# 12

# Heterologous Gene Expression in *Saccharomyces cerevisiae*

SUSAN M. KINGSMAN, ALAN J. KINGSMAN, MELANIE J. DOBSON, JANE MELLOR AND NICOLA A. ROBERTS

*Department of Biochemistry, University of Oxford, South Parks Road, Oxford, OX1 3QU, UK.*

## Introduction

Recombinant DNA technology can be used to isolate any eukaryotic gene encoding interesting and/or medically and commercially important polypeptides. In 1977 it was shown that a gene from a higher eukaryote could be expressed in a micro-organism, *Escherichia coli*, to produce a biologically active protein, somatostatin (Itekura *et al.*, 1977). The ability to express eukaryotic genes in micro-organisms allows the large-scale production of proteins which cannot be produced in significant quantities from natural sources. It appears, however, that *E. coli* may not be the most suitable host for the expression of all eukaryotic proteins. This may be particularly true for the production of human pharmaceuticals and food products because *E. coli* produces toxic and pyrogenic cell-wall components. In addition, the mechanisms of transcription, translation and post-translational processing in *E. coli* differ from those of eukaryotes so that the proteins produced from eukaryotic genes in *E. coli* may differ from the normal gene product and lack the required biological action or be insoluble until appropriately modified chemically (e.g. calf prochymosin, Emtage *et al.*, 1983; interferon-gamma (IFN-γ), Simons *et al.*, 1984). The yeast *Saccharomyces cerevisiae* may provide a suitable alternative host. Many of the eukaryotic proteins which are formed by *E. coli* in an inactive state are produced as soluble, biologically active proteins in *S. cerevisiae* (e.g. calf prochymosin, Mellor *et al.*, 1983; IFN-γ, Derynck, Singh and Goeddel, 1983). Furthermore, *S. cerevisiae* is not a pathogen and already has wide acceptability for use in the food industry.

Abbreviations: DAS, downstream activator sequence; ER, endoplasmic reticulum; GF, growth factor; GHRF, growth hormone releasing factor; HBsAg, hepatitis B surface antigen; hEGF, human epidermal growth factor; IFN, interferon; ILGF, insulin-like growth factor; kd, kilodaltons; ORF, open reading frame; PGK, phosphoglycerate kinase; t-PA, tissue plasminogen activator; UAS, upstream activator sequence.

*Biotechnology and Genetic Engineering Reviews* — Vol. 3, September 1985
0264-8725/85/03/377-40$10.00 + $0.00 

This feature should simplify its use on the manufacturing scale and greatly assist the production of food and pharmaceutical products free from toxic contaminants. *S. cerevisiae* has an additional advantage over *E. coli* in that it has a secretion system which is similar to higher eukaryotic systems (Schekman and Novick, 1982) and which can be manipulated to allow the secretion of heterologous proteins. *S. cerevisiae* also has an advantage over mammalian-cell systems which are being developed. These largely involve tumour or semi-transformed cell lines and tumour virus vectors which may reduce their acceptability. There are already well-established processes for the large-scale production of *Saccharomyces* and *Saccharomyces* products and these can be readily adapted for the production of heterologous proteins. The ease of genetic manipulation of *S. cerevisiae* coupled with its acceptability as a production organism and the ease of bulk fermentation make it the preferred organism for the production of many eukaryotic proteins.

### Transformation of *S. cerevisiae*

The manipulation of *S. cerevisiae* to produce foreign proteins was made possible by the development of techniques to introduce exogenous DNA into yeast cells (Beggs, 1978; Hinnen, Hicks and Fink, 1978). These transformation systems involve enzymatic removal of the cell wall to produce sphaeroplasts which can take up DNA on treatment with polyethylene glycol and calcium ions. Under appropriate conditions the cell walls can regenerate to permit propagation and selection of transformants in the normal way. The ability to detect the successful uptake of exogenous DNA depends on the presence of a genetic marker. The most common procedure is to use an auxotrophic host strain and incorporate the corresponding wild-type gene into the exogenous DNA. Transformants will take up the wild-type gene and can therefore be selected against the background of non-transformed auxotrophs. The *LEU2* (Beggs, 1978) and *TRP1* (Tschumper and Carbon, 1980) genes which encode 3-isopropylmalate dehydrogenase and *N*(5′-phosphoribosyl) anthranilate isomerase respectively, are commonly used as selectable markers and in addition, various dominant selectable genes which confer resistance to drugs are now available, e.g. thymidine kinase (McNiel and Friesen, 1981), chloramphenicol acetyltransferase (Cohen *et al.*, 1980), aminoglycoside 3′ phosphotransferase (G418 resistance) (Jimenez and Davies, 1980; Webster and Dickson, 1983), hygromycin B phosphotransferase (Gritz and Davies, 1983) and dihydrofolate reductase (tetrahydrofolate dehydrogenase; Miyajima *et al.*, 1984). These dominant selectable markers are most useful for transforming strains which lack suitable auxotrophic mutations. The best configuration for auxotrophic selection is to use a host strain with a complete deletion of the chromosomal copy of the gene to be used for selection (*see* Struhl, 1983, for methods of specific gene deletion). This ensures strain stability as there is no chromosomal site for integration by homologous recombination. In addition, if the expression of the selectable marker is reduced by mutation there may be selection for high plasmid copy

number to maintain a sufficient level of the essential gene product (*see* page 381).

Various plasmids have been designed to facilitate the genetic manipulation of *S. cerevisiae*. The most useful are the shuttle vectors which can replicate and be selected in *E. coli* as well as in *S. cerevisiae*. This permits all recombinant DNA preparative procedures to be carried out using the powerful *E. coli* technology. Two types of shuttle vectors are illustrated in *Figures 1* and *2*. The first type (*Figure 1*) relies on chromosomal DNA sequences for plasmid replication and maintenance. The plasmid YRp7 (Struhl *et al.*, 1979; *Figure 1a*) contains pBR322 to allow replication and selection in *E. coli* and a 1·45 kb fragment of yeast DNA which contains the

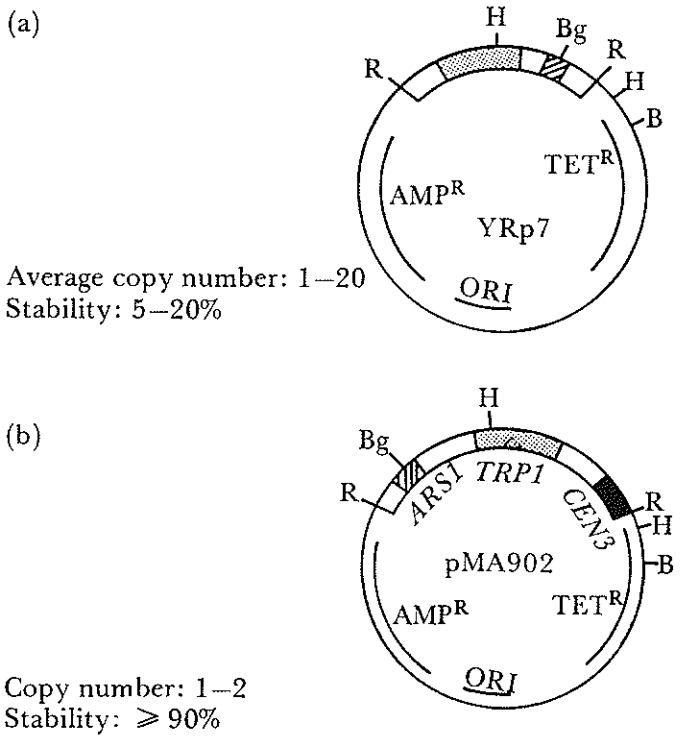


**Figure 1.** *E. coli*/*S. cerevisiae* shuttle vectors based on chromosomal replication and maintenance sequences. (a) plasmid YRp7 (Struhl *et al.*, 1979). (b) plasmid pMA902 (M.J. Dobson and B. Bowen, unpublished work): the *Eco* RI fragment from YRp7 was truncated beyond *ARS1* and fused to a truncated fragment from pYe*CDC10* (Clarke and Carbon, 1980) containing *CEN3*, to produce a 2·4 kb *Eco*RI fragment which was inserted into pBR322. Thin line = pBR322, the origin of replication (ORI) and ampicillin (AMP) and tetracycline (TET) resistance genes are indicated. The boxed region is yeast DNA, the stippled box = the *TRP1* coding region; hatched box = *ARS1* and dark box = *CEN3*. Selected restriction enzyme sites are indicated. B = *Bam* HI, Bg = *Bgl* II, H = *Hind* III, R = *Eco* RI. Average plasmid copy number per cell and stability as percentage of cells containing plasmid after selective growth are given.

*TRP1* gene as a selectable marker and *ARS1* (autonomously replicating sequence) which allows the plasmid to replicate in *S. cerevisiae*. *ARS*s are probably chromosomal origins of replication (Campbell, 1983); *ARS* plasmids are present at a high copy number, about 20–50 copies per cell (e.g. Hyman, Cramer and Rownd, 1982), although in selectively grown cultures only a fraction of the cells contain plasmid. The proportion of plasmid-bearing cells in a population grown under selective conditions varies but it can range from 5% (Murray and Szostak, 1983) to 80% (Kingsman *et al.*, 1979). This means that the average copy number of an *ARS* plasmid could be as low as one per cell. Furthermore, *ARS* plasmids are highly unstable: in the absence of selection about 20% of the cells lose plasmid per generation (Kingsman *et al.*, 1979). The maintenance of *ARS*-based plasmids can be stabilized by the addition of a second sequence (*CEN*) derived from the centromere region of a yeast chromosome (Clarke and Carbon, 1980). A typical *ARS/CEN* plasmid is shown in *Figure 1b*; it carries *TRP1*, *ARS1* and *CEN3* (Clarke and Carbon, 1980; Tschumper and Carbon, 1980). The *CEN* sequence reduces the plasmid copy number to one per cell, but the plasmids are more stably maintained during mitotic growth and in the absence of selection about 90% of the cells contain plasmid (Clarke and Carbon, 1980).

The second type of shuttle vector relies on DNA replication and maintenance sequences derived from an endogenous yeast plasmid, the 2 μm circle (Broach, 1982; *Figure 2a*). This is a 6318 bp plasmid (Hartley and Donelson, 1980) which contains two genes, *REP1* and *REP2*, a *cis*-acting region, *REP3*, and a replication origin, which are required for plasmid replication, maintenance and/or segregation (Jayaram, Li and Broach, 1983); a third gene, *FLP*, which catalyses a site-specific recombination event between the inverted repeat sequences, may also be involved in replication. The plasmid is present at 20–100 copies per cell (Clarke-Walker and Miklos, 1984; Gerbaud and Guerineau, 1980) and shows no segregation bias; therefore, all cells in the population contain plasmid at high copy number (Murray and Szostak, 1983). Beggs (1978) has constructed derivatives of the 2 μm circle (e.g. pJDB219) which contain the *LEU2* gene as a selectable marker and the plasmid, pMB9 for replication in *E. coli* (*Figure 2a*). Plasmid pJDB219 transforms *S. cerevisiae* at a frequency of about $10^5$ transformants/μg and is stably maintained at a high copy number, at least 100 copies/cell. A smaller, more convenient vector, pMA3a, has been constructed by inserting a 3·25 kb double *Eco*RI fragment from pJDB219 into pBR322 (Dobson *et al.*, 1982a; *Figure 2b*). This plasmid contains the *LEU2* selectable marker, the origin of replication and the *REP3* region from pJDB219. When this plasmid is introduced into a yeast strain containing endogenous 2 μm plasmid (a [$cir^+$] strain) it will replicate using the replication proteins (encoded by *REP1* and *REP2*) provided in *trans* by the endogenous 2 μm circle; however, it will not be stably maintained in a 2 μm plasmid-free [$cir^0$] strain. pMA3a is present at 100 copies per cell in [$cir^+$] strains and is stably maintained in the absence of selection (M.J. Dobson and N.A. Roberts, unpublished data). Several factors may influence the maximum copy number established by 2 μm-based plasmids such as pMA3a: in particular the *LEU2* gene from pJDB219 is often

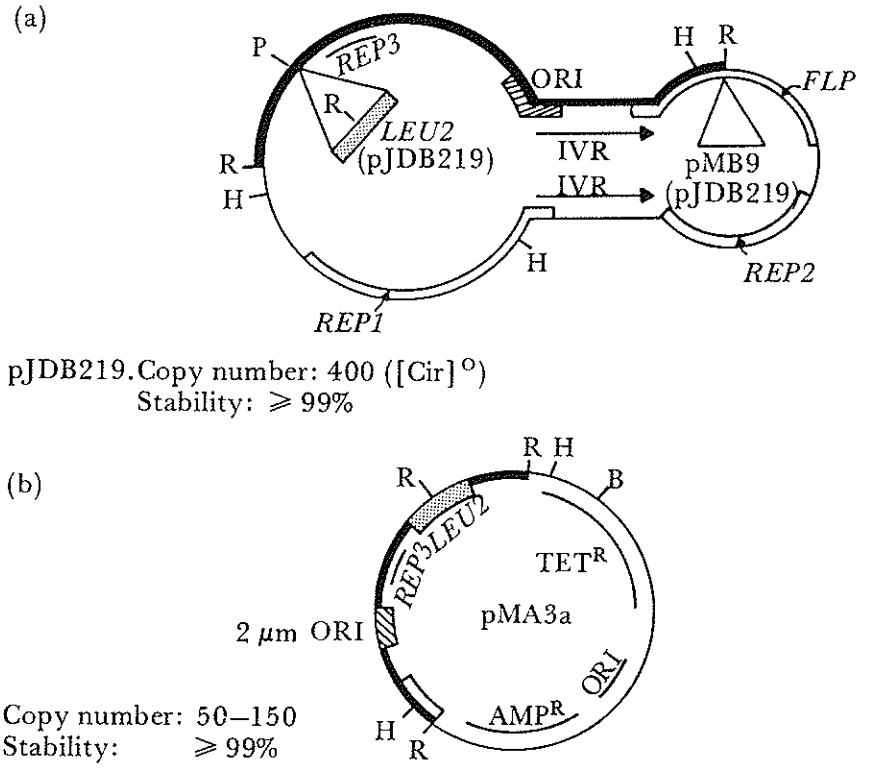


**Figure 2.** *E. coli*/*S. cerevisiae* shuttle vectors based on the replication and maintenance sequences of the endogenous 2 μm circle. (a) The 2 μm circle indicating sites of insertion of *S. cerevisiae* chromosomal sequences and *E. coli* plasmid, pMB9 sequences in the shuttle vector pJDB219 (Beggs, 1978). The 2 μm circle contains a hyphenated inverted repeat (IVR) and homologous recombination across this produces two forms (A and B); the B form is shown (Beggs, Guerineau and Atkins, 1978). The three coding regions, *REP1*, *REP2* and *FLP*, are indicated by open boxes; *REP3* is indicated by a line; hatched box = origin of replication; large triangles indicate the points of insertion of additional sequences into the 2 μm circle; a fragment containing the *S. cerevisiae* chromosomal *LEU2* gene (stippled box) was inserted at the *Pst* I site by AT tailing and pMB9 was inserted at the *Eco* RI site in *FLP* (Beggs, 1978). The thick line indicates the 3·25 kb double *Eco* RI fragment used to construct pMA3a. (b) pMA3a, a derivative of pJDB219. Symbols as in (a) except that thin line = pBR322 (*see Figure 1*); open box = partial *FLP* genes. Restriction sites as in *Figure 1*.

associated with high plasmid copy numbers. This *LEU2* gene has a truncated promoter which could result in inefficient expression: this might provide a selection system for high copy number to provide sufficient isopropylmalate dehydrogenase for growth (Erhart and Hollenberg, 1983). Alternatively, the location and/or properties of the *LEU2* DNA might affect the functioning of the replication origin or regulatory sequences (Broach, 1983). If pJDB219, which contains the entire 2 μm circle, is introduced into a yeast strain which lacks endogenous 2 μm plasmid (a [*cir*$^0$] strain, e.g. Dobson, Futcher and Cox, 1980) then copy numbers in the region of 400 per cell can be obtained (M.J. Dobson, unpublished data).This may be due in part to the lack of

competition with endogenous molecules for the replication machinery but is also related to the presence of the defective *LEU2* gene. Recently it has been shown that pJDB219 is stably maintained during 100 generations in non-selective continuous culture, although there was some host-strain-dependent variation in stability, the best results being obtained with a [cir$^0$] AH22 (Walmsley, Gardner and Oliver, 1983). It may also be possible to increase the copy number of 2 μm-based plasmids by manipulating the levels of expression of the 2 μ-encoded *trans*-acting gene products, for example, overproduction of the *REP1* gene product results in a copy number of up to 600 per cell although additional manipulations are required to improve plasmid stability (M.J. Dobson, unpublished results). Plasmid copy numbers in yeast transformants can be readily determined by comparing the levels of plasmid-specific restriction fragments in total DNA digests with those of the ribosomal DNA repeats; 2 μm plasmids can usually be detected by ethidium bromide staining but Southern blotting can be used to detect plasmids present at low copy numbers (Broach, 1983; *see also Figure 8*).

