Volume 16 Number 16 1988 Nucleic Acids Research

Extranuclear gene expression in yeast: evidence for a plasmid-encoded RNA polymerase of unique structure

Duncan W.Wilson+ and Peter A.Meacock*

Leicester Biocentre, University of Leicester, Leicester, LE1 7RH, UK

Received June 24, 1988; Revised and Accepted July 26, 1988 Accession no. X07946

ABSTRACT

Strains of the yeast *Kluyveromyces lactis* that produce killer-toxin have been found to contain two linear dsDNA plasmids, k_1 (8.9 Kb) and k_2 (13.4 Kb). The four transcribed open reading frames of plasmid k_1 contain no recognisable yeast nuclear expression signals. Moreover, a toxin subunit gene fused with the *lacZ* gene of *Escherichia coli* is not detectably expressed when introduced to *K.lactis* or *Saccharomyces cerevisiae* on a nuclear vector, even when native k_1 and k_2 are present in the cell. This and other evidence is consistent with the hypothesis that k_1 and k_2 reside in an extranuclear location, and do not utilise the nuclear RNA polymerases I, II or III for transcription of their genes. Sequencing of plasmid k_2, which is thought to encode factors necessary for the maintenance or expression of k_1, reveals an open reading frame predicted to encode a 974 amino acid polypeptide with homology to several DNA-directed RNA polymerases. We suggest that this is a component of a novel plasmid-specific extranuclear gene expression system.

INTRODUCTION

Many strains of the budding yeast *Kluyveromyces lactis* contain two linear double-stranded A/T rich DNA plasmids, k_1 (8.9 Kb) and k_2 (13.4 Kb)[1] that are associated with secretion of a multi-subunit protein toxin[2] which inhibits the growth of certain sensitive yeast species. Structurally, the plasmids belong to a class of extrachromosomal linear elements which includes adenovirus[3], bacteriophage phi 29[4] plasmids of *Streptomyces rochei*[5] and the S1/S2 mitochondrial plasmids associated with male sterility in maize[6]. All of these molecules have inverted terminal repeat sequences (202bp and 184bp in the case of k_1 and k_2[7]) and proteins covalently attached to the 5' end of each DNA strand (for k_1 and k_2 terminal proteins see reference 8). These may function as primers in the initiation of terminal DNA replication. In support of this hypothesis one of the four transcribed open reading frames of k_1[9,10] encodes a product with homology to the DNA polymerases of adenovirus and phi 29, and to a protein encoded by an ORF of the maize plasmid S1[11].

Plasmids k_1 and k_2 can be transferred to ρ^o (rho zero) strains of

© IRL Press Limited, Oxford, England.

Saccharomyces cerevisiae (which lack mitochondrial DNA) where they are maintained and express the killer phenotype[12,13]. However, they do not become established in the presence of mitochondrial DNA[14]. These observations, together with the high A/T content of the plasmid DNA, fluorescence staining of S.cerevisiae ρ^{o} derivatives containing the plasmids[13] and fractionation of yeast nuclei and cytoplasm by centrifugation techniques (reference 15, D.W. Wilson and P.A. Meacock unpublished observations) are all consistent with a cytoplasmic location for the plasmids. Plasmid curing experiments have shown that plasmid k_2 can be maintained in the absence of k_1, but plasmid k_1 is unable to exist independently[16] suggesting dependence upon k_2-encoded products. Such factors may be concerned with the replication or segregation of k_1, or for the expression of k_1-encoded genes essential for k_1 maintenance.

A variety of data suggest that these plasmids may utilise a novel system for gene transcription; viz. none of the ORFs of k_1 is preceded by recognisable yeast nuclear promoter elements, although all four are preceded by a motif identical with, or closely related to, the sequence ACT(A/T)AATATATGA. This has been termed the Upstream Conserved Sequence or UCS[9], and transcription is initiated approximately 14 bp downstream of this element (Romanos and Boyd, manuscript submitted). Toxin production cannot be detected from ρ^{o} strains of S.cerevisiae which have been transformed with the coding region of plasmid k_1 cloned into yeast 2 μm (micron)-based vectors (D.W. Wilson and P.A. Meacock, unpublished observations, reference 15). Furthermore, Northern blotting reveals that when k_1 DNA is introduced into K.lactis on nuclear vectors transcription of the k_1 gene ORF2 (which encodes two toxin subunits) is initiated at a number of sites distinct from those used by native linear k_1, and the transcript is prematurely terminated (Romanos and Boyd, manuscript submitted). Thus it appears that the yeast nuclear RNA polymerases I, II and III are unable to recognise and correctly transcribe the genes of these plasmids.

Here we demonstrate that expression of an ORF2/lacZ fusion gene, cloned into a yeast nuclear vector, cannot be detected in cells of K.lactis or S.cerevisiae which contain plasmids k_1 and k_2. This suggests that transcription of ORF2, and probably all plasmid-borne genes, occurs in an extranuclear cellular compartment. The provision of novel cytoplasmic expression factors could be one of the maintenance functions of k_2, and to investigate this we have begun to determine the nucleotide sequence of this plasmid. The predicted product of one of the ORFs encoded by k_2 has homology to two different subunits found within DNA-directed RNA polymerases, whilst

another has homology to a vaccinia virus helicase necessary for specific viral transcription.

MATERIALS AND METHODS

Strains and Media

Kluyveromyces lactis IFO1267 (prototrophic [$k_1^+k_2^+$]) was obtained from the National Collection of Yeast Cultures, Food Research Institute, Colney Lane, Norwich, UK. K.lactis SD11 (K.lactis lac4 trp1 [$k_1^+k_2^+$]) was kindly provided by Prof. C.Hollenberg, and K.lactis ABK802 (prototrophic, [k_1^o,k_2^+]) by Dr. A.Boyd. Saccharomyces cerevisiae SPK103 (a his3-Δ trp1-289 ura3-52 leu2-3 [k_1^+ k_2^+ L^+] ρ^o) was prepared by cytoduction of a ρ^o derivative of strain S150-2B (obtained from J.Hicks, Cold Spring Harbor Laboratory, New York) with JC25K (α ade2-1 his4-15 kar1-1 [k_1^+,k_2^+]) obtained from Dr. M.A.Romanos. Yeast were propagated at 30°C using either YPD medium (1 % yeast extract, 2 % peptone, 2 % glucose) or in minimal medium (0.67 % yeast nitrogen base, 2 % glucose) supplemented with nutrients as required. Plasmid constructions, bacterial assays for β-galactosidase activity, and superinfection for preparation of single stranded DNA made use of E.coli NM522 Δ(lac-proAB) thi^- supE hsdΔ5 [F' proAB $lacI^Q$ ZΔM15][17] (provided by Dr. A.Mileham), maintained on M9 minimal medium[18] supplemented with 0.001 % thiamine, in order to ensure maintenance of the F' episome and susceptibility to M13 amd M13K07 phage infection. M13K07 helper phage was from Pharmacia, Uppsala, Sweden.

