other one, which was acid labile, was similar to the *A.oryzae* taka-amylase. The acid stable amylase is 5 to 6 times lower than the acid-labile enzyme. It will be useful to study the physico-chemical properties of the products of the two cloned amylases by expressing each gene separately from the other. The prospect of actually carrying out site directed mutation to test the predictions of catalytic and substrate and calcium binding site is also possible with the cloning of the amylase-II gene. When the cDNA of the amylase-I gene was used as a probe for similar gene sequences in *A.niger* DNA, it appeared that only one copy exists, which contradicts the earlier report of two different enzymes. Under the hybridization conditions used

both amylase-I and II would have been detected. Gene dosage could explain the higher production of amylase by A.oryzae in comparison to A.niger. Recently Yamashita et al. (1985) reported the cloning of an amylase gene from Saccharomycopsis fibuligera by direct expression in S.cerevisiae. The structure of this gene is unknown at this time but the DNA hybridization analysis seems to indicate only one copy. It remains to be tested whether other strains of A.oryzae and A.niger also have the same copies of the amylase gene.

REFERENCES

- Akabori, S. and Ikenaka, T. (1955) Studies of Taka-amylase A. IV. On the amino terminal amino acid of Taka-amylase A. J. Biochem. 42 603-607.
- Amorese, D. (1986) Nonradioactive detection of RNA using the BlotGENETM system. Focus 8:4, 6-8.
- Angelova, M., Grigorov, I. and Nikolova, N. (1982) Effect of glucose on the biosynthesis of alpha-amylase induced by the strain Aspergillus oryzae 3000. Acta Microbiol. Bulg. 6, 52-58.
- Aviv, H. and Leder, P. (1972) Purification of biologically active globin messenger RNA by chromatography on oligo thymidilic acid-cellulose. Proc. Nat. Acad. Sci. 69, 1408-12.
- Benkel, B.F. and Hickey, D.A. (1986) Glucose repression of amylase gene expression in Drosophila melanogaster. Genetics 114, 137-144.
- Bennet, J.W. (1985) Prospects for a molecular mycology. in Gene manipulations in fungi, p 515, 527. J.W. Bennet and L.L. Lasure (eds). Academic Press.
- Bennetzen, J.L. and Hall, B.D. (1982) Codon selection in yeast. J. Biol. Chem. 257, 3026-3031.
- Benton, W.D. and Davies, R.W. (1977) Screening λ -gt recombinant clones by hybridization to single plaques in situ. Science 196, 180-182.
- Berger, S.L. (1984) The use of Cerenkov radiation for monitoring reactions performed in minute volumes: examples from recombinant DNA technology. Anal. Biochem. 136, 515-19.
- Biggin, M.D., Gibson, T.J. and Hong, G.F. (1983) Buffer gradients gels and ³⁵S label as an aid to rapid DNA sequence determination. Proc. Natl. Acad. Sci. USA 80, 3963-65.
- Birnboim, H.C. and Doly, J. (1979) A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Res. 7, 1513-1523.
- Boel, E., Hjort, I., Svensson, B., Norris, F., Norris, K.E. and Fiil, N.P. (1984)a. Glucoamylases G1 and G2 from Aspergillus niger are synthesized from two different but closely related mRNAs. The EMBO J. 3, 1097-1102.
- Boel, E., Hansen, M.T., Hjort, I., Hoegh, I. and Fiil, N.P. (1984)b. Two different types of intervening sequences in the glucoamylase gene from Aspergillus niger. EMBO J. 3, 1581-1585.
- Boer, P.H. and Hickey, D.A. (1986) An alpha amylase gene in Drosophila melanogaster: nucleotide sequence, gene structure and expression motifs. Nucleic Acids Res. 14 8399-8411.