All the plasmids which replicate independently of the yeast chromosome by virtue of the presence of an *ARS* or 2 μm sequences will transform *S. cerevisiae* at a high frequency, about $10^3$–$10^5$ transformants/μg. The plasmids can be readily rescued from the transformants and reisolated by transforming *E. coli* (e.g. Fergusson, Groppe and Reed, 1981). The ability to shuttle between hosts has several advantages: all plasmid manipulations can be done in *E. coli* which is more convenient to handle preparatively than *S. cerevisiae* and only the final plasmid construction is introduced into *S. cerevisiae*. Furthermore, the ability to rescue plasmids from *S. cerevisiae* into *E. coli* allows their structures to be confirmed to ensure that any results are not due to artefacts of plasmid rearrangement. It is not, however, essential to have any *E. coli* sequences present on the plasmid for efficient function in *S. cerevisiae*. Once the final molecule has been constructed the *E. coli* sequences can be removed by appropriate restriction enzyme digestion and subsequent purification of the *S. cerevisiae* sequences. These can then be religated and the ligation mix used directly to transform *S. cerevisiae*; transformation frequencies of about $10^3$–$10^4$ transformants/μg can be obtained (M.J. Dobson, unpublished results). In this case plasmid integrity is assayed by Southern hybridization.

Using the classical transformation procedures (Beggs, 1978; Hinnen, Hicks and Fink, 1978) the frequency of cotransformation is high with as many as 30 copies of plasmid being taken up per cell. Lithium ions can also be used to stimulate uptake of exogenous DNA (Ito *et al.*, 1983). This technique appears to result in a lower level of cotransformation and the overall transformation frequency is lower. In addition this technique appears to predispose the plasmid to mutations and may not, therefore, be generally useful.

Exogenous DNA can be inserted into the *S. cerevisiae* chromosome by using integrative vectors. These vectors lack any replication sequences and can be maintained only by integration. The transformation frequency is low, about 1–2 transformants/μg because integration is rare. Integrative vectors (*Figure 3*) insert into the chromosome by homologous recombination across

*S. cerevisiae* sequences. If multiple *S. cerevisiae* sequences are present on the vector, integration can occur at any of the homologous chromosomal locations, but if a double strand break is introduced into one of the *S. cerevisiae* sequences, recombination at this site is stimulated several hundred-fold. This allows the plasmid to be targeted to specific chromosomal locations

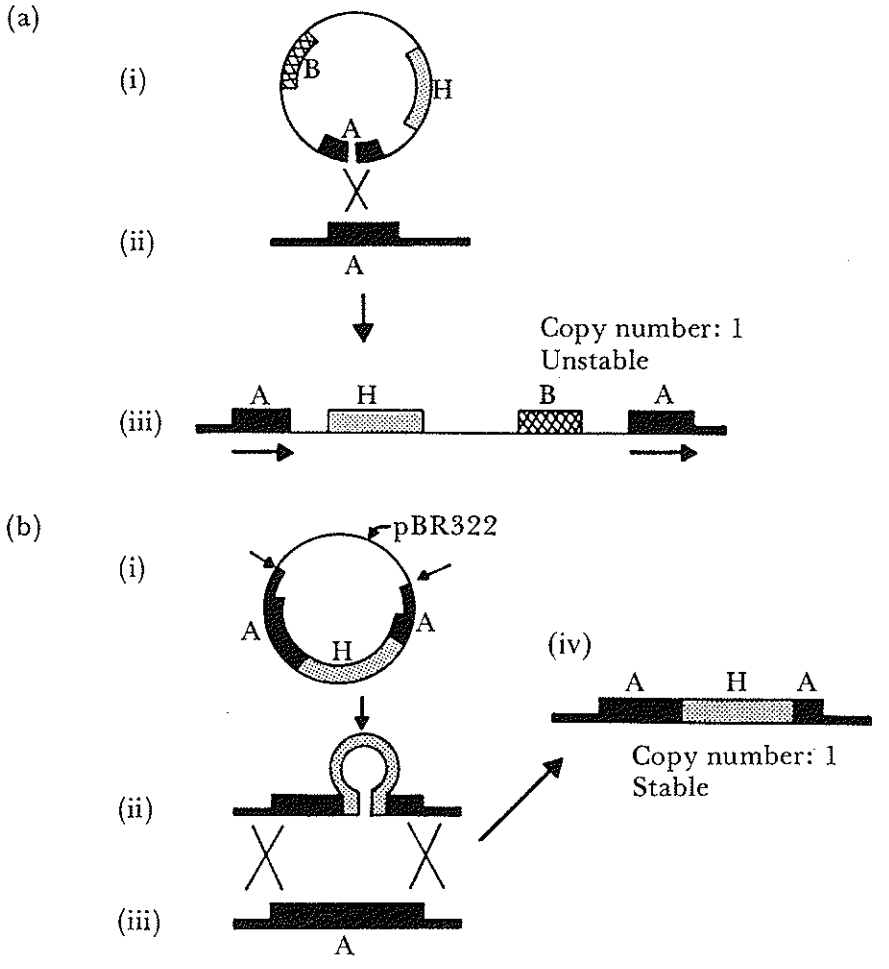


**Figure 3.** Integrative vectors. Integration of heterologous genes into the yeast chromosome. (a) An idealized cleaved plasmid integrative vector: thin line = pBR322; solid box = *S. cerevisiae* gene A with a double strand break; cross-hatched box = *S. cerevisiae* gene B (selective marker); stippled box = heterologous gene, H. (ii) The chromosomal gene A. (iii) The chromosomal gene A region after the integration. The plasmid integrates by homologous recombination across gene A. If the plasmid is not cleaved then integration would occur at a lower frequency at either chromosomal gene A or gene B. There is a duplication of the target chromosomal gene. (b) (i) An idealized linear integrative vector: a heterologous gene (H) is inserted into an *S. cerevisiae* sequence (A). (ii) A linear fragment is generated by cleavage in the flanking *S. cerevisiae* sequences indicated by arrows. (iii) The linear fragment integrates at the chromosomal sequence A by homologous recombination stimulated by the ends. (iv) The wild-type chromosomal sequence A is replaced with the sequence A containing the heterologous coding sequences (H). There is no duplication of target DNA.

(Orr-Weaver, Szostak and Rothstein, 1983). Homologous recombination across *S. cerevisiae* sequences in the vector results in insertion of the entire plasmid into the chromosome but creates a tandem duplication of the target sequence (*Figure 3a*). The strain is therefore genetically unstable because the plasmid sequences can be lost by excision. Rothstein (1983) has developed a procedure for the replacement of chromosomal genes with either mutant genes or heterologous DNA sequences, which does not generate tandem duplications. The heterologous DNA is inserted into a *S. cerevisiae* sequence and a linear fragment is generated which has ends within the *S. cerevisiae* sequences flanking the heterologous DNA. The ends would not be adjacent in the wild-type sequence. The ends are recombinogenic and target the fragment to the homologous chromosomal site where the normal gene is replaced by the recombinant gene by recombination at the ends of the fragment (*Figure 3b*). This procedure results in the stable introduction of a single copy of exogenous DNA into the *S. cerevisiae* chromosome. These techniques have been extended to produce multicopy integrative expression vectors (*see* 'Alternative vectors', pages 405–406). Various procedures for manipulating *S. cerevisiae* using recombinant plasmids have recently been reviewed (Struhl, 1983).

In order to express heterologous genes in *S. cerevisiae* one of these vector systems must be used to introduce the genes into the cell. The choice of vector will depend on the level of heterologous gene expression required and on the requirements for strain stability. If, for example, high levels of heterologous gene expression are required in a short fermentation, then a 2 μm-based vector will be favoured because of its high copy number. If, however, absolute strain stability is required, then an integrative vector might be used. Under some circumstances only a low level of heterologous gene expression will be required, as would be the case where new metabolic pathways have been constructed or where a low level of enzyme production is required to enhance a traditional process. In these cases *ARS/CEN* vectors or artificial chromosomes (*see* pages 379 and 406) might be useful.

### The expression of heterologous genes

As plasmid vectors became available, studies were undertaken to establish whether entire heterologous, i.e. non-yeast, genes could be expressed in *S. cerevisiae*. In the first study a genomic clone containing the rabbit β-globin gene inserted into pJDB219 was used (Beggs *et al.*, 1980). Rabbit β-globin homologous RNA was detected but it was aberrant: a different 5′ end was generated in *S. cerevisiae*, suggesting that the heterologous promoter was not used correctly. In addition, the introns had not been excised. Although several *S. cerevisiae* genes have introns (e.g. actin, Gallwitz and Sures, 1980; ribosomal proteins, Bollen *et al.*, 1982) heterologous introns are not reliably processed (Langford *et al.*, 1983). In order to express any heterologous coding sequences it is therefore important to use a cDNA. The ability of *S. cerevisiae* to recognize transcriptional signals in heterologous DNA is variable. In many cases, as with rabbit β-globin, there is some transcriptional

activity but the mRNA initiates at the wrong site in *S. cerevisiae*, e.g. *Drosophila melanogaster ADE8* (Henikoff and Furlong, 1983), phaseolin (Cramer, Lea and Slightom, 1985) and even the *ADH* gene from another, albeit distantly related, yeast *Schizosaccharomyces pombe* (Russell, 1983). In some cases no promoter activity can be detected, e.g. *Herpes simplex* thymidine kinase (Kiss *et al.*, 1982). There are some examples where the normal transcription initiation sites of the heterologous gene are used in *S. cerevisiae*, e.g. *Zea mays*, zein (Langridge *et al.*, 1984) and the *D. melanogaster, ADE8* transcript terminates correctly in *S. cerevisiae* (Henikoff and Cohen, 1984). The generation of specific transcripts in some cases implies that certain features of eukaryotic promoters and terminators might be conserved. It is clear however that while 'foreign' transcriptional signals might function correctly in *S. cerevisiae* this cannot be predicted at present and the efficiency of expression directed by heterologous signals in *S. cerevisiae* is generally low.

### *S. cerevisiae* gene-expression signals

From the studies outlined above (pages 384–385) it is clear that to ensure the efficient expression of any heterologous gene in *S. cerevisiae* it is necessary to replace the gene's own 'expression signals' with those from a *S. cerevisiae* gene. The important DNA sequences which are required for the expression of eukaryotic genes are still being identified. A 'promoter' region located upstream, i.e. 5′ to the coding region, is required to direct the initiation of transcription and this region may also contain sequences involved in determining the rate of transcription initiation and regulation. In addition, a region located downstream, i.e. 3′ of the coding region, is required for efficient termination of transcription. In general, the promoter regions of eukaryotic genes are larger and more complex than those of bacteria. The promoter regions from several *S. cerevisiae* genes are being analysed to identify the key sequences for expression. Such analyses allow the identification of expression signals which can be linked to a heterologous gene to ensure maximal expression in *S. cerevisiae*.

Different *S. cerevisiae* genes are expressed with different efficiencies: for example, the glycolytic enzymes each constitute 1–5% of total cell protein and their mRNAs are correspondingly abundant. The amino-acid biosynthetic enzymes are generally present at less than 0·1% of total cell protein and their mRNAs are rare. While there is no 'consensus' *S. cerevisiae* promoter, several features are shared by genes which are expressed at similar levels and these may therefore be important determinants of transcriptional and translational efficiency. Many of these features can be illustrated by examining the promoter from the gene encoding phosphoglycerate kinase (*PGK*) (Dobson *et al.*, 1982a) as an example of a highly expressed gene (*Figure 4a*) and the promoter from the gene encoding *N*(5′-phosphoribosyl)-anthranilate isomerase (*TRP1*) (Dobson *et al.*, 1982b; *Figure 4b*) as an example of a weakly expressed gene. The most important region of the *PGK* promoter is an upstream activator sequence (UAS) located at position −350 to −450 nucleotides upstream from the initiating ATG sequence (Kingsman *et al.*,

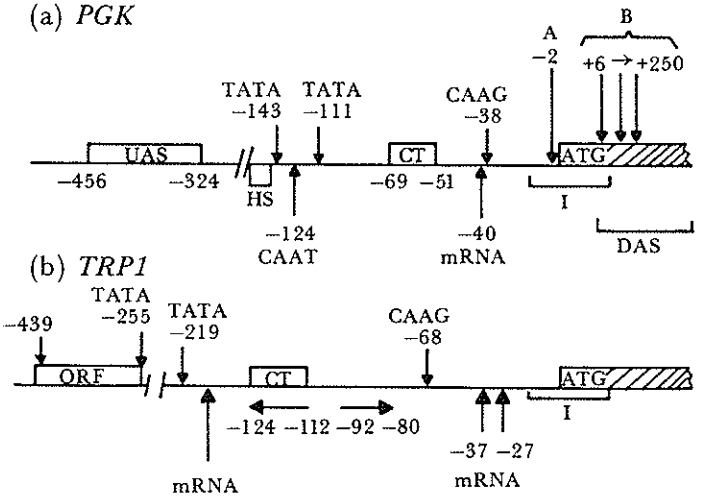


**Figure 4.** Features of two *S. cerevisiae* gene promoter regions. (a) The promoter region from the *PGK* gene (Dobson *et al.*, 1982a). Thin line = promoter region; hatched box = *N*-terminus of *PGK* coding region. The nucleotides are numbered from the initiating ATG as +1. Potential control sites are indicated by arrows and boxes. The end points of two classes of promoter fragments A and B used in expression vectors are indicated by long arrows, and are explained in the text. UAS = upstream activator sequence; TATA and CAAT refer to consensus sequences identified in higher eukaryotic gene promoters (Benoist *et al.*, 1980); CT = pyrimidine-rich tract; CAAG = mRNA start site; I = translation initiation environment; DAS = downstream activator sequence; HS = heat shock induction sequence (P. Piper and A. Lockheart, personal communication). (b) The promoter region from the *TRP1* gene (Dobson *et al.*, 1982b). ORF = upstream open reading frame; opposing arrows indicate a region of hyphenated dyad symmetry. There are three mRNA initiation sites: the most upstream maps in the region of −180. Other symbols as in (a). The drawings are not to scale.

1983; J. Ogden, unpublished results). Removal of this sequence reduces transcription by about 500-fold. Several *S. cerevisiae* genes contain UASs (Guarente, 1984; Sarokin and Carlson, 1984) which have been shown to be necessary for full transcriptional activity. It is not known whether all UASs function in the same way or if they are analogous to enhancers found in higher eukaryotic promoters (reviewed in Borrelli, Hen and Chambon, 1984). It is likely that other features of the promoter in addition to the UAS 'set' the maximum transcriptional activity. Some genes have been shown to have a negative regulatory region, e.g. *CYC7* (Wright and Zitomer, 1984) and *PGK* (J. Mellor and J. Ogden, unpublished results), and this may interact with the UAS to determine the balance of transcription (Guarente, 1984). The discovery of UASs and the general complexity of yeast transcriptional signals means that large promoter fragments, i.e. at least 500 bp, must be used to direct efficient heterologous gene expression; this is in contrast to the situation in *E. coli* where 70–100 bp is sufficient (reviewed in Kingsman and Kingsman, 1983). The discovery of negative elements suggests that promoter function might be enhanced by the deletion of these regions.

