



## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

<b>(51) International Patent Classification 5 :</b> C12N 15/29, C07K 13/00 C12N 1/21, 1/19		A1	<b>(11) International Publication Number:</b> WO 91/19801 <b>(43) International Publication Date:</b> 26 December 1991 (26.12.91)
<b>(21) International Application Number:</b>	PCT/GB91/00914		<b>(74) Agents:</b> SHEARD, Andrew, Gregory et al.; Kilburn & Strode, 30 John Street, London WC1N 2DD (GB).
<b>(22) International Filing Date:</b>	7 June 1991 (07.06.91)		
<b>(30) Priority data:</b> 9013016.2	11 June 1990 (11.06.90)	GB	<b>(81) Designated States:</b> AT (European patent), AU, BE (European patent), BR, CA, CH (European patent), DE (European patent), DK (European patent), ES (European patent), FI, FR (European patent), GB, GB (European patent), GR (European patent), HU, IT (European patent), JP, KR, LU (European patent), NL (European patent), NO, PL, RO, SE (European patent), SU, US.
<p><b>(71) Applicant (for all designated States except US):</b> MARS UK LIMITED [GB/GB]; 3D Dundee Road, Slough, Berkshire SL1 4JS (GB).</p> <p><b>(72) Inventors; and</b></p> <p><b>(75) Inventors/Applicants (for US only) :</b> SPENCER, Margaret, Elizabeth [GB/GB]; 24 Totley Brook Road, Sheffield S17 3QS (GB). HODGE, Rachel [GB/GB]; 16 Mount Avenue, Leicester LE5 3RN (GB). DEAKIN, Edward, Alfred [GB/GB]; The Oakes, Oakes Park, Norton, Sheffield S8 8BA (GB). ASHTON, Sean [GB/GB]; 5 Chamberlain Court, Burncross, Chapeltown, Sheffield S30 4ZU (GB).</p>			<p><b>Published</b>  <i>With international search report.</i>  <i>Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.</i></p>
<b>(54) Title:</b> RECOMBINANT 47 AND 31kD COCOA PROTEINS AND PRECURSOR			
<p><b>(57) Abstract</b></p> <p>47 kD and 31 kD proteins, and their 67 kD expression precursor, believed to be the source of peptide flavour precursors in cocoa (<i>Theobroma cacao</i>) have been identified. Genes coding for them have been probed, identified and sequenced, and recombinant proteins have been synthesised.</p>			

***FOR THE PURPOSES OF INFORMATION ONLY***

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	ES	Spain	MG	Madagascar
AU	Australia	FI	Finland	ML	Mali
BB	Barbados	FR	France	MN	Mongolia
BE	Belgium	GA	Gabon	MR	Mauritania
BF	Burkina Faso	GB	United Kingdom	MW	Malawi
BG	Bulgaria	GN	Guinea	NL	Netherlands
BJ	Benin	GR	Greece	NO	Norway
BR	Brazil	HU	Hungary	PL	Poland
CA	Canada	IT	Italy	RO	Romania
CF	Central African Republic	JP	Japan	SD	Sudan
CG	Congo	KP	Democratic People's Republic of Korea	SE	Sweden
CH	Switzerland	KR	Republic of Korea	SN	Senegal
CI	Côte d'Ivoire	LI	Liechtenstein	SU	Soviet Union
CM	Cameroon	LK	Sri Lanka	TD	Chad
CS	Czechoslovakia	LU	Luxembourg	TG	Togo
DE	Germany	MC	Monaco	US	United States of America
DK	Denmark				

1        RECOMBINANT 47 AND 31kD COCOA PROTEINS AND PRECURSOR

2

3        This invention relates to proteins and nucleic acids derived from or otherwise  
4        related to cocoa.

5

6        The beans of the cocoa plant (*Theobroma cacao*) are the raw material for cocoa,  
7        chocolate and natural cocoa and chocolate flavouring. As described by Rohan  
8        ("Processing of Raw Cocoa for the Market", FAO/UN (1963)), raw cocoa  
9        beans are extracted from the harvested cocoa pod, from which the placenta is  
10      normally removed, the beans are then "fermented" for a period of days, during  
11      which the beans are killed and a purple pigment is released from the cotyledons.  
12      During fermentation "unknown" compounds are formed which on roasting give  
13      rise to characteristic cocoa flavour. Rohan suggests that polyphenols and  
14      theobromine are implicated in the flavour precursor formation. After  
15      fermentation, the beans are dried, during which time the characteristic brown  
16      pigment forms, and they are then stored and shipped.

17

18      Biehl *et al*, 1982 investigated proteolysis during anaerobic cocoa seed  
19      incubation and identified 26kD and 44kD proteins which accumulated during  
20      seed ripening and degraded during germination. Biehl asserted that there were  
21      storage proteins and suggested that they may give rise to flavour-specific  
22      peptides.

23

24      Fritz *et al*, 1985 identified polypeptides of 20kD and 28kD appearing in the  
25      cytoplasmic fraction of cocoa seed extracts at about 100 days after pollination.  
26      It appears that the 20kD protein is thought to have glyceryl acyltransferase  
27      activity.

28

29      In spite of the uncertainties in the art, as summarised above, proteins apparently  
30      responsible for flavour production in cocoa beans have now been identified.  
31      Further, it has been discovered that, in spite of Fritz's caution that "cocoa seed

32

33

1 mRNA levels are notably low compared to other plants" (*loc. cit.*), it is possible  
2 to apply the techniques of recombinant DNA techniques to the production of  
3 such proteins.

4

5 According to a first aspect of the invention, there is provided a 67kD protein of  
6 *Theobroma cacao*, or a fragment thereof.

7

8 The 67kD protein appears to be a primary translation product of interest in  
9 proteins involved in flavour production in cocoa. The 67kD protein may be  
10 processed *in vivo* to form 47kD and 31kD polypeptides.

11

12 According to a second aspect of the invention, there is provided a 47kD protein  
13 of *Th. cacao*, or a fragment thereof.

14

15 According to a third aspect of the invention, there is provided a 31kD protein of  
16 *Th. cacao* or a fragment thereof.

17

18 The term "fragment" as used herein and as applied to proteins or peptides  
19 indicates a sufficient number of amino acid residues are present for the fragment  
20 to be useful. Typically, at least four, five, six or even at least 10 or 20 amino  
21 acids may be present in a fragment. Useful fragments include those which are  
22 the same as or similar or equivalent to those naturally produced during the  
23 fermentation phase of cocoa bean processing. It is believed that such fragments  
24 take part in Maillard reactions during roasting, to form at least some of the  
25 essential flavour components of cocoa.

26

27 Proteins in accordance with the invention may be synthetic; they may be  
28 chemically synthesised or, preferably, produced by recombinant DNA  
29 techniques. Proteins produced by such techniques can therefore be termed  
30 "recombinant proteins". Recombinant proteins may be glycosylated or  
31 non-glycosylated; non-glycosylated proteins will result from prokaryotic  
32 expression systems.

33

1      *Theobroma cacao* has two primary subspecies, *Th. cacao cacao* and *Th. cacao*  
2      *sphaerocarpum*. While proteins in accordance with the invention may be  
3      derived from these subspecies, the invention is not limited solely to these  
4      subspecies. For example, many cocoa varieties are hybrids between different  
5      species; an example of such a hybrid is the trinitario variety.

6

7      The invention also relates to nucleic acid, particularly DNA, coding for the  
8      proteins referred to above (whether the primary translation products, the  
9      processed proteins or fragments). The invention therefore also provides, in  
10     further aspects:

11

12        -      nucleic acid coding for a 67kD protein of *Th. cacao*, or for a  
13      fragment thereof;

14

15        -      nucleic acid coding for a 47kD protein of *Th. cacao*, or for a  
16      fragment thereof;

17

18        -      nucleic acid coding for a 31kD protein of *Th. cacao*, or for a  
19      fragment thereof;

20

21      Included in the invention is nucleic acid which is degenerate for the wild type  
22      protein and which codes for conservative or other non-deleterious mutants.  
23      Nucleic acid which hybridises to the wild type material is also included.

24

25      Nucleic acid within the scope of the invention will generally be recombinant  
26      nucleic acid and may be in isolated form. Frequently, nucleic acid in  
27      accordance with the invention will be incorporated into a vector (whether an  
28      expression vector or otherwise) such as a plasmid. Suitable expression vectors  
29      will contain an appropriate promoter, depending on the intended expression  
30      host. For yeast, an appropriate promoter is the yeast pyruvate kinase (PK)  
31      promoter; for bacteria an appropriate promoter is a strong lambda promoter.

32

33

1 Expression may be secreted or non-secreted. Secreted expression is preferred,  
2 particularly in eukaryotic expression systems; an appropriate signal sequence  
3 may be present for this purpose. Signal sequences derived from the expression  
4 host (such as that from the yeast alpha-factor in the case of yeast) may be more  
5 appropriate than native cocoa signal sequences.

6

7 The invention further relates to host cells comprising nucleic acid as described  
8 above. Genetic manipulation may for preference take place in prokaryotes.  
9 Expression will for preference take place in a food-approved host. The yeast  
10 *Saccharomyces cerevisiae* is particularly preferred.

11

12 The invention also relates to processes for preparing nucleic acid and protein as  
13 described above by nucleic acid replication and expression, respectively.

14

15 cDNA in accordance with the invention may be useful not only for obtaining  
16 protein expression but also for Restriction Fragment Length Polymorphism  
17 (RFLP) studies. In such studies, detectably labelled cDNA (eg radiolabelled) is  
18 prepared. DNA of a cultivar under analysis is then prepared and digested with  
19 restriction enzymes. Southern blotting with the labelled cDNA may then enable  
20 genetic correlations to be made between cultivars. Phenotypic correlations may  
21 then be deduced.

22

23 The invention will now be illustrated by the following non-limiting examples.  
24 The examples refer to the accompanying drawings, in which:  
25 Figure 1 shows a map of the coding region of the 67kD protein, together with  
26 the inter-relationship of plasmids pMS600, pMS700 and pMS800, from which  
27 sequence data were obtained;

28

29 Figure 2 shows the complete nucleotide sequence of cDNA coding for the 67kD  
30 protein and the deduced amino acid sequence;

31

32 Figure 3 shows the amino acid sequence referred to in Figure 2;

33

1      Figure 4 shows the relationship between the 67kD protein and seed storage  
2      proteins from other plants;  
3  
4      Figure 5 shows a map of plasmid pJLA502;  
5  
6      Figure 6 shows schematically the formation of plasmid pMS900;  
7  
8      Figure 7 shows two yeast expression vectors useful in the present invention;  
9      vector A is designed for internal expression and vector B is designed for  
10     secreted expression;  
11  
12     Figure 8a shows, in relation to vector A, part of the yeast pyruvate kinase gene  
13     showing the vector A cloning site, and the use of *Hin-Nco* linkers to splice in  
14     the heterologous gene;  
15  
16     Figure 8b shows, in relation to vector B, part of the yeast alpha-factor signal  
17     sequence showing the vector B cloning site, and the use of *Hin-Nco* linkers to  
18     create an in-phase fusion;  
19  
20     Figure 9a shows how plasmid pMS900 can be manipulated to produce plasmids  
21     pMS901, pMS903, pMS907, pMS908, pMS911, pMS912 and pMS914;  
22  
23     Figure 9b shows how plasmid pMS903 can be manipulated to produce plasmids  
24     pMS904, pMS905, pMS906, pMS909 and pMS916;  
25  
26     Figure 10 shows maps of plasmids pMS908, pMS914, pMS912, pMS906,  
27     pMS916 and pMS910;  
28  
29     Figure 11 shows the construction of a plasmid to express the 67kD protein by  
30     means of the AOX promoter on an integrated vector in *Hansenula polymorpha*;  
31     and  
32  
33

1      Figure 12 shows the construction of a plasmid to express the 67kD protein by  
2      means of the AOX promoter in conjunction with the yeast  $\alpha$ -factor secretory  
3      signal on an integrated vector in *Hansenula polymorpha*.

4

5      EXAMPLES

6

7      Example 1

8

9      *Identification of the Major Seed Proteins*

10

11     It is not practicable to extract proteins directly from cocoa beans due to the high  
12    fat and polyphenol contents, and proteins were, therefore, extracted from  
13    acetone powders made as follows. Mature beans from cocoa of West African  
14    origin (*Theobroma cacao amelonada*) were lyophilised and ground roughly in a  
15    pestle and mortar. Lipids were extracted by Soxhlet extraction with diethyl  
16    ether for two periods of four hours, the beans being dried and further ground  
17    between extractions. Polyphenols and pigments were then removed by several  
18    extractions with 80% acetone, 0.1% thioglycollic acid. After extraction the  
19    resulting paste was dried under vacuum and ground to a fine powder.

20

21     Total proteins were solubilised by grinding the powder with extraction buffer  
22    (0.05 M sodium phosphate, pH 7.2; 0.01 M 2-mercaptoethanol; 1% SDS) in a  
23    hand-held homogeniser, at 5mg/ml. The suspension was heated at 95°C for 5  
24    minutes, and centrifuged at 18 K for 20 minutes to remove insoluble material.  
25    The resulting clear supernatant contained about 1 mg/ml total protein.  
26    Electrophoresis of 25  $\mu$ l on an SDS-PAGE gel (Laemmli, 1970) gave three  
27    major bands, two of which were at 47 kD and 31 kD, comprising over 60% of  
28    the total proteins. The 47kD and 31kD proteins are presumed to be the  
29    polypeptide subunits of major storage proteins.

30

31

32

33

1     *Characteristics of the Storage Polypeptides*

2

3     The solubility characteristics of the 47 kD and 31 kD polypeptides were roughly  
4     defined by one or two quick experiments. Dialysis of the polypeptide solution  
5     against SDS-free extraction buffer rendered the 47 kD and 31 kD polypeptides  
6     insoluble, as judged by their ability to pass through a 0.22 micron membrane.  
7     Fast Protein Liquid Chromatography (FPLC) analysis also showed that the 47  
8     kD and 31 kD polypeptides were highly associated after extraction with  
9     McIlvaines buffer pH 6.8 (0.2 M disodium hydrogen phosphate titrated with  
10    0.1 M citric acid). The 47 kD and 31 kD polypeptides are globulins on the  
11    basis on their solubility.

12

13    *Purification of the 47 kD and 31 kD polypeptides*

14

15    The 47 kD and 31 kD polypeptides were purified by two rounds of gel filtration  
16    on a SUPEROSE-12 column of the PHARMACIA Fast Protein Liquid  
17    Chromatography system (FPLC), or by electroelution of bands after preparative  
18    electrophoresis. (The words SUPEROSE and PHARMACIA are trade marks.)  
19    Concentrated protein extracts were made from 50 mg acetone powder per ml of  
20    extraction buffer, and 1-2 ml loaded onto 2 mm thick SDS-PAGE gels poured  
21    without a comb. After electrophoresis the gel was surface stained in aqueous  
22    Coomassie Blue, and the 47 kD and 31 kD bands cut out with a scalpel. Gel  
23    slices were electroeluted into dialysis bags in electrophoresis running buffer at  
24    15 V for 24 hours, and the dialysate dialysed further against 0.1% SDS.  
25    Samples could be concentrated by lyophilisation.

