EXPRESSION OF A *CLOSTRIDIUM ACETOBUTYLICUM* ENDOGLUCANASE AND AN ENDOXYLANASE IN *SACCHAROMYCES CEREVISIAE*

ELIZABETH A. LEWIS

A dissertation submitted in partial fulfillment of the requirements for the degree of Master of Science in the Faculty of Science, University of Cape Town.

CAPE TOWN

FEBRUARY, 1989

Acknowledgements

I would like to thank my supervisors, Professor D.R. Woods for his advice, and Professor D.T. Jones for his encouragement during the course of this study.

I am very grateful to Dr. H. Zappe for his assistance and helpful advice.

My thanks to W. Jones for her assistance in the nucleotide sequencing of the α-mating factor and endoglucanase gene fusions, and to colleagues in the Microbiology Department who shared their knowledge and provided moral support.

Contents

Abstract

S. cerevisiae is widely used in industrial processes, in particular ethanol production. The aim of this study was to examine the feasability of cloning lignocellulase genes into yeast. These enzymes could possibly extend the substrate range of S. cerevisiae, thereby possibly making the process of ethanol production more cost effective.

An endoglucanase gene and an endoxylanase gene, previously isolated from C. acetobutylicum and characterised in E. coli by Zappe et al. (1986, 1987, 1988), were cloned into a variety of yeast-E. coli shuttle vectors. The manipulation of these genes was performed in E. coli. The recombinant plasmids were used to transform S. cerevisiae DBY746.

The endoglucanase gene was expressed in S. cerevisiae under the control of the yeast α-mating factor gene promoter. Active endoglucanase was produced and secreted into the growth medium by the yeast.

Neither endoglucanase nor xylanase gene expression was detected when the genes, cloned into the yeast, were under the control of the bacterial gene promoters.

Abbreviations

Ap	ampicillin
bp	base pair(s)
CsCl	caesium chloride
C-terminal	carboxy terminal end of a protein
CMC	carboxymethylcellulose
d	days
DTT	1,4-dithio-L-threitol
EDTA	ethylenediametetra-acetic acid
h	hours
IPTG	isopropyl β-D-thiogalactopyranoside
kb	kilobase pairs
LB	Luria Bertani medium
min	minute(s)
M_r	relative molecular mass
N-terminal	amino terminal end of a protein
p	plasmid
PC	citrate-phosphate buffer
pI	isoelectric point
r	resistance (superscript)
s	seconds
SDS	sodium dodecyl sulphate
Tc	tetracycline
Tris	Tris(hydroxymethyl)aminomethane
UV	utraviolet light
X-gal	5-Bromo-4-chloro-3-indolyl-β-D-galactopyranoside

Chapter 1

Introduction.

Recombinant DNA is a fully established technology enabling the isolation and manipulation of genes that encode medically and commercially important proteins. Expression of these genes in microorganisms allows for the large-scale and cost effective production of these proteins that are not produced in sufficient quantities in their natural host. The bacterium, Escherichia coli, is most commonly used in recombinant DNA technology. This organism, however, may not be the most suitable host for the expression of certain proteins, particularly those of eukaryotic origin. In addition, the mechanisms of transcription, translation and post-translational processing differ between prokaryotes and eukaryotes, such that proteins encoded by eukaryote genes in prokaryotes may lack activity or be insoluble until modified chemically, for example calf prochymosin (Emtage et al., 1983) and human interferon-γ (ifn-γ) (Simons et al., 1984).

The yeast Saccharomyces cerevisiae has been developed as an alternative eukaryotic host. Some eukaryotic proteins that were inactive when produced by E. coli, were soluble and biologically active when produced in S. cerevisiae. Mellor and co-workers (1983) found that calf chymosin was efficiently synthesised and active in yeast. Similarly, Ifn-γ was correctly processed in S. cerevisiae (Derynck et al., 1983). S. cerevisiae also has the advantage of

being non-pathogenic and widely used in the wine and food industry (Roman, 1982). Yeast can be manipulated in two ways : a) main aim is to produce a product, b) main aim is alcohol fermentation, where the addition of a gene that would improve the substrate range would not interfere or limit the alcohol fermentation.

Yeast plasmids

S. cerevisiae has been manipulated to produce foreign proteins by the development of techniques that enable the introduction of exogenous DNA into yeast cells (Beggs, 1978; Hinnen et al., 1978; Ito et al., 1983). Plasmid vectors that can replicate and be selected in E. coli as well as S. cerevisiae have been designed to enable recombinant DNA molecules to be manipulated in E. coli, and then transferred to S. cerevisiae, thereby making use of relatively simpler methodology for the faster growing E. coli. Yeast plasmids are classified according to their mode of replication or specific function in yeast (Parent et al., 1985). These vectors have been generally divided into eight groups.

Yeast integrating plasmids (YIp). These plasmids contain DNA sequences that are homologous to regions of the yeast chromosome. Transformation of yeast occurs at a relatively low frequency (1-100 colonies/μg DNA) since successful transformation requires integration of the plasmid into the host genome by homologous recombination (Parent et al., 1985). The plasmid, once integrated, becomes a heritable

component of the yeast genome and behaves as a Mendelian inheritance characteristic (Hinnen et al., 1978; Struhl et al., 1979), and is therefore highly stable. Integration of the plasmid by homologous recombination may occur at several different places in the yeast genome. Hinnen et al. (1978), using a mutant leu strain, found integration of plasmids carrying the leu2 gene (3-isopropylmalate dehydrogenase) into the leu2$^-$ region, as well as into two other locations as a result of rare meiotic recombination events.

Yeast episomal plasmids (YEp). In some yeast strains, a naturally occurring plasmid, called the 2 μm plasmid, has been isolated, and its nucleotide sequence determined by Hartley and Donelson (1980). It was found to contain two homologous inverted repeat regions (Broach, 1982). Recombination occurring in these regions results in two forms of the plasmid. As a consequence of this recombination, the 2 μm plasmids found in yeast are a mixed population of the two plasmid forms that differ in orientation of a unique region. These two forms have been designated A and B (Broach, 1982). YEp plasmids are able to autonomously replicate in yeast since they contain the entire naturally occurring 2 μm plasmid or part thereof (including the replication origin). YEp plasmids transform yeast at high frequencies (10^4 -10^5 transformants/μg DNA). In yeast they occur at high copy number (20-100 copies per cell) and exist as extrachromosomal elements, and can therefore be relatively unstable (Beggs, 1978; Gerbaud

et al., 1979; Broach et al., 1979; Parent et al., 1985). Murray and Szostak (1983) found that 2 μm derived plasmids showed no segregational bias, therefore cells in the population could contain the plasmid at high copy number. Broach et al. (1979) and Gerbaud et al. (1979) observed several low frequency rearrangements of these plasmids. The most common rearrangement was caused by insertion into an endogenous 2 μm plasmid. McNeil et al. (1980) identified the fragment of the 2 μm plasmid that is essential for high frequency transformation. Plasmids lacking the 2 μm ori transformed S. cerevisiae cir^+ strains by homologous recombination with the resident 2 μm plasmid. (Strains lacking 2 μm circles are indicated by (cir^o) while those with the wild-type 2 μm plasmid are indicated by (cir^+)).

Yeast replicating plasmids (YRp). YRp plasmids contain an autonomously replicating sequence (ARS). ARS elements from S. cerevisiae enable plasmids to be autonomously maintained in yeast cells as minichromosomes and replicate in the nucleus (Stinchcomb et al., 1979). ARSs are of chromosomal origin and probably function as replication origins on chromosomes as well as plasmids (Takagi et al., 1986; Palzkill and Newlon, 1988). These plasmids have similar properties to YEp plasmids. They transform yeast at high frequency (10^3 -10^4 colonies/ug DNA), due to the presence of these homologous or heterologous sequences that act as replication origins (Struhl et al., 1979; Parent et al., 1985). YRp plasmids are present at high copy number (20-50 copies per cell) (Hyman et al., 1982). YRp plasmids are

generally unstable and in the absence of selection, 20% of the cells lose their plasmids per generation (Kingsman et al., 1979).

Yeast centromeric plasmids (YCp). These are autonomously replicating plasmids that contain a functional yeast centromere (Parent et al., 1985). Clarke and Carbon (1980) identified a centromeric sequence (cen3) that, when present on a plasmid with an ARS element, allowed the plasmid to behave as a minichromosome in yeast. These plasmids, carrying cen3, are mitotically stable and segregate as chromosomes. The cen3 element also controls the copy number, maintaining the plasmid at approximately one copy per cell. In the absence of selection about 90% of the cells retained the plasmid after approximately 20-30 generations (Clarke and Carbon, 1980). Chlebowicz-Sledziewska and Sledziewski (1985) showed that the cen3 sequence could be regulated if the cen3 fragment was fused to a regulated promoter. Plasmids with the adh2-cen3 fusion could occur at high copy number when the alcohol dehydrogenase 2 (adh2) gene promoter interfered with cen3 functions. Repression of the adh2 promoter enabled cen3 to function normally and stabilize the plasmid. This type of plasmid is termed YCRp.

Yeast linear plasmids (YLp). These plasmids contain sequences of yeast chromosomes that behave as telomeres. Some YLp plasmids contain functional centromeres and behave as artificial chromosomes (Murray and Szostak, 1983),

although their stability is approximately a hundred fold lower than natural chromosomes (Parent et al., 1985). Additional sequences are spontaneously added to the ends of linear plasmids upon replication in yeast (Parent et al., 1985; Ascenzioni and Lipps, 1986).

Yeast promoter plasmids (YPp). These plasmids contain coding sequences for an easily assayed protein to which promoters containing transcriptional or translational signals can be fused. These plasmids are used in determining promoter functions and structure (Parent et al., 1985), and can be derived from any of the five types of plasmids described above.

Yeast expression plasmids (YXp). These plasmids contain the promoter region of a gene, usually of yeast origin. These promoters can be derived from a number of genes such as the α–mating pheromone gene (mfα) or the phosphoglycerate kinase (pgk) gene. Some plasmids also carry transcriptional terminator sequences. Homologous or heterologous gene sequences are fused between the promoter and terminator sequences for expression / excretion in yeast (Parent et al., 1985).

Yeast hybrid plasmids (YHp). These vectors contain hybrid genes composed of a variety of gene promoter, regulatory, coding and termination sequences. They are of particular interest when studying genotype expression in yeast, for example RNA processing and protein localization (Parent

et al., 1985; Emr et al., 1983). The β-galactosidase gene has been used in the construction of fusion vectors (Parent et al., 1985).

Many different types of yeast vectors are continually being constructed. These include the use of phage DNA sequences (Vernet et al., 1987) and transposons (Jacobs et al., 1988).

Gene expression in yeast.

In order to determine whether heterologous (non-yeast) genes could be expressed in S. cerevisiae, a variety of yeast plasmids were developed. Beggs et al. (1980) inserted the rabbit β-globin gene into pJDB219 (a 2 μm derived plasmid). Rabbit β-globin homologous RNA was detected. However, the RNA lacked 20-40 nucleotides from the 5′ end, contained a small intron and extended into half of the large intron where it terminated at an AT-rich region. This suggested that the heterologous promoter was not recognised correctly, and that the heterologous introns were not reliably processed. This was confirmed by Langford et al. (1983). Yeast is unable to excise foreign introns from hybrid genes either due to the splicing mechanisms or the recognition of intervening sequences not being universal. It is for this reason that most heterologous genes cloned into yeast are derived from mRNA via the cDNA cloning route. The ability of S. cerevisiae to recognize transcription promoters in heterologous DNA appears to be variable. With the rabbit β-globin gene, a defective mRNA was produced (Beggs

et al., 1980), whereas the Herpes simplex virus thymidine kinase gene was not transcribed in yeast (Kiss et al., 1982). In contrast, normal transcription was initiated at the heterologous promoter of zein maize storage protein gene (Langridge et al., 1984).

Eukaryotic and yeast gene-expression signals. In order to ensure that any heterologous gene is efficiently expressed in S. cerevisiae, it appears necessary to replace the gene's promoter with a promoter from a yeast gene. A S. cerevisiae gene promoter region, that is required to direct the initiation of transcription, may also be involved in determining the rate of transcription and regulation. A yeast gene 3' sequence is usually required for the efficient termination of the transcript. Promoters from several yeast genes were examined in order to identify the key sequences for expression. In general, it appears as if promoter regions from eukaryotic genes are more complex than bacterial promoters.

S. cerevisiae genes are obviously expressed with different efficiencies. Although no consensus yeast promoter has been defined, several features are common to genes that are expressed at similar levels and could be important in transcriptional and translational efficiency. An example of a highly expressed gene is the pgk gene (Dobson et al., 1982a) while the gene for N (5'-phosphoribosyl)-anthranilate isomerase (trp1) (Dobson et al., 1982b) is a weakly expressed gene.

A region upstream of some yeast genes has been implicated in gene regulation in S. cerevisiae. These naturally occurring AT-rich, cis-acting elements can lie hundreds of nucleotides upstream (5') from the region of transcription initiation (Guarente, 1984; Braus et al, 1988). These sequences are called upstream activation sites (UAS). The distance between the UAS and the initiation site is variable. This flexibility could be important in gene regulation (Guarente, 1984). Some genes (e.g. cyc1 gene (iso-1-cytochrome c)) have tandem UASs to respond to distinct stimuli, while other genes may have a collection of independently functioning UASs for a wide variety of regulatory responses. UASs may be positive or negative regulatory elements (Guarente, 1984).

Initiation of transcription in eukaryotes depends on the presence of the so-called "TATA box". It has been postulated that the TATA box is essential for recognition by RNA polymerase II (Zalkin and Yanofsky, 1982). In yeast the TATA box is found further upstream from the transcription initiation site than in mammalian genes (Dobson et al., 1982a). Dobson et al. (1982a) proposed two possible TATA boxes for the pgk gene, the first at position -152 and the second at -114. In the trp5 (tryptophan synthase) gene, a TATA sequence was found 37 base pairs (bp) upstream of the putative initiation site and another possible one at positions -3 to 3 (Zalkin and Yanofsky, 1982). Deletion of the TATA box has been shown to reduce the level of

transcription without affecting regulation, which suggests that the TATA box is part of the promoter (Guarente, 1984). Accurate mRNA initiation is determined by specific sequences downstream from the TATA box and not by the distance of the initiation site from the TATA element. This distance is variable with distances up to 100 bp being functional (Braus et al., 1988).

The CAAT box, found in many mammalian genes, is either absent or partially disguised in S. cerevisiae genes (Dobson et al., 1982a). The trp5 gene has the sequence CACT immediately after a run of thymidines (Zalkin and Yanofsky, 1982), while the pgk gene has a CAAT region between its two TATA boxes (Dobson et al., 1982a). Most efficiently expressed yeast genes have a CT-rich region (for the pgk gene, at position -49 to -69) followed by a CAAG sequence 8-12 bp later (Dobson et al., 1982a). In less efficiently expressed genes, such as the trp1 gene, the spacing between the CT-CAAG region and the transcription initiation site is much larger (Dobson et al., 1982a). It is also possible that the CT block and CAAG region contribute to the efficiency of a yeast gene promoter (Dobson et al., 1982a).

Downstream activator sequences (DAS) located within the coding region of the gene may be necessary for full promoter activity (Mellor et al., 1985). The pgk gene has a DAS, however it is not known whether this is specific for the pgk gene.

In addition to common promoter features, certain genes have specific regulatory sequences. Siciliano and Tatchell (1984) examined the mating type locus (mata and matα) in yeast. An intergenic regulatory region was found between the matα1 and matα2 genes. Some yeast gene promoters do not contain all the regulatory elements described. The mfα promoter does not have the CAAG sequence in a CT-rich block (Kurjan and Herskowitz, 1982). The trp genes from yeast have relatively complex promoters. Transcription is initiated at multiple sites. The trp1 gene has the initiation sites organised into two clusters with two and three different start sites respectively. Each cluster has a putative UAS sequence and a TATA box. The trp1 genes used in yeast vectors usually contain the second cluster only (Braus et al., 1988). The complete trp1 promoter is thought to contain a 5′ region involved in transcription termination which is required for efficient gene expression (Braus et al., 1988). Bennetzen and Hall (1982a), using the adh2 gene from S. cerevisiae, noted that this gene had the same transcription terminator sequence (5′-TAAATAA-Pu) (Pu-Purine) as several other yeast genes.

Most S. cerevisiae genes lack obvious ribosome binding sites. However, Zalkin and Yanofsky (1982) found that certain sequences in the 5′ region of the trp5 gene showed some homology with the 18S rRNA. They suggested that this complementarity with the 3′ end of the 18S rRNA may increase the efficiency of 40S ribosome binding. A Pu-CACACA sequence is found in the 5′ untranslated regions of a

variety of yeast genes. Although there is no evidence to support it, the Pu-CA sequence is believed to be involved in transcription initiation (Zalkin and Yanofsky, 1982).

All yeast genes studied so far have an ATG initiation codon, although Sherman and Stewart (1982) suggest that precise sequence conservation around the ATG codon is not important for efficient transcription. However, the first ATG codon adjacent to the 5′ end of the coding sequence is usually the translation initiation codon (Cigan and Donahue, 1987). There is usually an A residue at -3 (Dobson et al., 1982a; Cigan and Donahue, 1987).

Construction of vectors that facilitate the expression of heterologous genes in yeast.

The promoter regions from various yeast genes have been used in the construction of vectors for the expression of heterologous or foreign proteins in S. cerevisiae. The promoter of the pgk gene has been widely used (Tuite et al., 1982; Dobson et al., 1982a; Mellor et al., 1983; Mellor et al., 1985). Mellor et al. (1983) constructed two types of expression plasmid using the pgk promoter. A fragment of the promoter sequence, terminating one base pair upstream of the ATG initiation codon was inserted into the plasmid pMA3a. This could be used to express heterologous coding sequences that have their own initiation codon. The pgk promoter, in the second plasmid construct caused termination at sites within the pgk coding region. The resulting

plasmids could be used for the expression of fusion genes that lack their own initiation codon. Mellor et al. (1985) demonstrated the expression of human interferon-γ-2 (ifnγ2) using these plasmids. The production of Ifn when associated with the pgk and trp1 promoters (Dobson et al., 1982b) can be compared. The trp1 promoter directed Ifnγ2 synthesis that was a thousand fold lower than the pgk promoter directed synthesis. This illustrates the differences in transcriptional efficiency of the two promoters.

Transcriptional terminator sequences are important for maximal heterologous DNA expression. The ifn cDNA used by Mellor et al. (1985) had a fortuitous termination signal which was recognised in S. cerevisiae. If this signal was deleted, the yield of Ifn was considerably reduced. Vectors have been constructed which contain both the pgk promoter and termination sequences (Mellor et al., 1985).

Other S. cerevisiae promoter and terminator sequences have been used in the construction of yeast expression vectors. Table 1.1 lists some examples of promoters that have been used for the expression of heterologous genes.

Table 1.1. Examples of yeast gene promoters used in yeast expression vectors. (*enol* - enolase gene; *mel1* - β-galactosidase gene; *pho5* - acid phosphatase gene; *tpl* - triosephosphate gene)

Gene	Reference
adc1	Ruohonen *et al.*, 1987
adh1	Bennetzen and Hall, 1982a; Hitzeman *et al.*, 1981
adh2	Price *et al.*, 1987
cyc1	Neill *et al.*, 1987
eno1	Innis *et al.*, 1985
gal1	Goff *et al.*, 1984; St John and Davis, 1981
gal7	Handa *et al.*, 1987
gal10	Schultz *et al.*, 1987a
mel1	Curry *et al.*, 1988
mfα1	Kurjan and Herskowitz, 1982; Brake *et al.*, 1984
pgk	Dobson *et al.*, 1982a; Tuite *et al.*, 1982
pho5	Kramer and Anderson, 1980; Kramer *et al.*, 1984
tpl	Moir and Dumais, 1987
trp1	Dobson *et al.*, 1982a

Although a comparative study of all yeast promoters using the same heterologous gene has not been conducted, it appears as if promoters from glycolytic genes are the most efficient.

Factors affecting heterologous gene expression

The choice of yeast gene promoter is important for attaining maximal levels of heterologous gene expression. However, most yields of heterologous products are less than 5% of the total cell protein. Mellor et al. (1985) compared the yield of Ifnγ2 using the pgk promoter (1-3% of total cell protein) with the yield of Pgk from the homologous pgk gene on high copy number plasmids (50% of total yeast protein).

Several factors may limit the expression of heterologous proteins. Factors to take into consideration are efficiency in the synthesis of heterologous RNA, stability of this transcript, translational efficiency and stability of the final protein or a combination of these factors. It is possible that the discrepancy in homologous and heterologous protein yields could be a result of gene dosage caused by differences in plasmid copy number.

In addition to a yeast promoter sequence for transcription, there must be no intervening sequences (introns) that can not be processed in yeast (Langford et al., 1983; Hitzeman et al., 1981; Beggs et al., 1980). It is for this reason that cDNA is the method of choice when eukaryotic genes are cloned into yeast.

Some heterologous proteins are unstable in the yeast host. The human insulin gene is believed to be translated into protein but the small polypeptide was rapidly degraded

before it could be detected (Stepien et al., 1983; Keranen, 1986). Langley et al. (1988) noticed that proteolysis of the pre-s(2) region of a hepatitis B surface antigen occurred. When the protein was expressed in a protease-deficient host this problem was eliminated. Urdea et al. (1983) found that very low levels of human epidermal growth factor (hEgf) were produced even though the specific mRNA levels were sufficient for higher protein production.

Most yeast proteins are more stable than heterologous proteins. In support of this, Mellor et al. (1985) found the Ifn protein less stable than the homologous Pgk peptide. Although this can account for some reduction in protein levels, it is unlikely to be responsible for the large differences in protein steady state levels (Mellor et al., 1985). Further studies on Ifn by Mellor et al. (1985) found that the synthesis of the foreign protein is 15-500 times less efficient than the synthesis of homologous Pgk. The limiting stage could be translation or transcription of the Ifn mRNA. If translation was the rate limiting step then the mRNA levels should be the same, if the synthesis of the Ifn mRNA is inhibited or the mRNA degraded quicker than Pgk mRNA, then the steady state mRNA levels would not be comparable.

Inefficient translation, as a cause of limited gene expression can be due to codon usage. Highly expressed yeast genes have an extreme bias in codon usage (Bennetzen and Hall, 1982b). The Bennetzen and Hall Codon Bias Index

for the ifn gene is 0.32 and for the pgk gene is 0.91 (Tuite et al., 1982). The Codon Bias Index is calculated using the following formula (Bennetzen and Hall, 1982b) :

$$x = \frac{P - z}{Q - y}$$

Where x is the Codon Bias Index

P is the total number of occurrences of the preferred codons.

z is the number of codon usages expected if the code were used randomly

Q is the total number of amino acid residues in the protein (excluding methionine, aspartic acid and tryptophan)

y is the random expectation for usage of the preferred codons

Heterologous gene translation may be limited by the consequent requirement for rare charged tRNA species (Ikemura, 1982) which could cause a pause in the synthesis of the Ifn polypeptide (Mellor et al., 1985). Although the codon usage of heterologous genes have been compared to the codon bias of S. cerevisiae, most researchers have not found any direct evidence that inefficient translation is a cause of low yields of foreign protein (Mellor et al., 1985; Kniskern et al., 1986).

Mellor et al. (1985) noted that plasmid-directed Pgk RNA levels were five to ten fold higher than plasmid-directed

Ifn mRNA levels and fifty times higher than chromosomal Pgk mRNA levels. This suggests that the low yield of Ifn is due to the presence of heterologous DNA interrupting the pgk coding sequence. This phenomenon also occurred with calf prochymosin under pgk promoter control (Mellor et al., 1983). From this it can be deduced that low mRNA levels are not specific for any particular heterologous sequence. Initiation of Ifn mRNA and homologous Pgk mRNA occurs in the same region of the pgk promoter. Mellor and coworkers (1985) found that the expression of the ifn gene was the same irrespective of the termination signal used, either a fortuitous terminator in ifn cDNA or the pgk terminator. Therefore, structurally, the difference between the ifn gene and the pgk gene transcripts is the coding sequence. This result suggests that the coding sequence of the pgk gene acts as an internal enhancer and is an example of a DAS.

This discussion of pgk-directed Ifn synthesis may not hold for the synthesis of other heterologous proteins, where protein and transcript stability and codon usage may be significant. In fact, codon usage could become important if the levels of heterologous mRNA transcripts could be improved. The availability of rare tRNAs could limit the translation efficiency. Possibly the highest level of expression for a heterologous gene in yeast was reported by Kniskern et al. (1986). Hepatitis B virus core antigen was found to be in the region of 40% of the total cell protein. It is believed that this high level of expression can be attributed to the construct not containing any virus-derived

non-translated sequences. Other studies have also shown the importance of removing any non-yeast derived non-translated sequences from heterologous genes on expression plasmids (Kniskern et al., 1986).

Regulation of heterologous gene expression.

Some heterologous proteins have a detrimental effect on the host S. cerevisiae. High levels of foreign protein may reduce the growth rate and confer a selective advantage on cells that have reduced gene dosage. In order to overcome the problem of toxic proteins, regulation of their production can be achieved by using promoters that are part of a regulatory system and can be limited to expression at the end of the culture period.

Yeast gene promoters can be regulated by the carbon source used for the growth of S. cerevisiae. Non-glycolytic substrates such as glycerol, ethanol and acetate inhibit pgk gene activity whereas glucose is an inducer. The pgk system can be used in fermentations where carbon sources other than glucose are used. The pgk promoter can be induced by the addition of glucose at the end of the fermentation.