Enzymes

The Klenow fragment of DNA polymerase I, exonuclease III, exonuclease VII, and all restriction endonucleases used in this study were purchased from Bethesda Research Laboratories (BRL), Bethesda, Maryland, USA. T4 DNA ligase was purchased from Pharmacia, and Proteinase K from Boehringer (BCL), Mannheim, FRG. Zymolyase 100-T was from the Kirin brewery, Tokyo, Japan. All enzymes were used as recommended by the supplier.

Materials for Sequence Analysis

Deoxynucleotides and dideoxynucleotides were purchased from BCL. Acrylamide and bisacrylamide were "Electran" grade, purchased from BDH chemicals, Poole, UK. Ultrapure enzyme grade urea was from BRL. Sigmacote and N,N,N',N'-Tetramethylethylenediamine (TEMED) were obtained from Sigma, St Louis, Missouri, USA. Sequencing primers were 17 bp "universal primer" supplied by Pharmacia and custom-synthesised 17 bp oligodeoxynucleotides prepared by J.Keyte, Biochemistry department, Leicester University, using an Applied Biosystems DNA synthesiser. [α-^{35}S]dATP, at 10 μCi μl^{-1} and 650 Ci

mMol-1 were obtained from Amersham International, UK.

Preparation of Plasmid k_2 DNA

K.lactis ABK802 was grown in 1 litre of YPD broth to a density of $3x10^7$ cells ml^{-1}, cells were pelleted and washed once in water, then resuspended in 50 ml of SED buffer (1.2 M Sorbitol, 20 mM EDTA, 50 mM Dithiothreitol) and incubated at 37°C for 20 mins. To the suspension was added 1 ml of 10 mg ml^{-1} Zymolyase 100-T in SED buffer and incubation continued until spheroplasting was complete. Sphaeroplasts were pelleted and resuspended in 20 ml 10 mM EDTA, 50 mM Tris.HCl pH 8.0, then Sodium N-lauryl Sarcosinate added to a final concentration of 2 %. To the lysate was added 1 ml of 10 mg ml^{-1} Proteinase K, the mixture incubated at 37°C for 1 h, then the lysate chilled on ice, mixed with 5 ml of 5 M NaCl and left on ice for 1 h. After centrifugation at 18000 rpm for 30 mins at 4°C in a Sorvall SS34 rotor the supernatant was extracted three times with an equal volume of phenol, then twice with chloroform. Extracted supernatant was mixed with 2 ml 3 M sodium acetate pH 5.5 and 50 ml absolute ethanol, placed at -80°C for 30 mins, nucleic acid pelleted and the pellet rinsed with 10 ml of 70 % ethanol. The pellet was briefly dried under vacuum, resuspended in 5 ml TE buffer (1 mM EDTA, 10 mM Tris.HCl pH 7.6) then nucleic acid fractionated according to size in a 10 %-40 % sucrose gradient, containing 1 M NaCl, 5 mM EDTA, 20 mM Tris.HCl pH 8.0, by centrifugation at 26000 rpm for 24 h at 26°C in a Sorvall AH627 rotor. Fractions containing plasmid k_2 were identified by agarose gel electrophoresis, pooled and diluted with 2 volumes of water. DNA of plasmid k_2 was precipitated for 1 h at -80°C after the addition of 1/10th volume of 3 M sodium acetate pH 5.5 and 2.5 volumes of absolute ethanol, then pelleted, the pellet rinsed with 0.5 ml of 70 % ethanol, briefly vacuum dried and resuspended in 200 µl TE buffer.

Quantitative Assay for β-Galactosidase Activity

Yeast transformants were grown to between $1x10^7$ and $5x10^7$ cells ml^{-1} in 25 ml minimal medium, imposing selection for presence of plasmid-borne markers. Optical density of the cultures at 600 nm was determined, and cultures placed on ice for 20 mins. An aliquot was removed and plated onto complete media in order to determine, by subsequent replica plating onto complete and selective media, the percentage of cells within the population which contained plasmids. Chilled cells were pelleted and resuspended in 1 ml 50 mM KH_2PO_4. To 0.4 ml of the suspension was added an equal volume of 60 mM Na_2HPO_4, 40 mM NaH_2PO_4, 10 mM KCl, 1 mM $MgSO_4$, 50 mM β-mercaptoethanol pH 7.0, then 10 µl 1 % SDS and Phenylmethylsulphonyl fluoride (PMSF) to a

final concentration of 0.5 mM. The mixture was vortexed for 5 secs, incubated at 28°C for 10 mins then 30 μl chloroform added, the mixture vortexed again and incubation continued at 28°C for 10 mins. A 200 μl aliquot of 4 mg ml^{-1} ONPG (freshly dissolved in 100 mM KH_2PO_4) was warmed to 28°C then mixed with the treated cells, a timer started and incubation continued at 28°C. At the first appearance of a yellow colour within the mixture the reaction was stopped, by addition of 0.5 ml 1 M Na_2CO_3, and time of incubation recorded. Cells were removed from the mixture and optical density of the supernatant recorded at a wavelength of 420 nm. Units of β-galactosidase activity were expressed as OD_{420} min^{-1} $OD_{600}{}^{-1}$, then multiplied by a factor of 1000. When the percentage of plasmid-containing yeast cells was known, assay values were corrected to that expected if 100 % of cells had contained plasmid at the time of assay. Bacterial transformants were assayed in exactly the same manner, but using 10 ml cell cultures. Since bacterial transformants were grown under antibiotic selection, all cells contained plasmids.

Generation of Nested Overlapping Deletions From Cloned k_2 DNA

DNA (5 μg) of a plasmid containing the 7.7 Kb Bam HI-Xho I fragment of k_2 cloned within $pEMBL19^+$ was linearised by digestion with the restriction endonucleases Bam HI and Kpn I. The mixture was extracted with equal volumes of phenol and chloroform, and ethanol precipitated. The fragment was resuspended in TE buffer to a concentration of 1 μg DNA μl^{-1}.