26

27    Example 2

28

29    *Amino-acid Sequence Data from Proteins*

30

31    Protein samples (about 10  $\mu$ g) were subjected to conventional N-terminal  
32    amino-acid sequencing. The 47 kD and 31 kD polypeptides were N-terminally  
33    blocked, so cyanogen bromide peptides of the 47 kD and 31 kD peptides were

1 prepared, and some amino-acid sequence was derived from these. Cyanogen  
2 bromide cleaves polypeptide chains at methionine residues, and thus cleaved the  
3 47 kD and 31 kD polypeptides gave rise to 24 kD and 17 kD peptides. In  
4 addition the 47 kD polypeptide gave a 20 kD peptide. The 24 kD and 17 kD  
5 peptides had the same 9 N-terminal amino-acid residues. This fact, combined  
6 with the obvious one that the 31 kD could not contain both peptides  
7 consecutively, suggested that the 24 kD peptide arose for a partial digest, where  
8 full digestion would yield the 17 kD peptide. The other striking conclusion is  
9 that the 47 kD and 31 kD proteins are related, and the 31 kD could be a further  
10 processed form of the 47 kD. The 9 amino-acid sequence was used to construct  
11 an oligonucleotide probe for the 47 kD/31 kD gene(s).

12

13 Example 3

14

15 *Raising Antibodies to the 47 kD and 31 kD Polypeptides*

16

17 Polyclonal antibodies were prepared using the methodology of Catty and  
18 Raykundalia (1988). The serum was aliquoted into 1 ml fractions and stored at  
19 -20°C.

20

21 *Characterising Antibodies to the 47 kD and 31 kD Polypeptides*

22

23 Serum was immediately characterised using the Ochterloney double-diffusion  
24 technique, whereby antigen and antibody are allowed to diffuse towards one  
25 another from wells cut in agarose in borate-saline buffer. Precipitin lines are  
26 formed where the two interact if the antibody 'recognises' the antigen. This test  
27 showed that antibodies to both antigens had been formed, and furthermore that  
28 extensive cross-reaction took place between the 47 kD and 31 kD polypeptides  
29 and their respective antibodies. This is further indication that the 47 kD and 31  
30 kD polypeptides are closely related, as suggested by their cyanogen bromide  
31 cleavage patterns.

32

33

1 The gamma-globulin fraction of the serum was partially purified by  
2 precipitation with 50% ammonium sulphate, solubilisation in  
3 phosphate-buffered saline (PBS) and chromatography on a DE 52 cellulose  
4 ion-exchange column as described by Hill, 1984. Fractions containing  
5 gamma-globulin were monitored at 280 nm (OD<sub>280</sub> of 1.4 is equivalent to 1  
6 mg/ml gamma-globulin) and stored at -20°C.  
7 The effective titre of the antibodies was measured using an enzyme-linked  
8 immunosorbant assay (ELISA). The wells of a polystyrene microtitre plate  
9 were coated with antigen (10-1000 ng) overnight at 4°C in carbonate coating  
10 buffer. Wells were washed in PBS-Tween and the test gamma globulin added at  
11 concentrations of 10, 1 and 0.1 µg/ml (approximately 1:100, 1:1000 and  
12 1:10,000 dilutions). The diluent was PBS-Tween containing 2% polyvinyl  
13 pyrrolidone (PVP) and 0.2% BSA. Controls were preimmune serum from the  
14 same animal. Binding took place at 37°C for 3-4 hours. The wells were  
15 washed as above and secondary antibody (goat anti-rabbit IgG conjugated to  
16 alkaline phosphatase) added at a concentration of 1 µg/ml, using the same  
17 conditions as the primary antibody. The wells are again washed, and alkaline  
18 phosphatase substrate (p-nitrophenyl phosphate; 0.6 mg/ml in diethanol-amine  
19 buffer pH 9.8) added. The yellow colour, indicating a positive reaction, was  
20 allowed to develop for 30 minutes and the reaction stopped with 3M NaOH.  
21 The colour is quantified at 405 nm. More detail of this method is given in Hill,  
22 1984. The method confirmed that the antibodies all had a high titre and could  
23 be used at 1 µg/ml concentration.

24

25 Example 4

26

27 *Isolation of Total RNA from Immature Cocoa Beans*

28

29 The starting material for RNA which should contain a high proportion of  
30 mRNA specific for the storage proteins was immature cocoa beans, at about 130  
31 days after pollination. Previous work had suggested that synthesis of storage  
32 proteins was approaching its height by this date (Biehl *et al.*, 1982). The beans  
33 are roughly corrugated and pale pinkish-purple at this age.

1

2 The initial requirement of the total RNA preparation from cocoa beans was that  
3 it should be free from contaminants, as judged by the UV spectrum, particularly  
4 in the far UV, where a deep trough at 230 nm (260 nm : 230 nm ratio is  
5 approximately 2.0) is highly diagnostic of clean RNA, and is intact, as judged  
6 by agarose gel electrophoresis of heat-denatured samples, which should show  
7 clear rRNA bands. A prerequisite for obtaining intact RNA is scrupulous  
8 cleanliness and rigorous precautions against RNases, which are ubiquitous and  
9 extremely stable enzymes. Glassware is customarily baked at high  
10 temperatures, and solutions and apparatus treated with the RNase inhibitor  
11 diethyl pyrocarbonate (DEPC, 0.1%) before autoclaving.

12

13 The most routine method for extraction of plant (and animal) RNA is extraction  
14 of the proteins with phenol/chloroform in the presence of SDS to disrupt  
15 protein-nucleic acid complexes, and inhibit the RNases which are abundant in  
16 plant material. Following phenol extraction the RNA is pelleted on a caesium  
17 chloride gradient before or after ethanol precipitation. This method produced  
18 more or less intact RNA, but it was heavily contaminated with dark brown  
19 pigment, probably oxidised polyphenols and tannins, which always co-purified  
20 with the RNA. High levels of polyphenols are a major problem in *Theobroma*  
21 tissues.

22

23 A method was therefore adopted which avoided the use of phenol, and instead  
24 used the method of Hall *et al.* (1978) which involves breaking the tissue in hot  
25 SDS-borate buffer, digesting the proteins with proteinase K, and specifically  
26 precipitating the RNA with LiCl. This method gave high yields of reasonably  
27 clean, intact RNA. Contaminants continued to be a problem and the method  
28 was modified by introducing repeated LiCl precipitation steps, the precipitate  
29 being dissolved in water and clarified by microcentrifugation after each step.  
30 This resulted in RNA preparations with ideal spectra, which performed well in  
31 subsequent functional tests such as *in vitro* translation.

32

33

1      *Preparation of mRNA From Total RNA*

2

3      The mRNA fraction was separated from total RNA by affinity chromatography  
4      on a small (1 ml) oligo-dT column, the mRNA binding to the column by its  
5      poly A tail. The RNA (1-2 mg) was denatured by heating at 65°C and applied  
6      to the column in a high salt buffer. Poly A+ was eluted with low salt buffer,  
7      and collected by ethanol precipitation. The method is essentially that of Aviv  
8      and Leder (1972), modified by Maniatis *et al* (1982). From 1 mg of total  
9      RNA, approximately 10-20 µg polyA+ RNA was obtained (1-2%).

10

11     *In vitro Translation of mRNA*

12

13     The ability of mRNA to support *in vitro* translation is a good indication of its  
14     cleanliness and intactness. Only mRNAs with an intact polyA tail (3' end) will  
15     be selected by the oligo-dT column, and only mRNAs which also have an intact  
16     5' end (translational start) will translate efficiently. *In vitro* translation was  
17     carried out using RNA-depleted wheat-germ lysate (Amersham International),  
18     the *de novo* protein synthesis being monitored by the incorporation of [<sup>35</sup>  
19     S]-methionine (Roberts and Paterson, 1973). Initially the rate of *de novo*  
20     synthesis was measured by the incorporation of [<sup>35</sup> S]-methionine into  
21     TCA-precipitable material trapped on glass fibre filters (GFC, Whatman). The  
22     actual products of translation were investigated by running on SDS-PAGE,  
23     soaking the gel in fluor, drying the gel and autoradiography. The mRNA  
24     preparations translated efficiently and the products covered a wide range of  
25     molecular weights, showing that intact mRNAs for even the largest proteins had  
26     been obtained. None of the major translation products corresponded in size to  
27     the 47kD or 31kD storage polypeptides identified in mature beans, and it was  
28     apparent that considerable processing of the nascent polypeptides must occur to  
29     give the mature forms.

30

31

32

33

1    Example 5

2

3    *Identification of Precursor to the 47 kD and 31 kD Polypeptides by*  
4    *Immunoprecipitation*

5

6    Because the 47 kD and 31 kD storage polypeptides were not apparent amongst  
7    the translation products of mRNA from developing cocoa beans, the technique  
8    of immunoprecipitation, with specific antibodies raised to the storage  
9    polypeptides, was used to identify the precursors from the translation mixture.  
10   This was done for two reasons: first to confirm that the appropriate mRNA was  
11   present before cloning, and second to gain information on the expected size of  
12   the encoding genes.

13

14   Immunoprecipitation was by the method of Cuming *et al*, 1986. [<sup>35</sup> S]-labelled  
15   *in vitro* translation products were dissociated in SDS, and allowed to bind with  
16   specific antibody in PBS plus 1% BSA. The antibody-antigen mixture was then  
17   mixed with protein A-SEPHAROSE and incubated on ice to allow the IgG to  
18   bind to protein A. The slurry was poured into a disposable 1 ml syringe, and  
19   unbound proteins removed by washing with PBS +1% NONIDET P-40. The  
20   bound antibody was eluted with 1M acetic acid and the proteins precipitated  
21   with TCA. The antibody-antigen complex was dissociated in SDS, and subject  
22   to SDS-PAGE and fluorography, which reveals which labelled antigens have  
23   bound to the specific antibodies.

24

25   The results showed that the anti-47 kD and anti-31 kD antibodies both  
26   precipitated a 67 kD precursor. The precursor size corresponded to a major  
27   band on the *in vitro* translation products. The results with the 47 kD and 31 kD  
28   antibodies confirmed that the polypeptides are derived from a single precursor,  
29   or at least precursors of the same size. The large size of the precursor  
30   suggested that size-selection at mRNA or cDNA level may be necessary to  
31   obtain clones.

32

33

1    Example 6

2

3    *cDNA Synthesis From the mRNA Preparations*

4

5    cDNA synthesis was carried out using a kit from Amersham International. The  
6    first strand of the cDNA is synthesised by the enzyme reverse transcriptase,  
7    using the four nucleotide bases found in DNA (dATP, dTTP, dGTP, dCTP) and  
8    an oligo-dT primer. The second strand synthesis was by the method of Gubler  
9    and Hoffman (1983), whereby the RNA strand is nicked in many positions by  
10   RNase H, and the remaining fragments used to prime the replacement synthesis  
11   of a new DNA strand directed by the enzyme *E. coli* DNA polymerase I. Any  
12   3' overhanging ends of DNA are filled in using the enzyme T4 polymerase.  
13   The whole process was monitored by adding a small proportion of [<sup>32</sup>P]-dCTP  
14   into the initial nucleotide mixture, and measuring the percentage incorporation  
15   of label into DNA. Assuming that cold nucleotides are incorporated at the same  
16   rate, and that the four bases are incorporated equally, an estimate of the  
17   synthesis of cDNA can be obtained. From 1 µg of mRNA approximately 140  
18   ng of cDNA was synthesised. The products were analysed on an alkaline 1.4%  
19   agarose gel as described in the Amersham methods. Globin cDNA, synthesised  
20   as a control with the kit, was run on the same gel, which was dried down and  
21   autoradiographed. The cocoa cDNA had a range of molecular weights, with a  
22   substantial amount larger than the 600 bp of the globin cDNA.

23

24    Example 7

25

26    *Cloning of cDNA into a Plasmid Vector by Homopolymer Tailing*

27

28    The method of cloning cDNA into a plasmid vector was to 3' tail the cDNA  
29   with dC residues using the enzyme terminal transferase (Boehringer Corporation  
30   Ltd), and anneal into a *Pst*I-cut and 5' tailed plasmid (Maniatis *et al*, 1982  
31   Eschenfeldt *et al*, 1987). The optimum length for the dC tail is 12-20 residues.  
32   The tailing reaction (conditions as described by the manufacturers) was tested  
33

1 with a 1.5 kb blunt-ended restriction fragment, taking samples at intervals, and  
2 monitoring the incorporation of a small amount of [<sup>32</sup>P]-dCTP. A sample of  
3 cDNA (70 ng) was then tailed using the predetermined conditions.

4

5 A dG-tailed plasmid vector (3'-oligo(dG)-tailed pUC9) was purchased from  
6 Pharmacia. 15 ng vector was annealed with 0.5 - 5 ng of cDNA at 58°C for 2  
7 hours in annealing buffer: 5mM Tris-HCl pH 7.6; 1mM EDTA, 75 mM NaCl  
8 in a total volume of 50 µl. The annealed mixture was transformed into *E. coli*  
9 RRI (Bethesda Research Laboratories), transformants being selected on L-agar  
10 + 100 µg/ml ampicillin. Approximately 200 transformants per ng of cDNA  
11 were obtained. Transformants were stored by growing in 100 µl L-broth in the  
12 wells of microtitre plates. adding 100 µl 80% glycerol, and storing at -20°C.

13

14 Some of the dC tailed cDNA was size selected by electrophoresing on a 0.8%  
15 agarose gel, cutting slits in the gel at positions corresponding to 0.5, 1.0 and  
16 1.5 kb, inserting DE81 paper and continuing electrophoresis until the cDNA  
17 had run onto the DE81 paper. The DNA was then eluted from the paper with  
18 high salt buffer, according to the method of Dretzen *et al* (1981).

19

20

21 Example 8

22

23 *Construction of Oligonucleotide Probes for the 47/31 kD Gene*

24

25 The amino-acid sequence obtained from a cyanogen bromide peptide common to  
26 the 47 kD and 31 kD polypeptides is as follows:

27

28 Met-Phe-Glu-Ala-Asn-Pro-Asn-Thr-Phe

29

30 and the least redundant probe of 17 residues (a mixture of 32) is shown below:

31

32

33

7 The actual probe was made anti-sense so that it could also be used to probe  
8 mRNA. Probe synthesis was carried out using an Applied Biosystems  
9 apparatus.