The acid phosphatase (pho5) gene promoter provides a stringent regulatory system. It is repressed in the presence of an inorganic phosphate. Depletion of the inorganic phosphate in the culture medium causes its induction (Bostian et al., 1980; Kramer and Anderson, 1980;

Kramer et al., 1984). Kramer et al. (1984) reported that the expression of the pho5 gene can be regulated using temperature-sensitive mutants. The pho5 gene is positively regulated by the pho4 gene product and negatively controlled by the pho80 gene product. An example of this system is the use of a $pho4^{ts}$, pho80 mutant. At the non-permissive temperature ($36^{o}C$), the pho5 gene is repressed due to the absence of a positive regulator, while at $24^{o}C$, the permissive temperature, the pho5 gene product is constitutively produced independently of phosphate concentration due to the lack of repression by the pho80 mutant. Although the pho5 gene has a convenient regulatory system it cannot induce heterologous gene expression to the same levels produced by glycolytic promoters (Kramer et al., 1984).

The efficient gal1 promoter requires the gal4 product as a positive regulator and is induced by galactose (St John and Davis, 1981). Johnstone and Hopper (1982) reported that the overproduction of the gal4 gene products could induce maximal expression of the gal1 promoter on a multicopy plasmid. A fragment of the gal10 promoter has been fused to other promoters to confer galactose inducibility (Schultz et al., 1987b).

α-Factor expression from the mfα1 promoter can be regulated by the products of the sir (silent information regulator) genes (Herskowitz and Oshima, 1982; Brake et al., 1984). Mutation in any of the sir genes results in a non-mating

phenotype. Brake et al. (1984) used a temperature sensitive matα sir3 strain to examine the regulatory system of the mfα promoter. At the restrictive temperature (37°C) secretion of hEgf into the culture medium was on average 0.025 ng/ml. If the temperature was shifted to the permissive temperature (24°C) then 4 mg/ml of hEgf was produced.

Secretion of heterologous proteins.

Many proteins that are medically and commercially important are secreted from their natural host. Often the first amino acid of a protein is not a methionine. This means that a synthetic ATG codon must be attached to a mature protein's cDNA sequence for that protein to be expressed in yeast. The resultant protein has an additional methionine residue attached at the N-terminus. Methionine residues are removed from S. cerevisiae proteins if they precede certain amino acids, namely threonine, alanine and glycine, and not if the preceding amino acid is one of the following: isoleucine, leucine, aspartate, lysine, glutamine, arginine, methionine and sometimes valine (Sherman and Stewart, 1982). If the precursor protein could be secreted by S. cerevisiae then the initiation signal could be chemically or enzymatically removed from the precursor protein once in the culture medium.

In the S. cerevisiae secretory system, protein secretion is mediated by the N-terminal signal peptide. The precursor

proteins are translocated into the endoplasmic reticulum (ER) where N-glycosidically linked core oligosaccharides are added. Further modifications occur in the Golgi body complex. Proteins are packaged into secretory vesicles and transported to the cell surface (Schekman and Novick, 1982).

Several heterologous proteins have been secreted in yeast using their own secretory signals attached to an efficient yeast gene promoter. Table 1.2 lists some of these proteins.

Table 1.2. Examples of protein secretion from yeast using heterologous secretory signals.

Protein	Promoter	Reference
B. amyloliquefaciens α–amylase	adc1	Ruohonen *et al.*, 1987
bovine pancreatic phospholipase A	pho5	Tanaka *et al.*, 1988
chicken egg white lysozyme	unknown	Oberto and Davison, 1985
human Ifn-γ and Ifn-δ	adh1	Hitzeman *et al.*, 1983
human salivary α–amylase	pho5	Sato *et al.*, 1986
human salivary amylase	pho5	Nakamura *et al.*, 1986
wheat α–amylase	pgk	Rothstein *et al.*, 1987

Hitzeman *et al*. (1983) found that only a small fraction of the total preprotein synthesised was secreted and 64% of the secreted fraction still contained part of the signal sequence. These studies indicate that, although the secretion of proteins using heterologous signal sequences from yeast does occur, there were low yields and the proteins produced may be a collection of processed, unprocessed and incorrectly processed molecules.

If the mature coding region of a heterologous protein was fused to the secretion signal sequences of a secreted yeast protein, then the resulting fusion protein should be efficiently secreted and the signal sequence removed by the *S. cerevisiae* processing system. Three secretory systems commonly used utilize the secretion signals from *pho5*, invertase (*suc2*) and *mfα* genes.

Perlman and Halvorsen (1983) found that the *pho5* and *suc2* genes signal sequences could be subdivided into three domains. Domain I has a charged N-terminus, domain II has a central hydrophobic core and domain III has a consensus sequence for cleavage by a signal peptidase after an alanine residue. Both invertase and acid phosphatase are secreted into the periplasmic space of the cell. Moir and Dumais (1987) used the invertase secretion signals fused to the *tpl* promoter to obtain the expression of human α–1–antitrypsin in yeast. They found that only 20% of α–AT was secreted into the culture medium, the remaining 80% was inside the cell in the secretory pathway.

The yeast mating pheromone α–factor system may be more useful since the mature α–factor from the mfα genes is secreted into the culture medium. The α–factor precursor contains a 22 amino acid signal peptide for direction of the protein into the ER, and a 61 amino acid pro-segment which contains three glycosylation, sites and possibly directs the peptides into the extracellular secretory pathway, rather than into the periplasmic space (Kurjan and Herskowitz, 1982; Julius et al., 1983; Julius et al., 1984a).

The α–factor system was successfully used by Brake et al. (1984) to obtain biologically active hEgf that was accurately processed and secreted into the culture medium. Bitter et al. (1984) used the α–factor secretion signals for the secretion of a β–endorphine and a synthetic hIfn-γ. Both proteins were secreted into the culture medium. The α–factor signal, fused to the β–endorphine sequence contained additional dipeptides derived from the α–factor spacer. Some secreted proteins contained the additional dipeptides indicating incomplete processing by a dipeptidyl amino peptidase. Julius et al. (1983) reported a similar phenomenon when the α–factor is overproduced. This suggests that the protease is rate limiting in processing the signal peptide. The maximal protein yields using the α–factor promoter system are low. This may be increased by using a more efficient promoter in place of the α–factor promoter but still retaining the α–factor secretion signals. A

variety of promoters from yeast genes have been fused to the mfα gene secretion signals (Table 1.3).

Table 1.3. Examples of yeast promoters fused to mfα gene secretion signals

Gene	Promoter	References
Bo11-2	adh2	Price et al., 1987
EBV glycoprotein	gal10	Schultz et al., 1987b
human lysozyme	adh1	Castanon et al., 1988
MuGm-csf	adh2	Price et al., 1987

Cousens et al (1987) fused the UAS from the adh2 promoter to the region of the gapdh promoter containing the TATA box sequence. The human proinsulin (pi) gene sequence was fused to the 3′ end of the human superoxidase dismutase gene coding region. This was in turn fused to the hybrid promoter. Pi, produced as an intracellular protein, was more than 15% of the total cell protein.

Several factors other than yeast gene promoter strength and secretion signals may influence the secretion of heterologous proteins. The growth phase of the culture is important. Brake et al. (1984) found that cells grown to the stationary phase contained very little intracellular hEgf with more than 95% secreted into the culture medium. During exponential growth secretion was less efficient with 67% of hEgf produced remaining intracellular.

The introduction of a signal peptide to the N-terminal end of a protein does not always allow the secretion of the protein. Secretion also depends on the structure of the protein (Yoshizumi and Ashikari, 1987). It was initially thought that very large proteins could not be secreted by S. cerevisiae. The E. coli β–lactamase protein, using the suc2 gene signal sequence, was transported into the ER where it became trapped (Emr et al., 1984). This view has been subsequently proven incorrect. A 400 kD envelope protein from Epstein-Barr virus has been expressed and secreted in yeast using the mfα promoter and prepro-leader sequence (Schultz et al., 1987a). Examples of some secreted proteins and their size are listed in Table 1.4.

Table 1.4. Examples of proteins secreted from S. cerevisiae and their size. (EBV - Epstein Barr Virus; MuGm-csf - Murine granulocyte colony stimulating factor)

Protein	Size (kD)	Reference
Cellobiohydrolases	70-200	Penttila et al., 1988
EBV glycoprotein	400	Schultz et al., 1987a
Endoglucanases	70-80	Penttila et al., 1987
MuGm-csf	45	Price et al., 1987
Wheat α–amylase	44	Rothstein et al., 1987

There are certain advantages in secreting heterologous proteins from yeast. Brake et al. (1984) compared the yields of secreted hEgf (4.6 mg/ml) with intracellular hEgf

(30 μg/ml) (Urdea et al., 1983). This suggests that the protein could be degraded or toxic if expressed intracellularly. A rapid secretion of proteins may reduce proteolysis or the toxic effect. Yeast cells release very little extraneous protein into the culture medium, therefore purification of secreted heterologous proteins is simplified. The continuous secretion of protein may result in a higher yield than that which could be attained by intracellular synthesis (Brake et al., 1984).

Glycosylation.

Some degree of glycosylation of heterologous proteins occurs in yeast. The amount of glycosylation is variable but generally low, with some proteins having no detectable carbohydrate residues. The hepatitis B surface antigen is an example of the latter statement (Valenzuela et al., 1982). The primary secreted calf chymosin from yeast appeared to be heavily glycosylated (Smith et al., 1985). Yoshizumi and Ashikari (1987) noted that most proteins passing through the lumen of the ER are glycosylated. It appears as if glycosylation does not have any detrimental effect on the biological activity of heterologous proteins from S. cerevisiae.

Plasmids integration related to protein secretion

In some cases it may be necessary to ensure absolute culture homogeneity. This can be achieved by the stable integration

of heterologous genes into the host chromosomes. It is expected that protein yields would be reduced due to the decrease in gene dosage. Smith et al (1985) had the surprising result of an increase in the efficiency of secretion of calf prochymosin using the suc2 promoter and signal sequences when the heterologous gene was integrated into the yeast chromosome. These yeast strains produced approximately the same amount of prochymosin as strains with the gene on a multicopy plasmid, but secreted more of the protein. Multicopy integrative vectors using the transposable Ty element have been developed (Jacobs et al., 1988). Expression cassettes can be inserted into the Ty element located on a 2 μm-plasmid, transposition can occur to various sites in the yeast genome.

Methods used to obtain the expression of foreign genes in yeast has been discussed in detail. In many cases described, maximal gene expression was required. In this study an endoglucanase gene and a xylanase gene from the bacterium Clostridium acetobutylicum were cloned into yeast. The production of these two enzymes would enable yeast to aquire some enzymatic activity against the polysaccharides cellulose and xylan. In this case it would not necessary for the yeast to produce large amounts of enzyme, rather the constitutive expression of the endoglucanase and xylanase genes would presumably enable yeast to utilize a wider substrate range. Abundant plant waste material could become available for the production of ethanol in industrial yeast fermentations.

Lignocellulose and its degradation.

Lignocellulose waste, formed from plant biomass, is the most abundant renewable natural resource. There is considerable potential for the use of waste cellulosic material in the fermentation industry (Knowles et al., 1987).

Unfortunately the complexity of lignocellulose makes hydrolysis of the material difficult. Lignocellulose is a complex material composed of three main components which are present in varying proportions:

Lignin. This is a complex polymer that contains phenolic compounds. Lignin is the most rigid component of lignocellulose and is the most resistant of the components to enzymatic by degradative enzymes (Knowles et al., 1987).

Hemicellulose. The most common form of hemicellulose is xylan which forms the second most abundant group of lignocelluloses. Hemicellulose is a collective term for polysaccharides associated with cellulose. These polysaccharides have side chains containing different sugar residues, usually galactose, mannose or xylose. Classification is according to the sugar moiety (Dekker, 1985).

Xylan degradation. A number of bacteria and fungi have xylanolytic activity (Dekker, 1985). Some fungi that are

able to produce xylanases are Aspergillus, Botryodiploidia, Pestolotia, Penicillin and Trichoderma (Reese et al., 1973). Many bacterial species that produce xylanases are anaerobes, these include Bacteriodes, Butyrvibrio, Ruminococcus and Clostridium. Bacillus and Streptomyces are examples of aerobic xylanase producers (Coughlan, 1985).

Xylanases have been subdivided into three groups depending on their specificities (Reilly, 1981):

1. β–xylosidases - hydrolyse short xylooligosaccharides to xylose.

2. Exoxylanases - attack the polysaccharide from the non-reducing end to produce xylose.

3. Endoxylanases - cleave the xylan backbone internally (Biely, 1985).

A number of xylanases have been cloned including examples from Bacillus, Bacteriodes, Streptomyces and Clostridium species (see Zappe, 1988).

Detection of xylanase activity. Xylanase activity can be detected by the release of reducing sugar equivalents from a solution containing xylan. The methods of detection are similar to those described for the detection of cellulase activity.

Cellulose. The major component of lignocellulose is cellulose. The primary structure of cellulose is made from a linear polymer of β–1,4 linked anhydroglucose units (Coughlan, 1985). Each glucose unit is rotated 180^{o} with respect to its neighboring residues. The glucose chains are arranged in parallel to form insoluble fibrils. These glucose chains are held together by hydrogen bonding and Van der Waal's forces (Rees et al., 1982). Aggregates of the fibrils form the crystalline regions of cellulose and give cellulose its insolubility. The polysaccharide molecule also has intercrystalline regions of more random structure (Knowles et al., 1987). The structure of cellulose may be further distorted by the association of lignin and hemicellulose (Knowles et al., 1987).

In nature, microorganisms are responsible for the degradation of lignocellulose (Bisaria and Ghose, 1981). Due to the nature of this complex substrate, large enzyme complexes are required for hydrolysis (Knowles et al., 1987).

Cellulose degradation. Fungi are the main producers of cellulose. The fungus Trichoderma reesei has been studied extensively with respect to cellulase production (Coughlan, 1985; Knowles et al., 1987).

Bacterial cellulase producers appear to be divisible into two groups (Zappe, 1988). Those that produce a complete set of enzymes that are required for the hydrolysis of

crystalline cellulose include some species of Clostridium, Cellulomonas and Bacteriodes and certain Actinomycetes. The second group produces only some of the enzymes necessary for cellulose degradation and therefore can not hydrolyse crystalline cellulose. Examples from this group include Bacillus species and C. acetobutylicum (Zappe, 1988).

Cellulases all cleave the same chemical bond, β–1,4 glycosidic bonds. They have been divided into three classes depending on their substrate and product specificities (Knowles et al., 1987).

1. Cellobiase (β–glucosidase) - hydrolyses cellobiose to glucose

2. Exoglucanases or cellobiohydrolases - removes cellobiose units from the non-reducing end of cellulose

3. Endoglucanases - cleave the cellulose at random internal sites

The general model for cellulase activity, is based on the combined action of the endoglucanases, exoglucanases and cellobiases (Coughlan, 1985). The endoglucanase cleaves β–1,4-glycosidic bonds at random sites in the amorphous regions of the cellulose. The exoglucanase acts on the non-reducing ends generated by the endoglucanase action, by removing cellobiose units. Cellobiase cleaves the glycosidic bond in cellobiose releasing the two glucose

units. Cellobiose generally acts as an inhibitor of cellulase activity.

For the efficient degradation of insoluble cellulose, the organism must be in close contact with the substrate. Fungi are able to envelope the cellulose with mycelium and then secrete cellulases. Multienzyme complexes are commonly found in cellulolytic bacteria. C. thermocellum produces a cellulosome which can be extracellular or cell-associated (Lamed et al., 1985).

Structure, function and cloning of cellulase genes.

A number of genes coding for cellulolytic enzymes have been cloned from bacteria and fungi (Knowles et al., 1987). Nucleotide sequence analysis of several cloned genes revealed a divergence in the structure of different cellulolytic enzymes, even when a comparison is made between cellulases of similar function from the same organism. Knowles et al. (1987) compared four endoglucanases from C. thermocellum and four cellulases from T. reesei. Some common features in the protein architecture were found although the amino acid sequences were different. Regions of amino acid homology were found at either the N-terminal or C-terminal end of the protein. Of the four T. reesei cellulases, each had a region of approximately 30 amino acids showing 70% homology either at the C-terminal (Egl and Cbh1) or the N-terminal (Egl11 and Cbh11) end. A similar result was found with Cellulomonas fimi cellulases (Exh and

Eng) (Knowles et al., 1987). Knowles et al. (1987) presented a hypothetical model for a cellulolytic enzyme based on amino acid sequence data and low angle x-ray diffraction studies. The cellololytic enzymes have a hydrolytic domain, which hydrolyses the β-1,4-glycosidic bonds, probably joined to a "tail" region by a "hinge" (Fig 1.1).

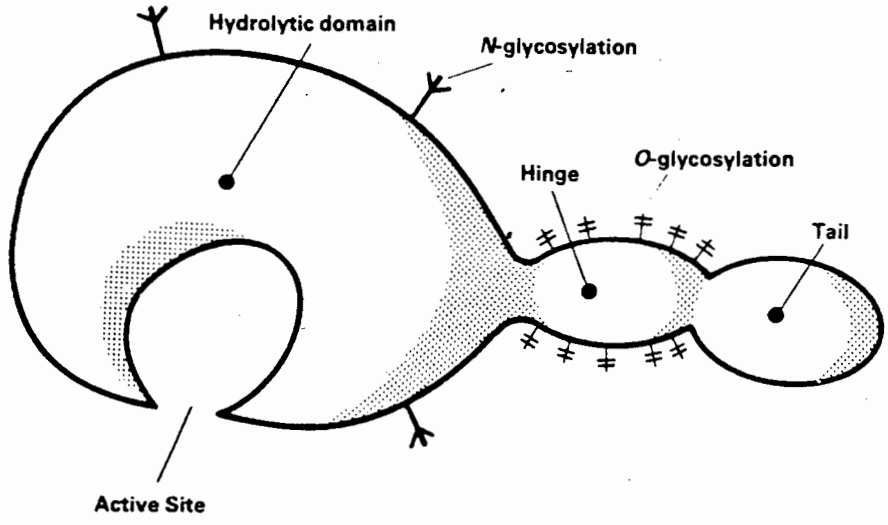


Fig 1.1 A hypothetical model of a cellulolytic enzyme based on amino acid sequence data and low angle X-ray diffraction studies (from Knowles et al., 1987).

The hydrolytic active site is found in the core protein since proteolytic removal of the terminal regions did not affect the hydrolysis of small soluble substrates, although the activity towards crystalline cellulose was reduced. This indicates that the terminal regions are involved in substrate binding or solubilization (Knowles et al., 1987). According to Knowles et al (1987), a comparison of the cellulases did not indicated any conserved region that could be an active site.

Regulation of cellulase activity. The presence of cellulose usually induces the synthesis of cellulases while glucose acts as a repressor of the synthesis (Coughlan, 1985). The regulation mechanism of cellulase synthesis is similar to that of other inducible enzyme systems. Usually, low glucose levels are associated with high cyclic AMP (cAMP) levels. cAMP and the catabolite activator protein (CAP) bind to the gene's promoter region and increase the rate of transcription from the promoter (Coughlan, 1985). According to Coughlan (1985), there is no evidence of cAMP being involved in cellulase synthesis in fungi.

Glycosylation of cellulases. Few bacterial proteins are glycosylated. However, many bacterial and fungal cellulases are extremely glycosylated (Knowles et al., 1987). This suggests that glycosylation is necessary for enzyme activity although there is no direct evidence to support this. Most of the cloned cellulases have been expressed in an unglycosylated form in E. coli and still possess cellulolytic activity. Cellulases from T. reesei that have been expressed in yeast were found to be overglycosylated (Van Arsdell et al., 1987; Penttila et al., 1987; Knowles et al., 1987). Knowles et al. (1987) suggest that glycosylation is therefore involved in substrate binding and not in the hydrolytic reaction.

Reducing sugar assays.

Soluble forms of cellulose (e.g. Carboxymethyl cellulose (CMC)) are used as substrates for endoglucanases. Activity is detected by the release of reducing sugar equivalents or by a reduction in the viscosity of the solution. Exoglucanase activity is detected by the formation of cellobiose which is released from acid swollen cellulose. β-glycosidase activity is measured according to the release of glucose or p-nitrophenol from p-nitrophenol-β-D-glucoside (Coughlan, 1985).

Evolution of cellulase genes.

Evolutionary relationships are evident between some cellulases which could have arisen by gene duplication. The shuffling and rearrangement of gene sequences that code for functional domains bordered by introns has been suggested as a cause of protein evolution, thus a number of cellulases have conserved terminal regions attached to unrelated core regions (Knowles et al., 1987).

In this laboratory the research emphasis has been on the isolation, cloning and characterization of genes involved in the acetone-butanol fermentation of C. acetobutylicum. Genes isolated include an endoglucanase, an endoxylanase, an alcohol dehydrogenase, a glutamine synthetase and two E. coli amino acid complementation genes argG6 and hisG1. The endoglucanase and xylanase genes were therefore

available for cloning into S. cerevisiae in order to improve the fermentation substrate range of yeast.

C. acetobutylicum. C. acetobutylicum produces the solvents acetone and butanol, which are important industrial chemicals. Jones and Woods (1986) reviewed the acetone-butanol fermentation of C. acetobutylicum. The production of acetone and butanol by fermentation could be more economically viable if alternative cheaper substrates could be utilised. C. acetobutylicum produces some components of the lignocellulosic system. Allcock and Woods (1981), Lee et al., (1985a, 1985b) and Zappe et al., (1986, 1987) have reported cellulolytic and xylanolytic activity from different strains of C. acetobutylicum.

The purpose of this study was to improve the fermentation ability of yeast by the introduction of cellulolytic enzymes. These enzymes could possibly enable yeast to utilize a fraction of the cellulose component in plant biomass as a carbon source.

A number of cellulolytic enzymes have already been cloned into S. cerevisiae and produced as heterologous proteins.

1. Cellobiases (β-glucosidases) from Aspergillus niger (Penttila et al., 1984) and Candida pelliculosa (Kohchi and Toh-e, 1986).

2. Exoglucanases (cellobiohydrolases): from C. fimi (Curry et al., 1988; Wong et al., 1988); T. reesei (Penttila et al., 1988).

3. Endoglucanases from C. fimi (Skipper et al., 1985); T. reesei (Van Arsdell et al., 1987; Penttila et al., 1987).

These cellulases will be discussed in Chapter 3. This study describes the cloning of a xylanase gene and a cellulase gene from C. acetobutylicum in S. cerevisiae.

Chapter 2

Subcloning of an endoglucanase gene and an endoxylanase gene from C. acetobutylicum into various yeast vectors.

2.0 Summary.

A DNA fragment, coding for the C. acetobutylicum xylanase gene, from the plasmid pHZ318 was subcloned into two yeast-E. coli shuttle vectors, YEp13 and YIp5. The recombinant plasmids were designated pYEL300 and pYIx318 respectively. E. coli cells, transformed with these plasmids, produced zones of clearing on Congo red stained xylan plates. An endo-β-1,4-glucanase gene from C. acetobutylicum, cloned in the plasmid pHZ117, was subcloned into YEp13. This plasmid was designated pYEL100. E. coli HB101 (pYEL100) colonies produced clear zones on Congo red stained CMC agar. A smaller C. acetobutylicum DNA fragment from the plasmid pHZ25T was subcloned into the yeast expression vector pMFα8. This DNA fragment, which lacked the putative endoglucanase gene promoter, was subcloned in the three different reading frames which were confirmed by nucleotide sequencing. These recombinant plasmids were designated pαEG1, pαEG2 and pαEG3. E. coli LK111 transformants of these plasmids did not produce zones of clearing on Congo red stained CMC agar.

2.1 Introduction.

Vector systems. Recombinant DNA technology has reached a high degree of refinement and has led to the manipulation of the genetic information of prokaryotes and eukaryotes at the molecular level. A wide variety of vector-host systems have been developed of which the E. coli system is the most extensively used. Plasmids are the most widely used vectors. They can be general purposes vectors, for example pBR322 (Bolivar et al., 1977) or constructed to contain bacterial or bacteriophage promoters (De Boer et al., 1983; Botterman and Zabeau, 1985) that control the expression of cloned genes adjacent to the promoter. Plasmid selection techniques have improved greatly. Whereas, initially plasmids were selected by the inactivation of an antibiotic resistance marker, positive selection plasmids can be selected by the inactivation of lethal (Kuhn et al., 1986), antibiotic sensitive (Dean, 1981) or metabolite sensitive genes (Ahmed, 1984), or the derepression of an antibiotic resistance gene (Nikolnikov et al., 1984).

Cellulose / Hemicellulose. The structure of lignocellulose and its major components, cellulose and hemicellulose, and the enzymes that mediate its degradation were discussed in Chapter 1. A number of bacteria and fungi possess xylanolytic and cellulolytic activity. The bacterium, C. acetobutylicum, under study in this laboratory, has been shown to produce components of these systems.