- Boguslawski, G. (1985) Yeast transformation in Gene Manipulations in Fungi. p 161-195. J.W. Bennett and L.L. Lasure (eds) Academic Press. Orlando.
- Breatnach, R. and Chambon, P. (1981) Organization and Expression of eukaryotic split genes coding for proteins. *Ann. Rev. Biochem.* 50, 349-83.
- Burke, J.F. (1984) High-sensitivity S1 mapping with single-stranded [32P] DNA probes synthesized from bacteriophage M13mp templates. *Gene* 30, 63-63.
- Cellini, A., Parker, R., McMahon, J., Guthrie, C. and Rossi, I. (1986) Activation of a cryptic TACTAAC box in the Saccharomyces cerevisiae actin intron. *Mol. Cell. Biol.* 6, 1571-1578.
- Chandler, P.M. (1982) The use of single-stranded phage DNA in hybrid arrest and release translation. *Anal. Biochem.* 127, 9-16.
- Chen, M.W., Jozef, A., Volckaert, G., Huysmans, E., Vandenberghe, A. and De Wachter, R. (1984) The nucleotide sequences of the 5'S rRNAs of seven molds and a yeast and their use in studying ascomycete phylogeny. *Nucleic Acids Res.* 12, 4881-4892.
- Chirgwin, J.M., Przybyla, A.E., MacDonald, R.J., Rutter, W.J. (1979) Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. *Biochem.* 19, 5294-99.
- Clarke, A.J. and Svensson, B. (1984) Identification of an essential tryptophanyl residue in the primary structure of glucoamylase G2 from Aspergillus niger. *Carlsberg Res. Commun.* 49, 559-566.
- Clarke, A.J. and Yaguchi, M. (1985) The role of carboxyl groups in the function of endo- β -1,4-glucanase from Schizophyllum commune. *Europ. J. Biochem.* 149, 233-238.
- Crerar, M.M., Swain, W.F., Pictet, R.L., Nikovits, W. and Rutter, W.J. (1983) Isolation and characterization of a rat amylase gene family. *J. Biol. Chem.* 258, 1311-1317.
- Cohen, S.N. (1982) Gene expression in heterospecific hosts, in Genetic Engineering to Biotechnology, W.J. Whelan and S. Black (eds), John Wiley, (London, New York).
- Craik, C.S., Largman, C., Fletcher, T., Rocznia, S., Barr, P.J., Fletterick, R. and Rutter, W.J. (1985) Redesigning trypsin: alteration of substrate specificity. *Science*, 228, 291-297.
- Dasinger, B.L., Fenton, D.M., Nelson, R.P., Roberts, F.F., Truesdell, S.J. (1985) Enzymic and microbial processes in the conversion of carbohydrates derived from starch. in Starch Conversion Technology. p 237-262. G.M.A. Van Beynum and J.A. Roels (eds). Marcel Dekker Inc.

- Davis, R.W., Botstein, D. and Roth, J.R. (1980) A manual for Genetic Engineering. Advanced Bacterial Genetics. CSH. New York. Cold Spring Harbor
- Denhardt, D.T. (1966) A membrane-filter technique for the detection of complementary DNA. *Biochem. Biophys. Res. Commun.* 23, 641-646
- Dons, J.J., Mulder, G.H., Rouwendal, G.H., Springer, J., Bermer, W. and Wessels, J.G. (1984) Sequence analysis of a split gene involved in fruiting from the fungus Schizophyllum commune. *The EMBO J.* 3, 2101-2106.
- Drake, J.W. (1969) Comparative rates of spontaneous mutation. *Nature* 221, 1132
- Dretzen, G., Bellard, M., Sassone-Corsi, P. and Chambon, P. (1981) A reliable method for the recovery of DNA fragments from agarose and acrylamide gels. *Anal. Biochem.* 112:295-298
- Erratt, S.A., Douglas, P.E., Moranelli, F. and Seligy, V.L. (1984) The induction of alpha amylase by starch in Aspergillus oryzae: evidence for controlled mRNA expression. *J. Biochem. Cell. Biol.* 62:678-690
- Erratt, J.A. and Nasim, A. (1986) Cloning and expression of a Saccharomyces diastaticus glucoamylase gene in Saccharomyces cerevisiae and Schizosaccharomyces pombe. *J. Bacteriol.* 166, 484-490.
- Fogarty, W.M. and Kelly, C.T. (1980) Amylases, amyloglucosidases and related glucanases. *Microbial Enzymes and Bioconversions*. 5:115-170.
- Friedmann, T. and Epstein, C.J. (1967) The role of calcium in the reactivation of reduced Taka-amylase. *J. Biol. Chem.* 242:5131-5140
- Gannon, F., O'Hare, K., Perrin, F., LePennec, J.P., Benoist, C., Cochet, M., Breathnach, R., Royal, A., Garapin, A., Cami, B. and Chambon, P. (1979) Organisation and sequences at the 5' end of a cloned complete ovalbumin gene. *Nature* 278, 428-434.
- Gemmell, R.M., Levy, J.N. and Doane, W.W. (1985) Molecular cloning of α -amylase genes from Drosophila melanogaster I. clone isolation by use of a mouse probe. *Genetics* 110, 299-312.
- Gerard, G.F. and Miller, K. (1986) Comparison of glyoxal and formaldehyde gels for sizing rRNAs. *Focus* 8:3, 5-6
- Gilbert, W., Marchionni, M. and McKnight, G. (1986) On the antiquity of introns. *Cell* 46, 151-154.
- Gillen, M.F., Banville, D., Rutledge, R.G., Narang, S., Seligy, V.L., Whitfield, J.F. and MacManus, P. (1986) The synthesis of a complete complementary DNA for the oncodevelopmental-binding protein, oncomodulin. *J. Biol. Chem.* in press.
- Glisin, V., Crkvenjkov, R. and Byus, C. (1974) Ribonucleic acid isolation by cesium chloride centrifugation. *Biochem.* 13, 2633-2637