The efficient and accurate initiation of transcription of higher eukaryotic genes depends upon the presence of a TATA or related sequence (Benoist *et al.*, 1980; Grosveld *et al.*, 1982) located 25 to 30 bp upstream from the mRNA initiation site. TATA boxes have been identified upstream of many *S. cerevisiae* genes and are important for expression (Struhl, 1981); however, in many cases they are found further upstream from the mRNA initiation site than in mammalian genes (Dobson *et al.*, 1982a). In the *PGK* gene there are TATA boxes at −111 and −143, i.e. 61 and 103 nucleotides upstream from the mRNA start (Dobson *et al.*, 1982a). Many mammalian genes have the sequence 5'GCC/TCAATCT (the CAAT box) located at about 70 nucleotides upstream from the TATA box (Benoist *et al.*, 1980; Grosveld *et al.*, 1982). Few *S. cerevisiae* promoters have a sequence analogous to the CAAT box although a sequence at −129 in the *PGK* promoter shows a partial match with the CAAT consensus. A feature shared by many efficiently expressed *S. cerevisiae* genes is a pyrimidine-rich tract (the CT block); this is located 8–12 bp upstream from the major mRNA initiation site (Dobson *et al.*, 1982a) which is often located in the sequence PyAAG (Dobson *et al.*, 1982a; Burke, Tekamp-Olsen and Najarian, 1983). In efficiently expressed genes the spacing between the CT block and the mRNA initiation site is between 8 and 12 bp whereas, although there is a CT–CAAG structure in the promoters of some less efficiently expressed genes such as *TRP1*, the spacing is much greater, e.g. 40 bp in *TRP1* (Dobson *et al.*, 1982b). It is possible that the spatial arrangement of the CT block and the PyAAG sequence may contribute to the efficiency of a promoter. In addition the full activity of the promoter may require downstream activator sequences, DASs, which are sequences located within the coding region of the transcriptional unit (*see* pages 394–395). The *PGK* promoter region contains some features which may not be shared with promoters from other efficiently expressed genes. For example the expression of the *PGK* gene can be increased two- to five-fold by a temperature shift from 25°C to 38°C and this may be mediated by a specific sequence in the promoter region (P. Piper and A. Lockheart, personal communication).

The important features of the inefficiently expressed *TRP1* gene are shown in *Figure 4b*. The structure differs from the *PGK* promoter in several respects. There are three major transcripts which initiate close to a region of hyphenated dyad symmetry ('hairpin loop'); a CT block is contained within the stem of this 'hairpin loop' but although there is a CAAG sequence this is located 40 bp downstream and is not a transcription initiation site (Dobson *et al.*, 1982b; J. Mellor and P. James, unpublished results). There is an open reading frame (ORF) which could encode a 52-amino-acid peptide, located upstream from the transcription start sites. It has been suggested that the gene product of upstream ORFs might be involved in regulation of expression of the downstream gene (Andreadis *et al.*, 1982), although deletion of the upstream ORF in the *LEU2* gene had little effect (Martinez-Arias, Yost and Casadaban, 1984). The fact that many genes encoding amino-acid biosynthetic enzymes contain an ORF-hairpin loop structure (Dobson *et al.*, 1982b) suggests, however, that it may have some function, and recent studies with the *TRP1* gene support this idea (S. Kim, A.J. Kingsman and S.M.

Kingsman, unpublished results). Although this type of structure is predominant in the promoter region of poorly expressed genes, it has also been found in the more efficiently expressed *GAP* gene (Edens *et al.*, 1984). Many of the genes encoding amino-acid biosynthetic enzymes are co-ordinately regulated by the absence of single amino acids and this general amino-acid control appears to be mediated by a specific sequence which has the core sequence 5'TGACTC–3' (Donahue *et al.*, 1983). The *TRP1* gene is not subject to general control and contains a region of only partial homology with this control sequence (Dobson *et al.*, 1982b).

In addition to the promoter features outlined for *PGK* and *TRP1* there will be specific regulatory sequences associated with different genes, e.g. the mating-type control sequences found on the promoter of the *MAT* genes (Siciliano and Tatchell, 1984). In addition, some promoters do not contain many of the elements discussed here, e.g. the Ty element delta promoter (Bowen *et al.*, 1984) and the *MF*α promoters (Kurjan and Herskowitz, 1982; Singh *et al.*, 1983). The analysis of these promoters may reveal additional elements which determine the efficiency and regulation of gene expression.

Several sequences have been implicated as important for transcription termination in *S. cerevisiae*. Bennetzen and Hall (1982a) have noted that transcription of several *S. cerevisiae* genes terminates at the sequence 5'-TAAATAAA/G, while Zaret and Sherman (1982) have shown that deletion of a different specific sequence in the 3' flanking DNA of the *CYC1* gene abolishes correct transcription termination. A related sequence has been found in other *S. cerevisiae* genes, and the consensus 5'TAG...TAGT...ATrich...TTT, where the spaces between the blocks is variable, has been proposed. The sequence 5'–TTTTTATA is located 50–90 bp upstream of the polyadenylation site in several *S. cerevisiae* genes and has been shown to be important for termination of transcription of a *Drosophila* gene (*ADE8*) in *S. cerevisiae* (Henikoff, Kelly and Cohen, 1983).

Unlike prokaryotic genes, which contain sequences in the 5' untranslated region of the transcript that form an essential ribosome-binding site (Shine and Dalgarno, 1975, reviewed in Kingsman and Kingsman, 1983) most *S. cerevisiae* genes lack any obvious ribosome-binding site although homologies with rRNA have been noted in some cases (Zalkin and Yanofsky, 1982). In addition, many studies suggest that precise sequence conservation around the ATG may not be important for efficient translation (Sherman and Stewart, 1982) although in accordance with Kozak's rules (Kozak, 1984) there is usually an adenine at −3 (Dobson *et al.*, 1982a).

### Vectors designed to facilitate the expression of heterologous genes

The 'promoter regions' from a number of different *S. cerevisiae* genes have now been used to construct vectors for the expression of heterologous proteins. One of the most efficient systems exploits the promoter region of the *PGK* gene: the *PGK* vectors (Dobson *et al.*, 1982a; Tuite *et al.*, 1982; Mellor *et al.*, 1983, 1985) will be discussed to illustrate various concepts of expression vector design and exploitation. Promoter fragments were gener-

ated from the *PGK* gene by removing the coding sequence by deletion with the exonuclease *Bal* 31 starting from a site within the *PGK* coding sequence (Dobson *et al.*, 1982a). This generated a series of fragments with end points in the *N*-terminal coding region of *PGK* and the 5′ flanking region. A synthetic oligonucleotide linker was ligated to the ends of the fragments to produce a range of convenient fragments on high copy number vectors based on the plasmid pMA3a (Dobson *et al.*, 1982a). Heterologous genes can be inserted at the unique restriction enzyme site at the end of the *PGK* promoter fragment. Two types of *PGK* promoter fragments have been used: one fragment (shown as A in *Figure 4a*) contains 1·5 kb of the promoter region terminating at position −2 upstream from the initiating ATG; it is present in expression vectors pMA301 (Mellor *et al.*, 1985; *Figure 5a*) and pMA91 (Mellor *et al.*, 1983) and is used to express coding sequences which have their own initiation codon. The second class of promoter fragments (labelled B in *Figure 4a*) terminate at different points within the *PGK* coding sequence: for example, in the expression vector pMA230 the end point is at +37 (Tuite *et al.*, 1982). These can be used to produce fusion proteins to express heterologous coding sequences which lack an initiating ATG. Fragments with different end points can be used to ensure that the correct reading frame for translation can be maintained for any heterologous coding sequence. To test these vectors an IFN-$\alpha_2$ cDNA was engineered to remove most of the signal sequence and then inserted into the unique *Bam*HI or *Bgl*II expression sites to produce IFN expression plasmids such as pMA301–1 (*Figure 5b*). IFN is produced at about $5 \times 10^6$ molecules per cell in *S. cerevisiae* strains transformed with such a vector, thereby allowing the production of at least 30 mg IFN per litre of simple batch culture. The IFN represents at least 1% of the total soluble protein (*see Figure 7*). When a *TRP1* promoter fragment was used to direct IFN-$\alpha_2$ expression on a similar plasmid, yields were only $2 \times 10^3$ molecules per cell (Dobson *et al.*, 1982b). The differences in IFN yield directly reflect differences in the respective IFN mRNA levels and illustrate the fact that *PGK* and *TRP1* promoters differ markedly in their transcriptional efficiencies.

The IFN cDNA used in these studies contains a fortuitous transcription-termination signal which is recognized in *S. cerevisiae* to produce a discrete transcript. If this termination signal is removed from the cDNA, or if a heterologous sequence which lacks any fortuitous termination signals is used in these vectors, the yields drop by at least tenfold (Mellor *et al.*, 1983, 1985). In the absence of transcription-termination signals proximal to the end of the gene, long transcripts are produced which terminate at distant sites in the *LEU2*/2 μm region (J. Mellor, unpublished results; Mellor *et al.*, 1985). The levels of these transcripts are low, probably, as suggested by Zaret and Sherman (1984), because long transcripts are unstable. To ensure that any heterologous DNA sequence can be maximally expressed, a vector, pMA91, has been constructed which contains both the *PGK* promoter and transcription-termination signals (Mellor *et al.*, 1983). Any DNA fragment can be inserted into the unique *Bgl*II site and discrete transcript is produced. *Figure 6* shows a mature IFN-coding sequence inserted into the vector pMA91.

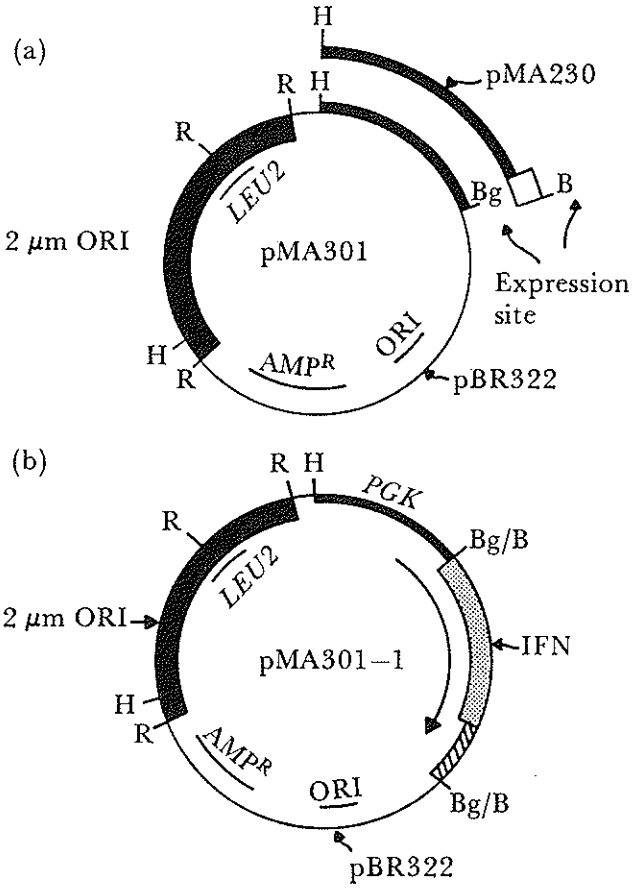


**Figure 5.** Expression vectors. (a) The vectors pMA301 and pMA230. Thin line = pBR322; closed box = *LEU2*/2 μm fragment; thick line = *PGK* 5′ flanking region; open box = *PGK* coding sequence; pMA301 contains *PGK* promoter fragment A (*see Figure 4*); pMA230 is the same as pMA301 except that it contains a class B promoter fragment terminating at +37 between the PGK coding sequence (*Figure 4*). The unique expression sites are indicated; heterologous genes can be inserted into the *Bgl* II site of pMA301 or the *Bam* HI site in pMA230. Restriction sites are as in previous figures. (b) The expression vector pMA301-1. A *Bam* HI fragment containing the coding (stippled box) and 3′ untranslated region (hatched box) of a human IFN-α cDNA is inserted into the *Bgl* II expression site of pMA301. The direction and size of the heterologous transcript is shown by the arrow. Bg/B = *Bgl* II/*Bam*HI junctions. This vector directs the production of IFN-α as about 1% of total cell protein (*see Figure 7*, lane d).

Transcription initiates in the *PGK* promoter, proceeds through the heterologous coding sequence and terminates in the *PGK* terminator; the mRNA is translated to produce authentic IFN. pMA91 is typical of the vectors currently being used to direct high-level expression of heterologous genes in *S. cerevisiae*. Different *S. cerevisiae* promoter and terminator sequences have been used to construct these vectors. *Table 1* lists promoters which have been cloned and characterized and which are potentially (or have proved to be) useful for the efficient expression of heterologous proteins. It is not possible

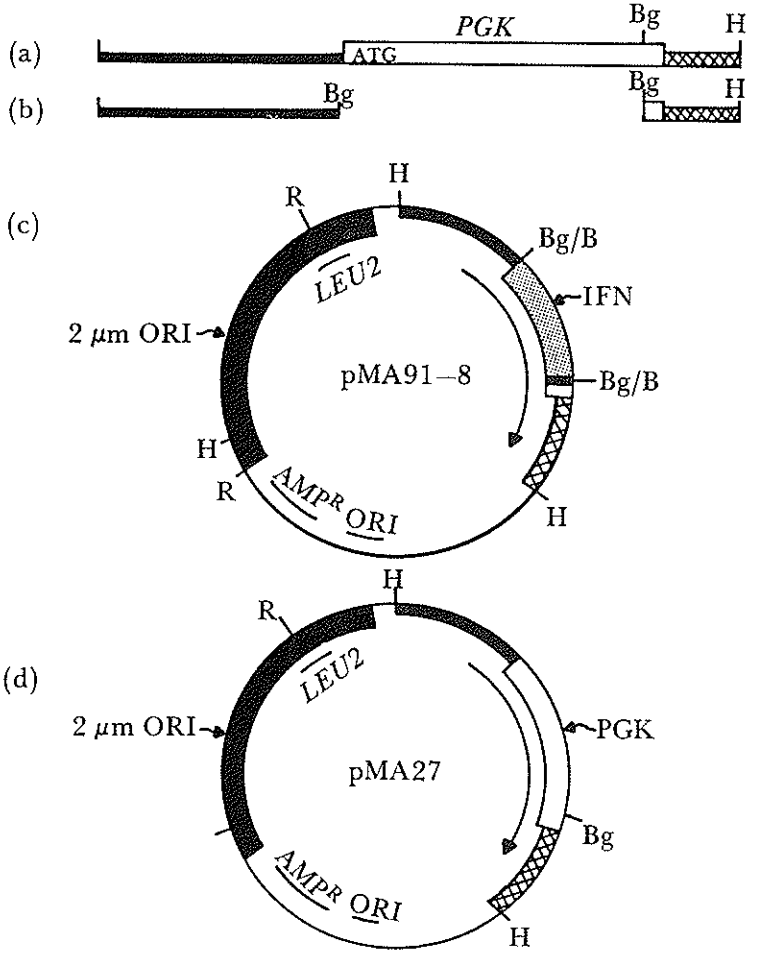


**Figure 6.** A typical high-efficiency heterologous gene expression vector and the corresponding homologous gene expression vector. (a) A 2·95 kb *Hind* III fragment containing the *PGK* gene; thick line = *PGK* promoter region; open box = *PGK* coding sequence; cross-hatched box = *PGK* transcription-termination region. (b) Fragments of the *PGK* gene used to construct the high-efficiency expression vector pMA91 (Mellor *et al.*, 1983). The promoter fragment terminates with a synthetic *Bgl* II linker at nucleotide −2, the terminator fragment begins with the natural *Bgl* II site in the carboxy-terminal *PGK* coding region. (c) The interferon (IFN) expression vector pMA91-8, consisting of pMA3a; the *PGK* signals shown in (b) and part of an IFN cDNA which contains only the coding sequence and no 5′ and 3′ non-coding DNA; this is fragment 8 of Mellor *et al.* (1985). The symbols are as in (a) and *Figure 5* except that the vertical box indicates the translation-termination signal which separates the IFN coding sequences from the *PGK* 3′ coding sequence which is therefore not translated. (d) Plasmid pMA27 which is similar to pMA91–8 except that the normal *PGK* promoter region is used and there is therefore no synthetic *Bgl* II linker and the *PGK* coding sequences are present in the place of IFN coding sequences. The drawings are not to scale.