C. acetobutylicum. Certain C. acetobutylicum strains produce some enzymes from the cellulase complex but cannot degrade crystalline cellulose (Allcock and Woods, 1981). Fermentation of hemicellulose to solvents by C. acetobutylicum has been investigated (see Jones and Woods, 1986). Compere and Griffith (1979) found Clostridium species that were able to produce solvents from xylan. Xylanolytic activity in various C. acetobutylicum strains was investigated by Lee et al. (1985a). Twenty strains were tested, of these seventeen hydrolysed larch xylan. Endoglucanase activity was also found in certain C. acetobutylicum strains. Lee et al. (1985b), on testing the seventeen xylanolytic strains further, found two that also showed endoglucanase activity. These strains NRRL B527 and ATCC 824 were able to hydrolyse CMC and produce zones of hydrolysis on CMC agar as visualised by using the Congo red staining technique of Teather and Wood (1982). Lee et al. (1985b) compared the endoglucanase activity levels of the two C. acetobutylicum strains with other endoglucanases. The activity levels were similar to other bacterial endoglucanases from Bacteroides succinogenes, C. thermocellum and C. fimi but considerably lower than endoglucanase activities detected in T. reesei QM6a. C. acetobutylicum P270 was reported to produce an endoglucanase when using molasses as a carbon source in fermentation (Allcock and Woods, 1981). This strain was able to degrade non-crystalline substrates such as CMC, but not Avicel, a more crystalline substrate.

C. acetobutylicum P262 was reported to possess xylanase and endoglucanase activity. A xylanase and an endoglucanase from C. acetobutylicum P262 have been cloned and characterised in E. coli (Zappe et al., 1987; Zappe et al., 1986).

The C. acetobutylicum P262 xylanase gene is expressed in E. coli and produces a zone of hydrolysis when grown on LB agar containing xylan and stained as described above. Xylan did not act as an inducer or repressor of the xylanase in E. coli, when added to the culture medium. Glucose and xylose, however, when added to the culture medium caused a reduction in xylanase activity in E. coli, indicating that the xylanase gene was subject to catabolite repression in the E. coli host (Zappe, 1988).

Characteristics of the xylanase isolated from E. coli include an optimal pH range of 5.5 to 6.5, with enzyme activity decreasing sharply above and below these values. However, the xylanase was stable at pH 3.5 and pH 8.5 since activity was restored when the pH was returned to pH 6.0. The temperature optimum was between 37^{o}C and 43^{o}C. The enzyme was relatively stable at the elevated temperature of 55^{o}C, but higher temperatures reduced activity (Zappe, 1988). The isoelectric point (pI) was approximately 10. The xylanase enzyme has an apparent Mr of approximately 28 000. Zappe (1988) showed that the C. acetobutylicum xylanase cloned in E. coli was localised in the cytoplasmic fraction (98%). The enzyme properties were similar to those

et al. (1987) for an endoxylanase (XynB) from C. acetobutylicum ATCC 824.

The endoglucanase from C. acetobutylicum P262, cloned into E. coli, produced a zone of hydrolysis (visualised as described above), when grown on LB agar containing CMC (Zappe et al., 1986). There was no marked effect on endoglucanase activity if CMC or cellobiose was added to the culture medium. The endoglucanase, similar to the xylanase, was subject to catabolite repression in E. coli. The addition of glucose to the culture medium caused a decrease in endoglucanase activity.

Zappe et al. (1986) reported the following characteristics for the endoglucanase from C. acetobutylicum P262 cloned in E. coli. Unlike the xylanase, 75% of endoglucanase activity was located in the periplasmic space, although there was no detectable endoglucanase activity in the culture supernatant. Maximal endoglucanase activity was found between pH 5.0 and 6.5. Activity decreased sharply below pH 5.0 and above pH 6.5. The endoglucanase activity could not be restored if the enzyme had been subject to both high (pH 8.5) and low (pH 3.5) pH, indicating that the enzyme was unstable at these values. The endoglucanase had a broad temperature activity range of 46^{o}C to 54^{o}C, with an optimum at 50^{o}C. The stability of the enzyme decreased above 50^{o}C with almost zero activity at 60^{o}C. The calculated Mr of the enzyme, deduced from the nucleotide sequence, was 49 000. The endoglucanase was irreversibly denatured by SDS. A

comparison of the deduced protein sequences of endoglucanase genes from Bacillus subtilis, alkalophilic Bacillus species and C. acetobutylicum showed extensive homology. It is of interest to note that these Gram positive bacteria are non-cellulolytic, whereas endoglucanases isolated from cellulolytic organisms such as C. thermocellum and T. reesei exhibit significantly less protein homology.

Yeast shuttle vectors. Three yeast E. coli shuttle vectors were used in this study. The use of shuttle vectors allows for the simplified construction of clones using the E. coli host before introduction into the yeast host. The first vector, YEp13 (Broach et al., 1979; Broach, 1983), is an example of a yeast episomal vector. The YEp shuttle vectors contain the β-lactamase marker of pBR322, the pBR322 origin of replication and the yeast leu2 gene. YEp13 has a 2.2 kb EcoRI fragment from the yeast 2 μm plasmid B-form which consists of the 2 μm origin of replication and the cis-acting rep3 locus. It lacks the trans-acting rep1 and rep2 loci, which are necessary for high-copy propagation of the 2 μm plasmid in yeast, and therefore is stably maintained at a high copy number only in strains of yeast containing endogenous 2 μm plasmids [cir^+] (Broach, 1983). A restriction endonuclease map of YEp13 is presented in Appendix C.

The copy number of plasmids such as YEp13 in yeast is 25 to 100 copies per cell (Gerbaud et al., 1979; Gonzalez et al., 1985). The frequency of transformation of S. cerevisiae by

YEp13 for leucine prototrophy was of the order of 1 x 10^3 1 x 10^4 colonies per μg of DNA (Broach et al., 1979; Gonzalez et al., 1985). This plasmid, YEp13, has been used in the construction of yeast genomic libraries (Broach et al., 1979; Webster and Dickenson, 1983; Jimenez and Davies, 1980; Chen et al., 1984; Gonzalez et al., 1985; Kochi and Toh-e, 1986). YEp13 has also been used in the characterization of genes, for example the pho8 gene of S. cerevisiae (Kaneko et al., 1987), a β-glucosidase from A. niger (Penttila et al., 1984), the human adenovirus (Handa et al., 1985) and vaccinia virus telomere (De Lange et al., 1984). YEp13 was used in the development of PEG-induced transformation of bacteria and yeasts by Klebe et al. (1983). This plasmid was also used in gene replacement experiments by co-transformation (Rudolph et al., 1985).

The second yeast E. coli shuttle vector used was YIp5 (Struhl et al., 1979). It carries the ura3 marker for selection in yeast, and the Tc^r and Ap^r genes from pBR322 for selection in E. coli (see Appendix C for the structure and a restriction endonuclease map of YIp5). YIp5 transforms yeast cells at low frequency (1-10 colonies per ug of DNA) since it integrates into the genome by homologous recombination. Once integrated, the plasmid is stable and replicates as part of the host chromosome. Transformants of YIp5 contain only one copy of the entire plasmid integrated at the ura3 locus (Struhl et al., 1979). This plasmid has been used for the expression of calf prochymosin in yeast

(Goff et al., 1984). YIp5 has been used in the construction of new plasmids. DNA sequences from Candida utilis, when cloned into YIp5, conferred on YIp5 the ability to replicate autonomously in S. cerevisiae (Hsu et al., 1983). Ma et al. (1987) used YIp5 to extend the series of YCp, YEp and YRp plasmids.

The expression plasmid pMFα8 (Miyajima et al., 1985) was the third shuttle vector used in this study. pMFα8 contains the expression and secretion signals of the mfα gene (Kurjan and Herskowitz, 1982; Julius et al. 1984a), the trp5 terminator for termination and polyadenylation of the transcript, the trp1 gene for selection in yeast, and the 2 μm plasmid origin of replication for plasmid maintenance. pMFα8 contains the β-lactamase gene and colE1 origin of replication derived from pBR322 sequences and can therefore be selected for Ap^r in E. coli. pMFα8 has a unique StuI cloning site. The restriction endonuclease StuI cleaves the DNA after an arginine codon (AGG) to produce a blunt end that is in-frame with the α-factor gene's secretion signals. Genes inserted at this site must therefore be fused in the correct reading frame in order to obtain gene expression (Miyajima et al., 1985). Appendix C presents a restriction map of pMFα8. The expression vector has been used by Miyajima and co-workers for the expression of three genes in S. cerevisiae: mouse interleukin-2 (Miyajima et al., 1985), murine and human granulocyte-macrophage colony stimulating factors (gm-csf) (Miyajima et al., 1986).

This chapter describes the subcloning of xylanase and endoglucanase genes, previously cloned from *C. acetobutylicum* P262, into these yeast-*E. coli* vectors.

2.2 Materials and Methods

2.2.1 Bacterial strains and plasmids. *E. coli* strains HB101 (*leuB6*, *trpE38*, *metE70*, *recA13*, *supE44*), (Boyer and Roulland-Dossoix, 1969) and LK111 (*lacI*$^-$, lacZΔM15, *lacY*$^+$), (Zabeau and Stanley, 1982) were used as recipient strains for recombinant plasmids. The *E. coli* strains were obtained from the strain collection in the Microbiology Department at the University of Cape Town. The plasmid YEp13 was obtained from the Microbiology Department at the University of Cape Town, the plasmid YIp5 from the Microbiology Department at the University of Stellenbosch and the plasmid pMFα8 from American Type Culture Collection (Maryland, USA).

Plasmids pHZ318 (*C. acetobutylicum* P262 xylanase gene cloned in pUC18); pHZ117, pHZ25 (various fragments of the *C. acetobutylicum* endoglucanase gene cloned in pUC19); were obtained from H. Zappe, University of Cape Town. Restriction endonuclease maps of these plasmids are presented in Appendix C.

2.2.2 Media and Buffers

All media and solutions not described in the text are listed in Appendix B.

2.2.3 Growth conditions. *E. coli* was grown in Luria-Bertani (LB) medium (Maniatis *et al.*, 1982). For the detection of clones expressing xylanase activity, LB agar

contained 0.1% (w/v) oat spelt xylan (Sigma x-0376; Lot 14F-0421). Endoglucanase activity was detected on LB agar containing 0.1% (w/v) CMC (medium viscosity; Sigma no. C4888, Degree of substitution was 0.7).

2.2.4 Restriction endonuclease digestion. Restriction digests were carried out in different volumes depending on the amount of DNA being digested. Normally 200-400 ng DNA was digested in a 20 μl volume using 1 u of enzyme for 1 h at 37^{o}C. The DNA was digested using one of four restriction buffers depending on the salt requirement of the restriction enzyme (restriction enzyme SmaI required a special buffer).

Provided the salt requirements of enzymes were compatible, multiple digests of the DNA using different enzymes could be done simultaneously. If this was not the case, then the digests were done sequentially using the enzyme with the lowest salt concentration requirement first, the salt concentration being adjusted before the addition of the next enzyme.

For analysis of the DNA by electrophoresis, the digestion reactions were stopped by the addition of sample loading solution. If the DNA was required for further enzyme reactions (e.g. ligation) it was purified by extraction with TE-saturated phenol (1/10 volume) followed by the addition of an equal volume of chloroform/isoamyl alcohol (24 : 1). The solution was vortexed briefly and the phases separated by centrifugation. The aqueous phase was re-extracted with

an equal volume chloroform/isoamyl alcohol. The DNA was precipitated from the aqueous phase by the addition of 5 M $NaClO_4$ (1/10 volume) and an equal volume of isopropanol. After 10 min on ice the DNA was pelleted by centrifugation in a microfuge for 15 min, washed with 70% ethanol, dried, and resuspended in TE buffer (Maniatis et al., 1982).

2.2.5 Agarose gel electrophoresis. Agarose (Sigma type II) was used at varying concentrations (0.5 - 2.0% (w/v) in Tris-acetate buffer (Maniatis et al., 1982) depending on the sizes of DNA fragments being investigated, 0.8% (w/v) was routinely used. The gels were stained by the addition of EtBr (0.5 μg/ml) to the molten agarose. Generally 300 ng of DNA was loaded per lane on the gel. A horizontal submerged gel system as described by Maniatis et al. (1982) was used for the electrophoresis. Gels were electrophoresed at 2 V/cm for 16h. or 5 V/cm for 4 h. Small gels, using a Hoefer Minnie Submarine agarose gel system (model ME33; Hoefer Scientific Instruments, San Francisco), were also used at 5 V/cm for 2 h. DNA bands were visualised using a short wave (254 nm) UV transilluminator (Chromato-Vue Model Ts-15, UV Products Inc., San Gabriel, California, U.S.A.).

The fragment sizes were calculated by comparison to a standard curve of the mobility of λ DNA fragments digested with PstI restriction enzyme. Approximately 1 μg of DNA was used as a standard on every gel.

2.2.6 Rapid subcloning method. DNA that was required for subcloning experiments was separated from extraneous fragments, using the rapid subcloning techniques described by Struhl (1985), with minor modifications. DNA fragments were separated by electrophoresis using low melting point (LMP) agarose (Sea Plague agarose, Marine Colloids, Rocklands, ME) in Tris-Acetate buffer (50 mM, pH 8.2) with 0.5 μg/ml ethidium bromide (EtBr). The DNA was visualised using a 310 nm transilluminator. The fragments were removed from the gel in as small a volume as possible. The gel slices were melted at 65°C for 2 min, equilibrated at 37°C then combined at the required ratios (vector-insert ratio of 1:5 for flush ends and 1:2 for overhanging ends) in a final volume of 10 μl. A further 10 μl containing 2x ligation buffer and T4 ligase (1 u) was added and held at room temperature for 2 h. The gel solution was melted at 65°C and diluted with 5 volumes of 0.1 M $CaCl_2$ before transformation of E. coli cells.

2.2.7 Klenow reaction for flush ends (fill-in). DNA (0.5-1.5 μg in 15 μl TE), 0.25 mM of each dideoxynucleotide (dXTP) and polymerase buffer were mixed before 1 u of DNA polymerase I large fragment (Klenow) was added. This solution was incubated at 37°C for 15 min. The enzyme and excess nucleotides were removed by phenol and chloroform isoamyl alcohol extraction as described above (2.2.4) (Maniatis et al.,1982).

2.2.8 Flush-ending using Mung Bean Nuclease. DNA (5 μg), Mung Bean Nuclease buffer and 15 u of Mung Bean Nuclease (New England Biolabs) were mixed in a volume of 200 μl and incubated for 30 min at 30°C. Thereafter the DNA was purified as described above (2.2.4).

2.2.9 Phosphorylation and addition of linkers. The method used was as described by Maniatis et al. (1982). The linkers were phosphorylated using T4 Polynucleotide Kinase. Linkers were added by ligation in a solution containing 0.4 μg of flush-ended DNA, linkers (1.0-2.0 μg), T4 ligase buffer and 1 u T4 DNA ligase. The solution was incubated at 22°C for 6 h. DNA was purified by phenol, chloroform isoamyl alcohol extraction (2.2.4). An excess of enzyme was used to digest the linkers because of the large molar excess of linkers. The excess linkers were removed by gel electrophoresis.

2.2.10 Preparation of competent E. coli cells. Competent E. coli cells were prepared essentially according to the method described by Dagert and Ehrlich (1979). An overnight culture of E. coli was diluted 1/1000 into 20 ml LB medium. The culture was grown with shaking to an optical density at 600 nm (OD_{600}) of 0.2, at 37°C, and cooled on ice for 5 min. The cells were collected by centrifugation (4000 x g for 5 min at 4°C), washed with 20 ml cold 0.1 M $MgCl_2$ and resuspended in 10 ml of cold 0.1 M $CaCl_2$. The cells were kept on ice for 20 min, collected by centrifugation and

2.2.11 Transformation of competent E. coli cells. Plasmid DNA was added to 100 μl samples of competent cells on ice. The DNA concentration used was 1-5 ng of purified plasmid DNA or 10-30 μl from an diluted LMP agarose gel ligation reaction. After 10 min on ice the cells were heat-shocked at 42^{o}C for 5 min to induce DNA uptake. LB medium (1 ml) was added and the cells and left at 42^{o}C for 30 min after which the transformation mix was plated onto LB agar containing the relevant antibiotics. Transformation controls included: competent cells with no DNA added and unrestricted plasmid to establish the transformation frequency. Transformation frequency was 3000-6000 colonies per ng of unrestricted plasmid DNA.

2.2.12 Congo Red staining technique. E. coli transformant colonies grown on Luria agar containing 0.1% (w/v) CMC or 0.1% (w/v) oat spelt xylan were removed by placing filter paper (Whatman No. 1) over the plates, pressing lightly and then peeling off the filter paper. The position of the colonies on the filter paper relative to their position on the agar plates was noted. The plates were stained using Congo Red stain for 15 min after which the stain was removed and the plates destained for 15 min using 1 M NaCl. Colonies showing cellulolytic or xylanolytic activity were identified by a zone of clearing around the colony (Teather and Wood, 1981). The required colonies could be picked from the filter paper.

2.2.13 Preparation of plasmid DNA.

2.2.13.1 Small scale (miniprep). Plasmids were isolated as described by Ish-Horowicz and Burke (1981). The cells from a 5 ml overnight culture (LB + Ap, 100 μg/ml) were collected in a 2 ml microfuge tube by centrifugation for 1 min in an Eppendorf microfuge. The pellet was resuspended in 0.2 ml of Solution I (50 mM glucose, 25 mM Tris-Cl, pH 8.0) and kept at room temperature for 5 min. Solution II (0.4 ml) (0.2 M NaOH, 1% (w/v) SDS) was added and the mixture vortexed briefly and placed on ice for 5 min. Precooled Solution III (0.3 ml) (5 M KOAc, pH 4.8) was added, the solution vortexed briefly and returned to ice for a further 5 min. Precipitated cellular debris and chromosomal DNA was pelleted by centrifugation for 8 min in a microfuge. The supernatant was collected (0.75 ml) and plasmid DNA was precipitated by the addition of 0.75 ml isopropanol. After 2 min at room temperature, the plasmid DNA was collected by centrifugation (5 min). The pellet was washed with 70% ethanol, dried and resuspended in 50 μl TE buffer.

Plasmid DNA prepared by this method was used for restriction enzyme analysis.

2.2.13.2 Large scale (maxiprep). Plasmid DNA was prepared from 200 ml overnight cultures (LB + Ap, 100 μg/ml). The cells were pelleted by centrifugation for 5 min at 4000 x g. The method of plasmid extraction is as described above with the volumes of the solutions increased 40x. The preparation

cells were pelleted by centrifugation for 5 min at 4000 x g. The method of plasmid extraction is as described above with the volumes of the solutions increased 40x. The preparation was centrifuged at 12000 x g for 10 min to remove the cell debris. The plasmid DNA was precipitated by the addition of 18 ml isopropanol and pelleted at 27000 x g for 15 min. The pellet was resuspended in 4.4 ml TE buffer. EtBr (0.41 ml of 10 mg/ml stock solution) and CsCl (4.51 g) were added and the solution centrifuged at 20000 x g for 15 min. The supernatant (refractive index of 1.396) was sealed in a Beckman Quickseal ultracentrifuge tube and centrifuged for 7-14 h at 55000 rpm at 15^{o}C in a Beckman Vti65.2 rotor. The band (visualised by long wave UV light - 350 nm) was removed in as small a volume as possible. The EtBr was removed by 3-fold extraction with an equal volume of NaCl-saturated isopropanol. The sample was diluted with 2 volumes of water to reduce the salt concentration. The DNA was precipitated by the addition of an equal volume of isopropanol and pelleted by centrifugation for 15 min. The plasmid DNA was resuspended in 0.2 ml TE buffer and the concentration determined spectophotometrically. The absorption spectrum of the DNA solutions was measured between 200 and 310 nm. The concentration of DNA was determined using the conversion where 1 absorbance unit at 260 nm is the equivalent of 50 μg DNA/ml (Maniatis et al., 1982). Plasmid DNA prepared by this method was used in cloning experiments and restriction endonuclease digests.

sequenase kit was obtained from United States Biochemical and used according to the manufacturer's specifications. A primer with the sequence: 5'-CCAACAGCACAAATAACGGG-3' which is homologous to a region of the α-factor leader sequence was synthesised in an Applied Biosystems 381A DNA synthesizer (Forester City, C.A., U.S.A.). This primer was purchased from Beckman Instruments (SA). A 8% denaturing polyacrylamide gel wedge (0.4 mm spacers at the top, 0.8 mm spacers at the bottom) was cast in a BRL apparatus (34 x 40 cm). A sharktooth comb (24 teeth) (BRL) was used to form the wells. After electrophoresis the gel was dried onto filter paper (Whatman No. 3) using a Dual Temperature Slab Gel Dryer (Model 1125B, Hoefer Scientific Instruments, San Francisco). The gel was placed in a X-Ray cassette under Cronex 4 film and exposed for 5 - 24 h. The autoradiographs were developed using Kodak GBX X-Ray developer and fixer.

2.3 Results

2.3.1 Subcloning of a xylanase gene from C. acetobutylicum into the yeast-E. coli shuttle vector YEp13, in E. coli.

The plasmid pHZ318 consists of an EcoRV/HindIII C. acetobutylicum DNA fragment, encoding a xylanase, cloned into the SmaI/HindIII sites of the pUC18 multiple cloning site (mcs) (Zappe, 1988).

The plasmid was digested with the restriction endonucleases EcoRI and HindIII thus releasing the xylanase encoding fragment, which included a portion of the mcs extending from the SmaI restriction site to the EcoRI restriction site (Fig 2.1). The yeast-E. coli shuttle vector YEp13 was linearised by digestion with the restriction endonuclease BamHI. The 5′ single stranded overhanging ends of both molecules were repaired using E. coli DNA polymerase I large fragment (Klenow). The xylanase encoding fragment from pHZ318 and the linear YEp13 vector were ligated and the ligation reaction used to transform E. coli HB101 to Ap^r. Recombinant plasmid transformants were distinguished from religated parental YEp13 transformants by zones of hydrolysis around the E. coli colonies after Congo Red staining. The presence of an insert in YEp13 was confirmed by restriction endonuclease analysis. The recombinant plasmid was designated pYEL300 (Fig 2.1).

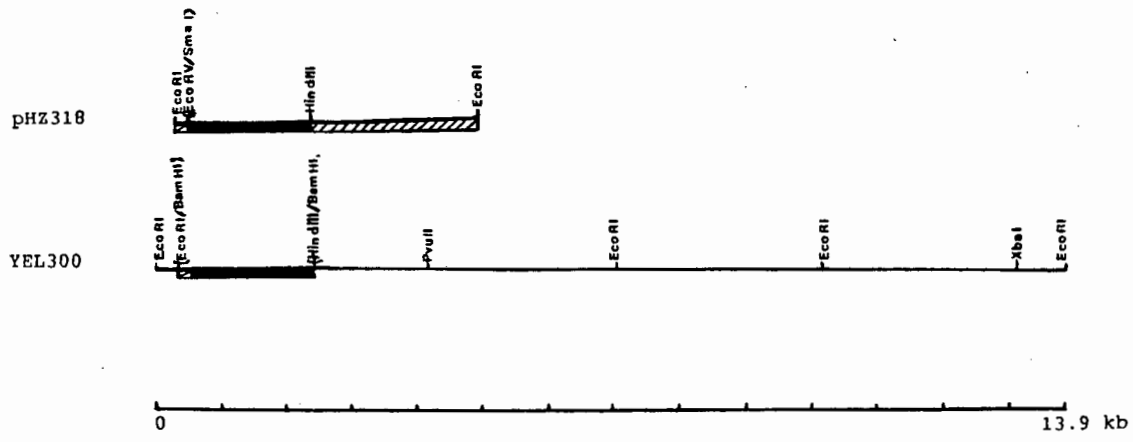


Fig 2.1 Restriction endonuclease maps of pHZ318 and pYEL300. The thick lines represent C. acetobutylicum DNA encoding a xylanase; the hatched regions represent DNA derived from pUC18 and the thin lines represent DNA from the vector YEp13. The EcoRI/HindIII DNA fragment containing the xylanase gene from pHZ318 was inserted into the BamHI restriction site of YEp13 by a flush-end ligation. The resulting recombinant plasmid was designated pYEL300.

2.3.2 Subcloning of an endoglucanase gene from C. acetobutylicum into a yeast-E. coli shuttle vector, YEp13, in E. coli. Plasmid pHZ117 consists of a HaeIII/PstI DNA fragment from pHZ100, that includes the endoglucanase gene from C. acetobutylicum, cloned into the SmaI/PstI sites of the mcs of pUC19 (Fig 2.2) (see Zappe et al. 1988).

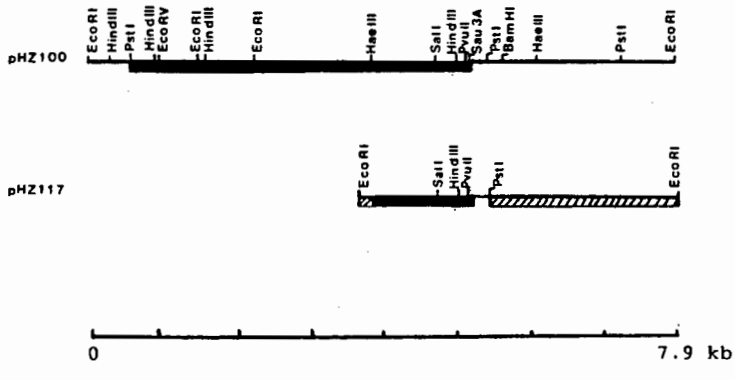


Fig 2.2 Restriction endonuclease maps of pHZ100 and pHZ117. The thick lines represent C. acetobutylicum DNA

encoding an endoglucanase. The hatched regions represent DNA derived from the vector pUC19 and the thin lines DNA from the vector pEcoR251. Deletion mapping of the C. acetobutylicum DNA fragment cloned into pHZ100 indicated that the endoglucanase gene was located between the HaeIII and PstI restriction sites of pHZ100. The HaeIII/PstI DNA fragment was subcloned into the vector pUC19 and designated pHZ117 (Zappe et al. 1988).