- Gopinathan, K.P., Weymouth, L.A., Kunkel, T.A. and Loeb, L.A. (1979) Mutagenesis in vitro by DNA polymerase from an RNA tumour virus. *Nature* 278, 857-859.
- Grunstein, M. and Wallis, J. (1979) Colony hybridization. *Methods Enzymol.* 68, 379-389
- Gubler, U. and Hoffman, B.J. (1983) A simple and very efficient method for generating cDNA libraries. *Gene* 25, 263-269
- Guilbot, A. and Mercier, C. (1983) Starch. *The Polysaccharides* Vol.3, p.209-282. G.O. Aspinall (ed). Academic Press Inc. Orlando
- Guthrie, C., Riedel, N., Parker, R., Swerdlow, H. and Paterson, B. (1986) Genetic of snRNAs and RNA processing in yeast. *UCLA Symp. Mol. Cell. Biol.* L33, 301-321.
- Hagenbuchle, O., Bovey, R. and Young, R. (1980). Tissue-specific expression of mouse α -amylase genes: nucleotide sequence of isoenzyme mRNAs from pancreas and salivary gland. *Cell* 21, 179-187.
- Hagenbluchle, O., Welauer, P.K., Cribbs, D.L, Schibler, U. (1984) Termination of transcription in the mouse alpha-amylase gene Amy-2a occurs at multiple sites downstream the polyadenylation site. *Cell* 38, 737-744.
- Hanahan, D. (1983) Studies on transformation of E. coli with plasmids. *J.Mol.Biol.* 166, 557-580
- Hanahan, D. and Meselson, M. (1980) Plasmid screening at high colony density. *Gene* 10, 63-67
- Harding, J.D. and Rutter, W.J. (1978) Rat pancreatic amylase mRNA. Tissue specificity and accumulation during embryonic development. *J. Biol. Chem.* 253, 8736-8740.
- Hawkes, R., Niddy, E. and Gordon, J. (1982) A dot-immunobinding assay for monoclonal and other antibodies. *Analytical Biochem.* 119, 142-147.
- Hohn, B. and Murray, K. (1977) Packaging recombinant DNA molecules into bacteriophage particles in vitro. *Proc.Natl.Acad.Sci. USA* 74, 3259-3266.
- Hohn, B. (1979) In vitro packaging of λ and cosmid DNA. *Methods enzymol.* 68, 299-309.
- Hu, N.T. and Messing, J. (1982) The making of strand specific M13 probes. *Gene* 17, 271-277.
- Ihara, H., Sasaki, T., Tsuboi, A., Yamagata, H., Tsukagoshi, N. and Uda, S. (1985) Complete nucleotide sequence of a thermophilic alpha-amylase gene: homology between prokaryotic and eukaryotic alpha-amylases at the active sites. *J.Biochem.* 98, 95-103