- 10
- 11
- 12      Example

## 14 Use of Oligonucleotides to Probe cDNA Library

16 The oligonucleotide probes were 5' end-labelled with gamma-[<sup>32</sup>P] dATP and  
17 the enzyme polynucleotide kinase (Amersham International). The method was  
18 essentially that of Woods (1982, 1984), except that a smaller amount of isotope  
19 (15  $\mu$ Ci) was used to label about 40 ng probe, in 10 mM MgCl<sub>2</sub>, 100 mM  
20 Tris-HCl, pH 7.6; 20 mM 2-mercaptoethanol.

21 The cDNA library was grown on GeneScreen (New England Nuclear) nylon  
22 membranes placed on the surface of L-agar + 100 µg/ml ampicillin plates. (The  
23 word GeneScreen is a trade mark.) Colonies were transferred from microtitre  
24 plates to the membranes using a 6 x 8 multi-pronged device, designed to fit into  
25 the wells of half the microtitre plate. Colonies were grown overnight at 37°C,  
26 lysed in sodium hydroxide and bound to membranes as described by Woods  
27 (1982, 1984). After drying the membranes were washed extensively in 3 x  
28 SSC/0.1% SDS at 65°C, and hybridised to the labelled probe, using a HYBAID  
29 apparatus from Hybaid Ltd, PO Box 82, Twickenham, Middlesex. (The word  
30 HYBAID is a trade mark.) Conditions for hybridisation were as described by  
31 Mason & Williams (1985), a  $T_d$  being calculated for each oligonucleotide  
32 according to the formula:

1

2  $T_d = 4^{\circ}\text{C}$  per GC base pair +  $2^{\circ}\text{C}$  per AT base pair.

3 At mixed positions the lowest value is taken.

4

5 Hybridisation was carried out at  $T_d -5^{\circ}\text{C}$ . Washing was in 6 x SSC, 0.1% SDS  
6 initially at room temperature in the HYBAID apparatus, then at the  
7 hybridisation temperature ( $T_d -5^{\circ}\text{C}$ ) for some hours, and finally at  $T_d$  for  
8 exactly 2 minutes. Membranes were autoradiographed onto FUJI X-ray film,  
9 with intensifying screens at  $-70^{\circ}\text{C}$ . (The word FUJI is a trade mark.) After 24  
10 - 48 hours positive colonies stood out as intense spots against a low background.

11

12 Example 10

13

14 *Analysis of Positive Clones for the 47 kD/31 kD Polypeptide*

15

16 Only one positive clone, pMS600, was obtained. This released two *Pst*I  
17 fragments on digestion, of total length 1.3 kb, insufficient to encode the 67 kD  
18 precursor. The total insert was removed from the vector on a *Hind*III-*Eco*RI  
19 fragment, nick-translated and used to probe the cDNA library, picking up a  
20 further two positive clones, pMS700 and pMS800. Restriction mapping of all  
21 three inserts suggested an overlapping map covering nearly 2.0 kb, sufficient to  
22 encode the 67 kD precursor (see Figure 1).

23

24 Example 11

25

26 *Sequencing the Cloned Inserts*

27

28 The sequencing strategy was to clone the inserts, and where appropriate  
29 subclones thereof, into the multiple cloning site of the plasmids  
30 pTZ18R/pTZ19R (Pharmacia). These plasmids are based on the better-known  
31 vectors pUC18/19 (Norrander *et al.* 1983), but contain a single-stranded origin  
32 of replication from the filamentous phage f1. When superinfected with phages  
33 in the same group, the plasmid is induced to undergo single-stranded

1 replication, and the single-strands are packaged as phages extruded into the  
2 medium. DNA can be prepared from these 'phages' using established methods  
3 for M13 phages (Miller, 1987), and used for sequencing by the method of  
4 Sanger (1977) using the reverse sequencing primer. The superinfecting phage  
5 used is a derivative of M13 termed M13K07, which replicates poorly and so  
6 does not compete well with the plasmid, and contains a selectable  
7 kanamycin-resistance marker. Detailed methods for preparing single-strands  
8 from the pTZ plasmids and helper phages are supplied by Pharmacia. DNA  
9 sequence was compiled and analysed using the Staden package of programs  
10 (Staden, 1986), on a PRIME 9955 computer. (The word PRIME is a trade  
11 mark.)

12

13 Example 12

14

15 *Features of the 47 kD/31 kD cDNA and Deduced Amino-acid Sequence of the 67  
16 kD Precursor*

17

18 DNA sequencing of the three positive clones, pMS600, pMS700, pMS800,  
19 confirmed the overlap presumed in Figure 1. No sequence differences were  
20 found in the overlapping regions (about 300 bp altogether), suggesting that the  
21 three cDNAs were derived from the same gene. The sequence of the combined  
22 cDNAs comprising 1818 bases is shown in Figure 2. The first ATG codon is  
23 found at position 14, and is followed by an open reading frame of 566 codons.  
24 There is a 104-base 3' untranslated region containing a polyadenylation signal at  
25 position 1764. The oligonucleotide probe sequence is found at position 569.

26

27 The open reading frame translates to give a polypeptide of 566 amino-acids  
28 (Figure 2), and a molecular weight of 65612, which is reasonably close to the  
29 67 kD measured on SDS-PAGE gels. The N-terminal residues are clearly  
30 hydrophobic and look like a characteristic signal sequence. Applying the rules  
31 of Von Heije (1983), which predict cleavage sites for signal sequences, suggests  
32 a cleavage point between amino-acids 20 and 21 (see Figure 3). The region  
33 following this is highly hydrophobic and contains four Cys-X-X-X-Cys motifs.

1 The N-terminus of the mature protein has been roughly identified as the  
2 glutamate (E) residue at 135 (Figure 3), on the basis of some tentative  
3 N-terminal sequence (EEPGSQFANPAYHF). This N-terminus would give a  
4 mature protein of 49068 kD, in rough agreement with that observed. There  
5 appears to be no glycosylation sites (Asn-X-Ser/Thr) in the mature protein of  
6 the sequence.

7

8 *Homologies Between the 67 kD Precursor and Other Known Proteins*

9

10 Searches through the PIR database, and through the literature, revealed close  
11 homologies between the 67 kD polypeptide and a class of seed storage proteins  
12 termed vicilins, one of two major classes of globulins found in seed (Borrotto  
13 and Dure, 1987). Alignments between the 67 kD polypeptide and vicilins from  
14 cotton (*Gossypium hirsutum*, Ghi), soybean (*Glycine max*, Gma), pea (*Pisum*  
15 *sativum*, Psa-c is convicilin, Psa-v is vicilin) and bean (*Phaseolus vulgaris*, Pvu)  
16 are shown in Figure 4 (Bown *et al*, 1988; Chlan *et al*, 1986; Doyle *et al*, 1986;  
17 Lycett *et al*, 1983). Identical residues are boxed.

18

19 All the vicilins have a mature molecular weight of around 47 kD, with the  
20 exception of soybean conglycinin alpha and alphal subunits, which are 72 kD  
21 and 76 kD respectively, and pea convicilin with a mature molecular weight of  
22 64kD. The pea and bean subunits (2 sub classes each) are synthesised as small  
23 precursors, around 50 kD. The most striking homology with the 67 kD is the  
24 cotton vicilin (Chlan *et al*, 1986). Cotton is also the most closely related to  
25 cocoa: both are members of the order Malvales. Interestingly cotton also has a  
26 large precursor, of 69 kD, comprising a short signal sequence, a large  
27 hydrophilic domain containing six Cys-X-X-X-Cys motifs, and a mature  
28 domain. It may therefore be possible to synthesise the corresponding cotton  
29 protein, by techniques analogous to those disclosed in this application and to use  
30 the cotton protein, or fragments of it, in the preparation of flavour components  
31 analogous to cocoa flavour components.

32

33

1      Example 13

2

3      *Expression of the 67 kD Polypeptide in E. coli*

4

5      Before the 67 kD coding region could be inserted into a expression vector the  
6      overlapping fragments from the three separate positive clones had to be spliced  
7      into a continuous DNA segment. The method of splicing is illustrated in Figure  
8      6: a *Hind*III-*Bgl*II fragment from pMS600, a *Bgl*II-*Eco*RI fragment from  
9      pMS700 and an *Eco*RI-*Sal*II fragment from pMS800 were ligated into pTZ19R  
10     cut with *Hind*III and *Sal*II. The resulting plasmid, containing the entire 67 kD  
11     cDNA, was termed pMS900.

12

13     An *Nco*I site was introduced at the ATG start codon, using the mutagenic  
14     primer:

15

16        5' TAG CAA CCA TGG TGA TCA 3'.

17

18     *In vitro* mutagenesis was carried out using a kit marketed by Amersham  
19     International, which used the method of Eckstein and co-workers (Taylor *et al*,  
20     1985). After annealing the mutagenic primer to single-stranded DNA the  
21     second strand synthesis incorporates alpha-thio-dCTP in place of dCTP. After  
22     extension and ligation to form closed circles, the plasmid is digested with *Nci*I,  
23     an enzyme which cannot nick DNA containing thio-dC. Thus only the original  
24     strand is nicked, and subsequently digested with exonuclease III. The original  
25     strand is then resynthesised, primed by the remaining DNA fragments and  
26     complementing the mutated position in the original strand. Plasmids are then  
27     transformed into *E. coli* and checked by plasmid mini preparations.

28

29     The 67 kD cDNA was then cloned into the *E. coli* expression plasmid, pJLA502  
30     (Figure 5), on an *Nco*I - *Sal*II fragment (pMS902).

31

32

33

1 pJLA502 (Schauder *et al.*, 1987) is marketed by Medac GmbH, Postfach  
2 303629, D-7000, Hamburg 36 and contains the strong lambda promoters,  $P_L$   
3 and  $P_R$ , and the leader sequence and ribosome binding site of the very  
4 efficiently translated *E. coli* gene, *atpE*. It also contains a temperature-sensitive  
5 cI repressor, and so expression is repressed at 30°C and activated at 42°C. The  
6 vector has an *NcoI* site (containing an ATG codon: CCATGG) correctly placed  
7 with respect to the ribosome binding site, and foreign coding sequences must be  
8 spliced in at this point.

9

10 The expression vector was transformed into *E. coli* UT580. The transformed  
11 strain was grown in L-broth + ampicillin (100  $\mu$ g/ml) at 30°C until log phase  
12 ( $OD_{610} = 0.5$ ) and the temperature was then shifted to 42°C and samples taken  
13 at intervals. Samples were dissociated by boiling in SDS loading buffer, and  
14 run on SDS-PAGE gels. The proteins were electroblotted onto nitrocellulose  
15 membranes (Towbin *et al.*, 1979) and Western blotting carried out using the  
16 anti-21 kD antibody prepared in Example 3 above (at 2  $\mu$ g/ml) and as a  
17 secondary antibody, goat anti-rabbit-IgG conjugated to alkaline phosphatase  
18 (Scott *et al.*, 1988).

19

20 A specific band at 67 kD was produced by pMS902, showing that a functional  
21 gene was present.

22

23 Example 14

24

25 *Expression of the 67 kD Polypeptide in Yeast*

26

27 Two yeast expression vectors were used, both based on a yeast-*E. coli* shuttle  
28 vector containing yeast and *E. coli* origins of replication, and suitable selectable  
29 markers (ampicillin-resistance for *E. coli* and leucine auxotrophy for yeast).  
30 Both vectors contain the yeast pyruvate kinase (PK) promoter and leader  
31 sequence and have a *Hind*III cloning site downstream of the promoter. One  
32 vector, A (YVA), is designed for internal expression, and the other, B (YVB),  
33 for secreted expression, having a portion of the signal sequence of the yeast

1 mating alpha-factor downstream of the promoter, with a *Hind*III site within it to  
2 create fusion proteins with incoming coding sequences. The vectors are  
3 illustrated in Figure 7.

4

5 To use the vectors effectively it is desirable to introduce the foreign coding  
6 region such that for vector A, the region from the *Hind*III cloning site to the  
7 ATG start is the same as the yeast PK gene, and for vector B, the remainder of  
8 the alpha-factor signal, including the lysine at the cleavage point. In practice  
9 this situation was achieved by synthesising two sets of *Hind*III - *Nco*I linkers to  
10 breach the gap between the *Hind*III cloning site in the vector and the *Nco*I at the  
11 ATG start of the coding sequence. This is illustrated in Figure 8.

12

13 In order to use the yeast vector B, the hydrophobic signal sequence must first be  
14 deleted from the 67 kD cDNA. Although direct evidence of the location of the  
15 natural cleavage site was lacking, the algorithm of Von Heije predicts a site  
16 between amino-acids 20 (alanine) and 21 (leucine). However it was decided to  
17 remove amino-acids 2-19 by deletion, so that the useful *Nco*I site at the  
18 translation start would be maintained.

19

20

21 For ease of construction of the yeast vectors, the strategy was to first clone the  
22 *Hind*III - *Nco*I linkers into the appropriate pTZ plasmids, and then to clone the  
23 linkers plus coding region into the yeast vectors on *Hind*III - *Bam*HI fragments.  
24 However the coding region contains an internal *Bam*HI which must be removed  
25 by in vitro mutagenesis, giving a new plasmid pMS903. The signal sequence  
26 was deleted from pMS903 using the mutagenic primer

27

28           5'    AGCATA GCA ACC ATGG TTG CTT GTT CT  3'

29

30 to give pMS904. The appropriate *Hind*III - *Nco*I linkers were then cloned into  
31 pMS903 and pMS904 to give pMS907 and pMS905 respectively, and the  
32 *Hind*III - *Bam*HI fragments (linkers + coding region) subcloned from these  
33

1 intermediate plasmids into YVA and YVB respectively to give the yeast  
2 expression plasmids pMS908 and pMS906. A diagrammatic scheme for these  
3 and other constructs is given in Figure 9.

4

5 Because the mature cocoa protein appears to lack the N-terminal hydrophilic  
6 domain, as described in Example 12, expression vectors have also been  
7 designed to express the mature protein directly. Yeast is unlikely to have the  
8 same processing enzymes as cocoa and optimum expression may be obtained for  
9 a protein as close as possible to that found naturally in cocoa. Hence the DNA  
10 encoding the hydrophilic domain (amino acids 20-134) was deleted from the  
11 intermediate plasmids pMS907 and pMS905 to give plasmids pMS911 and  
12 pMS909 respectively, and the *Hind*III - *Bam*HI fragments for these were cloned  
13 into YVA and YVB to give the expression plasmids pMS912 and pMS910  
14 (Figure 9).