The endoglucanase encoding DNA fragment was released from pHZ117 by digestion with EcoRI and PstI restriction endonucleases. The yeast-E. coli shuttle vector, YEp13, was linearised by BamHI endonuclease digestion. The resulting 5' overhangs were treated in the same way as described previously. The endoglucanase encoding fragment, from pHZ117 and YEp13 were ligated and the reaction used to transform E. coli HB101 to Ap resistance.

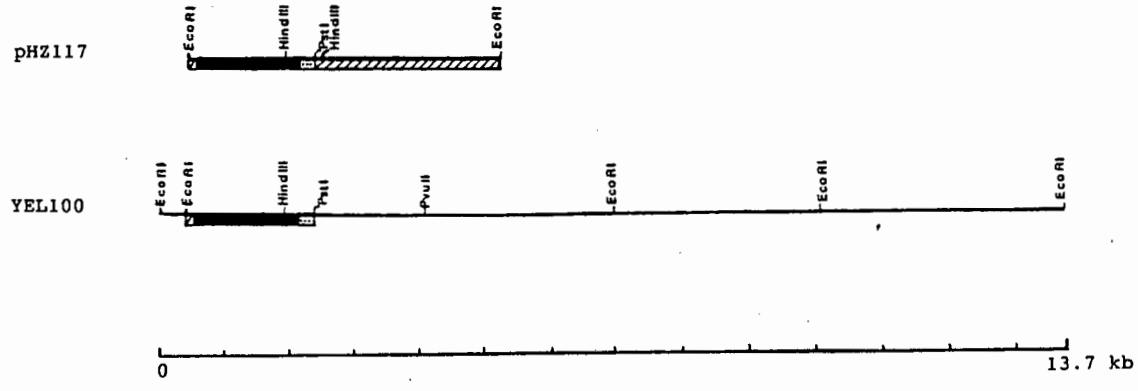


Fig 2.3 Restriction endonuclease maps of pHZ117 and pYEL100. The thick lines represent C. acetobutylicum DNA encoding an endoglucanase; the hatched regions represent DNA derived from the vector pUC19, the dotted regions DNA from pEcoR251 and the thin lines DNA from the vector YEp13. The EcoRI/PstI DNA fragment containing the endoglucanase gene from the plasmid pHZ117 was inserted into the BamHI restriction site of YEp13 by a flush-ended ligation. The resultant recombinant plasmid was designated pYEL100.

Congo red staining enabled transformant colonies harbouring recombinant plasmids and expressing the endoglucanase gene to be distinguished from YEp13 parental plasmids. The presence and orientation of the insert in YEp13 was confirmed by restriction endonuclease analysis. The recombinant plasmid carrying the endoglucanase gene was designated pYEL100 (Fig 2.3).

2.3.3 Subcloning of the C. acetobutylicum xylanase gene into the yeast-E. coli shuttle vector, YIp5, in E. coli.

The isolation of the xylanase encoding DNA fragment from pHZ318 has been described (2.3.1). This fragment was subcloned into YIp5 between the EcoRI and HindIII sites of this vector. This recombinant plasmid was designated pYIX318 (Fig 2.4).

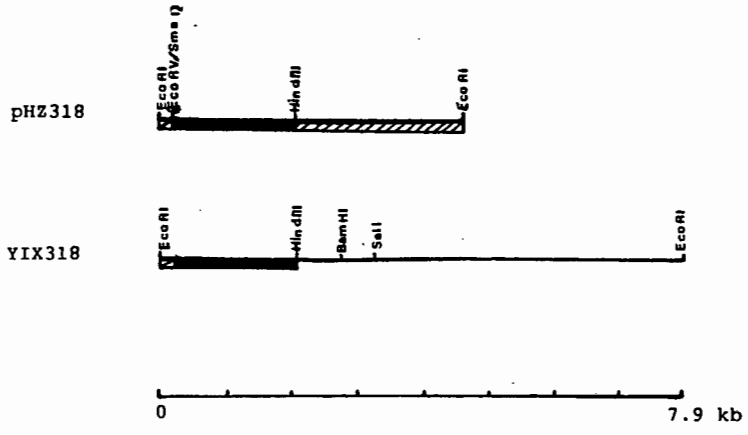


Fig 2.4. Restriction endonuclease maps of pHZ318 and pYIX318. The thick lines represent C. acetobutylicum DNA encoding a xylanase; the hatched regions represent pUC18 derived DNA and the thin lines represent DNA derived from the vector YIp5. The EcoRI/HindIII DNA fragment containing the xylanase gene from pHZ318 was subcloned between the EcoRI and PstI restriction sites of the vector YIp5. The resultant recombinant plasmid was designated pYIX318.

2.3.4 Subcloning of the endoglucanase from C. acetobutylicum into the yeast expression vector, pMFα8, in E. coli. The plasmid pMFα8 was linearised by restriction with the endonuclease StuI. A pUC19 derived plasmid, pHZ25, used in the nucleotide sequencing of the endoglucanase gene (Zappe et al., 1988) was used as a source of C. acetobutylicum DNA in the cloning experiment. The putative promoter and part of the signal peptide of the endoglucanase were removed by the action of Bal31 exonuclease (Zappe et al., 1988). The exact junction of the endoglucanase DNA and pUC19 DNA was determined by DNA sequencing.

Plasmid pHZ25 did not include the HindIII/Sau3A fragment of the endoglucanase gene cloned in pHZ117 (Fig 2.5). In order to reconstruct the endoglucanase gene, a HindIII/PstI fragment from pHZ117 was added to the 3′ end of the endoglucanase gene fragment in pHZ25. This construct, called pHZ25T, did not allow the formation of an active endoglucanase in E. coli LK111 since the putative endoglucanase promoter and part of the leader sequences were deleted (Fig 2.5), but did contain the coding sequences for the structural protein.

Due to the lack of suitable cloning sites in the endoglucanase gene, the following cloning strategies were employed.

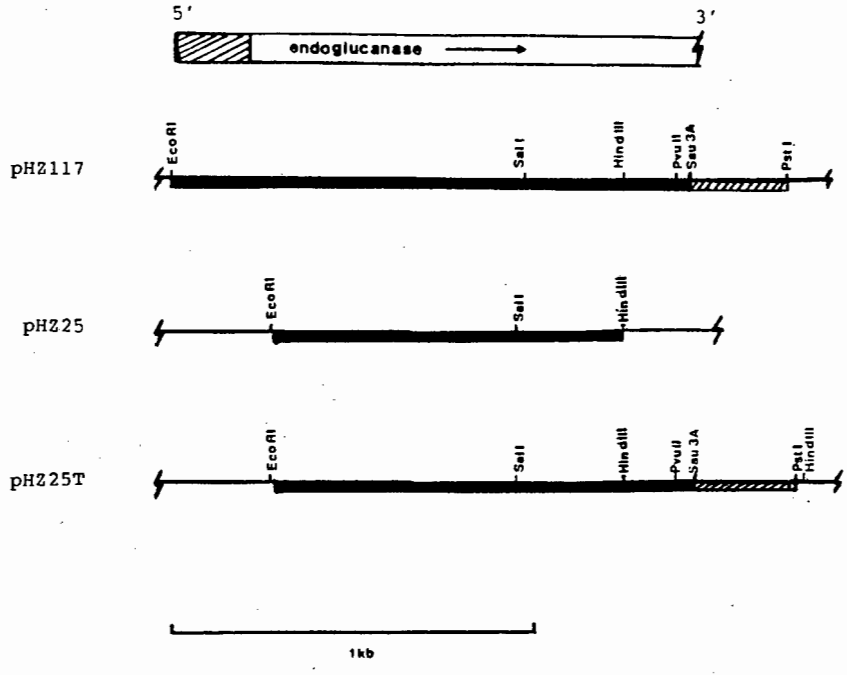


Fig 2.5. Restriction endonuclease maps of pHZ117, pHZ25 and pHZ25T. The thick lines represent C. acetobutylicum DNA encoding the endoglucanase, the hatched regions represent DNA derived from pEcoR251 and the thin lines DNA from the vector pUC19. pHZ25, a Bal31 deletion plasmid that was used in the determination of the endoglucanase gene nucleotide sequence (Zappe et al., 1988), does not include the putative promoter sequences and part of the leader peptide sequence (see Appendix D for the complete nucleotide sequence of the endoglucanase gene). pHZ25 also does not include the HindIII/Sau3A fragment present in pHZ117, at the 3' end of the gene. A HindIII/PstI fragment from pHZ117, which included this terminal region, was subcloned into pHZ25 to produce the plasmid pHZ25T, resulting in a promoterless but complete structural gene.

Cloning strategy 1

Two possible restriction sites, PvuII (at the 3' end of the endoglucanase gene) and PstI (downstream (3') from the end of the endoglucanase encoding DNA fragment), were considered. The PstI site was derived from 180 bp of pEcoR251 in the initial cloning of the endoglucanase gene (Zappe et al., 1988). The internal PvuII site located at the 3' end, is approximately 1.1 kb from the endoglucanase gene initiation codon (Fig 2.5).

The reading frame of the endoglucanase encoding fragment from pHZ25T was manipulated using a variety of restriction enzymes and techniques, including the addition of BamHI linkers. These strategies are outlined in Table 2.1. The endonuclease restriction sites used in cloning strategy 1 are indicated in Fig 2.6.

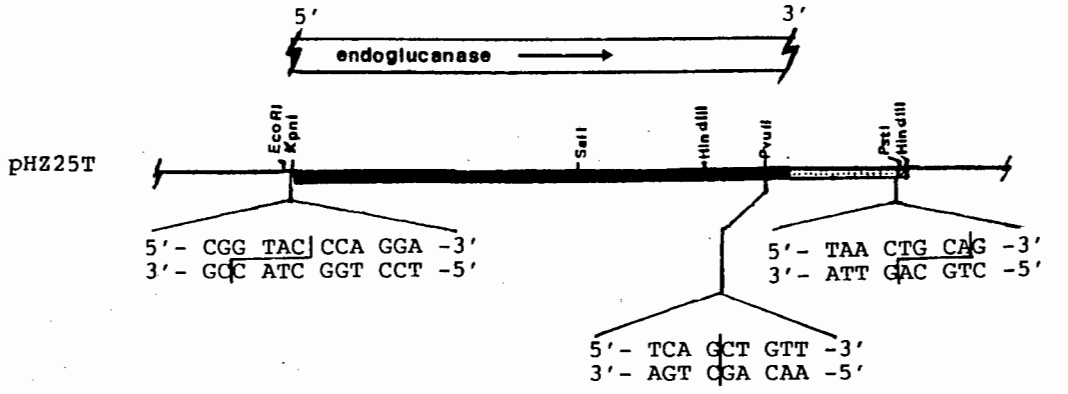


Fig 2.6. Restriction endonuclease map of pHZ25T showing the nucleotide sequence for the KpnI, PvuII and PstI restriction sites. These three restriction sites were used in cloning strategy 1, which is outlined in Table 2.1. The thick line represents the C. acetobutylicum endoglucanase encoding fragment, the dotted region represents pEcoR251 DNA and the thin lines pUC19 DNA. The orientation of the endoglucanase gene is indicated by a box shown above.

Table 2.1. Construction of recombinant plasmids containing the endoglucanase gene from C. acetobutylicum. To eliminate the bacterial gene promoter, plasmid pHZ25T (a Bal31 deletion plasmid, refer to Fig 2.5) was used. a) pHZ25T was cut at a unique PstI restriction site. The 3' overhangs were flush-ended using Mung Bean nuclease. BamHI linkers were ligated to these flush ends. The linkers were digested with BamHI endonuclease and the 5' overhangs treated with Mung Bean nuclease. pHZ25T was then cut at the unique KpnI restriction site and the 3' overhangs flushed-ended using Mung Bean nuclease. The flush-ended DNA fragment that contained the endoglucanase gene was ligated with pMFα8, linearised by StuI endonuclease. b) pHZ25T was cut at the

unique KpnI restriction site and the 3′ overhangs treated as described in a above. The PvuII restriction site on the endoglucanase DNA fragment was restricted to produce flush ends. The flush-ended DNA fragment which contained the endoglucanase gene was ligated to pMFα8, linearised by StuI endonuclease. c) pHZ25T was cut at the unique PstI and KpnI restriction sites and the 3′ overhangs treated as described in a above to produce flush ends. The flush-ended DNA fragment containing the endoglucanase gene was subcloned into the StuI restriction site of pMFα8. d) pHZ25T was cut at the PvuII restriction site. BamHI linkers were added to the flush ends and then digested with BamHI endonuclease. pHZ25T was then cut at the KpnI restriction site. The overhanging ends generated from the BamHI and KpnI restriction sites were flush-ended as described in a above. The flush-ended DNA fragment containing the endoglucanase gene was subcloned into the StuI restriction site of pMFα8.

a) PstI digestion	5′-TAA CTG CA 3′ 3′-ATT G 5′
Flush-end : Mung Bean nuclease	5′-TAA C 3′ 3′-ATT G 5′
BamHI linker addition	5′-TAA CGG GAT CCC 5′ 3′-ATT GCC CTA GGG 3′
BamHI digestion	5′-TAA CGG 3′ 3′-ATT GCC CTA G 5′
Flush-end : Mung Bean nuclease	5′-TAA CGG 3′ 3′-ATT GCC 5′
KpnI digestion	5′ CCA GGA-3′ 3′ C ATC GGT CCT-5′
Flush-end : Mung Bean nuclease	5′ CCA GGA-3′ 3′ GGT CCA-5′
b) KpnI digestion	5′ CCA GGA-3′ 3′ C ATC GGT CCT-5′
Flush-end : Mung Bean nuclease	5′ CCA GGA-3′ 3′ GGT CCA-5′
PvuII digestion	5′-TCA G 3′ 3′-AGT C 5′

c) PstI digestion	5′-TAA CTG CA 3′ 3′-ATT G 5′
Flush-end : Mung Bean nuclease	5′-TAA C 3′ 3′-ATT G 5′
KpnI digestion	5′ CCA GGA-3′ 3′ C ATC GGT CCT-5′
Flush-end : Mung Bean nuclease	5′ CCA GGA-3′ 3′ GGT CCA-5′
d) PvuII digestion	5′-TCA G 3′ 3′-AGT C 5′
BamHI linker addition	5′-TCA GGG GAT CCC 3′ 3′-AGT CCC CAT GGG 5′
BamHI digestion	5′-TCA CGG 3′ 3′-AGT GCC CTA G 5′
Flush-end : Mung Bean nuclease	5′-TCA CGG 3′ 3′-AGT GCC 5′
KpnI digestion	5′ CCA GGA-3′ 3′ C ATC GGT CCT-5′
Flush-end : Mung Bean nuclease	5′ CCA GGA-3′ 3′ GGT CCA-5′

DNA analysis of E. coli transformants of these constructs, however, only yielded parental pMFα8 plasmids (see Appendix C for a restriction endonuclease map of pMFα). The stages in each experiment were difficult to control. A revised cloning strategy, using restriction endonuclease sites derived from the E. coli plasmid Bluescript KS (pBSKS) (BRL), was devised to so that each stage in the strategy could be examined.

Cloning strategy 2

The plasmid pHZ25T was subjected to digestion by the restriction enzyme EcoRI and partial digestion by the endonuclease HindIII to obtain the entire gene (see Fig 2.5). The fragment containing the entire endoglucanase gene was isolated by cutting the fragment from a LMP agarose gel. This fragment was subcloned into two different vectors.

Frame 1:

The pHZ25T 1.25 kb fragment with 5' overhangs was flush-ended using the Klenow. This fragment was ligated into the StuI endonuclease site of pMFα8. Parental plasmids were distinguished from recombinant plasmids by restriction endonuclease analysis of miniprep plasmid DNA from Ap^r E. coli transformants (see below). A recombinant plasmid carrying the truncated endoglucanase gene fused to the α–factor promoter and secretory signals of pMFα8 in the correct orientation, and in the first of the three possible reading frames was designated pαEG1 (Fig 2.7).

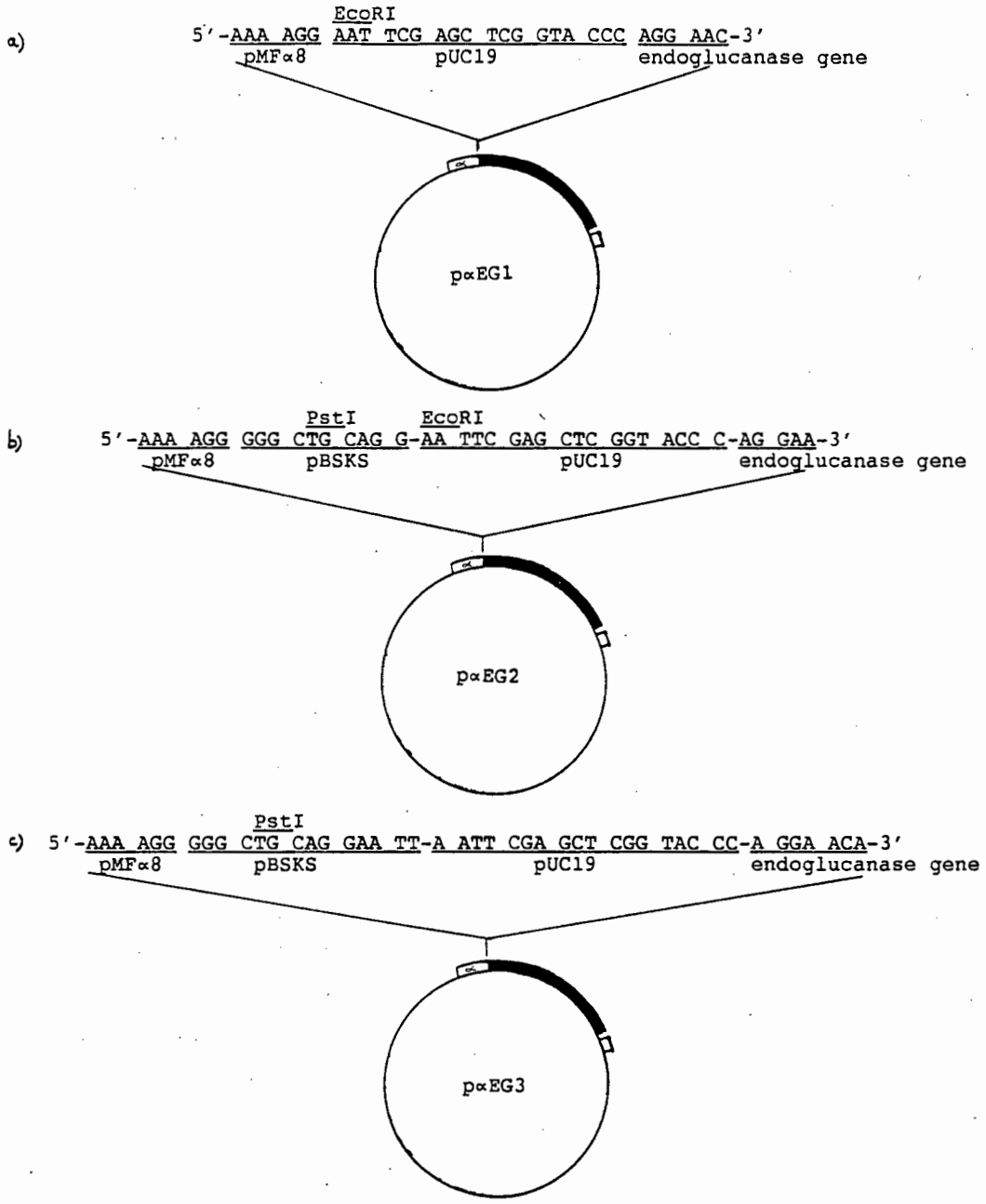


Fig 2.7. Nucleotide sequence of the junction of pMFα8 DNA and the C. acetobutylicum endoglucanase gene DNA in the plasmids pαEG1, pαEG2 and pαEG3. The thin lines represent pMFα8 DNA and the thick lines represent C. acetobutylicum DNA encoding an endoglucanase. a) pαEG1 had 18 nucleotides derived from the pUC19 mcs, between pMFα8 DNA and the endoglucanase encoding DNA fragment from C. acetobutylicum. The EcoRI restriction site from the pUC19 mcs was retained. The endoglucanase gene in this plasmid was the first of the three possible reading frames. b) pαEG2 had 10 nucleotides derived from the pBSKS mcs and 18 nucleotides from the pUC19 mcs, between pMFα8 DNA and the C. acetobutylicum endoglucanase DNA. The EcoRI restriction site was retained from the pUC19 mcs and the PstI restriction site from the pBSKS mcs. The endoglucanase gene in this plasmid is in the

second reading frame. c) pαEG3 had 14 nucleotides derived from the pBSKS mcs and 18 nucleotides from the pUC19 mcs, between pMFα8 DNA and the C. acetobutylicum endoglucanase DNA fragment. The PstI restriction site was retained from the pBSKS mcs, while the EcoRI restriction site, present in pαEG2, was filled in using Klenow. The endoglucanase gene in this plasmid is in the third reading frame. According to nucleotide sequence analysis, the pαEG3 was expected to contain the endoglucanase gene in the correct reading frame.

The pHZ25T 1.25 kb fragment was also inserted into the multiple cloning site of the vector pBSKS (BRL) at the EcoRI and HindIII restriction sites. Miniprep DNA analysis of the E. coli Ap^r transformants by restriction endonuclease digestion confirmed the presence of the insert in pBSKS (Fig 2.8). The purpose of this subcloning was to make available more restriction sites upstream of the endoglucanase gene. This plasmid, called pBSEG25T, was used in the construction of the two remaining reading frames of the endoglucanase gene in pMFα8.

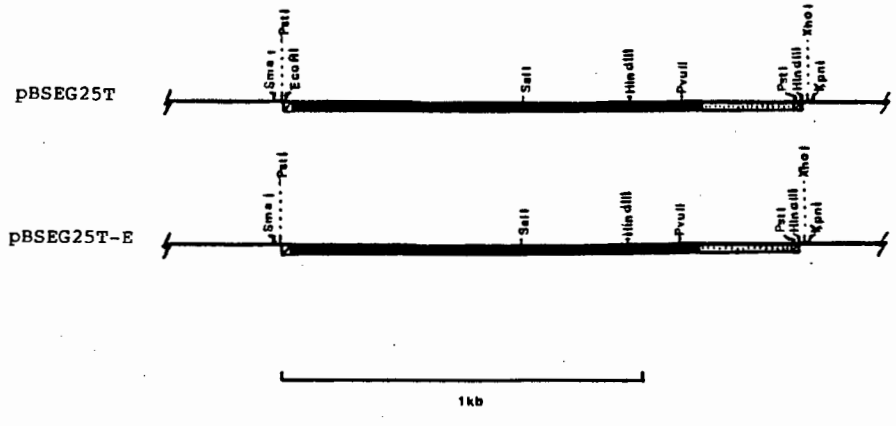


Fig 2.8. Restriction endonuclease maps of pBSEG25T and pBSEG25T-E. The thick lines represent C. acetobutylicum endoglucanase DNA, the dotted regions represent pEcoR251 derived DNA, the hatched regions pUC19 DNA and the thin lines part of the vector pBSKS. The EcoRI restriction site in pBSEG25T was digested by EcoRI endonuclease, the 5'

overhangs filled in using Klenow, and religated. The resultant plasmid, pBSEG25T-E, lacked the EcoRI restriction site.

Frame 2:

The second reading frame was obtained by SmaI endonuclease digestion and partial HindIII endonuclease digestion. This released the approximately 1.25 kb fragment from pBSEG25T with 10 nucleotides added at the 5' end. The HindIII 5' overhang was filled in using Klenow. This flush-ended fragment was cloned into the StuI restriction site of the vector pMFα8. Plasmids with inserts were distinguished from parental plasmids by restriction endonuclease analysis (see below). A plasmid with the insert in the correct orientation was selected and named pαEG2 (Fig 2.7).

Frame 3:

The plasmid pBSEG25T was restricted at the unique EcoRI retriction site. The resultant 5' overhangs were filled in using Klenow and religated using T4 DNA ligase. DNA analysis of Ap^r *E. coli* LK111 transformants yielded the plasmid pBSEG25T lacking the EcoRI site. This plasmid, called pBSEG25T-E (Fig 2.8), was subsequently restricted by SmaI endonuclease followed by partial digestion by HindIII endonuclease. This released the approximately 1.25 kb endoglucanase gene fragment with 14 additional nucleotides. The HindIII restriction site was filled in using Klenow and the fragment cloned into the StuI restriction site of pMFα8 as before. Plasmids harbouring inserts were detected by restriction endonuclease mapping. A plasmid with the

insertion in the correct orientation was called pαEG3 (Fig 2.7).

The three plasmids (pαEG1, pαEG2, pαEG3) could be distinguished from one another and from the parental pMFα8 by restriction endonuclease analysis. From Fig 2.9 it can be seen that an EcoRI restriction endonuclease digest distinguished the parental pMFα8 (lane 2) from pαEG1 (lane 3), pαEG2 (lane 4) and pαEG3 (lane 5). Furthermore, the absence of an EcoRI restriction site in pαEG3, distinguished pαEG3 (lane 5) from pαEG1 (lane 3) and pαEG2 (lane 4). The difference between pαEG1 (lane 7) and pαEG2 (lane 8) could be determined by a PstI restriction endonuclease digestion of the constructs.