- Innis, M.A., Holland, M.J., McCabe, P.C., Cole, G.E., Wittman, V.P., Tal, R., Watt, W.K., Gelfand, D.H., Holland, J.P. and Meade, J.H. (1985) Expression, glycosylation, and secretion of an Aspergillus glucoamylase by Saccharomyces cerevisiae. Science 228:21-26.
- Isemura, T., Takagi, T., Maeda, Y. and Yutan, K. (1963) Recovery of the intact structure of Taka-amylase A after reduction of all disulfide linkages in 8M urea. J.Biochem. 53, 155-161
- Kaufer, N.F., Simanis, V. and Nurse, P. (1985) Fission yeast Shizosaccharomyces pombe correctly excises a mammalian RNA transcript intervening sequence. Nature 318, 78-80.
- Kinnaird, J.H., Keighren, M.A., Kinsey, J.A., Eton, M. and Fincham, J.R.S. (1982) Cloning of the am (glutamate dehydrogenase) gene of Neurospora crassa through the use of a synthetic DNA probe. Gene 20, 387-396.
- Kluh, I. (1981) Amino acid sequence of hog pancreatic α -amylase isoenzyme I. FEBS 136:231-234.
- Kogane, F. and Yanagita, T. (1964) Isolation and purification of deoxyribonucleic acid from Aspergillus oryzae conidia. J. Gen. Appl. Microbiol. 10, 61-68
- Kozak, M. (1983) Comparison of initiation of protein synthesis in prokaryotes, eukaryotes and organelles. Microbiol. Rev 47, 1-45.
- Kozak, M. (1984) Compilation and analysis of sequences upstream from the translational start site in eukaryotic mRNAs. Nucleic Acids Res. 12, 857-863
- Laemmli, U.K. (1970) Cleavage of structural proteins during the Assembly of the Head of Bacteriophage T4. Nature 227, 680-685
- Langford, C. and Gallwitz, D. (1983) Evidence for an intron-contained sequence required for the splicing of yeast RNA polymerase II transcripts. Cell 33, 519-527.
- Langford, C.J., Klinz, F.J., Donath, C. and Gallwitz, D. (1984) Point mutations identify the conserved, intron-contained TACTAAC box as an essential splicing signal sequence in yeast. Cell 36, 645-653.
- Law, D.J. and Timberlake, W.E. (1980) Developmental regulation of Lacase levels in Aspergillus nidulans. J.Bacteriol. 144, 509-517
- Levy, J.N., Gemmill, R.M. and Doane, W.W. (1985) Molecular cloning of α -amylase genes from Drosophila melanogaster. II Clone organization and verification. Genetics 110, 313-324.
- MacDonald, R.J., Crefar, M.M., Swain, W.F., Picot, R.L., Thomas, G. and Rutter, W.J. (1980) Structure of a family of rat amylase genes. Nature 287, 117-122

- MacDonald, R.J., Stary, S.J. and Swift, G.H. (1982) Two similar, but nonallelic rat pancreatic trypsinogens. *J. Biol. Chem.* 257, 9724-9732.
- Mackay, R.M., Baird, S., Dove, M.J., Errat, J.A., Gines, M.J., Moranelli, F., Nasim, A., Willick, G.E., Yaguchi, M. and Seligy, V.L. (1985) Glucanase gene diversity in prokaryotic and eukaryotic organisms. *Biosystems* 18, 279-292.
- Mackay, R.M., Lo, A., Willick, G., Zuker, M., Baird, S., Dove, M., Moranelli, F. and Seligy, V. (1986) Structure of a Bacillus subtilis endo- β -1,4 glucanase gene. *Nucleic Acids Res.* 14, 9159-9170.
- McKnight, G.L., O'Hara, P.J. and Parker, M.L. (1986) Nucleotide sequence of the triphosphate isomerase gene from Aspergillus nidulans; Implications for a differential loss of introns. *Cell* 46, 143-147.
- Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982). Molecular Cloning C.S.H. Laboratory.
- Matheson, N.K. and McCleary, B.V. (1985) Enzymes metabolizing polysaccharides and their application to the analysis of structure and junction of glycans. *The Polysaccharides*. Vol. 3, p 1-100. G.O. Aspinall (ed) Academic Press Inc.
- Matsuura, Y., Kusunoki, M., Date, W., Harada, S., Bando, S., Tanaka, N. and Kakudo, M. (1979) Low resolution crystal structures of Taka-amylase and its complexes with inhibitors. *J. Biochem.* 86, 1773-1783
- Matsuura, Y., Kusunoki, M., Harada, W., Tanaka, N., Iga, Y., Yasuoka, N., Toda, H., Narita, K. and Kakudo, M. (1980) Molecular structure of Taka-amylase A. I. Backbone chain folding at 3A Resolution. *J. Biochem.* 87, 1555-1558.
- Matsuura, Y., Kusunoki, M., Harada, W. and Kakudo, M. (1984) Structure and catalytic residues of Taka-amylase. A. *J. Biochem.* 95, 697-702.
- Mavridis, I.M., Hatada, M.H., Tulinsky, A. and Leboide, L. (1982) Structure of 2-keto-3-deoxy-6-phosphogluconate aldolase at 2.8 angstroms resolution. *J. Mol. Biol.* 162, 419-444
- Messing, J. (1983) New M13 vectors for cloning. *Methods in Enzymology* 101, 20-78
- Messing, J., Crea, R., Seeburg, P.H. (1981) A system for shotgun DNA sequencing. *Nucl. Acids Res.* 9, 309-321
- Messing, J. and Vieira, J. (1982) A new pair of M13 vectors for selecting either DNA strand of double-digest restriction fragments. *Gene* 19 269-276.
- Meyrath, J., Bahn, M., Han, H.E. and Altman, H. (1971) Induction of amylase producing mutants in Aspergillus oryzae by different irradiations. *Radiat. Radioisot. Ind. Microorg. Proc. Symp.* 1971, 137-155.