15

16 A further modification was introduced by constructing expression in which the  
17 plant terminator had been removed and replaced with the yeast ADH terminator  
18 (present in YVA and YVB). The plant signal was removed by cutting the  
19 intermediate plasmids pMS907 and pMS905 at the *Pvu*II site immediately  
20 downstream of the coding region, at position 1716 in Figure 2. *Hind*III linkers  
21 were added and the entire coding region cloned into the yeast expression vectors  
22 on *Hind*III - *Hind*III fragments giving expression plasmids pMS914 (YVA) and  
23 pMS916 (YVB) (Figure 9). A summary of the constructs made is given in  
24 Figure 10.

25

26 The yeast expression plasmids were transferred into yeast spheroplasts using the  
27 method of Johnston (1988). The transformation host was the LEU<sup>-</sup> strain  
28 AH22, and transformants were selected on leucine-minus minimal medium.  
29 LEU<sup>+</sup> transformants were streaked to single colonies, which were grown in 50  
30 ml YEPD medium (Johnston, 1988) at 28 °C for testing the extent and  
31 distribution of foreign protein. Cells were harvested from cultures in  
32 preweighed tubes in a bench-top centrifuge, and washed in 10 ml lysis buffer  
33 (200 mM Tris, pH 8.1: 10% glycerol). The cell medium was reserved and

1 concentrated 10-25 x in an AMICON mini concentrator. (The word AMICON  
2 is a trade mark.) The washed cells were weighed and resuspended in lysis  
3 buffer plus protease inhibitors (1 mM phenyl methyl sulphonyl fluoride  
4 (PMSF); 1  $\mu$ g/ml aprotinin; 0.5  $\mu$ g/ml leupeptin) at a concentration of 1 g/ml.  
5 1 volume acid-washed glass-beads was added and the cells broken by vortexing  
6 for 8 minutes in total, in 1 minute bursts, with 1 minute intervals on ice. After  
7 checking under the microscope for cell breakage, the mixture was centrifuged at  
8 7000 rpm for 3 minutes to pellet the glass beads. The supernatant was removed  
9 to a pre-chilled centrifuge tube, and centrifuged for 1 hour at 20,000 rpm.  
10 (Small samples can be centrifuged in a microcentrifuge in the cold.) The  
11 supernatant constitutes the soluble fraction. The pellet was resuspended in 1 ml  
12 lysis buffer plus 10% SDS and 1% mercaptoethanol and heated at 90°C for 10  
13 minutes. After centrifuging for 15 minutes in a microcentrifuge the supernatant  
14 constitutes the particulate fraction.

15  
16 Samples of each fraction and the concentrated medium were examined by  
17 Western blotting. Considering first the plasmids designed for internal  
18 expression in YVA, pMS908 produced immunoreactive proteins at 67 kD and  
19 16 kD within the cells only. There was no evidence of the 67 kD protein being  
20 secreted under the influence of its own signal sequence. The smaller protein is  
21 presumed to be a degradation product. A similar result, but with improved  
22 expression, was obtained with pMS914, in which the plant terminator is  
23 replaced by a yeast terminator. However in pMS912, in which the coding  
24 region for the hydrophilic domain has been deleted, no synthesis of  
25 immunoreactive protein occurred.

26  
27 For industrial production of heterologous proteins in yeast a secreted mode is  
28 preferable because yeast cells are very difficult to disrupt, and downstream  
29 processing from total cell protein is not easy. The results from the vectors  
30 constructed for secreted expressed were rather complicated. From the simplest  
31 construct, pMS906, in which the yeast  $\alpha$ -factor signal sequence replaces the  
32 plant protein's own signal, immunoreactive proteins of approximately 47 kD, 28  
33 kD and 18-20 kD were obtained and secreted into the medium. At first sight

1 this is surprising because the coding region introduced should synthesise a 67  
2 kD protein. However the most likely explanation is that the yeast's KEX2  
3 protease, that recognises and cleaves the  $\alpha$ -factor signal at a Lys-Arg site is also  
4 cleaving the 67 KD protein at Lys-Arg dipeptides at positions 148 and 313 in  
5 the amino-acid sequence. The calculated protein fragment sizes resulting from  
6 cleavage at these positions are 47179 Daltons, 28344 Daltons and 18835  
7 Daltons, very close to the observed sizes.

8

9 When the plant terminator is replaced with a yeast terminator in pMS916 no  
10 expression is obtained in either cells or medium; it is possible that a mutation  
11 has been inadvertently introduced. From the construct pMS910, in which the  
12 hydrophilic domain has been deleted the main antigenic products were 28 KD  
13 and 18-20 kD, again secreted into the medium. It is presumed that the *de novo*  
14 47 kD product is immediately cleaved at the KEX2 site at position 313.

15

16 In summary, four of the six expression vectors constructed direct the synthesis  
17 of proteins cross-reacting with anti-47 kD antibodies. Two of the constructs  
18 secrete proteins into the medium.

19

20 Example 15

21

22 *Construction of Vectors Designed to Express the 67 kD Protein in Hansenula*  
23 *polymorpha*

24

25 The methylotropic yeast *Hansenula polymorpha* offers a number of advantages  
26 over *Saccharomyces cerevisiae* as a host for the expression of heterologous  
27 proteins (EP-A-0173378 and Sudbery *et al*, 1988). The yeast will grow on  
28 methanol as sole carbon source, and under these conditions the enzyme  
29 methanol oxidase (MOX) can represent up to 40% of the total cell protein.  
30 Thus the MOX promoter is a very powerful one that can be used in a vector to  
31 drive the synthesis of heterologous proteins, and it is effective even as a single  
32 copy. This gives the potential to use stable integrated vectors. *Hansenula* can  
33 also grow on rich carbon sources such as glucose, in which case the MOX

1 promoter is completely repressed. This means that cells containing the  
2 heterologous gene can be grown to a high density on glucose, and induced to  
3 produce the foreign protein by allowing the glucose to run out and adding  
4 methanol.

5

6 A plasmid, pHGL1, containing the MOX promoter and terminator, and a  
7 cassette containing the yeast  $\alpha$ -factor secretory signal sequence, were prepared.  
8 The 67 kD coding region was cloned into pHGL1 on a *Bam*HI - *Bam*HI  
9 fragment, replacing the *Bgl*II fragment which contains the 3' end of the MOX  
10 coding region. The whole promoter - gene - terminator region can then be  
11 transferred to YEp13 on a *Bam*HI - *Bam*HI fragment to give the expression  
12 plasmid pMS922. The details of the construction are illustrated in Figure 11.  
13 An analogous expression plasmid, pMS925, has been constructed with the yeast  
14  $\alpha$ -factor spliced onto the 67 kD coding region, replacing the natural plant  
15 signal. The *Bam*HI - *Hind*III cassette containing the  $\alpha$ -factor was ligated to the  
16 *Hind*III - *Bam*HI fragment used to introduce the 67 kD coding region into YVB.  
17 The  $\alpha$ -factor plus coding region was then cloned with pHGL1 on a *Bam*HI -  
18 *Bam*HI fragment, and transferred into YEP13 as before. Details are shown in  
19 Figure 12.

20

21 Both constructs have been transformed into *Hansenula* and grown under  
22 inducing conditions with 0.5% or 1% methanol. Both constructs directed the  
23 production of immunoreactive protein within the cells, and pMS925 secreted the  
24 protein into the medium under the influence of the  $\alpha$ -factor signal sequence.

25

26 *E. coli* Strains

27

28 RR1 F<sup>r</sup> *v*<sub>B</sub><sup>r</sup> *M*<sub>B</sub> *ara-14* *proA2* *leuB6* *lacY1* *galK2* *vpsL20* (str<sup>r</sup>)  
29 *xyl-5* *mtl-1* *supE44*

30

31 CAG629 *lac*<sub>am</sub> *tvp*<sub>am</sub> *pho*<sub>am</sub> *htpR*<sub>am</sub> *mal* *rpsL* *lon* *supC*<sub>ts</sub>

32

33

1      UT580      (*lac-pro*) *supE thi hsdD5 / F'tra D36 proA<sup>+</sup>B<sup>+</sup> lacI<sup>q</sup> lacZ*  
2      M15

3

4

5

6      References

7

8      Aviv, H., and Leder, P. *Proc. Natl. Acad. Sci. USA* **69**, 1408-1412 (1972).  
9      Purification of biologically active globin mRNA by chromatography on oligo  
10     dT cellulose

11

12     Biehl, B., Wewetzer, C., and Passern, D. *J. Sci. Food Agric.* **33**, 1291-1304  
13     (1982). Vacuolar (Storage) Proteins of Cocoa Seeds and their Degradation  
14     during Germination and Fermentation.

15

16     Borroto, K., and Dure, L. *Plant Molecular Biology* **8**, 113-131 (1987). The  
17     globulin seed storage proteins of flowering plants are derived from two ancestral  
18     genes.

19

20     Bown.D., Ellis, T.H.N., and Gatehouse, J.A. *Biochem. J.* **251**, 717-726  
21     (1988). The sequence of a gene encoding convicilin from pea (*Pisum sativum*)  
22     shows that convicilin differs from vicilin by an insertion near the N-terminus.

23

24     Catty, D. and Raykundalia, C. Production and Quality control of Polyclonal  
25     Antibodies in: "Antibodies: A Practical Approach" Vol I, IRL Press (1988)

26

27     Chlan, C.A., Pyle, J.B., Legocki, A.B., and Dure, L. *Plant Molecular*  
28     *Biology* **7**, 475-489 (1986). Developmental biochemistry of cotton seed  
29     embryogenesis and germination XVIII. cDNA and amino acid sequences of  
30     members of the storage protein families.

31

32

33

1 Cuming, A.C., Williams, R.S., and Cullimore, J.V. in "Immunology in Plant  
2 Science", Ed. Wang, T.L., Cambridge University Press, 1986. The use of  
3 Antibodies in Molecular Biology.

4

5 Doyle, J.J., Schuler, M.A., Godette, W.D., Zencer, V., Beachy, R.N., and  
6 Slightom, J.L. *J. Biol. Chem.* **261**, 9228-9238 (1986). The glycosylated seed  
7 storage proteins of *Glycine max* and *Phaseolus vulgaris* : Structural homologies  
8 of genes and proteins.

9

10 Dretzen, G., Bellard, M., Sassone-Corsi, P., and Chambon, P. *Analytical  
11 Biochemistry* **112**, 295-298 (1981). A reliable method for the recovery of DNA  
12 fragments from agarose and acrylamide gels.

13

14 Eschenfeldt, W.H., Puskas, R.S., and Berger, S.L. *Methods in Enzymology*  
15 **152**, 337-342 (1987). Homopolymeric Tailing.

16

17 Fritz *et al* (*J. Food Sci.* **50** 946-950 (1985))

18

19 Gubler, U., and Hoffman, B.J. *Gene* **25**, 263 (1983). A simple and very  
20 efficient method for generating cDNA libraries.

21

22 Hall, T.C., Ma, Y., Buchbinder, B.U., Pyrne J.W., Sun, S.M., and Bliss,  
23 F.A. *Proc. Natl. Acad. Sci. USA* **75**, 3196-3200 (1978). Messenger RNA for  
24 G1 protein of French bean seeds : cell-free translation and product  
25 characterisation.

26

27 Hill, S.A. "Methods in Plant Virology", Blackwell 1984.

28

29 Johnston, J.R. in "Yeast : A practical approach". Eds Campbell, I., and  
30 Duffus, J.H. IRL Press. 1988. Yeast Genetics, Molecular Aspects.

31

32 Kreil, G. *Annual Rev. Biochem.* **50**, 317-348 (1981). Transfer of proteins  
33 across membranes.

1  
2 Laemmli, U.K. *Nature* 227, 680 (1970). Cleavage of structural proteins  
3 during the assembly of the head of bacteriophage T4.  
4  
5 Lipman, D.J., and Pearson, W.R. *Science* 227, 1435-1441 (1985). Rapid  
6 protein sequence similarity searches.  
7  
8 Lycett, G.W., Delauney, A.J., Gatehouse, J.A., Gilroy, J., Croy, R.R.D., and  
9 Boulter, D. *Nucleic Acids Res.* 11, 2367-2380 (1983).  
10  
11 Maniatis, T., Fritsch, E.F., and Sambrook, J. "Molecular Cloning: A  
12 Laboratory Manual", Cold Spring Harbour Laboratory, 1982.  
13  
14 Mason, P.J., and Williams, J.G. in "Nucleic Acid Hybridisation : A Practical  
15 Approach". Ed. Hames, B.D., and Higgins, S.J. IRL Press 1985.  
16 Hybridisation in the Analysis of Recombinant DNA.  
17  
18 Meinkoth, J., and Wahl, G.M. *Analytical Biochemistry* 138, 267 (1984).  
19 Methods of Southern blotting and DNA probing.  
20  
21 Miller, H. *Methods in Enzymology* 152, 145-170 (1987). Practical Aspects of  
22 Preparing Phage and Plasmid DNA : Growth, Maintenance and Storage of  
23 Bacteria and Bacteriophage.  
24  
25 Norrander, J., Kempe, T., and Messing, J. *Gene* 26, 101 (1983).  
26 Construction of improved M13 vectors using oligodeoxynucleotide-directed  
27 mutagensis.  
28  
29 Proudfoot, N.J., and Brownlee, G.G. *Nature* 263, 211-214 (1976). 3'  
30 Non-coding region sequences in enkayotic messenger RNA.  
31  
32  
33

1      Roberts, B.E., and Paterson, B.M. *Proc. Natl. Acad. Sci. USA* **70**, 2330  
2      (1973). Efficient translation of tobacco mosaic virus RNA and rabbit globin 9S.  
3      RNA in a cell-free system from commercial wheat germ.

4

5      Sanger, F., Nicklen, S., and Coulson, A.R. *Proc. Natl. Acad. Sci. USA* **74**,  
6      5463-5467 (1977). DNA sequencing with chain-terminating inhibitors.

7

8      Schauder, B., Blocker, H., Frank, R., and McCarthy, J.E.G. *Gene* **52**,  
9      279-283 (1987). Inducible expression vectors incorporating the *E. coli* *aptE*  
10     translation initiation region.

11

12     Scott, R., Draper, J., Jefferson, R., Dury, G., and Jacob, L. in "Plant Genetic  
13     Transformation and Gene Expression: A Laboratory Manual". Eds. Draper,  
14     J., Scott, R., Armitage, P., Walden, R. Blackwell 1988. Analysis of gene  
15     organisation and expression in plants.

16

17     Staden, R. *Nucleic Acids Res.* **14**, 217-231 (1986). The current status and  
18     portability of our sequence handling software.

19

20     Sudbery, P.E., Gleeson, M.A., Veale, R.A., Lederboer, A.M., and  
21     Zoetmulder, M.C.M. *Biocem. Soc. Trans* **16** 1081-103 (1988). Hansenula  
22     polymorpha as a novel yeast system for the expression of heterologous genes.