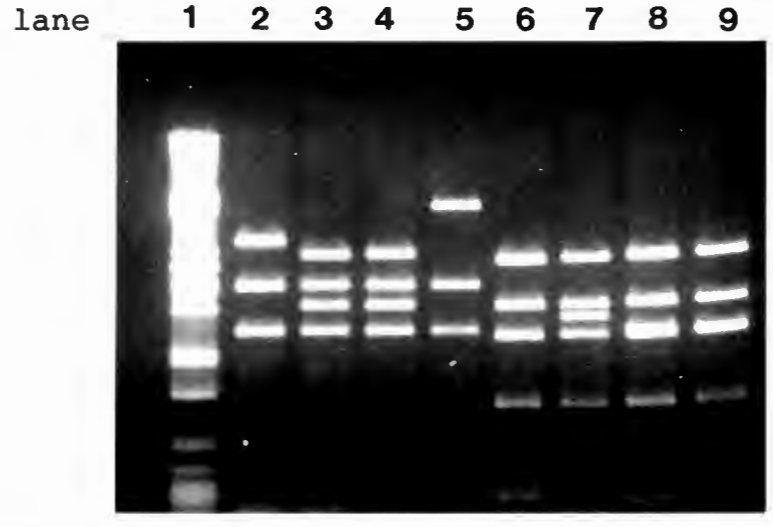


Fig 2.9. Agarose gel analysis of pMFα8, pαEG1, pαEG2 and pαEG3. pMFα8 (lane 2), pαEG1 (lane 3), pαEG2 (lane 4) and pαEG3 (lane 5) where digested with EcoRI endonuclease. Lanes 6 - 9 represent PstI endonuclease digest of pMFα8 (lane 6), pαEG1 (lane 7), pαEG2 (lane 8) and pαEG3 (lane 9). DNA standard (lane 1) was λ DNA digested with PstI endonuclease.

E. coli LK111 transfomants of these plasmids, pαEG1, pαEG2, pαEG3, were tested for the ability to produce an active endoglucanase when grown on LB agar containing CMC. The yeast promoter signals were not expressed in *E. coli* and no zones of hydrolysis were detected.

A large scale purified plasmid preparation (maxiprep) was made of the three α–factor plasmid constructs. This purified plasmid DNA was used to determine the nucleotide sequence at the junction of pMFα8 and insert DNA. A 20 bp primer, which was homologous to nucleotides 167 to 186 in the secretory signal of the *mfα* gene (Appendix D), was manufactured by Beckman Instruments (SA). The primer had the sequence:

5′- CC AAC AGC ACA AAT AAC GGG -3′

Sequencing confirmed the predicted nucleotide sequence implied by restriction analysis of the three constructs. Fig 2.10 shows the nucleotide sequence of pαEG3 at the junction of pMFα8 DNA and *C. acetobutylicum* DNA.

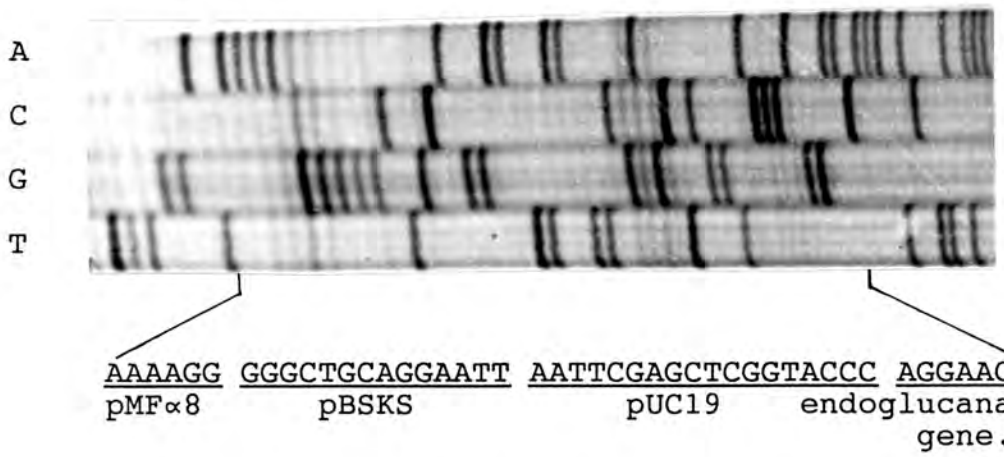


Fig 2.10. Nucleotide sequencing gel of pαEG3 at the junction of pMFα8 DNA and C. acetobutylicum DNA. The sequence can be read from the gel where lane 1 represents adenosine (A), lane 2 - cytosine (C), lane 3 - guanine (G) and lane 4 - thymidine (T) nucleotides. The nucleotide sequence given below the gel.

2.4 Discussion.

In this study two different strategies were used in the cloning of the endoglucanase gene into the yeast expression vector pMFα8.

The first method was devised to subclone the truncated endoglucanase gene from pHZ25 in the smallest possible fragment by minimizing DNA sequences from pUC19. Use of the KpnI restriction site from pUC19 would result in the addition of only two cytosine residues at the 5′ end, while the addition of BamHI linkers would result in the addition of two extra guanine residues to the 3′ end of the endoglucanase.

This method proved unsuccessful as no transformants were obtained from cells transformed with these ligations. This could be attributed to a number of factors : inefficient Mung Bean nuclease activity, poor ligations and poor restriction endonuclease digestion. The main drawback of this method is the lack of control at each stage in the strategy.

The second method was therefore designed to overcome these problems since each step in the construction could be examined by restriction endonuclease analysis. It was decided to construct three vectors, each with the endoglucanase gene fused in a different reading frame. If this gene were to be expressed in yeast, using the α–factor

promoter and secretory signals, then only one of the three constructs ought to be expressed.

The endoglucanase and endoxylanase genes from C. acetobutylicum do not have many useful restriction sites and there were a limited number of sites available in the multiple cloning site of pUC19. Convenient restriction sites were obtained by first subcloning the endoglucanase fragment into the the mcs of pBSKS. The three constructs, pαEG1, pαEG2 and pαEG3, could also be distinguished from one another and from the parental pMFα8 by restriction enzyme mapping (Fig 2.7). Sequencing of the junction of pMFα8 DNA and the three inserts confirmed the different reading frames of the three plasmids (Fig 2.10). In fact, only one of the three plasmids should produce an endoglucanase, the other two plasmids, although carrying the same DNA sequences, would be in the wrong reading frame for transcription, and would therefore provide negative controls for further experiments.

Chapter 3

Transformation of yeast and the expression of cloned genes.

3.0 Summary.

The plasmids pYEL300 and pYIX318, which contain a C. acetobutylicum xylanase encoding DNA fragment, were used to transform the yeast S. cerevisiae DBY746. Yeast transformants of these plasmids did not exhibit any xylanase activity. The plasmid pYEL100, which contains the C. acetobutylicum endoglucanase gene, including the bacterial promoter, and the plasmids pαEG1, pαEG2 and pαEG3, which each have the endoglucanase gene coding region fused to the α-factor gene promoter and secretion signals, in one of the three different reading frames, were used to transform yeast. Yeast transformants of pYEL100 did not exhibit any endoglucanase activity, whereas only the pαEG3 transformants of the three α-factor plasmids produced detectable endoglucanase activity. These results suggest that a yeast gene promoter is necessary for the expression of the endoglucanase and xylanase genes in yeast.

3.1 Introduction.

A number of yeast transformation methods have been developed but two methods are commonly used. The first method involves enzymatic removal of the cell wall thus producing spheroplasts (Beggs, 1978; Hinnen et al., 1978; Gerbaud et al., 1979; Sherman et al., 1986). The spheroplasts are stimulated to take up DNA by polyethylene glycol and calcium ion treatment. The cells are plated on isotonic plates and spontaneously regenerate cell walls. The second method, similar in principle to the E. coli method, uses alkali cations to induce transformation of intact yeast cells (Ito et al., 1983). Each method has certain advantages. Protoplasting, although relatively complex and time consuming, results in a high frequency of transformation of yeast cells by most plasmids. The LiCl protocol (Ito et al., 1983) has a lower transformation frequency, especially for plasmids with a 2 μm origin of replication, but is ideal for yeast cells that are resistant, or too sensitive, to lytic enzymes.

Whereas antibiotic resistance is efficiently used in bacterial transformation systems, selection of yeast transformants is dependent on the presence of a auxotrophic genetic marker carried on the plasmid, and a corresponding mutant strain (preferably a deletion mutant of the host). The leu2 gene described by Beggs (1978) and the trp1 gene (Tschumper and Carbon, 1980) are examples of auxotrophic markers. Various dominant selectable genes which confer

resistance to drugs are now available such as chloramphenicol acetyltransferase (Cohen et al., 1980) and aminoglycoside phosphotransferase-3'(I) (G418) (Jimenez and Davis, 1980) and are most useful for transforming strains which lack suitable auxotrophic mutations.

This study describes the cloning of two C. acetobutylicum genes, an endoglucanase gene and an endoxylanase gene, in yeast.

Cellulase gene expression in yeast

Cellulases are not produced by S. cerevisiae (Skipper et al., 1985). However a number of cellulase genes have been cloned and expressed in yeast.

β-glucosidases. Penttila et al. (1984) cloned a β-glucosidase gene from A. niger into S. cerevisiae. β-glucosidase activity was detected in yeast but the level of expression was too low to determine the cellular location and properties of the enzyme. Kohchi and Toh-e (1986) obtained expression of a C. pelliculosa β-glucosidase gene in S. cerevisiae. The expression of this gene was partially regulated by glucose in S. cerevisiae. β-glucosidase was secreted into the periplasmic space. Both β-glucosidases were expressed off their own promoters in the heterologous host. However, attempts to obtain efficient growth of the yeast strains on cellobiose using the expression of the

cellulases from A. niger (Penttila et al., 1984) and C. pelliculosa (Kohchi and Toh-e, 1986) were unsuccessful.

Exoglucanases. Penttila et al. (1988) cloned two different exoglucanase genes (celliobiohydrolases), cbh1 and cbh2, from T. reesei into S. cerevisiae. cDNAs of each gene were fused to the yeast pgk gene promoter. Both enzymes were efficiently secreted into the culture medium using their own secretion signal sequences. The estimated levels of CbhI in the culture medium was 1.6 μg/ml and for CbhII, about 10 μg/ml, when the yeast strains were grown in minimal media. Penttila et al. (1988) found that the levels of CbhII secreted, when the yeast was grown in rich medium, was 100 μg/ml. A C. fimi β-1,4-exoglucanase gene (cex) was fused to the S. cerevisiae mel1 (β-galactosidase) gene promoter and signal sequences (Curry et al., 1988). The exoglucanase gene was expressed and secreted as a 48.5 kD protein. The secreted fraction was 28% of the exoglucanase activity measured. The endoglucanase activity was substantially increased when the β-galactosidase signal peptide was spliced precisely to the mature exoglucanase protein (Curry et al., 1988).

Endoglucanases. The endoglucanase I gene from T. reesei was isolated and cloned into S. cerevisiae by Van Ardsell et al. (1987). Introns of higher eukaryotes are usually spliced in other eukaryotic hosts, however S. cerevisiae appears to be more stringent in its sequence specificity requirements and does not process the introns of foreign genes efficiently

(Langford et al., 1983). Therefore the introns of the endoglucanase gene were removed before the gene was fused to the yeast enolase gene promoter (eno1). The natural signal peptide of the endoglucanase was retained and efficient secretion (80%) of the recombinant enzyme into the culture medium was obtained. This enzyme was biologically active although it was extensively glycosylated with asparagine-linked sugars.

Penttila et al. (1987) obtained expression from the cDNA of two endo-β-1,4-glucanase genes (egl1 and egl3) from T. reesei in yeast. Both genes were under the control of the pgk gene promoter. EgI and EgIII were secreted into the growth medium in an active form. EgI was considerably larger and more heterogenous in size than the natural enzyme secreted from T. reesei, partly due to the difference in the extent of N-glycosylation. Although neither enzyme affected the growth rate of the respective strains, the authors noted that the yeast cells producing EgI and EgIII were larger and elongated compared to the control strain.

Skipper et al. (1985) fused cDNA of an endo-1,4-β-D-glucanase gene from C. fimi to the yeast adh1 gene promoter and the prepro-leader sequence of the K1 toxin gene. Fusion of the truncated cenA gene to the leader sequence of the preprotoxin gene sequence allowed for the expression of an active and secreted 50 kD endoglucanase from yeast. Wong et al. (1988) constructed a plasmid that contained the two C. fimi cellulase genes, cenA, and cex, in

tandem expression cartridges using the mel1 gene promoter and N-terminal signal peptide coding sequences. S. cerevisiae cells transformed with this plasmid co-expressed both the endoglucanase and exoglucanase as extracellular enzymes.

Hinchliffe and Box (1984) and Sacco et al. (1984) reported the cloning of endoglucanase genes from B. subtilis and C. thermocellum respectively, in yeast. Both enzymes were produced in an active form but were not secreted, presumably since they had bacterial signal sequences. Another gene, associated with cellulose degradation, that has been expressed and secreted by yeast host systems is an Aspergillus awamori glucoamylase (Innis et al., 1985).

In this study, the expression of an endoglucanase gene and xylanase gene from C. acetobutylicum were examined in yeast. Initially the C. acetobutylicum gene sequences that were cloned into yeast included their putative promoter sequences. From the above examples it would be possible that the endoglucanase and xylanase bacterial gene promoters may not mediate expression in yeast. It would therefore be necessary to replace the bacterial gene promoters with a yeast gene promoter. In this study, the yeast expression plasmid used was pMFα8, which contained gene promoter from the mating pheromone α-factor gene.

α–Factor processing.

S. cerevisiae can exist as any of three distinct cell types. Two mating cell types, a and α, can conjugate to form an aα diploid cell. The aα mating is initiated by the reciprocal exchange of diffusable a– and α–factor pheromones (Thorner, 1982; Herskowitz and Oshima, 1981). The pheromone released into the culture medium by cells is a polypeptide of 13 amino acid residues. However, it has been shown that the α–factor is initially synthesised as a substantially larger precursor protein, pre–pro–α–factor (pp–αF) (Kurjan and Herskowitz, 1982; Julius et al., 1983; Brake et al., 1984; Julius et al., 1984a). This precursor of the α–factor structural protein consists of 165 amino acids containing an N-terminal signal sequence of 22 hydrophobic amino acids, followed by an additional leader sequence of approximately 60 hydrophilic amino acids and then four identical tandem repeats of the mature pheromone sequence, each preceded by a short spacer octapeptide (Fig3.1) (see Appendix D for the nucleotide and amino acid sequences of the α–factor gene) (Kurjan and Herskowitz, 1982; Julius et al., 1983; Brake et al., 1984; Julius et al., 1984a). The precursor starts with ATG and ends with TAA immediately after the fourth α–factor sequence (Kurjan and Herskowitz, 1982). The ATG codon in position 1 is the only ATG codon that is in frame with the α–factor coding sequence and is presumed to be the initiation codon for the α–factor precursor. The mfα gene has the upstream sequence TATATAA (Kurjan and

implicated in processing yeast precursor proteins at pairs of basic amino acids (Julius et al., 1984b). The first processing event, catalysed by the kex2 gene product, occurs at the C-terminal side of -Lys-Arg- pairs of the spacer sequences (Julius et al., 1984b; Miyajima et al., 1985) (Fig 3.1).

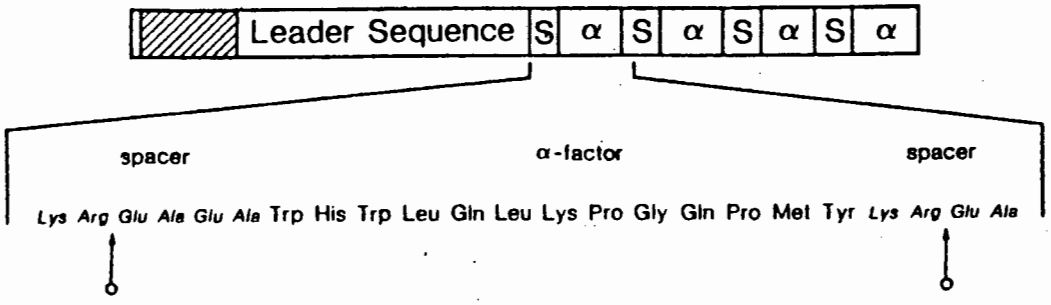


Fig 3.1. Structure of pre-pro-α-factor. Prepro-α-factor consists of N-terminal sequences of 20 amino acids. followed by a leader sequence of 60 amino acids and ending in four tandem repeats of mature α-factor (α) and spacer (S) (the first and part of the second spacers are indicated in italics). The amino acid sequence of the first α-factor spacer and part of the second spacer are shown. The kex2 product processing sites are indicated by the arrows (from Miyajima et al., 1985).

The carboxy terminus of the α-factors 1, 2 and 3 could then be trimmed by a carboxy peptidase, α-factor 4 does not require processing (Kurjan and Herskowitz, 1982).

A number of researchers have described the processing and secretion of heterologous proteins into the culture medium from yeast cells using the α-factor secretion system. Brake et al. (1984) investigated the expression, processing and secretion of hEgf from a chemically synthesised gene using the α-factor expression system. Emr et al. (1983) obtained

secretion of an invertase from yeast using an in frame fusion of α-factor pre-pro-leader sequence to the coding region of the *suc2* gene. Other proteins that have been secreted in *S. cerevisiae* using the α-factor expression system include β-endorphin, γ-interferon (Bitter *et al.*, 1984); mouse interleukin-2 (Miyajima *et al.*, 1985); murine and human granulocyte-macrophage colony stimulating factors (Miyajima *et al.*, 1986); the major envelope glycoprotein of Epstein Barr Virus (Schultz *et al.*, 1987a) and a synthetic human insulin-like growth factor (Bayne *et al.*, 1988).

The α-factor leader peptide has also been fused to the promoter of other yeast genes. Schultz *et al.* (1987b) fused the *gal10* gene promoter and α-mating factor pre-pro-leader to the structural gene Epstein Barr Virus glycoprotein. The *adh2* gene promoter was used by Price *et al.* (1987) for the expression of murine granulocyte-macrophage colony stimulating factor and bovine interleukin-2 genes in *S. cerevisiae*. Castanon *et al.* (1988) obtained human lysozyme expression from the fusion of the *adh1* gene promoter and α-factor leader sequence with the cDNA of the human lysozyme gene.

In general it appears that hybrid proteins are properly processed and the mature product secreted from yeast cells provided the first -Lys-Arg- processing site of the pre-pro-α-factor is retained in these fusion genes (Castanon *et al.*,

1988; Julius et al., 1984a). The yeast expression vector, pMFα8 (Miyajima et al., 1985), used in this study, contains the leader sequence which includes the coding region for the first -Lys-Arg- processing site which is cleaved by the kex2 gene product (see Fig 3.1 and Appendix C for a restriction endonuclease map of pMFα8).

3.2 Materials and Methods.

3.2.1 Yeast strains and plasmids. S. cerevisiae DBY746 (his3-1, leu2-3, leu2-112, ura3-52, trp1-289, cyh^r) was used as a recipient strain for yeast plasmids. This yeast strain was obtained from the strain collection in the Microbiology Department at the University of Cape Town. Plasmids pYEL100, pYEL300, pYIX318, pαEG1, pαEG2, pαEG3 were constructed as described in Chapter 2.

3.2.2 Media and buffers. All media and solutions not described in the text are listed in Appendix B.

3.2.3 Growth conditions. S. cerevisiae DBY746 was grown in YPD medium (Appendix B) (Sherman et al., 1986). S. cerevisiae DBY746 transformants were selected on SD agar (Appendix B) containing the relevant amino acids. Liquid culture minimal medium was SD broth plus amino acids and enriched with 2.5% YEP (Appendix B). Endoglucanase or xylanase activity of yeast transformants were detected by growning on SD plates (plus amino acids) containing 0.1% (w/v) CMC (medium viscosity, Sigma no. C4888, Degree of substitution is 0.7) or 0.1% (w/v) oat spelt xylan (Sigma x-0376, Lot 14F-0421). Zones of hydrolyses were detected using the Congo red staining technique. All yeast cultures were incubated at 30°C.

3.2.4 Yeast transformation. This method is based on the method described by Sherman et al. (1986) which is a modified version of the yeast transformation method described by Beggs (1978).

A culture of S. cerevisiae DBY746 was grown in 10 ml YPD for 24 h at 30°C. Three flasks containing 200 ml YPD were innoculated with 2, 10 and 25 µl of the yeast culture respectively. After 12 h growth at 30°C the cell number of each culture was determined microscopically. The culture with approximately 5 x 10^8 cells/ml was selected. The cells were pelleted by centrifugation for 5 min at 4000 x g and washed with sterile distilled water. The cells were resuspended in 20 ml SED (Appendix B), incubated for 10 min at 30°C, then pelleted by centrifuged for 5 min. After one wash in 20 ml Sorb (Appendix B), the cells were resuspended in 20 ml SCE (Appendix B). Zymolyase-100T (Seikagaku Kogyo Co. Ltd. Tokyo, Japan) (0.2 ml, 1 mg/ml) was added and the culture incubated at 30°C with gentle shaking for 1 h or until the yeast cells had formed spheroplasts. Spheroplast formation was monitored microscopically. The spheroplasts were centrifuged for 3 min at 700 x g, washed twice with 20 ml Sorb and once with 20 ml STC (Appendix B). The cells were resuspended in 1 ml STC and divided into 100 µl aliquots in 10 ml tubes. DNA (0.1 - 5 µg in 1 - 10 µl of TE buffer) was added to the spheroplasts and left at room

temperature for 10 min before 1 ml of PEG (4000) was added to each tube, mixed, and left for a further 10 min at room temperature. The cells were centrifuged at 1000 x g and resuspended in 150 µl SOS (Appendix B) and incubated at 30°C for 20 min. At this stage the cells were either plated or kept at 4°C for 1-3 d. Transformed cells (10 µl) were added to 5 ml of top agar held at 45°C, mixed briefly and poured onto pre-warmed plates (37°C). The plates were incubated at 30°C for 4 d. Transformation controls were cells only, on a plate containing all the amino acid requirements and on a plate lacking the selectable amino acid. In this study leucine, uracil and tryptophan auxotrophic markers were used.

3.2.5 Congo Red staining technique. This method is described in 2.2.12 except the Congo Red stain was buffered at pH 6.0 in 50 mM citrate-phosphate buffer (PC).

3.2.6 Isolation of protein from liquid culture medium. Methods were generally as described by Scopes (1982). The supernatant fraction from 300 ml yeast cultures was precipitated with 3 volumes of ice cold acetone at 4°C. The solution was stirred for 6 h. Precipitated protein was collected by centrifugation at 16000 x g for 10 min. The protein was resuspended in 2 ml of 50 mM PC buffer (pH 6). Samples (100 µl) were assayed for endoglucanase or xylanase

activity. The glucose content was determined using a Glucose Analyser 2 (Beckman Instruments S.A.).

3.2.7 Preparation of cell-free extracts from yeast. Yeast cells from 300 ml cultures were collected by centrifugation (6000 x g for 5 min). The cells were resuspended in 10 ml of 50 mM citrate-phosphate buffer (pH 6) before breaking with a French pressure cell press (American Instruments Co., Silver Spring, Maryland, U.S.A) (Hughes et al., 1971). Cell lysis was monitored microscopically. Samples (200 µl) were assayed for endoglucanase or xylanase activity. The glucose content of the cell-free extracts was determined using a Glucose Analyser 2 (Beckman Instruments, S.A.).

3.2.8 Enzyme assays. Enzyme activity was assayed by the release of glucose equivalents as detected by the dinitrosalicyclic acid reagent (DNS) for reducing sugars (Miller, 1959). A protein sample from the culture medium or cell-free extract was incubated with 0.5 ml 2% (w/v) CMC at 50°C for 30 min. DNS solution (Appendix B) (3 ml) was added and the samples mixed and placed in a boiling water bath for 5 min. The samples were centrifuged for 5 min at 5000 x g to remove any residual insoluble material. The samples were diluted 1/4 with water and the absorbance at 540 nm recorded. One unit of activity was defined as the amount of enzyme that releases 1 µmole glucose equivalents in 1 min. Specific activity was defined as units/mg protein.

3.2.9 Determination of protein concentrations. The Biuret reagent (Gornall et al., 1949) was used to determine the protein concentration of the cell-free extracts. Samples were diluted 1 : 10 with water and an equal volume Biuret reagent was added. The solutions were incubated at 37°C for 15 min to allow for the development of colour. The OD_{540} was measured and the protein concentration determined. BSA (Bovine Serum Albumin) was standard.

3.2.10 Isolation of plasmid DNA from yeast. A modified version of a small scale DNA isolation method reported by Keranen (1986) was used. A large colony (approximately 3 mm in diameter) was picked and resuspended in disruption buffer (Appendix B). Glass beads (0.45 mm in diameter) were added until just below the level of the liquid. The solution was vortexed for 1 min to disrupt the yeast cells. The sample was phenol extracted with an equal volume tris-buffered phenol (pH 8.0). The DNA was purified (2.2.4), resuspended in 20 μl TE and used to transform E. coli LK111 (2.2.11).

3.3 Results.

3.3.1 Transformation of yeast by the recombinant plasmids pYEL100, pYEL300 and pYIX318. The construction of plasmids pYEL100 (endoglucanase gene), pYEL300 and pYIX318 (xylanase gene) was described in Chapter 2. The C. acetobutylicum endoglucanase gene was cloned into pYEL100 and the xylanase gene was inserted into pYEL300 and pYIX318.