- Miller, J.H. (1972) Experiments in Molecular Genetics. Cold Spring Harbor, New York.
- Minoda, Y., Arai, M., Torigoe, Y. and Yamada, K. (1968) Acid-stable alpha-amylase of black Aspergilli. Part III. Separation of acid-stable alpha-amylase and acid-unstable alpha-amylase from the same mold amylase preparation. Agr. Biol. Chem. 32, 110-113.
- Moranelli, F., Barbier, J.R., Dove, M.J., Mackay, R.M., Seligy, V.L., Yaguchi, M. and Willick, G.E. (1986) A clone coding for Schizophyllum commune β -glucosidase: homology with a yeast β -glucosidase. Biochem. Internatl. 12, 905-912.
- Morosoli, R., Roy, C. and Yaguchi, M. (1986) Isolation and partial sequence of a xylanase from the yeast Cryptococcus albidus. Biochim. Biophys. Acta 870, 473-478.
- Mount, S.M. (1982) A catalogue of splice junction sequences. Nucleic Acids Res. 10, 459-472.
- Mullaney, E.J., Hamer, J.E., Roberti, K.A., Yelton, M. and Timberlake, W.E. (1985) Primary structure of the trpC gene from Aspergillus nidulans. Mol. Gen. Genet. 199, 37-45.
- Nakajima, R., Imanaka, T. and Aiba, S. (1985) Nucleotide sequence of the Bacillus stearothermophilus α -amylase gene. J. Bacteriol. 162, 401-406.
- Nakajima, R., Imanaka, T. and Aiba, S. (1986) Comparison of amino acid sequences of eleven different alpha-amylases. Appl. Microbiol. Biotechnol. 23, 355-360.
- Nakamura, Y., Ogawa, M., Nishide, T., Emi, M., Kosagi, C., Himeno, S. and Matsubara, K. (1984) Sequences of cDNA for human salivary and pancreatic α -amylases. Gene 28, 263-70.
- Narita, K. (1975) Amino acid sequence of the C-terminal sixty-six residues of Taka-amylase A. Proc. Japan Acad. 51, 285-290.
- Narita, K., Murakami, H. and Ikenaka, T. (1966) Reinvestigation of the amino acid composition and C-terminal group of taka-amylase A. J. Biochem. 59, 170-175.
- Nevins, J.R. (1983) The pathway of eukaryotic mRNA formation. Ann. Rev. Biochem. 52, 441-466.
- Nishide, T., Nakamura, Y., Emi, M., Yamamoto, T., Ogawa, M., Mori, T. and Matsubara, K. (1986) Primary structure of human salivary α -amylase gene. Gene 41, 299-304.
- Nunberg, J.H., Meade, J.M., Cole, G., Lawyer, F.C., McCabe, P., Scheickart, V., Tal, R., Wittman, V.P., Flatgaard, J.E. and Innis, M.A. (1984). Molecular cloning and characterization of the glucoamylase gene of Aspergillus awamori. Molecular and Cellular Biol. 4, 2306-2315.

- Oikawa, A. and Maeda, A. (1957) The role of calcium in Taka-amylase A. *J. Biochem.* 44, p 745
- Oikawa, A. (1959) The role of calcium in Taka-amylase A. II. The exchange reaction of calcium. *J. Biochem.* 46, 463-473
- Paice, M.G., Desrochers, M., Rho, D., Jurasek, L., Roy, C., Rollin, C.F., DeMiguel, E. and Yaguchi, M. (1984). Two forms of endoglucanase from the basidiomycete Schizophyllum commune and their relationship to other β -1,4-glycoside hydrolases. *Bio/technology* 2:535-539.
- Paice, M.G., D., Jurasek, L., Carpenter, M.R. and Smillie, L.B. (1978). Production, characterization, and partial amino acid sequence of xylanase A from Schizophyllum commune. *Appl. Environ. Microbiol.* 36:802-808.
- Penttila, M.E., Nevalainen, K.M.M., Raynal, A. and Knowles, J.K.C. (1984). Cloning of Aspergillus niger genes in yeast. Expression of the gene coding Aspergillus β -glucosidase. *Mol. Gen. Genet.* 194, 494-499.
- Pikielny, C.W., Teem, J.L. and Rosbash, M. (1983) Evidence for the biochemical role of an internal sequence in yeast nuclear mRNAs introns: implications for U1 RNA and metazoan mRNA splicing. *Cell* 34, 395-403.
- Reilly, P.J. (1985) Enzymic degradation of starch. in *Starch Conversion Technology*, p.101-142. G.M.A. Van Beynum and J.A. Roels (ed) Marcel Dekker Inc.
- Roghrig, K.L. (1984) Carbohydrate biochemistry and metabolism. AVI Publishing Co. Inc. Westport.
- Rogers, J.C. (1985) Conserved amino acid sequence domains in alpha-amylases from plants animals and bacteria. *Biochem. Biophys. Res. Commun.* 128, 470-476.
- Rogers, J. and Wall, R. (1980) A mechanism for RNA splicing. *Proc. Natl. Acad. Sci. USA* 77, 1877-1879.
- Rogers, J.C. and Milliman, C. (1983) Isolation and sequence analysis of a barley α -amylase cDNA clone. *J. Biol. Chem.* 258, 8169-8174.
- Rothstein, S.J., Lazarus, C.M., Smith, W.E., Baulcombe, D.C. and Gatenby, A.A. (1984) Secretion of a wheat α -amylase expressed in yeast. *Nature* 308, 662-665
- Rutledge, R. (1987) Ph.D. Thesis
- Sanger, F., Goulson, A.R., Barrell, B.G., Smith, A.J. and Roe, B. (1980). Cloning in single-stranded bacteriophage as an aid to rapid DNA sequencing. *J. Mol. Biol.* 143:161-178.
- Sanger, F., Nicklen, S., Coulson, A.R. (1977) DNA sequencing with chain terminating inhibitors.