**Table 1.** Promoters for use in expression vectors*

| Gene | Reference |
|---|---|
| *PGK* | Dobson *et al.*, 1982a; Tuite *et al.*, 1982. |
| *ADH1* | Bennetzen and Hall, 1982a; Hitzeman *et al.*, 1981. |
| *GAP3* | Holland and Holland, 1980; Edens *et al.*, 1984. |
| *PYK* | Burke, Tekamp-Olsen and Najarian, 1983. |
| *TPI* | Alber and Kawasaki, 1982. |
| *ENO* | Holland *et al.*, 1981. |
| *GAL1* | St John and Davis, 1981; Goff *et al.*, 1984. |
| *MFα* | Kurjan and Herskowitz, 1982; Brake *et al.*, 1984. |
| *PHO5* | Meyhack *et al.*, 1982; Kramer *et al.*, 1984. |
| *CUP1* | Karin *et al.*, 1984. |
| *HSP90* | Finkelstein and Strausberg, 1983. |

*This list is not comprehensive but indicates some of the best-characterized promoters which are likely to give high-efficiency expression.

to identify the most efficient promoter because comparative analyses of transcription, using the same heterologous coding sequence in the same genetic background, have not been done. The highest intracellular yields reported have, however, all been obtained using promoters from genes encoding glycolytic enzymes. Some of the terminator fragments which have been used are from *PGK* (Mellor *et al.*, 1983), human IFN-α (Tuite *et al.*, 1982), 2 μm *FLP* (Hitzeman *et al.*, 1983a), *TRP1* (Kramer *et al.*, 1984) and *PHO5* (Hinnen, Meyhack and Tsapis, 1983). There is some evidence that the use of different terminator fragments in expression vectors may influence RNA levels (Mellor *et al.*, 1985) and Zaret and Sherman (1984) have shown that variations in 3′ untranslated regions can affect both mRNA stability and translation. There has, however, been no detailed analysis of the precise termination sequence requirements for maximizing heterologous gene expression.

### Factors affecting heterologous gene expression

The choice of promoter is critical for achieving maximum levels of expression, but even the use of high-efficiency promoters results in yields of heterologous products which are generally less than 10% of total cell protein. For example, the yields of IFN-$\alpha_2$ directed by the *PGK* promoter are less than 5% of total cell protein (*Figure 7*). In contrast, when the entire *PGK* gene is incorporated into pMA3a (plasmid pMA27, *Figure 6d*), the PGK protein is produced as 50–80% of total cell protein, a level which directly reflects the increase in gene dosage. It seems likely that several factors may limit the yields of heterologous products and these may be different in each case. In determining the factors which might affect the yields of heterologous product, all stages of gene expression must be considered. These are the stability of the final protein product, the efficiency of translation of the heterologous RNA, the stability of the heterologous transcript and the efficiency of synthesis of heterologous RNA. In addition, the copy number of the expression vector must be determined to ensure that low yields are not simply a reflection of

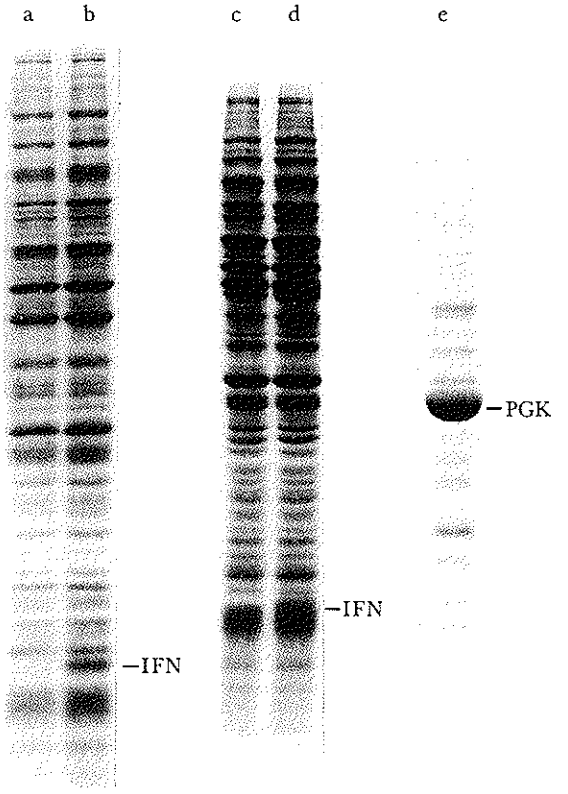


**Figure 7.** The synthesis of homologous and heterologous proteins in *S. cerevisiae*. A Coomassie blue-stained SDS-polyacrylamide gel of total soluble proteins extracted from control *S. cerevisiae* strains containing expression vectors pMA230 and pMA301 (lanes a and c), from strains containing the same vectors but with an IFN cDNA (fragment 1, Tuite *et al.*, 1982) inserted at the expression site to produce pMA230–1 (lane b) and pMA301–1 (lane d) and from a strain containing pMA27 (*Figure 6b*) which expresses the PGK protein. The IFN and PGK polypeptides are marked. These data show that IFN is produced in *S. cerevisiae* containing IFN expression plasmids; that similar yields are obtained when IFN is made as a fusion protein (lane b) and as an authentic protein (lane d) and that the levels of IFN are about 1% of total cell protein. They also show that the expression of the homologous gene is more efficient than expression of a heterologous gene because PGK protein is produced as at least 50% of total cell protein. These data are discussed in detail in Mellor *et al.* (1985).

reduced gene dosage. The factors limiting the yields of several heterologous proteins are currently being determined. Some heterologous proteins may be very unstable in *S. cerevisiae*: e.g. human insulin can be detected only when a large fusion is made with galactokinase, suggesting that the insulin sequence can be translated but that the small polypeptide is rapidly degraded (Stepien *et al.*, 1983). Instability might also explain the failure to detect rat growth hormone and the low levels of human epidermal growth factor (hEGF) which are produced despite the presence of specific mRNA (Ammerer *et al.*, 1981; Urdea *et al.*, 1983). Other proteins are, however, reasonably stable: pulse

labelling and chase experiments have shown that IFN-$\alpha_2$ has a half-life of 1–2 h; this is shorter than that for many yeast proteins, including PGK, but cannot account for all the differences in the steady-state yields. This is confirmed by comparing the levels of PGK and IFN-$\alpha_2$ after a 10-minute pulse label, which indicates at least a 10-fold difference. This must be attributable to differences in protein synthesis rather than degradation (Mellor *et al.*, 1985). The synthesis of heterologous proteins could be limited at translation or at transcription. Efficient initiation of translation of heterologous proteins is, however, easy to achieve because there are no stringent requirements for translation initiation in *S. cerevisiae* (Sherman and Stewart, 1982). This is supported by the finding that yields of human IFN-$\alpha_2$ produced as a PGK–IFN fusion protein show little difference from yields of IFN produced as the authentic protein (*Figure 7*). In the first case the normal translation initiation environment of PGK is used; in the second there is a synthetic linker immediately preceding the initiating ATG of the interferon sequence. This linker is, and should be, relatively A rich, as G-rich translation initiation environments may be slightly (three- to six-fold) less efficient, but there is not a rigorous sequence requirement (Kingsman and Kingsman, 1983). In some cases inefficient translation elongation might limit expression. Highly expressed *S. cerevisiae* genes show a marked bias in codon usage (Bennetzen and Hall, 1982b) and the abundance of charged tRNAs reflects this bias (Ikemura, 1982). Heterologous coding sequences do not show the same bias and therefore translation may be limited by the requirement for rare charged tRNA species. There is, however, little direct evidence to support the idea that codon bias is responsible for the low yields of heterologous products. In fact, in the case of IFN there is evidence that codon bias is not important. We have shown that the 10-fold difference in the synthesis of PGK and IFN can be explained by the finding that the steady-state levels of IFN mRNA are at least ten times lower than the levels of PGK mRNA (Mellor *et al.*, 1985; *Figure 8*). If the differences in protein levels had been due to inefficient translation, then the mRNA levels would have been the same. The plasmids expressing PGK and IFN-$\alpha_2$ are identical except for the coding sequence (*Figure 6, c, d*) and they are present at similar high copy numbers (*Figure 8*). This means that the presence of the heterologous coding sequence or the absence of the *PGK* coding sequence reduces the level of IFN-specific transcripts. Similar results have been obtained with a calf prochymosin coding sequence (Mellor *et al.*, 1983) implying that a low RNA level is not a heterologous-sequence-specific phenomenon. Our preliminary results suggest that IFN-$\alpha_2$ mRNA is not degraded more rapidly than *PGK* RNA (J. Mellor, unpublished results). Furthermore, in a construction where the entire *PGK* coding sequence has been fused to the entire IFN-$\alpha_2$ coding sequence to produce a hybrid PGK–IFN-$\alpha_2$ transcript, the levels of the hybrid transcript approach those of the *PGK* transcript (J. Mellor, M.J. Dobson, N.A. Roberts, A.J. Kingsman and S.M. Kingsman, unpublished work). These data suggest that the presence of *PGK* coding sequences restores transcriptional efficiency to the *PGK* promoter. This result is consistent with the findings of internal enhancers in other eukaryotic genes (Osborne *et al.*, 1984; Charnay *et*

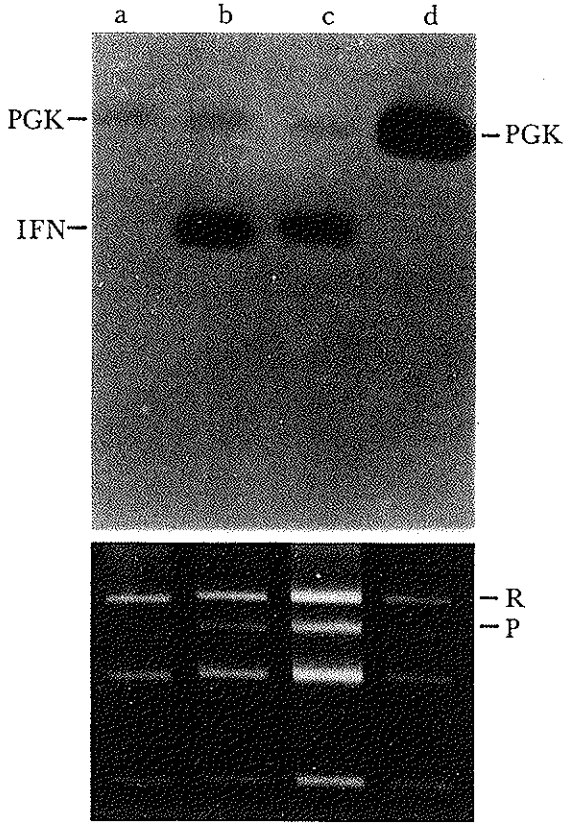


**Figure 8.** The synthesis of homologous and heterologous mRNA in *S. cerevisiae*. An autoradiograph of a Northern blot is shown. Equal amounts (20 μg) of total RNA from a control strain of *S. cerevisiae* containing no plasmid (lane a) and strains containing IFN expression plasmids pMA301–1 (lane b) and pMA230-1 (lane c) and PGK expression plasmid pMA27 (lane d). The filter was probed with IFN and PGK specific fragments labelled to the same specific activity. The IFN and PGK specific RNAs are indicated; the chromosomal *PGK* gene directs PGK specific RNA in all the strains. These data show that pMA27 directs about ten times more RNA than similar plasmids expressing a heterologous gene; they are discussed in detail in Mellor *et al.* (1985). The lower panel shows an ethidium bromide-stained gel of an *Eco* RI digest of total DNA isolated from each strain. The ribosomal repeated DNA (R) and a plasmid-specific band (P) are indicated. These data show that there is little difference in copy number of the plasmids in each strain.

*al.*, 1984; Wright *et al.*, 1984). We have suggested that a downstream activator sequence (DAS) is present in the *PGK* coding region (J. Mellor, M.J. Dobson, N.A. Roberts, A.J. Kingsman and S.M. Kingsman, unpublished work). The mechanism of action of the DAS is not clear but it may function as an enhancer. It is not known how far the data concerning expression of IFN directed by the *PGK* promoter will be relevant to other expression systems. It is possible, for example, that DASs will be present only in *PGK*. Furthermore, transcript stability and codon bias may be significant for some heterologous genes. Codon bias may also assume greater significance if the

levels of heterologous RNA can be improved. There may then be a limitation on the availability of charged tRNAs with a consequent reduction in translation efficiency. In our experience of analysing the expression of many different coding sequences directed by the *PGK* promoter, the predominant reason for low yields is low RNA levels, although for some proteins we have seen in addition high protein instability and reduced plasmid copy number, presumably attributable to selection against high doses of a toxic product.

**Plasmid copy number, stability and integrity**

The production of PGK as 50–80% of total cell protein has no effect on the growth characteristics of the *S. cerevisiae* strain, with doubling times in selective media being about 2·5 h. The production of IFN as 3% of total cell protein also has no effect. Plasmids expressing IFN are maintained at high copy number, about 100 copies per cell, and more than 99% of the cells contain plasmid even after prolonged growth (40 generations) in non-selective media (Mellor *et al.*, 1985). This is not true for the production of all heterologous proteins: for example, the production of prochymosin as 5% of total cell protein increases the doubling time to about 4 h, with the plasmid copy number being 100 per plasmid-bearing cell, and after 10 generations' growth in non-selective medium only 29% of the cells contain plasmid (M.J. Dobson, unpublished results). Selection must therefore be maintained to ensure high plasmid copy number and stability. The most extreme case that we have observed occurs during the production of influenza virus haemagglutinin which increases the doubling time to 11 h. In this case a high level of plasmid rearrangements, which result in the loss of the haemagglutinin sequences after prolonged culture, has also been observed. Rearrangements can be prevented by using a *rad52* (Game and Mortimer, 1974) host strain (N.A. Roberts, unpublished work). These data suggest that *S. cerevisiae* cannot tolerate high levels of some heterologous proteins: they reduce the growth rate and therefore confer a selective advantage on cells which have a reduced gene dosage. One way to overcome any problems with toxic proteins is to limit their production to the end of the culture period by using a regulatable system.