S. cerevisiae DBY746 was transformed by these plasmids and transformants of pYEL100 and pYEL300 were selected using complementation of the auxotrophic marker leu. The frequency of transformation of the yeast cells by these two episomal plasmids was approximately 2×10^4 colonies/μg DNA. pYIX300 yeast transformants were selected by complementation of the ura auxotrophic marker. The frequency of transformation by this integrating plasmid was approximately 8×10^2 colonies/μg DNA.

3.3.2 Endoglucanase and endoxylanase activity assays of S. cerevisiae DBY746 (pYEL100), DBY746 (pYEL300), and DBY746 (pYIX318).

Plate assay. S. cerevisiae DBY746 (pYEL100) was grown on SD agar containing selected amino acids and 0.1% (w/v) CMC. S. cerevisiae DBY746 (pYEL300) and DBY746 (pYIX318) were grown on the same medium which contained 0.1% (w/v) oat spelt xylan instead of CMC. The plates were stained using Congo red and examined for zones of hydrolysis. None of the

yeast strains produced zones of hydrolysis that could be attributed to endoglucanase or xylanase activity.

Reducing sugar assays of culture media. S. cerevisiae DBY746 (pYEL100), DBY746 (pYEL300) and DBY746 (pYIX318) were grown in minimal medium (SD plus amino acids) with 2.5% (v/v) enrichment with YEP for 40 h. Proteins precipitated from the culture medium were assayed for either endoglucanase or xylanase activity respectively, using the reducing sugar assay method (DNS assay). However, no endoglucanase or xylanase activity was detected, presumably indicating that the gene products were either not produced or secreted by yeast.

Reducing sugar assays of cell-free extracts. No endoglucanase or xylanase activity could be detected in cell-free extracts of S. cerevisiae DBY746 (pYEL100), DBY746 (pYIX318) and DBY746 (pYEL300) respectively. This suggested that neither the endoglucanase nor the xylanase enzymes were produced from their own promoters in yeast.

3.3.3 Isolation of plasmids pYEL100 and pYEL300 from yeast. Plasmid DNA isolated from S. cerevisiae DBY746 (pYEL100) and DBY746 (pYEL300) was used to transform E. coli LK111. A low transformation frequency of E. coli with these plasmid DNA preparations was obtained. This could be attributed to a poor yield of plasmids isolated, or to an inhibitory effect of cellular components that were present in the crude DNA preparation. Plate assays of the E. coli (pYEL100) and

E. coli (pYEL300) transformants showed endoglucanase and xylanase activity respectively. These plasmids (pYEL100 and pYEL300) did not appear to undergo any DNA rearrangements in the yeast host.

3.3.4 Transformation of yeast by the recombinant plasmids pαEG1, pαEG2 and pαEG3. The three pMFα8 derived plasmids, pαEG1, pαEG2 and pαEG3, that contain the endoglucanase encoding DNA fragment from C. acetobutylicum, each in one of the three different reading frames, transformed the yeast S. cerevisiae DBY746 with frequencies in the order of 8 x 10^4 colonies/μg DNA and were selected by complementation of the mutant trp gene of S. cerevisiae DBY746.

3.3.5 Endoglucanase activity assays of S. cerevisiae DBY746 (pαEG1), DBY746 (pαEG2) and DBY746 (pαEG3).

Plate assay. S. cerevisiae DBY746 (pαEG1), DBY746 (pαEG2) and DBY746 (pαEG3) were grown on SD agar containing the relevant amino acids and 0.1% (w/v) CMC. No zones were produced when the plates were stained using Congo red.

Reducing sugar assays of the liquid culture medium. S. cerevisiae DBY746 (pαEG1), DBY746 (pαEG2) and DBY746 (pαEG3) were grown in minimal medium with 2.5% (v/v) enrichment for 40 h. Protein was isolated from the supernatant fraction and was measured for endoglucanase activity using the DNS reducing sugar assay (Table 3.1).

Table 3.1. Endoglucanase activity in culture media of S. cerevisiae DBY746, DBY746 (pαEG1), DBY746 (pαEG2) and DBY746 (pαEG3) strains. Endoglucanase activity was determined by reducing sugar assays. The results presented are an average of assays done in triplicate.

Strain	Endoglucanase activity (U/ml)
S. cerevisiae	4.216×10^{-6}
S. cerevisiae (pαEG1)	3.961×10^{-6}
S. cerevisiae (pαEG2)	4.216×10^{-6}
S. cerevisiae (pαEG3)	1.431×10^{-5}

S. cerevisiae DBY746 itself does not have an endoglucanase gene, therefore the activity value of 4.216×10^{-6} can be attributed to a background effect. The activity readings for the strains DBY746 (pαEG1) and DBY746 (pαEG2) were of the same order as S. cerevisiae DBY746, indicating that the endoglucanase gene was not expressed in either DBY746 (pαEG1) or DBY746 (pαEG2). The activity from DBY746 (pαEG3) was greater than the activity obtained from S. cerevisiae DBY746, DBY746 (pαEG1) and DBY746 (pαEG2). This indicated that the endoglucanase gene, contained on the plasmid pαEG3, was expressed in the yeast and the gene product secreted into the culture medium.

Reducing sugar assays of cell-free extracts. Cell-free extracts of the yeast strains DBY746 (pαEG1), DBY746 (pαEG2) and DBY746 (pαEG3) were prepared and endoglucanase activity was determined. Table 3.2 presents the specific activities of the cell-free extracts.

Table 3.2. Endoglucanase specific activity from cell-free extracts of S. cerevisiae DBY746, DBY746 (pαEG1), DBY746 (pαEG2) and DBY746 (pαEG3) strains. Endoglucanase activity was determined by the reducing sugar assay.

Strain	Endoglucanase specific activity (U/mg protein)
S. cerevisiae	0.0032
S. cerevisiae (pαEG1)	0.0033
S. cerevisiae (pαEG2)	0.0040
S. cerevisiae (pαEG3)	0.0131

The results presented in Table 3.2 show that the endoglucanase specific activity of the cell-free extract from S. cerevisiae DBY746 (pαEG3) is greater than endoglucanase specific activity for the other yeast cell-free extracts. This increase, presumably due to endoglucanase gene expression in the yeast, is of similar magnitude in both the supernatant fraction and the cell-free extract (Table 3.1 and Table 3.2).

An experiment on S. cerevisiae DBY746 and S. cerevisiae DBY746 (pαEG3), to examine the amount of endoglucanase produced by DBY746 (pαEG3) according to time, was performed. The cells from five parallel cultures of each strain, grown in minimal medium with 2.5% (v/v) enrichment, were harvested at certain time intervals and cell-free extracts prepared. The specific activity of the endoglucanase in these samples was compared (Fig 3.2).

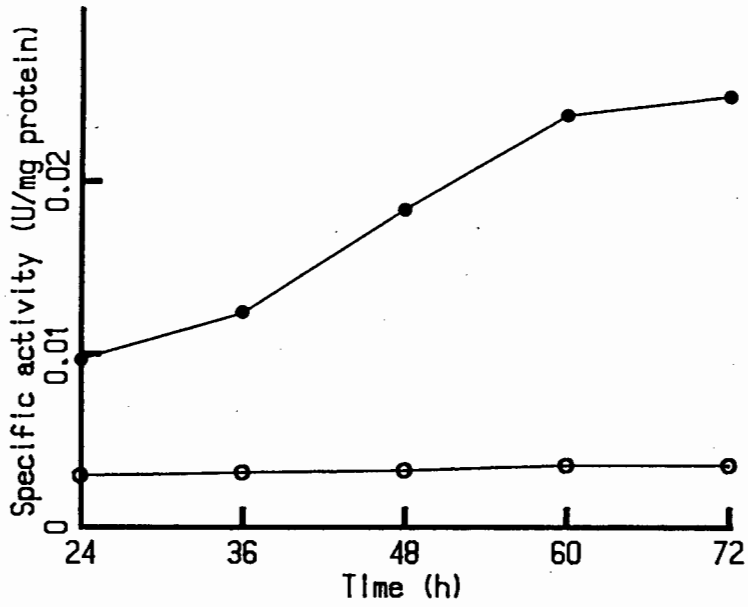


Fig 3.2 Endoglucanase activity profiles of cell-free extracts from S. cerevisiae DBY746 (○) and S. cerevisiae DBY746 (pαEG3) (●) taken over a 72 h period.

The effect of time on the amount of endoglucanase activity produced by S. cerevisiae DBY746 (pαEG3) was determined (Fig 3.2). There was a linear increase in endoglucanase activity from cell-free extracts of S. cerevisiae DBY746 (pαEG3) with respect to time. The endoglucanase specific activity from the cell-free extracts of the control strain, S. cerevisiae DBY746, was constant over all the time periods.

3.4 Discussion.

The S. cerevisiae DBY746 host showed no detectable endoglucanase activity. Yeast transformants of the plasmid pYEL100 did not exhibit any intracellular or extracellular endoglucanase activity against CMC. Similarly yeast cells harbouring the plasmids pYEL300 and pYIX318, showed no xylanase activity. Since the plasmids contained the entire bacterial gene sequences (promoter, initiation site and terminator), this indicates that the bacterial genes promoter and signal sequences were not recognised in the yeast host.

In this study the mating pheromone α-factor gene promoter and secretion signals mediated the expression and secretion of the C. acetobutylicum endoglucanase from the yeast cell. The endoglucanase gene, lacking its own promoter and initiation site, was fused in frame, to the 3′ end of α-factor pre-pro-leader secretion signal in the plasmid designated pαEG3.

Relatively low levels of endoglucanase activity were obtained from this endoglucanase producing yeast. A number of factors can account for this phenomenon.

Yeast heterologous gene expression is found to be considerably lower than homologous gene expression by many researchers. Usually the hybrid protein constitutes less than 5% of the total cell protein although the copy number

of the gene in the cell is high. The low levels of endoglucanase expression indicate that the amount of protein could be of this order.

It is possible that structural features and gene sequence play a significant role in the expression of the C. acetobutylicum DNA in yeast. The presence of DASs (discussed in Chapter 1) in the coding region of a yeast gene have been implicated in yeast gene expression. The insertion of heterologous gene sequences in place of the yeast coding sequences removes this possible enhancer region. The mfα1 gene promoter used in this study may or may not require the presence of a DAS for its efficient transcription.

The mfα1 gene is not a highly expressed yeast gene. It is usually expressed in the stationary growth phase. This means that any gene fused to the mfα1 promoter will only be expressed late in the growth of the culture. This could imply that high levels of endoglucanase would not be attained before late stationary phase of the culture.

Curry et al. (1988) noted that the removal of non-coding DNA sequences, derived from non-yeast genes, between the yeast gene promoter (mel1) and the mature coding sequences of the exoglucanase gene from C. fimi increased the level of expression of the foreign gene in the yeast host. In this study there were 20 extra amino acids (60 nucleotides) between pMFα8 α-factor leader peptide and the presumed

processing site of the C. acetobutylicum endoglucanase. The amino acid sequence -A-E-A- at positions 36-38 in the putative leader peptide of the endoglucanase gene product has a similar sequence to the cleavage site of other proteins isolated from Gram positive bacteria (Zappe, 1988). The removal of these excess nucleotides, derived from the subcloning strategy and endoglucanase gene leader sequence, may result in an increase in endoglucanase gene expression.

Smith et al. (1985) found the integration of the calf prochymosin gene fused to the suc2 gene promoter produced a high level of expression relative to the low copy number of the gene in the cell. This had the advantage of being a stable event as well as producing a high level of expression. This, however, could be a phenomenon specific to the calf prochymosin gene and may not produce an increase in endoglucanase expression if the fusion gene was inserted into the chromosome.

Physiological factors may also be implicated in the low gene expression. The optimum assay conditions for the endoglucanase from yeast may not be the same as the requirements of the enzyme when characterised in E. coli. Zappe (1988) suggested the possibility that the endoglucanase from C. acetobutylicum P262 could be modified in some way in E. coli, to exhibit different pH and temperature optima when compared to its natural host. The same suggestion could apply to the endoglucanase when expressed in the yeast system.

Since the endoglucanase is secreted by the α–factor secretion pathway and is not the product of a highly expressed gene, the enzyme may not reach levels high enough to inhibit cell growth. This system is ideal when the protein produced is known to be harmful to the cell. It is not known whether the expression of the endoglucanase gene had a detrimental effect on the growth of the yeast culture.

Another alternative is the accumulation of endoglucanase intracellularly, possibly as a result of slow inefficient secretion of the enzyme or the protein being trapped in the ER. This phenomenon was reported by Emr et al. (1984) for the E. coli β–lactamase protein produced in S. cerevisiae.

The level of endoglucanase in the yeast cells increased linearly with time up to 60 h. This could be the effect of increased expression of the endoglucanase gene off the mfα1 gene promoter as the culture reaches late log phase and the mfα1 gene promoter activity increases.

The expression of a foreign protein in yeast may change the morphology of the yeast cell. Penttila et al. (1987) noticed that yeast cells that expressed cloned T. reesei endoglucanase genes were larger and more irregular in shape than the yeast control cells. This phenomenon was present at all stages in the growth cycle. On the basis of this information, the morphology of yeast cells secreting the C. acetobutylicum endoglucanase were examined

microscopically. The cells appeared slightly larger than the control strain. However, this result could not be considered conclusive since a comprehensive analysis was not conducted.

These results show the expression and secretion of an endoglucanase gene from C. acetobutylicum in the yeast S. cerevisiae DBY746 using the promoter and secretion signals of the mfα1 gene promoter.

Chapter 4

General Discussion.

The yeast S. cerevisiae is used extensively in many biotechnical processes. Lignocellulose material and its components of cellulose and hemicellulose provide a source of inexpensive substrates for fermentation. Pretreatment procedures are necessary for the cellulose and hemicellulose to be available to the enzymatic action of cellulases. Certain microorganisms are able to ferment these complex substrates. S. cerevisiae, however is not an example of a cellulolytic organism.

A yeast that is able to utilize a greater percentage of raw or partially degraded plant material could make the industrial process of ethanol production more cost effective. Cellulases from cellulolytic organisms such as T. reesei (Penttila et al.,1987; van Arsdell et al.,1988; Penttila et al.,1988) and C. fimi (Skipper et al.,1985; Curry et al.,1988; Wong et al.,1988) have been used to confer exoglucanase or endoglucanase activity to yeast. The aim of this study was to introduce two genes (an endoglucanase gene and a xylanase gene) from C. acetobutylicum, into yeast, with the aim of low level constitutive expression of these genes in yeast to allow the breakdown of some components of cellulose and hemicellulose.

A number of different approaches were used to subclone xylanase and endoglucanase encoding DNA fragments from C. acetobutylicum into a variety of yeast-E. coli shuttle vectors. These clones were constructed in E. coli, since techniques for easier manipulation have been developed for the E. coli host system. Initially the entire bacterial genes were inserted into a shuttle vector and transfered to a yeast host.

An active endoglucanase or xylanase could not be detected in yeast strains carrying recombinant plasmids which contained the entire bacterial gene (promoter, coding sequences and termination signals). A similar phenomenon was reported by Skipper et al. (1985) for the endoglucanase from C. fimi. It was assumed that transcriptional initiation failed to occur as the yeast RNA polymerase would be unable to recognize the promoter and initiation sequences of these bacterial genes. To overcome this problem an, endoglucanase encoding DNA fragment lacking a promoter, initiation codon and part of the putative leader peptide sequence was manipulated using sites from the multiple cloning cassettes of various bacterial plasmids. This resulted in a plasmid that contained an in-frame fusion of the α-mating factor promoter and secretion signals with the endoglucanase encoding DNA fragment from C. acetobutylicum, in the yeast expression vector pMFα8 (Miyajima et al.,1985). The replacement of the bacterial promoter by a yeast promoter resulted in the production of an active endoglucanase which was also secreted into the culture medium.

Although the aim of this study was to produce a yeast strain that could utilize some components of cellulose, and not the production of a cellulolytic organism, the levels of endoglucanase expression were relatively low. It might be desirable to increase this activity. This can be achieved by a number of ways.

The mfα1 promoter used in this study possesses the secretory signals that allow for the secretion of protein into the culture medium (Brake et al., 1984) and not just into the periplasmic space as is the case with the suc2 and pho5 secretion systems (Perlman and Halvorsen, 1983). The fusion of a known highly expressed yeast gene promoter such as the pgk gene promoter (Dobson et al., 1982a) could produce a highly expressed heterologous protein that is efficiently secreted. The highly expressed adh1 gene promoter (Castanon et al., 1988) and gal10 gene promoter (Shultz et al.,1987b) have been fused to the α-factor secretion sequences to mediate efficient secretion of the gene products into the culture medium. The adh2 gene promoter has also been fused to the α-factor secretion signals. The adh2 promoter is an example of an inducible promoter that can be used to regulate the expression of foreign genes in yeast whose product could have a toxic effect on the host cell.

The fusion of the α-factor promoter to the endoglucanase encoding DNA fragment resulted in the addition of 60 extra nucleotides to the prepro-leader sequence of the

heterologous gene. These nucleotides translated to amino acids are compared below.

```
                  Kex2
α-factor  : - K R E A E A -

                  Kex2
pMFα8     : - K R P -

                  Kex2
pαEG3     : - K R G L Q E L I R A R Y P G T N T Y K A E A -
            pMFα8  pBSKS      pUC19     C. acetobutylicum
```

The Kex2 endopeptidase Lys-Arg processing site is retained in the protein encoded by the endoglucanase gene in the plasmid pαEG3. This suggests that the α-factor pre-pro-leader sequence would be removed by the normal α-factor processing system. However, the 20 additional amino acids after the Kex2 processing site may or may not be removed by the yeast processing mechanism. If not removed, these amino acids could adversely effect the activity of the protein by inhibiting the progress of the protein in the secretary pathway, preventing the protein taking up the correct conformation required for full efficiency or making the protein more susceptible to attack by the yeast host proteases. It would be of interest to fuse the α-factor prepro-leader sequences precisely to the mature coding region of the endoglucanase in order to determine whether the presence or absence of the amino acid sequence of the leader peptide region is critical for expression.

The yields of endoglucanase could be improved using alternative host strains. S. cerevisiae DBY746 was used as

a plasmid recipient strain as it has a fairly large variety of auxotrophic markers, including the trp marker, that is complemented by the trp1 gene on the pMFα8 plasmid, and the α-mating factor genotype. Alternative host strains could be used to investigate the effect of strain differences on the expression of the endoglucanase gene carried on the plasmid pαEG3, on the condition that they have the α-mating type. Hyper-secreting yeast mutants can be used to boost the secretion of the endoglucanase into the culture medium.

Another factor to take into consideration is the suitability of the plasmid used in conjunction with the purpose for which the yeast is required. In general, high copy number plasmids, such as 2 μm derived plasmids, should allow for a higher level of expression of a heterologous gene in yeast as a function of copy number. This is not always the case as discovered by Smith et al. (1985) (discussed in Chapter 1). Integrating plasmids carrying a foreign gene are usually present at a low copy number (approximately one per cell). This is advantageous when a homogeneous yeast culture is required for particular purposes. In most cases, multicopy plasmids require a fairly strict selection procedure. It would be of interest to determine the levels of endoglucanase expression using an integrating plasmid constructed using the endoglucanase encoding DNA fragment attached to the α-factor promoter sequences.

Yeast is essentially an organism that is involved in fermentation. The aim is not to change yeast into a

cellulolytic organism but rather to increase the substrate range of yeast. Low constitutive expression of cellulase and hemicellulase genes could facilitate the degradation of a small fraction of the substrate material not usually fermented. This would result in an increase in the carbon source available for the production of ethanol and hence result in a more cost effective process.

Appendix A

Maintenance of bacterial and yeast strains.

A.1. *E. coli*

Stock cultures of *E. coli* LK111 and HB101 were maintained at -70°C. Overnight cultures (5 ml LB medium containg Ap) were divided into 500 µl aliquots. Sterile glycerol was added (15% v/v final concentration) and the cultures stored at -70°C. *E. coli* strains were initially obtained from the Microbiology Department, University of Cape Town, culture collection

A.2. *S. cerevisiae*

S. cerevisiae DBY746 from an overnight culture (5 ml YPD medium) was aliquoted into 500 µl ammounts. Sterile glycerol (15% v/v final concentration) was added and the culture stored at -70°C. *S. cerevisiae* DBY746 was obtained from the culture collection of the Microbiology Department, University of Cape Town.

Appendix B

Media, Buffers and Solutions

B.1. Bacterial media

B.1.1. Luria-Bertani medium (LB)

Bacto tryptone	10 g
Yeast extract	5 g
NaCl	5 g
Distilled water	1000 ml

Solid media contained 1.5% (w/v) agar

B.1.2. YT medium (x2)

Bacto tryptone	16 g
Yeast extract	10 g
NaCl	5 g
Distilled water	1000 ml

B.2 Bacterial media additives

Media were cooled to 50°C before the addition of ampicillin, X-gal or IPTG.

B.2.1. Ampicillin

Ampicillin (sodium salt; Sigma)	1 g
Distilled water	10 g

The solution was filter sterilized and stored in 1 ml aliquotes at -20°C.

B.2.2. IPTG (isopropyl-β-D-thio-galactopyranoside)

IPTG (100 mM)	23.8 g
Distilled water	1 ml

The solution was stored in aliquots at -70°C.

B.2.3. X-gal (5-bromo-4-chloro-3-indolyl-β-galactoside)

X-gal (2%,w/v)	0.2 g
Dimethylformamide	10 ml

The solution was stored at -70°C.

B.3. Yeast media

B.3.1. YPD

Bacto yeast extract	10 g
Bacto peptone	20 g
Dextrose	20 g
Distilled water	1000 ml

Solid media contained 2% (w/v) agar.

B.3.2. SD

Bacto yeast nitrogen base w/o amino acids	6.67 g
Dextrose	20 g
Distilled water	1000 ml

Solid media contained 2% (w/v) agar.

B.4. Yeast media additives

B.4.1. Yeast nitrogen base w/o amino acids (YNB) (x20)

Bacto YNB	13.5 g
Distilled water	100 ml

The solution was stored in silver foil at 4°C.

B.4.2. Amino acid stocks.

Solutions were stored at 4°C.

Stock solution	conc.	per 100 ml
L-leucine	3.6 mg/ml	0.8 ml
L-histidine	2.4 mg/ml	0.8 ml
Uracil	2.4 mg/ml	0.8 ml
L-tryptophan	2.4 mg/ml	0.8 ml
Adenine	1.2 mg/ml	1.8 ml

B.5. Buffers and solutions

B.5.1. Biuret reagent (Gornall et al. 1949)

Copper sulphate ($CuSO_4.5H_2O$)	1.50 g
Sodium potassium tartrate ($NaKC_4H_4O_6.4H_2O$)	6 g
Distilled water	500 ml
NaOH (10%)	300 ml

B.5.2. DNA polymerase I buffer (Klenow). The buffer was made according to the following table and stored at -20°C.

Stock solution	final conc.	per 100 ml
Tris-Cl (1 M, pH 7.6)	0.1 M	1 ml
$MgCl_2$ (1 M)	60 mM	1 ml
NaCl (5 M)	0.5 M	1 ml
2-mercaptoethanol	0.7 M	50 µl
Distilled water		6.95 ml

B.5.3. DNA sample loading solution (x6)

Bromophenol blue	0.25 g
Sucrose	40 g
Distilled water	to 100 ml

The solution was stored at 4°C.

B.5.4. Dinitrosalicylic acid solution (DNS) (Miller et al. 1959)

3,5 Dinitrosalic acid	10.6 g
NaOH	19.8 g
Rochelle Salts (Na K Tartrate)	306 g
Phenol	7.6 ml
Na-meta bisulphite	8.3 g
Distilled water	1416 ml

Dinitrosalicylic acid, NaOH and Rochelle Salts were dissolved in water before adding the other constituents. Phenol is melted at 50°C. The end-point of a 3 ml sample was established using 5-6 ml 0.1 M HCl. If less HCl was required, the solid NaOH was added to the DNS solution at a rate of 2 g/ml of HCl less than 5 ml. DNS solution was stored in a dark bottle under N_2.

B.5.5. Ethidium bromide solution (2.7-diamino-10-ethyl-9-phenol-phenanthridinium bromide) (EtBr). A solution of 10 mg/ml was made in distilled water and stored in a dark bottle.

B.5.6. Isopropanol solution, salt saturated. Isopropanol was saturated with aqueous 5 M NaCl, 10 mM Tris-Cl and 1 mM EDTA (pH 8.5) (Maniatis et al., 1982).

B.5.7. Ligase dilution buffer. The buffer was made according to the following table and stored at -20°C.

Stock solution	final conc.	per 10 ml
Tris-Cl (1 M, pH 7.6)	20 mM	0.2 ml
EDTA (0.5 M, pH 8.0)	1 mM	2 µl
DTT (0.5 M)	5 mM	10 µl
KCl (1 M)	60 mM	0.6 ml
Glycerol	44% (w/v)	4.4 ml
Distilled water		4.788 ml

B.5.8. Ligation buffer (x10). The buffer was made according to the following table and stored at -20°C.

Stock solution	final conc.	per ml
Tris-Cl (1 M, pH 7.6)	66 mM	0.66 ml
$MgCl_2$ (1 M)	6 mM	66 µl
ATP (0.1 M)	1 mM	0.1 ml
DTT	0.1 M	15.4 mg
Distilled water		0.174 ml

B.5.9. Linker kinase buffer (x10). The buffer was made according to the following table and stored at -20°C.