Schechtman, M.G. and Yanofsky (1983) Structure of the trifunctional *trp-1* gene from Neurospora crassa and its aberrant expression in *Escherichia coli*. *J. Mol. Appl. Genet.* 2:83-99.

Scheele, G.A. (1980) Biosynthesis, segregation and secretion of exportable proteins by the exocrine pancreas. *Am. J. Physiol.* 238, G467-G477.

Schekman, R. and Novick P. (1982) The secretory process and yeast cell-surface assembly. in *The Molecular Biology of the yeast Saccharomyces*. Metabolism and gene expression. p. 361-393. J.N.Strathern, E.W.Jones and J.R.Broach (eds). CSH.

Schibler, U., Pittet, A.M., Young, R.A., Hagenbuchle, Tosi, M., Gellman, S. and Wellauer, P. (1982) The mouse α -amylase multigene family. *J. Mol. Biol.* 155:247-266.

Seligy, V.L., Baird, S., Bernier, R., Desrocher, M., Dove, M.J., Lo, A., MacKay, R.M. and Willick, G.E. (1986) Functional mapping and expression of a Bacillus subtilis endo-1,4- β -D-glucanase gene in Escherichia coli and the yeasts Saccharomyces cerevisiae and Schwanniomyces alluvius. in the *Biochemistry and Molecular Biology of Industrial Yeast*. Stewart, G.G., Russell, I., Klein, R.C. and Hiebsch R.R. (eds). Uniscience Series, CRC Press, Boca Raton, Fl. (in press)

Seligy, V.L., Barbier, J., Dimock, K., Dove, M.J., Moranelli, F., Morosoli, R., Willick, G.E. and Yaguchi, M. (1983). Current status in the application of recombinant DNA technology in the construction of cellulolytic yeast strains. pp 167-185. In M. Korhola and E. Vaisanen (ed). *Gene expression in yeast*. Proc. of the Alko Yeast Symp. Helsinki.

Seligy, V.L., Miki, B.L., Thomas, D.V. and Poon, N.H. (1980) Chromatin structure and intracellular location of the yeast plasmid *Scp* or *Z* um DNA. In *Current Developments in yeast research (ISY-5)* G.G. Stewart and I. Russell (eds) pp 165-70.

Shoemaker, S.P., Gelfand, D.H., Kwok, S., Schweickart, V., Watt, K., van Arsdell and Innis, M.A. (1984) Expression of *Trichoderma reesei* L27 cellulases in Saccharomyces cerevisiae *Am. Chem. Soc., Natl. Meet.*, 188th. (abstr.) 71

Shoemaker, S., Schweickart, V., Ladner, M., Gelfand, D., Kwok, S., Myambo, K. and Innis, M. (1983). Molecular cloning of exo-cellobiohydrolase I derived from Trichoderma reesei strain L27. *Bio/Technology* 1:691-696.