For deletion of the 3'recessed strand, 5 μl of linearised plasmid fragment were mixed with 27 μl water, 4 μl 10xExoIII buffer (10 mM $MgCl_2$, 660 mM Tris.HCl pH 8.0) and brought to reaction temperature by incubating at 37°C for 5 mins. Then 4 μl of 10 units μl^{-1} exonuclease III were added. A 2 μl sample was immediately removed and mixed with 2 μl 10xExoVII buffer (300 mM KCl, 100 mM EDTA, 100 mM Tris.HCl pH 7.5) on ice. Samples were similarly taken at 30 second intervals for 5 mins, at which time 20 μl of prewarmed 1xExoIII buffer containing 40 units Exonuclease III were added to the incubation. Ten subsequent samples of 4 μl each were removed at 30 s intervals into 2 μl of ice-cold 10xExoVII buffer. The twenty samples were each diluted to 19 μl with water and mixed with 1 μl exonuclease VII diluted to 1.5 units μl^{-1} in 1xExoVII buffer. The reactions were incubated at 37°C for 2 h to remove single-stranded regions of DNA. Samples were phenol extracted, chloroform extracted and ethanol-precipitated, resulting pellets were drained, vacuum dried and resuspended in 15 μl of 10 mM $MgCl_2$, 10 mM Dithiothreitol, 1mM ATP, 100 μg ml^{-1} Bovine Serum Albumin, 50 mM Tris.HCl pH 7.4, 50 μm dATP, 50 μm dTTP, 50 μm dCTP, 50 μm dGTP, 3 units T4

DNA Ligase, 0.5 unit Klenow fragment. Ligation was allowed to proceed overnight at 30°C. Twenty aliquots of competent E.coli NM522 were transformed with 5 μl of each ligation diluted by addition of 100 μl of 0.1 M $CaCl_2$. Transformants were picked and plasmid DNA prepared from 1.5 ml cultures by standard methods. Degree of deletion within each plasmid was determined by agarose gel electrophoresis of restriction digestion products.

Superinfection of pEMBL-containing E.coli

A 1 ml aliquot of M9 media, supplemented with 0.001% thiamine, 50 μg ml^{-1} ampicillin, 0.2 % glucose, was inoculated with 10 μl of stationary phase culture of E.coli NM522 transformed with a deletion-derivative of a pEMBL/k_2 clone. The culture was grown at 37°C overnight, and a 20 μl aliquot used to inoculate 2 ml of 2xYT broth, 0.001 % thiamine, 150 μg ml^{-1} ampicillin prewarmed to 37°C. This culture was shaken at 37°C until reaching an optical density of between 0.5 and 1.0 at a wavelength of 660 nm. A 1 ml aliquot of this culture was infected with M13K07 helper-phage at a multiplicity of 10 M13K07 plaque forming units/bacterial cell. The cell/phage mixture was shaken at 37°C for 1 h, then 400 μl added to 10 ml prewarmed 2xYT broth, containing 0.001 % thiamine, 150 μg ml^{-1} ampicillin, 70 μg ml^{-1} kanamycin. The culture was grown overnight at 37°C, cells pelleted and cell-free supernatants stored until required for template preparation.

DNA Sequencing

Sequencing was by the dideoxy method[19] but using [α-^{35}S]dATP. All reactions were carried out at 50°C. Dideoxy-terminated fragments were electrophoresed upon a 6 % polyacrylamide/urea sequencing gel, in 1xTBE buffer (5.5 g l^{-1} boric acid, 0.93 g l^{-1} EDTA, 10.8 g l^{-1} Tris base). Electrophoresis was for 2.5 h or up to 7.5 h and gels dried under vacuum (without prior fixing) at 80°C for 90 mins then exposed to Kodak XAR-5 film for at least 12 h.

RESULTS AND DISCUSSION

Killer Plasmid Genes Are Not Expressed Within The Yeast Nucleus

Previous studies (reference 15. Wilson and Meacock, unpublished observations) have shown that k_1 toxin genes, cloned in circular autonomously replicating vectors, do not confer a killer phenotype upon their host. However, none of these experiments could be performed with yeast strains that also contained both k_1 and k_2 as native linear plasmids. Thus plasmid-encoded factors necessary for k_1 toxin gene expression may have been absent. In order to test, quantitatively, whether linear plasmid-encoded genes contained within nuclear vectors can be expressed when the endogenous plasmids k_1 and

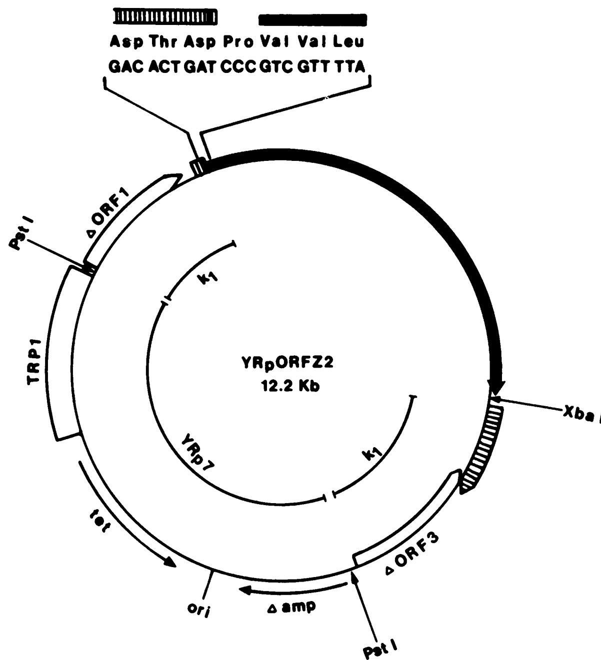

Fig. 1. Structure of the plasmid YRpORFZ2. Regions of the plasmid which are derived from plasmids k_1 and YRp7 are indicated. The fusion junction between the amino-terminal coding region of ORF2 and lacZ is shown. Solid box: lacZ coding region. Hatched box: Remainder of ORF2 coding region. Open box: k_1 genes and TRP1 gene of YRp7. Curved arrows: Bacterial genes. Arrow heads indicate direction of transcription of genes. Δ (delta): Partial gene. Bacterial origin of replication, ori, is also shown.

k_2 are also present in the cell, we assayed expression of a cloned k_1 gene using a gene fusion.

A translational fusion was made between the amino-terminal 36 codons of ORF2 and the lacZ gene of Escherichia coli. As the predicted lacZ-fusion product retains the ORF2 signal sequence it may become membrane associated. However analogous gene fusions between lacZ and the GAL2 permease of S.cerevisiae have been successfully assayed in yeast[20]. The ORF2-lacZ gene fusion retained 1.8 Kb of k_1 DNA normally upstream of ORF2 and 1.5 Kb of k_1 DNA normally downstream of the ORF2 carboxy terminus. The fusion gene was transferred to the S.cerevisiae vector YRp7[21], to form plasmid YRpORFZ2. See Figure 1. During construction of plasmid YRpORFZ2, the ARS1 sequence, necessary for autonomous replication of YRp7 in S.cerevisiae, was removed. However the A/T rich k_1-derived DNA fortuitously provides elements[22] enabling