Stock solution	final conc.	per ml
Tris-Cl (1 M, pH 7.6)	0.66 M	0.66 ml
ATP (1 M)	10 mM	0.1 ml
Spermadine (1 M)	10 mM	0.1 ml
$MgCl_2$ (1 M)	0.1 M	0.1 ml
DTT (1 M)	150 mM	150 µl
Gelatin or BSA	2 mg/ml	2 mg

B.5.10. Mung Bean nuclease buffer (x5). The buffer was made according to the following table and stored at -20°C.

Stock solution	final conc.	per ml
NaAc (1 M, pH 5.0)	150 mM	0.15 ml
NaCl (1 M)	250 mM	0.25 ml
$ZnCl_2$ (1 M)	5 mM	5 µl
Glycerol	-	25%

B.5.11. Mung Bean nuclease dilution buffer (x10). The buffer was made according to the following table and stored at -20°C.

Stock solution	final conc.	per ml
NaAc (1 M, pH 5.0)	0.1 M	0.1 ml
ZnAc (1 M)	1 mM	1 µl
Cysteine (1 M)	10 mM	10 µl
Triton X-100	-	0.1%
Glycerol	-	50%

B.5.12. Phenol (buffer saturated). Phenol (200 g, Merck) was melted at 65°C and 0.34 g of 8-hyroxyquinoline was added. The phenol was extracted three times with 1 M Tris-Cl (pH 8.0) or until the pH of the aqueous phase was approximately pH 7.6. The phenol was stored under 0.1 M Tris-Cl (pH 8.0) at -20°C.

B.5.13. Restriction enzyme core buffers (x10)

Stock solution	final concentration
Tris-Cl (1 M, pH 7.9)	0.1 M
$MgCl_2$ (1 M)	0.1 M
DTT (0.5 M)	10 mM
BSA (10 mg/ml)	1 mg/ml
Glycerol	44% (v/v)
NaCl	0, 50, 100 or 150 mM

The buffers were made using the following table and stored at -20°C.

	Salt concentration (mM)			
Stock solution	0	50	100	150
Tris-Cl (pH 7.9)	1 ml	1 ml	1 ml	1 ml
$MgCl_2$	1 ml	1 ml	1 ml	1 ml
DTT	0.2 ml	0.2 ml	0.2 ml	0.2 ml
BSA	1 ml	1 ml	1 ml	1 ml
Glycerol	4.4 ml	4.4 ml	4.4 ml	4.4 ml
Sterile water	2.4 ml	1.4 ml	0.4 ml	2.4 ml
NaCl (5 M)	-	1 ml	2 ml	87.7 mg

B.5.14. Restriction enzyme dilution buffer. The buffer was made according to the following table and stored at -20°C.

Stock solution	final conc.	per 10 ml
Tris-HCl (1 M, pH 7.5)	10 mM	0.1 ml
NaCl (5 M)	50 mM	0.1 ml
Distilled water	-	5.3 ml

This solution was filter sterilized and then the following constituents added:

2-mercaptoethanol	10 mM	7 µl
Gelatin (10 mg/ml)	100 µg/ml	0.1 ml
Glycerol	44% (v/v)	4.4 ml

Sma1 restriction endonuclease buffer (x10)

Stock solution	final conc.	per 10 ml
Tris-Cl (1 M, pH 8.0)	0.1 M	1 ml
KCl (1 M)	0.2 M	1 ml
$MgCl_2$ (1 M)	0.1 M	1 ml
DTT (0.5 M)	10 mM	0.2 ml
Glycerol	44% (v/v)	4.4 ml
Distilled water	-	1.4 ml

B.5.15. TE (Tris-EDTA) buffer (x100)

Tris-Cl (pH 7.6)	121 g
EDTA (0.5 M, pH 8.0)	200 ml
Distilled water	to 1000 ml

The buffer was autoclaved before use and appropriately diluted with sterile distilled water.

B.5.16. Tris-acetate buffer (x50)

Tris base	242 g
Acetic acid	57.1 ml
EDTA (0.5 M, pH 8.0)	100 ml
Distilled water	to 1000 ml

B.5.17. Yeast transformation buffers and solutions

B.5.17.1. PEG

PEG 4000	4.0 g
$CaCl_2$ (1 M)	0.2 ml
Tris-Cl (1 M, pH 7.4)	0.2 ml
Distilled water	to 20 ml

The solution was filter sterilized.

B.5.17.2. SCE

Sorbitol (1 M)	49.5 ml
Sodium citrate (pH 5.8)	1.47 g
EDTA (1 M)	0.5 ml

B.5.17.3. SED

Sorbitol (1 M)	18.5 ml
EDTA (1 M, pH 8.0)	0.5 ml
Dithiotheitol* (DTT) (1 M)	1 ml

Filter sterilize solution. Add DTT immediately before use.

*DTT was stored at -20°C.

B.5.17.4. Sorb

Sorbitol (1 M)	36.43 g
Distilled water	to 200 ml

B.5.17.5. SOS

Sorbitol (2 M)	10 ml
YEP	6.7 ml
$CaCl_2$ (1 M)	0.13 ml
Amino acid to be selected	27 µl
Distilled water	3.17 ml

The solution was filter sterilized.

B.5.17.6. STC

Sorbitol (1 M)	98 ml
$CaCl_2$ (1 M)	1 ml
Tris-Cl (1 M, pH 7.5)	1 ml

B.5.17.7. Top agar

Sorbitol	18.2 g
Agar	2.0 g
YNB (x20 stock)*	5.0 ml
Glucose	2.0 g
Adenine*	0.1 ml
Amino acids as required*	0.4 ml
Distilled water	to 100 ml

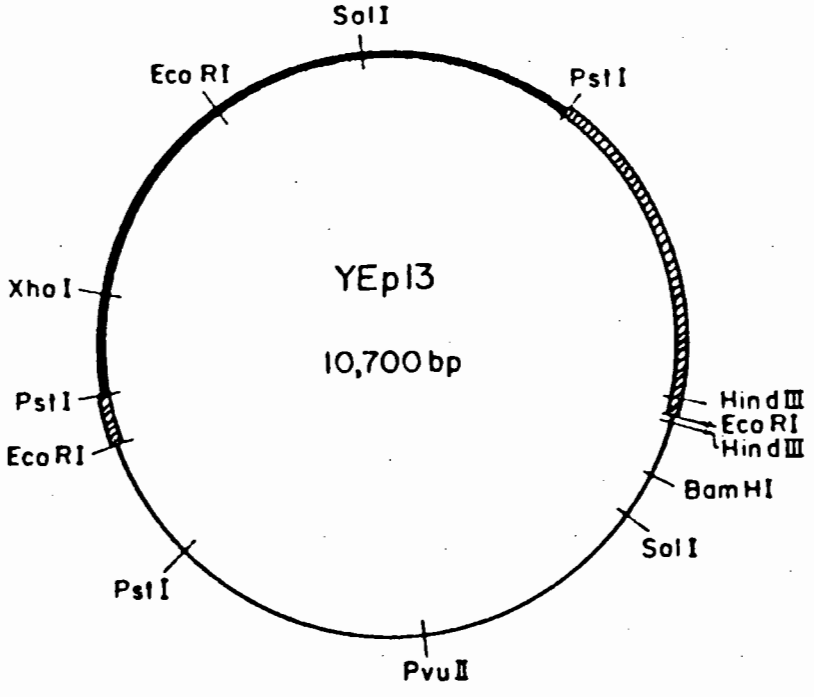
*add after autoclaving

B.5.17.8. YEP

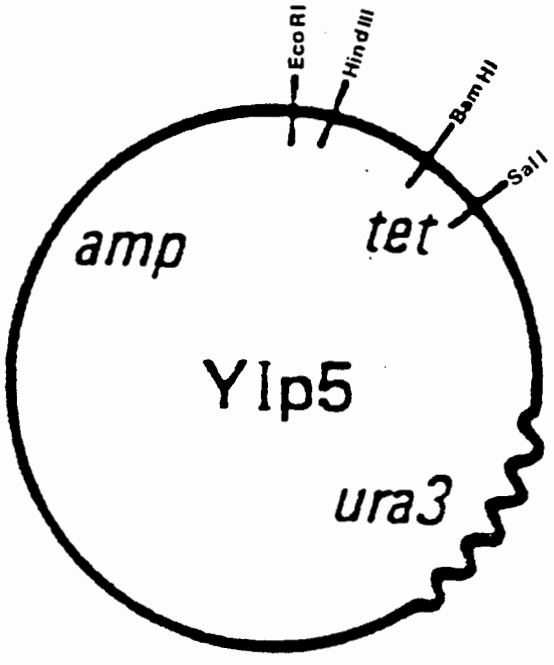
Bacto yeast extract	0.5 g
Bacto peptone	1.0 g
Distilled water	50 ml

Appendix C

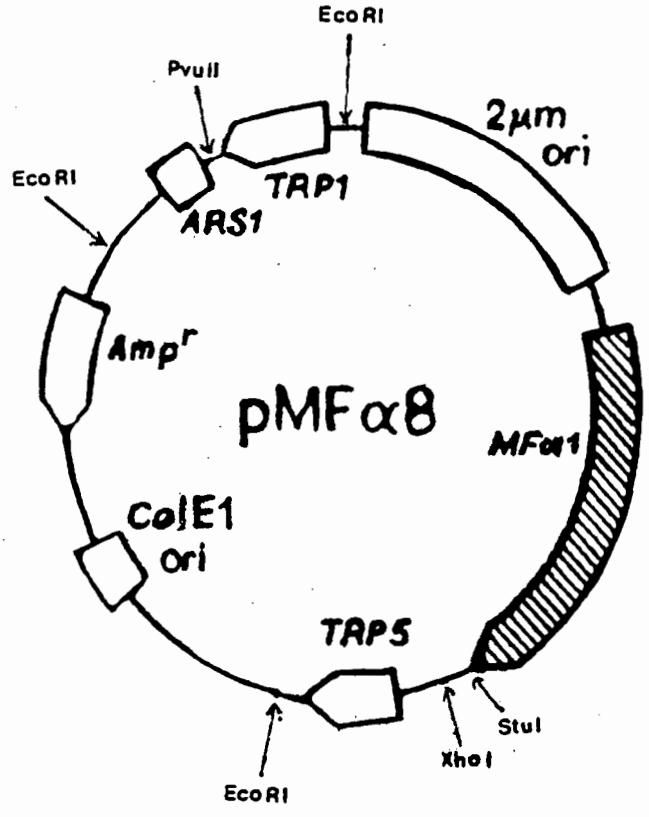
Vector endonuclease restriction maps



C.1 Restriction endonuclease map of YEp13 (Broach et al., 1978). The line line indicates pBR322 sequences, the thick line sequences derived from the plasmid pYeleu10, and the hatched region 2 μm sequences (from Broach et al., 1978).



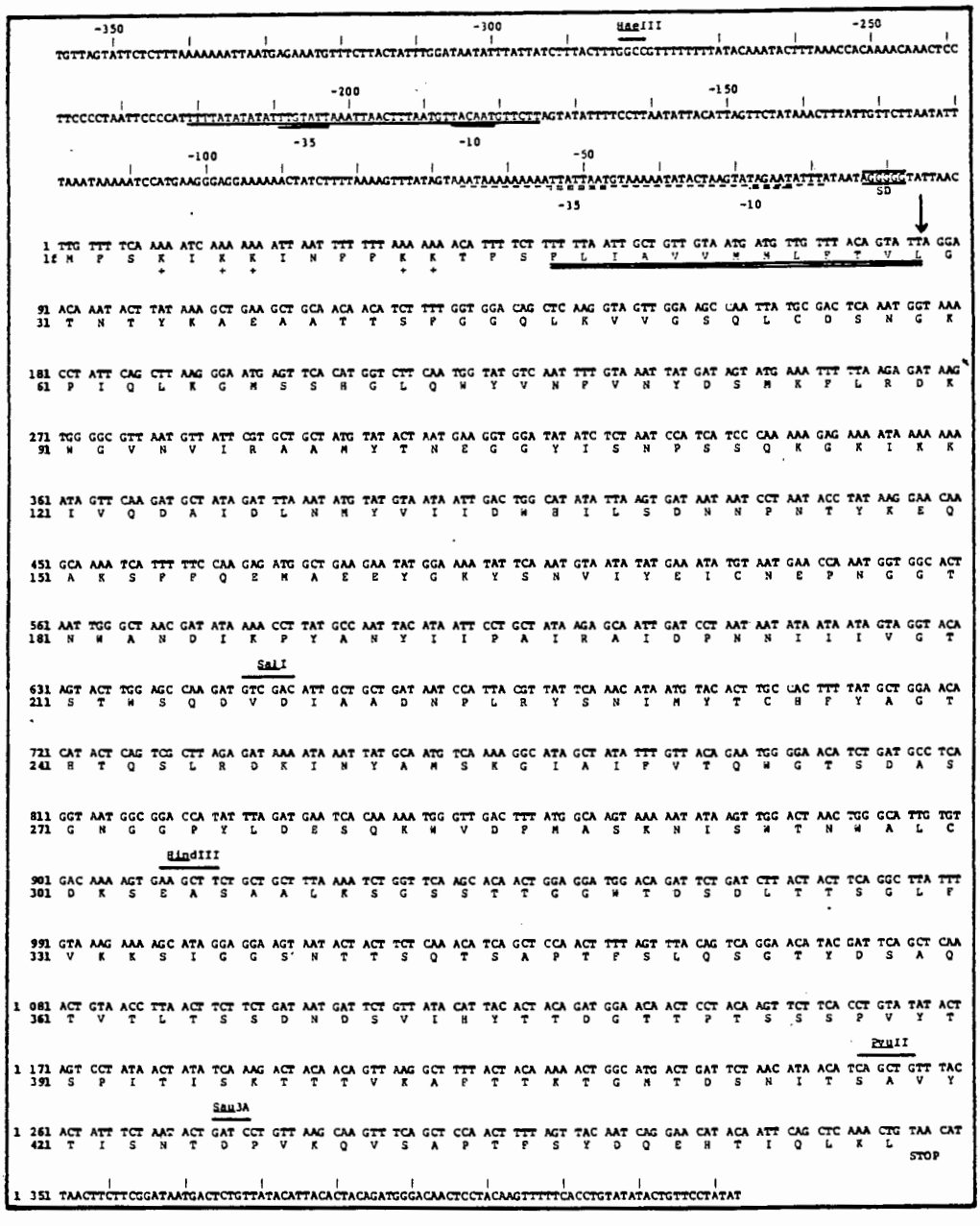
C.2 Restriction endonuclease map of YIp5 (Struhl et al., 1979). The thin line represents pBR322 DNA and the wavy line yeast chromosomal DNA (from Struhl et al., 1979).



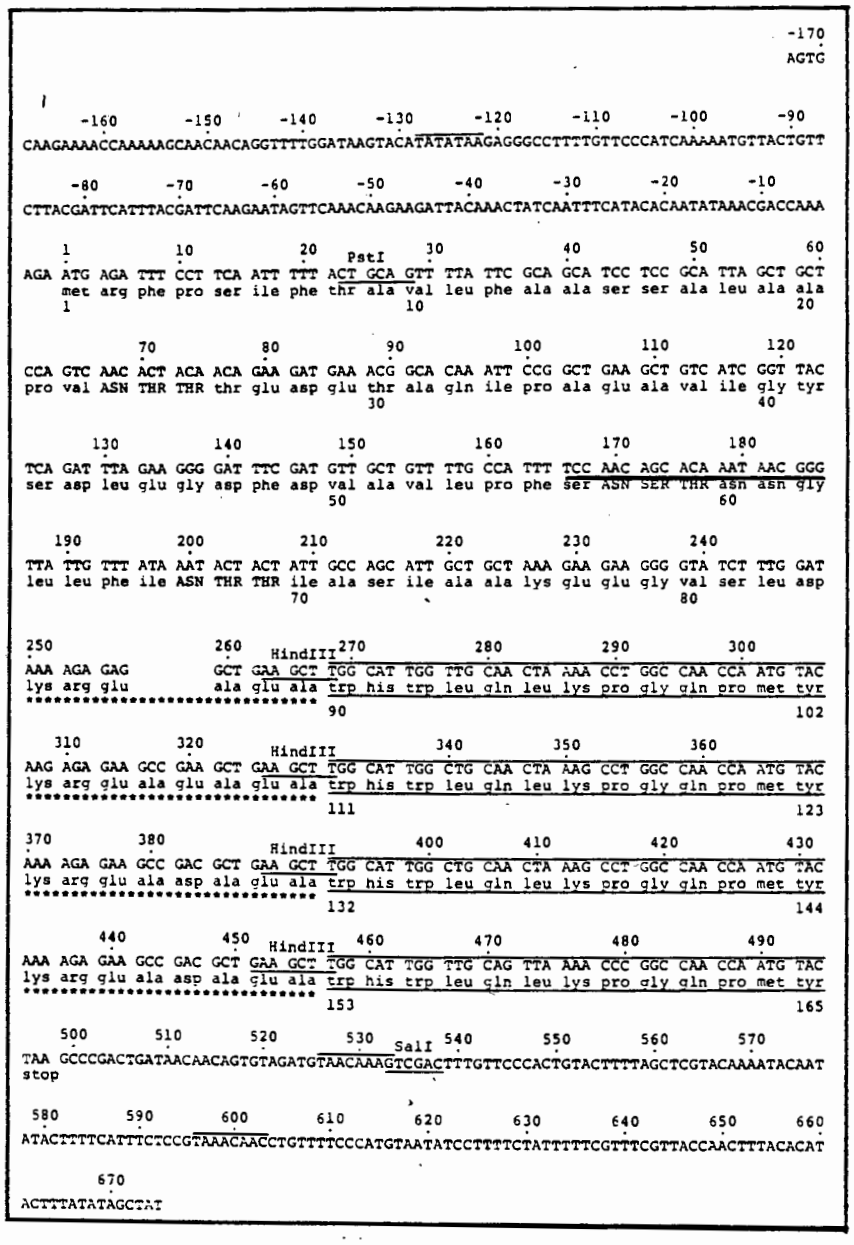
C.3 Restriction endonuclease map of pMFα8 (Miyajima et al., 1985). The hatched region represnts the mfα1 promoter and leader sequence. Also indicated are the 2 μm ori, the trp1 gene, ARS1, Ap^r gene, colE1 ori and the trp5 gene terminator region (from Miyajima et al., 1985)

Appendix D

Nucleotide sequence of the C. acetobutylicum endoglucanase gene and the S. cerevisiae α-mating factor gene.



D.1 Complete nucleotide sequence of the C. acetobutylicum endoglucanase gene. The putative promoter sequence is underlined with a solid line and a second putative promoter with a broken line. The -10, -35 and Shine-Dalgarno regions are indicated. The start of the nucleotide sequence for the endoglucanase gene fragment cloned into pHZ25, is indicated by the arrow (from Zappe, 1988).



D.2 Nucleotide sequence of mfα. The over and underlined regions represent the 4 tandem repeats of the α-factor coding sequences, the spacer regions are also indicated (*) (Kurjan and Herskowitz, 1982). The underlined region of 20 nucleotides indicates the region to which a 20 mer homologous primer was made.

Appendix E

Three and one letter codes used for amino acids.

Amino acid	Code		Code	Amino acid
Alanine	Ala	A	A	Alanine
Arginine	Arg	R	C	Cysteine
Asparagine	Asn	N	D	Aspartic acid
Aspartic acid	Asp	D	E	Glutamic acid
Cysteine	Cys	C	F	Phenylalanine
Glutamine	Gln	Q	G	Glycine
Glutamic acid	Glu	E	H	Histidine
Glycine	Gly	G	I	Isoleucine
Histidine	His	H	K	Lysine
Isoleucine	Ile	I	L	Leucine
Leucine	Leu	L	M	Methionine
Lysine	Lys	K	N	Asparagine
Methionine	Met	M	P	Proline
Phenylalanine	Phe	F	Q	Glutamine
Proline	Pro	P	R	Arginine
Serine	Ser	S	S	Serine
Threonine	Thr	T	T	Threonine
Tryptophan	Trp	W	V	Valine
Tyrosine	Tyr	Y	W	Tryptophan
Valine	Val	V	Y	Tyrosine

References

Ahmed A. 1984. Plasmid vectors for positive galactose-resistant selection of cloned DNA in Escherichia coli. Gene **28**:37-43.

Allcock E., and Woods D.R. 1981. Carboxymethyl cellulase and cellobiase production by Clostridium acetobutylicum in an industrial fermentation medium. Appl. Environ. Microbiol. **41**:539-541.

Ascenzioni F., and Lipps H.J. 1986. A linear shuttle vector for yeast and hypotrichous ciliate Strylonychia. Gene **46**:123-126.

Bayne M.L., Applebaum J., Chicchi G.G., Hayes N.S., Green B.G., and Cascieri M.A. 1988. Expression, purification and characterization of recombinant human insulin-like growth factor I in yeast. Gene **66**:235-244.

Beggs J.D. 1978. Transformation of yeast by a replicating hybrid plasmid. Nature **275**:104-109.

Beggs J.D., Van den Berg J., Van Ooyen A., and Weissman C. 1980. Abnormal expression of chromosomal rabbit β-globin gene in Saccharomyces cerevisiae. Nature **283**:835-840.

Bennetzen J.L., and Hall B.D. 1982a. The primary structure of the Saccharomyces cerevisiae gene for alcohol dehydrogenase 1. J. Biol. Chem. **257**:3018-3026.

Bennetzen J.L., and Hall B.D. 1982b. Codon selection in yeast. J. Biol. Chem. **257**:3026-3031.

Biely P. 1985. Microbial xylanolytic systems. Trends in Biotechnology **3**:286-290.

Bisaria V.S., and Ghose T.K. 1981. Biodegradation of cellulosic materials: substrates, microorganisms, enzymes and products. Enzyme Microb. Technol. **3**:90-104.

Bitter G.A., Chen K.K., Banks A.R., and Lai P-H. 1984. Secretion of foreign proteins from Saccharomyces cerevisiae directed by α-factor gene fusion. Proc. Natl. Acad. Sci. USA. **81**:5330-5334.

Bolivar F., Rodriguez R.L., Green P.J., Betlach M.C., Heyneker H.L., Boyer H.W., Crosa J.H., and Falkow S. 1977. Constuction and characterization of new cloning vehicles II. A multipurpose cloning system. Gene **2**:95-113.

Bostian K.A., Lemire J.M., Cannon L.E., and Halvorsen H.O. 1980. In vitro synthesis of repressible yeast acid phosphatase : identification of multiple mRNAs and products. Proc. Natl. Acad. Sci. USA. **77**:4504-4508.

Botterman J., and Zabeau M. 1985. High-level production of the EcoRI endonuclease under the control of the P_L promoter of bacteriophage lambda. Gene **37**:229-239.

Boyer H.W., and Roulland-Dussoix D. 1969. A complementation analysis of the restriction and modification of DNA in Escherichia coli. J. Mol. Biol. **41**:459-472.

Brake A.J., Merryweather J.P., Coit D.G., Heberlein U.A., Masiarz F.R., Mullenbach G.T., Urdea M.S., Valenzuela P., and Barr P.J. 1984. α-Factor-directed synthesis and secretion of mature foreign proteins in Saccharomyces cerevisiae. Proc. Natl. Acad. Sci. USA. **81**:4642-4646.

Braus G., Paravicini G., and Hutter R. 1988. A consensus transcription termination sequence in the promoter region is necessary for efficient gene expression of the trp1 gene of Saccharomyces cerevisiae. Mol. Gen. Genet. **212**:495-504.

Broach J.R. 1982. The Yeast Plasmid 2 μm circle. p445-470 In Strathern J.N., Jones E.W., Broach J.R. (eds.). The Molecular Biology of the Yeast Saccharomyces: Life Cycle and Inheritance. Cold Spring Harbour, New York.

Broach J.R. 1983. Construction of high copy number yeast vectors using 2 μm circle sequences. p307-325 In Wu R., Grossman L., Moldave K. (eds.). Methods in Enzymology **101**, Academic Press, New York.

Broach J.R., Strathern J.N., and Hicks J.B. 1979. Transformation in yeast : development of a hybrid cloning vector and isolation of the can1 gene. Gene **8**:121-133.

Castanon M.J., Spevak W., Adolf G.R., Chlebowicz-Sledziewska E., and Sledziewski A. 1988. Cloning of human lysozyme gene and expression in the yeast Saccharomyces cerevisiae. Gene **66**:223-234.

Cauchon N., and LeDuy A. 1984. Effect of dilution on carboxymethylcellulase and xylanase assays. Biotechnol. Bioeng. **26**:988-991.

Chen C.Y., Opperman H., and Hitzeman R.A. 1984. Homologous versus heterologous gene expression in the yeast, Saccharomyces cerevisiae. Nucleic Acids Res. **12**:8951-8970.

Chlebowicz-Sledziewska E., and Sledziewski A.Z. 1985. Constuction of multicopy yeast plasmids with regulated centromere function. Gene **39**:25-31.

Cigan A.M., and Donahue T-F. 1987. Sequence and structural features associated with translational initiator regions in yeast - a review. Gene **59**:1-18.

Clarke L., and Carbon J. 1980. Isolation of a yeast centromere and construction of functional small circular chromosomes. Nature **257**:504-509.

Cohen J.D., Eccleshall T.R., Needleman R.B., Federoff H., Buchferer B.A., and Marmur J. 1980. Functional expression in the yeast of the Escherichia coli plasmid gene coding for chloramphenicol acetyltransferase. Proc. Natl. Acad. Sci. USA. **77**:1078-1082.

Compere A.L., and Griffith W.L. 1979. Evaluation of subsrates for butanol production. Dev. Ind. Microbiol. **20**:509-517.

Coughlan M.P. 1985. The properties of fungal and bacterial cellulases with comment on their production and application. Biotechnology and Genetic Engineering Reviews. **3**:39-109.