23

24     Taylor, J.W., Ott, J., and Eckstein, F. *Nucleic Acids Res.* **13**, 8765-8785  
25     (1985). The rapid generation of oligonucleotide-directed mutations at high  
26     frequency using phosphorothioate-modified DNA.

27

28     Towbin, H., Staehelin, T., and Gordon, J. *Proc. Natl. Acad. Sci. USA* **76**,  
29     4350-4534 (1979). Electrophoretic transfer of proteins from polyacrylamide  
30     gels to nitrocellulose sheets : procedure and some applications.

31

32     Von Heije, G. *Eur. J. Biochem* **133**, 17-21 (1983). Patterns of Amino-acids  
33     near Signal-Sequence Cleavage Sites.

1  
2 Wahl, G.M., and Berger, S.L. *Methods in Enzymology* 152, 415-423 (1987).  
3 Screening Colonies or Plaques with Radioactive Nucleic Acid Probes.  
4 Woods, D.E. *Focus (Bethesda Research Labs)* 6, 3 (1984). Oligonucleotide  
5 Screening of cDNA Libraries.  
6  
7 Woods, D.E., Markham, A.F., Ricker, A.T., Goldberger, G., and Colten,  
8 H.R. *Proc. Natl. Acad. Sci. USA* 79, 5561 (1982).  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33

## **CLAIMS**

3 1. A 67kD protein of *Theobroma cacao*, or a fragment thereof.  
4  
5 2. A 47kD protein of *Th. cacao*, or a fragment thereof.  
6  
7 3. A 31kD protein of *Th. cacao*, or a fragment thereof.  
8  
9 4. A protein as claimed in claim 1, 2 or 3, having at least part of the  
10 sequence shown in Figure 2.  
11  
12 5. A fragment as claimed in any one of claims 1 to 4, which comprises at  
13 least four amino acids.  
14  
15 6. A protein or fragment as claimed in any one of claims 1 to 6, which is  
16 recombinant.  
17  
18 7. Recombinant or isolated nucleic acid coding for a protein or fragment as  
19 claimed in any one of claims 1 to 5.  
20  
21 8. Nucleic acid as claimed in claim 7 which is DNA.  
22  
23 9. Nucleic acid as claimed in claim 8, having at least part of the sequence  
24 shown in Figure 2.  
25  
26 10. Nucleic acid as claimed in claim 7, 8 or 9, which is in the form of a  
27 vector.  
28  
29 11. Nucleic acid as claimed in claim 10, wherein the vector is an expression  
30 vector and the protein- or fragment-coding sequence is operably linked to a  
31 promoter.  
32  
33

1       12. Nucleic acid as claimed in claim 11, wherein the expression vector is a  
2 yeast expression vector and the promoter is a yeast pyruvate kinase (PK)  
3 promoter.

4

5       13. Nucleic acid as claimed in claim 11, wherein the expression vector is a  
6 bacterial expression vector and the promoter is a strong lambda promoter.

7

8       14. Nucleic acid as claimed in claim 11, 12 or 13, comprising a signal  
9 sequence.

10

11       15. A host cell comprising nucleic acid as claimed in any one of claims 10 to  
12 14.

13

14       16. A host cell as claimed in claim 15 which is *Saccharomyces cerevisiae*.

15

16       17. A host cell as claimed in claim 15 which is *E. coli*.

17

18       18. A process for the preparation of a protein or fragment as claimed in any  
19 one of claims 1 to 5, the process comprising coupling successive amino acids by  
20 peptide bond formation.

21

22       19. A process for the preparation of a protein or fragment as claimed in any  
23 one of claims 1 to 5, the process comprising culturing a host cell as claimed in  
24 claim 15, 16 or 17.

25

26       20. A process for the preparation of nucleic acid as claimed in any one of  
27 claims 7 to 14, the process comprising coupling together successive nucleotides  
28 and/or ligating oligo- or poly-nucleotides.

29

30

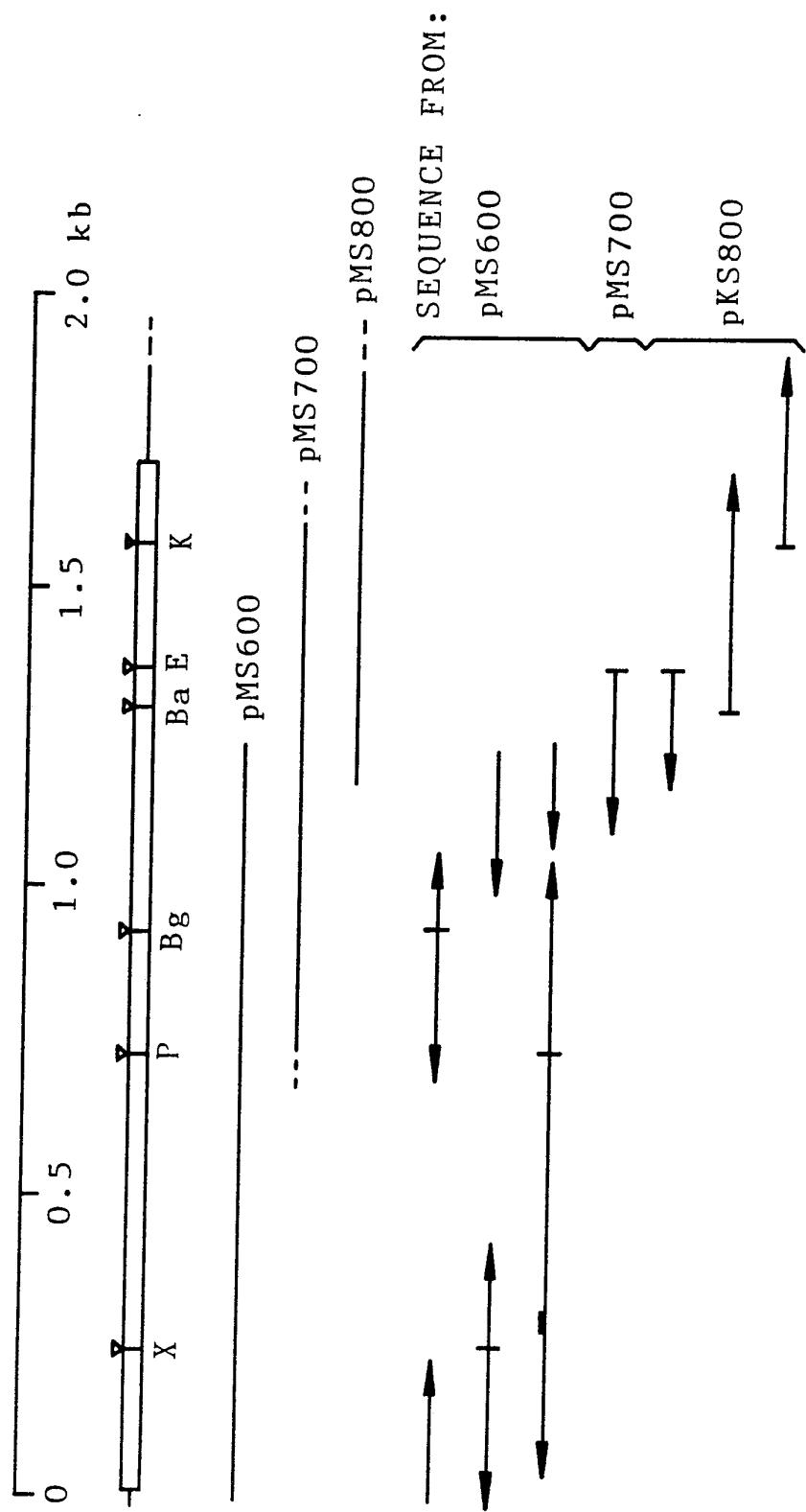
31

32

33

1/22

FIG. 1



2/22

FIG. 2A

3/22

R	D	E	E	G	N	F	K	I	L	Q	R	F	A	E	N	S	P	P	L
CAGGGATGAAGAGGGCAACTTCAAGATCCACTTCCAGAGGTTGCTGAGAACTCTCCTCCACT	490	500	510	520	530	540													
CAAAGGGCATCAACCGATTACCGCTTGGCCATGTTGAGCAAATCCCACACTTTATCT	550	560	570	580	590	600													
P	H	H	C	D	A	E	A	I	Y	F	V	T	N	G	K	G	T	I	T
TCGGCACCACTGTGATGCTGAGGCAATTACTTCGTGACAAACGGAAAGGGACAATTAC	610	620	630	640	650	660													
F	V	T	H	E	N	K	E	S	Y	N	V	Q	R	G	T	V	V	S	V
GTTRTGTGACTCATGAAACAAAGACTTAAATGTCAGCGTGGAACAGTAGTCAGCGT	670	680	690	700	710	720													
P	A	G	S	T	V	Y	V	V	S	Q	D	N	Q	E	K	L	T	I	A
TCCTGCAGGAAGCACTGTTACGTGGTTAGCCAAAGACAAACCAAGAGCTAACCATAGC	730	740	750	760	770	780													
V	L	A	L	P	V	N	S	P	G	K	Y	E	L	F	F	P	A	G	N
TGTGCTGCCCTGGCTTAATTCTCCCTGGCAAATATGAGTTATTCTGAAGTCTTGGCAA	790	800	810	820	830	840													
N	K	P	E	S	Y	Y	G	A	F	S	Y	E	V	L	E	T	V	F	N
TAATAAACCTGAATCATATTACGGAGGCCATTCTGAGCTATGAAGTCTTGAGACCGTCTTCAA	850	860	870	880	890	900													
T	Q	R	E	K	L	E	E	I	L	E	E	Q	R	G	Q	K	R	Q	Q
TACACAAAGAGAAGCTGGAGGAGATCTGGAGGAACAGAGAGGGCAGAAAGAGGGCAGCA	910	920	930	940	950	960													

FIG. 2B

4/22

G	Q	Q	G	M	F	R	K	A	K	P	E	Q	I	R	A	I	S	Q	Q
GGGCAGCGGGTATGTTCCGGAAAGCCAAACCAAGAGGAGATAAGAGCAATAAGCCAAACA	970	980	990	990	1000	1000	1010	1010	1020	1020									
A	T	S	P	R	H	R	G	G	E	R	L	A	I	N	L	S	Q	S	
AGCTACTTCTCCAAAGGCACAGGGGAGAGACTTGCCTCAATCTATTGAGGACTTCAGTCA	1030	1040	1050	1050	1060	1060	1070	1070	1080	1080									
P	V	Y	S	N	Q	N	G	R	F	F	E	A	C	P	E	D	F	S	Q
GCCTGCTACTCCAAACCAAACGGGACCGCTTCTTGAAGGCTTGTCCCTGAGGACTTCAGTCA	1090	1100	1110	1110	1120	1120	1130	1130	1140	1140									
F	Q	N	M	D	V	A	V	S	A	F	K	L	N	Q	G	A	I	F	V
ATTTCAAGAACATGGATGTCGCTGTTTCAGCCTTCAAACACTGAATTCAGGGAGCCATATTGT	1150	1160	1170	1170	1180	1180	1190	1190	1200	1200									
P	H	Y	N	S	K	A	T	F	V	V	F	V	T	D	G	Y	G	Y	A
GCCACACTACAAATTCTAAGGCTACATTCTGGTGTGTTGTCAAGGACGGATATGGGTACGC	1210	1220	1230	1230	1240	1240	1250	1250	1260	1260									
Q	M	A	C	P	H	L	S	R	Q	S	Q	G	S	Q	S	G	R	Q	D
TCAAATGGCTTGCCTGCATCTCCAGACAGAGCCAGGGATCCCAGAGGAGACATTGGAGAATGGCAAGA	1270	1280	1290	1290	1300	1300	1310	1310	1320	1320									
R	R	E	Q	E	E	S	E	E	T	F	G	E	F	Q	Q	V	K		
CAGAAGAGAACAAAGAACAGAGACTTCAGAAGAGGAGACATTGGAGAATGGCAAGGTC	1330	1340	1350	1350	1360	1360	1370	1370	1380	1380									
A	P	L	S	P	G	D	V	F	V	A	P	A	G	H	A	V	T	F	F
AGCCCCATTGTCACCTGGTGAAGGTCTTGTAGCCCCGGCAGGCCATGCAGTTACATTC	1390	1400	1410	1410	1420	1420	1430	1430	1440	1440									

FIG. 2C

5/22

A	S	K	D	Q	P	L	N	A	V	A	F	G	L	N	A	Q	N	N	Q
TGCAATCCAAGAACCGCCCTGAATGCAGTTGCCTTGGACTCAACGCCAGAACCAACCA																			
1450	1460	1470	1480	1490	1500														
R	I	F	L	A	G	K	K	N	L	V	R	Q	M	D	S	E	A	K	E
GAGAATTTCCTTGCAGGGAAAAAGAACCTGGTCAGACAAATGGATAGCCAGGAAAGGA																			
1510	1520	1530	1540	1550	1560														
L	S	F	G	V	P	S	K	L	V	D	N	I	F	N	N	P	D	E	S
GTATCATTTGGGTACCATCGAAATTGGTAGATAATATTCAACAAACCCGGATGAGTC																			
1570	1580	1590	1600	1610	1620														
Y	F	M	S	F	S	Q	Q	R	Q	R	R	D	E	R	R	G	N	P	L
GTATTCATGTTCTCTCAACAGAGGCAGCGTCCAGATGAAAGGAGGGCAATCCCTT																			
1630	1640	1650	1660	1670	1680														
A	S	I	L	D	F	A	R	L	F	*									
GGCCTCAATTCTGGACTTTGCCGCTTGTCTAAGCAGCTGCTTCCACTTTGTATCAGA																			
1690	1700	1710	1720	1730	1740														
CATGCAGAGGCATGTAATGCAATAATAAGTTGCCCCATGTTAAAGAGGAGAGTTTGCT																			
1750	1760	1770	1780	1790	1800														
TTTGTCTGTTCTAACCTTGTAACTAGTAAACTTTCAATGAAATGAGAGTTGTTAT																			
1810	1820	1830	1840	1850	1860														
CTTTCTA																			