Table 1. Expression of an ORF2-lacZ fusion gene in yeast and E.coli

Plasmid	E.coli NM522	S.cerevisiae SPK103	K.lactis SD11	K.lactis IFO1267
pUC19	2.57	—	—	—
pUCORFZ2	3.14	—	—	—
pLG669Z	11.4	776.88	—	—
pPA13	0.04	39.20	—	—
YRpORFZ2	0.43	0.71	0.58	—
YRp7	0.01	0.49	—	—
KRp2	—	—	0.52	—
—	—	—	—	8.35

Units of β-galactosidase activity expressed by various plasmids in strains of yeast and E.coli. For calculation of activity values, and genotypes of strains, see materials and methods. SD11 carries an inactivating lesion within the endogenous K.lactis β-galactosidase, IFO1267 does not, but was grown under conditions which would not induce expression of the enzyme beyond basal levels. Plasmid pUCORFZ2 carries the ORF2-lacZ fusion expressed from the lac promoter of pUC19. YRpORFZ2 carries the ORF2-lacZ fusion cloned into the yeast high copy number vector YRp7. pLG669Z expresses the lacZ gene from the S.cerevisiae CYC1 nuclear promoter. pPA13 expresses lacZ from the cauliflower mosaic virus 19S promoter. YRp7 & KRp2 are, respectively, S.cerevisiae and K.lactis autonomously replicating plasmids which carry no lacZ gene.

autonomous replication of the plasmid in both S.cerevisiae and K.lactis.

Plasmid YRpORFZ2 was transformed into strains of K.lactis and S.cerevisiae which harboured linear plasmids k_1 and k_2, using a whole-cell transformation procedure[23]. Transformed cells were permeabilised with chloroform, and β-galactosidase activity measured by spectrophotometric assay of the rate of hydrolysis of o-nitrophenyl-β (beta)-D-galactoside (ONPG). No enzyme activity could be detected in either yeast (Table 1). Detection of activity in E.coli strain NM522, when transformed with YRpORFZ2 or with pUCORFZ2 (which expresses the gene fusion from the lac promoter) confirmed that, when expressed, the fusion protein did retain normal β-galactosidase activity. Also, other controls indicated that, under these conditions, it was

```
                 UCS
           ┌──────────────┐
ORF 1      ACTATAATA-TATGA................AAAATG
ORF 2      ACT-TAATA-TATGA..................AAAATG
ORF 3      ACT-AAATA-TATGA..................AAAATG
ORF 4        TAAAATAATCTGA.................................AAAATG
ORF 974            TA-TATGA................AAAATG
ORF 579            TA-TCTGA................AAAATG
ORF 336           AAAATATGA.....................................ATAATG
ORF 158           ATATTTTGA..........AATATG
ORF 132            TA-TGTGA.................AAAATG
ORF 112            TA-TTTGA...................................ATTATG
ORF 103         AAATA-TATGA................AAAATG
Consensus          TAATATGA
                   A T T
                     - C
                       G
```

Fig. 2. Comparison of the upstream conserved sequence (UCS) found before each of the ORFs of plasmid k_1 with upstream regions of plasmid k_2 ORFs identified during this study. The predicted initiation codon (ATG) and three preceding nucleotides are shown in each case. Each intervening nucleotide is represented by a dot (.). Note the requirement for an adenine residue at the -3 position, and preference for an AAAATG initiation context. A consensus UCS is shown.

possible to detect β-galactosidase activity in yeast when expressed at high (plasmid pLG669Z) or low (plasmid pPA13 and non-induced IFO1267) levels.

We suggest that expression of at least ORF2, and probably all of the k_1 genes, therefore requires factors unavailable within the nuclear compartment, even when native linear k_1 and k_2 are also present in the cell.

Plasmid k_2 Encodes a Product With Homology to Two RNA Polymerase Subunits

In order to investigate the novel extranuclear expression system which may be encoded by k_2 we are determining the complete nucleotide sequence of this plasmid. DNA of plasmid k_2 was prepared from *K. lactis* ABK802 [k_1^o,k_2^+] and digested to completion with the restriction endonucleases *Bam HI* and *Xho I*. This yielded two *Bam HI*-*Xho I* fragments, of 7.7 Kb and 5.5 Kb, which after modification of terminal restriction sites were cloned in both possible orientations into the sequencing vector pEMBL19$^+$. An overlapping series of deletion-derivatives were prepared from the 7.7 Kb fragment, in both directions, using the exonuclease III/exonuclease VII deletion method[24]. Deletion-derived clones were sequenced, and sequence data assembled manually using the University of Wisconsin genetics computer group (UWGCG) programs[25].

Analysis of 7688 bp of continuous sequence data obtained from plasmid k_2 reveals seven ORFs, with potential to encode products of 974, 579, 336, 158,

MetAspTyrGlyGlnIleGluIleTyrAsnAspTyrPheArgAsnValAspLysLysThrIleLysTyrAsnProMetGluCysIleGluTyrLysLeuSer
AGCTATATGATCGTATTAGATATACCTAAAATGGATTATGGACAAATAGAAATTTATAACGATTATTTTAGAAATGTTGATAAGAAAACTATAAAATACAATCCCATGGAGTGCATAGAATATAAACTAAGT

TyrCysAlaIleTyrAsnAspArgGluLeuProIleMetIleGlySerSerLeuAspIleSerLysGluAsnProLeuAsnLeuLysGlyTyrPheIleIleAspGlyIleCysLysSerValAsnAsnIle
TACTGTGCTATATACAATGACAGAGAACTTCCTATAATGATTGGTAGTTCACTGGATATTAGCAAAGAAAATCCTTTAAATTTAAAAGGTTATTTCATAATTGATGGTATATGTAAATCAGTTAATAATATA

100
LysIleLysGluLysIleHisPheSerLysAspArgAlaTyrLeuSerAspAspSerIleIleGlnIleLysAspMetPheAsnMetThrIleLysAsnAsnLysAsnThrLeuLysTrpAsnLeuProLeu
AAAATAAAAGAAAAAATACATTTTTCGAAAGATAGAGCTTATCTTAGCGATGATTCTATAATACAAATAAAAGACATGTTTAACATGACAATAAAAAATAATAAAAATACATTAAAATGGAATTTACCATTA

AsnTrpSerAspIleIleLysTyrSerAspHisLysLysGluLeuGluAsnHisLeuLysIleIleAspSerPheAsnLysLysAlaGluLysProIleLysAsnGluIleAspLeuIleSerLeuCysTyr
AATTGGTCAGATATAATTAAATATAGCGACCATAAAAAGGAATTAGAGAATCATTTAAAAATAATAGATTCTTTTAATAAAAAAGCGGAAAAACCTATTAAAAATGAGATAGATTTAATATCATTATGTTAC