**Regulation of heterologous gene expression**

The first regulated system to be described used the *PGK* promoter to direct expression of IFN. The activity of the *PGK* promoter is reduced when *S. cerevisiae* is grown on non-glycolytic carbon sources such as glycerol, ethanol or acetate and induced by growth on glucose. By growing cultures in acetate, IFN yields were reduced to $7 \times 10^4$ molecules/cell and by 8 h after transferring the culture to glucose they were increased to about $10^6$ molecules/cell. This system has advantages in fermentations where carbon sources other than glucose are used or where high concentrations of ethanol are produced. The *PGK* promoter can be activated at the end of the fermentation by the introduction of glucose. This system, however, produces only about a 20-fold

induction ratio and the uninduced levels of synthesis may still be too high for the efficient production of toxic proteins. A more stringent regulation can be achieved using the *PHO5* promoter. The transcription of *PHO5* is tightly repressed when inorganic phosphate is present in the growth medium and induced by depletion of inorganic phosphate (Bostian *et al.*, 1980; Kramer and Andersen, 1980). When IFN-$\alpha_1$ is fused to the *PHO5* promoter the level of IFN produced in low-phosphate medium is about 200-fold higher than in high-phosphate medium (Kramer *et al.*, 1984). Depletion of phosphate may not be convenient on a large scale and *PHO5*-directed expression can also be regulated by using temperature-sensitive regulatory mutants. The expression of acid phosphatase is positively regulated by the *PHO4* gene product and repressed by the *PHO80* product. In a *pho4*$^{ts}$, *pho80* mutant at the non-permissive temperature (36°C), little acid phosphatase is produced because of the lack of the positive regulator, whereas at the permissive temperature (24°C) acid phosphatase is produced constitutively and independently of phosphate concentration because of the lack of repressor. *PHO5* promoter-directed heterologous gene expression can be induced 50-fold in this mutant by reducing the temperature from 36°C to 24°C. Although the *PHO5* promoter can be regulated, the maximum induced yields appear to be 10- to 20-fold lower than with 'glycolytic gene' promoters such as *PGK* (Kramer *et al.*, 1984). The *GAL1* promoter is efficient and can also be regulated: expression requires a positive regulator (*GAL4*) and there is about a 1000-fold induction by galactose (St John and Davis, 1981). A 335 bp fragment of the *GAL-10* divergent promoter region confers galactose inducibility and will function with other promoters (Guarente, Yocum and Gifford, 1982). Maximum expression directed by the *GAL1* promoter on a multicopy plasmid can, however, be achieved only by ensuring overproduction of the *GAL4* gene products in the same cell (Johnston and Hopper, 1982). Furthermore, to ensure low uninduced levels of expression of toxic proteins the *GAL80* repressor protein may also have to be overproduced. Expression directed by the *MFα1* promoter can be regulated: α-factor expression requires the products of the *SIR* genes (Herskowitz and Oshima, 1982); α-factor expression is repressed in a *sir*$^{ts}$ mutant at 37°C but expression is induced 1000-fold by a shift down to 24°C. Using the *MFα1* promoter, human epidermal growth factor (hEGF) was produced at less than 10 ng/litre at 37°C and levels increased to 4 mg/litre within several hours of shifting to 24°C (Brake *et al.*, 1984). The *MFα* gene promoter can be used only in haploid strains as it is repressed in a/α diploids. Other potentially useful regulated promoters are a heat-shock protein gene which is induced 50-fold by heat shock, anoxia or high cell density (Brazzell and Ingolia, 1984) and the *CUP1* gene encoding copper resistance which has a 20-fold induction ratio (Karin *et al.*, 1984). It may also be possible to isolate genes which show regulation under specific conditions by 'shotgunning' into *lacZ* vectors (Ruby, Szostak and Murray, 1983). For example, genes which are switched on only at the end of a batch fermentation may be useful as the majority of yeast genes are switched off and therefore purification of the heterologous product may be simplified: the high level of intracellular degradative enzymes in stationary-

phase cultures must, however, be considered in designing the precise expression strategy.

Another approach to regulating expression is to use bacterial regulation signals. The *E. coli lexA* operator has been inserted between $UAS_G$ of *GAL1* and the mRNA initiation site in a *S. cerevisiae* strain which has been engineered to produce *lexA* protein. The activity of $UAS_G$ is reduced four- to 10-fold in the presence of *lexA* (Brent and Ptashne, 1984). Although the biological basis of this effect on $UAS_G$ function is not clear, the experiments suggest the possibility of designing an effective regulated promoter. For example, the insertion of *lacO* into an efficient promoter in a *S. cerevisiae* strain that produces a *lac* repressor might provide a much higher induction ratio in response to the simple addition of lactose (Reznikoff and Abelson, 1980).

### Secretion of heterologous proteins

For many proteins, particularly pharmaceuticals, it is important that the completely authentic protein is produced. If it is not, there may be problems of aberrant biological activity and antigenicity. Many eukaryotic proteins of medical and commercial interest are normally secreted and are therefore produced as precursor proteins with an amino-terminal signal sequence that is proteolytically removed to produce the mature protein. The first amino acid of the mature protein is not necessarily a methionine and consequently a synthetic ATG has to be added to the mature coding sequence if it is to be expressed in *S. cerevisiae*; this will result in the production of a protein with an additional methionine residue at the *N*-terminus. The methionine residues are removed from *S. cerevisiae* proteins if they precede certain amino acids, notably threonine, alanine and glycine, but *N*-terminal methionines are not removed if they precede isoleucine, leucine, methionine, aspartate, lysine, arginine, glutamine and in some cases valine (Sherman and Stewart, 1982). It is not clear how reliably the methionine is removed from heterologous proteins. One way of circumventing this problem is to produce the precursor protein, for example, pre-interferon, and then remove the signal. This would be efficiently achieved if the precursor protein could be secreted by *S. cerevisiae*.

The *S. cerevisiae* secretion system appears to be remarkably similar to that in higher eukaryotic cells. The same organelles are used; the cotranslational translocation into the endoplasmic reticulum (ER) of proteins destined for secretion is mediated by an *N*-terminal signal peptide on the protein; *N*-glycosidically linked core oligosaccharides are added in the ER lumen and additional glycosyl modifications occur in a Golgi-like complex. The proteins are packaged into secretory vesicles and delivered to the cell surface (secretion is reviewed in detail in Schekman and Novick, 1982). Several studies have now shown that heterologous signal sequences will direct secretion in *S. cerevisiae*. In these studies efficient *S. cerevisiae* promoters such as *PGK* have been placed in front of sequences coding for the precursor protein (*Figure 9a(i)*). Heterologous proteins are secreted only if the signal

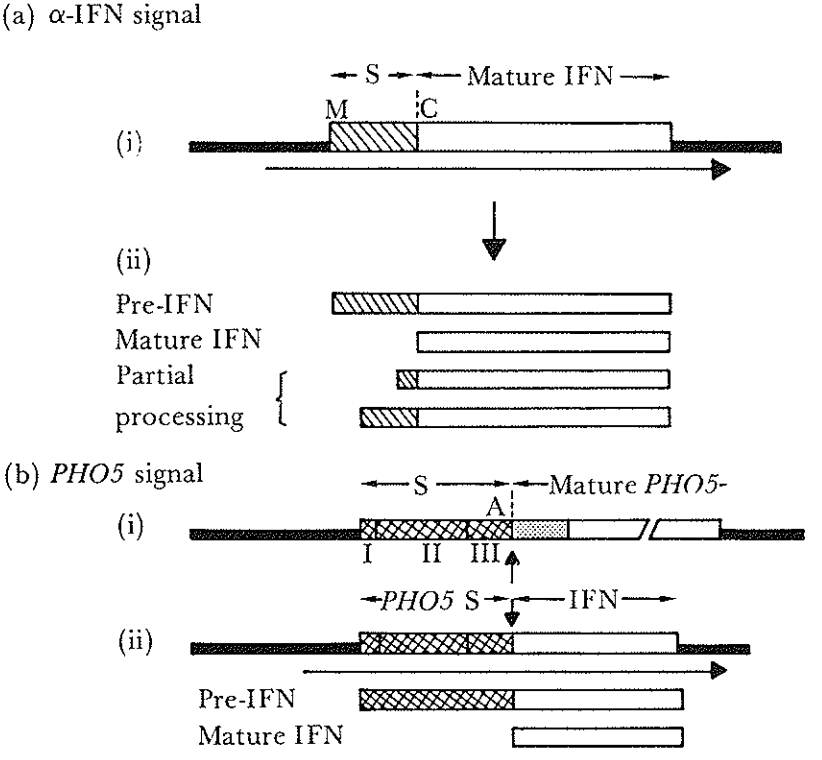


**Figure 9.** Secretion vectors using heterologous and homologous 'classical' signal sequences. (a) The secretion of a heterologous protein, IFN, using the heterologous gene's signal sequence. (i) The expression configuration; thick line = *S. cerevisiae* gene promoter and terminator regions; hatched box = IFN signal sequence; open box = mature IFN sequence. In this example IFN-α has a 21-amino-acid signal sequence starting at the initiating methionine (M) residue and a 166-amino-acid mature sequence starting with a cysteine (C). The arrow indicates the transcript. (ii) Indicates the protein products; these are preIFN and various processed derivatives including mature IFN (*see* text). (b) The secretion of a heterologous protein using a 'classical' *S. cerevisiae* signal sequence from acid phosphatase (*PHO5*) (i) The *PHO5* gene, thick lines = promoter and terminator regions; cross-hatched box = signal sequence which has three domains terminating with an alanine (A) residue adjacent to the cleavage site indicated by the arrows; stippled box = part of the *PHO5* coding region which might be required for efficient cleavage; open box = *PHO5* coding region. (ii) A *PHO5* secretion-vector configuration; the *PHO5* promoter (thick line) and signal sequence up to the terminal alanine residue (hatched box) are fused to a mature IFN-coding sequence; the gene products are pre-IFN and mature IFN.

sequence is present, indicating that a specific process, rather than cell lysis, is occurring. Secretion has been detected using the signal sequences of human IFN-α and IFN-γ (Hitzeman *et al*, 1983a), *E. coli* β-lactamase (Roggenkamp, Kustermann-Kuhn and Hollenberg, 1981), wheat α-amylase (Rothstein *et al.*, 1984), plant thaumatin (Edens *et al.*, 1984) and mouse immunoglobulin light and heavy chains (Wood *et al.*, 1985). In most cases the overall levels of synthesis were significantly lower when the preprotein sequence was expressed, as compared with the mature protein (e.g. Hitzeman *et al.*, 1983a). Furthermore, there is some evidence from studies of β-lactamase secretion in

yeast that the signal sequence reduces the translational efficiency five- to 10-fold (R. Roggenkamp, H. Dargatz and C.P. Hollenberg, personal communication). In addition, only a small fraction of the total preprotein synthesized is secreted: for example less than 10% of human IFN-α is found in the culture medium, 10% is found in association with the cell wall, and the remainder is intracellular. The yield of truly secreted interferon therefore represents less than 0·01% of total cell protein (Hitzeman *et al.*, 1983a). Similarly, secreted β-lactamase represents less than 0·2% of total cell protein (Roggenkamp, Dargatz and Hollenberg, 1985). Most of the precursor β-lactamase is associated with polysomes and this intracellular form has no biological activity, implying incorrect folding. The preprotein is also more susceptible to proteolysis (Roggenkamp, Dargatz and Hollenberg, 1985). The lower yields of preproteins may therefore be due to a combination of reduced translational efficiency and enhanced proteolysis due to the retention of the incorrectly folded precursor in the cytoplasm. The fact that much of the precursor protein is not cotranslationally translocated to the ER implies either that the heterologous signal is inefficiently recognized or that some component of the secretory pathway is limiting. In many cases a molecular weight estimation from SDS polyacrylamide gels is consistent with the heterologous signal sequence being removed in *S. cerevisiae*. In the case of IFN, however, although 64% of the secreted molecules were processed accurately, the remainder had retained three amino-acid residues of the signal sequence. Furthermore, the intracellular protein had also been processed to yield the authentic IFN, plus three-residue IFN and an additional species with eight signal amino acids (Hitzeman *et al.*, 1983a, *Figure 9a (ii)*). Signal sequences are not always removed: for example, pre-$\alpha_1$-antitrypsin is not processed or secreted (Cabezon *et al.*, 1984). These studies indicate that secretion of heterologous proteins by yeast is possible but that using heterologous signal sequences results in low yields and may produce a collection of unprocessed, processed and incorrectly processed molecules. An alternative approach is to fuse the signal sequence from secreted *S. cerevisiae* proteins on to the mature coding sequence of a heterologous protein. The resulting fusion protein should be efficiently secreted and the *S. cerevisiae* signal sequence should be removed accurately by the normal processing enzymes. To date, three systems have been developed: these use the signal sequences from acid phosphatase (*PHO5*), invertase (*SUC2*) or α-factor (*MF*α). Acid phosphatase and invertase possess 'classical' signal sequences (Perlman and Halvorsen, 1983) with three domains: a charged *N*-terminus (I), a central hydrophobic core (II), and a consensus sequence for cleavage by the signal peptidase after an alanine residue (III) (*Figure 9b (i)*). The 17-amino-acid signal sequence of acid phosphatase will direct heterologous proteins to the periplasmic space (Hinnen, Meyhack and Tsapis, 1983) although there is some evidence that efficient cleavage by the signal peptidase may require some amino acids from the *N*-terminus of mature acid phosphatase (Haguenauer-Tsapis and Hinnen, 1984). If additional amino acids are required to maximize secretion, these may have to be removed from the heterologous protein *in vitro*, thereby removing one of the advantages of a

secretion system. Invertase and acid phosphatase are both normally secreted into the periplasmic space, with less than 5% of the total protein being found in the culture medium. At present there is no strong evidence that their signal sequences will direct other proteins into the culture medium rather than the periplasmic space. Secreted proteins located in the periplasmic space will be more difficult to extract and purify than those in the culture medium.

Secretion vectors based on these homologous 'classical' signal sequences may be subject to the same yield limitations as observed with heterologous signals: for example, the over-production of acid phosphatase on a high-copy-number plasmid (Haguenauer-Tsapis and Hinnen, 1984) results in the intracellular accumulation of precursor protein. It is possible that, irrespective of the levels of synthesis which can be achieved using a high-copy-number plasmid and a high-efficiency promoter, only a low absolute number of molecules may be secreted. This might be overcome by identifying the limiting factor(s); if, as suggested by Haguenauer-Tsapis and Hinnen (1984), it is the signal peptidase, this could be over-produced in the host cell.

A potentially more promising secretion system uses the secretion and processing signals from the yeast mating pheromone, α-factor. This is a 13-amino-acid peptide which is secreted into the culture medium and conditions cells of the opposite mating type for conjugation (reviewed in Herskowitz and Oshima, 1982). It is synthesized as a 165-amino-acid precursor protein; there is a 22-amino-acid putative signal peptide which probably directs the protein to the ER and a 61-amino-acid pro-segment, the function of which is not clear but which contains three glycosylation signals and may direct the precursor into extracellular, rather than periplasmic or integral membrane, secretory pathways. The α-factor peptide is repeated four times and repeats are separated by spacer peptides which have the structure, lys–arg, followed by two or three dipeptides which are either glu, ala or asp, ala. The α-factor peptide is produced from the precursor by three processing events: there is a cathepsin B or trypsin-like cleavage after the lys–arg residues; the *N*-terminal dipeptides are removed by a dipeptidyl-aminopeptidase, and the *C*-terminal basic residues are removed by carboxy-peptidase B (*Figure 10a*). The cloning of the α-factor gene, *MFα1* and the characterization of the precursor and processing events are described in Kurjan and Herskowitz (1982), Julius *et al.* (1983) and Julius, Schekman and Thorner (1984).

Brake *et al.* (1984) have taken a fragment which carries the α-factor promoter, the prepro segment and part of the first spacer peptide terminating after the arg residue, and have fused this to a sequence encoding mature hEGF (*Figure 10b*). Biologically active hEGF was secreted into the culture medium at about 5 μg/ml. Furthermore, *N*-terminal sequencing indicated that the pre-pro–lys–arg segment had been accurately removed from 100% of the secreted hEGF molecules. Bitter *et al.* (1984) have used a similar construction to express β-endorphin and synthetic human-IFN-α coding sequence. In these constructions, however, two dipeptides of the spacer had been retained to produce a 'pre-pro–lys–arg–glu–ala–glu–ala–IFN' and the dipeptides were incompletely removed from the *N*-terminus of the heterologous proteins. A

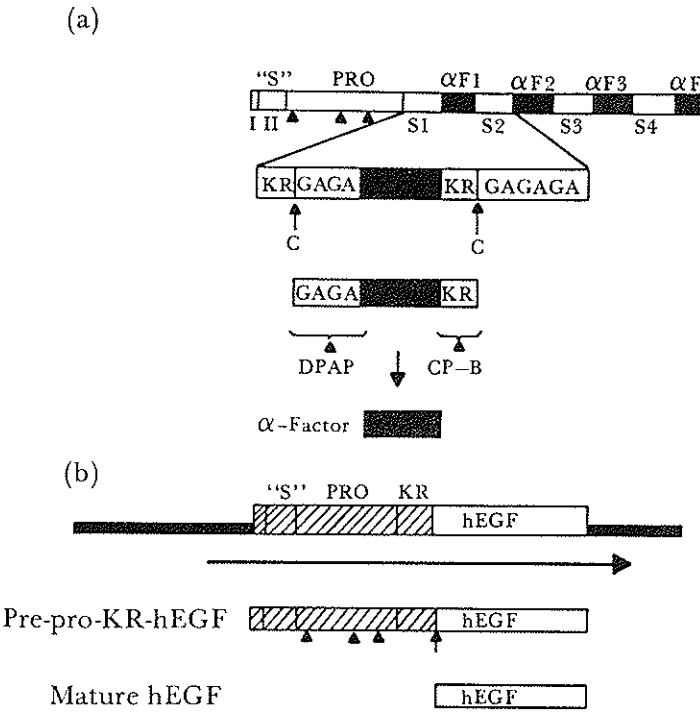


**Figure 10.** The α-factor secretion vectors. (a) The processing pathway for α-factor. A large precursor protein is produced; "S" = signal sequence; this has two domains but lacks a consensus cleavage site and may not be removed. PRO = 61-amino-acid prosegment; solid triangles indicate glycosylation sites; S1–S4 = spacer peptides; αF1–αF4 = α-factor peptides. The expanded region shows the amino acid sequence of S1 (lys, arg, glu, ala, glu, ala) and S2 (lys, arg, (glu ala)$_3$) flanking αF1. α-Factor is produced by cleavage at lys arg by a cathepsin-B-like protease (C), followed by processing with a dipeptidyl aminopeptidase (DPAP) and carboxypeptidase B (CP-B). (b) An α-factor secretion-vector configuration; thick line = *S. cerevisiae* gene promoter and terminator sequences; hatched box = the "S"-PRO-KR sequence from *MF*α; open box = coding sequence for mature hEGF. The mature hEGF is generated by cleavage of the pre-Pro-KR-hEGF precursor.

similar incomplete processing occurs if α-factor is overproduced, suggesting that the dipeptidyl-aminopeptidase is limiting (Julius *et al.*, 1983). It is important, therefore, to construct α-factor secretion vectors which retain only the lys-arg residues of the spacer; it is not clear whether all of the prosegment is necessary.