FIG. 2D

6/22

10 MVISKSPPFIV LIFSLLLSFA LLCSSGVSAYG RKQYERDPRQ QYEQCQRRCSE SEATEEREQE  
 20 QCEQRRCEREF KEQQRQQEEE LQRQYQQCQG RCQEQQQGGQR EQQQCQRKCW EQYKEQERGE  
 30 HENYHNNHKKN RSEEEEGQQR NNPYYFPKRR SFQTRFRDFF GNFKILQRFRA ENSPPLKGIN  
 40 DYRLAMFEAN PNTFILPHHC DAEAIYFVTN GKGTITFVTH ENKESYNVQR GTVVSVPAGS  
 50 TVYVVVSQDNQ EKLTIAVLAL PVNSPGKYEL FFPAGNNKPE SYYGAFSYEV LETVFNTQRE  
 60 KLEEELEEQR GQKRQQGQQG MFRKAKPEQI RAISQQATSP RHRGGERLAI NLLSQSPVYS  
 70 NQNGRFFEAC PEDFSQFQNM DVAVSAFKLN QGAIFVPHYN SKATFVVVFVT DGYGYAQMAC  
 80 PHLSRQSQGS QSGRQDRREQ EEESEEEETFG EFQQVKAPLS PGDVFVAPAG HAVTFFASKD  
 90 QFPLNAVAFGL NAQNNQRIFL AGKKNLVRQM DSEAKELSFG VPSKLVNDNIF NNPDESYFMS  
 100 FSQQRQRDE RRGNPIASIL DFARLIF

FIG.3

7/22

FIG. 4A

## FIG. 4B

8/22

Tca	-	Q Q C Q R K C W E Q Y K E Q E R G E H E N Y H K K N R S E E E G E -	160	Q Q R N N P Y F P K	200
Ghi	R K Q Q C V A E C R E R Y Q E N P W R R E E A E E E E T E E G E -	170	Q Q S H N P F H F - H	190	Q Q R N N P Y F P K
Gma	R Q F P F P R P P H Q K E E R N E E E E E E E E D E E Q Q R E E E	180	Q E Q S H N P F L F F S	210	Q E H R N P F L F K S
Psa-c	-	-	220	- Q E H R N P F L F K S	240
Psa-v	-	-	230	- R S D Q E N P F L F K S	250
Pvu	-	-	260	- T S L R E E E S Q D N P F Y F N S	270
Tca	R R S F Q T R F R D E E G N F K I L Q R F A E N S P P L K G I N D Y R L A M F E A N P N T F I L P H	280	290	300	310
Ghi	R R S F Q S R F R E H G N F R V L Q R F A S R H P I L R G I N E F R L S I L E A N P N T F I L P H	290	-	-	-
Gma	N R - F E T L F K N Q Y G R I R V L Q R F D K N Q L R D Y R I L E F N S K P N T L L P N	300	-	-	-
Psa-c	N K - F L T L F E N E N G H I R R L Q R F D K R S D L F E N L Q N Y R L V E Y K S K P H T L F L P Q	310	-	-	-
Psa-v	N R - F Q T L Y E N E N G H I R L Q R F D K R S K I F E N L Q N Y R L L E Y K S K P H T L F L P Q	320	-	-	-
Pvu	D N S W N T L F K N Q Y G H I R V L Q R F D Q Q S K R L Q N L E D Y R L V E F R S K P E T L L P Q	330	-	-	-
Tca	H C D A E A I Y F V T N G K G T I T F V T H E N K E S Y N V Q R G T V V -	340	350	360	370
Ghi	H C D A E K I Y L V T N G R G T L T F L T H E N K E S Y N V V P G V V -	340	-	-	-
Gma	H A D A D Y L I V I L N G T A I L S L V N N D D R D S Y R L Q S G D A L -	350	-	-	-
Psa-c	H I D A D L I L V V L N G K A I L T V L S P N D R N S Y N L E R G D T I -	360	-	-	-
Psa-v	Y T D A D F I L V V L S G K A T L T V L K S N D R N S E N L E R G D A I -	370	-	-	-
Pvu	Q A D A E L L V V R S G S A I L V L V K P D D R R E Y F F L T S D N P I F S D H Q K I P A G T I F	380	-	-	-

9/22

FIG. 4C

10 / 22

FIG. 4D

11/22

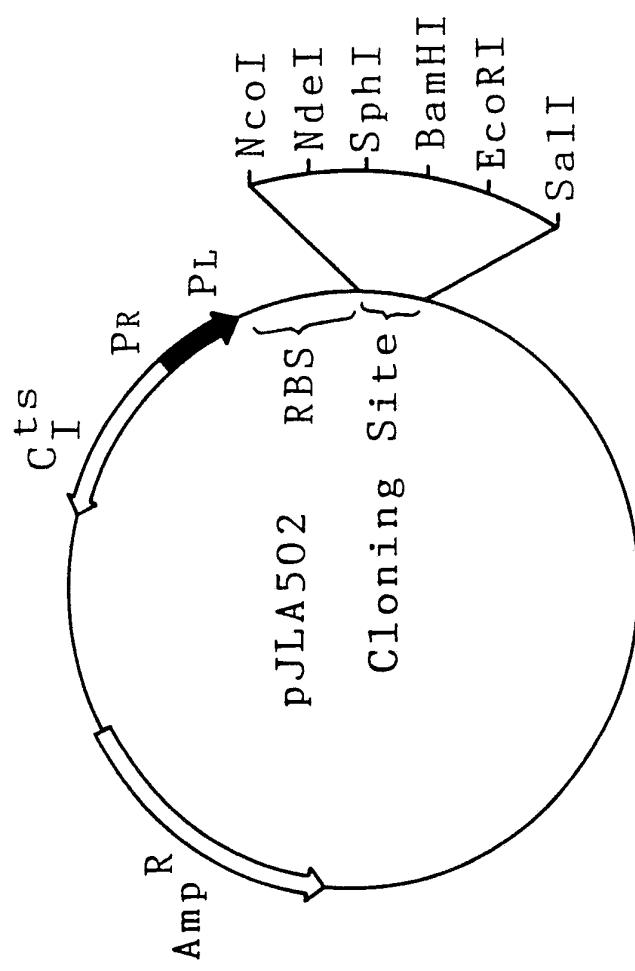
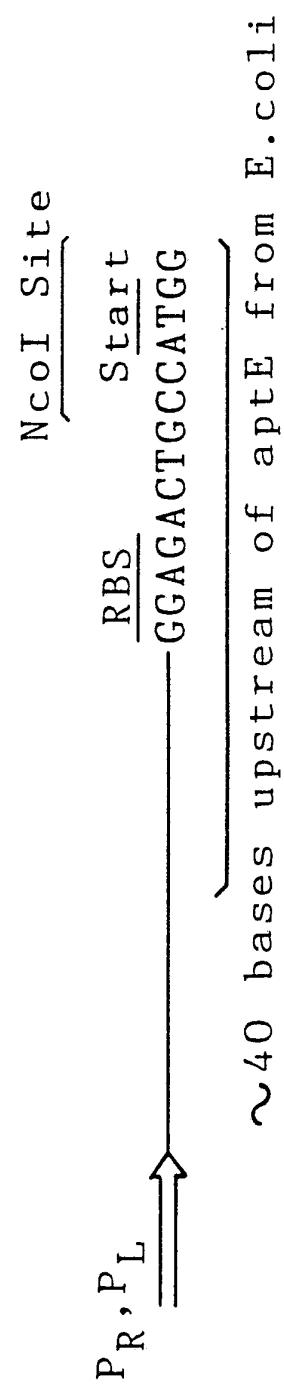
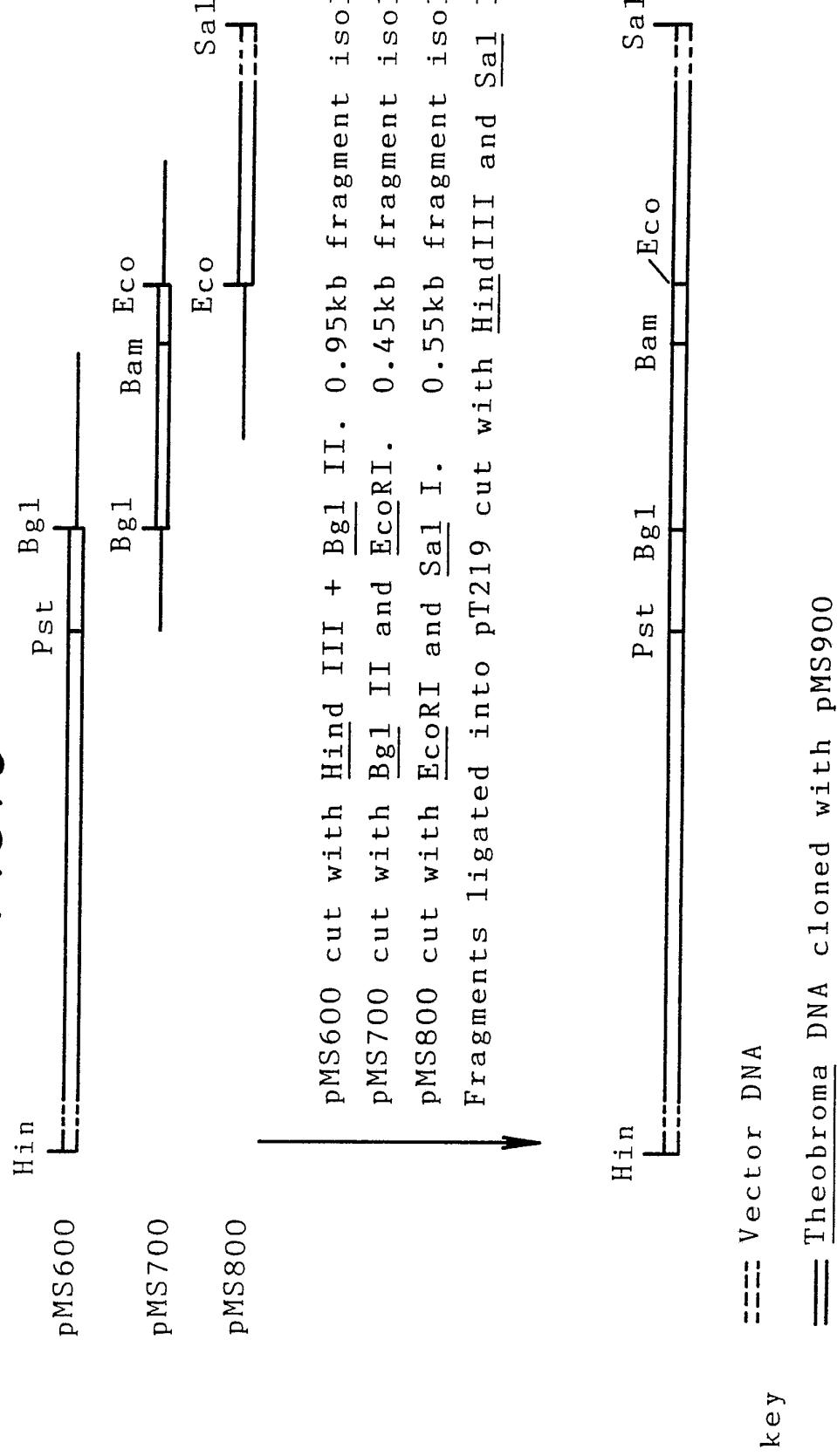


FIG. 5



12/22

## FIG. 6



13/22

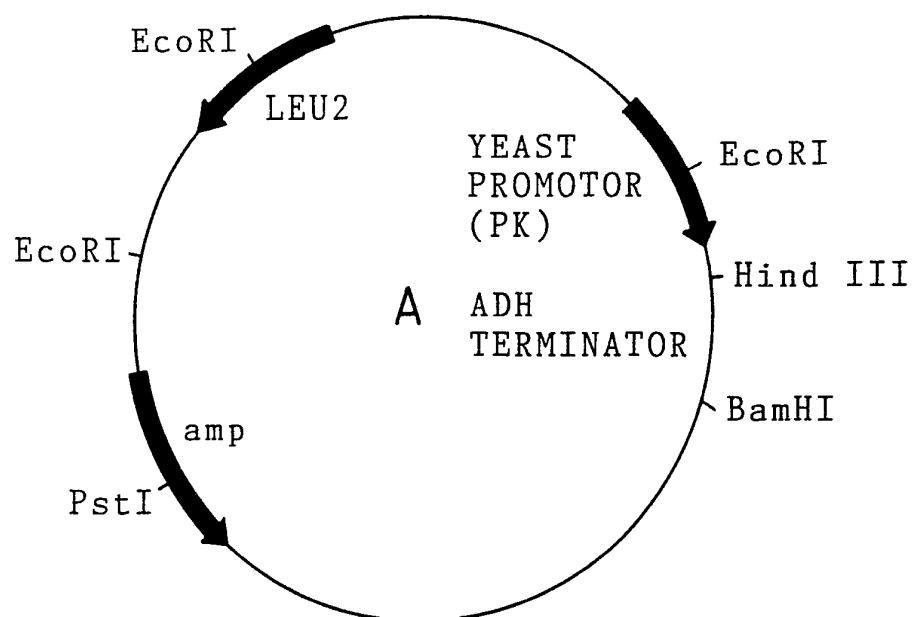
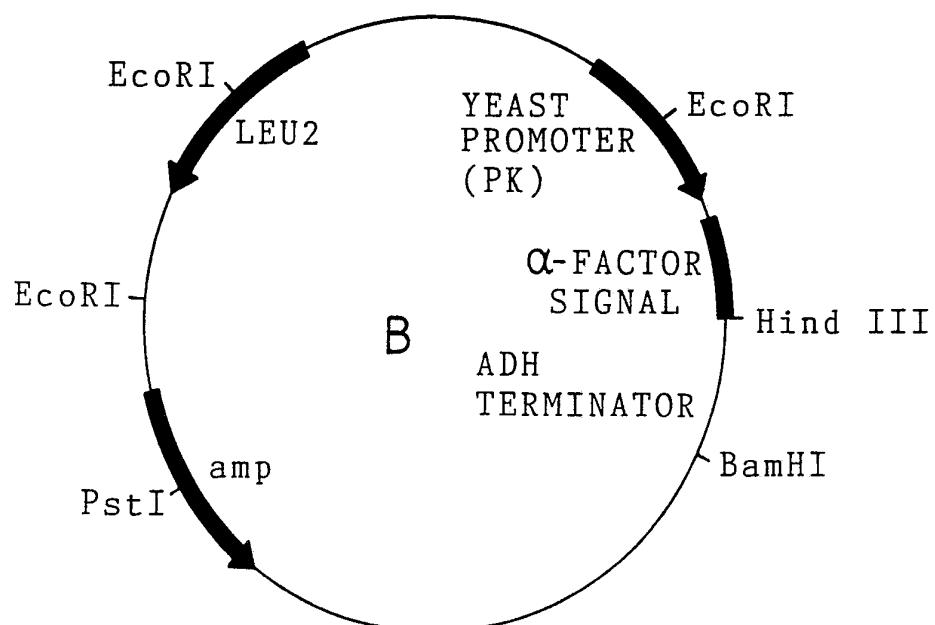