Sills, A.M., Sauder, M.E., Stewart, G.G. (1984) Amylase activity in certain yeasts and a fungal species. *Devel. Industrial Microbiol.* 24, 295-303

Slightom, J.L., Blechl, A.E., Smithies, O. (1980) Human fetal G γ - and A-globin genes: Complete nucleotide sequences suggest that DNA can be exchanged between these duplicated genes. *Cell* 21:627-638

- Smith, B.W. and Roe, J.H. (1949) A photometric method for the determination of α -amylase in blood and urine with use of the starch iodine color. J. Biol. Chem. 179, 53-59.
- Southern, E.M. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. 98, 503-517
- St. John's, T.P. and Davis, R.W. (1981) The organization and transcription of the galactose gene cluster of Saccharomyces. J. Mol. Biol. 152, 285-315.
- Stewart, G.G., Panchal, C.J., Russell, I. and Sills, A.M. (1982) Advances in ethanol from sugars and starches, p.4-57. in H.E. Duckworth and E.A. Thompson (ed.) International Symposium on Ethanol from Biomass. The Royal Society of Canada, Toronto
- Swinkels, J.J.M. (1985) Sources of starch, its chemistry and physics. in starch conversion technology p 15-46. G.M.A. Van Beynum and J.A. Roels (eds) Marcel Dekker Inc. N.Y.
- Tagaki, T. and Isemura, T. (1965) Necessity of calcium for the renaturation of reduced taka-amylase A. J. Biochem. 57, 89-95.
- Tagaki, T. and Isemura, T. (1962) Studies on the denaturation of Taka-amylase A and on its reversibility. II urea denaturation and its reversal by removal or dilution of urea. J. Biochem. 52, 314-323.
- Tagaki, T. and Toda, H. (1962) Studies of the denaturation of Taka-amylase A and on its reversibility. I. Acid-denaturation and its reversibility. J. Biochem. 52, 16-27.
- Takamine, J. (1894) United States Patents, 525, 820 and 525, 823. as referred to by J.W. Bennet (1985) in Molecular Genetics of Filamentous Fungi, p.345-366. W.E. Timberlake (ed). Alan R. Liss, Inc
- Takkinen, K., Pettersson, R.F., Kalkkinnen, N., Palva, I., Soderlund, H. and Kaariainen, L. (1983) Amino acid sequence of α -amylase from Bacillus amyloliquefaciens deduced from the nucleotide sequence of the cloned gene. J. Bio. Chem. 258, 1007-1013.
- Taylor, G.R. (1985) Southern transfer of native DNA using nylon membranes. Anal. Biochem. 148, 524-26.
- Thomas, P.G., Russell, A.J. and Fersht, A.R. (1985) Tailoring the pH dependence of enzyme catalysis using protein engineering. Nature, 318, 375-376.
- Thomas, P.S. (1980) Hybridization of denatured RNA transferred or dotted to nitrocellulose paper. Proc. Natl. Acad. Sci. USA 77, 5201-5207
- Thomsen, K.K. (1983) Mouse α -amylase synthesized by Saccharomices cerevisiae is released into the culture medium. Carlsberg Res. Commun. 48, 545-555

- Toda, H., Kato, I. and Narita, K. (1968) Correlation of the masked sulfhydryl group with the essential calcium in Taka-amylase A. *J. Biochem.* 63, 295-301.
- Toda, H., Kondo, K. and Narita, K. (1980) N-terminal amino acid sequence of Taka-amylase A from Aspergillus oryzae. *Agric. Biol. Chem.* 44, 1945-1947.
- Toda, H., Kondo, K. and Narita, K. (1982) The complete amino acid sequence of Taka-amylase A. *Proc. Japan Acad.*, 58 (ser B):208-212.
- Tosi, M., Young, R.A., Hagenbuchle, O. and Schibler, U. (1981) Multiple adenylation sites in a mouse α -amylase gene. *Nucleic Acids Res.* 9, 2313-2323.
- Van Beynum, G.M.A. and Roels, J.A. (1985) The multidimensionality of starch conversion technology. in *Starch Conversion Technology*, p.1-14. G.M.A. Van Beynum and J.A. Roels (eds) Marcel Dekker Inc. N.Y.
- Viera, J. and Messing, J. (1982) The pUC8 plasmids, an M13 mp7-derived system for insertion mutagenesis and sequencing with synthetic universal primers. *Gene* 19, 259-68.
- Wallace, R.B., Johnson, M.J., Hirose, T., Miyake, T., Kawashima, E., Itamura, K. (1981) The use of synthetic oligonucleotide as hybridization probes. II Hybridization of oligonucleotides of mixed sequence to rabbit B-globin DNA. *Nucleic Acids Res.* 9, 879-94.
- Weaver, L.H., Grutter, M.G., Remington, S.J., Gray, T.M., Isaacs, N.W., and Matthews, B.W. (1985). Comparison of Goose-type, chicken-type, and phage-type lysozymes illustrates the changes that occur in both amino acid sequence and three-dimensional structure during evolution. *J. Mol. Evol.* 21, 97-111.
- White, T.J., Meade, J.H., Shoemaker, S., Kotha, K.E. and Innis, M.A. (1984). Enzyme cloning for the food fermentation industry. Food Technology 83, 90-93.
- Willick, G.E., Morosoli, R., Seligy, V.L., Yaguchi, M. and Desrochers, M. (1984) Extracellular proteins secreted by the Basidiomycete, Schizophyllum commune in response to carbon source. *J. Bacteriol.* 159, 294-299.
- Willick, G.E., and Seligy, V.L. (1985) Multiplicity in cellulases of Schizophyllum commune derives in part from transcript and glycosylation heterogeneity. *Europ. J. Biochem.* 151, 89-96.
- Woods, D. (1984) Oligonucleotide screening of cDNA libraries. *Focus* 6:3, 1-3
- Woudt, F. and Planta, R.J. (1983) The genes coding for histone H3 and H4 in Neurospora crassa are unique and contain intervening sequences. *Nucleic Acids Res.* 11, 5347-5360