200
MetPheGluCysTrpLeuGlyLeuArgGluGluProArgLeuTyrTyrArgLeuValThrProGlyGluLeuIleHisAspIleIleLysGluLysArgSerValIleLysCysPheArgThrAsnThrTrp
ATGTTCGAATGCTGGCTAGGTCTGAGAGAAGAACCTAGATTGTATTATAGATTAGTTACTCCTGGAGAACTGATACATGACATTATAAAAGAAAAAAGAAGTGTAATAAAATGTTTTAGAACAAATACTTGG

ThrValLysAsnIleArgAsnValPheCysValSerGluAspMetLysHisTyrAsnLysIleGlyAspLeuGluSerIleArgArgIleThrIleProSerSerArgSerAsnAlaProMetLysGluArg
ACAGTTAAAAATATAAGAAATGTTTTCTGTGTTAGTGAAGACATGAAACACTATAACAAAATAGGAGATTTAGAATCTATAAGAAGAATAACAATACCATCTTCTAGATCTAATGCACCTATGAAAGAAAGG

ArgValGluAspAsnAspLysTyrLysIleCysProIleGlnThrSerAspGlySerLeuCysGlyThrIleThrTyrLeuCysLysAspAlaLysIleThrThrLysIleAspThrIleLysValGluLys
AGAGTAGAAGACAATGATAAATATAAAATATGTCCAATACAAACATCCGATGGTTCTTTATGCGGTACTATTACATATTTATGTAAAGATGCAAAAATAACTACAAAAATTGATACTATAAAAGTTGAAAAA

300
AlaAspAsnGlyLeuHisThrPhePheAsnAsnThrTyrLeuGlyLysValLysValLysThrLysGluSerAspIleTyrCysIleAspLysIleCysTyrIlePheGlyAsnIleGlyArgIleIleLys
GCAGATAATGGATTGCATACATTTTTCAATAATACCTATTTGGGAAAAGTAAAAGTAAAAACTAAAGAATCTGATATATACTGTATTGATAAAATATGTTACATATTTGGTAACATAGGAAGGATAATTAAA

GlyAsnTyrLeuCysSerTyrThrGlyGluLeuLeuAspTyrLysAsnTyrAsnProProIleArgSerMetPheValCysAsnMetIleLysGlnAlaIleThrGlyAspSerValTyrSerGlyAspIle
GGAAATTATTTATGTAGTTATACAGGAGAATTATTAGATTATAAAAACTATAATCCTCCAATTAGGAGTATGTTTGTATGTAACATGATAAAACAAGCTATAACAGGAGATTCTGTTTATTCTGGTGATATA

400
LeuHisAsnThrLysSerLeuIleAsnGlyAspLeuGlyHisLysLeuGluIleAlaIleMetProTrpLysLysPheAsnIleGluAspAlaIleValIleSerAsnLysValSerLysLeuPheArgSer
TTGCACAATACTAAATCTTTAATTAATGGAGATTTAGGACATAAACTAGAAATAGCTATAATGCCGTGGAAAAAATTTAATATAGAAGATGCTATAGTTATAAGTAATAAAGTATCTAAATTATTTAGATCT

LysArgThrIleIleHisArgGluAsnAspValLysIleLeuAsnValTyrValAsnIleAsnGlnAsnIleThrLysGlyAspLeuLeuTyrArgThrTyrAspProMetGlnValLysThrIleAsnPhe
AAGAGAACAATAATACATAGAGAAAATGATGTAAAAATATTAAATGTTTATGTAAATATAAATCAAAATATAACTAAAGGAGATCTGTTGTATAGAACTTATGATCCAATGCAAGTGAAAACCATTAATTTT

500
ValTyrAlaGluHisAspGlyLysValLysGluIleIleArgGluGluLysTyrLeuLysLeuIleLeuIleLysGluArgAspLeuGluValGlyAspLysMetSerSerMetHisGlyGlnLysGlyIle
GTTTATGCAGAACATGATGGAAAGGTAAAAGAAATAATAAGAGAAGAAAAATATTTAAAATTAATATTAATAAAAGAAAGAGATCTAGAAGTAGGAGATAAAATGAGTAGTATGCATGGACAAAAAGGTATA

ValSerLeuIleGluAsnAsnMetProTyrTyrTyrLysAsnGlyGluLysValGluIleAspLeuIleIleAsnProHisAlaPheProSerArgMetThrLeuGlyGlnIleLysGluMetGlyGluLys
GTATCACTTATAGAAAACAATATGCCATATTATTACAAAAATGGAGAAAAGGTAGAAATAGATTTAATAATAAATCCTCACGCATTCCCTTCTAGAATGACTCTAGGACAAATAAAAGAAATGGGTGAGAAA

600
GluGluTyrValTyrIleGluAspThrLysLeuGluAsnLysIleIleValGlyLysCysLeuTyrLeuAlaLeuArgHisGlnValAspAspLysValGlnPheArgAsnGlyGlyAsnIleAspIleVal
GAAGAATATGTATACATAGAAGATACAAAATTAGAAAATAAAATAATAGTAGGTAAGTGTTTATATCTAGCTTTAAGACATCAGGTAGATGATAAAGTCCAATTTAGAAATGGAGGAAACATAGATATAGTT

ThrLysGlnProValSerGlyArgLysArgSerGlyGlyLeuArgPheGlyGlnMetGluArgAspIleLeuIleGlyLeuGlyAlaTrpAsnThrIleLysGluIleTrpSerIleAspArgSerLysIle
ACCAAACAGCCAGTTTCTGGTAGGAAAAGAAGTGGAGGTTTAAGATTTGGACAGATGGAAAGAGATATACTTATAGGTCTTGGAGCATGGAATACTATAAAAGAAATATGGTCCATAGATAGAAGTAAAATA

LysIleAspSerLysThrGlyArgIleAsnTrpIleArgTyrAspLysGluMetGluIleGlyGlnTyrPheLysIleCysLeuSerTyrMetArgSerLeuGlyArgAspIleLeuIleLysAsnAspLys
AAAATAGATAGTAAAACAGGAAGAATAAATTGGATAAGGTATGATAAAGAAATGGAAATAGGACAATACTTCAAAATATGTTTAAGTTATATGAGATCTTTAGGAAGAGATATATTAATAAAAAATGATAAG

700
TyrSerIleValGluPheAspAsnSerTyrLeuProLysThrAspThrMetLysPheGlyAspLeuAspValLeuAspLeuArgIleTyrLysGlyIleValMetLeuProLeuCysLeuArgSerThrTyr
TATTCCATAGTAGAATTCGATAATTCATACCTACCCAAAACAGATACAATGAAATTTGGAGATTTAGATGTTTTAGATTTAAGAATATATAAAGGAATAGTTATGTTACCATTATGTCTAAGATCTACATAT