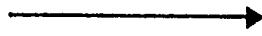
FIG. 7



## FIG. 8A

YEAST PYRUVATE  
KINASE GENE

TTTACAAGACACCAATCAAATAAAACAAATAACATCATCACAAATGTCTAGA  
MetSerArg



SEQUENCE ALTERED  
TO CREATE CLONING  
SITE

TTTACAAGCTCCAATCAAATAAAACAAATAACATCATCACAAATGTCTAGA  
MetSerArg  
14/22  
AGCT  
Hin

+

HIN-NCO LINKERS

AGCTTCCAATCAAATAAAACATCATCAC  
AGGTTAGTTTTGTTTATTTGTAGTAGTGGTAC



67kD EXPRESSION  
VECTOR A

TTTACAAGCTCCAATCAAATAAAACATCATCACCCATGGTGATC  
MetValIle  
Hin  
Nco

## FIG. 8B

YEAST ALPHA-FACTOR  
SIGNAL SEQUENCE

1 231  
 Met----GluGlyValSerLeuAspLysArgGlu  
 ATG----GAAGGGTAAGCTTGGATAAAAGAGAG  
 Hin

## HIN-NCO LINKERS

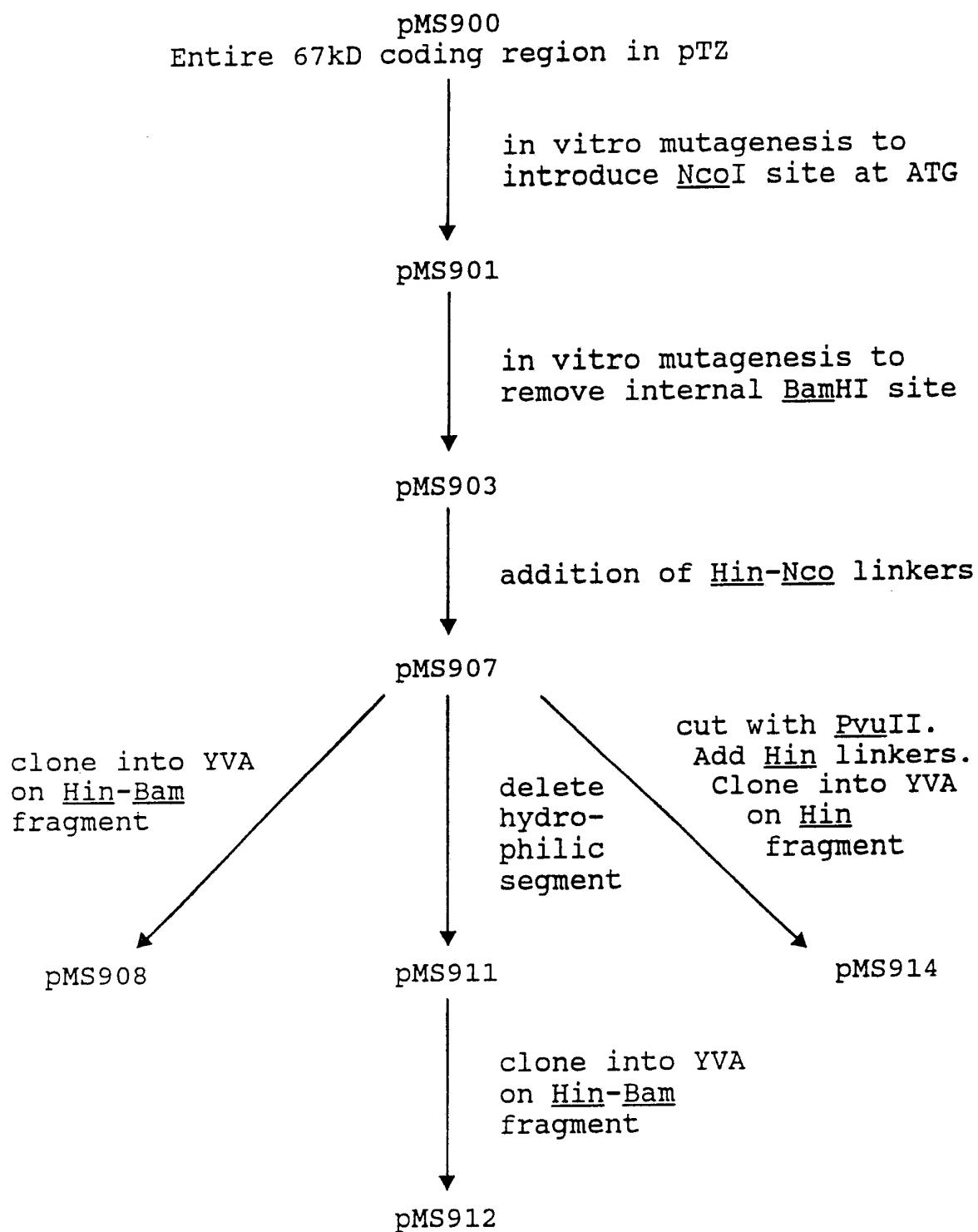
AGCTTGGATAAAAGAGC  
 ACCTATTCTCGGTAC

IN-PHASE FUSION OF  
 67KD CODING REGION

Met----GluGlyValSerLeuAspLysArgAlaMetAlaLeu  
 ATG----GAAGGGTAAGCTTGGATAAAAGAGCCCATGGCGTTG  
 Hin Nco

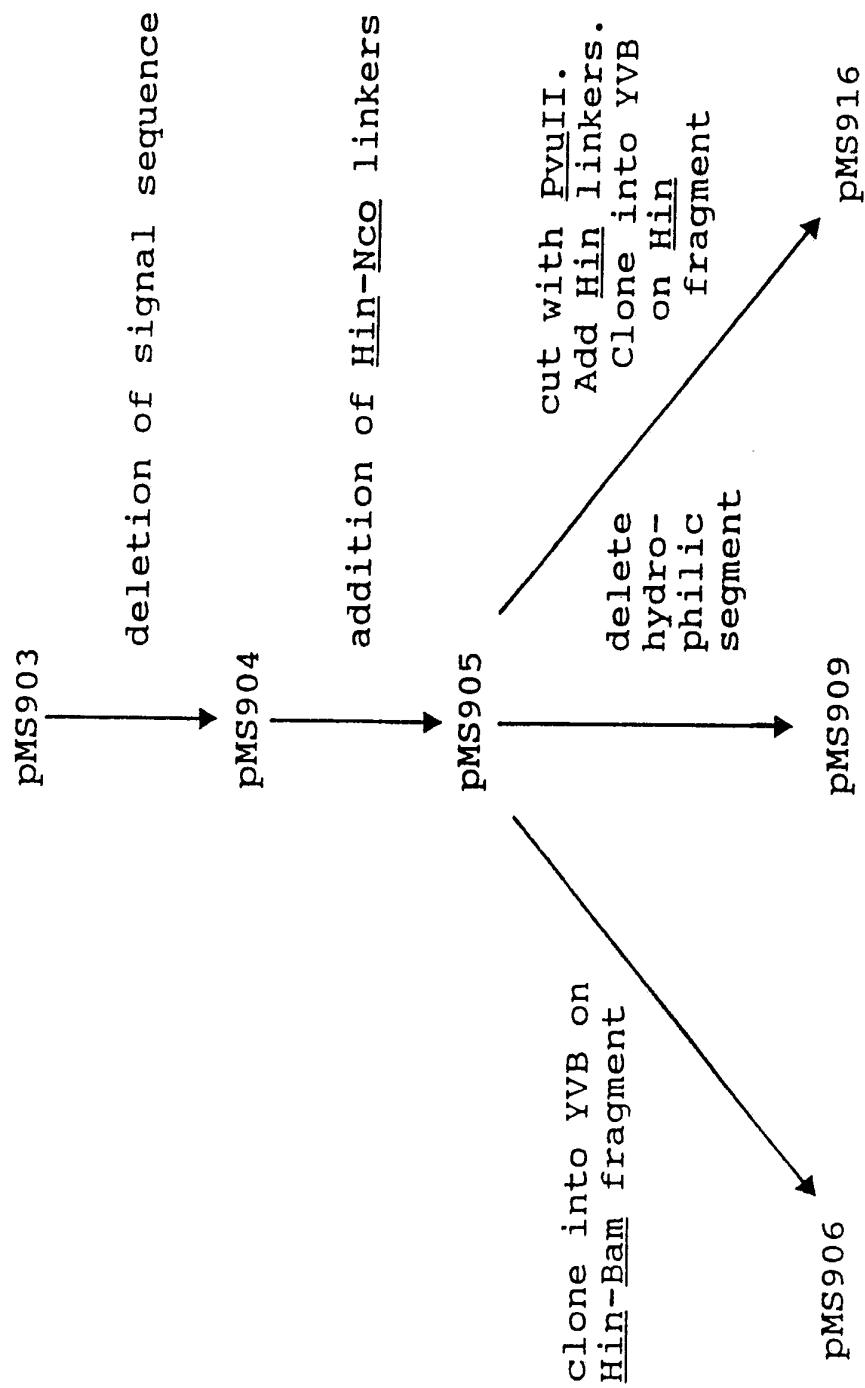
16/22

## FIG. 9A



17/22

## FIG. 9B



18/22

YVA SERIES

pMS908

*FIG. 10*

pMS914



pMS912

YVB SERIES

pMS906



pMS916

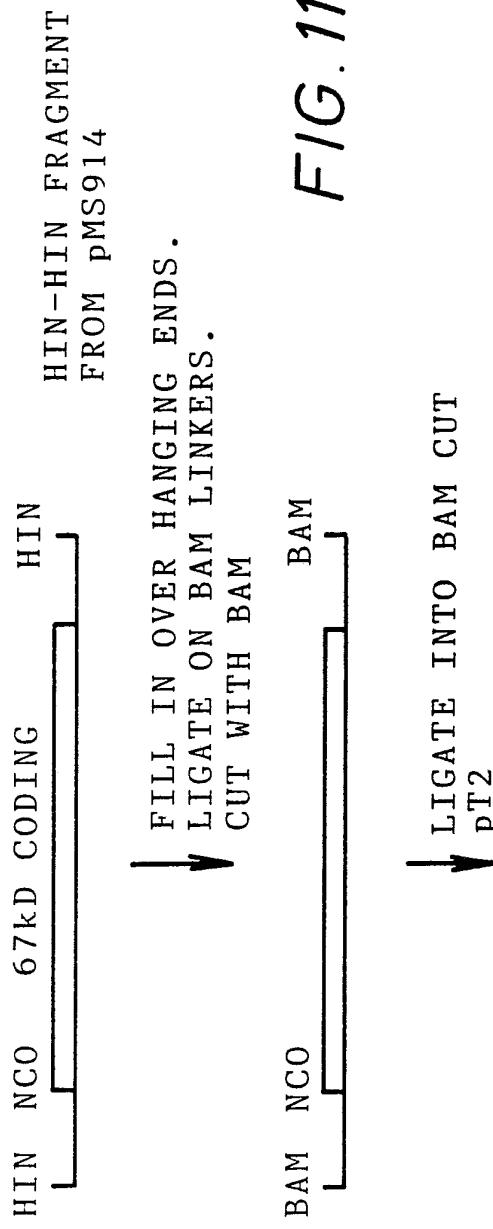


pMS910

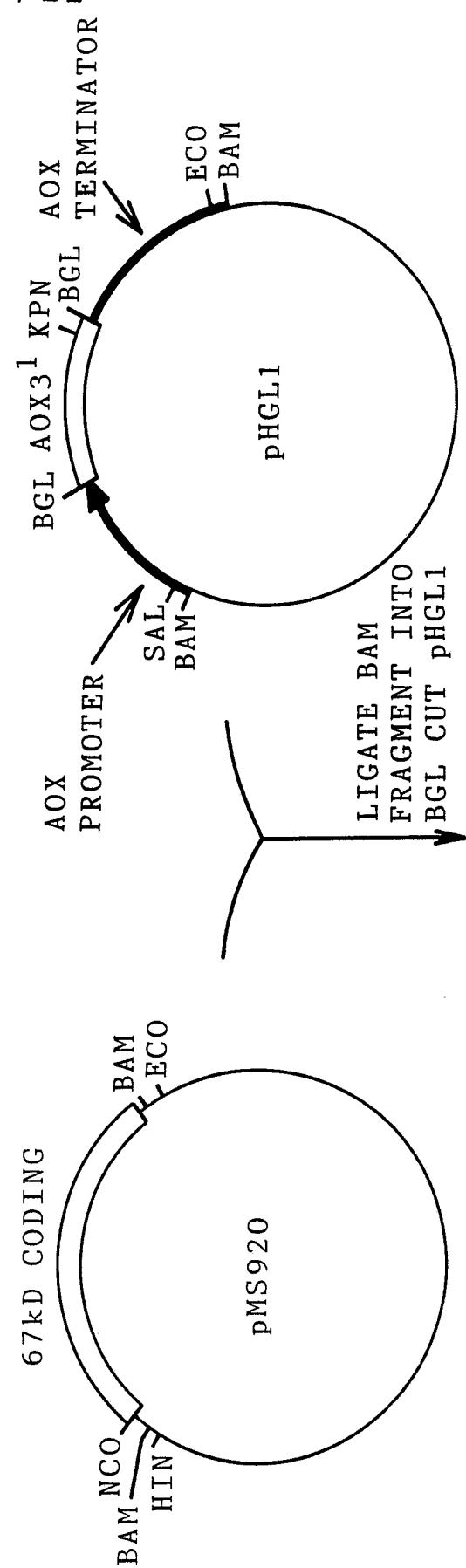


-  COCOA 67kD SIGNAL
-  YEAST  $\alpha$ -FACTOR SIGNAL
-  HYDROPHILIC DOMAIN
-  47kD CODING REGION
-  COCOA TERMINATOR
-  YEAST TERMINATOR