One problem with the α-factor system is that the cathepsin B-like protease can attack some sites within the mature protein next to arg residues, resulting in the production of subfragments (Bitter *et al.*, 1984). The degree of internal cleavage is low but could generate contaminating peptides with undesirable biological activities. The structural gene (*KEX2*) for the lys–arg cleavage endopeptidase has been isolated (Julius *et al.*, 1984) and by using a *kex2* host and a *KEX2* expressing system it may be possible to manipulate the levels of the peptidase or increase its specificity to minimize internal cleavages. The

maximum yields of proteins secreted by the α-factor system might be increased by replacing the α-factor promoter with a more efficient promoter. Although as much α-factor is produced on a molar basis as phosphoglycerate kinase (Thorner, 1982) there are two α-factor genes (Singh *et al.*, 1983) each producing four α-factor peptides and therefore the promoter may be at least eight times less efficient than the *PGK* promoter. The α-factor signals have also been used to direct the secretion of interleukin 2, insulin-like growth factor (ILGF) and growth hormone releasing factor (GHRF) (Brake *et al.*, 1984).

Several factors, in addition to promoter strength and the secretion signals which are used, may affect the levels of secretion of heterologous proteins. The most significant is the growth phase of the culture: the secretion of IFN-α directed by its own signal sequence (Hitzeman *et al.*, 1983b) and α-factor-directed secretion (Brake *et al.*, 1984) are both maximal from stationary-phase cultures. In the case of α-factor-directed secretion, maximum yields are obtained 24 h after the culture reaches stationary phase, when more than 99% of the hEGF is found in the culture medium; this represents 7% of the total protein produced by the culture. Secretion was much less efficient from actively growing cells, which retained at least 67% of the hEGF intracellularly (Brake *et al.*, 1984). In addition, there is some evidence that high levels of secretion are obtained if rich medium is used (Rothstein *et al.*, 1984) although the extent of this effect varies with different host strains (Wood *et al.*, 1985). It is also possible that very large proteins will not be secreted by *S. cerevisiae*: for example, the invertase signal sequence will transport *E. coli* β-galactosidase to the ER but it becomes trapped and none is found in the periplasmic space (Emr *et al.*, 1984). To date the *S. cerevisiae* secretion systems have been used extensively only for small (less than 25 kd) polypeptides which are effectively secreted into the culture medium. There is no reason why average-sized heterologous proteins should not also be efficiently secreted into the medium in some systems: for example, wheat α-amylase (44 kd) and *S. diastaticus* amylo-α-1,4-glucosidase (about 60 kd) are both secreted by *S. cerevisiae* (Tubb *et al.*, 1983; Rothstein *et al.*, 1984).

There are several general advantages in secreting heterologous proteins in addition to generating completely authentic proteins. Some proteins have been shown, or are suspected, to be unstable in *S. cerevisiae*, e.g. insulin (Stepien *et al.*, 1983) and rat growth hormone (Ammerer *et al.*, 1981). In these cases rapid secretion may reduce proteolysis. Although there is no direct evidence for instability, this might explain why, when hEGF is synthesized intracellularly under the direction of the efficient *GAP* promoter, the yields are only 30 μg/litre (Urdea *et al.*, 1983) but if it is secreted the yields are about 4 mg/litre (Brake *et al.*, 1984). Similarly, plant thaumatin can be detected only if the signal sequence is present to allow secretion (Edens *et al.*, 1984). Furthermore, there is some evidence that some proteins require secretion for the formation of their correct molecular configuration.

*S. cerevisiae* secretes very few of its own proteins into the culture medium and therefore purification of heterologous secreted proteins away from contaminating proteins may be simplified. The use of continuous secretion

processes may also result in higher yields per unit biomass than can be obtained by intracellular synthesis. An advantage of intracellular synthesis is, however, the ease of concentrating the heterologous product. It might be more desirable to produce high levels of heterologous proteins inside the cell using a regulated system; the production of authentic product could be ensured by engineering a specific cleavage site adjacent to the authentic *N*-terminus which could be used in a one-step purification and processing stage. For example, factor Xa cleaves at an ile–glu–gly–arg specific cleavage site: if this is engineered into a hybrid β-globin protein produced in *E. coli*, the factor Xa can be used to produce the authentic mature β-globin *in vitro* (Nagai and Thogersen, 1984).

### Subcellular localization of heterologous proteins

The presence or absence of a 'classical' signal sequence on a heterologous protein does not necessarily determine the subcellular localization. Calf chymosin, for example, is produced as a precursor protein which has a signal and pro-sequence. Preprochymosin, prochymosin and chymosin have all been expressed in *S. cerevisiae*: they are all found in the cell membrane and the cell wall and none is present in the cytoplasm (Mellor *et al.*, 1983). Similarly, hepatitis B surface antigen (HBsAg), which lacks a signal sequence, also appears to be membrane or cell-wall associated (Valenzuela *et al.*, 1982; Hitzeman *et al.*, 1983b). Cell-wall-associated proteins can be readily solubilized, for example by mechanical disruption (Mellor *et al.*, 1983). The process whereby proteins lacking signal sequences can accumulate in the cell wall is unknown but they may pass through the vacuole. This pathway may be used for enhancing the degradation of 'foreign' proteins. The finding that yields of prochymosin are significantly increased by using a strain with the *pep4–3* mutation which is deficient in vacuolar proteases (Zubenko, Park and Jones, 1982) supports this idea (Mellor *et al.*, 1983). Yields of a number of proteins (e.g. β-endorphin, Bitter *et al.*, 1984; $\alpha_1$-antitrypsin, Rosenberg *et al.*, 1984) are increased in a *pep4–3* background although it has little effect on other proteins, e.g. IFN (M.J. Dobson, M.F. Tuite, A.J. Kingsman and S.M. Kingsman, unpublished results). Fluorescent antibody staining has also shown that a proportion of both mature and precursor mouse immunoglobulin chains are found in the vacuole (Wood *et al.*, 1985).

### Glycosylation of heterologous proteins

There is increasing evidence that some degree of glycosylation of heterologous proteins occurs in *S. cerevisiae*. High-molecular-weight forms of several proteins, e.g. $\alpha_1$-antitrypsin (Cabezon *et al.*, 1984) and mouse immunoglobulins (Wood *et al.*, 1985), have been detected and these are not found in tunicamycin-treated cultures or after endoglycosidase H or chemical deglycosylation. However, the percentage glycosylation is low and heterogeneous and some proteins such as HBsAg, which are normally glycosylated, have no detectable carbohydrate in *S. cerevisiae* (Valenzuela *et al.*,

1982). *S. cerevisiae* glycoproteins are of the high-mannose type (Ballou, 1982) whereas mammalian glycoproteins contain a variety of glycosyl residues with complex branching (e.g. Staneloni and Leloir, 1982). It is therefore unlikely that glycosylation of heterologous proteins in *S. cerevisiae* will contribute to any biological activity which requires complex and specific carbohydrate modifications.

### Alternative vectors

In some fermentation processes it may be essential to ensure the absolute genetic homogeneity of the culture by stably integrating the heterologous gene into the chromosome. Yields would, however, be reduced if only a single copy was integrated and multiple copies in tandem would be unstable. A multicopy integrative vector which is dispersed throughout the genome is being developed using the transposable Ty element Ty1–15 (Kingsman *et al.*, 1981; W. Wilson, M.J. Dobson, J. Mellor, S.M. Kingsman and A.J. Kingsman, unpublished results; *Figure 11*). The element has been engineered to contain two selectable markers, *TRP1* (Tschumper and Carbon, 1980) and *LEU2* from pMA3a, and the *PGK* expression signals from pMA91 (Dobson *et al.*, 1982a; Mellor *et al.*, 1983) with an IFN-$\alpha_2$ coding sequence (Tuite *et al.*, 1982) at the expression site (*Figure 11*). A single copy of the engineered Ty is integrated into the genome using a linear fragment to stimulate recombination across the ends of the element and thereby replace an endogenous element. Transformants are selected for the *TRP1* marker. Few transfor-

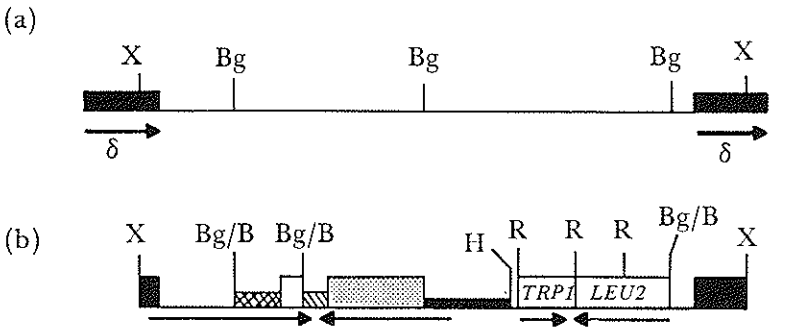


**Figure 11.** A multicopy integrative vector based on the *S. cerevisiae* Ty element. (a) The *S. cerevisiae* Ty element, Ty1–15 (Kingsman *et al.*, 1981). (b) Ty integrative vector. The *Hind* III–*Bam* HI fragment from pM91–1 (Mellor *et al.*, 1985) containing the *PGK* expression signals described in *Figure 6b* and the interferon cDNA described in Tuite *et al.* (1982) and *Figure 5*, is inserted at the promoter proximal *Bgl* II site of Ty1–15 (Bowen *et al.*, 1984). The two internal *Bgl* II fragments of Ty1–15 are removed and a selection cassette is inserted; this is the *TRP1*/*LEU2* cassette from pMA134 (M.J. Dobson, unpublished work). Closed box = Ty delta sequences; thin line = Ty epsilon sequences; thick line = *PGK* promoter; stippled box = IFN; hatched box = IFN transcription terminator; open box = *PGK* 3′ coding region; cross-hatched box = *PGK* 3′ untranslated and flanking sequences. The arrows indicate the direction and location of transcripts, the IFN transcription terminator is bidirectional (J. Mellor and A.M. Fulton, unpublished observations). Restriction sites as in previous figures: X = *Xho* I.

mants are obtained by selecting for *LEU2* as insufficient enzyme is produced by a single copy of this gene. The transformant is then grown in decreasing concentrations of leucine to select for an increase in the copy number of the *LEU2* gene, presumably by spread of the Ty element throughout the genome by gene conversion and transposition (Roeder and Fink, 1983). A stable strain has been constructed which produces $8 \times 10^5$ molecules of IFN per cell; this is intermediate between yields from single copy *ARS/CEN* vectors ($10^5$ molecules/cell) and from multicopy vectors such as pMA91 ($6 \times 10^6$ molecules/cell) (W. Wilson, M.J. Dobson and J. Mellor, unpublished results).

Another approach to vector construction is to create an artificial *S. cerevisiae* chromosome. Large linear fragments which carry *CEN, ARS* and telomeric (*TEL*) sequences are stably maintained and propagated as normal chromosomes (Szostak, 1983). The construction of artificial chromosomes may be useful for assembling novel metabolic pathways in *S. cerevisiae*.

### A summary of heterologous eukaryotic proteins made in *S. cerevisiae*.

A large number of different polypeptides have now been expressed in *S. cerevisiae* and we have attempted to collate the data that are available (*Table* 2). Proteins have been produced in different yields but they are all biologically active. There are some notable examples: HBsAg is produced as particles which are morphologically indistinguishable from serum-derived particles (Valenzuela *et al.*, 1982). These particles are antigenic and protect primates against hepatitis infection (McAleer *et al.*, 1984; Murray *et al.*, 1984). Calf prochymosin, which is a zymogen and normally activated by the acid gastric environment in the calf, is also produced as an inactive protein in *S. cerevisiae* and is readily activated by acidification, indicating that the protein is correctly folded in *S. cerevisiae* (Mellor *et al.*, 1983). *S. cerevisiae* strains containing plasmids expressing both the heavy and light chains of a mouse immunoglobulin secrete active antibody (Wood *et al.*, 1985). These data suggest that *S. cerevisiae* can be used to produce heterologous proteins with fully authentic biological activities. To date, with the exception of extraordinarily stable proteins which accumulate to very high levels, the maximum levels of heterologous proteins in *S. cerevisiae* are in the region of 5% of total cell protein.

## Future prospects

Many of the essential concepts for the expression of heterologous proteins in *S. cerevisiae* have been defined and vectors are available to direct high levels of intracellular synthesis, to regulate expression and to direct secretion. There is now intensive research into the structure and function of *S. cerevisiae* signals which are required for maximum expression. In particular the manipulation of UASs, DASs and negative elements may result in higher levels of transcription. Hybrid signals are being constructed, for example the efficiency of the α-factor secretion system can be improved by replacing the α-factor promoter with the *PGK* promoter. In addition, sequences required

**Table 2.** Summary of heterologous eukaryotic proteins made in *S. cerevisiae*

| Protein | Promoter | References* |
|---|---|---|
| Human-IFN-α | *ADH1* | Hitzeman *et al.*, 1981 |
| | *PGK* | Tuite *et al.*, 1982; Hitzeman *et al.*, 1983a |
| | *TRP1* | Dobson *et al.*, 1982b |
| | *PHO5* | Hinnen, Meyhack and Tsapis, 1983; Kramer *et al.*, 1984 |
| | *MF*α | Bitter *et al.*, 1984 |
| Human-IFN-γ | *PGK* | Derynck, Singh and Goeddel, 1983 |
| HBsAg | *ADH1* | Valenzuela *et al.*, 1982 |
| | *PGK* | Hitzeman *et al.*, 1983b |
| | *PHO5* | Miyanohara *et al.*, 1983 |
| Calf chymosin | *PGK* | Mellor *et al.*, 1983 |
| | *GAL1* | Goff *et al.*, 1984 |
| Insulin | *GAL1* | Stepien *et al.*, 1983 |
| Human-$\alpha_1$-antitrypsin | *PHO5* | Rosenberg *et al.*, 1984 |
| | *ARG3* | Cabezon *et al.*, 1984 |
| hEGF | *MF*α | Brake *et al.*, 1984 |
| β-endorphin | *MF*α | Bitter *et al.*, 1984 |
| Interleukin-2 | *MF*α | cited in Brake *et al.*, 1984 |
| Insulin-like GF | *MF*α | |
| GHRF | *MF*α | |
| *H. simplex* TK | *HIS3*† | McNeil and Friesen, 1981 |
| | *GAP* | Zhu, Ward and Weissbach, 1984 |
| Mouse immunoglobulin | *PGK* | Wood *et al.*, 1985 |
| Polyoma middle T | *PGK* | G. Belsham, personal communication |
| Influenza H3 | *PGK* | S.M. Kingsman *et al.*, unpublished results |
| t-PA | *PGK* | S.M. Kingsman *et al.*, unpublished results |
| Wheat α-amylase | *PGK* | Rothstein *et al.*, 1984 |
| Plant thaumatin | *GAP* | Edens *et al.*, 1984 |
| Maize zein | OWN‡ | Langridge *et al.*, 1984 |
| Bean phaseolin | OWN‡ | Cramer, Lee and Slightom, 1985 |

*Only one or two key references are given.
†The promoter was 3′ to the *HIS3* gene.
‡OWN refers to the situation in which a heterologous gene is expressed from its own promoter.

for regulation are being identified and these might be 'plugged in' to a variety of different promoters; systems which have a high induction ratio with a simple, convenient inducer will be favoured for the production of toxic proteins. An increased understanding of the importance of the various sequences discussed on pages 385–388 will allow promoter strengths to be manipulated to provide a precise level of expression. This will be important for the construction of entire regulated metabolic pathways or for the production of low levels of enzymes in multistage processes. Current analyses of the replication of the 2 μm plasmid will lead to plasmids with increased copy number and ultimately strains will be constructed in which heterologous gene expression and plasmid amplification are co-ordinately controlled. More studies are needed to establish systems for the secretion of high-molecular-weight proteins and these may involve using the signal sequence from large *Saccharomyces* proteins which are normally extracellular. Another secretion system which might be explored is the use of the killer factor signals (Bostian *et al.*, 1984; Skipper, Thomas and Lau, 1984). There will also be further

manipulations of the genetic background of *S. cerevisiae* to improve yields, plasmid stability, secretion and product extraction.