LeuAsnLysLeuTyrValAspArgLysTyrSerGluAlaGluLysGluValThrLysLeuLeuLysSerLysAsnGlyAlaTyrHisThrLeuValGluGlyHisArgValAspArgCysIleArgSerVal
TTAAATAAATTATATGTAGATAGAAAATATTCAGAAGCTGAAAAGGAAGTAACTAAACTACTTAAAAGTAAAAATGGGGCATATCATACTTTAGTAGAGGGACATAGAGTAGATAGATGTATAAGATCTGTT

800
IleValProAspProThrLeuAspIleAspThrIleLysIleProPheGlyAlaAsnIleGlyCysGluTyrGlyLeuLeuAsnArgGlnProSerLeuAsnValAspSerIleLysLeuValLysLeuLys
ATTGTTCCAGATCCTACTCTAGATATAGACACAATTAAAATACCATTTGGAGCAAATATAGGTTGTGAATATGGACTTCTTAATAGACAACCATCTTTAAATGTAGATTCTATTAAACTAGTAAAACTAAAA

GlnGlySerAsnLysThrIleAlaIleAsnProLeuLeuCysGlnSerPheAsnAlaAspPheAspGlyAspGluMetAsnIleTyrGlyIleArgAsnLysGluSerIleGluGluMetLysIleLeuAla
CAGGGGTCCAATAAAACTATAGCTATAAATCCGTTGTTGTGTCAATCATTTAATGCAGATTTTGATGGAGATGAAATGAATATATACGGTATAAGAAATAAAGAAAGTATAGAAGAAATGAAAATACTTGCA

900
LysValIleGluAsnLysThrGlnAspTyrIleLeuAspLysLysAspLeuThrAlaAsnLysLysGlyIleLeuGluMetIleGluLysGlySerLysGlyLysMetPheAsnCysGluHisMetPheLys
AAAGTAATAGAAAATAAAACTCAAGATTATATATTAGATAAGAAAGATTTAACTGCTAACAAAAAAGGAATACTAGAAATGATAGAAAAAGGATCAAAAGGAAAAATGTTTAATTGTGAACATATGTTTAAA

GluIleGlyLysValThrIleAsnLysLysGluIleAsnIleLysGlyCysTyrAsnAsnSerIleMetAsnGluGluTrpTyrGluMetCysLysValAlaArgGluAsnAlaAlaSerIleSerIleAsn
GAAATAGGTAAAGTAACAATAAATAAAAAGGAAATAAATATAAAAGGATGTTATAATAATTCCATCATGAATGAAGAATGGTATGAAATGTGTAAAGTAGCAAGAGAAAATGCTGCTTCTATAAGTATAAAT

ThrProIleThrGlyTyrLeuGluAsnIleCysAsnGluMetTyrLeu
ACTCCAATAACAGGTTATTTAGAAAATATATGTAATGAAATGTACTTATAA

Fig. 3. Sequence of the ORF 974 coding region aligned with that of its predicted product, showing upstream DNA. The ORF was translated using the universal genetic code.

132, 112 and 103 amino acids. All seven ORFs exhibit the same codon usage as the ORFs of plasmid k_1, all are preceded by UCS-like elements and the initiation codon of all ORFs is in a similar context (figure 2). We suggest that this indicates a common expression mechanism for the genes of plasmids k_1 and k_2. Alignment of k_1 and k_2 UCS elements leads to definition of a new minimal consensus UCS; (A/T)A(A/T/-)TNTGA, where - indicates the possible absence of a base, and N may be any base.

Table 2. Codon Usage of ORF 974 and k_1 ORFs

aa	%codon usage	k_1 ORFs	ORF 974	aa	%codon usage	k_1 ORFs	ORF 974
Ala	GCT	53	48	Leu	TTA	64	57
	GCC	8	0		TTG	9	9
	GCA	37	48		CTT	13	12
	GCG	2	4		CTC	1	0
					CTA	9	17
Arg	CGT	4	0		CTG	4	5
	CGC	1	0				
	CGA	5	0	Lys	AAA	85	91
	CGG	0	0		AAG	15	9
	AGA	88	88				
	AGG	2	12	Met	ATG	100	100
Asn	AAT	85	88	Phe	TTT	83	73
	AAC	15	12		TTC	17	27
Asp	GAT	87	89	Pro	CCT	65	38
	GAC	13	11		CCC	7	8
					CCA	27	46
Cys	TGT	89	88		CCG	1	8
	TGC	11	13				
				Ser	TCT	44	38
Gln	CAA	94	76		TCC	4	9
	CAG	6	24		TCA	21	16
					TCG	0	2
Glu	GAA	86	92		AGT	30	29
	GAG	14	8		AGC	0	5
Gly	GGT	52	33	Thr	ACT	48	45
	GGC	3	0		ACC	4	7
	GGA	41	64		ACA	46	48
	GGG	3	4		ACG	2	0
His	CAT	90	81	Trp	TGG	100	100
	CAC	10	19				
				Tyr	TAT	88	77
Ile	ATT	40	14		TAC	12	23
	ATC	3	1				
	ATA	57	85	Val	GTT	48	39
					GTC	6	2
					GTA	38	57
					GTG	8	2

Percentage codon usage. Average codon usage of the four open reading frames of plasmid k_1 are shown, compared with codon usage of ORF 974.
aa: Amino acid.