- Yaguchi, M., Roy, C., Rollin, C.F., Paice, M. and Jurasek, L. (1983) A fungal cellulase shows sequence homology with the active site of hen egg-white lysozyme. Biochem. Biophys. Res. Comm. 116:408-411.
- Yamaguchi, H., Ikenaka, T. and Matsushima, Y. (1971) The complete sequence of a glycopeptide obtained from taka-amylase A. J. Biochem. 10, 587-594.
- Yamashita, I., Itoh, T. and Fukui, S. (1985a) Cloning and expression of the Saccharomycopsis fibuligera glucoamylase gene in Saccharomyces cerevisiae. Appl. Microbiol. Biotechnol. 23, 130-133.
- Yamashita, I., Itoh, T. and Fukui, S. (1985b) Cloning and expression of the Saccharomycopsis fibuligera α -amylase gene in Saccharomyces cerevisiae. Agric. Biol. Chem. 49, 3089-3091.
- Yamashita, I., Suzuki, K. and Fukui, S. (1984) Nucleotide sequence of the extracellular glucoamylase gene STAl in the yeast Saccharomyces diastaticus. J. Bacteriol. 567-73.
- Yang, M., Galizzi, A., Henner, D. (1983) Nucleotide sequence of the amylase gene from Bacillus subtilis. Nucleic Acids Res. 11, 237-249.
- Yanisch-Perron, C., Vieira, J. and Messing, J. (1985) Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13-mp18 and pUC19 vectors. Gene 33, 103-19.
- Yelton, M.M., Hamer, J.E. and Timberlake, W.E. (1984) Transformation of Aspergillus nidulans by using a trpC plasmid. Proc. Natl. Acad. Sci. USA 81, 1470-74.
- Zantinge, B., Dons, H. and Wessels, G.H. (1979) Comparison of poly(A)-containing RNAs in different cell types of the lower eukaryote Schizophyllum commune. Eur. J. Biochem. 101, 251-260.
- Zaret, K.S. and Sherman, F. (1982) DNA sequence required for efficient transcription termination in yeast. Cell 28, 563-573.

END

29-106-89

FIN

(47.5%) when the sequences flanking the coding region are taken into account. The EcoRI-PstI fragment is particularly high in A. Comparison of nucleotide content with that of total A.oryzae DNA (Kogane and Yanagita, 1964) indicates that the molar ratio (1.05) is lower for the Amylase-II structural gene than for the total (1.11) which reflects a more even distribution in nucleotides. The fragment EcoRI-HindIII which includes the upstream A-rich region, is similar to total DNA. These differences may be correlated with specific codon biases for Amylase II.

The codon usage of the 1500-base Amylase II coding region which includes the 21 codons for the secretion signal and stop signal but not the introns is given in Table X. There appears to be a bias of codon usage which is similar to other Ascomyceteous glycosidases. The frequency for preference for one codon over another is high (65-89%) for Asn, Cys, Gln, Ile, Lys, Phe and Tyr. When more than two codons can be used for the same amino acid, the trend is much less pronounced. The exception is the lack of usage of AGA for Arg, while the other five combinations are in similar proportions.

(d) Comparison of 5' and 3' untranslated regions of Amylase genes I and II

The nucleotide sequences of the DNA located 5' and 3' of the coding region have been analyzed in detail and presented in Figures 17 and 19.

The nucleotide sequences of the Amylase I and II genes between EcoRI and SalI have been compared and they are virtually identical. The N-termini of both the mature protein which is the same as TAA and the precursor (Met at -21 relative to Ala-1) are located within the 341 base pair PstI-SalI fragment. The putative start signal is 63 base pairs upstream of the GCA, Ala-1 codon. It is assumed that of the two inframe MET codons, the distal most codon is the true start signal. Further