Cousens L.S., Shuster J.R., Gallegos C., Ku L., Stempien M.M., Urdea M.S., Sanchez-Pescador R., Taylor A., and Tekamp-Olson P. 1987. High level expression of proinsulin in the yeast Saccharomyces cerevisiae. Gene **61**:265-275.

Curry C., Gilkes N., O'Neill G., Miller R.C., and Skipper N. 1988. Expression and secretion of a Cellulomonas fimi exoglucanase in Saccharomyces cerevisiae. Appl. Environ. Microbiol. **54**:476-484.

Dagert M., and Ehrlich S.D. 1979. Prolonged incubation in calcium chloride improved the competence of Escherichia coli cells. Gene **6**:23-28.

De Boer H.A., Comstock L.J., and Vasser M. 1983. The tac promoter : a functional hybrid derived from the trp and lac promoters. Proc. Natl. Acad. Sci. USA. **80**:21-25.

DeLange A.M., Futcher B., Morgan R., and McFadden G. 1984. Cloning of the vaccinia virus telomere in a yeast plasmid vector. Gene **27**:13-21.

Dean D. 1981. A plasmid cloning vector for the direct selection of strains carrying recombinant plasmids. Gene **15**:99-102.

Dekker R.F.H. 1985. Biodegradation of the hemicelluloses. p505-533 In Biosynthesis and Biodegradation of Wood Components. Chapter 18. Academic Press Inc., New York.

Derynck R., Singh A., and Goeddel D.V. 1983. Expression of the human interferon-γ cDNA in yeast. Nucleic Acids Res. **11**:1819-1837.

Dobson M.J., Tuite M.F., Mellor J., Roberts N.A., King R.M., Burke D.C., Kingsman A.J., and Kingsman S.M. 1982b. Expression in Saccharomyces cerevisiae of human interferon-alpha directed by trp1 5′ region. Nucleic Acids Res. **11**:2287-2302.

Dobson M.J., Tuite M.F., Roberts N.A., Kingsman A.J., Kingsman S.M., Perkins R.E., Conroy S.C., Dunbar B., and Fothergill L.A. 1982a. Conservation of high efficiency promoter sequences in Saccharomyces cerevisiae. Nucleic Acids Res. **10**:2625-2637.

Emr S.D., Schauer I., Hansen W., Esmon P., and Schekman R. 1984. Invertase β-galactosidase hybrid proteins fail to be transported from the endoplasmic reticulum in Saccharomyces cerevisiae. Mol. Cell. Biol. **4**:2347-2355.

Emr S.D., Schekman R., Flessel M.C., and Thorner J. 1983. An mfα1-suc2 (α-factor-invertase) gene fusion forthe study of protein localization and gene expression in yeast. Proc. Natl. Acad. Sci. USA. **80**:7080-7084.

Emtage J.S., Angal S., Doel M.T., Harris T.J.R., Jenkins B., Lilley G., and Lowe P.A. 1983. Synthesis of calf prochymosin (prorennin) in Escherichia coli. Proc. Natl. Acad. Sci. USA. **80**:3671-3675.

Gerbaud C., Fournier P., Blanc H., Aigle M., Heslot H., and Guerineau M. 1979. High frequency of yeast transformation by plasmids carrying part or entire 2 μm yeast plasmid. Gene **5**:233-253.

Goff C.G., Moir D.T., Kohno T., Gravius T.C., Smith R.A., Yamasaki E., and Tauton-Rigby A. 1984. Expression of calf prochymosin in Saccharomyces cerevisiae. Gene **27**:35-46.

Gonzalez A., Davila G., and Calva E. 1985. Cloning of a DNA sequence that complements glutamine auxotrophy in Saccharomyces cerevisiae. Gene **36**:123-129.

Gornall A.G., Bardwill C.J., and David M.M. 1949. Determination of serum proteins by means of the Biuret method. J. Biol. Chem. **177**:751-766.

Guarente L. 1984. Yeast promoters : positive and negative elements. Cell **36**:799-800.

Handa H., Mizumoto K., Oda K., Okamoto T., and Fukasawa T. 1985. Transcription of the human adenovirus Ela gene in Saccharomyces cerevisiae. Gene **33**:159-168.

Handa H., Toda T., Tajima M., Wada T., Iida H., and Fukasawa T. 1987. Expression of human adenovirus Ela product in yeast. Gene **58**:127-136.

Hartley J.L., and Donelson J.E. 1980. Nucleotide sequence of the yeast plasmid. Nature **286**:860-864.

Herskowitz I., and Oshima Y. 1982. Control of cell mating type interconversion. p181-210 In Strathern J.N., Jones E.W., Broach J.R. (eds) The Molecular Biology of the Yeast Saccharomyces: Life Cycle and Inheritance. Cold Spring Harbour, New York.

Hinchliffe E., and Box W.G. 1984. Expression of the cloned endo-1,3-1,4-β-glucanase gene of Bacillus subtilis in Saccharomyces cerevisiae. Curr. Genet. **8**:471-475.

Hinnen A., Hicks J.B., and Fink G.R. 1978. Transformation of yeast. Proc. Natl. Acad. Sci. USA. **75**:1929-1933.

Hitzeman R.A., Chen C.Y., Hagie F.E., Patzer E.G., Lui C-C., Estell D.A., Miller J.V., Yaffe A., Kleid D.G., Levinson A.D., and Opperman H. 1983. Expression of Hepatitis B virus surface antigens in yeast Nucleic Acids Res. 11:2745-2763.

Hitzeman R.A., Hagie F.F., Levine H.L., Goeddel D.W., Ammerer G., and Hall B.D. 1981. Expression of human gene for interferon in yeast. Nature **293**:717-722.

Hsu W.H., Magee P.T., Magee B.B., and Reddy C.A. 1983. Constuction of a new yeast cloning vector containing autonomous replication sequences from Candida utilis. J. Bacteriol. **154**:1033-1039.

Hughes D.E., Wimpenny J.W.T., and Lloyd D. 1971. The disintergration of micro-organisms. p1-54 In Norris J.R., Ribbons D.W. (eds.) Methods in Microbiology **5B**. Academic Press, New York.

Hyman B.C., Cramer J.H., and Rownd R.H. 1982. Properties of a Saccharomyces cerevisiae mtDNA segment conferring high frequency yeast transformation. Proc. Natl. Acad. Sci. USA. **79**:1578-1582.

Ikemura T. 1982. Correlation between the abundance of yeast transfer RNAs and the occurrence of the respective codons in protein genes. J. Mol. Biol. **158**:573-597.

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.

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

Ito H., Fukuda Y., Murata K., and Kimura A. 1983. Transformation of intact yeast cells treated with alkali cations. J. Bacteriol. **153**:163-168.

Jacobs E., Dewerchin M., and Boeke J.D. 1988. Retrovirus-like vectors for Saccharomyces cerevisiae : intergration of foreign genes controlled by efficient promoters into yeast chromosomal DNA. Gene **67**:259-269.

Jimenez A., and Davis J. 1980. Expression of a transposable antibiotic resistance element in Saccharomyces. Nature **287**:869-871.

Johnstone S., and Hopper J.E. 1982. Isolation of the regulatory gene gal4 and analysis of its dosage effects on the galactose-melibiose region. Proc. Natl. Acad. Sci. USA. **79**:6971-6975.

Jones D.T., and Woods D.R. 1986. The acetone butanol fermentation revisited. Microbiol. Rev. **50**:484-524.

Julius D., Blair L., Brake A., Sprague G., and Thorner J. 1983. Yeast α-factor is processed from a larger precursor polypeptide : the essential role of a membrane bound dipeptidyl aminopeptidase. Cell **32**:839-852.

Julius D., Brake A., Blair L., Kunisawa R., and Thorner J. 1984b. Isolation of the putative structural gene for the lysine-arginine-cleaving endopeptidase required for processing of yeast pre-pro-α-factor. Cell **37**:1075-1089.

Julius D., Schekman R., and Thorner J. 1984a. Glycosylation and processing of pre-pro-α-factor through the yeast secretory pathway. Cell **36**:309-318.

Kaneko Y., Hayashi N., Toh-e A., Banno I., and Oshima Y. 1987. Structural characteristics of the pho8 gene encoding repressible alkaline phosphatase in Saccharomyces cerevisiae. Gene **58**:137-148.

Keranen S. 1986. Synthesis and processing of Semliki forest virus polyprotein in Saccharomyces cerevisiae : a yeast type glycosylation of E1 envelope protein. Gene **48**:267-275.

Kingsman A.J., Clarke L., Mortimer R.K., and Carbon J. 1979. Replication in Saccharomyces cerevisiae of plasmid pBR313 carrying DNA from the yeast trp1 region. Gene **7**:141-152.

Kiss G.B., Pearlman R.E., Cornish K.V., Friesen J.D., and Chan V.L. 1982. The Herpes simplex virus thymidine kinase gene is not transcribed in Saccharomyces cerevisiae. J. Bacteriol. **149**:542-547.

Klebe R.J., Harriss J.V., Sharp Z.D., and Douglas M.G. 1983. A general method for polyethylene-glycol-induced genetic transformation of bacteria and yeast. Gene **25**:333-341.

Kniskern P.J., Hagopian A., Montgomery D.L., Burke P., Dunn N.R., Hoffman K.J., Miller W.J., and Ellis R.W. 1986. Usually high-level expression of a foreign gene (hepatitis B virus core antigen) in Saccharomyces cerevisiae. Gene **46**:135-141.

Knowles J., Lehtovaara P., and Teeri T. 1987. Cellulase families and their genes. Trends Biotechnol. **5**:255-261.

Kohchi C., and Toh-e A. 1986. Cloning of Candida pelliculosa β-glucosidase gene and its expression in Saccharomyces cerevisiae.Mol. Gen. Genet. **203**:89-94.

Kramer R.A., and Anderson N. 1980. Isolation of yeast genes with mMRA levels controlled by phosphate concentration. Proc. Natl. Acad. Sci. USA. **77**:6541-6545.

Kramer R.A., Dechiara T.M., Schaber M.D., and Hilliker S. 1984. Regulated expression of a human interferon gene in yeast : control by phosphate concentration or temperature. Proc. Natl. Acad. Sci. USA. **81**:367-370.

Kuhn I., Stephenson F.H., Boyer H.W., and Green P.J. 1986. Positive selection vectors utilizing lethality of the EcoRI endonuclease. Gene **44**:253-263.

Kurjan J., and Herskowitz I. 1982. Structure of a yeast pheromone gene (mfα) : A putative α-factor precursor contains four tandem copies of mature α-factor. Cell **30**:933-943.

Lamed R., Kenig R., and Setter E. 1985. Major characteristics of the cellulolytic system of Clostridium thermocellum coincide with those of the purified cellulosome. Enzyme Microb. Technol. **7**:37-41.

Langford C., Nellen N., Niessing J., and Gallwitz D. 1983. Yeast is unable to excise foreign intervening sequences from hybrid gene transcripts. Proc. Natl. Acad. Sci. USA. **80**:1496-1500.

Langley K.E., Egan K.M., Barendt J.M., Parker C.G., and Bitter G.A. 1988. Characterization of a purified hepatitis B surface antigen containing pre-s(2) epitopes expressed in Saccharomyces cerevisiae. Gene **67**:229-245.

Langridge R., Eibel H., Brown J.W.S., and Feix G. 1984. Transcription from maize storage protein gene promoters in yeast. EMBO J. **3**:2467-2471.

Lee S.F., Forsberg C.W., and Gibbins L.N. 1985a. Cellulolytic activity of Clostridium acetobutylicum. Appl. Environ. Microbiol. **50**:220-228.

Lee S.F., Forsberg C.W., and Gibbins L.N. 1985b. Xylanolytic activity of Clostridium acetobutylicum. Appl. Environ. Microbiol. **50**:1068-1076.

Lee S.F., Forsberg C.W., and Rattray J.B. 1987. Purification and properties of two endoxylanases from Clostridium acetobutylicum ATCC 824. Appl. Environ. Microbiol. **53**:644-650.

Ma H., Kunes S., Schatz P.J., and Botstein D. 1987. Plasmid constuction by homologous recombination in yeast. Gene **58**:201-216.

Maniatis T., Fritsch E.F., and Sambrook J. 1982. Molecular cloning : A laboratory manual. Cold Spring Harbour Laboratory, Cold Spring Harbour, New York.

McNeil J.B., Storms R.K., and Freisen J.D. 1980. High frequency recombination and the expression of genes cloned on chimeric yeast plasmids : Identification of a fragment of 2 μm essential for transformation. Curr. Genet. **2**:17-23.

Mellor J., Dobson M.J., Roberts N.A., Kingsman A.J., and Kingsman S.M. 1985. Factors affecting heterologous gene expression in Saccharomyces cerevisiae. Gene **33**:215-226.

Mellor J., Dobson M.J., Roberts N.A., Tuite M.F., Emtage J.S., White S., Lowe P.A., Patel T., Kingsman A.J., and Kingsman S.M. 1983. Efficient synthesis of enzymatically active calf chymosin in Saccharomyces cerevisiae. Gene **24**:1-14.

Miller G.L. 1959. Use of dinitrosalicylic acid reagent for determination of reducing sugar. Anal. Chem. **31**:426-428.

Miyajima A., Bond M.W., Otsu K., Arai K., and Arai N. 1985. Secretion of a mature mouse interleukin-2 by Saccharomyces cerevisiae : use of a general secretion vector containing promoter and leader sequences of the mating pheromone α-factor. Gene **37**:155-161.

Miyajima A., Otsu K., Schreurs J., Bond M.W., Abrams J.S., and Arai K. 1986. Expression of murine and human granulocyte-macrophage colony-stimulating factors in Saccharomyces cerevisiae : mutagenesis of the potential glycosylation sites. EMBO J. **5**:1193-1197.

Moir D.T., and Dumais D.R. 1987. Glycosylation and secretion of human alpha-1-trypsin by yeast. Gene **56**:209-217.

Murray A.W., and Szostak J.W. 1983. Pedigree analysis of plasmid segregation in yeast. Cell **34**:911-970.

Nakamura Y., Sato T., Emi M., Miyanohara A., Nishide T., and Matsubara K. 1986. Expression of human salivary α-amylase gene in Saccharomyces cerevisiae and its secretion using the mammalian signal sequence. Gene **50**:239-245.

Neill J.D., Litts J.C., Anderson O.D., Green F.C., and Stiles J.I. 1987. Expression of a wheat α-gliadin gene in Saccharomyces cerevisiae. Gene **55**:303-317.

Nikolnikov S., Posfai G., and Sain B. The construction of a versatile plasmid vector that allows direct selection of fragments cloned into six unique sites of the cI gene of coliphage 434. Gene **30**:261-265.

Oberto J., and Davison J. 1985. Expression of chicken egg white lysozyme by Saccharomyces cerevisiae. Gene **40**:57-65.

Palzkill T.G., and Newlon C.S. 1988. A yeast replication origin consists of multiple copies of a small conserved sequence. Cell **53**:441-450.

Parent S.A., Fenimore C.M., and Bostian K.A. 1985. Vector systems for the expression, analysis and cloning of DNA sequences in S. cerevisiae. Yeast **1**:83-138.

Penttila M., Andre L., Lehtovaara P., Baily M., Teeri T.T., and Knowles J.K.C. 1988. Efficient secretion of two fungal cellobiohydrolases by Saccharomyces cerevisiae. Gene **63**:103-112.

Penttila M.E., Andre L., Saloheimo M., Lehtovaara P., and Knowles J.K.C. 1987. Expression of two Trichoderma reesei endoglucanases in the yeast Saccharomyces cerevisiae. Yeast **3**:175-185.

Penttila M.E., Navalainen K.M.H., 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.

Perlman D., and Halvorsen H.O. 1983. A putative signal site in eukaryotic and prokaryotic signal peptides. J. Mol. Biol. **167**:391-409.

Price V., Mochizuki D., March C.J., Cosman D., Deeley M.C., Klinke R., Clevenger W., Giles S., Baker P., and Urdal D. 1987. Expression, purification and characterization of recombinant murine granulocyte-macrophage colony-stimulating factor and bovine interleukin-2 from yeast. Gene **55**:287-293.

Rees D.A., Morris E.R., Thom D., and Madden J.K. 1982. Shapes and interactions of carbohydrate chains. p195-290. In Polysaccharides **1** Academic Press, New York.

Reese E.T., Maguire A., and Parrish F.W. 1973. Production of β-D-xylopyranosidases by fungi. Can. J. Microbiol. **19**:1065-1074.

Reilly P.J. 1981. Xylanases: structure and function. p111-129 In Hollaender A., Rabson R., Rogers P., San Pietro A., Valentine R. and Wolfe R. (eds.) Trends in the biology of fermentations for fuels and chemicals. Plenum Press, New York.

Roman H. 1982. Development of yeast as an experimental organism. p1-9 In Strathern J.N.,Jones E.W.,Broach J.R.(eds.). The Molecular Biology of the yeast Saccharomyces : Life Cycle and Inheritance. Cold Spring Harbour, New York.

Rothstein S.J., Lahners K.N., Lazarus C.M., Baulcombe D.C., and Gatenby A.A. 1987. Synthesis and secretion of wheat α-amylase in Saccharomyces cerevisiae. Gene **55**:353-356.

Rudolf H., Koenig-Rauseo I., and Hinnen A. 1985. One-step gene replacement in yeast by cotransformation. Gene **36**:87-95.

Ruohonen L., Hackman P., Lehtovaara P., Knowles J.K.C., and Keranen S. 1987. Efficient secretion of Bacillus amyloliquefaciens α-amylase cells by its own signal peptide from Saccharomyces cerevisiae host. Gene **59**:161-170.

Sacco M., Millet J., and Aubert J.P. 1984. Cloning and expression in Saccharomyces cerevisiae of a cellulase gene from Clostridium thermocellum. Ann. Microbiol. **135A**:485-488.

Sato T., Tsunasawa S., Nakamura Y., Emi M., Sakiyama F., and Matsubara K. 1986. Expression of the human salivary α-amylase gene in yeast and characterization of the secreted protein. Gene **50**:247-257.

Schekman R., and Novick P. 1982. The secretory process and yeast cell surface assembly. p361-398 In Strathern J.N.,Jones E.W.,Broach J.R.(eds.). The Molecular Biology of the yeast Saccharomyces : Metabolism and Gene Expression. Cold Spring Harbour, New York.

Schultz L.D., Hofmann K.J., Mylin L.M., Montgomery D.L., Ellis R.W., and Hopper J.E. 1987b. Regulated overproduction of the gal4 gene product greatly increases expression from galactose-inducible promoters on multi-copy expression vectors in yeast. Gene **61**:123-133.

Schultz L.D., Tanner J., Hofmann K.J., Emini E.A., Condra J.H., Jones R.E., Kieff E., and Ellis R.W. 87a. Expression and secretion in yeast of a 400 kDa envelope glycoprotein derived from Epstein-Barr virus. Gene **54**:113-123.

Scopes R.K. 1982. Protein purification. Springer-Verlag, New York.

Sherman F., Fink G.R., and Hicks J.B. 1986. Methods in yeast genetics. Cold Spring Harbour Laboratory, Cold Spring Harbour, New York.

Sherman F., and Stewart J.W. 1982. Mutations altering the initiation of translation of yeast iso-1-cytochrome c: contrast between the eukaryotic and prokaryotic initiation process. p301-355 In Strathern J.N.,Jones E.W.,Broach J.R.(eds.). The Molecular Biology of the Yeast Saccharomyces : Metabolism and Gene Expression. Cold Spring Habour, New York.

Siciliano P.G., and Tatchell K. 1984. Transcription and regulatory signals of the mating type locus in yeast. Cell **37**:969-978.

Simons G., Remut E., Allet B., Devos R., and Fiers W. 1984. High level expression of human interferon gamma in Escherichia coli under control of the P1 promoter of bacteriophage lambda. Gene **28**:55-64.

Skipper N., Sutherland M., Davis R.W., Kilburn D., Miller R.C.Jr., Warren A., and Wong R. 1985. Secretion of a bacterial cellulase in yeast. Science **230**:958-961.

Smith R.A., Duncan M.J., and Moir D.T. 1985. Heterologous protein secretion from yeast. Science **229**:1219-1224.

St. John T., and Davis R.W. 1981. The organization and transcription of the gal operon. J. Mol. Biol. **152**:285-315.

Stepien P.P., Brousseau R., Wu R., Narang S., and Thomas D.Y. 1983. Synthesis of a human insulin gene VI. Expression of the synthetic proinsulin gene in yeast. Gene **24**:289-297.

Stinchcomb D.T., Struhl K., and Davis R.W. 1979. Isolation and characterization of a yeast chromosomal replicator. Nature **282**:39-43.

Struhl K. 1985. A rapid method for creating recombinant DNA molecules. BioTechniques **3**:452-453.

Struhl K., Stinchcomb D.T., Scherer S., and Davis R.W. 1979. High frequency transformation of yeast : autonomous replicating hybrid molecules. Proc. Natl. Acad. Sci. USA. **76**:1035-1039.

Tabor S., and Richardson C.C. 1987. DNA sequence analysis with a modified bacteriophage T7 DNA polymerase. Proc. Natl. Acad. Sci. USA. **84**:4767-4771.

Takagi M., Kawai S., Chang M.C., Shibuya I., and Yano K. 1986. Construction of a host-vector system in Candida maltosa by using an ARS site isolated from its genome. J. Bacteriol. **167**:551-555.

Tanaka T., Kimura S., and Ota Y. 1988. Secretion of the proenzyme and active bovine pancreatic phospholipase A2 enzyme by Saccharomyces cerevisiae : design and use of a synthetic gene. Gene **64**:257-264.

Teather R.M., and Wood P.J. 1982. Use of Congo red-polysaccharide interactions in eunumeration and characterization of cellulolytic bacteria from the bovine rumen. Appl. Environ. Microbiol. **43**:777-780.

Thorner J. 1982. Pheromonal regulation of development in Saccharomyces cerevisiae. p143-180 In Strathern J.N., Jones E.W., Broach J.R.(eds.). The Molecular Boilogy of the Yeast Saccharomyces : Life Cycle and Inheritance. Cold Spring Harbour, New York.

Tschumper G., and Carbon J. 1980. Sequence of a yeast DNA fragment containing a chromosomal replicator and trp1 gene. Gene **10**:157-166.

Tuite M.F., Dobson M.J., Roberts N.A., King R.M., Burke D-C., Kingsman S.M., and Kingsman A.J. 1982. Regulated high efficiency of human interferon-alpha in Saccharomyces cerevisiae. EMBO J. **1**:603-608.

Urdea M.S., Merryweather J.P., Mullenbach D.C., Coit D., Heberlein U., Valenzuela P., and Barr P.J. 1983. Chemical synthesis of a gene for human epidermal growth factor urogastrone and its expression in yeast. Proc. Natl. Acad. Sci. USA. **80**:7461-7465.

Valenzuela P., Medina A., Rutter W.J., Ammerer G., and Hall B.D. 1982. Synthesis and assembly of hepatitis B virus surface antigen particles in yeast. Nature **298**:347-350.

Van Arsdell J.N., Kwok S., Schweickart V.L., Ladner M.B., Gelfand D.H., and Innis M.A. 1987. Cloning, characterization, and expression in Saccharomyces cerevisiae of endoglucanase I from Trichoderma reesei. Bio/Technol. **5**:60-64.

Vernet T., Dignard D., and Thomas D.Y. 1987. A family of yeast expression vectors containing phage f1 intergenic region. Gene **52**:255-233.

Webster T.D., and Dickenson R.C. 1983. Direct selection of Saccharomyces cerevisiae resistant to the antibiotic G418 following transformation with a DNA vector carrying the kanamycin-resistant gene of Tn 903. Gene **26**:243-252.

Wong W.K.R., Curry C., Parekh R.S., Parekh S.R., Wayman M., Davies R.W., Kilburn D.G., and Skipper N. 1988. Wood hydrolysis by Cellulomonas fimi endoglucanase and exoglucanase coexpressed as secreted enzymes in Saccharomyces cerevisiae. Biotechnol. **6**:713-719.

Yoshizumi H., and Ashikari T. 1987. Expression, glycosylation and secretion of fungal hydrolases in yeast. Trends Biotechnol. **5**:277-281.

Zabeau M., and Stanley K.K. 1982. Enhanced expression of cro-beta-galactosdase fusion proteins under the control of the P_R promotor of bacteriophage lambda. EMBO J. **1**:1217-1224.

Zalkin H., and Yanofsky C. 1982. Yeast gene trp5 : structure, function, regulation. J. Biol. Chem. **257**:1491-1500.

Zappe H. 1988. The cloning and characterisation of an endoglucanase and an endoxylanase from Clostridium acetobutylicum in Escherichia coli. Ph D Thesis University of Cape Town

Zappe H., Jones D.T., and Woods D.R. 1986. Cloning and expression of Clostridium acetobutylcum endoglucanase, cellobiase and amino acid biosynthesis genes in Escherichia coli. J. Gen. Microbiol. **132**:1367-1372.

Zappe H., Jones D.T., and Woods D.R. 1987. Cloning and expression of a xylanase gene from Clostridium acetobutylicum P262 in Escherichia coli. J. Microbiol. Biotechnol. **27**:57-63.

Zappe H., Jones W.A., Jones D.T., and Woods D.R. 1988. Stucture of an endo-β-1,4-glucanase gene from Clostridium acetobutylicum P262 showing homology with endoglucanase genes from Bacillus spp. Appl. Environ. Microbiol. **54** :1289-1292.