Novel proteins may be synthesized in *S. cerevisiae*: for example, an $\alpha_1$-antitrypsin cDNA has been modified by *in vitro* mutagenesis and expressed in *S. cerevisiae* to produce a more stable protein for the possible treatment of emphysema (Rosenberg *et al.*, 1984); a completely synthetic IFN-α comprising a consensus of all the IFN-α sequences has been produced in *S. cerevisiae* and may have a higher antiviral activity than the natural IFN-α proteins (Bitter *et al.*, 1984). The discovery that HBsAg assembles into antigenic particles has led to the idea of 'embedding' antigenic fragments from other viruses into the HBsAg coding region to allow efficient assembly into particles and the production of compound vaccines (Valenzuela *et al.*, 1985).

The first heterologous gene was expressed in *S. cerevisiae* in 1981 (Hitzeman *et al.*, 1981) and in 1985 at least one *S. cerevisiae*-heterologous product, HBsAg, is going into clinical trials. The future should see much of the fundamental research described in this review transferred to industry, and the large-scale production of many natural and novel proteins from *S. cerevisiae* is imminent.

## Acknowledgements

The authors would like to thank all their colleagues who communicated results before publication. The authors' own research cited here was supported by Celltech and by a Celltech/SERC co-operative grant (GR/C/94513) and a MRC project grant (G979/822/CB).

## References

ALBER, T. AND KAWASAKI, G. (1982). The nucleotide sequence of the triose phosphate isomerase gene of *Saccharomyces cerevisiae*. *Journal of Molecular and Applied Genetics* **1**, 419–434.

AMMERER, G., HITZEMAN, R., HAGIE, F., BARTA, A. AND HALL, B.D. (1981). The functional expression of mammalian genes in yeast. *Recombinant DNA, Proceedings of the Third Cleveland Symposium on Macromolecules* (A.G. Walton, Ed.), pp. 188–197. Elsevier, Holland.

ANDREADIS, A., HSU, Y.-P., KOHLHAN, G.B. AND SCHIMMEL, P. (1982). Nucleotide sequence of yeast *LEU2* shows 5'-noncoding region has sequences cognate to leucine. *Cell* **3**, 319–325.

BALLOU, C.E. (1982). Yeast cell wall and cell surface. In *Molecular Biology of the Yeast* Saccharomyces. *Volume 2: Metabolism and Gene Expression* (J. Strathern, E.W. Jones and J.R. Broach, Eds), pp. 335–360. Cold Spring Harbor Laboratory, New York.

BEGGS, J.D. (1978). Transformation of yeast by a replicating hybrid plasmid. *Nature* **275**, 104–109.

BEGGS, J.D., GUERINEAU, M. AND ATKINS, J.F. (1978). A map of the restriction targets in yeast 2 micron plasmid DNA cloned on bacteriophage lambda. *Molecular and General Genetics* **148**, 287–294.

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.

BENOIST, C., O'HARE, K., BREATHNACH, R. AND CHAMBON, P. (1980). The ovalbumin gene-sequence of putative control regions. *Nucleic Acids Research* **8**, 127–142.

BENNETZEN, J.L. AND HALL, B.D. (1982a). The primary structure of the *Saccharomyces cerevisiae* gene for alcohol dehydrogenase I. *Journal of Biological Chemistry* **257**, 3018–3026.

BENNETZEN, J.L. AND HALL, B.D. (1982b). Codon selection in yeast. *Journal of Biological Chemistry* **257**, 3026–3031.

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 fusions. *Proceedings of the National Academy of Sciences of the United States of America* **81**, 5330–5334.

BOLLEN, G.H.P.M., MOLENAAR, C.M.T., COHEN, L.H., VAN RAAMSDONK-DUIN, M.M.C., MAGER, W.H. AND PLANTA, R.J. (1982). Ribosomal protein genes of yeast contain intervening sequences. *Gene* **18**, 29–37.

BOSTIAN, K.A., LEMIRE, J.M., CANNON, L.E. AND HALVORSON, H.O. (1980). *In vitro* synthesis of repressible yeast acid phosphatase: identification of multiple mRNAs and products. *Proceedings of the National Academy of Sciences of the United States of America* **77**, 4504–4508.

BOSTIAN, K.A., ELLIOTT, Q., BUSSEY, H., BURN, V., SMITH, A. AND TIPPER, D.J. (1984). Sequence of the preprotein dsRNA gene of type I killer yeast: multiple processing events produce a two component toxin. *Cell* **36**, 741–751.

BORRELLI, E., HEN, R. AND CHAMBON, P. (1984). Adenovirus-2 EIA products repress enhancer induced stimulation of transcription. *Nature* **312**, 608–612.

BOWEN, B.A., FULTON, A.M., TUITE, M.F., KINGSMAN, S.M. AND KINGSMAN, A.J. (1984). Expression of Ty-*lacZ* fusions in *Saccharomyces cerevisiae*. *Nucleic Acids Research* **12**, 1627–1640.

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*. *Proceedings of the National Academy of Sciences of the United States of America* **81**, 4642–4646.

BRAZZELL, C. AND INGOLIA, T.D. (1984). Stimuli that induce a yeast heat shock gene fused to β-galactosidase. *Molecular and Cellular Biology* **4**, 2573–2584.

BROACH, J.R. (1982). The yeast plasmid 2μ circle. *Cell* **28**, 203–204.

BROACH, J.R. (1983). Construction of high copy number yeast vectors using 2 μm circle sequences. In *Methods in Enzymology* (R. Wu, L. Grossman and K. Moldave, Eds) volume 101, pp. 307–325. Academic Press, New York.

BRENT, R. AND PTASHNE, M. (1984). A bacterial repressor protein or a yeast transcriptional terminator can block upstream activation of a yeast gene. *Nature* **312**, 612–615.

BURKE, R.L., TEKAMP-OLSEN, P. AND NAJARIAN, R. (1983). The isolation, characterisation and sequence of the pyruvate kinase gene of *Saccharomyces cerevisiae*. *Journal of Biological Chemistry* **258**, 2193–2201.

CABEZON, T., DE WILDE, M., HERION, P., LORIAU, R. AND BOLLEN, A. (1984). Expression of human $\alpha_1$-antitrypsin cDNA in the yeast *Saccharomyces cerevisiae*. *Proceedings of the National Academy of Sciences of the United States of America* **81**, 6594–6598.

CAMPBELL, J.L. (1983). Yeast DNA replication. In *Genetic Engineering Principles and Methods* (J.K. Setlow and A. Hollaender, Eds), volume 5, pp. 109–146. Plenum, New York.

CHARNAY, P., TREISMAN, R., MELLON, P., CHAD, M., AXER, R. AND MANIATIS, T. (1984). Differences in human α and β globin gene expression in mouse erythroleukaemia cells: The role of intragenic sequences. *Cell* **38**, 251–263.

CLARKE, L. AND CARBON, J. (1980). Isolation of a yeast centromere and construction of functional small circular chromosomes. *Nature* **257**, 504–509.

CLARKE-WALKER, G.D. AND MIKLOS, G.L.G. (1974). Localisation and quantification of circular DNA in yeast. *European Journal of Biochemistry* **41**, 359–365.

COHEN, J.D., ECCLESHALL, T.R., NEEDLEMAN, R.B., FEDEROFF, H., BUCHFERER, B.A. AND MARMUR, J. (1980). Functional expression of the *Escherichia coli* plasmid

gene coding for chloramphenicol acetyltransferase. *Proceedings of the National Academy of Sciences of the United States of America* **77**, 1078–1082.

CRAMER, J.H., LEA, K. AND SLIGHTOM, J.R. (1985). Expression of phaseolin cDNA genes in yeast under control of natural plant DNA sequences. *Proceedings of the National Academy of Sciences of the United States of America* **82**, 334–338.

DERYNCK, R., SINGH, A. AND GOEDDEL, D.V. (1983). Expression of the human interferon-α cDNA in yeast. *Nucleic Acids Research* **11**, 1819–1837.

DOBSON, M.J., FUTCHER, A.B. AND COX, B.S. (1980). Loss of 2 μm DNA from *Saccharomyces cerevisiae* transformed with the chimaeric plasmid pJDB219. *Current Genetics* **2**, 201–205.

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 Research* **10**, 2625–2637.

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 the *TRP1* 5′ region. *Nucleic Acids Research* **11**, 2287–2302.

DONAHUE, T.F., DAVIES, R.S., LUCCHINI, G. AND FINK, G.R. (1983). A short nucleotide sequence required for the regulation of *HIS4* by the general control system of yeast. *Cell* **32**, 89–98.

EDENS, L., BOM, I., LEDEBOER, A.M., MAAT, J., TOONEN, M.Y., VISSER, C. AND VERRIPS, C.T. (1984). Synthesis and processing of the plant protein thaumatin in yeast. *Cell* **37**, 629–633.

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*. *Molecular and Cellular Biology* **4**, 2347–2355.

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*. *Proceedings of the National Academy of Sciences of the United States of America* **80**, 3671–3675.

ERHART, E. AND HOLLENBERG, C.P. (1983). The presence of a defective *LEU2* gene in 2 μm DNA recombinant plasmids of *Saccharomyces cerevisiae* is responsible for curing and high copy number. *Journal of Bacteriology* **156**, 625–635.

FERGUSSON, J., GROPPE, J.C. AND REED, S.I. (1981). Construction and characterisation of three yeast–*Escherichia coli* shuttle vectors designed for rapid subcloning of yeast genes on small DNA fragments. *Gene* **16**, 191–197.

FINKELSTEIN, D.B. AND STRAUSBERG, S. (1983). Heat shock regulated production of *Escherichia coli* β-galactosidase in *Saccharomyces cerevisiae*. *Molecular and Cellular Biology* **3**, 1625–1633.

GALLWITZ, D. AND SURES, I. (1980). Structure of a split yeast gene: complete nucleotide sequence of the actin gene in *Saccharomyces cerevisiae*. *Proceedings of the National Academy of Sciences of the United States of America* **77**, 2546–2550.

GAME, J.C. AND MORTIMER, R.K. (1974). A genetic study of X-ray sensitive mutants in yeast. *Mutation Research* **24**, 281–292.

GERBAUD, C. AND GUERINEAU, M. (1980). 2 μm plasmid copy number in different yeast strains and repartition of endogenous and 2 μ chimeric plasmids in transformed strains. *Current Genetics* **1**, 219–228.

GOFF, C.G., MOIR, D.T., KOHNO, T., GRAVIUS, T.C., SMITH, R.A., YAMASAKI, E. AND TAUNTON-RIGBY, A. (1984). The expression of calf prochymosin in *Saccharomyces cerevisiae*. *Gene* **27**, 35–46.

GRITZ, L. AND DAVIES, J. (1983). Plasmid encoded hygromycin B resistance: the sequence of hygromycin B phosphotransferase and its expression in *Escherichia coli* and *Saccharomyces cerevisiae*. *Gene* **25**, 179–188.

GROSVELD, G.C., DE BOER, E., SHEWMAKER, C.K. AND FLAVELL, R.A. (1982). DNA sequences necessary for transcription of the rabbit β-globin gene *in vivo*. *Nature* **295**, 120–126.

GUARENTE, L. (1984). Yeast promoters: positive and negative elements. *Cell* **36**, 799–800.

GUARENTE, L., YOCUM, R. AND GIFFORD, P. (1982). A *GAL10-CYC1* hybrid yeast promoter identifies the *GAL4* regulator as an upstream site. *Proceedings of the National Academy of Sciences of the United States of America* **79**, 7410–7414.

HAGUENAUER-TSAPIS, R. AND HINNEN, A. (1984). A deletion that includes the signal peptidase cleavage site impairs processing, glycosylation and secretion of cell surface yeast acid phosphatase. *Molecular and Cellular Biology* **4**, 2668–2675.

HARTLEY, J.L. AND DONELSON, J.E. (1980). Nucleotide sequence of the yeast plasmid. *Nature* **286**, 860–865.

HENIKOFF, S. AND COHEN, E.H. (1984). Sequences responsible for transcription termination on a gene segment in *Saccharomyces cerevisiae*. *Molecular and Cellular Biology* **4**, 1515–1520.

HENIKOFF, S. AND FURLONG, C. (1983). Sequence of a *Drosophila* DNA segment that functions in *Saccharomyces cerevisiae* and its regulation by a yeast promoter. *Nucleic Acids Research* **11**, 789–800.

HENIKOFF, S., KELLY, J.D. AND COHEN, E.H. (1983). Transcription terminates in yeast distal to a control sequence. *Cell* **33**, 607–614.

HERSKOWITZ, I. AND OSHIMA, Y. (1982). Control of cell type in *Saccharomyces cerevisiae*. Mating type and mating type interconversion. In *The Molecular Biology of the Yeast* Saccharomyces. *Volume 1: Life Cycle and Inheritance* (J.N. Strathern, E.W. Jones and J.R. Broach, Eds), pp. 181–210. Cold Spring Harbor Laboratory, New York.

HINNEN, A., HICKS, J.B. AND FINK, G.R. (1978). Transformation of yeast. *Proceedings of the National Academy of Sciences of the United States of America* **75**, 1929–1933.

HINNEN, A., MEYHACK, B. AND TSAPIS, R. (1983). High expression and secretion of foreign proteins in yeast. In *Gene Expression in Yeast, Foundation for Biotechnical and Industrial Fermentation Research* (M. Korhola and E. Vaisanen, Eds), volume 1, pp. 157–166. Kauppakirjapaino Oy, Helsinki.

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.

HITZEMAN, R.A., CHEN, C.Y., HAGIE, F.E., PATZER, E.G., LIU, C-C., ESTELL, D.A., MILLER, J.V., YAFFE, A., KLEID, D.G., LEVINSON, A.D. AND OPPERMAN, H. (1983a). Expression of Hepatitis B virus surface antigens in yeast. *Nucleic Acids Research* **11**, 2745–2763.

HITZEMAN, R.A., LEUNG, D.W., PERRY, L.J., KOHR, W.J., LEVINE, H.L. AND GOEDDEL, D.V. (1983b). Secretion of human interferons by yeast. *Science* **219**, 620–625.

HOLLAND, J.P. AND HOLLAND, M.J. (1980). Structural comparison of two nontandemly repeated yeast glyceraldehyde-3-phosphate dehydrogenase genes. *Journal of Biological Chemistry* **255**, 2596–2605.

HOLLAND, M.J., HOLLAND, J.P., THILL, G.P. AND JACKSON, K.A. (1981). The primary structures of two yeast enolase genes. *Journal of Biological Chemistry* **256**, 1385–1395.

HYMAN, B.C., CRAMER, J.H. AND ROWND, R.H. (1982). Properties of a *Saccharomyces cerevisiae* mtDNA segment conferring high frequency yeast transformation. *Proceedings of the National Academy of Sciences of the United States of America* **79**, 1578–1582.

IKEMURA, T. (1982). Correlation between the abundance of yeast transfer RNAs and the occurrence of the respective codons in protein genes. *Journal of Molecular Biology* **158**, 573–597.