The complete sequence of this region of k_2 has been deposited within the EMBL database under accession number X07946. Analysis of the ORFs contained within it will be reported elsewhere (D.W.W., PhD thesis, University of

```
Homology region A
Species
& subunit
E.coli β     F E R P V T V G Y M Y M L K L N H L V D D K M H A R S T G S Y S L V T Q Q P L G
Nt chl β     F E Q P V I I G K P Y I L K L I H Q V D D K I H G R S S G H Y A L V T Q Q P L R    continued
Mp chl β     F E Q P I T I G K A Y M L K L I H Q V D D K I H A R S S G P Y A L V T Q Q P L R      below
ORF974       L E N K I I V G K C L Y L A L R H Q V D D K V Q F R N G G N I D I V T K Q P V S
             *       *   *                               * *                 *           *

E.coli β     G K A Q F G G Q R F G E M E V W A L E A Y G A A Y T L Q E M L T V K S D D V N G  1300
Nt chl β     G R A K Q G G Q R V G E M E V W A L E G F G V A H I L Q E M L T Y K S D H I R A  1015
Mp chl β     G R S R R G G Q R V G E M E V W A L E G F G V A Y I L Q E M L T I K S D H I R A  1010
ORF974       G R K R S G G L R F G Q M E R D I L I G L G A W N T I K E I W S I D R S K I K I  552
                   *               *                 *           *     *   * *           *

Homology region II
Dm Pol II    L K G K E G R I R G N L M G K R V D F S A R T V I T P D P N L R I D Q V G V P R S I A Q  377
Sc Pol III   L K G K Q G R F R G N L S G K R V D F S G R T V I S P D P N L S I D E V A V P D R V A K  400
Sc Pol II    L K G K E G R I R G N L M G K R V D F S A R T V I S G D P N L E L D Q V G V P K S I A K  372
E.coli β'    I K G K Q G R F R Q N L L G K R V D Y S G R S V I T V G P Y L R L H Q C G L P K K M A L  374
Mp chl β'    I E G K E G R F R E N L L G K R V D Y S G R S V I V V G P F L P L H Q C G L P R E M A I  403
ORF974       L K S K N G A Y H T L V E G H R V D R C I R S V I V P D P T L D I D T I K I P F G A N I  802
                     *     * *     *                                           *   *           *

Homology region III
Sc Pol III   V V L F N R Q P S L H R L S I L S H Y A K I R P W R T F R L N E C V C T P Y N A D F D G D E M N L H V P Q T E E A R A E  530
Sc Pol II    P V L F N R Q P S L H K M S M M A H R V K V I P Y S T F R L N L S V T S P Y N A D F D G D E M N L H V P Q S E E T R A E  500
E.coli β'    P V L L N R A P T L H R L G I Q A F E P V L I E G K A I Q L H P L V C A A Y N A D F D G D Q M A V H V P L T L E A Q L E  479
Mp chl β'    P I L L N R A P T L H R L G I Q A F Q P I L V N G R A I H L H P L V C G G F N A D F D G D Q M A V H I P L S L E A Q A E  508
ORF974       Y G L L N R Q P S L N V D S I K L V K L K Q G S N K T I A I N P L L C Q S F N A D F D G D E M N I Y G I R N K E S I E E  865
                                 *                             *       *       *       *                     *       *
```

Fig. 4. Homologies between product of ORF 974 and several RNA polymerase subunits. Origin of subunits abbreviated as follows: Nt chl, Nicotiana tabacum chloroplast. Mp chl, Marchantia polymorpha chloroplast. Sc, Saccharomyces cerevisae. Dm, Drosophila melanogaster. Numbers refer to location of final residue of homology region within corresponding polypeptide. Boxes and asterisks, respectively, indicate where a residue of the ORF 974 product is identical to, or conservatively different from, corresponding amino acids within two or more aligned RNA polymerase subunits. By this criterion, region A of ORF 974 product has 36.3 % identity with region A of the other polypeptides. This rises to 53.8 % if conservative substitutions are also taken into account. Region II shows 43.2 % identity rising to 56.8 %. Region III shows 48.3 % identity rising to 58.3 %.

Leicester). Using the algorithm of Lipman and Pearson[26] to search the National Biomedical Research Foundation (NBRF) protein sequence database, we have found significant homologies between the predicted ORF 974 product and the subunits of several DNA-directed RNA polymerases. The sequence of the predicted ORF 974 product is shown in figure 3. The codon usage of ORF 974 is shown in table 2.

Computer-assisted comparison of the predicted protein sequences of several RNA polymerase subunits with the ORF 974 product reveals an 80 amino acid region of homology (which we have termed region A) in common with the β subunits of RNA polymerases from E.coli[27], the chloroplasts of tobacco Nicotiana tabacum[28] and the chloroplasts of the liverwort Marchantia polymorpha[29]. This region is also homologous to one of nine conserved regions shared between the β-subunit of *E.coli* RNA polymerase and the 140 kD subunit of *S.cerevisiae* RNA polymerase II[30]. Additionally there are two other regions, of 43 amino acids and 59 amino acids, with homology to regions of the E.coli RNA polymerase β' subunit[31] the M.polymorpha chloroplast RNA

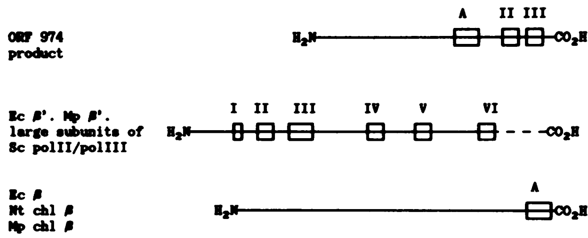

Fig. 5. Structure of predicted ORF 974 product and several RNA polymerase subunits. Abbreviations as in Figure 4 legend. Homology regions A and I to VI are boxed. Dotted line; carboxy terminal repeat region of Sc Pol II[32].

polymerase β' subunit[29], the 215 kD subunits of S.cerevisiae32 and Drosophila melanogaster[33] RNA polymerase II and the 160 Kd subunit of S.cerevisiae RNA polymerase III[32]. These latter two regions of homology correspond to the RNA polymerase large subunit conserved regions II and III defined by Allison and coworkers[32]. See Figure 4. However regions A, II and III are arranged differently within the ORF 974 product, as is shown in Figure 5.

We have been unable to detect any homology between the ORF 974 product and regions I,IV,V and VI common to S.cerevisiae RNA Pol II, Pol III and E.coli subunit β'. We have also failed to detect homology between the amino terminal 570 residues of the ORF 974 product and other RNA polymerase subunits. The ORF 974 product bears no primary sequence homology with the mitochondrial RNA polymerase of S.cerevisiae[34]. This enzyme resembles the RNA polymerases of bacteriophage T3 and T7, and the predicted product of an ORF within the mitochondrial linear plasmid S2 of Zea mays[35]. That the ORF 974 product is unrelated to this class of polymerases is consistent with an extramitochondrial location for the killer plasmids, as would be expected from their maintenance and expression within ρ^o strains of S.cerevisiae, and their capacity to encode proteins which enter the secretory pathway.

The apparent cytoplasmic location of k_1 and k_2 and their lack of recognisable expression signals implies that an RNA polymerase of novel properties may be required to transcribe their genes. We suggest that the ORF 974 polypeptide has a role in this task, perhaps via recognition of the UCS elements found 5' to all k_1 and k_2 genes. There may of course be additional plasmid-encoded or cellular proteins required for the process of transcription. In this regard we have detected homology between the product of a second k_2 ORF, ORF 579 (which is able to encode a polypeptide of 579 amino acid residues), and proteins D-569 and C-637 of the poxvirus vaccinia.

Vaccinia is a cytoplasmic DNA virus which carries its own gene expression system, and product D-569 has been identified as the DNA-dependent ATPase[36] which may unwind the vaccinia duplex during transcription[37]. Perhaps the product of ORF 579 performs a similar function for the killer plasmids. Like ORF 974, ORF 579 is preceded by a UCS-like sequence and its predicted initiation codon lies in an AAAATG context (figure 2). We note that a consequence of our interpretation of this data is that ORF 974 and ORF 579 products would be required for their own expression.