19/22

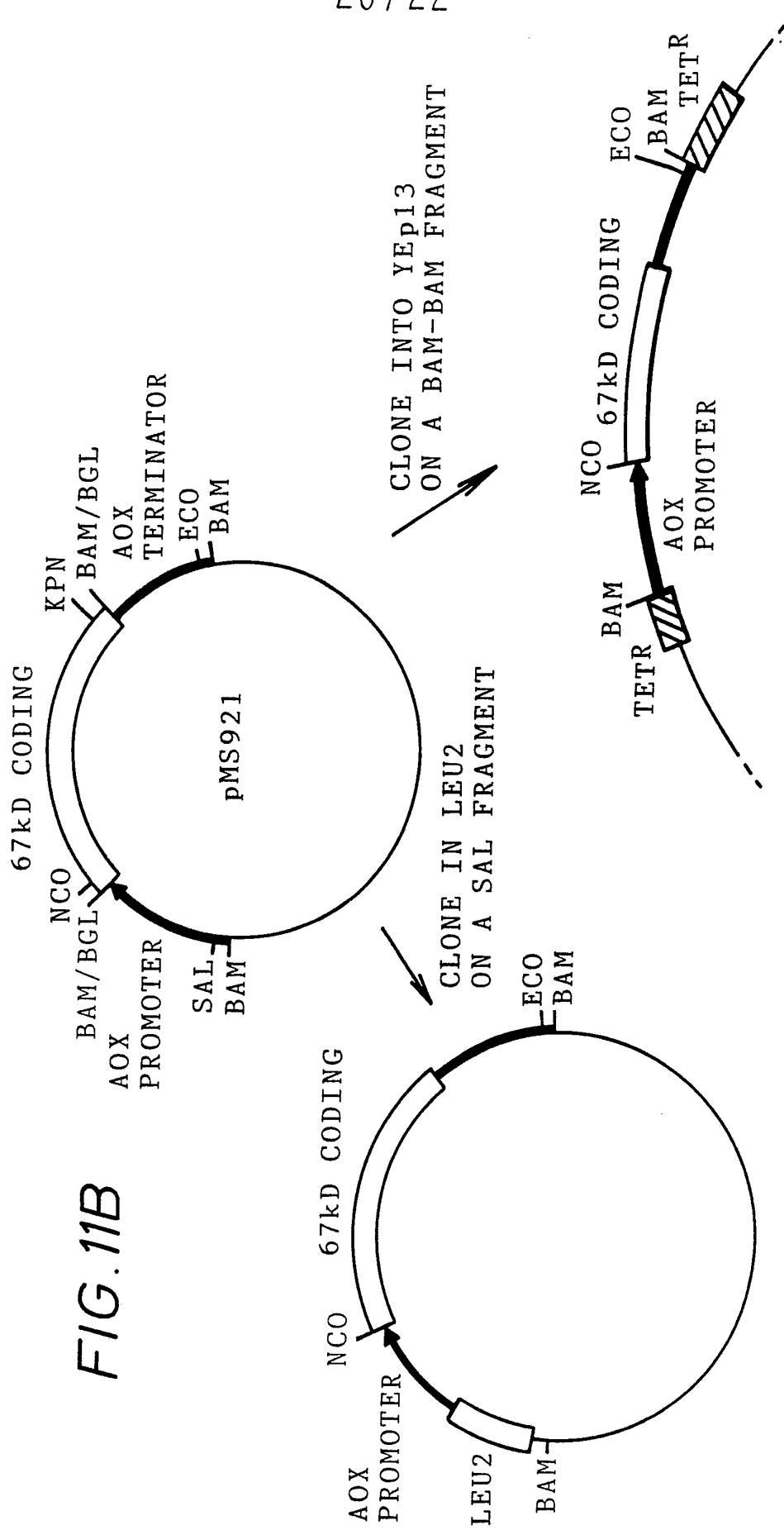


SUBSTITUTE SHEET



20/22

FIG. 11B



21/22

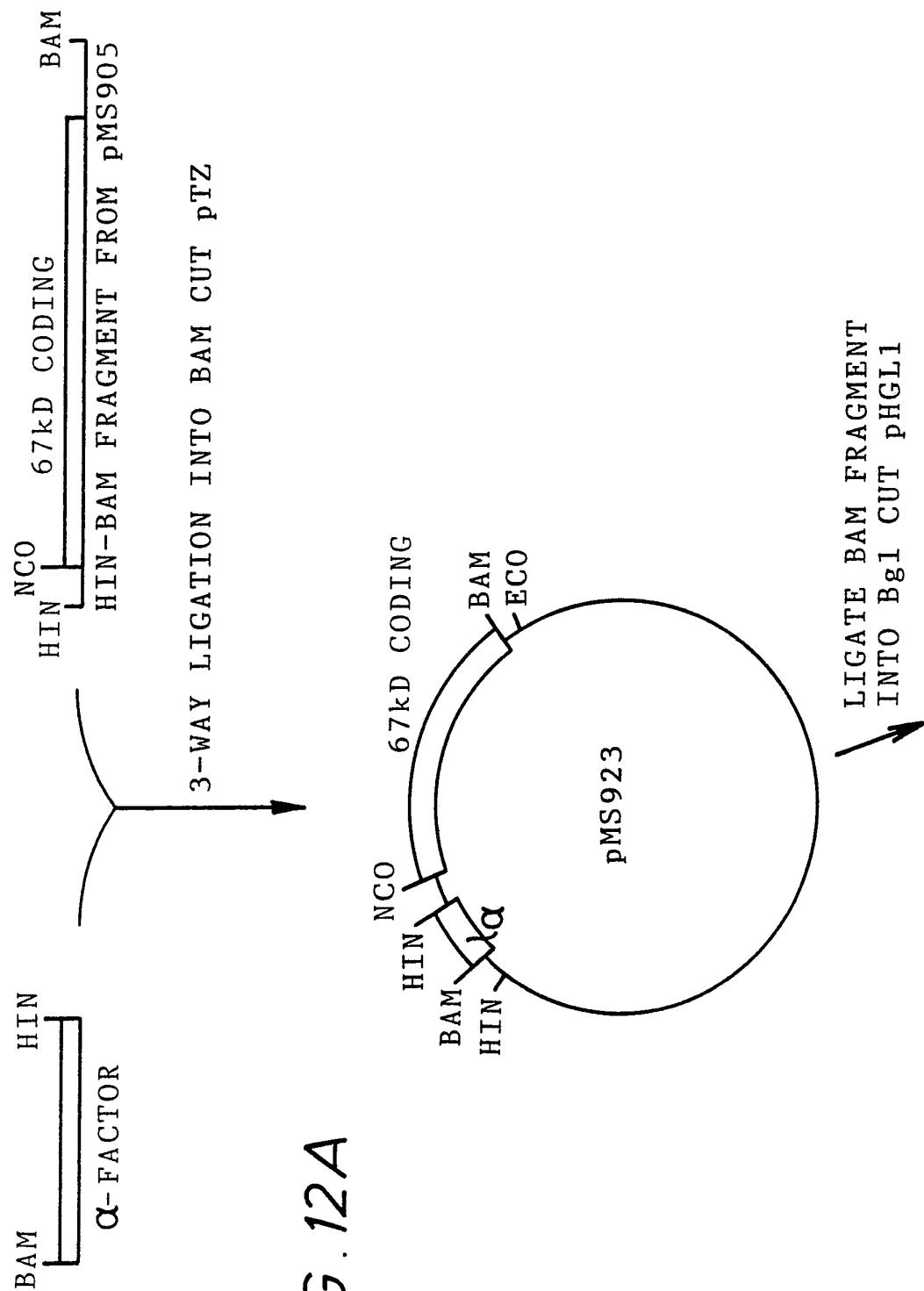
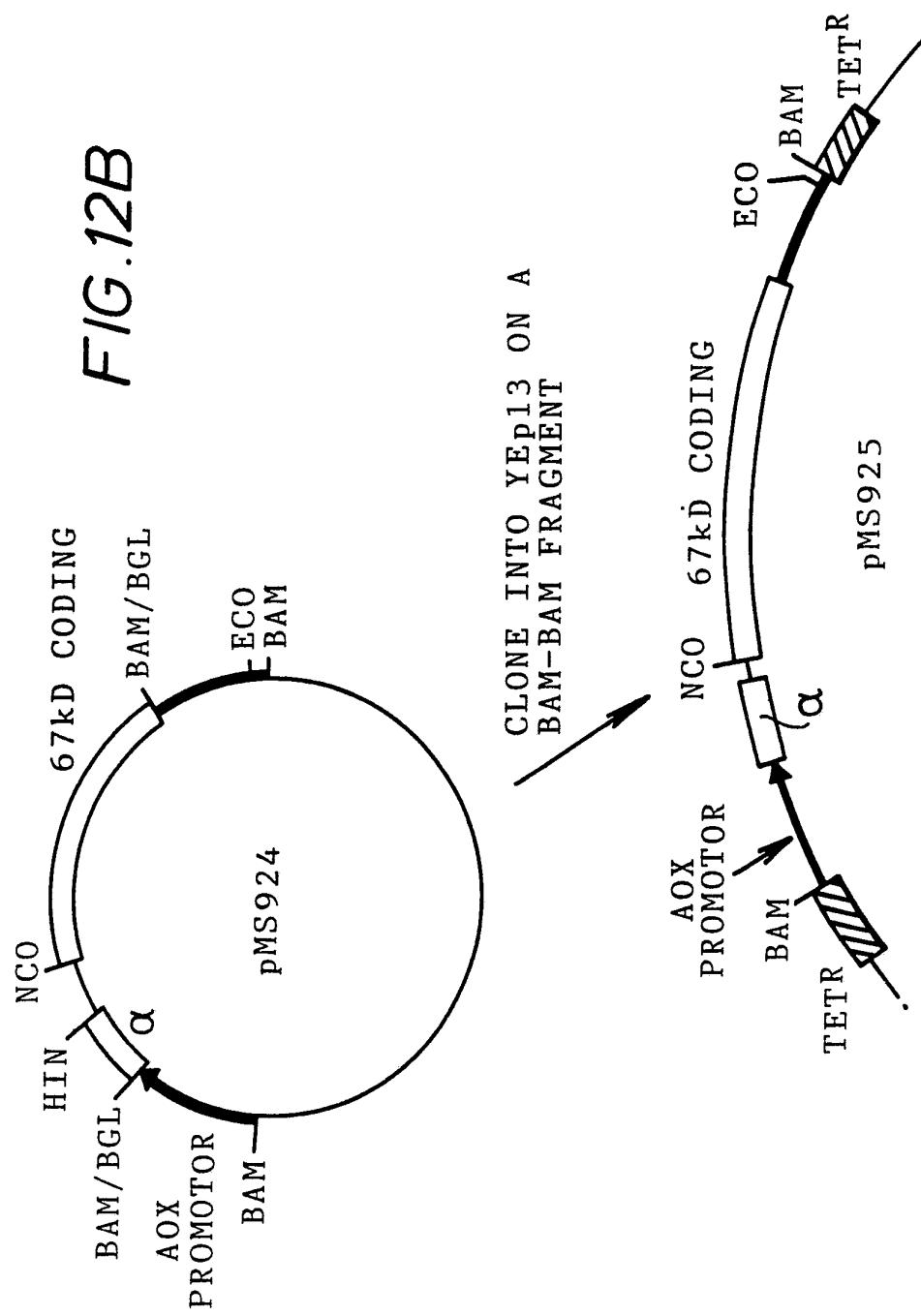


FIG. 12A

22/22



## INTERNATIONAL SEARCH REPORT

International Application No

PCT/GB 91/00914

I. CLASSIFICATION OF SUBJECT MATTER (if several classification symbols apply, indicate all)<sup>6</sup>

According to International Patent Classification (IPC) or to both National Classification and IPC

Int.C1. 5 C12N15/29 ; C07K13/00 ; C12N1/21 ; C12N1/19

## II. FIELDS SEARCHED

Minimum Documentation Searched<sup>7</sup>

Classification System	Classification Symbols
Int.C1. 5	C07K ; C12N

Documentation Searched other than Minimum Documentation  
to the Extent that such Documents are Included in the Fields Searched<sup>8</sup>III. DOCUMENTS CONSIDERED TO BE RELEVANT<sup>9</sup>

Category <sup>10</sup>	Citation of Document, <sup>11</sup> with indication, where appropriate, of the relevant passages <sup>12</sup>	Relevant to Claim No. <sup>13</sup>
X	J. SCI. FOOD AGRIC. vol. 33, 1982, pages 1291 - 1304; BIEHL, B., ET AL: 'Vacuolar storage proteins of cocoa seeds and their degradation during germination and fermentation' see the whole document ---	1-6
X	J. FOOD SCIENCE vol. 50, 1985, pages 946 - 950; FRITZ, P. J., ET AL: 'Cocoa seeds: Changes in protein and polysomal RNA during development' see the whole document ---	1-6 -/-

° Special categories of cited documents :<sup>10</sup>

- "A" document defining the general state of the art which is not considered to be of particular relevance
- "E" earlier document but published on or after the international filing date
- "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- "O" document referring to an oral disclosure, use, exhibition or other means
- "P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

## IV. CERTIFICATION

Date of the Actual Completion of the International Search

Date of Mailing of this International Search Report

5

30 SEPTEMBER 1991

20. 11. 91

International Searching Authority

EUROPEAN PATENT OFFICE

Signature of Authorized Officer

MADDOX A. D.

III. DOCUMENTS CONSIDERED TO BE RELEVANT (CONTINUED FROM THE SECOND SHEET)		
Category °	Citation of Document, with indication, where appropriate, of the relevant passages	Relevant to Claim No.
X	CAFE CACAO THE vol. 34, no. 1, January 1990, pages 23 - 26; PETTIPHER G. L.: 'The extraction and partial purification of cocoa storage proteins' see the whole document ----	1-6
Y	ABSTR. PAP. AM. CHEM. SOC. vol. 188, 1984, BIOL 148 WILSON, M. R., ET.AL.: 'Cocoa theobroma -cacao seed complementary DNA library' see the abstract 148 ----	7-20
Y	PLANT MOL. BIOL. vol. 9, no. 6, 1987, pages 533 - 546; CHLAN C. A., ET. AL.: 'Developmental biochemistry of cottonseed embryogenesis and germination. XIX. Sequences and genomic organization of the alpha-globulin, vicilin, genes of cottonseed.' see figures 3,4 ----	4-11, 14, 15, 17, 20
X	CHEMICAL ABSTRACTS, vol. 109, no. 17, Columbus, Ohio, US; abstract no. 143880s, WATSON, MARTIN D. 'Isolation and expression of a pea vicilin cDNA in the yeast <i>Saccharomyces</i> <i>cerevisiae</i> ' see abstract ----	4-11, 14-16, 19
X	PROC. NATL. ACAD. SCI. U. S. A. vol. 82, no. 2, January 1985, pages 334 - 338; CRAMER J H: 'Expression of phaseolin cDNA genes in yeast under control of natural plant DNA sequences.' see the whole document ----	4-11, 14-16, 19
X	CHEMICAL ABSTRACTS, vol. 106, no. 13, Columbus, Ohio, US; abstract no. 98280, CRAMER, JANE HARRIS 'Signal peptide specificity in posttranslational processing of the plant protein phaseolin in <i>Saccharomyces cerevisiae</i> ' see abstract ----	4-11, 14-16, 19
		-/-

III. DOCUMENTS CONSIDERED TO BE RELEVANT (CONTINUED FROM THE SECOND SHEET)		
Category *	Citation of Document, with indication, where appropriate, of the relevant passages	Relevant to Claim No.
X	PLANT MOL BIOL vol. 11, 1988, pages 683 - 695; HIGGINS T.J.V., ET.AL.: 'The sequence of a pea vicilin gene and its expression in transgenic tobacco plants' see page 684 methods section ----	4-11, 15, 17, 19