ITEKURA, K., HIROSE, T., CREA, R., RIGGS, A.D., HEYNEKER, H.L., BOLIVER, F. AND BOYER, H.W. (1977). Expression in *Escherichia coli* of a chemically synthesised gene for the hormone somatostatin. *Science* **198**, 1056–1063.

ITO, H., FUKUDA, Y., MURATA, K. AND KIMURA, A. (1983). Transformation of intact yeast cells treated with alkali cations. *Journal of Bacteriology* **153**, 163–168.

JAYARAM, M., LI, Y.-Y. AND BROACH, J.R. (1983). The yeast plasmid 2 μ circle encodes components required for its high copy number propagation. *Cell* **34**, 95–104.

JIMENEZ, A. AND DAVIES, J. (1980). Expression of a transposable antibiotic resistance element in *Saccharomyces*. *Nature* **287**, 869–871.

JOHNSTON, S. AND HOPPER, J.E. (1982). Isolation of the yeast regulatory gene *GAL4* and analysis of its dosage effects on the galactose-melibiose region. *Proceedings of the National Academy of Sciences of the United States of America* **79**, 6971–6975.

JULIUS, D., SCHEKMAN, R. AND THORNER, J. (1984). Glycosylation and processing of prepro-α-factor through the yeast secretory pathway. *Cell* **36**, 309–318.

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. (1984). Isolation of the putative structural gene for the lys–arg-cleavage endopeptidase required for processing of yeast prepro-α-factor. *Cell* **37**, 1075–1089.

KARIN, M., NAJARIAN, R., HASLINGER, A., VALENZUELA, P., WELCH, J. AND FOGEL, S. (1984). Primary structure and transcription of an amplified genetic locus. The *CUP1* locus of yeast. *Proceedings of the National Academy of Sciences of the United States of America* **81**, 337–342.

KINGSMAN, S.M. AND KINGSMAN, A.J. (1983). The production of interferon in bacteria and yeast. In *Interferons, Society for General Microbiology Symposium* (D.C. Burke and A. Morris, Eds), volume 35, pp. 211–254. Cambridge University Press.

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.

KINGSMAN, A.J., GIMLICH, R.L., CLARKE, L., CHINAULT, A.C. AND CARBON, J. (1981). Sequence variation in dispersed repetitive sequence in *Saccharomyces cerevisiae*. *Journal of Molecular Biology* **145**, 619–632.

KINGSMAN, S.M., DOBSON, M.J., TUITE, M.F., MELLOR, J., ROBERTS, N.A. AND KINGSMAN, A.J. (1983). High efficiency expression vectors. In *Gene Expression in Yeast, Foundation for Biotechnical and Industrial Fermentation Research* (M. Korhola and E. Vaisanen, Eds), volume 1, pp. 95–114. Kauppakirjaipaino Oy, Helsinki

KISS, G.B., PEARLMAN, R.E., CORNISH, K.V., FRIESEN, J.D. AND CHAN, V.L. (1982). The *Herpes simplex* virus thymidine kinase is not transcribed in *Saccharomyces cerevisiae*. *Journal of Bacteriology* **149**, 542–547.

KOZAK, M. (1984). Compilation and analysis of sequences upstream from the translational start in eukaryotic mRNAs. *Nucleic Acids Research* **12**, 857–879.

KRAMER, R.A. AND ANDERSEN, N. (1980). Isolation of yeast genes with mRNA levels controlled by phosphate concentration. *Proceedings of the National Academy of Sciences of the United States of America* **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. *Proceedings of the National Academy of Sciences of the United States of America* **81**, 367–370.

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.

LANGFORD, C., NELLEN, N., NIESSING, J. AND GALLWITZ, D. (1983). Yeast is unable to excise foreign intervening sequences from hybrid gene transcripts. *Cell* **33**, 519–524.

LANGRIDGE, R., EIBEL, H., BROWN, J.W.S. AND FEIX, G. (1984). Transcription from maize storage protein gene promoters in yeast. *EMBO Journal* **3**, 2467–2471.

MCALEER, W.J., BUYNAK, E.B., MAIGETTER, R.Z., WAMPLER, D.E., MILLER, W.J. AND HILLEMAN, M.R. (1984). Human hepatitis B vaccine from recombinant yeast. *Nature* **307**, 178–180.

MCNEIL, J.B. AND FRIESEN, J. (1981). Expression of the *Herpes simplex* virus thymidine kinase gene in *Saccharomyces cerevisiae*. *Molecular and General Genetics* **184**, 386–393.

MARTINEZ-ARIAS, A., YOST, H.J. AND CASADABAN, M.J. (1984). Role of an upstream regulatory element in leucine repression of the *Saccharomyces cerevisiae LEU2* gene. *Nature* **307**, 740–742.

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.

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.

MEYHACK, B., BAJWA, W., RUDOLPH, H. AND HINNEN, A. (1982). Two yeast acid phosphatase structural genes are the results of a tandem duplication and show different degrees of homology in their promoter and coding sequence. *EMBO Journal* **1**, 675–680.

MIYAJIMA, A., MIYAJIMA, I., ARAI, K.-I. AND ARAI, N. (1984). Expression of plasmid R388 encoded type II dihydrofolate reductase as a dominant selective marker in *Saccharomyces cerevisiae*. *Molecular and Cellular Biology* **4**, 407–414.

MIYANOHARA, A., TOH-E, A., NOZAKI, C., HAMADA, F., OHTOMO, N. AND MATSUBARA, K. (1983). Expression of hepatitis B surface antigen gene in yeast. *Proceedings of the National Academy of Sciences of the United States of America* **80**, 1–5.

MURRAY, A.W. AND SZOSTAK, J.W. (1983). Pedigree analysis of plasmid segregation in yeast. *Cell* **34**, 911–970.

MURRAY, K., BRUCE, S.A., HINNEN, A., WINGFIELD, P., VAN ERD, P.M.C.A., DE REUS, A. AND SCHELLEKENS, H. (1984). Hepatitis B virus antigens made in microbial cells immunise against viral infection. *EMBO Journal* **3**, 645–651.

NAGAI, K. AND THOGERSEN, H.C. (1984). Generation of β-globin by sequence specific proteolysis of a hybrid protein produced in *Escherichia coli*. *Nature* **309**, 810–812.

ORR-WEAVER, T.L., SZOSTAK, J.W. AND ROTHSTEIN, R.J. (1983). Genetic applications of yeast transformations with linear and gapped plasmids. In *Methods in Enzymology* (R. Wu, L. Grossman and K. Moldave, Eds) volume 101, pp. 228–245. Academic Press, New York.

OSBORNE, T.F., ARVIDSON, D.N., TYAU, E.S., DUNSWORTH-BROWNE, M. AND BERK, A.J. (1984). Transcription control region within the protein-coding portion of Adenovirus E1A genes. *Molecular and Cellular Biology* **4**, 1293–1305.

PERLMAN, D. AND HALVORSEN, H.O. (1983). A putative signal peptidase recognition site in eukaryotic and prokaryotic signal peptides. *Journal of Molecular Biology* **167**, 391–409.

REZNIKOFF, W.S. AND ABELSON, J.N. (1980). The *lac* promoter. In *The Operon* (J.H. Miller and W.S. Reznikoff, Eds), pp. 31–88. Cold Spring Harbor Laboratory, New York.

ROEDER, G.C. AND FINK, G.R. (1983). Transposable elements in yeast. In *Mobile Genetic Elements* (J.A. Shapiro, Ed.), pp. 300–328. Academic Press, New York.

ROGGENKAMP, R., DARGATZ, H. AND HOLLENBERG, C.P. (1985). Precursor of β-lactamase is enzymatically inactive: Accumulation of the preprotein in *Sacchar-*

*omyces cerevisiae*. *Journal of Biological Chemistry*, **260**, 1508–1512.

ROGGENKAMP, R., KUSTERMANN-KUHN, B. AND HOLLENBERG, C.P. (1981). Expression and processing of bacterial β-lactamase in the yeast *Saccharomyces cerevisiae*. *Proceedings of the National Academy of Sciences of the United States of America* **78**, 4466–4470.

ROSENBERG, S., BARR, P.J., NAJARIAN, R.C. AND HALLEWELL, R.A. (1984). Synthesis in yeast of a functional oxidation resistant mutant of human $\alpha_1$-antitrypsin. *Nature* **312**, 77–80.

ROTHSTEIN, R.J. (1983). One-step gene disruption in yeast. In *Methods in Enzymology*, (R. Wu, L. Grossman and K. Moldave, Eds.) volume 101, pp. 202–211. Academic Press, New York.

ROTHSTEIN, S.J., LAZARUS, C.M., SMITH, W.E., BAULCOMBE, D.C. AND GATENBY, A.A. (1984). Secretion of a wheat α-amylase expressed in yeast. *Nature* **308**, 662–665.

RUBY, S.W., SZOSTAK, J.W. AND MURRAY, A.W. (1983). Cloning regulated yeast genes from a pool of *lacZ* fusions. In *Methods in Enzymology* (R. Wu, L. Grossman and K. Moldave, Eds), volume 101, pp. 253–269. Academic Press, New York.

RUSSELL, P.R. (1983). Evolutionary divergence of the mRNA transcription initiation mechanism in yeast. *Nature* **301**, 167–169.

ST JOHN, T. AND DAVIS, R.W. (1981). The organisation and transcription of the *GAL* operon. *Journal of Molecular Biology* **152**, 285–315.

SAROKIN, L. AND CARLSON, M. (1984). Upstream region required for regulated expression of the glucose-repressible *SUC2* gene of *Saccharomyces cerevisiae*. *Molecular and Cellular Biology* **4**, 2750–2757.

SCHEKMAN, R. AND NOVICK, P. (1982). The secretory process and yeast cell surface assembly. In *The Molecular Biology of the Yeast* Saccharomyces. *Volume 2: Metabolism and Gene Expression* (J. Strathern, E.W. Jones and J.R. Broach, Eds), pp. 361–398. Cold Spring Harbor Laboratory, New York.

SHERMAN, F. AND STEWART, J.W. (1982). Mutations altering initiation of translation of yeast iso-l-cytochrome C: contrast between the eukaryotic and prokaryotic initiation process. In *The Molecular Biology of the Yeast* Saccharomyces. *Volume 2: Metabolism and Gene Expression* (J. Strathern, E.W. Jones and J.R. Broach, Eds), pp. 301–355. Cold Spring Harbor Laboratory, New York.

SHINE, J. AND DALGARNO, L. (1975). Determinant of cistron specificity in bacterial ribosomes. *Nature* **254**, 34–38.

SICILIANO, P.G. AND TATCHELL, K. (1984). Transcription and regulatory signals at the mating type locus in yeast. *Cell* **37**, 969–978.

SIMONS, G., REMAUT, E., ALLET, B., DEVOS, R. AND FIERS, W. (1984). High level expression of human interferon gamma in *Escherichia coli* under control of the $P_L$ promoter of bacteriophage lambda. *Gene* **28**, 55–64.

SINGH, A., CHEN, E.-Y., LUGOVY, J.M., CHANG, C.N., HITZEMAN, R.A. AND SEEBURG, P. (1983). *Saccharomyces cerevisiae* contains two discrete genes coding for the α-factor pheromone. *Nucleic Acids Research* **11**, 4049–4063.

SKIPPER, N., THOMAS, D.Y. AND LAU, P.C.K. (1984). Cloning and sequencing of the preprotein-coding region of the yeast M1 double-stranded RNA. *EMBO Journal* **3**, 107–111.

STANELONI, R.J. AND LELOIR, L.F. (1982). The biosynthetic pathway of the asparagine-linked oligosaccharides of glycoproteins. *Critical Reviews in Biochemistry* **12**, 298–326.

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.

STRUHL, K. (1981). Deletion mapping a eukaryotic promoter. *Proceedings of the National Academy of Sciences of the United States of America* **78**, 4451–4465.

STRUHL, K. (1983). The new yeast genetics. *Nature* **305**, 391–396.

STRUHL, K., STINCHCOMB, D.T., SCHERER, S. AND DAVIS, R.W. (1979). High frequency transformation of yeast: autonomous replication of hybrid molecules. *Proceedings of the National Academy of Sciences of the United States of America* **76**, 1035–1039.

SZOSTAK, J.W. (1983). A rapid procedure for the construction of linear yeast plasmids. In *Methods in Enzymology* (R. Wu, F. Grossman and K. Moldave, Eds) volume 101, pp. 245–252. Academic Press, New York.

THORNER, J. (1982). Pheromonal regulation of development in *Saccharomyces cerevisiae*. In *The Molecular Biology of the Yeast* Saccharomyces. *Volume 1: Life Cycle and Inheritance* (J.N. Strathern, E.W. Jones and J.R. Broach, Eds), pp. 143–180. Cold Spring Harbor Laboratory, New York.

TSCHUMPER, G. AND CARBON, J. (1980). Sequence of a yeast DNA fragment containing a chromosomal replicator and the *TRP1* gene. *Gene* **10**, 157–166.

TUBB, R.S., SEARL, B.A., OGDEN, K., MEADEN, P.G.T. AND LUKER, M.A. (1983). Constructing amylolytic strains of yeast for commercial applications. In *Gene Expression in Yeast, Foundation for Biochemical and Industrial Fermentation Research* (M. Korhola and E. Vaisanen, Eds) volume 1, pp. 229–231. Kauppakirjapaino Oy, Helsinki.

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 expression of human interferon-alpha in *Saccharomyces cerevisiae*. *EMBO Journal* **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. *Proceedings of the National Academy of Sciences of the United States of America* **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.

VALENZUELA, P., COIT, D., MEDINA-SELBY, M.A., KUO, C.H., VAN NEST, G., BURKE R.L., BULL, P., URDEA, M.S. AND GRAVES, P.V. (1985). Antigen engineering in yeast: Synthesis and assembly of hybrid hepatitis B surface antigen–Herpes simplex 1 gD particles. *Biotechnology* **3**, 323–329.

WALMSLEY, R.M., GARDNER, D.C. AND OLIVER, S.G. (1983). Stability of a cloned gene in yeast grown in chemostat culture. *Molecular and General Genetics* **194**, 361–365.

WEBSTER, T.D. AND DICKSON, R.C. (1983). Direct selection of *Saccharomyces cerevisiae* resistant to the antibiotic G418 following transformation with a DNA vector carrying the kanamycin resistance gene of Tn903. *Gene* **26**, 243–252.

WOOD, C.R., BOSS, M.A., KENTON, J.M., CALVERT, J.E., ROBERTS, N.A. AND EMTAGE, J.S. (1985). The synthesis and *in vivo* assembly of functional antibodies in yeast. *Nature*, **314**, 446–449.

WRIGHT, C.F. AND ZITOMER, R.S. (1984). A positive regulatory site and a negative regulatory site control the expression of the *Saccharomyces cerevisiae CYC7* gene. *Molecular and Cellular Biology* **4**, 2023–2030.

WRIGHT, S., ROSENTHAL, A., FLAVELL, R. AND GROSVELD, F. (1984). DNA sequences required for regulated expression of β-globin genes in murine erythroleukaemia cells. *Cell* **38**, 265–273.

ZALKIN, H. AND YANOFSKY, C.L. (1982). Yeast gene *TRP5:* Structure, function and regulation. *Journal of Biological Chemistry* **257**, 1491–1500.

ZARET, K. AND SHERMAN, F. (1982). DNA sequence required for efficient transcription termination in yeast. *Cell* **28**, 563–573.

ZARET, K. AND SHERMAN, F. (1984). Mutationally altered 3′ ends of yeast *CYC1* mRNA affect transcript stability and translational efficiency. *Journal of Molecular Biology* **177**, 107–136.

ZHU, X.L., WARD, C. AND WEISSBACH, A. (1984). Control of *Herpes simplex* virus thymidine kinase gene expression in *Saccharomyces cerevisiae* by a yeast promoter sequence. *Molecular and General Genetics* **194**, 31–41.

ZUBENKO, G.S., PARK, F.J. AND JONES, E.W. (1982). Genetic properties of mutations at the *PEP-4* locus in *Saccharomyces cerevisiae*. *Genetics* **102**, 679–690.