The roles performed by the various subunits of prokaryotic or eukaryotic RNA polymerases are poorly understood, although there is some evidence to suggest that the β' subunit of *E.coli* RNA polymerase may have a role in DNA binding[38] whilst the β subunit may bind nucleotides[39]. Because the product of ORF 974 shares homology with some of the regions found in both types of subunit, its structure is unique. Comparison of its biological activity with known RNA polymerases of conventional structure will lead to an increased understanding of the roles of each subunit and the conserved regions within them.

ACKNOWLEDGEMENTS

We thank Michael Romanos and Alan Boyd for communicating ORF2 transcript data prior to publication, and Michael Stark and Michael Pocklington for advice concerning interpretation of homology data. Plasmid pPA13 was a gift from Jennifer Richardson. This work was supported by the core research programme of the Leicester Biocentre, and the award of an SERC studentship to D.W.W.

+Present address: Department of Molecular Biology, Lewis Thomas Laboratory, Princeton, NJ 08544, USA

*To whom correspondence should be addressed

REFERENCES

1. Gunge, N., Tamaru, A., Ozawa, F., Sakaguchi, K. (1981). J. Bact. 145, 382-390
2. Stark, M.J.R., Boyd, A. (1986). EMBO J. 5, 1995-2002
3. Challberg, M.D., Kelly, T.J. (1979). J. Mol. Biol. 135, 999-1012
4. Garcia, J., Penalva, M., Blanco, L., Salas, M. (1984). Proc. Natl. Acad. Sci. USA. 81, 80-84
5. Hirochika, H., Sakaguchi, K. (1982). Plasmid 7, 59-65
6. Kemble, R.J., Thompson, R.D. (1982). Nucl. Acids. Res. 10, 8181-8190
7. Sor, F., Wesolowski, M., Fukuhara, H. (1983). Nucl. Acids. Res. 11, 5037-5044

8. Kikuchi, Y., Hirai, K., Hishinuma, F. (1984). Nucl. Acids. Res. 12, 5685-5692
9. Stark, M.J.R., Mileham, A.J., Romanos, M.A., Boyd, A. (1984). Nucl. Acids. Res. 12, 6011-6030
10. Sor, F., Fukuhara, H. (1985). Curr. Genet. 9, 147-155
11. Jung, G., Leavitt, M.C., Ito, J. (1987). Nucl. Acids. Res. 15, 9088
12. Gunge, N., Sakaguchi, K. (1981). J. Bact. 147, 155-160
13. Gunge, N., Murata, K., Sakaguchi, K. (1982). J. Bact. 151, 462-464
14. Gunge, N., Yamane, C. (1984). J. Bact. 159, 533-539
15. Stam, J.C., Kwakman, J., Meijer, M., Stuitje, A.R. (1986). Nucl. Acids. Res. 14, 6871-6884
16. Niwa, O., Sakaguchi, K., Gunge, N. (1981). J. Bact. 148, 988-990
17. Gough, J.A., Murray, N.E. (1983). J. Mol. Biol. 166, 1-19
18. Maniatis, T., Fritsch, E.F., Sambrook, J. (1982). Molecular Cloning. A Laboratory Manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York
19. Sanger, F., Nicklen, S., Coulson, A.R. (1977). Proc. Natl. Acad. Sci. USA. 74, 5463-5476
20. Tschopp, J.F., Emr, S.D., Field, C., Schekman, R. (1986). J. Bact. 166, 313-318
21. Struhl, K., Stinchcomb, D.T., Scherer, S., Davis, R.W. (1979). Proc. Natl. Acad. Sci. USA. 76, 1035-1039
22. Thompson, A., Oliver S.G. (1986). Yeast. 2, 179-191
23. Ito, H., Fukuda, Y., Murata, K., Kimura, A. (1983). J. Bact. 153, 163-168
24. Henikoff, S. (1984). Gene 28, 351-359
25. Devereux J., Haeberli, P., Smithies, O. (1984). Nucl. Acids. Res. 12, 387-395
26. Lipman, D.J., Pearson, W.R. (1985). Science 227, 1435-1441
27. Ovchinnikov, Y.A., Monastyrskaya, G.S., Gubanov, V.V., Guryev, S.O., Chertov, O.Y., Modyanov, N.N., Grinkevich, V.A., Makarova, I.A., Marchenko, T.V., Polovnikova, I.N., Lipkin, V.M., Sverdlov, E.D., (1981). Eur. J. Biochem. 116, 621-629
28. Shinozaki, K., Ohme, M., Tanaka, M., Wakasugi, T., Hayashida, N., Matsubayashi, T., Zaita, N., Chunwongse, J., Obokata, J., Yamaguchi-Shinozaki, K., Ohto, C., Torazawa, K., Meng, B.Y., Sugita, M., Deno, H., Kamogashira, T., Yamada, K., Kusuda, J., Takaiwa, F., Kato, A., Tohdoh, N., Shimada, H., Sugiura, M. (1986). EMBO J. 5, 2043-2049
29. Ohyama, K., Fukuzawa, H., Kohchi, T., Shirai, H., Sano, T., Sano, S., Umesona, K., Shiki, Y., Takeuchi, M., Chang, Z., Aota, S., Inokuchi, H., Ozeki, H. (1986). Nature 322, 572-574
30. Sweetser, D., Nonet, M., Young, R.A. (1987). Proc. Natl. Acad. Sci. USA. 84, 1192-1196
31. Ovchinnikov, Y.A., Monastyrskaya, G.S., Gubanov, V.V., Guryev, S.O., Salomatina, I.S., Shuvaeva, T.M., Lipkin, V.M., Sverdlov, E.D. (1982). Nucl. Acids. Res. 10, 4035-4044
32. Allison, L.A., Moyle, M., Shales, M., Ingles, C.J. (1985). Cell 42, 599-610
33. Biggs, J., Searles, L.L., Greenleaf, A.L. (1985) Cell 42, 611-621
34. Masters, B.S., Stohl L.L., Clayton D.A. (1987). Cell 51, 89-99
35. Kuzmin, E.V., Levchenko I.V., Zaitseva G.N. (1988). Nucl. Acids. Res. 16, 4177
36. Broyles, S.S., Moss, B. (1987). J. Virol. 61, 1738-1742
37. Broyles, S.S., Moss, B. (1987). Molec. Cell. Biol. 7, 7-14
38. Fukuda, R., Ishihama, A. (1974). J. Mol. Biol. 87, 523-540

39. Armstrong, V.W., Sternbach, H., Eckstein, F. (1976). Biochemistry 15, 2086-2091

NOTE ADDED IN PROOF

Since submission of this manuscript Tommasino *et al*. published the complete nucleotide sequence of plasmid k_2 (Nucl.Acids.Res. 16, 5863-5878). Their predictions concerning the genetic organisation of k_2 ORFs are in good agreement with our own, except that they have found ORFs 112 and 336 to be joined into a single ORF of 453 amino acids.